UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Plant Molecular and Cellular Biology Joint Research Institute (IBMCP)

Study of the evolution of the transposable element space in plant genomes.

Master's Thesis

Master's Degree in Plant Molecular and Cellular Biotechnology

AUTHOR: Zhu, Yujie

Tutor: Mulet Salort, José Miguel

Experimental director: BOMBARELY GOMEZ, AURELIANO

ACADEMIC YEAR: 2023/2024

Contents

Introduction

Transposable Elements (TEs) are DNA sequences capable of changing their position within the genome (Bourque et al., 2018). Once considered "junk DNA," the availability of genome sequences and the growth of genomic databases have accelerated the study of TEs, now recognized as evolutionary features (Ramakrishnan et al., 2022).

Based on their transposition mechanisms, TEs are classified into two major classes:

Class I: RNA transposons, also known as Retrotransposons, move through a "copy-paste" mechanism. Initially, they are transcribed from DNA to RNA, and the resulting RNA intermediate is reverse-transcribed into a cDNA copy, which is then inserted into a new position in the genome using TE enzyme machinery (Bourque et al., 2018).

Class II: DNA transposons move directly via a "cut-paste" mechanism without an RNA intermediate. They insert into new genomic locations through a DNA intermediate (Muñoz-López & García-Pérez, 2010). Given that genome size, complexity, and instability are often associated with TE copy number, RNA transposons play a more significant role (Wicker et al., 2007).

These two major classes of transposons are further classified into subclasses based on their chromosomal integration mechanisms. The subclasses for RNA transposons include LTR, DIRS, and Non-LTR. Long Terminal Repeats (LTR) retrotransposons integrate into the genome through an integration enzyme-catalyzed cleavage and strand transfer process, similar to retroviruses (Brown et al., 1987). In contrast, non-LTR integration occurs via a process called target-primed reverse transcription and reverse transcription coupling (Luan et al., 1993).

Each subclass is further divided into superfamilies, with LTR's superfamilies being Copia, Gypsy, and ERV, and Non-LTR's Superfamilies being LINE, SINE, and PLE. These superfamilies are often present across diverse organisms but share common genetic organization and monophyletic origins (Bourque et al., 2018). The two primary superfamilies of LTR retrotransposons, Copia and Gypsy, are found in nearly all major eukaryotic lineages (Malik & Eickbush, 2001). Plant SINEs are derived from tRNAs, which hold a special position, as these small non-coding and non-autonomous elements of several hundred base pairs utilize the transposition mechanism of LINEs to ensure their amplification (Mhiri et al. 2022). Each superfamily is further divided into families. In principle, each TE sequence in the genome can be assigned to a family, superfamily, subclass, and class (Bourque et al., 2018).

Transposable elements play significant roles in organismal evolution and genomic dynamics, with their importance in the evolutionary process attributable to several key properties:

- Genomic Plasticity and Diversity: TEs induce changes and reshaping of the genome structure by mobilizing to new positions within the genome or influencing gene expression. Such genomic alterations provide genetic diversity, which is crucial for the evolutionary process.

- Functional Innovation of Genes: TEs possess diverse structures and functions, for instance, in plants, they can act as promoters and enhancers (Ramakrishnan et al., 2022). When TEs are inserted into genes or their regulatory regions, they can alter gene functions or expression patterns, occasionally leading to the emergence of novel functions, thereby equipping organisms with adaptability to new environments or lifestyles.

- Genome Size and Evolution: The vast and varying sizes of genomes are primarily due to the proliferation of TEs (Haley & Mueller, 2022). TEs are also considered significant contributors to other mechanisms like recombination rates and polyploidy (Mhiri et al., 2022). Transposition is an efficient mechanism for genome expansion, which is counterbalanced over time by DNA removal. The balance between these processes is a driving force for the evolution of eukaryotic genome sizes (Schubert & Vu, 2016).

- Source of Genetic Variability: TEs serve as extensive sources of genetic and hereditary polymorphism. TEs occupy a considerable portion of a species' genome, including a significant portion unique to that species (Bourque et al., 2018). The insertion and activation of TEs can trigger genetic variations, laying the groundwork for the effects of natural selection and influencing population adaptability.

- Genomic Stability and Safeguarding: To mitigate potential detrimental consequences from excessive transposition, evolutionary processes have developed various epigenetic mechanisms to silence and control TEs, such as small RNAs, KRAB domain-containing zinc finger proteins, DNA methylation, histone modifications, splicing inhibition, and RNA modifications (Almeida et al., 2022). TEs are influenced by multiple regulatory sources, including the TEs themselves (regulatory motifs, biology) and host plant features (epigenetic control, genome size, ploidy level, sequence elimination mechanisms, and reproductive systems). The collective interplay of these evolutionary forces might lead to a state of equilibrium (Mhiri et al., 2022).

To better understand the landscape and dynamics of TEs within genomes, sophisticated computational tools have been developed. In this study, we employed a comprehensive approach utilizing tools such as RepeatModeler, RepeatMasker, and TEsorter for TE re-annotation. These tools facilitate the identification, classification, and masking of repetitive elements, offering insights into their distribution, diversity, and potential functional implications within the genome.

Given the intricate nature and profound impact of Transposable Elements on genomic evolution, these dynamic DNA segments have emerged as focal points of research across various biological disciplines. As our understanding of TEs deepens, it becomes increasingly evident that vast territories of the TE landscape remain uncharted, brimming with potential insights. Recognizing the pressing need to delve further into this captivating realm and bridge existing knowledge gaps, the present study embarks on a rigorous exploration. To unravel the complexities and nuances of TE dynamics across plant lineages, we articulate the following hypotheses:

- *Hypothesis 1*: The most divergent lineages harbor the most unique TE landscape. Given the vast evolutionary distances and distinct ecological niches occupied by different plant lineages, it is plausible to speculate that their TE profiles have evolved in unique directions, leading to distinct TE compositions and activities.

- *Hypothesis 2*: There exists a correlation between genome size and the TE landscape across plant species. Larger genomes may offer more "space" for TE insertions, potentially influencing gene density and spatial organization. Consequently, this may affect the accumulation and distribution patterns of TEs within these genomes.

- *Hypothesis 3*: Rapidly evolving plant lineages may exhibit heightened TE activity or mutation rates. In comparison to more primitive lineages like Chlorophytes, advanced groups like Angiosperms might manifest increased TE diversity and variability, potentially reflecting adaptive responses to diverse environmental pressures.

Objectives

Transposable elements, dynamic segments of DNA within the genome, play significant roles in shaping genomic composition through processes like gene disruption and recombination rate modulation. Our hypothesis posits varying TE diversity among different plant species, confined within specific ranges tied to distinct evolutionary clades. The project aims to explore this diversity across plant lineages, spanning from algae to angiosperms. The selected 138 plant genomes, representing diverse lineages (Chlorophytes, Charophytes, Liverworts, Mosses, Hornworts, Lycophytes, Ferns, Gymnosperms, and Angiosperms), will undergo TE re-annotation using advanced tools such as RepeatModeler2, RepeatMasker, and TEsorter. This comprehensive re-annotation will offer an updated understanding of TE landscapes within each genome.

The overarching goal of the project is to unravel the diversity and specificity parameters of TE landscapes, contributing to our understanding of how TEs have influenced genomic evolution across diverse plant lineages. This research holds significance in shedding light on the functional implications of TEs and their roles in genomic processes, thus contributing to the broader fields of genomics, evolutionary biology, and plant science.

The specific goals of this project are:

- Determine the speciation and diversity of the transposable elements within 138 plant species along the plant tree of life.

- Compare the transposable landscape of these 138 species using common data visualization tools such as boxplots, heat maps, and Principal Component Analysis (PCA).

- Validate the hypothesis that the most divergent lineages possess the most unique TE landscape.

- Investigate the correlation between genome size and the TE landscape across plant species.

- Assess the potential functional implications of observed transposable element variations on specific genomic features or evolutionary traits.

- Provide insights into the evolutionary significance of transposable elements in shaping plant genome architecture and function.

Materials & Methods

1. Input Data Acquisition

To comprehensively explore the landscape of transposable elements in plant genomes, a diverse dataset was curated. Genomic data from 42 Chlorophytes were downloaded from NCBI using the "curl" command-line tool.

SPECIES	LINEAGE	SOURCE	ASSEMBLY
Astrephomene gubernaculifera	Chlorophyte	NCBI	Astre_guber_v1.0
Auxenochlorella protothecoides (0710)	Chlorophyte	NCBI	ASM73321v1
Auxenochlorella protothecoides (UTEX 25)	Chlorophyte	NCBI	ASM370936v1
Bathycoccus prasinos	Chlorophyte	NCBI	ASM222023v1
Chlamydomonas eustigma	Chlorophyte	NCBI	C.eustigma genome v1.0
Chlamydomonas incerta	Chlorophyte	NCBI	ASM1683460v1
Chlamydomonas reinhardtii	Chlorophyte	NCBI	Chlamydomonas_reinhardtii_v5.5
Chlamydomonas schloesseri	Chlorophyte	NCBI	ASM1683459v1
Chlamydomonas sp. UWO 241	Chlorophyte	NCBI	CUWO241_v1.0_nuclear
Chlorella desiccata (nom. nud.) (UTEX 2437)	Chlorophyte	NCBI	LANL_v2
Chlorella desiccata (nom. nud.) (UTEX 2526)	Chlorophyte	NCBI	LANL_Cdes_v1.1
Chlorella ohadii	Chlorophyte	NCBI	ASM2502687v1
Chlorella sorokiniana	Chlorophyte	NCBI	Chlorella_sorokiniana 2.0
Chlorella variabilis	Chlorophyte	NCBI	v 1.0
Chloropicon primus (CCMP 1205)	Chlorophyte	NCBI	ASM785969v1
Chloropicon primus (RCC138)	Chlorophyte	NCBI	ASM2320587v1
Coccomyxa sp. Obi	Chlorophyte	NCBI	COCOBI_1.0
Coccomyxa subellipsoidea C-169	Chlorophyte	NCBI	Coccomyxa subellipsoidae v2.0
Dunaliella salina	Chlorophyte	NCBI	Dunsal1 v. 2
Edaphochlamys debaryana	Chlorophyte	NCBI	ASM1685814v1
Haematococcus lacustris	Chlorophyte	NCBI	Lacustris_1.0
Helicosporidium sp. ATCC 50920	Chlorophyte	NCBI	Helico_v1.0
Micractinium conductrix	Chlorophyte	NCBI	ASM224581v2
Micromonas commoda	Chlorophyte	NCBI	ASM9098v2
Micromonas pusilla CCMP1545	Chlorophyte	NCBI	Micromonas pusilla CCMP1545 v2.0
Monoraphidium minutum	Chlorophyte	NCBI	Monmin1
Monoraphidium neglectum	Chlorophyte	NCBI	mono_v1
Ostreobium quekettii	Chlorophyte	NCBI	Ostreobium_1D_genome
Ostreococcus lucimarinu	Chlorophyte	NCBI	ASM9206v1
Ostreococcus tauri (RCC1115)	Chlorophyte	NCBI	Ostta1115_2
Ostreococcus tauri (RCC4221)	Chlorophyte	NCBI	version 140606
Pedinophyceae sp. YPF-701	Chlorophyte	NCBI	Pedinophyceae_YPF-701_genome
Picochlorum sp. BPE23	Chlorophyte	NCBI	ASM2520934v1
Picochlorum sp. BPE23	Chlorophyte	NCBI	ASM2520937v1
Pycnococcus provasolii	Chlorophyte	NCBI	Ppro_1.0
Raphidocelis subcapitata	Chlorophyte	NCBI	Rsub_1.0
Scenedesmus sp. NREL 46B-D3	Chlorophyte	NCBI	Scesp_1
Tetrabaena socialis	Chlorophyte	NCBI	TetSoc1
Trebouxia sp. A1-2	Chlorophyte	NCBI	ASM863618v1
Volvox africanus	Chlorophyte	NCBI	Vafri_1.0
Volvox reticuliferus (NIES 3786)	Chlorophyte	NCBI	Vretimale_1.0
Volvox reticuliferus (NIES-3785)	Chlorophyte	NCBI	Vretifemale_1.0

Genomic data from 96 plant species representing various lineages, including Charophyte (1), Charales (1) Liverworts (2), Mosses (5), Hornworts (1), Lycophytes (4), Ferns (2), Gymnosperms (5), and Angiosperms (75), were downloaded from renowned repositories, including NCBI, CNCBI, owned databases, and SGN.

