

ADVANCES IN MODERN ECOLOGY AND EVOLUTIONARY BIOLOGY

TRANSPOSABLE ELEMENT CONTENT IN YELLOW STARHISTLE

By

Adelina Isabella Lilani Lane

---

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree  
With Honors in

Ecology and Evolutionary Biology

THE UNIVERSITY OF ARIZONA

MAY 2018

Approved by:

---

Dr. Katrina Dlugosch  
Department of Ecology and Evolutionary Biology

## Abstract

Species invasions can be influenced by various ecological and genomic factors. One contribution that may influence the ability for new variation to arise comes from transposable element (TE) activity. TEs may provide organisms with favorable novel genomic mutations that allow for adaptation. This study investigates the contribution of TE content in *Centaurea solstitialis*, commonly known as yellow starthistle (YST). YST is known for its invasive ability in the Americas, where it has established large invaded ranges. In this study, seeds from native and invaded populations were reared in a common environment and traits involving growth, reproduction, and genome size were measured. Correlations between genome size and growth rate, dry aboveground biomass, number of flowering heads, and days to first start flowering were tested. In addition bioinformatic analysis of genomic data from a native and invading YST were used to estimate average number and average length of long terminal repeat (LTR) sequences. Results from genomic analysis show substantial amounts of intraspecific size variation in genomic content, commonly attributed by class I TE activity, as well as larger genome size being positively correlated with longer generation times. These results demonstrate that TE activity contributing to genome size correlates with important ecological traits.

## Introduction and Background

An invasive species is one that has colonized a novel environment and grows to high abundance, commonly crowding out native species and reducing native populations (Beck *et al.* 2008). Understanding mechanisms that enable an introduced species to become invasive is critical to their successful control and management. In particular YST has evolved increased ability to use resources in its invaded ranges, which aids its ability to flourish in its established non-native regions (Dlugosch *et al.* 2015). YST is a thistle plant that originated from a region in the Mediterranean and has now colonized regions in both North and South America (Barker *et al.* 2017). A notable feature is that YST's native range and invaded range populations include substantially different phenotypes, even when grown in to the same environmental conditions (Eriksen *et al.* 2012; Dlugosch *et al.* 2015). Through common garden experiments it has been shown that plants from the invaded range are much larger, develops faster and have a greater number of flowering heads (Eriksen *et al.* 2012; Dlugosch *et al.* 2015). Faster growth rates as well as less time required to reproduction have classically been described as traits of the "ideal weed" (Baker, 1965). However, the genetic basis of these differences between native and invasive genotypes remains unclear. Genomic analysis and comparison between native and invasive range populations may help explain the differences in phenotype as well as invasiveness.

## Study System

YST is a species of angiosperm that has been introduced to the Americas from a population that likely originated in Western Europe (Barker *et al.* 2017). Genomic analysis supports the likelihood of YST originating near Turkey, somewhere in the eastern Mediterranean (Barker *et al.* 2017). According to this analysis the European native population has proliferated and spread from their original Turkey region (Barker *et al.* 2017). It is understood that this proliferation swept all the way throughout Western Europe as well as

parts of Eurasia (Barker *et al.* 2017). In its native range YST is observed to grow at low density, at around a density of 5 plants/m<sup>2</sup> (Andonian *et al.* 2011). It is likely that alfalfa seed contamination is what brought YST to the Americas in 1848 (Gerlach, 1997). YST has since developed ranges throughout various parts of Western North America and South America. It continues to expand its ranges and is estimated to range across 14 million acres in California alone (Pitcairn *et al.* 2006). The invasions in California and Argentina are incredibly invasive with large and very dense patches often above densities of 200 plants/m<sup>2</sup> (Hierro *et al.* 2006). However, not all YST invasions are the same. The invasion in Chile, for instance, is slow spreading, grows in smaller patches, and does not grow as densely as the invasion in California (Andonian *et al.* 2011). In Chile the density of patches grow around 20 plants/m<sup>2</sup> (Andonian *et al.* 2011). As evident by its densities and spread through the Americas, YST's invasiveness varies quite substantially and its invasions in North and parts of South America are quite different.

### *Transposable Elements*

When looking at species invasion mechanisms there are noteworthy genetic factors to consider. Adaptation to new environments requires genetic variation on which natural selection can act. Studies show that successful species invasions generally consist of minimal reduction in genetic diversity, suggesting multiple introductions and large founding populations are important in establishing in new invasions (Zayed, 2007). Introduced populations may undergo genetic bottlenecks, as new populations can be founded by very few individuals (Zayed, 2007). However, there is currently no evidence of the YST invasions in America being subject to reduced genetic diversity. YST was likely introduced in large numbers, due to its historically frequent contamination of alfalfa seed shipments (Pitcairn, 2006) and is an obligate outcrosser (Maddox *et al.*, 1996), which allowed it to maintain high variation across its populations. Furthermore, invasive species often have a remarkable ability to adapt to novel environments, regardless of small introduction sizes (Frankham, 2004). Genetic variation is important when considering ability to response to environmental pressures (Stapley *et al.* 2015). Ultimately mutation is the source for all genetic variation and includes mechanisms with adaptive potential, such as whole genome duplications or point mutations (Casacuberta, 2013). Another potential source of new variation, in introduced populations, is transposable element (TE) activity (Schrader, 2014; Stapley *et al.* 2015).

Transposable elements are short repeating sequences of DNA that have the ability to copy or cut and reinsert themselves into the genome (Wicker *et al.* 2007). They represent a diverse group of genetic elements that can be classified into two broad classes of TEs (Wicker *et al.* 2007). Class I is composed of retrotransposons or TEs that have the ability to copy and paste themselves throughout the genome via an RNA intermediate (Wicker *et al.* 2007). It is due to this capacity that genomic size differences are generally attributed to class I TE proliferation (Wicker *et al.* 2007). Class II TEs, on the other hand, are composed of DNA transposons that use a cut and paste method to move via a DNA intermediate (Wicker *et al.* 2007). Each class of TE can further be classified by order and family, which are defined by overall organization, specific insertion mechanism and enzymology (Wicker *et al.* 2007).

