This project is part of the PRIMA Programme supported by the European Union

Population-scale variation and coordinated dynamics of transposable elements in Mediterranean fig (*Ficus carica L.)*

Castellacci M., Usai G., Mascagni F., Vangelisti A., Simoni S., Natali L., Cavallini A., Giordani T.

Department of Agriculture, Food and Environment (DAFE), University of Pisa, Via del Borghetto 80, Pisa, Italy

Introduction

Transposable elements (TEs) are major components of plant genomes and key drivers of genome evolution, however the determinants underlying intraspecific variation in TE abundance remain poorly understood, particularly at TE class or lineage level. Here, we present a population-scale analysis of the TE landscape in the Mediterranean fig (*Ficus carica* L.) one of the earliest domesticated fruit trees. Using whole-genome Illumina resequencing (WGR) data from 286 Mediterranean fig genotypes (**Figure 1 and 2**), we quantified genome-wide TE abundance and variability and investigated their co-variation patterns. Moreover, we identified presence/absence variants referred to TEs and determined the age of TE insertions.

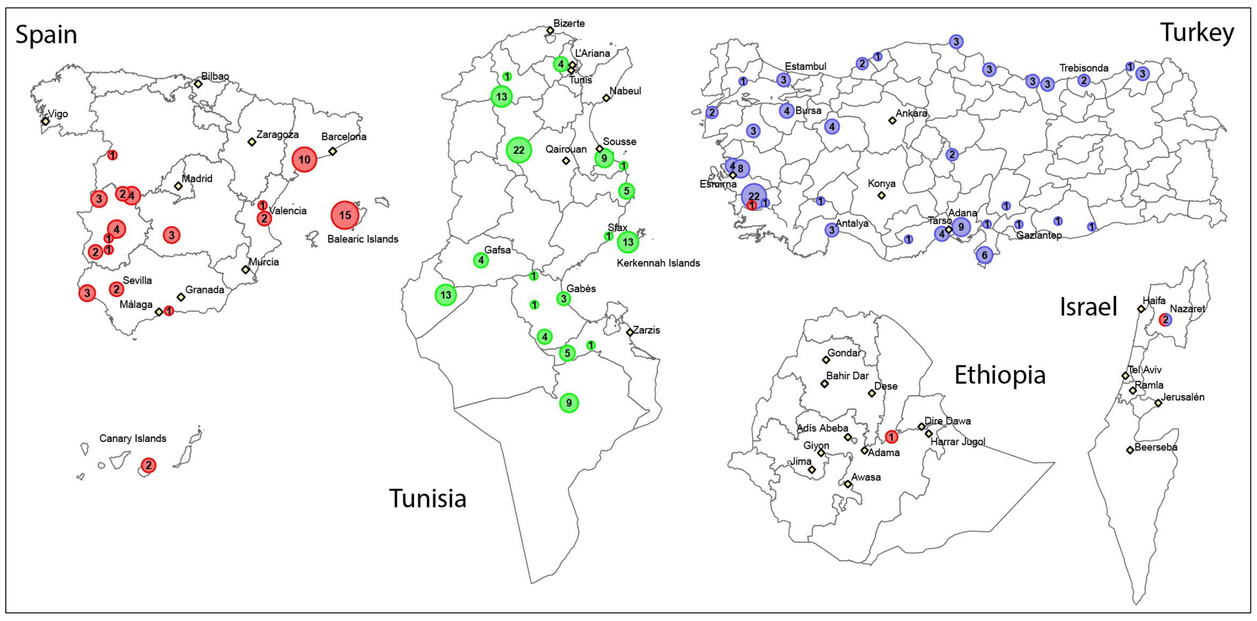

Figure 1. Geographical origin of 286 fig tree genotypes. Genotypes shown in red, blue, and green belong to the Spanish, Turkish, and Tunisian germplasm collections, respectively.

Figure 2. Representative examples of genotype diversity in fig germplasm.