SPECIES	LINEAGE	SOURCE	ASSEMBLY
Acer_negundo	Angiosperm (Eudicots)	NCBI	GCA_025594385.1_ASM2559438v1_genomic.fna
Acer_yangbiense	Angiosperm (Eudicots)	NCBI	GCA_008009225.1_AYv1.1_genomic.fna
Actinidia_chinensis	Angiosperm (Eudicots)	NCBI	GCA_003024255.1_Red5_PS1_1.69.0_genomic.fna
Adansonia_digitata	Angiosperm (Eudicots)	NCBI	GCA_029448705.1_ASM2944870v1_genomic.fna
Adiantum_capillus	Fern	NCBI	GCA_014529385.2_ASM1452938v2_genomic.fna
Amborella_trichopoda	Angiosperm (Basal)	NCBI	GCF_000471905.2_AMTR1.0_genomic.fna
Ananas_comosus	Angiosperm (Monocots)	NCBI	GCF_001540865.1_ASM154086v1_genomic.fna
Annona_glabra	Angiosperm (Magnolids)	CNCB	GWHBCKB00000000_Annona.genome.fna
Anthoceros_angustus	Hornwort	NCBI	GCA_010909165.1_ASM1090916v1_genomic.fna
Arabidopsis_arenosa	Angiosperm (Eudicots)	NCBI	GCA_905216605.1_AARE701a_genomic.fna
Arabidopsis_suecica	Angiosperm (Eudicots)	NCBI	GCA_019202805.1_ASM1920280v1_genomic.fna
Arabidopsis_thaliana	Angiosperm (Eudicots)	NCBI	GCF_000001735.4_TAIR10.1_genomic.fna
Arabis_alpina	Angiosperm (Eudicots)	NCBI	GCA_900128785.1_MPIPZ.v5_genomic.fna
Boechera_stricta	Angiosperm (Eudicots)	NCBI	GCA_018361395.1_NTU_Bstr_SAD12_2.2_genomic.fna
Brachypodium_distachyon	Angiosperm (Monocots)	NCBI	GCF_000005505.3_Brachypodium_distachyon_v3.0_genomic.fna
Brassica_carinata	Angiosperm (Eudicots)	NCBI	GCA_016771965.1_ASM1677196v1_genomic.fna
Brassica_juncea	Angiosperm (Eudicots)	NCBI	GCA_018703725.1_ASM1870372v1_genomic.fna
Brassica_napus	Angiosperm (Eudicots)	NCBI	GCF_020379485.1_Da-Ae_genomic.fna
Brassica_nigra	Angiosperm (Eudicots)	NCBI	GCA_016432835.1_Bnig_sang_1.1_genomic.fna
Brassica_oleracea	Angiosperm (Eudicots)	NCBI	GCF_000695525.1_BOL_genomic.fna
Brassica_rapa	Angiosperm (Eudicots)	NCBI	GCF_000309985.2_CAAS_Brap_v3.01_genomic.fna
Camelina_sativa	Angiosperm (Eudicots)	NCBI	GCF_000633955.1_Cs_genomic.fna
Camellia_sinensis	Angiosperm (Eudicots)	NCBI	GCF_004153795.1_AHAU_CSS_1_genomic.fna
Capsicum_annuum	Angiosperm (Eudicots)	NCBI	GCF_002878395.1_UCD10Xv1.1_genomic.fna
Carica_papaya	Angiosperm (Eudicots)	NCBI	GCF_000150535.2_Papaya1.0_genomic.fna
Ceratodon_purpureus	Moss	NCBI	GCA_014871385.1_CpurpureusR40_1_0_genomic.fna
Ceratopteris_richardii	Fern	NCBI	GCA_020310875.1_C.richardii_v2_genomic.fna
Chara_braunii	Algae (Charales)	NCBI	GCA_003427395.1_Cbr_1.0_genomic.fna
Chlamydomonas_reinhardtii	Algae (Chlorophyta)	NCBI	GCF_000002595.2_Chlamydomonas_reinhardtii_v5.5_genomic.fna
Cinnamomum_micranthum	Angiosperm (Magnolids)	NCBI	GCA_003546025.1_ASBRC_Ckan_1.0_genomic.fna
Citrullus_lanatus	Angiosperm (Eudicots)	NCBI	GCA_029034555.1_KOR._cv.242-1_genomic.fna
Citrus_sinensis	Angiosperm (Eudicots)	NCBI	GCF_022201045.2_DVS_A1.0_genomic.fna
Cucumis_melo	Angiosperm (Eudicots)	NCBI	GCF_025177605.1_USDA_Cmelo_AY_1.0_genomic.fna
Cucumis_sativus	Angiosperm (Eudicots)	NCBI	GCF_000004075.3_Cucumber_9930_V3_genomic.fna
Cycas_panzhihuaensis	Gymnosperm	NCBI	GCA_023213395.1_ASM2321339v1_genomic.fna
Cydonia_oblonga	Angiosperm (Eudicots)	NCBI	GCA_015708375.1_ASM1570837v1_genomic.fna
Cynara_cardunculus	Angiosperm (Eudicots)	NCBI	GCF_001531365.2_CcrdV1.1_genomic.fna
Cyperus_esculentus	Angiosperm (Monocots)	NCBI	GCA_030179965.1_Ces_v0.1_genomic.fna
Datura_stramonium	Angiosperm (Eudicots)	CNCB	GWHBKBC00000000_Datura.genome.fna
Daucus_carota	Angiosperm (Eudicots)	NCBI	GCF_001625215.1_ASM162521v1_genomic.fna
Dendrobium_officinale	Angiosperm (Monocots)	NCBI	GCA_019514585.1_ASM1951458v1_genomic.fna
Fragaria_ananassa	Angiosperm (Eudicots)	NCBI	GCA_019022445.1_NIHHS_FaW_1.0_genomic.fna
Fragaria_vesca	Angiosperm (Eudicots)	NCBI	GCF_000184155.1_FraVesHawaii_1.0_genomic.fna
Ginkgo_biloba	Gymnosperm	NCBI	GCA_024626585.1_ASM2462658v1_genomic.fna
Glycine_max	Angiosperm (Eudicots)	NCBI	GCF_000004515.6_Glycine_max_v4.0_genomic.fna
Glycine_soja	Angiosperm (Eudicots)	NCBI	GCF_004193775.1_ASM419377v2_genomic.fna
Gnetum_montanum	Gymnosperm	NCBI	GCA_015680685.1_Gmon01_genomic.fna
Gossypium_arboreum	Angiosperm (Eudicots)	NCBI	GCF_025698485.1_ASM2569848v2_genomic.fna
Gossypium_hirsutum	Angiosperm (Eudicots)	NCBI	GCF_007990345.1_Gossypium_hirsutum_v2.1_genomic.fna
Gossypium_raimondii	Angiosperm (Eudicots)	NCBI	GCF_025698545.1_ASM2569854v1_genomic.fna
Ipomoea_batatas	Angiosperm (Eudicots)	NCBI	GCA_002525835.2_ipoBat4_genomic.fna
Isoetes_engelmannii	Lycopods	NCBI	GCA_011763485.2_ASM1176348v2_genomic.fna
Isoetes_taiwanensis	Lycopods	NCBI	GCA_021234155.1_ASM2123415v1_genomic.fna
Isothecium_myosuroides	Moss	NCBI	GCA_951799445.1_cbIsoMyos1.1_genomic.fna
Lactuca_sativa	Angiosperm (Eudicots)	NCBI	GCF_002870075.4_Lsat_Salinas_v11_genomic.fna
Lemna_minuta	Angiosperm (Monocots)	NCBI	GCA_024174645.1_Salk_lm5633_a03_genomic.fna
Lycium_ferocissimum	Angiosperm (Eudicots)	NCBI	GCA_029784015.1_AGI_CSIRO_Lferr_CH_V1_genomic.fna
Lycopodium_clavatum	Lycopods	CNCB	GWHBJYW00000000_Lycopodium.genome.fna
Malus_domestica	Angiosperm (Eudicots)	NCBI	GCF_002114115.1_ASM211411v1_genomic.fna
Mangifera_indica	Angiosperm (Eudicots)	NCBI	GCF_011075055.1_CATAS_Mindica_2.1_genomic.fna
Marchantia_paleacea	Liverworts	NCBI	GCA_014180765.2_ASM1418076v2_genomic.fna
Marchantia_polymorpha	Liverworts	NCBI	GCA_003032435.1_Marchanta_polymorpha_v1_genomic.fna
Medicago_truncatula	Angiosperm (Eudicots)	NCBI	GCF_003473485.1_MtrunA17r5.0-ANR_genomic.fna
Musa_acuminata	Angiosperm (Monocots)	NCBI	GCF_000313855.2_ASM31385v2_genomic.fna
Nelumbo_nucifera	Angiosperm (Eudicots)	NCBI	GCF_000365185.1_Chinese_Lotus_1.1_genomic.fna
Nicotiana_benthamiana	Angiosperm (Eudicots)	owned	Niben261_genome_complete01.fna
Nymphaea_colorata	Angiosperm (Basal)	NCBI	GCF_008831285.2_ASM883128v2_genomic.fna
Olea_europaea	Angiosperm (Eudicots)	NCBI	GCA_902713445.1_OLEA9_genomic.fna
Oryza_sativa	Angiosperm (Monocots)	NCBI	GCF_001433935.1_IRGSP-1.0_genomic.fna.gz
Phalaenopsis_equestris	Angiosperm (Monocots)	NCBI	GCF_001263595.1_ASM126359v1_genomic.fna
Phaseolus_vulgaris	Angiosperm (Eudicots)	NCBI	GCF_000499845.1_PhaVulg1_0_genomic.fna
Phoenix_dactylifera	Angiosperm (Monocots)	NCBI	GCF_009389715.1_palm_55x_up_171113_PBpolish2nd_filt_p_genomic.fna
Physalis_pubescens	Angiosperm (Eudicots)	CNCB	GWHANUX00000000_Physalis.genome.fna
Physcomitrium_patens	Moss	NCBI	GCF_000002425.4_Phypa_V3_genomic.fna
Pistacia_vera	Angiosperm (Eudicots)	NCBI	GCF_008641045.1_PisVer_v2_genomic.fna
Populus_trichocarpa	Angiosperm (Eudicots)	NCBI	GCF_000002775.5_P.trichocarpa_v4.1_genomic.fna
Prunus_persica	Angiosperm (Eudicots)	NCBI	GCF_000346465.2_Prunus_persica_NCBIv2_genomic.fna
Punica_granatum	Angiosperm (Eudicots)	NCBI	GCF_007655135.1_ASM765513v2_genomic.fna
Salvia_hispanica	Angiosperm (Eudicots)	NCBI	GCF_023119035.1_UniMelb_Shisp_WGS_1.0_genomic.fna
Salvia_splendens	Angiosperm (Eudicots)	NCBI	GCF_004379255.2_SspV2_genomic.fna
Selaginella_moellendorffii	Lycopods	NCBI	GCF_000143415.4_v1.0_genomic.fna
Setaria_italica	Angiosperm (Monocots)	NCBI	GCF_000263155.2_Setaria_italica_v2.0_genomic.fna
Solanum_lycopersicum	Angiosperm (Eudicots)	NCBI	GCF_000188115.5_SL3.1_genomic.fna
Solanum_melongena	Angiosperm (Eudicots)	SGN	Eggplant_SGN_V4.1.fna
Solanum_pimpinellifolium	Angiosperm (Eudicots)	NCBI	GCA_014964335.1_ASM1496433v1_genomic.fna
Solanum_stenotomum	Angiosperm (Eudicots)	NCBI	GCF_019186545.1_ASM1918654v1_genomic.fna
Solanum_tuberosum	Angiosperm (Eudicots)	NCBI	GCF_000226075.1_SolTub_3.0_genomic.fna
Sphagnum_fallax	Moss	NCBI	GCA_021442195.1_S.fallax_v1.1_genomic.fna
Sphagnum_magellanicum	Moss	NCBI	GCA_021904315.1_S.magellanicum_v1.1_genomic.fna
Thuja_plicata	Gymnosperm	NCBI	GCA_018584345.1_redcedar-v3_genomic.fna
Utricularia_gibba	Angiosperm (Eudicots)	NCBI	GCA_002189035.1_U_gibba_v2_genomic.fna
Vanilla_planifolia	Angiosperm (Monocots)	NCBI	GCA_023846275.1_ASM2384627v1_genomic.fna
Vicia_sativa	Angiosperm (Eudicots)	NCBI	GCA_021764765.1_ASM2176476v1_genomic.fna
Vitis_vinifera	Angiosperm (Eudicots)	NCBI	GCF_000003745.3_12X_genomic.fna
Welwitschia_mirabilis	Gymnosperm	CNCB	CNA0022760_Wmirabilis_genomic.fna
Zea_mays	Angiosperm (Monocots)	NCBI	GCF_902167145.1_Zm-B73-REFERENCE-NAM-5.0_genomic.fna

2. Genome Metrics Retrieval

To retrieve the genome metrics two programs were used.

2.1 Quast v5.10: Assessing Genome Quality

Quast, the Quality Assessment Tool for Genome Assemblies, was significant in providing a comprehensive evaluation of the quality and completeness of the assembled genomes. Quast v5.10 (Mikheenko et al., 2018) was employed to extract essential genome metrics, including Assembly Size and Number of Contigs.

```
# Activate the conda environment with Quast installed
conda activate quast_env

# Run Quast
quast -o <out_directory_name> -m 1 -g <genomic_gff> <assembly_fasta>
```

2.2 agat_sp_statistics.pl: Detailed Genome Metrics

Additionally, agat_sp_statistics.pl v1.2.0 (Dainat et al., 2023) was employed for more detailed genome metrics retrieval, including providing the Number of Genes, contributing to a comprehensive understanding of the assembled genomes.

```
# Run agat_sp_statistics.pl
agat_sp_statistics.pl -gff <genomic_gff> -o statistics.txt
```

This dual-program approach ensured a thorough assessment of the genomic landscape, laying the groundwork for subsequent analyses.

3. Repetitive Element Identification

The intricate world of repetitive elements was unraveled through a two-fold process utilizing advanced tools.

3.1 RepeatModeler2: Unveiling the Unknown Repetitive Landscape

RepeatModeler2 v2.0.4 (Flynn et al., 2020) played a significant role in identifying unknown repetitive elements within the genomic sequences. This tool employed a de novo approach, discovering repetitive motifs without relying on pre-existing libraries. RepeatModeler's primary goal was to comprehensively analyze the entire genome, automating the identification, modeling, and classification of various types of repetitive elements. Its output was a library containing identified repetitive element families. Of particular significance was its capability to unearth Long Terminal Repeats (LTRs), shedding light on crucial genomic regions characterized by repetitive dynamics.

```
# Start a Docker instance for DFAM tools
screen -L -Logfile RunDFAMToolsDATE.log /data/Software/dfam-tetools.sh

# Generate the RepeatModeler sequence database
BuildDatabase --<name> <my_genome> <my_genome.fasta>

# Run RepeatModeler2
RepeatModeler -database <my_genome> -threads 4 -LTRStruct
```

3.2 RepeatMasker: Annotating the Repetitive Landscape

Building upon the insights gained from RepeatModeler2, RepeatMasker v4.1.5 (A.F.A. Smit et al., RepeatMasker at http://repeatmasker.org) enriched the annotation process, unraveling the repetitive landscape encoded within the genome. This tool annotated the identified repetitive elements in the genome FASTA sequence using the DFAM database and the elements found by RepeatModeler2. RepeatMasker's focus was to mark the positions of repetitive elements in the genome sequence using a known repetitive element library (often generated by RepeatModeler).

```
# Run RepeatMasker
RepeatMasker -lib <my_genome-families.fa> -xsmall -gff -pa 4
my_genome.fasta

# Exit from the Docker instance
exit;
```

RepeatMasker not only annotated repetitive elements but also provided crucial information on their genomic locations and possible families. This tandem approach facilitated a comprehensive exploration of the repetitive elements' distribution and composition within the genomes.

3.3 TEsorter: Sorting and Categorizing Repetitive Elements

Following RepeatMasker, TEsorter v1.4.6 (Zhang et al., 2022) was employed to sort individual repeats, providing a refined and categorized view of the repetitive landscape.

```
# Extract the repetitive elements previously annotated by RepeatMasker
python3 /data/users/collaborators/zhuy/miniconda3/bin/RepeatMasker.py
out2seqs <my_genome.fasta.out_from_RepeatMasker> <my_genome.fasta> >
<my_genome.IndividualRepeats.fasta>

# Run TEsorter
TEsorter <my_genome.IndividualRepeats.fasta> -p 48
```

4. Comparative Analysis of Transposable Elements

The comprehensive re-annotation of TE landscapes extended beyond identification to estimating diversity parameters and specificity. Diversity parameters, including Shannon's entropy (Hj) and Kolmogorov complexity (KC), were calculated, gauging variations within TE landscapes. Specificity, assessed through the specialization index (δj index) and divergence with respect to the entire TE landscape using Kullback–Leibler divergence (Divj), offered insights into TE behavior.

```
# Estimation of Diversity Parameters
python
~/trabajo/diversity_tools/diversity_tools/shannon_index_operations.py
-i <GenomesAnalysis_count_matrix.csv> -o diversity -f <olivia_results>

# Estimation of Specificity Parameters
python
~/trabajo/diversity_tools/diversity_tools/shannon_index_operations.py
-i    <GenomesAnalysis_count_matrix.csv>    -o    specificity    -f
<olivia_results>
```

An R script was developed to create horizontal stacked bar charts depicting the diversity and specificity patterns across different plant species. The ggplot2 package v3.4.4 (Wickham et al., 2022) was utilized for its versatile features and customization options.

```
library(ggplot2)
# Create a Horizontal Stacked Bar Chart of Diversity
horizontal_stacked_bar_chart_diversity                                <-
ggplot(Diversity_Specificity_of_species, aes(x = interaction(SPECIES,
LINEAGE), y = DIVERSITY, fill = LINEAGE)) +
  geom_bar(stat = "identity", position = "stack") +
  labs(title = "Horizontal Stacked Bar Chart of Diversity",
       x = "Species",
       y = "Diversity",
       fill = "Lineage") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 4),
        plot.title = element_text(hjust = 0.5),
        legend.position = "bottom") +
  scale_fill_viridis_d()
theme_set(theme_classic())
print(horizontal_stacked_bar_chart_diversity)
# Create a Horizontal Stacked Bar Chart of Specificity
horizontal_stacked_bar_chart                                          <-
ggplot(Diversity_Specificity_of_species, aes(x = interaction(SPECIES,
LINEAGE), y = SPECIFICITY, fill = LINEAGE)) +
```

```
  geom_bar(stat = "identity", position = "stack") +
  labs(title = "Horizontal Stacked Bar Chart of Specificity",
       x = "Species",
       y = "Specificity",
       fill = "Lineage") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 4),
        plot.title = element_text(hjust = 0.5),
        legend.position = "bottom") +
  scale_fill_viridis_d()
theme_set(theme_classic())
print(horizontal_stacked_bar_chart_specificity)
```

Subsequently, scatter plots were generated in R to visually depict the diversity and specificity patterns across various TE superfamilies.

```
# Sort the data frame by diversity in ascending order when creating it
Diversity_Specificity_of_TEs <-
Diversity_Specificity_of_TEs[order(Diversity_Specificity_of_TEs$DIVER
SITY), ]
# Creat Scatter Plot of Diversity
scatter_plot <- ggplot(Diversity_Specificity_of_TEs, aes(x =
SUPERFAMILY, y = DIVERSITY, color = SUPERFAMILY)) +
  geom_point() +
  ggtitle("Scatter Plot of Diversity") +
  labs(x = "Superfamily", y = "Diversity", color = "Superfamily") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "top",
        plot.title = element_text(hjust = 0.5))
print(scatter_plot_divercity)
# Creat Scatter Plot of Specificity
scatter_plot_specificity <- ggplot(Diversity_Specificity_of_TEs,
aes(x = SUPERFAMILY, y = SPECIFICITY, color = SUPERFAMILY)) +
  geom_point() +
  ggtitle("Scatter Plot of Specificity") +
  labs(x = "Superfamily", y = "Specificity", color = "Superfamily") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "top",
        plot.title = element_text(hjust = 0.5))
print(scatter_plot_specificity)
```

To explore the correlation between Genome Size and Bases Masked in plant genomes, the following R script was utilized to generate a comprehensive scatter plot. This scatter

plot not only visually represents the distribution of data points but also includes an accompanying trend line. The trend line serves as a valuable tool, allowing for the identification of potential patterns or trends within the relationship between genome size and the extent of masked bases. This integrated visualization provides a nuanced understanding of the interplay between these two genomic parameters.

```
# Create scatter plot with trend line
ggplot(GenomeSize, aes(x = `GENOME SIZE (Mb)`, y = `BASES MASKED (Mb)`,
color = LINEAGE)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, linetype = "solid", color =
"grey") +
  labs(title = "The Correlation between Genome Size and Bases Masked",
       x = "Genome Size (Mb)",
       y = "Bases Masked (Mb)",
       color = "Lineage") +
  theme_classic() +
  theme(plot.title = element_text(hjust = 0.5))
```

The analysis also utilized the following programs:

https://github.com/AgustinAmata/Repeattools

Python programs were employed to analyze TEsorter results, generating visualizations such as box plots, heatmaps, and Principal Component Analysis (PCA) plots. These visualizations provided a holistic understanding of the dynamics of TEs within plant genomes.