Across the flowering plants, TEs have been particularly influential in determining genome size. Flowering plant genome sizes (the total nuclear DNA content of a cell) vary across three orders of magnitude, even though the number of genes is relatively conserved, and much

of this is known to be mediated by variation in TE content (Tenaillon *et al.* 2010). Species with small genome sizes, such as the model organism *Arabidopsis thaliana*, have relatively low TE content, while larger genomes found in maize and barley have upwards of 80% of their genomes comprised of TEs (Tenaillon *et al.* 2010). Furthermore, different families of TEs contribute differently to genome size expansions (Hawkins, 2006). There are many different types of TEs, but one important order is known as long terminal repeats (LTRs). This particular class is a retrotransposon, or class I, TE (Hawkins, 2006). This means it is capable of making copies of itself and inserting them throughout the genome. Additionally LTRs are among the longest types of TEs, which means they likely contribute to the majority of genome size variation (Hawkins, 2006). For example, examinations of the maize genome have found 33-62% of the genome was composed of LTR sequences (SanMiguel, 1998) with a 22% difference in genome size between individuals, the majority of which was due to variation in LTR copy number (Vielle-Calzada *et al.* 2009). In this way, TE activity may have indirect ecological consequences, as differences in genome size affect traits related to growth and reproduction. Specifically, larger genome sizes may incur higher replicative costs and require larger cell sizes, with concomitant effects of increasing time for reproduction, growth rates, (Francis, 2008) as well as increased time required to diversify (Knight, 2005). In fact, genome contraction, or reduction in genome size, has the ability to facilitate rapid evolutionary changes in invasive plants that may be significant in phenotypic evolution (Lavergne, 2010).

Specific TE activity may have the potential to alter plant phenotypes (Wei & Cao, 2016). Within a genome, TEs are generally thought to have deleterious effects, in part, due to their ability to insert themselves into genes vital for organismal function (Keightley & Eyre, 1999). Similarly, certain retroviruses are observed to have similar ability to disrupt genes (Boeke & Stoye, 1997). Not surprisingly class II TEs, otherwise known as retrotransposons, likely share a common evolutionary origin with retroviruses (Boeke & Stoye, 1997). As such, host species have evolved mechanisms to regulate and mitigate TE proliferation (Wei & Cao, 2016). Host defense mechanisms, also known as immune mechanisms, against TEs are not well understood, but are believed to include DNA removal through unequal-intra-strand homologous recombination (Tenaillon *et al.* 2010) as well as epigenetic silencing (Cui & Cao, 2014; Lisch, 2009; Tenaillon *et al.* 2010). Although these mechanisms can suppress TE activity they are sensitive towards environmental and genomic stressors (Stapley *et al.* 2015). Host stressors may affect TEs in such a way that these genomic areas are less regulated and therefore more able to proliferate (Negi, 2016). Stressors include factors like low genetic diversity (Stapley *et al.* 2015), hybridization (Ellstrans & Schierenbeck 2000; O'Neill *et al.* 1998;), inbreeding (Vergeer *et al.* 2012), novel pathogens (Grandbastien, 2005; Wessler, 1996), and abiotic stress such as temperature changes (Gonzalez, 2010; Vieira *et al.* 1998). Up until recently it was assumed TE proliferation was simply a drawback of overwhelming stress. However, recent data suggests there may be an adaptive advantage for regulating TE activity under certain amounts of stress (Stapley *et al.* 2015). TE activity, like other mutational processes, has the potential to create novel genetic variation, facilitating adaptation (Stapley *et al.* 2015). This capacity for adaptation is particularly important when looking at novel species introductions and could potentially provide a mechanism to explain the invasive species genetic paradox.

YST is known for its invasive ability, but the mechanism of its invasiveness is still unclear. An important finding is that intraspecific size variation has been found in YST (Miskella, 2014).

The YST genome sizes were found to range between 1.50-1.76 pg with  $\sigma = 0.07$  (Miskella, 2014). This is significant, because differences in genome size, within a ploidy level, are often attributed to differential TE activity (Stapley *et al.* 2015). Therefore, the size variation found in the YST genome point toward the potential for a differential TE activity mechanism. TE activity may also explain the invasiveness of the species invasion throughout the Americas. Genomic analysis of the invaded range and native range populations TE content should shed light on this particular mechanism as an explanation for YST invasiveness in the Americas.

In this study seeds were collected from 19 different populations, including 4 invaded and 15 native locations, and individuals from each population were grown in a common garden greenhouse. Common garden measurements of significant traits included growth rate, dry biomass, flower number, and first days to flower. Estimates of individuals' genome size were made using flow cytometry. Additionally two draft genomes were analyzed, via a bioinformatics program, to determine estimates of the number and average length of LTR sequences found in each.



Figure 1. Native range of YST shown in yellow and invaded ranges shown in red. Figure by KM Dlugosch, after Gerlach, 1997.

## Methods

### *Common Garden*

Common garden data were collected and analyzed to compare phenotypic changes between populations from different ranges. The common garden included 19 different populations of YST from across both native and invaded ranges (Table 1). The native range is represented by four western European populations that are confirmed to be part of the source invasion in North and South America. The invaded ranges include fourteen populations from California including coastal, central valley and the leading edge regions. The coastal regions represent the first site of invasion, the central valley regions have the most severe YST invasion, and the leading edge regions are where the populations are spreading to new areas. Additionally data were collected for an invading population from Argentina, which represents a replicate invasion from Western Europe. Measurements were collected from 4-34 individuals from each population amounting to a total of n=457 individuals. There were n=424 invaded range individuals and n=33 native range individuals.

