BANIAGA : LTR DYNAMICS

**Nuclear Genome Size is Positively Correlated with Mean LTR Insertion Date in Fern
and Lycopphyte Genomes**

ANTHONY E. BANIAGA

Department of Ecology & Evolutionary Biology

University of Arizona

PO BOX 210088

Tucson, AZ, 85721

USA

*author for correspondence: abaniaga@email.arizona.edu

MICHAEL S. BARKER

Department of Ecology & Evolutionary Biology

University of Arizona

PO BOX 210088

Tucson, AZ, 85721

USA

msbarker@email.arizona.edu

BANIAGA : LTR DYNAMICS

ABSTRACT.—Nuclear genome size is highly variable in vascular plants. The composition of long terminal repeat retrotransposons (LTRs) is a chief mechanism of long term change in the amount of nuclear DNA. Compared to flowering plants, little is known about LTR dynamics in ferns and lycophytes. Drawing upon the availability of recently sequenced fern and lycophyte genomes we investigated these dynamics and placed them in the context of vascular plants. We found that similar to seed plants, mean LTR insertion dates were strongly correlated with haploid nuclear genome size. Fern and lycophyte species with small genomes such as those of the heterosporous *Selaginella* and members of the Salviniaceae had recent mean LTR insertion dates, whereas species with large genomes such as homosporous ferns had old mean LTR insertion dates intermediate between angiosperms and gymnosperms. This pattern holds despite methylation and life history differences in ferns and lycophytes compared to seed plants, and our results are consistent with other patterns of structural variation in fern and lycophyte genomes.

KEY WORDS.—Long Terminal Repeat Retrotransposons, Monilophytes, Pteridophytes, *Selaginella*

BANIAGA : LTR DYNAMICS

Vascular plants exhibit an astonishing diversity of genome sizes relative to other eukaryotes. Plant genomes vary more than 2,400 fold in size, from the minute genomes of *Genlisea tuberosa* [1C=0.06 Gbp (Fleischmann et al. 2014; Rivadavia et al. 2013)] to the extremely large genomes of *Paris japonica* [1C=152.2 Gbp (Pellicer et al. 2010)]. The rate of plant genome size evolution appears to be lineage specific. For example, the genomes of *Genlisea* spp. range from 0.06–1.7 Gbp, with recent reductions in the smallest genomes following recent whole genome duplication (Fleischmann et al. 2014). This stands in contrast to the small genomes of the Selaginellaceae that have a narrower range of genome size from 1C 0.082–0.184 Gbp (Baniaga et al. 2016) despite deep divergences among subgenera dating to the Permian-Triassic boundary (Arrigo et al. 2013; Kenrick and Crane 1997; Korall and Kenrick 2002; Westrand et al. 2016; Zhou et al. 2016).

The primary mechanisms driving variation in genome size in vascular plants are whole genome duplication (WGD) and the proliferation of transposable elements. Although WGDs are common and phylogenetically widespread throughout the history of vascular plants (Barker et al. 2016; Cui et al. 2006; Jiao et al. 2011; Kagale et al. 2014; Landis et al. 2018; Li et al. 2015; Li et al. 2018), genome size is often reduced following polyploidization by fractionation and diploidization (Freeling et al. 2012; Murat et al.

BANIAGA : LTR DYNAMICS

2014), partially due to species specific maximum genome size thresholds (Zenil-Ferguson *et al.* 2016). Consequently, most of the variation in genome size is caused by the expansion and contraction of Class I transposable elements, specifically Long Terminal Repeat retrotransposons (LTRs) which increase in copy number via a copy-and-paste mechanism (Feschotte *et al.* 2002). LTRs can comprise a majority of a vascular plant genome (Michael 2014), and total LTR content is positively correlated with genome size in angiosperms (Flavell 1980; Michael 2014; Tiley and Burleigh 2015; Wendel *et al.* 2016). For example, in the extremely small genome of *Utricularia gibba*, LTRs comprise only 2% of the genome (Ibarra-Laclette *et al.* 2013), but in *Zea mays* they comprise more than 75% of the genome (Schnable *et al.* 2009). These dynamics lead LTRs to be the dominant driver of genome size evolution in vascular plants.

Comparative analyses of LTR insertion dates can be used to reveal the dynamics of LTRs and their impacts on genome evolution. It is possible to estimate LTR insertion dates because LTRs have identical 5' and 3' ends at initial insertion. After insertion, they evolve neutrally and given an estimate of the synonymous substitution rate for a lineage or taxon the divergence can be used to date LTR insertion (SanMiguel *et al.* 1998). Most angiosperms have young and active LTRs inserted 0–4 mya with few full length LTRs recovered beyond 4 mya in monocots and eudicots (El Baidouri and Panaud 2013). Exceptions to this pattern are found in the genomes of *Fritillaria* spp. These plants have some of the largest known angiosperm genomes, and are characterized by the slow accumulation of diverse repeat elements without their subsequent removal (Kelly *et al.*

BANIAGA : LTR DYNAMICS

2015). Similar LTR age distributions are found in gymnosperms, which also have large genome sizes. Their genomes are characterized by numerous diverse repeats with a uniform distribution of LTR insertion times dating back to more than 60 mya (Kovach *et al.* 2010; Nystedt *et al.* 2013; Voronova *et al.* 2017; Zuccolo *et al.* 2015). The persistence of LTRs over millions of years is a common feature of the largest genomes of seed plants.

In contrast to seed plants, we know relatively little about the dynamics of LTRs and genome size evolution in the ferns and lycophytes. Fern nuclear genome size ranges from 1C=250 Mbp in the heterosporous water fern *Salvinia cucullata* (Li *et al.* 2018) to 1C=147.3 Gbp in the whisk-fern *Tmesipteris obliqua* (Hidalgo *et al.* 2017). On average, most ferns have large genome sizes with a mean of 1C=13.98 Gbp (Clark *et al.* 2016), intermediate between those of gymnosperms and angiosperms. Genome sizes in lycophytes range from the extremely small genomes of 1C=81.5 Mbp in the spike-moss *Selaginella selaginoides* (Baniaga *et al.* 2016) to 1C=11.7 Gbp in the quillwort *Isoetes lacustris* (Hanson and Leitch 2002), with an average of 1.66 Gbp for lycophytes (Leitch and Leitch 2013). Unlike seed plants, nuclear genome size in ferns and lycophytes correlates strongly with chromosome number (Bainard *et al.* 2011; Barker 2013; Clark *et al.* 2016; Leitch and Leitch 2013; Nakazato *et al.* 2008). We may thus expect that LTRs may play a lesser role in the evolution of fern and lycophyte genome size compared to seed plants.

BANIAGA : LTR DYNAMICS

Investigations of LTR dynamics in ferns and lycophytes are in their infancy, but have already revealed some important details. Generally, the proportion of the genome comprised of LTRs is within the range observed in seed plants (Banks *et al.* 2011; Li *et al.* 2018; Van Buren *et al.* 2018; Wolf *et al.* 2015; Xu *et al.* 2018). LTR insertion dates have only been estimated for two heterosporous ferns and species of *Selaginella*. In the heterosporous ferns, which have much smaller genome sizes than their homosporous relatives, a high amount of recent activity is found with relatively low amounts of genetic divergence between LTR 5' and 3' ends (Li *et al.* 2018). In *Selaginella* the majority of LTRs are also relatively young (Banks *et al.* 2011; Van Buren *et al.* 2018; Xu *et al.* 2018), and are strongly linked to the high haplotype variation observed in *S. lepidophylla* (Van Buren *et al.* 2018). The role of LTRs in influencing haplotype variation in *Selaginella* may explain the high heterozygosity and assembly difficulties of other *Selaginella* genomes despite their minute nature.