This multi-tiered approach positioned the study to make substantial contributions to unraveling the complex interplay between TEs and plant genome evolution. The integration of advanced tools and custom analyses provided a nuanced view of the repetitive element landscape, offering insights into their roles in genomic evolution

Results

1. Repetitive Element Identification

To comprehensively understand the repetitive elements (RE) landscape within the plant genomes, a series of analyses were conducted. Initially, the program Quast v5.10 was employed to obtain the assembly size (genome size) and the number of contigs (number of scaffolds) for Chlorophytes. Subsequently, the program agat_sp_statistics.pl was employed to obtain the number of genes for Chlorophytes. Finally, RepeatModeler2 and RepeatMasker were executed, providing detailed metrics such as the bases masked and their corresponding percentages in the genomes, the summarized results are provided below:

SPECIES	LINEAGE	N_GENES	GENOME SIZE (Mb)	N_SCAFFOLDS	BASES MASKED (Mb)	PERCENTAGE
Astrephomene gubernaculifera	Chlorophyte	13724	103.9	207	29.5	28.44%
Auxenochlorella protothecoides (0710)	Chlorophyte	13724	103.9	374	29.6	28.52%
Auxenochlorella protothecoides (UTEX 25)	Chlorophyte	5852	21.2	217	0.7	3.29%
Bathycoccus prasinos	Chlorophyte	7941	15.1	21	1.0	6.90%
Chlamydomonas eustigma	Chlorophyte	14068	66.6	520	3.9	5.89%
Chlamydomonas incerta	Chlorophyte	16350	129.2	453	34.9	27.03%
Chlamydomonas reinhardtii	Chlorophyte	17742	111.1	53	23.2	20.91%
Chlamydomonas schloesseri	Chlorophyte	15571	130.2	457	39.4	30.27%
Chlamydomonas sp. UWO 241	Chlorophyte	16018	211.6	2458	109.4	51.71%
Chlorella desiccata (nom. nud.) (UTEX 2437)	Chlorophyte	8418	20.7	29	1.7	8.01%
Chlorella desiccata (nom. nud.) (UTEX 2526)	Chlorophyte	9308	21.6	18	1.7	8.04%
Chlorella ohadii	Chlorophyte	10866	57.1	486	3.5	6.13%
Chlorella sorokiniana	Chlorophyte	9526	59.6	159	6.5	10.90%
Chlorella variabilis	Chlorophyte	9780	46.2	414	5.2	11.27%
Chloropicon primus (CCMP 1205)	Chlorophyte	8639	17.4	20	1.4	8.24%
Chloropicon primus (RCC138)	Chlorophyte	8627	17.6	20	1.6	8.85%
Coccomyxa sp. Obi	Chlorophyte	11588	50.4	21	2.2	4.35%
Coccomyxa subellipsoidea C-169	Chlorophyte	9834	48.8	29	2.6	5.31%
Dunaliella salina	Chlorophyte	16527	343.7	5512	125.7	36.58%
Edaphochlamys debaryana	Chlorophyte	19227	142.1	527	26.4	18.59%
Haematococcus lacustris	Chlorophyte	28279	309.4	9693	159.6	51.58%
Helicosporidium sp. ATCC 50920	Chlorophyte	6033	12.4	5666	0.4	2.92%
Micractinium conductrix	Chlorophyte	9217	61.0	300	12.1	19.77%
Micromonas commoda	Chlorophyte	10147	21.1	19	1.1	5.22%
Micromonas pusilla CCMP1545	Chlorophyte	10238	22.0	21	3.5	15.42%
Monoraphidium minutum	Chlorophyte	15335	68.2	511	13.2	19.33%
Monoraphidium neglectum	Chlorophyte	16755	69.7	6720	6.0	8.55%
Ostreobium quekettii	Chlorophyte	10833	151.9	3134	53.8	35.42%
Ostreococcus lucimarinu	Chlorophyte	7603	13.2	21	1.2	9.04%
Ostreococcus tauri (RCC1115)	Chlorophyte	8099	14.8	70	1.7	11.46%
Ostreococcus tauri (RCC4221)	Chlorophyte	7869	13.0	22	0.8	5.81%
Pedinophyceae sp. YPF-701	Chlorophyte	7940	27.9	32	2.3	8.33%
Picochlorum sp. BPE23	Chlorophyte	7227	14.9	16	1.4	9.59%
Picochlorum sp. BPE23	Chlorophyte	7227	14.9	12	1.4	9.59%
Pycnococcus provasolii	Chlorophyte	11297	22.7	43	1.5	6.59%
Raphidocelis subcapitata	Chlorophyte	13383	51.2	300	6.6	12.88%
Scenedesmus sp. NREL 46B-D3	Chlorophyte	17398	151.9	2661	44.0	28.98%
Tetrabaena socialis	Chlorophyte	14296	135.8	5856	17.3	12.77%
Trebouxia sp. A1-2	Chlorophyte	13851	52.9	217	3.0	5.71%
Volvox africanus	Chlorophyte	13577	129.3	448	38.1	29.46%
Volvox reticuliferus (NIES 3786)	Chlorophyte	13772	134.0	230	39.8	29.72%
Volvox reticuliferus (NIES-3785)	Chlorophyte	13898	133.1	200	39.0	29.31%

Within the presented table, a comprehensive dataset related to Repetitive Element identification is provided. The table details key parameters for each genome, including the number of genes, genome size, number of scaffolds, bases masked, and their respective percentages. Based on the table:

- *Dunaliella salina* has the largest genome size (343.7 Mb).
- *Haematococcus lacustris* presents the highest number of genes, the highest number of scaffolds (9693), the highest number of bases masked (159.6 Mb), and the maximum percentage of bases masked (51.58%).
- *Helicosporidium sp. ATCC 50920* has the smallest genome size (12.4 Mb), the lowest number of bases masked (0.4 Mb), and the minimum percentage of bases masked (2.92%).
- *Auxenochlorella protothecoides (UTEX 25)* has the lowest number of genes (5852).

- *Picochlorum sp. BPE23* has the lowest number of scaffolds (12).

It is evident from the data that within the Chlorophyte lineage, there is a positive correlation between genome size and the relative content of Transposable Elements. For instance, Haematococcus lacustris, with a genome size of 309.4M, exhibits a TE content of 51.58%. Conversely, smaller genomes within this lineage tend to have lower TE content. For example, Ostreococcus tauri (RCC4221), having a genome size of 13.0M, has a TE content of 5.81%.

I continued the analyses using Repeat Modeler 2 and Repeat Masker until detailed data was obtained for all the studied plants, including the number of contigs, assembly size, bases masked, and their corresponding percentages. The results are presented below:

SPECIES	LINEAGE	N_CONTIGS	GENOME SIZE (Mb)	BASES MASKED (Mb)	PERCENTAGE
Acer_negundo	Angiosperm (Eudicots)	108	442.4	272.2	61.51%
Acer_yangbiense	Angiosperm (Eudicots)	280	665.9	472.0	70.88%
Actinidia_chinensis	Angiosperm (Eudicots)	1234	553.8	232.5	41.98%
Adansonia_digitata	Angiosperm (Eudicots)	232407	687.0	382.1	55.61%
Adiantum_capillus	Fern	601	4822.6	4118.5	85.40%
Amborella_trichopoda	Angiosperm (Basal)	5746	706.5	436.0	61.71%
Ananas_comosus	Angiosperm (Monocots)	3129	382.1	218.9	57.29%
Annona_glabra	Angiosperm (Magnolids)	66	1027.3	593.6	57.78%
Anthoceros_angustus	Hornwort	289	119.3	57.6	48.25%
Arabidopsis_arenosa	Angiosperm (Eudicots)	8	149.7	32.9	21.98%
Arabidopsis_suecica	Angiosperm (Eudicots)	269	272.3	67.6	24.83%
Arabidopsis_thaliana	Angiosperm (Eudicots)	7	119.7	20.4	17.05%
Arabis_alpina	Angiosperm (Eudicots)	8	311.6	166.1	53.30%
Boechera_stricta	Angiosperm (Eudicots)	22	190.5	70.7	37.12%
Brachypodium_distachyon	Angiosperm (Monocots)	11	271.3	101.0	37.22%
Brassica_carinata	Angiosperm (Eudicots)	1636	1087.0	653.6	60.13%
Brassica_juncea	Angiosperm (Eudicots)	1544	933.5	501.4	53.71%
Brassica_napus	Angiosperm (Eudicots)	3167	1001.9	575.2	57.41%
Brassica_nigra	Angiosperm (Eudicots)	1054	534.2	308.6	57.77%
Brassica_oleracea	Angiosperm (Eudicots)	32886	489.0	236.1	48.30%
Brassica_rapa	Angiosperm (Eudicots)	1100	353.0	170.4	48.27%
Camelina_sativa	Angiosperm (Eudicots)	37212	641.4	255.2	39.78%
Camellia_sinensis	Angiosperm (Eudicots)	14028	3105.4	2267.0	73.00%
Capsicum_annuum	Angiosperm (Eudicots)	81202	3212.5	2612.7	81.33%
Carica_papaya	Angiosperm (Eudicots)	17766	370.4	149.8	40.45%
Ceratodon_purpureus	Moss	151	362.5	169.1	46.66%
Ceratopteris_richardii	Fern	6172	7462.5	6463.4	86.61%
Chara_braunii	Algae (Charales)	11654	1751.2	1019.7	58.23%
Chlamydomonas_reinhardtii	Algae (Chlorophyta)	53	111.1	23.9	21.47%
Cinnamomum_micranthum	Angiosperm (Magnolids)	2150	730.4	403.7	55.27%
Citrullus_lanatus	Angiosperm (Eudicots)	14	361.5	216.7	59.96%
Citrus_sinensis	Angiosperm (Eudicots)	11	299.8	151.2	50.44%
Cucumis_melo	Angiosperm (Eudicots)	1310	438.4	280.4	63.96%
Cucumis_sativus	Angiosperm (Eudicots)	88	226.6	82.9	36.56%
Cycas_panzhihuaensis	Gymnosperm	923	10482.7	8796.7	83.92%
Cydonia_oblonga	Angiosperm (Eudicots)	303932	488.4	268.7	55.01%
Cynara_cardunculus	Angiosperm (Eudicots)	8174	725.0	457.6	63.12%
Cyperus_esculentus	Angiosperm (Monocots)	1190	296.6	124.7	42.03%
Datura_stramonium	Angiosperm (Eudicots)	13	1974.3	1678.6	85.02%
Daucus_carota	Angiosperm (Eudicots)	4826	421.5	184.5	43.76%
Dendrobium_officinale	Angiosperm (Monocots)	1621	1228.7	895.7	72.90%
Fragaria_ananassa	Angiosperm (Eudicots)	204	805.7	359.3	44.60%
Fragaria_vesca	Angiosperm (Eudicots)	3048	214.4	63.1	29.42%
Ginkgo_biloba	Gymnosperm	265987	2638.1	998.4	37.84%
Glycine_max	Angiosperm (Eudicots)	284	978.9	527.0	53.84%
Glycine_soja	Angiosperm (Eudicots)	1120	1013.8	558.1	55.05%
Gnetum_montanum	Gymnosperm	38363	2147.7	1770.0	82.41%
Gossypium_arboreum	Angiosperm (Eudicots)	948	1621.4	1356.4	83.66%
Gossypium_hirsutum	Angiosperm (Eudicots)	1027	2306.1	1792.7	77.74%
Gossypium_raimondii	Angiosperm (Eudicots)	289	751.0	524.0	69.76%
Ipomoea_batatas	Angiosperm (Eudicots)	28461	837.0	392.0	46.83%
Isoetes_engelmannii	Lycopods	319260	641.0	281.5	43.92%
Isoetes_taiwanensis	Lycopods	1113	1658.3	857.9	51.73%
Isothecium_myosuroides	Moss	256	388.5	222.7	57.32%
Lactuca_sativa	Angiosperm (Eudicots)	93	2590.4	2210.7	85.34%
Lemna_minuta	Angiosperm (Monocots)	2381	360.4	245.4	68.08%
Lycium_ferocissimum	Angiosperm (Eudicots)	14905	1219.4	896.3	73.51%
Lycopodium_clavatum	Lycopods	7102	2304.7	1853.8	80.43%
Malus_domestica	Angiosperm (Eudicots)	807	703.4	391.0	55.59%
Mangifera_indica	Angiosperm (Eudicots)	250	392.0	197.0	50.26%
Marchantia_paleacea	Liverworts	192	250.8	112.2	44.75%
Marchantia_polymorpha	Liverworts	2957	225.8	55.8	24.70%
Medicago_truncatula	Angiosperm (Eudicots)	42	430.0	221.3	51.46%
Musa_acuminata	Angiosperm (Monocots)	7259	472.2	173.0	36.63%
Nelumbo_nucifera	Angiosperm (Eudicots)	3603	804.6	424.4	52.74%
Nicotiana_benthamiana	Angiosperm (Eudicots)	17640	3035.8	2192.4	72.22%
Nymphaea_colorata	Angiosperm (Basal)	786	408.9	177.2	43.33%
Olea_europaea	Angiosperm (Eudicots)	9753	1316.7	821.6	62.40%
Oryza_sativa	Angiosperm (Monocots)	58	374.4	181.9	48.58%
Phalaenopsis_equestris	Angiosperm (Monocots)	89584	1064.2	704.9	66.24%
Phaseolus_vulgaris	Angiosperm (Eudicots)	708	521.1	271.1	52.02%
Phoenix_dactylifera	Angiosperm (Monocots)	2391	773.2	447.9	57.92%
Physalis_pubescens	Angiosperm (Eudicots)	327	1389.3	1161.7	83.61%
Physcomitrium_patens	Moss	359	472.1	283.1	59.98%
Pistacia_vera	Angiosperm (Eudicots)	1865	671.3	428.9	63.89%
Populus_trichocarpa	Angiosperm (Eudicots)	47	392.3	182.0	46.38%
Prunus_persica	Angiosperm (Eudicots)	192	227.6	96.1	42.24%
Punica_granatum	Angiosperm (Eudicots)	474	320.5	160.5	50.08%
Salvia_hispanica	Angiosperm (Eudicots)	1556	321.5	141.8	44.12%
Salvia_splendens	Angiosperm (Eudicots)	1162	806.1	475.6	59.00%
Selaginella_moellendorffii	Lycopods	757	212.3	108.1	50.90%
Setaria_italica	Angiosperm (Monocots)	337	405.9	206.1	50.79%
Solanum_lycopersicum	Angiosperm (Eudicots)	3066	828.0	517.2	62.47%
Solanum_melongena	Angiosperm (Eudicots)	13	1164.4	807.5	69.34%
Solanum_pimpinellifolium	Angiosperm (Eudicots)	127	808.1	568.4	70.34%
Solanum_stenotomum	Angiosperm (Eudicots)	17084	846.4	503.1	59.44%
Solanum_tuberosum	Angiosperm (Eudicots)	14853	705.9	433.3	61.38%
Sphagnum_fallax	Moss	36	395.1	185.5	46.94%
Sphagnum_magellanicum	Moss	31	439.0	217.5	49.55%
Thuja_plicata	Gymnosperm	67899	9095.9	6448.2	70.89%
Utricularia_gibba	Angiosperm (Eudicots)	518	100.7	29.7	29.52%
Vanilla_planifolia	Angiosperm (Monocots)	3874	1416.4	1072.3	75.71%
Vicia_sativa	Angiosperm (Eudicots)	18	1653.6	1346.0	81.40%
Vitis_vinifera	Angiosperm (Eudicots)	1907	486.2	264.6	54.43%
Welwitschia_mirabilis	Gymnosperm	22	6867.0	5231.1	76.18%
Zea_mays	Angiosperm (Monocots)	687	2182.8	1823.4	83.54%

Based on the table:

- *Isoetes_engelmannii* (Lycopods) has the highest number of contigs (319,260).
- *Cycas_panzhihuaensis* (Gymnosperm) has the largest assembly size (10,482.7 Mb) and the highest number of bases masked (8,796.7 Mb).
- *Ceratopteris_richardii* (Fern) has the maximum percentage of bases masked (86.61%).
- *Arabidopsis_thaliana* (Angiosperm)has the lowest number of contigs (7) and the minimum percentage of bases masked (17.05%).
- *Chlamydomonas_reinhardtii* (Algae) has the smallest assembly size (111.1 Mb).
- *Utricularia_gibba* (Angiosperm) has the lowest number of bases masked (29.7 Mb).