Common garden data were collected for each of these individuals by growing them from seeds that were collected from each of their regions mentioned above. The seeds were collected and brought over by others. The seeds were grown in the same greenhouse and therefore were exposed to similar environmental conditions. They were grown in uniform soil, water, and light conditions and maintained into senescence. Individuals from all regions were organized in blocks throughout the greenhouse to remove effects of differences within the greenhouse. They were planted in December 2016 and harvested by September and October of 2017. The data collected included flowering time, flowering number, and aboveground biomass. Linear mixed effect models were used to test for significant correlations between genome size and these variables, with fixed effects of range (invaded or native), genome size and population and random effects of block. All statistical analyses were conducted in R v3.3.1 (R core team, 2016), using the R software packages *lmerTest* (Kuznetsova *et al.* 2017) to run linear mixed effects models and *lsmeans* (Lenth 2017).

Table 1. Individuals grown in the common garden based on the populations, regions and ranges from which seeds were collected.

Range	Region	Population	Individual Count
Invaded	Coast	Diablo	30
		Gilroy	30
		Napa	29
		Marin	28
	Central Valley	Red Bluff	30

		Clovis	24
		Triangle	17
	Sierras	Sanflat	34
		Arnold	33
		Belden	30
		Vet	30
		Placerville	29
		Quincy	28
		Lyons	22
	Argentina	Argentina	30
Native	Europe	Salamanca	19
		Canales	5
		Cuenca	5
		Granada	4

### *Genome size estimation*

Genomic size data were obtained to quantify variation in genomic content in relation to phenotypic variances. Genome size data were collected and estimated by flow cytometry using a FACSCanto II instrument (BD Biosciences, San Jose, CA, USA) equipped with a blue (488-nm), air-cooled, 20-mW solid state laser and a red (633-nm) 17-mW helium neon laser for UV excitation. Flow cytometry is a separation technique that uses fluorescent tags for comparison markers (Galbraith, 1983). Propidium iodide was the fluorescent stain used for genomic comparison. Radish was used as the internal standard, since radish genomes are already well characterized (Mitsui, 2015). Genome size estimates were obtained for n=391 individuals. Additionally an Otto buffer was used for better characterization of genomic results.

### Quantifying LTR Content

Analysis of TE content was obtained from two YST individuals that did not come from the common garden. One individual came from the invaded range in California (Triangle; Table 1) and the other individual from a native range population in Turkey (TK23). A draft genome was obtained for both individuals (KM Dlugosch, unpublished data). These draft genomes were subsequently analyzed for TE content via the bioinformatics program 'LTR\_finder' (Xu & Wang, 2007). LTR\_finder is a program that finds *de novo* LTR sequences. LTRs are among the longest types of TEs and likely contribute the most to size variation within a ploidy level (Hawkins, 2006) The program finds candidate LTR sequences through its ability to detect certain LTR regions, target site repeat (TSR) regions, primer binding site (PBS) regions, polypurine tract (PPT) regions as well as specific protein domains. An initial candidate is found by identifying 5'LTR and 3'LTR regions, which should be somewhat homologous in LTRs (Xu & Wang, 2007). These two regions of the retrotransposon should be identical when the element first inserts itself into a host genome and will subsequently begin to evolve (Xu & Wang, 2007). The 5' end of the 5'LTR will typically have a region called the 'TG..CG box', while the 3' end of the 3'LTR region will often have a 'CA' (Xu & Wang, 2007). Once a candidate is found the program will subsequently analyze for PBS, PPT as well as RT. PBS is a region found close to the 3' end of the 5'LTR (Xu & Wang, 2007). This region is important as it binds tRNA and initiates the process of reverse transcription, which is vital for its copy and paste mechanism (Xu & Wang, 2007). Two different tRNA databases were used to identify PBS regions. These databases were from *Arabidopsis thaliana* and *Triticum sp.* plants (Chan & Lowe, 2008). PPT is another such region important for reverse transcription (Xu & Wang, 2007). Additionally the program scans to see if specific protein domains are found within the LTR boundaries. The protein domains that the program looks for are meant to code for enzymes important in reverse transcription and insertion of the transposon back into the host genome (Xu & Wang, 2007). The program will subsequently provide a confidence level depending on how many of these sequences it's able to identify in the LTR candidate (Xu & Wang, 2007).

### Results

A linear mixed effects model demonstrated geographic origin had a significant effect on growth ( $p=0.044$ ), where invasive genotypes exhibit faster early growth rate (least squares means (LSM) =  $13.56 \pm 0.82$ ) than native genotypes ( $7.775 \pm 1.658$ ) (Fig. 2). Similarly, invaders grew significantly larger over their lifetimes (invasive =  $3344.77 \pm 115.27$ , native =  $2471.74 \pm 252.08$ ,  $p=0.003$ ) (Fig. 3). Linear mixed effects models also demonstrated a marginally significant effect of geographic origin on YST reproduction. Invasive genotypes required fewer days to first flower (invasive =  $193.30 \pm 3.92$ , native =  $198.38 \pm 8.40$ ,  $p=0.065$ ), resulting in greater number of flowering heads produced over their lifetime (invasive =  $8.023 \pm 0.447$ , native =  $6.398 \pm 0.917$ ,  $p=0.09$ ) (Fig. 4).

Mean genome size for the native and invaded ranges were relatively similar (Table 2). Genome size does not have a significant relationship with growth rate, dry above ground biomass or number of flowering heads. Genome size had no significant relationship with growth rate ( $p=0.73$ ). Genome size had no significant relationship with dry aboveground biomass ( $p=0.51$ ). Genome size had no significant relationship with flowering heads ( $p=0.53$ ).

However, there is a weakly positive, but significant effect of genome size on the days to first flower ( $p=0.04$ ), in which individuals with larger genome sizes flower later (Figure 5).

Table 2. Means and standard deviations of genome sizes, days to flower, number of flowers and dry biomass for invaded and native ranges.