Given that chromosome number is strongly correlated with nuclear genome size in ferns and lycophytes (Bainard *et al.* 2011; Barker 2013; Clark *et al.* 2016; Nakazato *et al.* 2008; Leitch and Leitch 2013), we sought to test if genome size is also well correlated with LTR insertion activity. With the advent of newly available genome-wide data for several fern and lycophyte taxa (Banks *et al.* 2011; Li *et al.* 2018; VanBuren *et al.* 2018; Wolf *et al.* 2015; Xu *et al.* 2018), we were able to explore the LTR dynamics of ferns and lycophytes. The addition of these phylogenetically diverse genomic data allowed us to test if genome size was correlated with mean LTR insertion date in ferns and lycophytes, as

BANIAGA : LTR DYNAMICS

demonstrated previously for seed plants (Michael 2014). If genome size is correlated with mean LTR insertion date, then we predict that 1) homosporous ferns should have intermediate LTR insertion dates between angiosperms and gymnosperms, and 2) the heterosporous *Selaginella* with their extremely small genomes and the relatively small genomes of the heterosporous water ferns should have recent LTR insertion dates comparable to angiosperms.

MATERIALS AND METHODS

LTR discovery.—Genome assemblies used in analyses were downloaded from published and publicly available datasets (Li *et al.* 2018; Phytozome v11; VanBuren *et al.* 2018; Wan *et al.* 2018; Wolf *et al.* 2015; Xu *et al.* 2018). Five homosporous fern taxa (*Ceratopteris richardii*, *Cystopteris protrusa*, *Dipteris conjugata*, *Plagiogyria formosana*, *Pteridium aquilinum*), two heterosporous fern taxa (*Azolla filiculoides*, *Salvinia cucullata*), and three lycophyte taxa (*Selaginella lepidophylla*, *S. moellendorffii*, *S. tamariscina*) were analyzed.

Only *Selaginella* taxa were available to represent lycophytes. In general, the proportion of the genome comprised of LTRs in ferns and lycophytes is similar to that observed in other seed plants (Table 1). In addition to the seven fern taxa and three species of *Selaginella*, five representative angiosperms (*Amborella trichopoda*, *Sorghum bicolor*, *Arabidopsis thaliana*, *Erythranthe guttata*, *Solanum lycopersicum*) and three representative gymnosperms (*Gnetum montanum*, *Picea abies*, *Pinus sylvestris*) were included in our analysis.

BANIAGA : LTR DYNAMICS

Assembled genomes were scanned using LTR_Finder (Xu and Wang 2007) with the following parameters; “Maximum distance between LTRs=15000”, “Minimum distance between LTRs=1000”, “Maximum LTR length=3500”, “Minimum LTR length=100”, “Length of exact match pairs=15”, “Extension cutoff=0.5”, “Reliable extension=0.5”. Nested LTRs were excluded from all analyses. Predicted full length LTRs were extracted from scaffolds using custom perl scripts (https://bitbucket.org/barkerlab/baniaga_fern_lycophyte_ltrs) and blasted to RepBase (Bao *et al.* 2015) using the tblastx algorithm. The highest scoring hit to the repbase database from each predicted LTR sequence was used in annotations. Predicted full length LTRs with no significant hit to RepBase were blasted to the GenBank nr database using the tblastx algorithm, and significant hits (e-value < 1e⁻⁵) to protein coding genes were removed. A custom perl script was used to extract 5’ and 3’ LTR pairs from sequences based on their predicted start and end location from the LTR_Finder output (https://bitbucket.org/barkerlab/baniaga_fern_lycophyte_ltrs).

Unlike most investigations of LTR dynamics in vascular plants, we chose to exclude LTRs with a history of intraelement gene conversion. This is because the estimation of LTR insertion age assumes that the 5’ and 3’ ends of an LTR experience mutations independently after insertion. Intraelement gene conversion in LTRs is common in plant genomes (Cossu *et al.* 2017) and acts to homogenize the 5’ and 3’ ends of an LTR. This leads to a reduction in the amount of sequence divergence, and an underestimation of LTR insertion dates. Our exclusion of LTRs with evidence of intraelement gene

BANIAGA : LTR DYNAMICS

conversion confronts this potential bias. LTRs with significant evidence of intraelement gene conversion ($p < 0.05$) were identified with GENECONV (Sawyer 1989) using the following parameters “/w123/lp/f/eb/g2 -include_monosites”. These LTRs were excluded from the dataset and all subsequent analyses.

LTR insertion date estimation.—Insertion dates were calculated following the methods of SanMiguel *et al.* (1998). LTR 5' and 3' pairs were aligned in MUSCLE (Edgar 2004) and divergence between LTR pairs was calculated in fastphylo v1.0.1 (Khan *et al.* 2013) using the Kimura two-parameter (K2P) and the GTR+ γ nucleotide substitution models in FastTree v2.1 (Price *et al.* 2010). Insertion dates were converted into million years using lineage specific synonymous substitution rates per site per year sourced from the literature (Appendix 1), and the arithmetic mean, and standard deviation of insertion dates were calculated for each taxon. In our analyses, the GTR+ γ nucleotide model consistently estimated younger insertion dates for all taxa, with an average 58% reduction in the mean estimate of LTR insertion date relative to the K2P model. We chose to describe the results of the GTR+ γ model because it allowed a greater amount of parameters to vary in how nucleotides change over time in comparison to the more simpler K2P model (Tavaré 1986).

Relationship between mean LTR insertion date and genome size.—Haploid nuclear genome size estimates were sourced from the literature. Due to the large variation in haploid genome size across taxa, a regression was performed on mean LTR insertion date

BANIAGA : LTR DYNAMICS

against $\log_{10}$ [haploid nuclear genome size Mbp]. In addition, because of the shared evolutionary history of the species included in our analyses a phylogenetic independent contrast was performed on this dataset. Sequences for the *rbcL* plastid marker were downloaded from NCBI GenBank for each species and the outgroup *Physcomitrella patens* (Appendix 1). An *rbcL* sequence for *Plagiogyria formosana* was not available, and instead a sequence from the congener *P. yakumonticola* Nakaike was used. Sequences were aligned in MUSCLE, and an ultrametric phylogenetic tree was inferred using BEAST v2.0 (Bouckaert *et al.* 2014). The outgroup was pruned from the tree and the phylogenetic independent contrast was performed on both the $\log_{10}$ [haploid nuclear genome size Mbp] and mean LTR insertion date using the ‘pic’ function in the R ape package (Paradis *et al.* 2004). A nexus file of the phylogenetic tree is available at TreeBASE (<http://purl.org/phylo/treebase/phyloids/study/TB2:S24023>).

RESULTS

Fern LTR dynamics.—We observed different patterns of LTR dynamics in homosporous versus heterosporous ferns. Across all five homosporous fern taxa analyzed, estimated LTR insertion dates followed a broad unimodal distribution with three common attributes (Fig. 1). Homosporous ferns had few recent LTRs inserted within the past million years. Less than 3% of the LTR insertions occurred within the past one million years in *Ceratopteris richardii* and *Pteridium aquilinum*, and 1% or less of LTR insertions occurred within the past million years in *Dipteris*, *Plagiogyria*, and *Cystopteris*. More recent activity was found in *Cystopteris* with roughly 12% of the insertions occurring

BANIAGA : LTR DYNAMICS

between 1–2 mya. The majority of the LTRs in homosporous ferns were estimated to be inserted between 5–13 mya, with a mean of 7.49 mya in *Plagiogyria formosana*, 8.16 mya in *Ceratopteris*, 9.04 mya in *Cystopteris*, 8.15 mya in *Dipteris*, and 9.52 mya in *Pteridium aquilinum*. Finally, we found numerous LTR insertions beyond 15 mya across all homosporous fern taxa, comprising roughly 10% of the total recovered LTR insertions for each taxon. The oldest full length LTR recovered for each taxon was 39.28 mya in *Pteridium*, 34.57 mya in *Ceratopteris*, 28.68 mya in *Dipteris*, 25.42 mya in *Cystopteris* and 24.99 mya in *Plagiogyria*.

LTR dynamics in the two heterosporous ferns were markedly different from the homosporous ferns. In both *Azolla* and *Salvinia*, estimated LTR insertion dates were similar to those observed in angiosperms with high amounts of recent activity (Fig. 1). In contrast to the homosporous ferns, an estimated 44% of LTR insertions occurred within the past one million years in *Azolla* and 24% in *Salvinia*. Mean insertion dates were also much younger with an estimated mean of 1.70 mya in *Azolla* and 3.11 mya in *Salvinia*. In addition, fewer than 0.1% of LTRs were older than 15 mya in *Azolla* and *Salvinia*.