In this table, plant lineages such as Angiosperms, Ferns, Hornwort, Moss, Gymnosperms, and Lycopods are represented.

It is evident that similar to Chlorophyte, there exists a positive correlation between genome size and TE content across these lineages.

Gymnosperms typically have large genomes with elevated TE content. For instance, *Cycas panzhihuaensis* has an assembly size of 10,482.7 Mb with a TE content of 83.92%. Similarly, *Gnetum montanum* possesses an assembly size of 2,147.7 Mb and a TE content of 82.41%. It's noteworthy to mention an outlier, *Ginkgo biloba*, with a genome size of 2638.1 Mb, yet a TE content of only 37.84%.

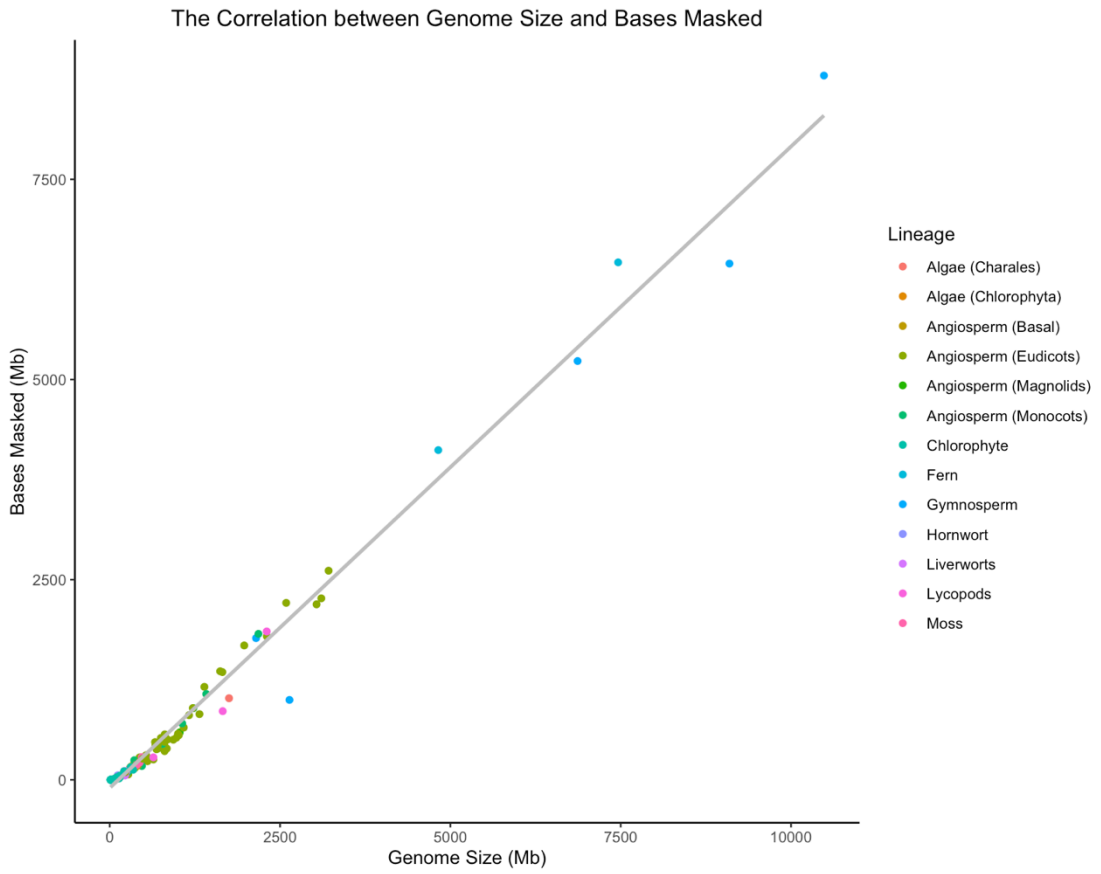
Ferns also tend to have expansive genomes and high TE content. *Ceratopteris richardii*, with an assembly size of 7,462.5 Mb, has a TE content of 86.61%. In comparison, *Adiantum capillus* has an assembly size of 4,822.6 Mb and a TE content of 86.61%.

For Angiosperms, the genome sizes vary considerably, as do the TE contents. *Capsicum annuum* has an assembly size of 3,212.5 Mb with a TE content of 81.33%, while *Nicotiana benthamiana*, with an assembly size of 3,035.8 Mb, has a TE content of 72.22%. On the other end of the spectrum, *Arabidopsis thaliana*, with an assembly size of 119.7 Mb, has a TE content of 17.05%, while *Utricularia gibba*, with an assembly size of 100.7 Mb, has a TE content of 29.52%.

In contrast to the aforementioned lineages, Mosses generally exhibit smaller genomes and lower TE content. For example, *Ceratodon purpureus* has an assembly size of 362.5 Mb and a TE content of 46.66%, while *Sphagnum fallax* possesses an assembly size of 395.1 Mb with a TE content of 46.94%.

Visualizing the Correlation between Genome Size and Bases Masked

To explore the correlation between Genome Size and Bases Masked and validate their relationship, a scatter plot was generated for effective visualization of the results.



From this scatter plot, it is evident that Gymnosperms exhibit the largest Genome Size and the highest content of Bases Masked, while Chlorophytes have the smallest Genome Size and the lowest Bases Masked content. However, regardless of lineage, there's a clear positive correlation between Genome Size and TE content. The presence of a grey reference line further emphasizes this trend, with data points clustered around the line, highlighting the consistent relationship between Genome Size and Bases Masked across all lineages.

2. Comparative and analysis

2.1 Distribution Metrics and Box Plots Representations

The analysis provides insight into the distribution and characteristics of various repetitive element classes in the genome. Within our dataset, we have captured fundamental metrics for key repetitive elements, such as LTRs and DNAs, including their quantity, length, and respective proportions.

To visualize these data comprehensively, we employed box plots, showcasing the divergence percentages of specific repetitive elements across all analyzed plant genomes. Through these box plots, insights into the variability, central tendencies, and potential outliers within the dataset are revealed, offering a clear and concise representation of comparative genomics concerning repetitive elements.

2.1.1 Retroelement Comparative and Analysis

Retrotransposons stand as paramount transposable elements due to their ability to amplify their copy numbers with each transposition event, thereby enhancing their presence within a genome. This amplification strategy underscores their significance, especially evident in the rich diversity observed in plant genomes. Within the realm of retrotransposons, the LTR elements emerge as the most prominent subclass, further accentuating their important role in genome dynamics.

The subsequent table lists the number, length, and genome percentage of Retroelements and LTR for each of the 138 plant genomes.

SPECIES	LINEAGE	Retroelements (number of elements)	Retroelements (lenth occupied)	Retroelements (percentage of sequence)	LTR elements (number of elements)	LTR elements (lenth occupied)	LTR elements (percentage of sequence)
Acer negundo	Angiosperm (Eudicots)	126284	108153961 bp	24.44%	107755	94969041 bp	21.46%
Acer yangbiense	Angiosperm (Eudicots)	304250	300756057 bp	45.17%	286105	284898812 bp	42.78%
Actinidia_chinensis	Angiosperm (Eudicots)	168776	80994548 bp	14.62%	156310	76690563 bp	13.85%
Adansonia_digitata	Angiosperm (Eudicots)	129083	95171537 bp	13.85%	124980	93001000 bp	13.54%
Adiantum_capillus	Fern	2626336	2577489961 bp	53.45%	2343681	2325126100 bp	48.21%
Amborella_trichopoda	Angiosperm (Basal)	246584	174102641 bp	24.64%	176416	143810231 bp	20.36%
Ananas_comosus	Angiosperm (Monocots)	133349	86985233 bp	22.77%	127366	83775114 bp	21.93%
Annona_glabra	Angiosperm (Magnolids)	298681	190968875 bp	18.59%	266212	173077730 bp	16.85%
Anthoceros_angustus	Hornwort	29517	20820774 bp	17.45%	23734	18843450 bp	15.79%
Arabidopsis_aromosa	Angiosperm (Eudicots)	16650	11996279 bp	8.02%	11844	9149909 bp	6.11%
Arabidopsis_suecica	Angiosperm (Eudicots)	31503	23548574 bp	8.65%	23728	19357696 bp	7.11%
Arabidopsis_thaliana	Angiosperm (Eudicots)	9167	8569281 bp	7.16%	6850	7206069 bp	6.02%
Arabis_alpina	Angiosperm (Eudicots)	89990	93621774 bp	30.04%	73488	82567377 bp	26.49%
Boechera_stricta	Angiosperm (Eudicots)	46312	36510203 bp	19.16%	39001	33100811 bp	17.37%
Brachypodium_distachyon	Angiosperm (Monocots)	53741	54509144 bp	20.09%	40667	48157771 bp	17.75%
Brassica_carinata	Angiosperm (Eudicots)	391064	336756525 bp	30.98%	352722	308421386 bp	28.37%
Brassica_juncea	Angiosperm (Eudicots)	241104	233374128 bp	25.00%	215078	214307868 bp	22.96%
Brassica_napus	Angiosperm (Eudicots)	251480	250019728 bp	24.95%	209312	216493746 bp	21.61%
Brassica_nigra	Angiosperm (Eudicots)	147898	169278854 bp	31.69%	129613	156442747 bp	29.28%
Brassica_oleracea	Angiosperm (Eudicots)	136918	89575654 bp	18.32%	111877	72071780 bp	14.74%
Brassica_rapa	Angiosperm (Eudicots)	64642	69883453 bp	19.80%	52268	60497844 bp	17.14%
Camelina_sativa	Angiosperm (Eudicots)	170190	121378804 bp	18.93%	133751	99072655 bp	15.45%
Camellia_sinensis	Angiosperm (Eudicots)	2239746	1525217008 bp	49.12%	2169372	1487753032 bp	47.91%
Capsicum_annuum	Angiosperm (Eudicots)	2037324	1717370983 bp	53.46%	1948635	1668572981 bp	51.94%
Carica_papaya	Angiosperm (Eudicots)	83531	66296790 bp	17.90%	62509	58868000 bp	15.89%
Ceratodon_purpureus	Moss	36467	34221432 bp	9.44%	29307	28516970 bp	7.87%
Ceratopteris_richardii	Fern	2895272	4220538644 bp	56.56%	2607186	4011182274 bp	53.75%
Chara_braunii	Algae (Charales)	597011	378537583 bp	21.62%	408071	302920585 bp	17.30%
Chlamydomonas_reinhardtii	Algae (Chlorophyta)	7610	4833342 bp	4.35%	2150	803213 bp	0.72%
Cinnamomum_micranthum	Angiosperm (Magnolids)	297428	189630832 bp	25.96%	232072	169058689 bp	23.15%
Citrullus_lanatus	Angiosperm (Eudicots)	114291	83792964 bp	23.18%	99756	78484908 bp	21.71%
Citrus_sinensis	Angiosperm (Eudicots)	72900	77348665 bp	25.80%	65176	72890621 bp	24.31%
Cucumis_melo	Angiosperm (Eudicots)	98784	88531705 bp	20.19%	84062	73597501 bp	16.79%
Cucumis_sativus	Angiosperm (Eudicots)	48848	286414454 bp	12.64%	34374	23667544 bp	10.44%
Cycas_panzhihuaensis	Gymnosperm	3657414	5840400182 bp	55.71%	3371197	5460793725 bp	52.09%
Cydonia_oblonga	Angiosperm (Eudicots)	194496	94987254 bp	19.45%	173856	88494700 bp	18.12%
Cynara_cardunculus	Angiosperm (Eudicots)	314338	245307740 bp	33.84%	302988	240004514 bp	33.11%
Cyperus_esculentus	Angiosperm (Monocots)	47411	27131554 bp	9.15%	36575	20753460 bp	7.00%
Datura_stramonium	Angiosperm (Eudicots)	964669	1288224295 bp	65.25%	930087	1268310423 bp	64.24%
Daucus_carota	Angiosperm (Eudicots)	103470	80177183 bp	19.02%	98001	76534034 bp	18.16%
Dendrobium_officinale	Angiosperm (Monocots)	712553	519734100 bp	42.30%	610235	448897704 bp	36.54%
Fragaria_ananassa	Angiosperm (Eudicots)	272575	197918641 bp	24.57%	249338	188066071 bp	23.34%
Fragaria_vesca	Angiosperm (Eudicots)	34583	22169814 bp	10.34%	28489	19813785 bp	9.24%
Ginkgo_biloba	Gymnosperm	909435	414305036 bp	15.70%	792852	359312066 bp	13.62%
Glycine_max	Angiosperm (Eudicots)	269973	274809994 bp	28.07%	243156	263206332 bp	26.89%
Glycine_soja	Angiosperm (Eudicots)	309093	316760621 bp	31.25%	278983	302272523 bp	29.82%
Gnetum_montanum	Gymnosperm	1123984	1209745542 bp	56.33%	991060	1142204795 bp	53.18%
Gossypium_arboreum	Angiosperm (Eudicots)	883668	891540712 bp	54.99%	859534	876596174 bp	54.07%
Gossypium_hirsutum	Angiosperm (Eudicots)	1180243	908953042 bp	39.42%	1143328	886030685 bp	38.42%
Gossypium_raimondii	Angiosperm (Eudicots)	411480	265031815 bp	35.29%	386632	251921492 bp	33.54%
Ipomoea_batatas	Angiosperm (Eudicots)	303424	127835361 bp	15.27%	257017	107775405 bp	12.88%
Isoetes_engelmannii	Lycopods	208596	89667531 bp	13.99%	91831	42168719 bp	6.58%
Isoetes_taiwanensis	Lycopods	660855	412359670 bp	24.87%	500148	317698243 bp	19.16%
Isotheciun_myosuroides	Moss	43941	29873530 bp	7.69%	38342	24898929 bp	6.41%
Lactuca_sativa	Angiosperm (Eudicots)	1524278	1415709237 bp	54.65%	1500370	1404411729 bp	54.22%
Lemna_minuta	Angiosperm (Monocots)	207428	144937857 bp	40.21%	197789	139129548 bp	38.60%
Lycium_farocissimum	Angiosperm (Eudicots)	784090	475377383 bp	38.99%	740042	452147529 bp	37.08%
Lycopodium_clavatum	Lycopods	1649330	1253559560 bp	54.39%	1584705	1204562564 bp	52.26%
Malus_domestica	Angiosperm (Eudicots)	366230	260993834 bp	37.11%	347424	249112549 bp	35.42%
Mangifera_indica	Angiosperm (Eudicots)	92567	69973374 bp	17.85%	89602	66983416 bp	17.09%
Marchantia_paleacea	Liverworts	89221	75574358 bp	30.13%	77191	67017555 bp	26.72%
Marchantia_polymorpha	Liverworts	32656	23916746 bp	10.59%	29704	21586169 bp	9.56%
Medicago_truncatula	Angiosperm (Eudicots)	139608	88891116 bp	20.67%	90938	73298225 bp	17.05%
Musa_acuminata	Angiosperm (Monocots)	128250	102395441 bp	21.68%	120275	98827874 bp	20.93%
Nelumbo_nucifera	Angiosperm (Eudicots)	382120	223138929 bp	27.73%	279023	176720871 bp	21.96%
Nicotiana_benthamiana	Angiosperm (Eudicots)	1559691	1371929835 bp	45.19%	1415995	1279329657 bp	42.14%
Nymphaea_colorata	Angiosperm (Basal)	89048	66371837 bp	16.23%	77746	60605936 bp	14.82%
Olea_europaea	Angiosperm (Eudicots)	267988	319040996 bp	24.23%	250278	310251709 bp	23.56%
Oryza_sativa	Angiosperm (Monocots)	89261	81247802 bp	21.70%	78155	76500715 bp	20.43%
Phalaenopsis_equestris	Angiosperm (Monocots)	385471	286037644 bp	26.88%	305995	245274952 bp	23.05%
Phaseolus_vulgaris	Angiosperm (Eudicots)	162297	140949605 bp	27.05%	162297	140949605 bp	27.05%
Phoenix_dactylifera	Angiosperm (Monocots)	266584	308813102 bp	39.94%	233335	291702991 bp	37.73%
Physalis_pubescens	Angiosperm (Eudicots)	495621	786447376 bp	56.61%	464400	767057869 bp	55.21%
Physcomitrium_patens	Moss	186935	234049057 bp	49.58%	185573	232903586 bp	49.34%
Pistacia_vera	Angiosperm (Eudicots)	299672	273470045 bp	40.74%	293826	269721484 bp	40.18%
Populus_trichocarpa	Angiosperm (Eudicots)	80654	58590584 bp	14.93%	73167	54459028 bp	13.88%
Prunus_persica	Angiosperm (Eudicots)	48229	34348462 bp	15.09%	41550	32601176 bp	14.33%
Punica_granatum	Angiosperm (Eudicots)	64866	61632920 bp	19.23%	59907	58335806 bp	18.20%
Salvia_hispanica	Angiosperm (Eudicots)	86911	50640195 bp	15.75%	80075	48703390 bp	15.15%
Salvia_splendens	Angiosperm (Eudicots)	310331	256463420 bp	31.82%	295040	239798646 bp	29.75%
Selaginella_moellendorffii	Lycopods	55676	42610188 bp	20.07%	51260	40169230 bp	18.92%
Setaria_italica	Angiosperm (Monocots)	102770	111202280 bp	27.40%	91248	104474231 bp	25.74%
Solanum_lycopersicum	Angiosperm (Eudicots)	314154	270745220 bp	32.70%	290900	259555545 bp	31.35%
Solanum_melongena	Angiosperm (Eudicots)	609151	470873526 bp	40.44%	530723	435626114 bp	37.41%
Solanum_pimpinellifolium	Angiosperm (Eudicots)	322691	307347865 bp	38.03%	293249	293647709 bp	36.34%
Solanum_stenotomum	Angiosperm (Eudicots)	335021	270491984 bp	31.96%	292764	255211907 bp	30.15%
Solanum_tuberosum	Angiosperm (Eudicots)	313204	215515370 bp	30.33%	274651	200270957 bp	28.37%
Sphagnum_fallax	Moss	43574	28389396 bp	7.18%	34964	23578149 bp	5.97%
Sphagnum_magellanicum	Moss	56959	44986679 bp	10.25%	41884	38593352 bp	8.79%
Thuja_plicata	Gymnosperm	4258994	4125788269 bp	45.36%	3914103	3891527676 bp	42.78%
Utricularia_gibba	Angiosperm (Eudicots)	11013	8267568 bp	8.21%	10910	8222061 bp	8.17%
Vanilla_planifolia	Angiosperm (Monocots)	260019	182985341 bp	12.92%	181897	127659550 bp	9.01%
Vicia_sativa	Angiosperm (Eudicots)	783561	981218887 bp	59.34%	728365	928261470 bp	56.14%
Vitis_vinifera	Angiosperm (Eudicots)	172049	129381341 bp	26.61%	141575	108612661 bp	22.34%
Welwitschia_mirabilis	Gymnosperm	3804488	2998342067 bp	43.66%	3390620	2722817332 bp	39.65%
Zea_mays	Angiosperm (Monocots)	906347	1403681099 bp	64.31%	875079	1382635184 bp	63.34%
Astrephomene gubernaculifera	Chlorophyte	7912	5142283 bp	4.95%	3963	2215737 bp	2.13%
Auxenochlorella protothecoides (0710)	Chlorophyte	6978	4361325 bp	4.20%	3552	1892903 bp	1.82%
Auxenochlorella protothecoides (UTEX 25)	Chlorophyte	16	2074 bp	0.01%	0	0	0.00%
Bathycoccus prasinos	Chlorophyte	0	0	0.00%	0	0	0.00%
Chlamydomonas asymjena	Chlorophyte	3159	1598873 bp	2.40%	2437	1289364 bp	1.94%
Chlamydomonas incerta	Chlorophyte	13871	10383340 bp	8.03%	7617	3565071 bp	2.76%
Chlamydomonas reinhardtii	Chlorophyte	7943	4692980 bp	4.22%	2375	817308 bp	0.74%
Chlamydomonas schloesseri	Chlorophyte	18332	11166777 bp	8.58%	3314	1828165 bp	1.40%
Chlamydomonas sp. UWO 241	Chlorophyte	100522	35253546 bp	16.66%	45858	11432131 bp	5.40%
Chlorella desiccata (nom. nud.) (UTEX 2437)	Chlorophyte	1031	762167 bp	3.68%	481	477291 bp	2.30%
Chlorella desiccata (nom. nud.) (UTEX 2526)	Chlorophyte	899	688559 bp	3.20%	464	420911 bp	1.95%
Chlorella ohadii	Chlorophyte	418	131233 bp	0.23%	155	26743 bp	0.05%
Chlorella sorokiniana	Chlorophyte	689	729007 bp	1.22%	160	92032 bp	0.15%
Chlorella variabilis	Chlorophyte	1555	756130 bp	1.64%	769	366151 bp	0.79%
Chloropicon primus (CCMP 1205)	Chlorophyte	0	0	0.00%	0	0	0.00%
Chloropicon primus (RCC138)	Chlorophyte	32	4714 bp	0.03%	0	0	0.00%
Coccomyxa sp. Obi	Chlorophyte	2274	872154 bp	1.73%	163	284727 bp	0.56%
Coccomyxa subellipsoidea C-169	Chlorophyte	1303	780014 bp	1.60%	0	0	0.00%
Dunaliella salina	Chlorophyte	114367	53909452 bp	15.68%	25840	7867582 bp	2.29%
Edaphochlamys debaryana	Chlorophyte	5305	5593255 bp	3.93%	1153	2601152 bp	1.83%
Haematococcus lacustris	Chlorophyte	101213	54235413 bp	17.53%	76186	35123253 bp	11.35%
Helicosporidium sp. ATCC 50920	Chlorophyte	14	1853 bp	0.01%	0	0	0.00%
Micractinium conductrix	Chlorophyte	2278	1452065 bp	2.38%	585	736873 bp	1.21%
Micromonas commoda	Chlorophyte	0	0	0.00%	0	0	0.00%
Micromonas pusilla CCMP1545	Chlorophyte	0	0	0.00%	0	0	0.00%
Monoraphidium minutum	Chlorophyte	3562	2355047 bp	3.45%	337	307286 bp	0.45%
Monoraphidium neglectum	Chlorophyte	755	145674 bp	0.21%	19	3362 bp	0.00%
Ostreobium quekettii	Chlorophyte	22817	25738896 bp	16.94%	19859	24748776 bp	16.29%
Ostreococcus lucimarinu	Chlorophyte	0	0	0.00%	0	0	0.00%
Ostreococcus tauri (RCC1115)	Chlorophyte	255	653270 bp	4.43%	255	653270 bp	4.43%
Ostreococcus tauri (RCC4221)	Chlorophyte	62	19246 bp	0.15%	62	19246 bp	0.15%
Pedinophyceae sp. YPF-701	Chlorophyte	892	506208 bp	1.81%	474	198039 bp	0.71%
Picochlorum sp. BPE23	Chlorophyte	414	428997 bp	2.87%	179	152738 bp	1.02%
Picochlorum sp. BPE23	Chlorophyte	414	428997 bp	2.87%	179	152738 bp	1.02%
Picocystis provasolii	Chlorophyte	199	70386 bp	0.31%	199	70386 bp	0.31%
Raphidocelis subcapitata	Chlorophyte	0	0	0.00%	0	0	0.00%
Scenedesmus sp. NREL 46B-D3	Chlorophyte	36878	15535419 bp	10.23%	4559	2285121 bp	1.50%
Tetrabaena socialis	Chlorophyte	4715	1602995 bp	1.18%	1589	628003 bp	0.46%
Treboxia sp. A1-2	Chlorophyte	3150	873129 bp	1.65%	637	250089 bp	0.47%
Volvox africanus	Chlorophyte	13271	12234895 bp	9.46%	8542	10125734 bp	7.83%
Volvox reticuliferus (NIES 3786)	Chlorophyte	14215	12434484 bp	9.28%	9472	9973979 bp	7.45%
Volvox reticuliferus (NIES-3785)	Chlorophyte	14296	12842426 bp	9.65%	8961	10245678 bp	7.70%