Means and Standard Deviations				
Range	Genome size (pg)	Days to flower	Flower number	Dry biomass
Invaded	1.722 +/-0.045	191.140 +/-30.049	7.899 +/-3.716	3322.8 +/-1003.2
Native	1.725 +/-0.041	198.133 +/-25.454	7.125 +/-4.316	2502.8 +/-719.8

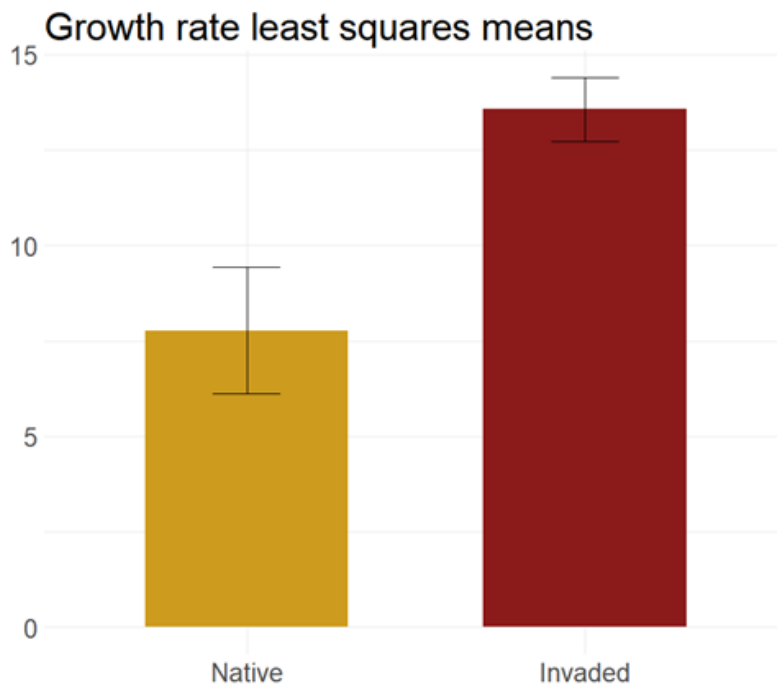


Figure 2. Least squares means (+/- standard error) of daily increase in a linear size index (ln mm) from a mixed effects model of growth rate differences between invaded and native ranges. Invasive genotypes grew significantly faster ( $p=0.044$ ).

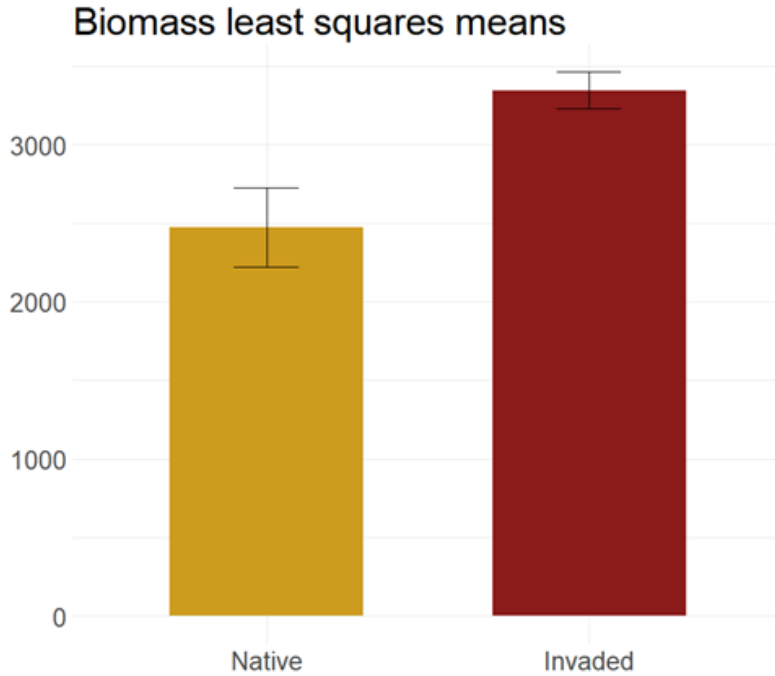


Figure 3. Least squares means ( $\pm$  standard error) from a mixed effects model of dry aboveground biomass (mg) differences between invaded and native ranges. Invasive genotypes grew significantly larger over their life time ( $p=0.003$ ).

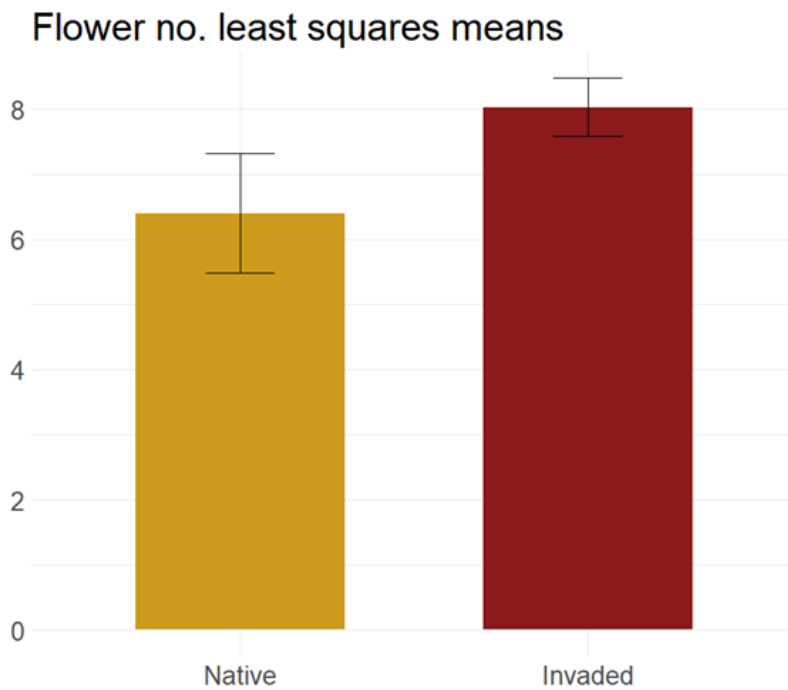


Figure 4. Least squares means ( $\pm$  standard error) from a mixed effects model of flowering number between invaded and native ranges. Invasive genotypes produced more flowering heads over their lifetimes, with marginally significant difference between ranges ( $p=0.09$ ).