Lycophyte (Selaginella) LTR dynamics.—All three *Selaginella* species shared similar patterns of LTR activity. These LTR dynamics are comparable to angiosperms and the heterosporous ferns (Fig. 1), in which a large proportion of the estimated LTRs insertions occurred 0–4 mya. Mean LTR insertion dates for the *Selaginella* species were 0.86 mya in *S. lepidophylla*, 1.73 mya in *S. tamariscina*, 1.88 mya in *S. moellendorffii*. In all

BANIAGA : LTR DYNAMICS

Selaginella genomes sampled, a majority of LTRs have insertion dates within the past 1 million years. An estimated 39% of LTRs are inserted within the past 1 million years in *S. moellendorffii*, 41% in *S. tamariscina*, and 75% in *S. lepidophylla*. Few LTRs inserted beyond 10 mya were recovered for any *Selaginella* species, the oldest estimated LTR insertion for *S. lepidophylla* was 20.75 mya, 22.2 mya for *S. moellendorffii*, and 11.58 mya for *S. tamariscina*.

Relationship between mean LTR insertion date and genome size.—A large range of mean LTR insertion dates and genome sizes were observed in our analyses. Mean LTR insertion dates range from 0.45 mya in *Erythranthe guttata* to 15.74 mya in *Pinus sylvestris*, and genome sizes range from 86 Mbp in *Selaginella moellendorffii* to 22.47 Gbp in *Pinus sylvestris*. A strong significant positive correlation is observed between the haploid nuclear genome size and mean LTR insertion date (Fig. 2). Species with small genome sizes consistently have relatively recent mean LTR insertion dates, and species with large genomes have older mean LTR insertion dates. This is evident for the linear regression of mean LTR insertion date against $\log_{10}$ [haploid nuclear genome size] (Adj- $R^2=0.77$, $y=5.012x - 10.3312$, $p<0.01$, $df=16$, $f\text{-statistic}=56.89$), as well as when phylogenetic relationships are taken into account (Adj- $R^2=0.55$, $y=4.132x$, $p<0.01$, $df=16$, $f\text{-statistic}=21.81$).

BANIAGA : LTR DYNAMICS

DISCUSSION

Across the ten fern and lycophyte genomes we analyzed, we found that haploid nuclear genome size was positively correlated with mean LTR insertion dates. Combined with data from representative seed plants, we found that the age distribution of LTR insertions was well correlated with vascular plant haploid nuclear genome size. Across vascular plants, we found that species with large genomes have older mean LTR insertion dates such as those seen in gymnosperms and homosporous ferns. Species with relatively small genomes had more recently inserted LTRs such as those seen in *Selaginella* and the heterosporous ferns. This first survey of LTR insertion dates for homosporous ferns and other species bridges a significant phylogenetic gap in understanding the evolution of vascular plant genome size. This result is consistent with previous research from seed plants that found genome size and LTR activity are well correlated.

Previous analyses have found a strong positive correlation between haploid nuclear genome size and chromosome number in ferns and lycophytes (Bainard *et al.* 2011; Barker 2013; Clark *et al.* 2016; Leitch and Leitch 2013; Nakazato *et al.* 2008). This pattern is absent from seed plants (Nakazato *et al.* 2008), and instead most of the variation in their nuclear genome size is due to LTR activity (El Baidouri and Panaud 2013; Michael 2014). The high correlation of chromosome number and genome size combined with uniquely small and uniform sized chromosomes in ferns and lycophytes (Nakazato *et al.* 2008; Wagner and Wagner, 1980;) suggested that transposable element activity may be

BANIAGA : LTR DYNAMICS

relatively low and not contribute significantly to genome size variation (Nakazato *et al.* 2008). However, our results suggest that LTR dynamics also strongly influence the nuclear genome sizes of ferns and lycophytes (*Selaginella* spp.).

Our current results indicate that the relationship among chromosome number, LTR activity, and genome size is not clearly understood in these plants. It may be the rate of LTR activity is sufficiently low enough that it does not interfere with the strong correlation of genome size and chromosome number generated by the (unknown) forces driving chromosome size uniformity. We currently have too few sampled fern and lycophyte genomes to formally analyze if chromosome number or LTR activity explains more variation in genome size, as well as their joint effect. However, the proportion of variation in genome size explained by LTR activity across the fern and lycophyte species we sampled ($\text{adj-R}^2=0.39$) is positive but possibly less than that explained by chromosome number (e.g., $\text{R}^2=0.83$ Nakazato *et al.* 2008; $\rho=0.5$ Clark *et al.* 2016). Future analyses leveraging better sampled fern and lycophyte genomes are needed to explore the relationships among chromosome number, LTR activity, and genome size. Comparative and population genomic approaches of LTR activity and chromosomal evolution are needed in this clade to understand the unique pattern of genome evolution.

Previous analyses have proposed that the high correlation of genome size and chromosome number, as well as the uniform size of fern chromosomes (Nakazato *et al.*

BANIAGA : LTR DYNAMICS

2008; Wagner and Wagner 1980;), is caused by the suppression of transposable elements in homosporous ferns (Nakazato *et al.* 2008). The older average LTR insertion times we observed in homosporous ferns may support this hypothesis. The lifecycle of a LTR inside a host genome involves replication, insertion into a different part of the genome, and potential silencing and deactivation of its activity via epigenetic and methylation processes from the host genome. The position of LTRs inside the host genome can affect the expression of proximal genes, as genes proximal to methylated LTRs often experience higher methylation and reduced expression (Hollister and Gaut 2009; Niederhuth *et al.* 2016). Cytosine methylation is one type of DNA methylation. Across vascular plants, the highest cytosine methylation found to date is in homosporous ferns with an estimated 57% of their genes having greater than 90% of their cytosines methylated (Takuno *et al.* 2016). In contrast, cytosine methylation in *Selaginella* appears to be localized to a subset of approximately 10% of genes (Takuno *et al.* 2016; VanBuren *et al.* 2018). The reason for high methylation in homosporous ferns is unclear. However, homosporous ferns are unique among vascular plants because following whole genome duplication chromosomes appear to be largely retained (Barker and Wolf 2010; Barker 2013; Nakazato *et al.* 2008). It may be that the high methylation in homosporous ferns is associated with silencing LTRs and genes on these retained chromosomes. Future analyses of homosporous fern genomes are needed to disentangle the relationships among these processes and determine the forces ultimately driving their unique patterns of genome evolution.

BANIAGA : LTR DYNAMICS

Older average insertion times of LTRs is consistent with observed patterns of structural turnover in homosporous fern genomes. Unlike many angiosperm lineages, ferns and lycophytes have highly uniform chromosomes in both size and structure (Wagner and Wagner 1980), and some lineages exhibit evidence of conserved genome size over long periods of geologic time (Baniaga *et al.* 2016; Clark *et al.* 2016; Schneider *et al.* 2015). Even fossilized and extant chromosomes of the royal ferns (Osmundaceae) do not show significant difference in size despite more than 180 million years (Bomfleur *et al.* 2014), and sterile hybrids may be generated between species of different genera that diverged more than 60 million years ago (Rothfels *et al.* 2015). Our observation of older average insertion times of LTRs in homosporous ferns is consistent with a slow rate of structural turnover in their genomes.

Why do we observe less active LTRs in homosporous fern genomes? A possible explanation is their unique mating systems. Although most diploid homosporous ferns sampled to date are highly outcrossing (Haufler 1987; Sessa *et al.* 2016; Soltis and Soltis 1992), homosporous ferns are capable of sporophytic and gametophytic selfing (Haufler *et al.* 2016; Klekowski and Baker 1966; Klekowski and Lloyd 1968). Mating system plays a large role in transposable element (TE) dynamics, of which LTRs are one type (Agren *et al.* 2014; Agren and Clark 2018; Boutin *et al.* 2012; Charlesworth and Langley 1986).