From the table, it's evident that the majority of the Retroelements across the plant lineages are LTR elements.

In lineages such as Angiosperms, Ferns, Hornwort, Moss, Gymnosperms, and Lycopods, the plant species with the highest proportion of Retroelements are *Datura stramonium* (Angiosperm) and *Zea mays* (Angiosperm). For *Datura stramonium*, 65.25% of its genome consists of Retroelements, out of which 64.24% are LTR elements. This implies that a staggering 98.5% of its retrotransposons are LTRs. Similarly, *Zea mays* has 64.31% of its genome as Retroelements, with 63.34% being LTR elements, indicating that 98.5% of its retrotransposons are LTRs.

In contrast, *Arabidopsis thaliana* (Arabidopsis) has a much lower percentage, with Retroelements making up 7.16% of its genome and LTR elements accounting for 6.02%. This means that 84.1% of its retrotransposons are LTRs.

In the Fern and Gymnosperm lineages, the number, length, and percentage of retrotransposons and LTRs are generally high. Mosses, however, exhibit a relatively lower content of retrotransposons.

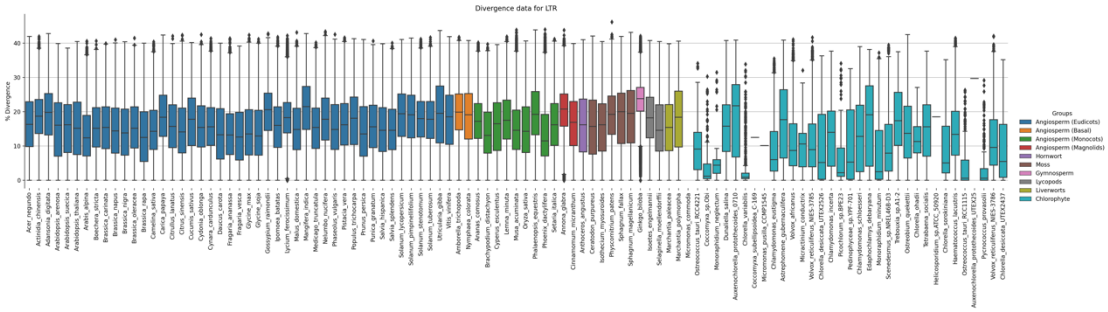
However, across all these lineages, it's notable that most plant genomes have LTR elements comprising over 80% of their Retroelements.

In contrast to the other lineages, Chlorophyte lineage genomes have notably fewer retrotransposons and LTRs. Within Chlorophyte, six genomes lack retrotransposons entirely, and approximately one-fourth lack LTR elements. In the remaining genomes, the majority have LTRs making up less than 3% of their genome.

However, there are exceptions within Chlorophyte. Notably, *Haematococcus lacustris* has retrotransposons comprising 17.53% of its genome and LTRs making up 11.35%. *Ostreobium quekettii* has retrotransposons at 16.94% and LTRs at 16.29% of its genome. A particularly unusual case is *Dunaliella salina*, where retrotransposons account for a significant 15.68% of its genome, while LTRs only constitute 2.29%.

2.1.1.1 LTR Repetitive Element Analysis

LTR element divergence data was analyzed and visually represented for each plant genome using boxplots.



This panel of boxplots displays the distribution of %Divergence values for LTR elements in all plant genomes. Each boxplot represents a single genome, allowing for a comparative assessment of divergence patterns across different species. The central line within each box denotes the median %Divergence, while the box's boundaries indicate the interquartile range (IQR). Whiskers extend to show the data's range, excluding any potential outliers.

Species-Specific Divergence: Genomes from certain plant species (e.g., *Ginkgo biloba (Gymnosperm)*), exhibited higher median %Divergence values, indicating potentially faster evolutionary rates in their LTR elements.

Distinct Low Median Divergence: Conversely, some plant genomes (e.g., *Ostreococcus tauri RCC1115* (Chlorophyte)) presented notably lower median %Divergence values. This could imply a slower evolutionary rate or distinct evolutionary constraints acting upon the LTR elements in these species.

Uniform Divergence Patterns: A subset of genomes demonstrated narrower IQRs (e.g., *Ostreococcus tauri RCC1115* (Chlorophyte)), suggesting a more uniform divergence pattern within those species.

Variable Divergence Patterns: In contrast to the aforementioned uniform patterns, certain genomes (e.g., *Auxenochlorella protothecoides 0710* (Chlorophyte)) exhibited wider interquartile ranges (IQRs). This broader IQR suggests a greater variability in the divergence rates of LTR elements within these species. Such variability could arise from a myriad of factors, including differential TE activity, genomic structural variations, or varying evolutionary pressures across populations of these plants.

Outliers and Unique Evolutionary Trajectories: Beyond the central tendencies, certain genomes (e.g., *Ginkgo biloba* (Gymnosperm)) displayed outliers, signifying either exceptionally high or low %Divergence values. These outliers might be indicative of unique evolutionary trajectories, genomic events, or selective pressures influencing the evolution and divergence of LTR elements in those specific plants.

Comparative Evolutionary Dynamics: The juxtaposition of these divergence patterns across diverse plant genomes offers a glimpse into the varied evolutionary dynamics at play. The differential %Divergence values underscore the complexity of evolutionary processes and the intricate interplay between genomic structures and evolutionary pressures.

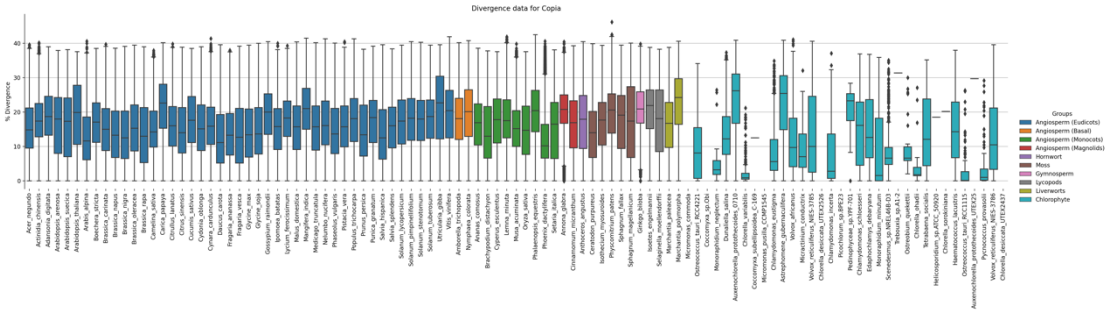
Overall, when examining the LTR elements across various plant lineages, a distinct pattern emerges. Within the Chlorophyte lineage, the median %Divergence ranges from 0 to 22, displaying significant diversity. A majority of these values predominantly lie below 15, underscoring the nuanced evolutionary patterns within the Chlorophyte lineage. The interquartile range (IQR) within this group exhibits variability, with some

species presenting narrower IQRs, while others showcase wider ones.

Conversely, for plant species outside the Chlorophyte lineage, the pattern is notably consistent. The median %Divergence for these species tends to be higher, clustering around 15 to 20. The IQR for these non-Chlorophyte species is relatively uniform, generally ranging between 10 and 15. Additionally, the whisker lengths for these species span from 0 to 40, indicating a broader range of variation in LTR elements across these lineages.

LTR Superfamily-Specific Divergence Analysis:
To delve deeper into the specific evolutionary patterns of LTR elements, we further segregated and analyzed the two predominant Superfamilies, Copia and Gypsy. These two superfamilies, renowned for their widespread distribution and significant impact on genome evolution, offer unique insights into the evolutionary trajectories and dynamics within the plant genomes.

Copia Repetitive Element Analysis:
Presented below are the boxplots representing the %Divergence of Copia elements across the diverse set of analyzed plant species.



For a majority of the plant lineages, spanning Angiosperms, Hornwort, Moss, Gymnospern, Lycopods, and Liverworts, the median %Divergence of Copia elements predominantly falls within the range of 15-20. This suggests a relatively conserved evolutionary pattern across these lineages. However, notable exceptions arise in specific species such as *Utricularia gibba* (Angiosperm - Eudicots) and *Marchantia polymorpha* (Liverworts). These species exhibit a higher median %Divergence, approximately 25, indicating potentially distinct evolutionary pressures or histories for the Copia elements within these genomes.

The interquartile range (IQR) for %Divergence remains broadly consistent across most lineages. Nevertheless, outliers in some species, such as *Annona glabra* (Angiosperm - Magnolids) and *Ginkgo biloba* (Gymnosperms), exhibit narrower IQRs. *Annona glabra* stands out further due to its particularly high median %Divergence and the presence of numerous outliers, suggesting unique evolutionary dynamics for Copia elements in this species.