TE abundance, variability and co-variation patterns

Methods

Paired-end WGR reads from 286 fig genotypes were obtained from Castellacci et al. (2026), filtered with *fastp* (Chen et al., 2018), and aligned with BWA (Li and Durbin, 2009) to the TE library from the fig reference genome (Usai et al., 2025). TE relative abundance was estimated as average coverage of mapped reads. Lineage-level correlations were computed and visualized in R using *metan* (Olivoto and Lúcio, 2020) and *corrplot* (Wei and Simko, 2017).

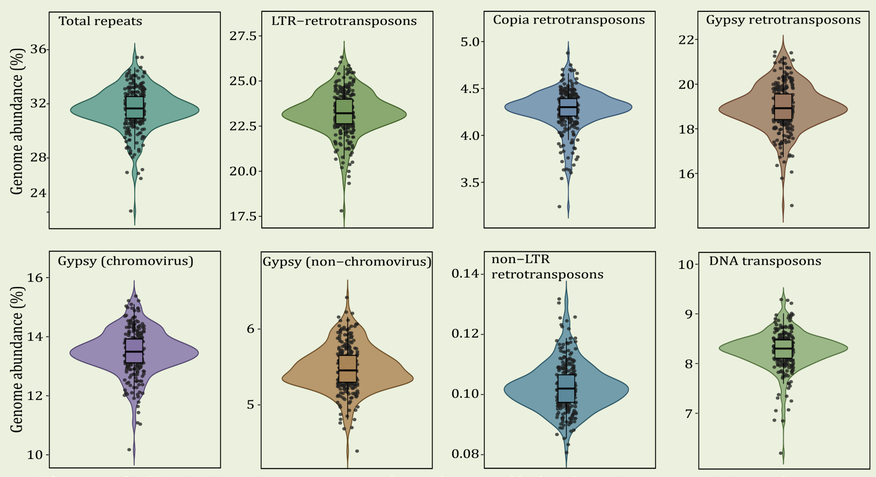

Figure 3. Variation in genome abundance (%) of major repetitive element classes across 286 fig genotype. Violin plots show density; boxes show median and interquartile range.

Results

On average, TEs accounted for 31.6% of the genome, with moderate variation among genotypes (**Figure 3**). Retrotransposons were the predominant component, with *Gypsy* long-terminal-repeat (LTR) retrotransposons, especially *Chromovirus*-related families such as *Tekay*, representing the largest TE fraction, followed by *Copia* retrotransposons. DNA transposons, especially Helitrons, also contributed substantially to genome composition (**Table 1**). Highly abundant lineages (e.g. *Tekay*, *Helitron*, *CRM*, *Athila*, *Tat/Retand*, *Ale*, *Angela*) showed relatively stable abundance across genotypes, whereas low-abundance lineages (e.g. *Galadriel* and *Alesia*) displayed higher relative variability (**Table 1**).

Correlation analyses revealed structured co-variation patterns, with strong positive correlations among some specific *Copia, Gypsy* and DNA transposon lineages (**Figure 4**), indicating that variation in these lineages tended to occur in a coordinated manner across genotypes.

Table 1. Summary of TE abundance by lineage across fig genotypes

					Min - Max	Mean	CV (%)
Total repeats					24.08 - 35.44	31.62	4.75
Class I				Total	17.89 - 26.45	23.35	5.11
	LTR-retrotransposons			Total	17.81 - 26.32	23.24	5.11
		Copia		Total	3.24 - 4.88	4.27	4.92
				Ale	1.09 - 1.57	1.43	4.63
				Alesia	0.005 - 0.009	0.007	14.29
				Angela	0.93 - 1.45	1.23	6.35
				Bianca	0.06 - 0.1	0.08	6.41
				Ikeros	0.3 - 0.46	0.39	5.36
				Ivana	0.04 - 0.07	0.06	8.33
				SIRE	0.03 - 0.05	0.04	7.14
				TAR	0.6 - 0.93	0.81	5.55
				Tork	0.18 - 0.26	0.23	4.80
		Gypsy		Total	14.57 - 21.44	18.97	5.26
			Chromovirus	Total	10.17 - 15.38	13.49	5.49
				CRM	2.65 - 4.69	3.70	7.41
				Galadriel	0.04 - 0.13	0.07	20.27
				Reina	0.010 - 0.016	0.013	7.69
				Tekay	7.47 - 11.08	9.71	5.56
			non-Chromovirus	Total	4.39 - 6.42	5.48	5.37
				Athila	2.24 - 3.53	2.96	5.45
				Tat/Retand	2.03 - 3.2	2.52	7.92
	non-LTR			Total	0.08 - 0.13	0.10	7.66
				LINE	0.07 - 0.12	0.09	7.69
				SINE	0.01 - 0.014	0.012	8.33
Class II				Total	6.2 - 9.29	8.27	4.48
	Subclass 1		TIR	Total	1.94 - 2.91	2.52	5.50
				EnSpm/CACTA	1.29 - 1.99	1.69	6.03
				hAT	0.13 - 0.21	0.16	7.41
				MuDR/Mutator	0.37 - 0.7	0.51	9.77
				PIF/Harbinger	0.12 - 0.19	0.15	7.89
	Subclass 2			Helitron	4.25 - 6.46	5.75	4.76