Theoretical models suggest that if TEs are deleterious to host fitness, the activity and accumulation of TEs should be reduced in inbreeding species. This is because selfers reduce the spread of TEs between individuals, and, compared to outcrossers, more

BANIAGA : LTR DYNAMICS

efficiently purge insertions with deleterious recessive effects (Morgan 2001; Wright *et al.* 2008; Wright and Schoen 1999). Outcrossing may also increase the activity and genetic diversity of LTRs during transposition burst cycles (Sanchez *et al.* 2017). Expectations of these models parallel those concerning genetic load in homosporous ferns (Hedrick 1987), and TEs, if deleterious, may comprise a large component of genetic load.

Applying these theoretical models of mating system and TE dynamics to homosporous ferns we would expect that strongly outcrossing species should have more recent LTR insertional activity and species with a history of inbreeding via selfing should have fewer recent LTR insertions. Interestingly, despite relatively limited sampling, our analyses of LTR dynamics in ferns and lycophytes are consistent with these predictions. The heterosporous *Selaginella* and the two aquatic ferns *Azolla filiculoides* and *Salvinia cucullata* share similar LTR dynamics characterized by a predominance of recent insertions. Although there may be several adaptive roles of heterospory in vascular plants (Petersen and Burd 2018), it does function to promote outcrossing. In contrast, LTR dynamics in the homosporous ferns which are capable of a greater array of selfing mechanisms, have relatively fewer recent insertions. As the costs to sequencing decrease it would be worthwhile to experimentally manipulate and compare the TE and LTR dynamics of homosporous ferns with different mating systems such as *C. richardii* and its sister species to empirically verify at a finer scale these macroevolutionary patterns. Similarly, it would also be useful to compare LTR dynamics in homosporous ferns with high rates of gametophytic selfing, such as species in the Ophioglossaceae

BANIAGA : LTR DYNAMICS

(Barker and Hauk 2003), to species with low rates of self-fertilization. Ferns and lycophytes may prove to fertile ground for testing hypotheses of LTR dynamics and mating systems in general because of their diverse mating systems.

Another interesting result of our analysis is evidence that we did not observe an effect of generation time on the insertion dynamics of LTRs in homosporous ferns. For example, the minimum generation time of *P. aquilinum* is two to three years (Klekowski 1972), whereas in *C. richardii* the minimum generation time is three months (Banks 1999; Hickok *et al.* 1995; Hickok and Warne 2004). If generation time had a strong effect on activity of LTRs, we would expect to observe a greater amount of recent LTR insertions in *C. richardii*, or a much older mean LTR insertion date in the other fern taxa sampled. Instead, we found that *C. richardii* has a similar LTR insertion distribution as observed in other fern taxa.

LTRs are important components of vascular plant genomes. Our analysis investigated the LTR dynamics of ferns and lycophytes and placed the results in the context of LTR dynamics in other vascular plants. Both the heterosporous *Selaginella* and members of the Salviniaceae share similar LTR insertion patterns to angiosperms, characterized by a predominance of insertions up to 5 mya. In contrast, the homosporous ferns share older average LTR insertion dates intermediate between angiosperms and gymnosperms. We also show that across vascular plants LTR insertion dates are strongly associated with nuclear genome size. Species with small genomes have rapid LTR turnover rates with

BANIAGA : LTR DYNAMICS

many young insertions, while species with larger genomes consistently have older LTRs with relatively little recent activity. Ferns and lycophytes are located at key locations to understand LTR dynamics in vascular plants, and elements of their reproductive process such as the capacity for gametophytic selfing in homosporous ferns make them unique study systems for future inquiries into LTR and other transposable element dynamics.

ACKNOWLEDGEMENTS

The author are grateful for L. Zheng and S. A. Jorgensen for comments on earlier versions of the manuscript. This work was supported in part by NSF grants IOS-1339156 and EF-1550838 to M.S.B.

Data Availability

TreeBase

*.nex

BitBucket data and scripts

LITERATURE CITED—

AGREN, 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.

BANIAGA : LTR DYNAMICS

AGREN, J. A., AND A. G. CLARK. 2018. Selfish genetic elements. PLoS Genetics 14:e1007700.

AMBORELLA GENOME PROJECT. 2013. The *Amborella* genome and the evolution of flowering plants. Science 342:1241089.

ARRIGO, N., J. THERIEN, C. L. ANDERSON, M. D. WINDHAM, C. H. HAUFLE, AND M. S. BARKER. 2013. A total evidence approach to understanding phylogenetic relationships and ecological diversity in *Selaginella* subg. *Tetragonostachys*. American Journal of Botany 100:1672–1682.

BATNARD, J. D., T. A. HENRY, L. D. BATNARD, AND S. G. NEWMASER. 2011. DNA content variation in monilophytes and lycophytes: large genomes that are not endopolyploid. Chromosome Research 19:763–775.

BANIAGA, A. E., N. ARRIGO, AND M. S. BARKER. 2016. The small nuclear genomes of *Selaginella* are associated with a low rate of genome size evolution. Genome Biology and Evolution 8:1516–1525.

BANKS, J. A. 1999. Gametophyte development in ferns. Annual Review of Plant Physiology and Plant Molecular Biology 50:163–186.

BANIAGA : LTR DYNAMICS

BANKS, J. A., T. NESHTYAMA, M. HASEBE, J. L. BOWMAN, M. GREBSKOV, C. DEPAMPHILIS, V. A. ALBERT, N. AONO, T. AOYAMA, B. A. AMBROSE, N. W. ASHTON, M. J. AXTELL, E. BARKER, M. S. BARKER, J. L. BENNETZEN, N. D. BONAWITZ, C. CHAPPLE, C. CHENG, L. G. G. CORREA, M. DACRE, J. DEBARRY, I. DREYER, M. ELIAS, E. M. ENGSTROM, M. ESTELLE, L. FENG, C. FINET, S. K. FLOYD, W. B. FROMMER, T. FUJITA, L. GRAMZOW, M. GUTENSOHN, J. HARHOLT, M. HATTORI, A. HEYL, T. HIRAI, Y. HIWATASHI, M. ISHIKAWA, M. IWATA, K. G. KAROL, B. KOEHLER, U. KOLUKTSAOGLU, M. KUBO, T. KURATA, S. LALONDE, K. LI, Y. LI, A. LITT, E. LYONS, G. MANNING, T. MARUYAMA, T. P. MICHAEL, K. MIKAMI, S. MIYAZAKI, S.-I. MORINAGA, T. MURATA, B. MUELLER-ROEBER, D. R. NELSON, M. OBARA, Y. OGURI, R. G. OLMSTEAD, N. ONODERA, B. L. PETERSEN, B. PILLS, M. PRIGGE, S. A. RENSING, D. M. RIAÑO-PACHÓN, A. W. ROBERTS, Y. SATO, H. V. SCHELLER, B. SCHULZ, C. SCHULZ, E. V. SHAKIROV, N. SHIBAGAKI, N. SHINOHARA, D. E. SHEPPEN, I. SØRENSEN, R. SOTOOKA, N. SUGIMOTO, M. SUGITA, N. SUMIKAWA, M. TANURDZIC, G. THEISSEN, P. ULVSKOV, S. WAKAZUKI, J.-K. WENG, W. G. T. WILLATS, D. WIPF, P. G. WOLF, L. YANG, A. D. ZIMMER, Q. ZHU, T. MITROS, U. HELLSTEN, D. LOQUÉ, R. OTELLAR, A. SALAMOV, J. SCHMUTZ, H. SHAPIRO, E. LINDQUIST, S. LUCAS, D. ROKHSAR, AND I. V. GRIGORIEV. 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332:960–963.

BAO, W., KOJIMA, K. K., AND O. KOHANY. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6:11.

BANIAGA : LTR DYNAMICS

BARKER, M. S., AND W. D. HAUKE. 2003. An evaluation of *Sceptridium dissectum* (Ophioglossaceae) with ISSR markers: implications for *Sceptridium* systematics. American Fern Journal 93:1–19.

BARKER, M. S. 2009. Evolutionary genomic analyses of ferns reveal that high chromosome numbers are a product of high retention and fewer rounds of polyploidy relative to angiosperms. American Fern Journal 99:136–141.

BARKER, M. S., AND P. G. WOLF. 2010. Unfurling fern biology in the genomics age. BioScience 60:177–185.