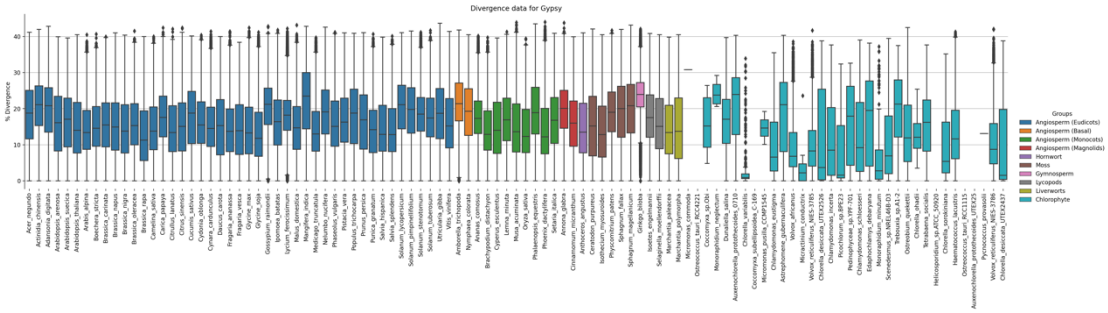
Furthermore, while the whisker lengths across the analyzed lineages remain relatively

consistent, spanning from 0 to 40, exceptions like *Phoenix dactylifera* (Angiosperm - Monocots) showcase shorter whiskers. Intriguingly, this species also presents an abundance of outliers surpassing the maximum %Divergence, indicative of potential rapid or distinct evolutionary events for its Copia elements.

However, within the Chlorophyte lineage, distinct patterns emerge in the %Divergence characteristics of Copia elements. The boxplot distributions for Chlorophyte species manifest considerable variability, evidenced by a spectrum of median %Divergence values. Notably, species such as *Auxenochlorella protothecoides 0710* and *Astrephomene gubernaculifera* present elevated median %Divergence values, accompanied by extended interquartile ranges (IQRs) and whisker lengths ranging from 0 to around 40. Conversely, some species exhibit very low median values, such as *Ostreococcus_tauri_RCC1115*, which exhibit a median %Divergence of zero. Moreover, the boxplots for *Ostreococcus tauri_RCC1115*, *Pycnococcus provasoli*, and *Chlorella variabilis* depict notably narrow IQRs and a pronounced presence of outliers exceeding the maximum %Divergence value.

In summary, while a majority of the analyzed plant lineages exhibit a relatively stable and conserved %Divergence for Copia elements, specific species, especially within the Chlorophyte lineage, present unique evolutionary patterns, emphasizing the diverse evolutionary dynamics of the Copia superfamily across plant genomes.

Gypsy Repetitive Element Analysis:
Presented below are the boxplots representing the %Divergence of Gypsy elements across the diverse set of analyzed plant species.



For a majority of the plant lineages, spanning Angiosperms, Hornwort, Moss, Gymnospern, Lycopods, and Liverworts, the median %Divergence of Gypsy elements predominantly clusters within the range of 15-25. Noteworthy species in this context include *Mangifera indica* (Angiosperm - Ludicots) and *Ginkgo biloba* (Gymnosperms), both showcasing a median %Divergence around 25. Compared to the broader trend where most plant species exhibit an IQR of approximately 10, *Ginkgo biloba* stands out with a notably narrow IQR. Additionally, while the whisker lengths for the majority of species range from 0 to 40, *Ginkgo biloba*'s whiskers span approximately 12-37, accompanied by numerous outliers at both extremes.

Overall, analogous to Copia, these various lineages manifest a pattern of relatively close

and stable %Divergence for Gypsy elements.

However, within the Chlorophyte lineage, the boxplot distributions for Gypsy elements remain notably diverse. Drawing parallels to observations in the Copia distributions, species like *Auxenochlorella protothecoides 0710* and *Astrephomene gubernaculifera* once again display elevated median %Divergence values, complemented by broad IQR and whiskers spanning from 0 to 40. Conversely, *Chlorella variabilis* continues to present a strikingly low median %Divergence. Its boxplot is characterized by a narrow IQR, shorter whiskers, and a significant representation of outliers exceeding the maximum %Divergence value.

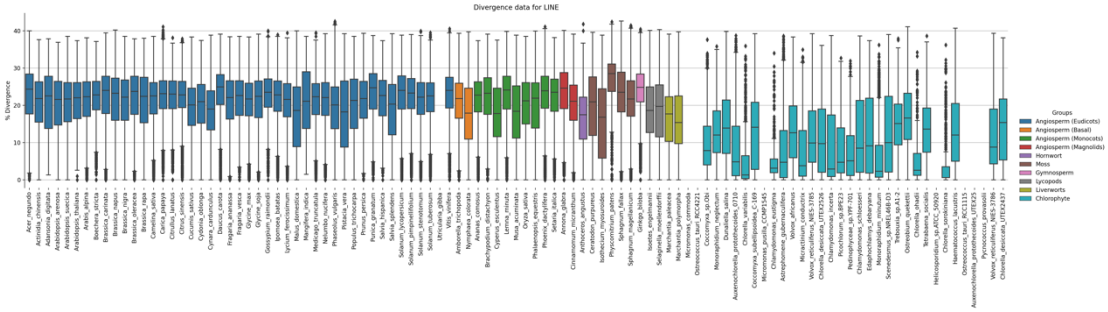
In summary, while the majority of plant lineages depict a relatively consistent %Divergence trend for Gypsy elements, specific species, particularly within the Chlorophyte lineage, highlight distinct evolutionary trajectories, underscoring the intricate evolutionary dynamics of the Gypsy superfamily across plant genomes.

2.1.1.2 Non-LTR Repetitive Element Analysis

Within the realm of Non-LTR transposons, LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements) are recognized as prominent superfamily members. These elements have left indelible imprints on plant genomes due to their transposition dynamics.

LINE Repetitive Element Analysis

The following boxplots detail the %Divergence profiles of LINEs across diverse plant lineages.



For the majority of plant groups, including Angiosperms, Hornwort, Moss, Gymnosperms, Lycopods, and Liverworts, the median %Divergence typically falls within the 15 to 25 range. However, an exception is *Physcomitrium patens* (Moss), which exhibits a slightly elevated median value of around 28.

Regarding the Interquartile Range (IQR), most lineages maintain a consistent range, roughly around 10. Yet, species like *Physcomitrium patens* (Moss) and *Ginkgo biloba* (Gymnosperms) stand out with narrower IQRs, accompanied by numerous outliers falling below the minimum value.

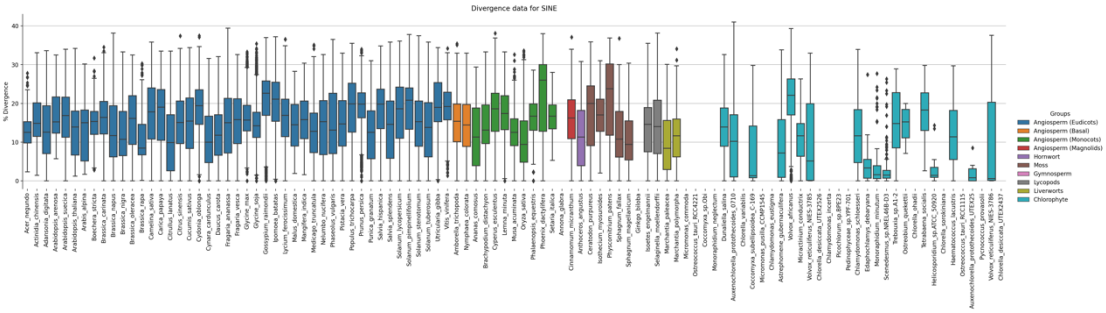
The whisker lengths vary among these lineages. While the maximum whisker value is

approximately 40 for many plants, the minimum values span between 0 and 15. Intriguingly, around half of the species across these lineages showcase outliers that fall below this minimum value.

Conversely, within the Chlorophyte lineage, there's a distinct profile for LINE elements. The median %Divergence values are notably lower, predominantly between 0 and 15. Notably, species like *Chlorella variabilis*, *Chlamydomonas_eustigma*, *Monoraphidium minutum*, *Chlorella ohadii*, and *Chlorella sorokiniana* display narrow IQRs and median values below 5, accompanied by a significant number of outliers exceeding the maximum value. In contrast, the majority of other Chlorophyte plants depict wider IQRs, roughly around 15, with whisker lengths spanning from 0 to 40.

SINE Repetitive Element Analysis

The following boxplots detail the %Divergence profiles of SINEs across diverse plant lineages.



For SINE elements, the boxplot distributions present a more varied landscape compared to LINE elements. Across the Angiosperms, Hornwort, Moss, Lycopods, and Liverworts lineages, the median %Divergence for SINE elements ranges from 8 to 26. Notably, *Brassica rapa* (Angiosperm) and *Marchantia palearea* (Liverworts) exhibit the lowest median values, both at 8, while *Phoenix dactylifera* (Angiosperm) displays the highest median value at 26.

The IQR values among these lineages vary significantly. For instance, species like *Phoenix dactylifera* (Angiosperm) and *Physcomitrium patens* (Moss) have broader IQRs, ranging between 15 and 20. In contrast, species such as *Acer negundo* (Angiosperm) present narrower IQRs, approximately around 5.

Regarding whisker lengths, they vary across these lineages, with minimum values starting from 0 and maximum values up to 39. A minority of species within these lineages exhibit outliers on both ends of the distribution.

Within the Chlorophyte lineage, the SINE elements showcase a median %Divergence ranging from 2 to 23. Notably, *Volvox africanus* presents the highest median value, while *Volvox reticuliferus NIES-3786* exhibits the lowest. The IQR values for Chlorophyte species vary, with species like *Volvox reticuliferus NIES-3786* having an IQR exceeding 20. Conversely, species like *Scenedesmus sp. NREL46B-D3* and

Monoraphidium minutum have narrower IQRs, less than 5. Additionally, these two species have notably shorter whiskers, with a considerable number of outliers surpassing the maximum %Divergence value.

2.1.2 DNA Transposon Comparative and Analysis

DNA transposons are crucial agents in genomic evolution, the s following table lists the number, length, and genome percentage of DNA transposons for each of the 138 plant genomes.