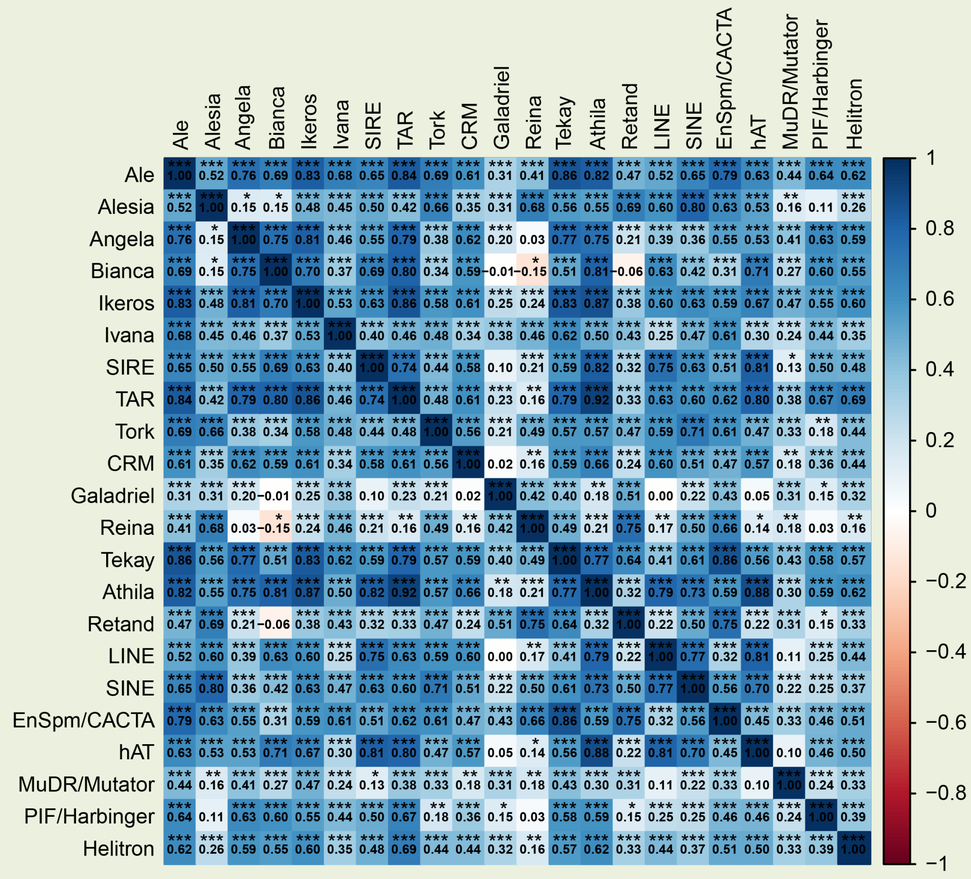

Figure 4. Pearson correlation heatmap among repetitive element classes. Statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Presence/absence variations (PAVs)

Methods

Paired-end WGR reads from 286 fig genotypes were mapped using Bowtie2 (Langmead and Salzberg, 2012) to the fig reference genome (Usai et al., 2025). TE presence variants were detected using SplitReader (Baduel et al., 2021), whereas TE absence variants were identified using TEPID (Stuart et al., 2016). Insertion age was estimated with PLINK 2 (Chang et al 2015) from SNP accumulation within 5-kb windows flanking each TE PAV, and divergence values were divided by the fig substitution rate per year of 2.54×10^{-9}.