BARKER, M. S. 2013. Karyotype and genome evolution in pteridophytes. Pp. 245–253, in: J. Greilhuber, J. Dolezel, and J. Wendel (eds.), *Plant Genome Diversity Volume 2*. Springer, Vienna.

BARKER, M. S., Z. LI, T. I. KIDDER, C. R. REARDON, Z. LAI, L. O. OLIVETRA, M. SCASCITELLI, AND L. H. REESEBERG. 2016. Most Compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. American Journal of Botany 103:1203–1211.

BANIAGA : LTR DYNAMICS

BOMFLEUR, B., S. McLOUGHLIN, AND V. VAJDA. 2014. Fossilized nuclei and chromosomes reveal 180 million years of genomic stasis in royal ferns. *Science* 343:1376–1377.

BOUCKAERT, R., J. HELED, D. KUHNERT, T. VAUGHAN, C.-H. WU, D. XIE, M. A. SUCHARD, A. RAMBAUT, AND A. J. DRUMMOND. 2014. BEAST 2: a software for bayesian evolutionary analysis. *PLoS Computational Biology* 10:e1003537.

BOUTIN, T. S., A. LE ROUZIC, AND P. CAPY. 2012. How does selfing affect the dynamics of selfish transposable elements?. *Mobile DNA* 3:5.

CHARLESWORTH, B., AND C. H. LANGLEY. 1986. The evolution of self-regulated transposition of transposable elements. *Genetics* 112:359–383.

CLARK, J., O. HEDALGO, J. PELLICER, H. LIU, J. MARQUARDT, Y. ROBERT, M. CHRISTENHUSZ, S. ZHANG, M. GIBBY, I. J. LEITCH, AND H. SCHNEIDER. 2016. Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytologist* 210:1072–1082.

COSSU, R. M., C. CASOLA, S. GIACOMELLO, A. VIDALIS, D. G. SCOTFIELD, AND A. ZUCCOLO. 2017. LTR retrotransposons show low levels of unequal recombination and high rates of intraelement gene conversion in large plant genomes. *Genome Biology and Evolution* 9:3449–3462.

BANIAGA : LTR DYNAMICS

CUT, L., P. K. WALL, J. H. LEEBENS-MACK, B. G. LINDSAY, D. E. SOLTIS, J. J. DOYLE, P. S.

SOLTIS, J. E. CARLSON, K. ARUMUGANATHAN, A. BARAKAT, V. A. ALBERT, H. MA, AND C. W.

DEPAMPHILIS. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16:738–749.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797.

EL BAIDOURI, M., AND O. PANAUD. 2013. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biology and Evolution* 5:954–965.

FESCHOTTE, C., N. JIANG, AND S. R. WESSLER. 2002. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* 3:329–341.

FLAVELL, R. 1980. The molecular characterization and organization of plant chromosomal DNA sequences. *Annual Review of Plant Physiology and Plant Molecular Biology* 31:569–596.

FLETSCHMANN, A., T. P. MICHAEL, F. RIVADAVIA, A. SOUSA, W. WANG, E. M. TEMSCH, J.

GREILHUBER, K. F. MÜLLER, AND G. HEUBL. 2014. Evolution of genome size and chromosome

BANIAGA : LTR DYNAMICS

number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Annals of Botany* 114:1651–1663.

FREELING, M., M. R. WOODHOUSE, S. SUBRAMANTAM, G. TURCO, D. LETSCH, AND J. C. SCHNABLE. 2012. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Current Opinion in Plant Biology* 15:131–139.

GAUT, B. S., B. R. MORTON, B. C. MCCATG, AND M. T. CLEGG. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proceedings of the National Academy of Sciences of the United States of America* 93:10274–10279.

HANSON, L. AND I. J. LETTCH. 2002. DNA amounts for five pteridophyte species fill phylogenetic gaps in C-value data. *Botanical Journal of the Linnean Society* 140:169–173.

HAUFLER, C. H. 1987. Electrophoresis is modifying our concepts of evolution in homosporous pteridophytes. *American Journal of Botany* 74:953–966.

HAUFLER, C. H. 2014. Ever since Klekowski: testing a set of radical hypotheses revives the genetics of pteridophytes. *American Journal of Botany* 101:2036–2042.

BANIAGA : LTR DYNAMICS

HAUFLER, C. H., K. M. PRYER, E. SCHUETTPPELZ, E. B. SESSA, D. R. FARRAR, R. MORAN, J. J.

SCHNELLER, J. E. WATKINS, AND M. D. WINDHAM. 2016. Sex and the single gametophyte: revising the homosporous vascular plant life cycle in light of contemporary research. *BioScience* 66:928–937.

HEDRICK, P. 1987. Genetic load and the mating system in homosporous ferns. *Evolution* 41:1282–1289.

HICKOK, L. G., T. R. WARNE, AND R. S. FRIEDBOURG. 1995. The biology of the fern *Ceratopteris* and its use as a model system. *International Journal of Plant Sciences* 156:332–345.

HICKOK, L. G., AND T. R. WARNE. 2004. *C-Fern Manual: Teaching and Research*. Carolina Biological Supply, Burlington, N.C..

HEDALGO, O., J. PELLICER, M. J. M. CHRISTENHUSZ, H. SCHNEIDER, AND I. J. LEITCH. 2017. Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*. *Botanical Journal of the Linnean Society* 183:509–514.

HOLLISTER, J. D., AND B. S. GAUT. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* 19:1419–1428.

BANIAGA : LTR DYNAMICS

IBARRA-LACLETTE, E., E. LYONS, G. HERNÁNDEZ-GUZMÁN, C. A. PÉREZ-TORRES, L.

CARRETERO-PAULET, T.-H. CHANG, T. LAN, A. J. WELCH, M. J. A. JUÁREZ, J. SIMPSON, A.

FERNÁNDEZ-CORTÉS, M. ARTEAGA-VÁZQUEZ, E. GÓNGORA-CASTILLO, G. ACEVEDO-HERNÁNDEZ, S.

C. SCHUSTER, H. HIMMELBAUER, A. E. MINOCHE, S. XU, M. LYNCH, A. OROPEZA-ABURTO, S. A.

CERVANTES-PÉREZ, M. DE JESÚS ORTEGA-ESTRADA, J. I. CERVANTES-LUEVANO, T. P. MICHAEL, T.

MOCKLER, D. BRYANT, A. HERRERA-ESTRELLA, V. A. ALBERT, AND L. HERRERA-ESTRELLA. 2013.

Architecture and evolution of a minute plant genome. *Nature* 498:94–98.

JIAO, Y., N. J. WICKETT, S. AYYAMPALAYAM, A. S. CHANDERBALT, L. LANDHERR, P. E. RALPH, L. P.

TOMSHO, Y. HU, H. LIANG, P. S. SOLTIS, D. E. SOLTIS, S. W. CLIFTON, S. E. SCHLARBAUM, S. C.

SCHUSTER, H. MA, J. LEEBENS-MACK, AND C. W. DEPAMPHILIS. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100.

KAGALE, S., S. J. ROBINSON, J. NIXON, R. XIAO, T. HUEBERT, J. CONDIE, D. KESSLER, W. E.

CLARKE, P. E. EDGER, M. G. LINKS, A. G. SHARPE, AND I. A. P. PARKIN. 2014. Polyploid evolution of the Brassicaceae during the Cenozoic Era. *Plant Cell* 26:2777–2791.

KELLY, L. J., S. RENNY-BYFELD, J. PELLICER, J. MACAS, P. NOVAK, P. NEUMANN, M. A. LYSAK, P.

D. DAY, M. BERGER, M. F. FAY, R. A. NICHOLS, A. R. LETTCH, AND I. J. LETTCH. 2015. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytologist* 208:596–607.

BANIAGA : LTR DYNAMICS

KENRICK, P., AND P. R. CRANE. 1997. *The origin and early diversification of land plants*.

Smithsonian Institution Press, Washington.

KHAN, M. A., I. ELIAS, E. SJOLUND, K. NYLANDER, R. V. GUTMERA, R. SCHOBESBERGER, P.