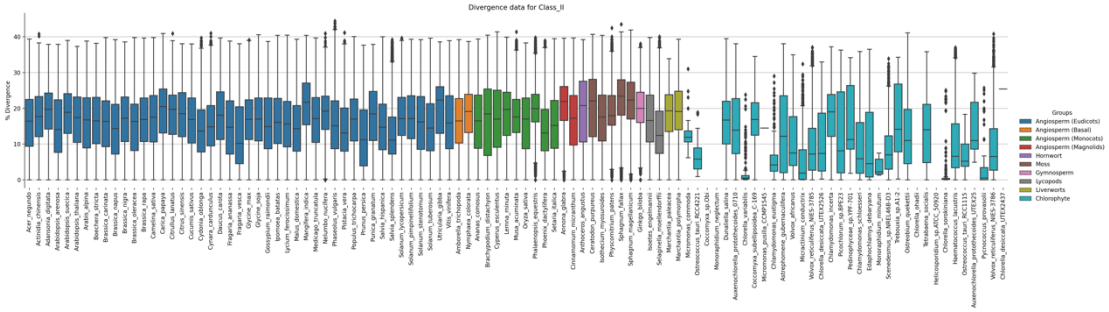
SPECIES	LINEAGE	DNA transposons (number of elements)	DNA transposons (lenth occupied)	DNA transposons (percentage of sequence)
Acer_negundo	Angiosperm (Eudicots)	28079	22687843 bp	5.13%
Acer_yangbiense	Angiosperm (Eudicots)	34847	22145145 bp	3.33%
Actinidia_chinensis	Angiosperm (Eudicots)	15649	5025946 bp	0.91%
Adansonia_digitata	Angiosperm (Eudicots)	21277	9929957 bp	1.45%
Adiantum_capillus	Fern	140142	137239168 bp	2.85%
Amborella_trichopoda	Angiosperm (Basal)	31356	14818905 bp	2.10%
Ananas_comosus	Angiosperm (Monocots)	8605	4486366 bp	1.17%
Annona_glabra	Angiosperm (Magnolids)	25657	20847775 bp	2.03%
Anthoceros_angustus	Hornwort	3437	1965941 bp	1.65%
Arabidopsis_arenosa	Angiosperm (Eudicots)	2788	1405650 bp	0.94%
Arabidopsis_suecica	Angiosperm (Eudicots)	8107	4658733 bp	1.71%
Arabidopsis_thaliana	Angiosperm (Eudicots)	2984	1920405 bp	1.60%
Arabis_alpina	Angiosperm (Eudicots)	16338	11973228 bp	3.84%
Boechera_stricta	Angiosperm (Eudicots)	12589	6401409 bp	3.36%
Brachypodium_distachyon	Angiosperm (Monocots)	7723	6342986 bp	2.34%
Brassica_carinata	Angiosperm (Eudicots)	85589	63851737 bp	5.87%
Brassica_juncea	Angiosperm (Eudicots)	41813	28303004 bp	3.03%
Brassica_napus	Angiosperm (Eudicots)	59041	47108121 bp	4.70%
Brassica_nigra	Angiosperm (Eudicots)	30036	21113120 bp	3.95%
Brassica_oleracea	Angiosperm (Eudicots)	35262	15686904 bp	3.21%
Brassica_rapa	Angiosperm (Eudicots)	13956	9020307 bp	2.56%
Camelina_sativa	Angiosperm (Eudicots)	33285	17442410 bp	2.72%
Camellia_sinensis	Angiosperm (Eudicots)	136002	82099863 bp	2.64%
Capsicum_annuum	Angiosperm (Eudicots)	111672	91257570 bp	2.84%
Carica_papaya	Angiosperm (Eudicots)	1508	639083 bp	0.17%
Ceratodon_purpureus	Moss	21246	21881291 bp	6.04%
Ceratopteris_richardii	Fern	700805	276888413 bp	3.71%
Chara_braunii	Algae (Charales)	24602	19031977 bp	1.09%
Chlamydomonas_reinhardtii	Algae (Chlorophyta)	92	113813 bp	0.10%
Cinnamomum_micranthum	Angiosperm (Magnolids)	29649	18022208 bp	2.47%
Citrullus_lanatus	Angiosperm (Eudicots)	19744	8613034 bp	2.38%
Citrus_sinensis	Angiosperm (Eudicots)	13969	7164638 bp	2.39%
Cucumis_melo	Angiosperm (Eudicots)	37829	22749575 bp	5.19%
Cucumis_sativus	Angiosperm (Eudicots)	7150	3582145 bp	1.58%
Cycas_panzhihuaensis	Gymnosperm	31104	17769842 bp	0.17%
Cydonia_oblonga	Angiosperm (Eudicots)	26497	8060790 bp	1.65%
Cynara_cardunculus	Angiosperm (Eudicots)	27138	11142394 bp	1.54%
Cyperus_esculentus	Angiosperm (Monocots)	6570	3794589 bp	1.28%
Datura_stramonium	Angiosperm (Eudicots)	18010	12733199 bp	0.64%
Daucus_carota	Angiosperm (Eudicots)	20398	8833851 bp	2.10%
Dendrobium_officinale	Angiosperm (Monocots)	53725	24786543 bp	2.02%
Fragaria_ananassa	Angiosperm (Eudicots)	39861	26574682 bp	3.30%
Fragaria_vesca	Angiosperm (Eudicots)	6574	4042406 bp	1.89%
Ginkgo_biloba	Gymnosperm	13928	5231600 bp	0.20%
Glycine_max	Angiosperm (Eudicots)	52355	37001649 bp	3.78%
Glycine_soja	Angiosperm (Eudicots)	46229	33166672 bp	3.27%
Gnetum_montanum	Gymnosperm	10918	4694992 bp	0.22%
Gossypium_arboreum	Angiosperm (Eudicots)	37704	22759193 bp	1.40%
Gossypium_hirsutum	Angiosperm (Eudicots)	61576	40838996 bp	1.77%
Gossypium_raimondii	Angiosperm (Eudicots)	24357	21626986 bp	2.88%
Ipomoea_batatas	Angiosperm (Eudicots)	29242	9622488 bp	1.15%
Isoetes_engelmannii	Lycopods	26177	9098791 bp	1.42%
Isoetes_taiwanensis	Lycopods	63161	41043338 bp	2.48%
Isothecium_myosuroides	Moss	25252	15953978 bp	4.11%
Lactuca_sativa	Angiosperm (Eudicots)	61984	32148664 bp	1.24%
Lemna_minuta	Angiosperm (Monocots)	5655	3405085 bp	0.94%
Lycium_ferocissimum	Angiosperm (Eudicots)	32113	19780550 bp	1.62%
Lycopodium_clavatum	Lycopods	137085	104243086 bp	4.52%
Malus_domestica	Angiosperm (Eudicots)	27620	13272521 bp	1.89%
Mangifera_indica	Angiosperm (Eudicots)	37851	18456744 bp	4.71%
Marchantia_paleacea	Liverworts	1413	622907 bp	0.25%
Marchantia_polymorpha	Liverworts	5440	3069328 bp	1.36%
Medicago_truncatula	Angiosperm (Eudicots)	22620	8037179 bp	1.87%
Musa_acuminata	Angiosperm (Monocots)	7992	2427137 bp	0.51%
Nelumbo_nucifera	Angiosperm (Eudicots)	22620	8378896 bp	1.04%
Nicotiana_benthamiana	Angiosperm (Eudicots)	103844	68164679 bp	2.25%
Nymphaea_colorata	Angiosperm (Basal)	25524	9616460 bp	2.35%
Olea_europaea	Angiosperm (Eudicots)	65866	37386672 bp	2.84%
Oryza_sativa	Angiosperm (Monocots)	17153	14539055 bp	3.88%
Phalaenopsis_equestris	Angiosperm (Monocots)	57005	16838695 bp	1.58%
Phaseolus_vulgaris	Angiosperm (Eudicots)	24958	15143968 bp	2.91%
Phoenix_dactylifera	Angiosperm (Monocots)	40516	27993003 bp	3.62%
Physalis_pubescens	Angiosperm (Eudicots)	20473	9102245 bp	0.66%
Physcomitrium_patens	Moss	3541	2374183 bp	0.50%
Pistacia_vera	Angiosperm (Eudicots)	31656	24951778 bp	3.72%
Populus_trichocarpa	Angiosperm (Eudicots)	17118	12993860 bp	3.31%
Prunus_persica	Angiosperm (Eudicots)	13747	16312160 bp	7.17%
Punica_granatum	Angiosperm (Eudicots)	6254	3244159 bp	1.01%
Salvia_hispanica	Angiosperm (Eudicots)	24728	12169842 bp	3.79%
Salvia_splendens	Angiosperm (Eudicots)	76040	53353608 bp	6.62%
Selaginella_moellendorffii	Lycopods	23300	23523787 bp	11.08%
Setaria_italica	Angiosperm (Monocots)	26641	22395623 bp	5.52%
Solanum_lycopersicum	Angiosperm (Eudicots)	28570	15859714 bp	1.92%
Solanum_melongena	Angiosperm (Eudicots)	13443	6766115 bp	0.58%
Solanum_pimpinellifolium	Angiosperm (Eudicots)	25814	13190001 bp	1.63%
Solanum_tuberosum	Angiosperm (Eudicots)	25737	12925525 bp	1.53%
Solanum_tuberosum	Angiosperm (Eudicots)	22507	9897002 bp	1.40%
Sphagnum_fallax	Moss	27429	14202558 bp	3.59%
Sphagnum_magellanicum	Moss	29317	15596520 bp	3.55%
Thuja_plicata	Gymnosperm	334111	263861406 bp	2.90%
Utricularia_gibba	Angiosperm (Eudicots)	3442	1452433 bp	1.44%
Vanilla_planifolia	Angiosperm (Monocots)	2199	697584 bp	0.05%
Vicia_sativa	Angiosperm (Eudicots)	78506	49529771 bp	3.00%
Vitis_vinifera	Angiosperm (Eudicots)	19462	12494505 bp	2.57%
Welwitschia_mirabilis	Gymnosperm	67861	42157800 bp	0.61%
Zea_mays	Angiosperm (Monocots)	50124	46563403 bp	2.13%
Astrephomene gubernaculifera	Chlorophyte	0	0	0.00%
Auxenochlorella protothecoides (0710)	Chlorophyte	41	72408 bp	0.07%
Auxenochlorella protothecoides (UTEX 25)	Chlorophyte	14	3451 bp	0.02%
Bathycoccus prasinos	Chlorophyte	0	0	0.00%
Chlamydomonas eustigma	Chlorophyte	90	36841 bp	0.06%
Chlamydomonas incerta	Chlorophyte	1189	260191 bp	0.20%
Chlamydomonas reinhardtii	Chlorophyte	170	143750 bp	0.13%
Chlamydomonas schloesseri	Chlorophyte	342	213990 bp	0.16%
Chlamydomonas sp. UWO 241	Chlorophyte	8794	840942 bp	0.40%
Chlorella desiccata (nom. nud.) (UTEX 2437)	Chlorophyte	0	0	0.00%
Chlorella desiccata (nom. nud.) (UTEX 2526)	Chlorophyte	27	43564 bp	0.20%
Chlorella ohadii	Chlorophyte	0	0	0.00%
Chlorella sorokiniana	Chlorophyte	316	1092893 bp	1.83%
Chlorella variabilis	Chlorophyte	235	137400 bp	0.30%
Chloropicon primus (CCMP 1205)	Chlorophyte	0	0	0.00%
Chloropicon primus (RCC138)	Chlorophyte	0	0	0.00%
Coccomyxa sp. Obi	Chlorophyte	0	0	0.00%
Coccomyxa subellipsoidea C-169	Chlorophyte	626	380477 bp	0.78%
Dunaliella salina	Chlorophyte	5896	1223833 bp	0.36%
Edaphochlamys debaryana	Chlorophyte	178	358582 bp	0.25%
Haematococcus lacustris	Chlorophyte	8464	3116883 bp	1.01%
Helicosporidium sp. ATCC 50920	Chlorophyte	0	0	0.00%
Micractinium conductrix	Chlorophyte	818	653231 bp	1.07%
Micromonas commoda	Chlorophyte	0	0	0.00%
Micromonas pusilla CCMP1545	Chlorophyte	0	0	0.00%
Monoraphidium minutum	Chlorophyte	0	0	0.00%
Monoraphidium neglectum	Chlorophyte	0	0	0.00%
Ostreobium quekettii	Chlorophyte	3115	3160941 bp	2.08%
Ostreococcus lucimarinu	Chlorophyte	0	0	0.00%
Ostreococcus tauri (RCC1115)	Chlorophyte	56	41913 bp	0.28%
Ostreococcus tauri (RCC4221)	Chlorophyte	34	19657 bp	0.15%
Pedinophyceae sp. YPF-701	Chlorophyte	0	0	0.00%
Picochlorum sp. BPE23	Chlorophyte	96	175748 bp	1.18%
Picochlorum sp. BPE23	Chlorophyte	96	175748 bp	1.18%
Pycnococcus provasolii	Chlorophyte	67	25558 bp	0.11%
Raphidocelis subcapitata	Chlorophyte	0	0	0.00%
Scenedesmus sp. NREL 46B-D3	Chlorophyte	270	137745 bp	0.09%
Tetrabaena socialis	Chlorophyte	292	139488 bp	0.10%
Treubaxia sp. A1-2	Chlorophyte	60	29089 bp	0.05%
Volvox africanus	Chlorophyte	646	508891 bp	0.39%
Volvox reticuliferus (NIES 3786)	Chlorophyte	1698	891663 bp	0.67%
Volvox reticuliferus (NIES-3785)	Chlorophyte	907	552913 bp	0.42%

Upon examining the distribution of DNA transposons across the analyzed plant lineages, it becomes evident that these transposons are not predominant in any lineage. However, a notable pattern emerges where the lineages of Angiosperms, Ferm, Hornwort, Moss, Gymnosperms, and Lycopods consistently harbor a considerably higher abundance of DNA transposons compared to the Chlorophyte lineage. Within the Chlorophyte genomes studied, approximately one-third of the species lack DNA transposons entirely. Among the remaining Chlorophyte genomes, the majority possess a representation of DNA transposons accounting for less than 1%. Yet, mirroring the LTR element trend, *Ostreobium quekettii* stands out with the highest proportion of DNA transposons at 2.08%.

Across the other examined lineages, the proportion of DNA transposons varies considerably, ranging from 0.17% in *Cycas panzhihuaensis* (Gymnosperm) to a noteworthy 11.08% in *Selaginella moellendorffii* (Lycopod).

For most Angiosperms, the DNA transposon representation falls within the range of 1%-5%. Conversely, most Gymnosperms display a more subdued representation, with the DNA transposon proportion typically being less than 1%.

The presented boxplots delineate the %Divergence metrics for these elements.



For most plant lineages, including Angiosperms, Hornwort, Moss, Gymnosperms, Lycopods, and Liverworts, the median %Divergence of DNA predominantly falls within the range of 15-20. This consistent median %Divergence across diverse lineages may suggest a fundamental evolutionary constraint or a shared genomic stability mechanism across these plant groups. However, there are exceptions. For instance, the median values for *Mangifera indica* and *Utricularia gibba* (both Angiosperm - Eudicots) are approximately 22, while those for *Sphagnum fallax* and *Sphagnum magellanicum* (both Moss) are around 23 and 24, respectively. Such deviations could be indicative of lineage-specific evolutionary pressures or unique genomic dynamics in these species, potentially related to environmental adaptations or specific genomic rearrangements.

Across these lineages, the Interquartile Range (IQR) remains relatively consistent, approximately around 15. Yet, certain lineages exhibit narrower IQRs. Notably, species such as *Phalaenopsis equestris* (Angiosperm - Monocots), *Physcomitrium patens*, *Sphagnum fallax*, *Sphagnum magellanicum* (all Moss), *Ginkgo biloba* (Gymnosperms), and *Annona alabra* (Angiosperm - Magnolids) display narrower IQRs, with their

median values also relatively elevated. This could imply a more constrained genomic variability within these species or lineages, possibly reflecting specialized ecological niches or specific developmental pathways.

Furthermore, while the whisker lengths across most lineages range from 0 to 40, exceptions like *Salvia splendens* (Angiosperm - Ludicots) present shorter whiskers and exhibit numerous outliers surpassing the maximum divergence percentage. Such outliers may signify rapid genomic changes or unique evolutionary trajectories in these particular species, warranting further investigation into their genomic dynamics and potential adaptive significance.

In the Chlorophyte lineage, a distinct pattern emerges in the %Divergence of DNA transposons. The median values across species in this lineage vary widely, ranging from 0 to 18. Particularly, *Chlamydomonas incerta* exhibits the highest median value, whereas *Chlorella sorokiniana* and *Chlorella variabilis* approach a median value near 0, with numerous outliers beyond the maximum value. Additionally, the IQR lengths differ among species in this lineage, with *Picochlorum sp.BPE23* having an IQR of approximately 15, while *Chlorella sorokiniana*'s IQR is close to 0. The whisker lengths for these species also vary, spanning from 0 to 42, with a majority having outliers beyond the maximum value. These variations may reflect diverse genomic architectures or differential evolutionary pressures within the Chlorophyte lineage, highlighting the need for comprehensive genomic studies to elucidate the underlying mechanisms and functional implications.

2.1.3 Unclassified Element Comparative and Analysis

Unclassified elements refer to segments within the genomes that, despite being identified as repetitive elements, do not fit into the conventional categories or classifications like LTR elements or DNA transposons. Their functional significance or origin may not be fully elucidated, leading to their current unclassified status.
The following table lists the number, length, and genome percentage of unclassified elements for each of the 138 plant genomes.