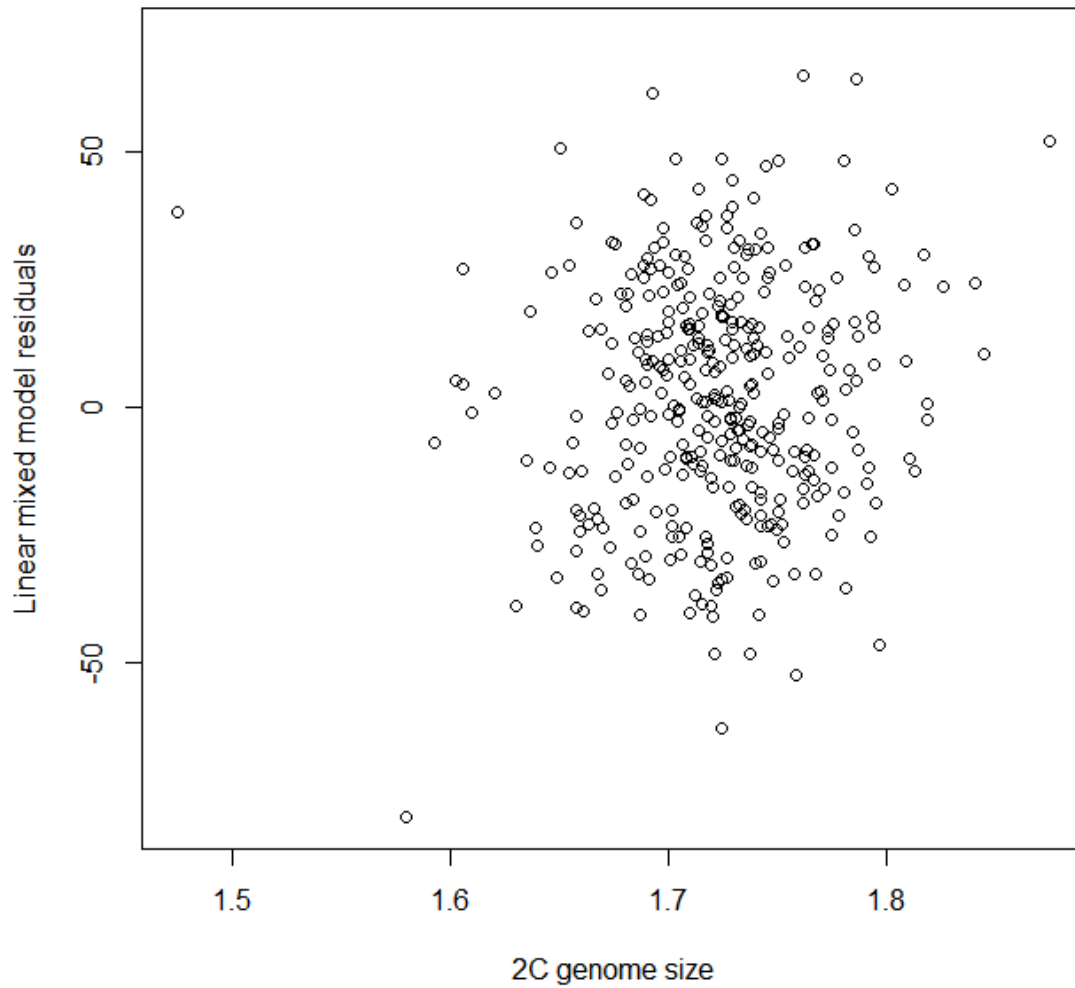


Figure 5. Residuals of linear mixed effects model of flowering time (days to first flower head) against estimated genome sizes. Genome size has a weakly positive, but significant effect on days to first flower ( $p=0.042$ ).