SCHMITZBERGER, J. LAGERGREN, AND L. ARVESTAD. 2013. Fastphylo: fast tools for phylogenetics.

BMC Bioinformatics 14:334.

KLEKOWSKI, E. J., AND H. G. BAKER. 1966. Evolutionary significance of polyploidy in the

Pteridophyta. *Science* 153:305–307.

KLEKOWSKI, E. J., AND R. M. LLOYD. 1968. Reproductive biology of the Pteridophyta 1.

General considerations and a study of *Onoclea sensibilis* L. *Botanical Journal of the*

Linnean Society 60:315–324.

KLEKOWSKI, E. J. 1972. Evidence against genetic self-incompatibility in the homosporous

fern *Pteridium aquilinum*. *Evolution* 26:66–73.

KOCH, M. A., B. HAUBOLD, AND T. MITCHELL-OLDS. 2000. Comparative evolutionary

analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and

related genera (Brassicaceae). *Molecular Biology and Evolution* 17:1483–1498.

BANIAGA : LTR DYNAMICS

KORALL, P., AND P. KENRICK. 2002. Phylogenetic relationships in Selaginellaceae based on *rbcL* sequences. *American Journal of Botany* 89:506–517.

KOVACH, A., J. L. WGRZYN, G. PARRA, C. HOLT, G. E. BRUENING, C. A. LOOPSTRA, J. HARTIGAN, M. YANDELL, C. H. LANGLEY, I. KORF, AND D. B. NEALE. 2010. The *Pinus taeda* genome is characterized by diverse and highly repetitive sequences. *BMC Genomics* 11:420.

LANDIS, J. B., D. E. SOLTIS, Z. LI, H. E. MARX, M. S. BARKER, D. C. TANK, AND P. S. SOLTIS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* 105:348–363.

LEITCH I. J., AND A. R. LEITCH. 2013. Genome size diversity and evolution in land plants. Pp. 307–322, in I. J. Leitch, J. Greilhuber, J. Doležal, and J. F. Wendel, (eds.), *Plant Genome Diversity, Volume 2*. Springer-Verlag, Vienna.

LI, F.-W., P. BROUWER, L. CARRETERO-PAULET, S. CHENG, J. DE VRIES, P. M. DELAUX, A. EITLY, N. KOPPERS, L. Y. KUO, Z. LI, M. SIMENC, I. SMALL, E. WAFULA, S. ANGARITA, M. S. BARKER, A. BRÄUTIGAM, C. DEPAMPHILES, S. GOULD, P. S. HOFMANT, Y. M. HUANG, B. HUETTEL, Y. KATO, X. LIU, S. MAERE, R. MCDOWELL, L. A. MUELLER, K. G. J. NIEROP, S. A. RENSING, T. ROBISON, C. J. ROTHFELS, E. M. STIGEL, Y. SONG, P. R. TIMILSENA, Y. VAN DE PEER, YVES, H. WANG, P. K. I. WILHELMSSON, P. G. WOLF, X. XU, J. P. DER, H. SCHLUEPMANN, G. K. WONG, AND K.M. PRYER.

BANIAGA : LTR DYNAMICS

2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* 4:460–472.

LI, Z., A. E. BANIAGA, E. B. SESSA, M. SCASCITELLI, S. W. GRAHAM, L. H. REESEBERG, AND M. S. BARKER. 2015. Early genome duplications in conifers and other seed plants. *Science Advances* 1:e1501084.

MA, J., AND J. L. BENNETZEN. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences United States of America* 101:12404–12410.

MICHAEL, T. P. 2014. Plant genome size variation: bloating and purging DNA. *Briefings in Functional Genomics* 13:308–317.

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

MURAT, F., R. ZHANG, S. GUTZARD, R. FLORES, A. ARMERO, C. PONT, D. STEINBACH, H. QUESNEVILLE, R. COOKE, AND J. SALSE. 2014. Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16K protogenes. *Genome Biology and Evolution* 6:12–33.

BANIAGA : LTR DYNAMICS

NAKAZATO, T., M. S. BARKER, L. H. REESEBERG, AND G. J. GASTONY. 2008. Evolution of the nuclear genome of ferns and lycophytes. Pp. 175–198, in T. A. Ranker and C. H. Haufler (eds.) *Biology and Evolution of Ferns and Lycophytes*. Cambridge University Press, Cambridge, England.

NIEDERHUTH, C. E., A. J. BEWICK, L. JI, M. S. ALABADY, K. D. KIM, Q. LI, N. A. ROHR, A. RAMBANT, J. M. BURKE, J. A. UDALL, C. EGEST, J. SCHMUTZ, J. GRIMWOOD, S. A. JACKSON, N. M. SPRINGER, AND R. J. SCHMITZ. 2016. Widespread natural variation of DNA methylation within angiosperms. *Genome Biology* 17:194.

NYSTEDT, B., N. R. STREET, A. WETTERBOM, A. ZUCCOLO, Y.-C. LIN, D. G. SCOTTELD, F. VEZZI, N. DELHOMME, S. GIACOMELLO, A. ALEXEYENKO, R. VICEDOMINI, K. SAHLIN, E. SHERWOOD, M. ELFSTRAND, L. GRAMZOW, K. HOLMBERG, J. HÄLLMAN, O. KEECH, L. KLASSON, M. KORTABINE, M. KUCUKOGLU, M. KÄLLER, J. LUTHMAN, F. LYSHOLM, T. NIETTYLÄ, A. OLSON, N. RELAKOVIC, C. RITLAND, J. A. ROSSELLÓ, J. SENA, T. SVENSSON, C. TALAVERA-LÓPEZ, G. THEISSEN, H. TUOMINEN, K. VANNESTE, Z.-Q. WU, B. ZHANG, P. ZERBE, L. ARVESTAD, R. BHALERAO, J. BOHLMANN, J. BOUSQUET, R. G. GIL, T. R. HVEDSTEN, P. DE JONG, J. MACKAY, M. MORGANTE, K. RITLAND, B. SUNDBERG, S. L. THOMPSON, Y. VAN DE PEER, B. ANDERSSON, O. NILSSON, P. K. INGVARSSON, J. LUNDEBERG, AND S. JANSSON. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497:579–84.

BANIAGA : LTR DYNAMICS

PARADES, E., J. CLAUDE, AND K. STRIMMER. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.

PELLICER, J., M. F. FAY, AND I. J. LETTCH. 2010. The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society* 164:10–15.

PETERSEN, K. B., AND M. BURD. 2018. The adaptive value of heterospory: evidence from *Selaginella*. *Evolution* 72:1080–1091.

PRICE, M. N., P. S. DEHAL, AND A. P. ARKIN. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490.

RIVADAVIA, F., P. M. GONELLA, AND A. FLETSCHMANN. 2013. A new and tuberous species of *Genlisea* (Lentibulariaceae) from the Campos Rupestres of Brazil. *Systematic Botany* 38:464–470.

ROTHFELS, C. J., A. K. JOHNSON, P. H. HOVENKAMP, D. L. SWOFFORD, H. C. ROSKAM, C. R. FRASER-JENKINS, M. D. WINDHAM, AND K. M. PRYER. 2015. Natural hybridization between genera that diverged from each other approximately 60 million years ago. *American Naturalist* 185:433–442.

BANIAGA : LTR DYNAMICS

SANCHEZ, D. H., J. GAUBERT, H.-G. DROST, N. R. ZABET, AND J. PASZKOWSKI. 2017.

High-frequency recombination between members of an LTR retrotransposon family during transposition bursts. *Nature Communications* 8:1283.

SANMIGUEL, P., B. S. GAUT, A. TEKHONOV, Y. NAKAJIMA, AND J. L. BENNETZEN. 1998. The paleontology of intergene retrotransposons in maize. *Nature Genetics* 20:43–45.

SAWYER, S. 1989. Statistical tests for gene conversion. *Molecular Biology and Evolution* 6:526–538.

SCHNABLE, P. S., D. WARE, R. S. FULTON, J. C. STEIN, F. WEI, S. PASTERNAK, C. LIANG, J. ZHANG,

L. FULTON, T. A. GRAVES, P. MINX, A. D. REILLY, L. COURTNEY, S. S. KRUCHOWSKI, C.

TOMLINSON, C. STRONG, K. DELEHAUNTY, C. FRONICK, B. COURTNEY, S. M. ROCK, E. BELTER, F.