Results

23,414 TE insertion variants and 561 TE deletion variants relative to the reference genome were identified. PAVs associated with LTR retrotransposons account for over 95% of the total, with a marked prevalence of Gypsy elements (75%), especially belonging to the Tekay lineage (43%). DNA TEs account for a smaller fraction of the PAVs, mostly represented by elements of the CACTA family (4%) (**Figure 5**). The inferred ages of TE PAVs ranged up to 35 MY, with a mean age of 5.7 MY; 1.3% of dated events occurred within the last 10,000 years (**Figure 6**).

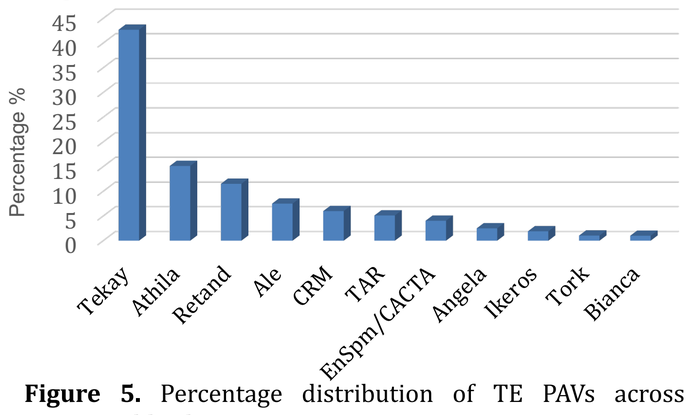

Figure 5. Percentage distribution of TE PAVs across transposable element groups.

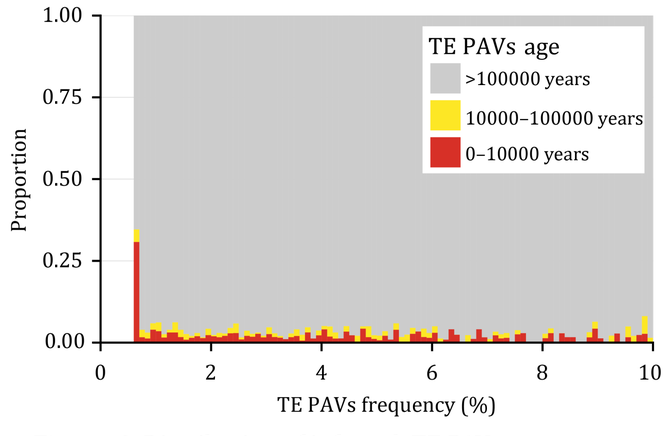

Figure 6. Distribution of inferred TE PAVs ages.

Conclusions

Overall, this study provides the first comprehensive population-level analysis of TE abundance variation in fig and establishes a framework for investigating genetic control of TE dynamics in a perennial, clonally propagated crop.

Aknowledgements

This study was conducted within the framework of the AGROFIG/PRIMA24_00100 project, which is part of the PRIMA Programme supported by the European Union through a national MUR (Italy) grant.

References

Baduel et al. (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. Genome biology, 22(1), 138. **Castellacci et al. (2026).** Exploring the genetic diversity of Mediterranean fig trees highlights genes associated with fruit traits. Frontiers in Plant Science, 17, 1750632. **Chang et al. (2015).** Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 4(1), s13742-015. **Chen et al. (2018).** fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), i884-i890. **Langmead & Salzberg (2012).** Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4):357-359. **Li and Durbin (2009).** Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 25(14), 1754-1760. **Olivoto and Lúcio (2020).** metan: An R package for multi-environment trial analysis. Methods in Ecology and Evolution, 11(6), 783-789. **Stuart et al. (2016).** Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. elife, 5, e20777. **Usai et al. (2025).** Haplotype-resolved genome assembly of Ficus carica L. reveals allele-specific expression in the fruit. The Plant Journal, 121(4), e70012. **Wei et al. (2017).** Package 'corrplot'. Statistician, 56(316), e24.