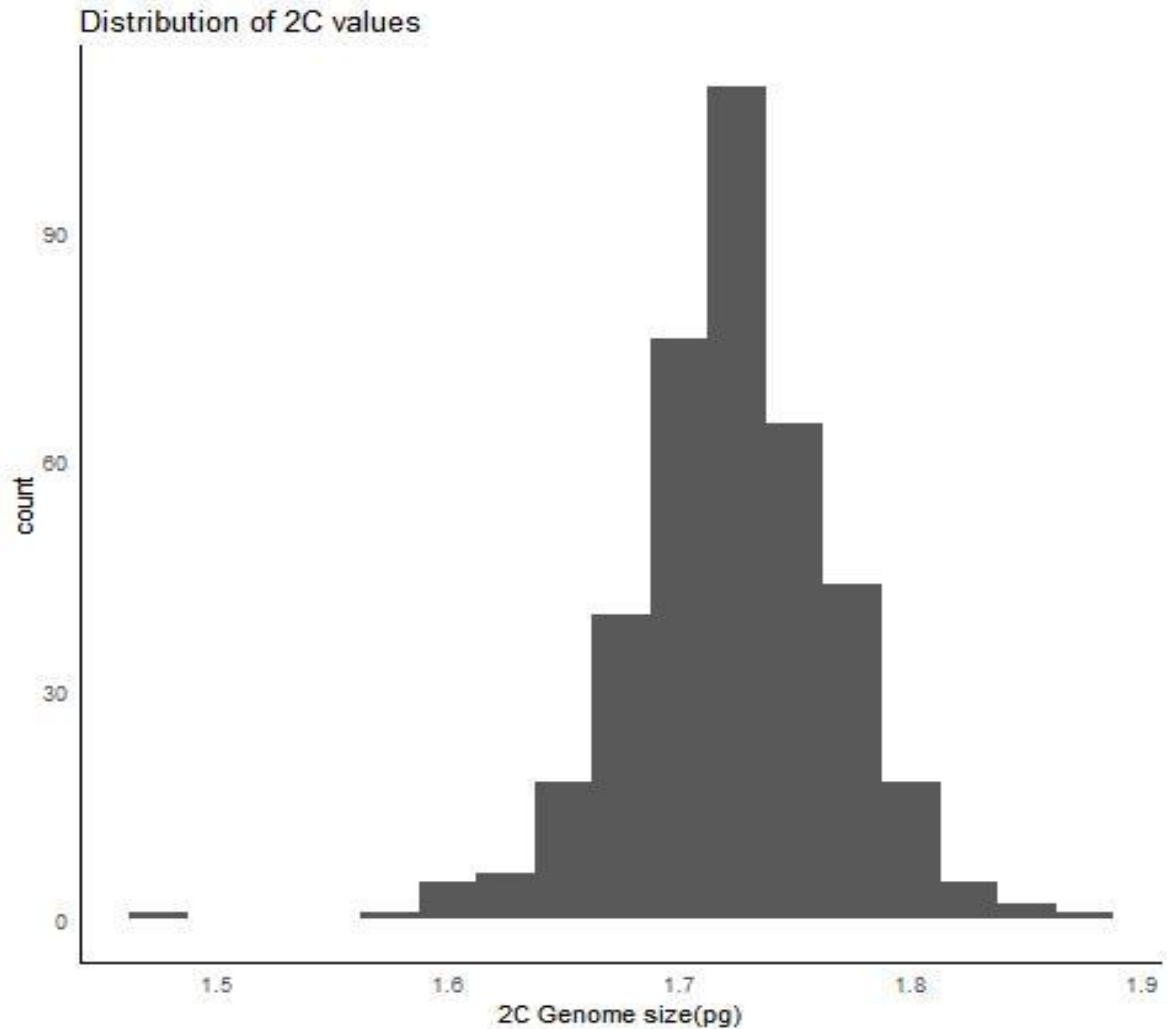


Figure 6. Overall 2C genome size variation of both invaded and native range YST genotypes in picograms. Genome size estimates were obtained for n=391 individuals and vary between 1.47-1.87 pg, demonstrating +/-23% variation around the mean.

Table 3. The number of LTRs found as well as average length for the invaded range and native range using comparisons to the tRNA databases of *Arabidopsis thaliana* and *Triticum sp.*. This data shows greater number and greater average length of LTRs for the invaded individual in both database comparisons. N50 refers to the contig length found at 50% of the total assembly lengths (Videvall, 2017).

	LTR number		Average length (bp)		N50
	Arabidopsis	Wheat	Arabidopsis	Wheat	
Invaded	3809	4279	7337.66	7315.62	34kb
Native	599	699	4824.84	4853.35	7kb

There is a substantial range of size variation found in the YST individuals (Figure 6). The number of LTR sequences as well as their average length of sequence were different in the draft assemblies from the native range and invaded range individuals (Table 3).

## Discussion

Phenotypic data from the common garden replicate previously published results that demonstrated increased growth and reproduction in the invaded range (Dlugosch *et al.* 2015). Growth in a common environment provides valuable information, since the environment is being controlled. This means that differences in phenotype can be attributed to genetic contribution and not environmental factors. Furthermore, the individuals were set up in a block approach, which should resolve any differences in environmental conditions throughout the greenhouse. Table 2 results show that genotypes from the invaded range grew larger, produced more flowers and required fewer days to begin flowering than its native counterpart. Similar work in other systems has also demonstrated shifts towards larger sizes in the invaded range, as in a common garden study of invasive *Mahonia aquifolium* and *M. repens* that grew longer stems, more leaves, and greater above ground biomass than genotypes from their native ranges (Ross and Auge, 2008). These are all factors that could aid invasive success in the invaded ranges. Of course, there are other hypotheses, regarding invasive species success, that may play roles in the locations these plants are commonly found. These hypotheses include enemy release (Keane and Crawley, 2002), as well as the evolution of increased competitive ability that states enemy release is coupled with resource reallocation between defense and growth mechanisms (Blossey and Notzold, 1995).

Of the measured traits, in this study, number of days to first flower had a significant relationship with genome size. The relationship between these two factors shows that larger genome size is positively correlated with longer minimum reproduction time. Similar results were found in a study of pines, in which genome size was positively correlated with cell size, seed size, and minimum generation time, and negatively correlated with growth rate (Grotkopp *et al.* 2004). This is consistent with the hypothesis that larger genome sizes experience higher replicative costs and therefore longer overall cell cycles (Francis, 2008). Moreover, another study found significant data that larger genome size, in maize, is negatively correlated with the rate at which cells are produced (Bilinski *et al.* preprint). This means individuals with larger maize genomes will produce new cells slower than those with smaller maize genomes. These studies all provide supporting evidence that larger genome size correlates with longer generation time.

The 2C genome size data (Figure 6) showed +/-23% variation around the mean. Observing such genome size variation indicates a basis for analyzing factors that affect genome size, such as class I TEs. The variation in Figure 6 was not significantly different between native range and invaded range populations. An interesting area of further study would be to compare the Chilean invasion and California invasion to address differences in the types of invasions seen in the Americas. If substantial genomic differences are found between the two invasions this may indicate particular factors playing a role in the extent of invasive ability in an invasion.

My survey of LTRs in available draft genomes provides preliminary evidence that the invaded range individual may have a greater amount as well as longer average length of LTRs than the native range individual. This was seen in both the *Arabidopsis thaliana* and *Triticum sp.*

database comparisons. Of course, more accurate estimates could be made if the comparison databases were more closely related to YST, such as by using a sunflower database. Additionally, at this point, there are no size estimates for the genomes for the two draft assemblies. This means that even if the LTR\_finder program finds differences in LTR, we cannot tell how they contribute to the overall genome size. Additionally, the N50 of the native range individual was shorter than the invader. LTR\_Finder discovers elements by identifying the 5' and 3' ends of individual LTRs within a given sequence, so it is highly influenced by the average contig length. This could have caused the candidate identification of the LTR software to find fewer pairs of LTRs in the TK23 assembly. Extensions of this work may include building consensus sequences and annotating them to determine families from the LTRs identified by LTR\_Finder in each individual. These consensus sequences may be mapped onto the raw sequence data and the contributions of individual LTR families to genome size could be estimated from the proportion of sequence matches. This project does not address the effect of an individual TE insertion; rather it considers the wholesale effect of overall TE content and its contribution to genome size. However, insertion of an individual element may have important phenotypic effects, depending on its site of insertion. Future work should consider differences in other types of TEs, including class I TEs or other non-LTR retrotransposons.