DU, K. KIM, R. M. ABBOTT, M. COTTON, A. LEVY, P. MARCHETTO, K. OCHOA, S. M. JACKSON, B.

GILLAM, W. CHEN, L. YAN, J. HIGGINBOTHAM, M. CARDENAS, J. WALTGORSKI, E. APPLEBAUM, L.

PHELPS, J. FALCONE, K. KANCHI, T. THANE, A. SCIMONE, N. THANE, J. HENKE, T. WANG, J.

RUPPERT, N. SHAH, K. ROTTER, J. HODGES, E. INGENTHON, M. CORDES, S. KOHLBERG, J. SGRO, B.

DELGADO, K. MEAD, A. CHINWALLA, S. LEONARD, K. CROUSE, K. COLLURA, D. KUDRNA, J. CURRIE,

R. HE, A. ANGELOVA, S. RAJASEKAR, T. MUELLER, R. LOMELT, G. SCARA, A. KO, K. DELANEY, M.

WISSOTSKI, G. LOPEZ, D. CAMPOS, M. BRAIDOTTI, E. ASHLEY, W. GOLSER, H. KIM, S. LEE, J. LIN,

Z. DUJMIC, W. KIM, J. TALAG, A. ZUCCOLO, C. FAN, A. SEBASTIAN, M. KRAMER, L. SPTEGEL, L.

NASCIMENTO, T. ZUTAVERN, B. MILLER, C. AMBROISE, S. MULLER, W. SPOONER, A. NARECHANIA, L.

BANIAGA : LTR DYNAMICS

REN, S. WEL, S. KUMAR, B. FAGA, M. J. LEVY, L. MCMAHAN, P. VAN BUREN, M. W. VAUGHN, K. YING, C.-T. YEH, S. J. EMREICH, Y. JIA, A. KALYANARAMAN, A.-P. HSTA, W. B. BARBAZUK, R. S. BAUCOM, T. P. BRUTNELL, N. C. CARPITA, C. CHAPARRO, J.-M. CHIA, J.-M. DERAGON, J. C. ESTELL, Y. FU, JEFFREY A. JEDDELOH, Y. HAN, H. LEE, P. LI, D. R. LITSCH, S. LTU, Z. LTU, D. H. NAGEL, M. C. MCCANN, P. SANMIGUEL, A. M. MYERS, D. NETTLETON, J. NGUYEN, B. W. PENNING, L. PONNALA, K. L. SCHNEIDER, D. C. SCHWARTZ, A. SHARMA, C. SODERLUND, N. M. SPRINGER, Q. SUN, H. WANG, M. WATERMAN, R. WESTERMAN, T. K. WOLFGRUBER, L. YANG, Y. YU, L. ZHANG, S. ZHOU, Q. ZHU, J. L. BENNETZEN, R. K. DAWE, J. JTANG, N. JTANG, G. G. PRESTING, S. R. WESSLER, S. ALURU, R. A. MARTIENSSSEN, S. W. CLIFTON, W. R. MCCOMBIE, R. A. WING, AND R. K. WILSON. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115.

SCHNEIDER, H., H. LTU, J. CLARK, O. HEDALGO, J. PELLICER, S. ZHANG, L. J. KELLY, M. F. FAY, AND I. J. LETTCH. 2015. Are the genomes of royal ferns really frozen in time? Evidence for coinciding genome stability and limited evolvability in royal ferns. *New Phytologist* 207:10–13.

SESSA, E. B., W. L. TESTO, AND J. E. WATKINS. 2016. On the widespread capacity for, and functional significance of, extreme inbreeding in ferns. *New Phytologist* 211:1108–1119.

SOLTIS, D. E., AND P. S. SOLTIS. 1992. The distribution of selfing rates in homosporous ferns. *American Journal of Botany* 79:97–100.

BANIAGA : LTR DYNAMICS

TAKUNO, S., J.-H. RAN, AND B. S. GAUT. 2016. Evolutionary patterns of genic DNA methylation vary across land plants. *Nature Plants* 2:15222.

TAVARÉ, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* 17:57–86.

TILEY, G. P., AND J. G. BURLEIGH. 2015. The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC Evolutionary Biology* 15:194.

VANBUREN, R., C. M. WAT, S. OU, J. PARDO, D. BRYANT, N. JIANG, T. C. MOCKLER, P. EDGER, AND T. P. MICHAEL. 2018. Extreme haplotype variation in the desiccation-tolerant clubmoss *Selaginella lepidophylla*. *Nature Communications* 9:13.

VORONOVA, A., V. BELEVICH, A. KORTICA, AND D. RUNGIS. 2017. Retrotransposon distribution and copy number variation in gymnosperm genomes. *Tree Genetics and Genomes* 13:88.

WAGNER, W. H., JR., AND F. S. WAGNER. 1980. Polyploidy in Pteridophytes. Pp. 199–214, in W. H. Lewis (ed.), *Polyploidy: Biological Relevance*. Plenum Press, New York.

BANIAGA : LTR DYNAMICS

WAN, T., Z.-M. LIU, L.-F. LI, A. R. LEITCH, I. J. LEITCH, R. LOHAUS, Z.-J. LIU, H.-P. XIN, Y.-B. GONG, Y. LIU, W.-C. WANG, L.Y. CHEN, Y. YANG, L. J. KELLY, J. YANG, J.-L. HUANG, Z. LI, P. LIU, L. ZHANG, H.-M. LIU, H. WANG, S.-H. DENG, M. LIU, J. LI, L. MA, Y. LIU, Y. LI, W. XU, L.-Q. WU, F. LIU, Q. MA, X.-R. YU, Z. JIANG, G.Q. ZHANG, S.-H. LI, R.-Q. LI, Z.-Z. ZHANG, Q.-F. WANG, Y. VAN DE PEER, J.-B. ZHANG, AND X.-M. WANG. 2018. A genome for gnetophytes and early evolution of seed plants. *Nature Plants* 4:82–89.

WENDEL, J. F., S. A. JACKSON, B. C. MEYERS, AND R. A. WING. 2016. Evolution of plant genome architecture. *Genome Biology* 17:37.

WESTRAND, S., AND P. KORRALL. 2016. Phylogeny of Selaginellaceae: there is value in morphology after all! *American Journal of Botany* 103:2136–2159.

WOLF, P. G., E. B. SESSA, D. B. MARCHANT, F.-W. LI, C. J. ROTHFELS, E. M. SIGEL, M. A. GITZDANNER, C. J. VISGER, J. A. BANKS, D. E. SOLTIS, P. S. SOLTIS, K. M. PRYER, AND J. P. DER. 2015. An exploration of fern genome space. *Genome Biology and Evolution* 7:2533–2544.

WRIGHT, S. I., AND D. J. SCHOEN. 1999. Transposon dynamics and the breeding system. *Genetica* 107:139–148.

WRIGHT, S. I., R. W. NESS, J. P. FOXE, AND S. C. H. BARRETT. 2008. Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences* 169:105–118.

BANIAGA : LTR DYNAMICS

XU, Z., AND H. WANG. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* 35:W265–W268.

XU, Z., Y. XIN, D. BARTELS, Y. LI, W. GU, H. YAO, S. LIU, H. YU, X. PU, J. ZHOU, J. XU, C. XU, H. LI, J. SONG, AND S. CHEN. 2018. Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. *Molecular Plant* 11:983–994.

ZENIL-FERGUSON, R., J. M. PONCIANO, AND J. G. BURLEIGH. 2016. Evaluating the role of genome downsizing and size thresholds from genome size distributions in angiosperms. *American Journal of Botany* 103:1175–1186.

ZHOU, X.-M., C. J. ROTHFELS, L. ZHANG, Z.-R. HE, T. LE PECHON, H. HE, N. T. LU, R. KNAPP, D. LORENCE, X.-J. HE, X.-F. GAO, AND L.-B. ZHANG. 2016. A large-scale phylogeny of the lycophyte genus *Selaginella* (Selaginellaceae: Lycopodiopsida) based on plastid and nuclear loci. *Cladistics* 32:360–389.

ZUCCOLO, A., D. G. SCOFIELD, E. DE PAOLI, AND M. MORGANTE. 2015. The *Ty1-copia* LTR retroelement family PARTC is highly conserved in conifers over 200 MY of evolution. *Gene* 568:89–99.

BANIAGA : LTR DYNAMICS

TABLE 1. Summary of LTR Composition in Fern and Lycophyte Genomes. Percent of the genome comprised of LTRs was sourced from the literature.

Species	Authority	Family	Percent LTRs	Reference
<i>Azolla filiculoides</i>	Lam.	Salviniaceae	37.8	Li <i>et al.</i> 2018
<i>Ceratopteris richardii</i>	Brongn.	Pteridaceae	37.6	Wolf <i>et al.</i> 2015
<i>Cystopteris protrusa</i>	(Weath.) Blasdel	Cystopteridaceae	31.4	Wolf <i>et al.</i> 2015
<i>Dipteris conjugata</i>	Reinw.	Dipteridaceae	22.6	Wolf <i>et al.</i> 2015
<i>Plagiogyria formosana</i>	Nakai	Plagiogyriaceae	24.9	Wolf <i>et al.</i> 2015
<i>Polypodium glycyrrhiza</i>	D.C. Eaton	Polypodiaceae	28.5	Wolf <i>et al.</i> 2015
<i>Pteridium aquilinum</i>	(L.) Kuhn	Dennstaedtiaceae	27.8	Wolf <i>et al.</i> 2015
<i>Salvinia cucullata</i>	Roxb.	Salviniaceae	19.2	Li <i>et al.</i> 2018
<i>Selaginella lepidophylla</i>	(Hook. & Grev.) Spring	Selaginellaceae	13.07	Van Buren <i>et al.</i> 2018
<i>Selaginella moellendorffii</i>	Hieron.	Selaginellaceae	35.6	Banks <i>et al.</i> 2011
<i>Selaginella tamariscina</i>	(P. Beauv.) Spring	Selaginellaceae	25.47	Xu <i>et al.</i> 2018

BANIAGA : LTR DYNAMICS

TABLE 2. LTR Summary Statistics. Number of full length LTRs recovered in the genome of each species from our analysis. Arithmetic mean and standard deviation of estimated LTR insertion dates based on lineage specific substitution rates (Appendix 1).

Summaries reported for both K2P and GTR + γ nucleotide substitution models.

Species	Authority	Haploid Nuclear Genome Size (Mbp)	Full Length LTRs	GTR + γ Mean	K2P Mean
<i>Selaginella lepidophylla</i>	(Hook. & Grev.) Spring	109	797	0.86	1.49
<i>Selaginella moellendorffii</i>	Hieron.	86	2,159	1.88	3.27
<i>Selaginella tamariscina</i>	(P. Beauv.) Spring	150	2,675	1.73	3.04
<i>Azolla filiculoides</i>	Lam.	750	17,186	1.70	3.04
<i>Ceratopteris richardii</i>	Brongn.	11,205	125	8.18	12.58
<i>Cystopteris protrusa</i>	(Weath.) Blasdell	4,230	33	9.04	16.3
<i>Dipteris conjugata</i>	Reinw.	2,450	76	8.15	14.39
<i>Plagiogyria formosana</i>	Nakai	14,810	74	7.49	13.62
<i>Pteridium aquilinum</i>	(L.) Kuhn	15,650	391	9.52	15.99
<i>Salvinia cucullata</i>	Roxb.	260	1,711	3.11	5.52
<i>Gnetum montanum</i>	Markgr.	4,200	10,977	9.69	16.91
<i>Picea abies</i>	(L.) H. Karst	19,570	36,868	13.15	22.54
<i>Pinus sylvestris</i>	L.	22,474	89	15.74	25.48
<i>Amborella trichopoda</i>	Baill.	868	1,166	4.16	6.75
<i>Arabidopsis thaliana</i>	(L.) Heynh.	135	286	0.75	1.31
<i>Erythranthe guttata</i>	(Fisch. ex DC.) G.L. Nesom	362	2,535	0.45	0.79
<i>Solanum lycopersicum</i>	L.	950	3,916	1.35	2.12
<i>Sorghum bicolor</i>	(L.) Moench	697	12,980	0.64	1.13

BANIAGA : LTR DYNAMICS

Appendix 1. List of synonymous substitution rates per site per year used to estimate LTR insertion dates. List of GenBank accession numbers used to infer evolutionary relationships for a phylogenetic independent contrast (* no available sequence for *Plagiogyria formosana* was available and the congener *P. yakumonticola* was used).

Species	Synonymous Substitution Rate/site/year	Reference	NCBI GenBank Accession Number
<i>Selaginella lepidophylla</i>	6.50E-09	Gaut <i>et al.</i> 1996	AF419051.1
<i>Selaginella moellendorffii</i>	6.50E-09	Gaut <i>et al.</i> 1996	KY023084.1
<i>Selaginella tamariscina</i>	6.50E-09	Gaut <i>et al.</i> 1996	AB574655.1
<i>Azolla filiculoides</i>	4.79E-09	Barker 2009	KM360662.1
<i>Ceratopteris richardii</i>	4.79E-09	Barker 2009	EU352297.1
<i>Cystopteris protrusa</i>	4.79E-09	Barker 2009	JX874051.1
<i>Dipteris conjugata</i>	4.79E-09	Barker 2009	EF588692.1
<i>Plagiogyria formosana</i> *	4.79E-09	Barker 2009	AB574748.1
<i>Pteridium aquilinum</i>	4.79E-09	Barker 2009	AY300097.1
<i>Salvinia cucullata</i>	4.79E-09	Barker 2009	U05649.1
<i>Gnetum montanum</i>	2.20E-09	Nystedt <i>et al.</i> 2014	AY056575.1
<i>Picea abies</i>	2.20E-09	Nystedt <i>et al.</i> 2014	EU364777.1
<i>Pinus sylvestris</i>	2.20E-09	Nystedt <i>et al.</i> 2014	AB097775.1
<i>Amborella trichopoda</i>	6.50E-09	Amborella Genome Project 2013	L12628.2
<i>Arabidopsis thaliana</i>	1.50E-08	Koch <i>et al.</i> 2000	KU739560.1
<i>Erythranthe guttata</i>	1.50E-08	Koch <i>et al.</i> 2000	KF997284.1
<i>Solanum lycopersicum</i>	1.30E-08	Ma and Bennetzen 2004	XM_010327541.3
<i>Sorghum bicolor</i>	1.30E-08	Ma and Bennetzen 2004	AM849341.1
<i>Physcomitrella patens</i>	N/A	N/A	AB013656.2

BANIAGA : LTR DYNAMICS

Figure 1. Estimated Distribution of LTR Insertion Times in Vascular Plants. Ridgeline density plots of species across vascular plants with a focus on ferns and lycophytes. The heterosporous *Selaginella* and members of Salviniaceae share similar LTR insertion dynamics with angiosperms. Homosporous ferns share similar dynamics to gymnosperms. Minimum density displayed 0.005, LTR insertion dates are based on the GTR + γ nucleotide substitution model and lineage specific substitution rates.

Figure 2. Genome Size by LTR Decay Rate. Species with small genome sizes consistently have relatively recent LTR insertion dates, and species with large genomes have relatively older LTR insertion dates. A) The $\log_{10}$ (haploid nuclear genome size Mbp) and mean LTR insertion date (mya) across vascular plants. The mean age of LTR insertions is significantly correlated with genome size (Adj- $R^2=0.77$, $y=5.012x - 10.3312$, $p<0.01$, $df=16$, $f\text{-statistic}=56.89$). B) The phylogenetically corrected $\log_{10}$ (haploid nuclear genome size Mbp) and mean LTR insertion date (mya) across vascular plants. The phylogenetically corrected mean age of LTR insertions is also significantly correlated with genome size (Adj- $R^2=0.55$, $y=4.132x$, $p<0.01$, $df=16$, $f\text{-statistic}=21.81$).