## Conclusions

The common garden experiment demonstrated that YST invaded range individuals grow faster, produce more flowers, and become larger overall. These are all factors that could increase the invasiveness of the invaded range. There was a significant relationship between genome size and the days it took to first begin flowering. This relationship was weakly positive and demonstrates that larger genome size is positively correlated with longer minimum reproductive time. Additionally a considerable amount of genomic size variation was observed in both the native and invaded ranges. Differences in genomic sizes may be attributed to differences in the amounts of TE content. TE content was quantified in draft genomes from two individuals; one from a native range, and one from an invaded range. Results showed preliminary data indicating a greater number as well as longer average length of LTRs in the invaded range individual. However, the quality of genome assemblies between the two individuals differ greatly and therefore comparisons of TEs between invaded and native ranges require further study. Further research may consider looking at other non-LTR transposons as well as aim to identify families of TEs and their specific contribution to genome size in YST.

## Acknowledgements

Faculty advisor Dr. Katrina Dlugosch, Graduate student Alice Cang, Undergraduate student Jenny Wong, and The University of Arizona

## Literature Cited

- Andonian, K., Hierro, J. L., Khetsuriani, L., Becerra, P., Janoyan, G., Villarreal, D., ... Callaway, R. M. (2011). Range-expanding populations of a globally introduced weed experience negative plant-soil feedbacks. *PLoS ONE*, 6(5). doi:10.1371/journal.pone.0020117
- Baker, H. G., Baker, H. G., & Stebbins, G. L. (1965). Characteristics and modes of origin of weeds. *The genetics of colonizing species.*, 147-168.

- Bancheva, S., & Greilhuber, J. (2006). Genome size in Bulgarian *Centaurea s.l.* (Asteraceae). *Plant Systematics and Evolution*, 257(1-2), 95–117. doi:10.1007/s00606-005-0384-7
- Barker, B. S., Andonian K., Swope S. M., Luster D. G., Dlugosch K. M. (2017). Population genomic analyses reveal a history of range expansion and trait evolution across the native and invaded range of yellow starthistle (*Centaurea solstitialis*). *Molecular Ecology*, Volume 26 (4), pp. 1131-1147. doi 10.1111/mec.13998
- Beck, K. G., Zimmerman, K., Scharadt, J. D., Stone, J., Lukens, R. R., Reichard, S., Randall, J., Cangelosi, A. A., Cooper, D., & Thompson J. P.. (2008). *Invasive Plant Science and Management*. pp 414-421. doi:10.1614/IPSM-08-089.1
- Bilinski, P., Albert, P. S., Berg, J. J., Birchler, J., Grote, M., Lorant, A., Quezada, J., Swarts, K., Yang, J., Ross-Ibarra, J. (2017). Parallel Altitudinal Clines Reveal Adaptive Evolution of Genome Size In *Zea mays*. Preprint. Doi: <https://doi.org/10.1101/134528>
- Blossey, B., Notzold R. (1995) Evolution of increased competitive ability in invasive non-indigenous plants: a hypothesis. *J Ecol* 83:887-889
- Boeke JD, Stoye JP (1997) Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Coffin JM, Hughes SH, Varmus H, editors. *Retroviruses*. Cold Spring Harbor (New York): Cold Spring Harbor Laboratory Press. pp. 343–435. pp.
- Casacuberta, E., and Gonzalez, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular Ecology*. 22, 1503-1517. doi:10.1111/mec.12170
- Chan, P. P., & Lowe, T. M. (2008). GtRNADB: a database of transfer RNA genes detected in genomic sequence. *Nucleic acids research*, 37(suppl\_1), D93-D97. doi: 10.1093/nar/gkn787
- Charlesworth D. & Charlesworth B. Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* 18, 237–268 (1987).
- Cui, X., and Cao, X. (2014). Epigenetic regulation and functional exaptation of transposable elements in higher plants. *Curr Opin Plant Biol* 21c, 83–88.
- Dlugosch, K. M., Cang, A. F., Barker, B. S., Andonian, K., Swope, S. M., & Rieseberg, L. H. (2015). Evolution of invasiveness through increased resource use in a vacant niche. *Nature Plants*, 1(6), 15066. doi:10.1038/nplants.2015.66
- Ellstrand, N. C., & Schierenbeck, K. a. (2000). Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America*, 97(13), 7043–50. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=34382&tool=pmcentrez&rendertype=abstract>
- Eriksen, R. L., Desronvil, T., Hierro, J. L. & Kesseli, R. (2012). Morphological differentiation in a common garden experiment among native and non-native specimens of the invasive weed yellow starthistle (*Centaurea solstitialis*). *Biol. Invasions* 14, 1459–1467.
- Eriksen, R. L., Hierro, J. L., Eren, Ö., Andonian, K., Török, K., Becerra, P. I., ... Kesseli, R. (2014). Dispersal pathways and genetic differentiation among worldwide populations of the invasive weed *Centaurea solstitialis* L. (Asteraceae). *PLoS ONE*, 9(12), 1–20. doi:10.1371/journal.pone.0114786
- Frankham, R. (2004). Resolving the genetic paradox in invasive species. *Heredity*, 94, 385.
- Francis, D., Davies M. S., Barlow, P.W. (2008). A strong nucleotypic effect on the cell cycle regardless of ploidy level. *Annals of Botany*. vol. 101(6). pp 747-757. doi:10.1093/aob/mcn038

- Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, N. M., Sharma, D. P., & Firoozabady, E. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science*, 220(4601), 1049–1051.
- Gerlach, J. (1997). How the west was lost: reconstructing the invasion dynamics of yellow starthistle and other plant invaders of western rangelands and natural areas. *Proceedings of the California Exotic Pest Plant Council*, 6. Retrieved from [http://www.cal-ipc.org/symposia/archive/pdf/1997\\_symposium\\_proceedings1937.pdf](http://www.cal-ipc.org/symposia/archive/pdf/1997_symposium_proceedings1937.pdf)
- Gonzalez J, Karasov TL, Messer PW, Petrov DA (2010) Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLOS Genetics*, 6, e1000905.
- Grandbastien, M.A., Audeon, C., Bonnivard, E., Casacuberta, J.M., Chalhour, B., Costa, A.P., Le, Q.H., Melayah, D., Petit, M., Poncet, C., Tam, S.M., van Sluys, M.A. and Mhiri, C. (2005) Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* 110:229–241.
- Grotkopp, E., Rejmanek, M., Sanderson, M. J., & Rost, T. L. (2004). Evolution of Genome Size in Pines ( Pinus ) and Its Life-History Correlates : Supertree Analyses. *Evolution*, 58(8), 1705–1729.
- Hawkins, J. S., Kim, H., Nason, J. D., Wing, R. A., & Wendel, J. F. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Research*, 16(10), pp 1252–1261. doi:10.1101/gr.5282906
- Hierro, J.L., Villareal, D., Eren, O., Graham, J.M., Callaway, R.M. (2006) Disturbance facilitates invasion: the effects are stronger abroad than at home. *The American Naturalist* 168: 144–156.
- Keane R. M. and Crawley M. J. (2002). Exotic plant invasions and the enemy release hypothesis. *Trends Ecol Evol* 17:164–170.
- Keightley, P. D., and A. Eyre-Walker, (1999). Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* 153: 515–523.
- Knight CA, Molinari NA, Petrov DA. The large genome constraint hypothesis: evolution, ecology and phenotype, *Annals of Botany* , 2005, vol. 95 (pg. 177–190)
- Kuznetsova A, Brockhoff PB and Christensen RHB (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software*, 82(13), pp. 1–26. doi: 10.18637/jss.v082.i13
- Lavergne, S., Muenke, N. J., & Molofsky, J. (2010). Genome size reduction can trigger rapid phenotypic evolution in invasive plants. *Annals of Botany*, 105(1), 109–116. doi:10.1093/aob/mcp271
- Lenth, R., and Love, J. (2017). Least-Squares Means. Version 2.27-61. [Computer Program] Retrieved from <https://cran.r-project.org/web/packages/lsmmeans/lsmmeans.pdf>
- Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* 60, 43–66.
- Maddox, D.M., Joley, D.B., Supkoff, D.M., Mayfield, A., 1996. Pollination biology of yellow starthistle (*Centaurea solstitialis*) in California. *Can. J. Bot.* 74, 262–267.
- Miskella, J. (2014). *Hybridization between Yellow Starthistle (Centaurea solstitialis) and Meadow Knapweed (Centaurea x moncktonii)*.
- Mitsui, Y. *et al.* The radish genome and comprehensive gene expression profile of tuberous root formation and development. *Sci. Rep.* 5, 10835; doi: 10.1038/srep10835 (2015).

- Negi, P., Rai, A. N., & Supreasanna, P. (2016). Moving through the Stressed Fenome: Emerging Regulatory Roles for Transposons in Plant Stress Response. *Front Plant Science*. doi: 10.3389/fpls.2016.01448
- O'Neill RJW, O'Neill MJ, Graves JAM (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*, 393, 68–72.
- Pitcairn, M. J., Schoenig, S., Yacoub, R., & Gendron, J. (2006). Yellow starthistle continues its spread in California. *California Agriculture*, 60(2), 83–90. doi:10.3733/ca.v060n02p83
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ross, C. A., and Auge, H. (2008) Invasive Mahonia plants outgrow their native relatives. *Plant Ecol* 199:21-31. doi:10.1007/s11258-008-9408-z
- SanMiguel P., Bennetzen J.L., Bennetzen J.L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot. (Lond.)* 1998;82:37–44.
- Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., ... Oettler, J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, 5, 5495. doi:10.1038/ncomms6495
- Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*, 24(9), 2241–2252. doi:10.1111/mec.13089
- Tenaillon, M. I., Hollister, J. D., & Gaut, B. S. (2010). A triptych of the evolution of plant transposable elements. *Trends in Plant Science*, 15(8), 471–478. doi:10.1016/j.tplants.2010.05.003
- Vergeer P, Wagemaker N, Ouborg NJ (2012) Evidence for an epigenetic role in inbreeding depression. *Biology Letters*, 8, 798–801.
- Videvall, E. (2017) What's N50?. *The molecular Ecologist*. Retrieved from <http://www.molecularecologist.com/2017/03/whats-n50/>
- Vieira C, Aubry P, Lepetit D, Bie mont C (1998) A temperature cline in copy number for 412 but not roo/B104 retrotransposons in populations of *Drosophila simulans*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265, 1161–1165.
- Vielle-Calzada, J.-P., Martínez de la Vega, O., Hernández-Guzmán, G., Ibarra-laclette, E., Alvarez-mejía, C., Vega-arreguín, J. C., ... Herrera-estrella, A. (2009). The Palomero Genome Suggests Metal effects on Domestication. *Science*, 326, 1078. doi:10.1126/science.1178437
- Wei, L., Cao, X. (2016). The effect of transposable elements on phenotypic variation: insights from plants to humans. *Science China Life Sciences*, Vol 59 (1), pp 24-37. doi:10.1007/s11427-015-4993-2
- Wessler, S. R. (1996). Plant retrotransposons. Turned on by stress. *Current Biology*, 6(959-961), 959–961.
- Wicker, T., Sabor, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H. (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982
- Wicker, T., Narechania, A., Sabot, F., Stein, J., Vu, G. T. H., Graner, A., ... Stein, N. (2008). Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. *BMC Genomics*, 9, 518. doi:10.1186/1471-2164-9-518

Xu, Z., & Wang, H. (2007). LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, 35(Web Server issue), W265–W268.  
<http://doi.org/10.1093/nar/gkm286>

Zayed A, Constantin SA, Packer L. Successful biological invasion despite a severe genetic load, *PLoS One* , 2007, vol. 2 pg. e868 doi:10.1371/journal.pone.0000868