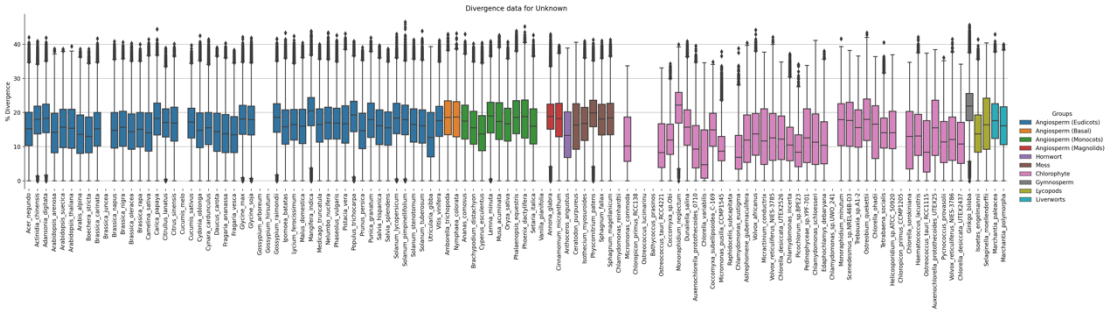
SPECIES	LINEAGE	Unclassified (number of elements)	Unclassified (lenth occupied)	Unclassified (percentage of sequence)
Acer_negundo	Angiosperm (Eudicots)	544760	130037607 bp	29.39%
Acer_yangbiense	Angiosperm (Eudicots)	577503	135738787 bp	20.38%
Actinidia_chinensis	Angiosperm (Eudicots)	744075	138377348 bp	24.98%
Adansonia_digitata	Angiosperm (Eudicots)	1144391	264414665 bp	38.49%
Adiantum_capillus	Fern	3749055	1318017673 bp	27.33%
Amborella_trichopoda	Angiosperm (Basal)	945500	232627118 bp	32.93%
Ananas_comosus	Angiosperm (Monocots)	452670	116046036 bp	30.37%
Annona_glabra	Angiosperm (Magnolids)	1439494	371633160 bp	36.18%
Anthoceros_angustus	Hornwort	86679	33324172 bp	27.92%
Arabidopsis_arenosa	Angiosperm (Eudicots)	53727	15920018 bp	10.64%
Arabidopsis_suecica	Angiosperm (Eudicots)	88525	32856968 bp	12.07%
Arabidopsis_thaliana	Angiosperm (Eudicots)	21921	7459066 bp	6.23%
Arabis_alpina	Angiosperm (Eudicots)	170695	54991419 bp	17.65%
Boechera_stricta	Angiosperm (Eudicots)	89134	24168195 bp	12.68%
Brachypodium_distachyon	Angiosperm (Monocots)	175368	37044778 bp	13.65%
Brassica_carinata	Angiosperm (Eudicots)	720593	230304662 bp	21.19%
Brassica_juncea	Angiosperm (Eudicots)	573264	217141236 bp	23.26%
Brassica_napus	Angiosperm (Eudicots)	637744	239012010 bp	23.86%
Brassica_nigra	Angiosperm (Eudicots)	290455	106523763 bp	19.94%
Brassica_oleracea	Angiosperm (Eudicots)	438284	122432596 bp	25.04%
Brassica_rapa	Angiosperm (Eudicots)	212763	83552437 bp	23.67%
Camelina_sativa	Angiosperm (Eudicots)	330593	103209815 bp	16.09%
Camellia_sinensis	Angiosperm (Eudicots)	2224045	592622309 bp	19.08%
Capsicum_annuum	Angiosperm (Eudicots)	2357972	782159937 bp	24.35%
Carica_papaya	Angiosperm (Eudicots)	177433	69989504 bp	18.89%
Ceratodon_purpureus	Moss	209537	108539954 bp	29.94%
Ceratopteris_richardii	Fern	4742183	1521578784 bp	20.39%
Chara_braunii	Algae (Charales)	2239563	526321037 bp	30.05%
Chlamydomonas_reinhardtii	Algae (Chlorophyta)	27758	6751890 bp	6.08%
Cinnamomum_micranthum	Angiosperm (Magnolids)	849081	180647370 bp	24.73%
Citrullus_lanatus	Angiosperm (Eudicots)	473072	116557476 bp	32.25%
Citrus_sinensis	Angiosperm (Eudicots)	207264	60778753 bp	20.27%
Cucumis_melo	Angiosperm (Eudicots)	418573	142851878 bp	32.58%
Cucumis_sativus	Angiosperm (Eudicots)	178666	44910196 bp	19.82%
Cycas_panzhihuaensis	Gymnosperm	7435038	2743644965 bp	26.17%
Cydonia_oblonga	Angiosperm (Eudicots)	744719	157526361 bp	32.25%
Cynara_cardunculus	Angiosperm (Eudicots)	891700	191402492 bp	26.40%
Cyperus_esculentus	Angiosperm (Monocots)	326916	73495076 bp	24.78%
Datura_stramonium	Angiosperm (Eudicots)	997980	352427016 bp	17.85%
Daucus_carota	Angiosperm (Eudicots)	328592	89948690 bp	21.34%
Dendrobium_officinale	Angiosperm (Monocots)	1052446	328746339 bp	26.76%
Fragaria_ananassa	Angiosperm (Eudicots)	418155	119347769 bp	14.81%
Fragaria_vesca	Angiosperm (Eudicots)	124784	33812268 bp	15.77%
Ginkgo_biloba	Gymnosperm	2095638	568088060 bp	21.53%
Glycine_max	Angiosperm (Eudicots)	659719	197258115 bp	20.15%
Glycine_soja	Angiosperm (Eudicots)	657508	190551905 bp	18.80%
Gnetum_montanum	Gymnosperm	1517945	537394856 bp	25.02%
Gossypium_arboreum	Angiosperm (Eudicots)	1085306	419635130 bp	25.88%
Gossypium_hirsutum	Angiosperm (Eudicots)	2229684	806786366 bp	34.99%
Gossypium_raimondii	Angiosperm (Eudicots)	827870	226686515 bp	30.18%
Ipomoea_batatas	Angiosperm (Eudicots)	1341028	234230149 bp	27.98%
Isoetes_engelmannii	Lycopods	894777	171125709 bp	26.70%
Isoetes_taiwanensis	Lycopods	1188150	380920875 bp	22.97%
Isothecium_myosuroides	Moss	582520	172208054 bp	44.33%
Lactuca_sativa	Angiosperm (Eudicots)	1854922	721243084 bp	27.84%
Lemna_minuta	Angiosperm (Monocots)	262617	89694251 bp	24.88%
Lycium_ferocissimum	Angiosperm (Eudicots)	1376314	388427203 bp	31.86%
Lycopodium_clavatum	Lycopods	1419861	465823119 bp	20.21%
Malus_domestica	Angiosperm (Eudicots)	490811	107174828 bp	15.24%
Mangifera_indica	Angiosperm (Eudicots)	404622	100428467 bp	25.62%
Marchantia_paleacea	Liverworts	91949	32983442 bp	13.15%
Marchantia_polymorpha	Liverworts	62461	26266591 bp	11.63%
Medicago_truncatula	Angiosperm (Eudicots)	554946	115755563 bp	26.92%
Musa_acuminata	Angiosperm (Monocots)	206778	62323735 bp	13.20%
Nelumbo_nucifera	Angiosperm (Eudicots)	944578	183285065 bp	22.78%
Nicotiana_benthamiana	Angiosperm (Eudicots)	2439044	729796446 bp	24.04%
Nymphaea_colorata	Angiosperm (Basal)	352541	93276435 bp	22.81%
Olea_europaea	Angiosperm (Eudicots)	2451453	449864431 bp	34.17%
Oryza_sativa	Angiosperm (Monocots)	361697	79787375 bp	21.31%
Phalaenopsis_equestris	Angiosperm (Monocots)	1123275	387859297 bp	36.45%
Phaseolus_vulgaris	Angiosperm (Eudicots)	387921	106343957 bp	20.41%
Phoenix_dactylifera	Angiosperm (Monocots)	385815	101231084 bp	13.09%
Physalis_pubescens	Angiosperm (Eudicots)	702077	355823893 bp	25.61%
Physcomitrium_patens	Moss	105865	34376740 bp	7.28%
Pistacia_vera	Angiosperm (Eudicots)	450625	114620514 bp	17.07%
Populus_trichocarpa	Angiosperm (Eudicots)	399467	103447188 bp	26.37%
Prunus_persica	Angiosperm (Eudicots)	151840	41006384 bp	18.02%
Punica_granatum	Angiosperm (Eudicots)	269529	89840239 bp	28.03%
Salvia_hispanica	Angiosperm (Eudicots)	327757	72564732 bp	22.57%
Salvia_splendens	Angiosperm (Eudicots)	627415	155463347 bp	19.29%
Selaginella_moellendorffii	Lycopods	79062	35867924 bp	16.89%
Setaria_italica	Angiosperm (Monocots)	239840	67729328 bp	16.69%
Solanum_lycopersicum	Angiosperm (Eudicots)	710452	219993742 bp	26.57%
Solanum_melongena	Angiosperm (Eudicots)	962324	319299623 bp	27.42%
Solanum_pimpinellifolium	Angiosperm (Eudicots)	725276	238023301 bp	29.45%
Solanum_stenotomum	Angiosperm (Eudicots)	787416	209450434 bp	24.75%
Solanum_tuberosum	Angiosperm (Eudicots)	751728	197190543 bp	27.93%
Sphagnum_fallax	Moss	528940	133199787 bp	33.71%
Sphagnum_magellanicum	Moss	566915	144987832 bp	33.03%
Thuja_plicata	Gymnosperm	5112945	1987481838 bp	21.85%
Utricularia_gibba	Angiosperm (Eudicots)	46760	17420498 bp	17.30%
Vanilla_planifolia	Angiosperm (Monocots)	750493	488967057 bp	34.52%
Vicia_sativa	Angiosperm (Eudicots)	875356	294064471 bp	17.78%
Vitis_vinifera	Angiosperm (Eudicots)	368225	113188427 bp	23.28%
Welwitschia_mirabilis	Gymnosperm	6734176	2104850097 bp	30.65%
Zea_mays	Angiosperm (Monocots)	837546	359709792 bp	16.48%
Astrephomene gubernaculifera	Chlorophyte	40070	11349855 bp	10.93%
Auxenochlorella protothecoides (0710)	Chlorophyte	48692	12594245 bp	12.13%
Auxenochlorella protothecoides (UTEX 25)	Chlorophyte	648	319987 bp	1.51%
Bathycoccus prasinos	Chlorophyte	1201	354784 bp	2.35%
Chlamydomonas eustigma	Chlorophyte	7174	1751837 bp	2.63%
Chlamydomonas incerta	Chlorophyte	80379	11320685 bp	8.76%
Chlamydomonas reinhardtii	Chlorophyte	33671	7146352 bp	6.43%
Chlamydomonas schloesseri	Chlorophyte	49746	12415401 bp	9.54%
Chlamydomonas sp. UWO 241	Chlorophyte	286445	51069248 bp	24.13%
Chlorella desiccata (nom. nud.) (UTEX 2437)	Chlorophyte	1410	608452 bp	2.93%
Chlorella desiccata (nom. nud.) (UTEX 2526)	Chlorophyte	1935	659116 bp	3.06%
Chlorella ohadii	Chlorophyte	3131	804465 bp	1.41%
Chlorella sorokiniana	Chlorophyte	2347	1240924 bp	2.08%
Chlorella variabilis	Chlorophyte	3707	1023134 bp	2.22%
Chloropicon primus (CCMP 1205)	Chlorophyte	2969	737946 bp	4.24%
Chloropicon primus (RCC138)	Chlorophyte	3451	850942 bp	4.84%
Coccomyxa sp. Obi	Chlorophyte	5931	884838 bp	1.76%
Coccomyxa subellipsoidea C-169	Chlorophyte	9132	1062252 bp	2.18%
Dunaliella salina	Chlorophyte	323978	61259470 bp	17.82%
Edaphochlamys debaryana	Chlorophyte	36483	11210618 bp	7.89%
Haematococcus lacustris	Chlorophyte	264151	75750219 bp	24.49%
Helicosporidium sp. ATCC 50920	Chlorophyte	1131	223095 bp	1.80%
Micractinium conductrix	Chlorophyte	12812	4256088 bp	6.98%
Micromonas commoda	Chlorophyte	1604	347194 bp	1.64%
Micromonas pusilla CCMP1545	Chlorophyte	10660	1519793 bp	6.92%
Monoraphidium minutum	Chlorophyte	8628	2049762 bp	3.01%
Monoraphidium neglectum	Chlorophyte	10102	1442908 bp	2.07%
Ostreobium quekettii	Chlorophyte	77186	23765604 bp	15.65%
Ostreococcus lucimarinu	Chlorophyte	1033	570506 bp	4.32%
Ostreococcus tauri (RCC1115)	Chlorophyte	1185	404736 bp	2.74%
Ostreococcus tauri (RCC4221)	Chlorophyte	860	235602 bp	1.81%
Pedinophyceae sp. YPF-701	Chlorophyte	2127	778867 bp	2.79%
Picochlorum sp. BPE23	Chlorophyte	957	588300 bp	3.94%
Picochlorum sp. BPE23	Chlorophyte	957	588300 bp	3.94%
Pycnococcus provasolii	Chlorophyte	2302	510174 bp	2.25%
Raphidocelis subcapitata	Chlorophyte	3252	772664 bp	1.51%
Scenedesmus sp. NREL 46B-D3	Chlorophyte	107676	17972991 bp	11.83%
Tetrabaena socialis	Chlorophyte	69821	10668108 bp	7.86%
Trebouxia sp. A1-2	Chlorophyte	8064	1417979 bp	2.68%
Volvox africanus	Chlorophyte	43997	12486256 bp	9.65%
Volvox reticuliferus (NIES 3786)	Chlorophyte	47983	17130208 bp	12.79%
Volvox reticuliferus (NIES-3785)	Chlorophyte	45359	16178329 bp	12.16%

Across the spectrum of plant species analyzed, unclassified elements are universally present. However, a pronounced pattern emerges wherein the lineages of Angiosperms, Ferm, Hornwort, Moss, Gymnosperms, and Lycopods consistently exhibit a notably higher abundance of unclassified elements in terms of quantity, length, and proportion compared to the Chlorophyte lineage.

Within the Chlorophyte genomes, the proportion of unclassified elements spans a range from 1.51% in *Auxenochlorella protothecoides (UTEX 25)* to a substantial 24.49% in *Haematococcus lacustris*. The majority of these Chlorophyte species typically present a representation of unclassified elements falling between 1-10%.

In contrast, the other examined lineages showcase a broader spectrum of unclassified element proportions, ranging from 6.23% in *Arabidopsis thaliana* (Angiosperm) to a striking 44.33% in *Isothecium myosuroides* (Moss). Most plant lineages outside Chlorophytes display a prevalence of unclassified elements ranging between 10-35%.

The presented boxplots delineate the %Divergence metrics for these elements.



For the Unknown elements within the Angiosperms, Hornwort, Gymnosperms, Lycopods, and Liverworts lineages, the %Divergence exhibits overall stability, with median values ranging between 15 and 20%. However, Gymnospern stands out with a notably higher median value, approximately around 23%. Across these lineages, the IQR values hover around 10, and whisker lengths span from 0 to approximately 35. It's noteworthy that, except for Hornwort, the other lineages showcase a significant number of outliers surpassing the maximum %Divergence value.

In contrast, within the Chlorophyte lineage, the Unknown elements present a distinctly lower median %Divergence range, predominantly falling between 5 and 15%. A notable exception is observed with *Monoraphidium neglectum*, which boasts the highest median value, approximately 22%. The IQR values among Chlorophyte species vary, with narrower ranges observed in species like *Micromonas pusilla CCMP1545*, approximately around 5, and broader ranges in species like *Auxenochlorella protothecoides (UTEX25)*, which is approximately 15. The whisker lengths among these species are diverse, and roughly half of them exhibit outliers that exceed the maximum %Divergence value.

2.1.4 Small RNA Comparative and Analysis

Small RNAs are short RNA molecules that play important roles in regulating gene expression at the post-transcriptional level. They are integral components in various biological processes, including gene silencing, defense against viruses, and developmental regulation, among others. Given their critical roles, even small shifts in their abundance can have significant implications for cellular functions and organismal development.

The following table lists the number, length, and genome percentage of small RNA for each of the 138 plant genomes.

SPECIES	LINEAGE	Small RNA (number of elements)	Small RNA (lenth occupied)	Small RNA (percentage of sequence)
Acer_negundo	Angiosperm (Eudicots)	1229	344086 bp	0.08%
Acer_yangbiense	Angiosperm (Eudicots)	2653	812371 bp	0.12%
Actinidia_chinensis	Angiosperm (Eudicots)	982	190582 bp	0.03%
Adansonia_digitata	Angiosperm (Eudicots)	1101	348887 bp	0.05%
Adiantum_capillus	Fern	13029	2339337 bp	0.05%
Amborella_trichopoda	Angiosperm (Basal)	2589	642699 bp	0.09%
Ananas_comosus	Angiosperm (Monocots)	1056	326812 bp	0.09%
Annona_glabra	Angiosperm (Magnolids)	1196	588969 bp	0.06%
Anthoceros_angustus	Hornwort	537	149303 bp	0.13%
Arabidopsis_arenosa	Angiosperm (Eudicots)	566	203027 bp	0.14%
Arabidopsis_suecica	Angiosperm (Eudicots)	1577	912640 bp	0.34%
Arabidopsis_thaliana	Angiosperm (Eudicots)	304	30991 bp	0.03%
Arabis_alpina	Angiosperm (Eudicots)	384	81516 bp	0.03%
Boechera_stricta	Angiosperm (Eudicots)	528	120697 bp	0.06%
Brachypodium_distachyon	Angiosperm (Monocots)	1293	496317 bp	0.18%
Brassica_carinata	Angiosperm (Eudicots)	2529	2284441 bp	0.21%
Brassica_juncea	Angiosperm (Eudicots)	3684	7772346 bp	0.83%
Brassica_napus	Angiosperm (Eudicots)	8989	25015006 bp	2.50%
Brassica_nigra	Angiosperm (Eudicots)	1752	2808871 bp	0.53%
Brassica_oleracea	Angiosperm (Eudicots)	766	257836 bp	0.05%
Brassica_rapa	Angiosperm (Eudicots)	1867	1247835 bp	0.35%
Camelina_sativa	Angiosperm (Eudicots)	2252	596318 bp	0.09%
Camellia_sinensis	Angiosperm (Eudicots)	5136	2416588 bp	0.08%
Capsicum_annuum	Angiosperm (Eudicots)	6714	2923770 bp	0.09%
Carica_papaya	Angiosperm (Eudicots)	6849	7906534 bp	2.13%
Ceratodon_purpureus	Moss	2396	518026 bp	0.14%
Ceratopteris_richardii	Fern	0	0	0.00%
Chara_braunii	Algae (Charales)	109909	16036476 bp	0.92%
Chlamydomonas_reinhardtii	Algae (Chlorophyta)	55	26640 bp	0.02%
Cinnamomum_micranthum	Angiosperm (Magnolids)	3310	703207 bp	0.10%
Citrullus_lanatus	Angiosperm (Eudicots)	1929	498500 bp	0.14%
Citrus_sinensis	Angiosperm (Eudicots)	1261	426145 bp	0.14%
Cucumis_melo	Angiosperm (Eudicots)	15884	19873977 bp	4.53%
Cucumis_sativus	Angiosperm (Eudicots)	707	277779 bp	0.12%
Cycas_panzhihuaensis	Gymnosperm	5104	9428006 bp	0.09%
Cydonia_oblonga	Angiosperm (Eudicots)	2154	323450 bp	0.07%
Cynara_cardunculus	Angiosperm (Eudicots)	4351	384022 bp	0.05%
Cyperus_esculentus	Angiosperm (Monocots)	11245	15297332 bp	5.16%
Datura_stramonium	Angiosperm (Eudicots)	6334	14583826 bp	0.74%
Daucus_carota	Angiosperm (Eudicots)	1830	245203 bp	0.06%
Dendrobium_officinale	Angiosperm (Monocots)	283	200621 bp	0.02%
Fragaria_ananassa	Angiosperm (Eudicots)	4615	2742672 bp	0.34%
Fragaria_vesca	Angiosperm (Eudicots)	406	94557 bp	0.04%
Ginkgo_biloba	Gymnosperm	993	209037 bp	0.01%
Glycine_max	Angiosperm (Eudicots)	4636	1965247 bp	0.20%
Glycine_soja	Angiosperm (Eudicots)	4775	946282 bp	0.09%
Gnetum_montanum	Gymnosperm	2544	1960244 bp	0.09%
Gossypium_arboreum	Angiosperm (Eudicots)	5990	5875927 bp	0.36%
Gossypium_hirsutum	Angiosperm (Eudicots)	4571	3594270 bp	0.16%
Gossypium_raimondii	Angiosperm (Eudicots)	2699	2266898 bp	0.30%
Ipomoea_batatas	Angiosperm (Eudicots)	9279	851299 bp	0.10%
Isoetes_engelmannii	Lycopods	4864	830939 bp	0.13%
Isoetes_tatvanensis	Lycopods	6829	3297570 bp	0.20%
Isothecium_myosuroides	Moss	1117	1838770 bp	0.47%
Lactuca_sativa	Angiosperm (Eudicots)	2597	2827872 bp	0.11%
Lemna_minuta	Angiosperm (Monocots)	1847	433122 bp	0.12%
Lycium_ferocissimum	Angiosperm (Eudicots)	4173	1088988 bp	0.09%
Lycopodium_clavatum	Lycopods	6775	2441199 bp	0.11%
Malus_domestica	Angiosperm (Eudicots)	4369	878289 bp	0.12%
Mangifera_indica	Angiosperm (Eudicots)	384	847763 bp	0.22%
Marchantia_paleacea	Liverworts	1855	1299484 bp	0.52%
Marchantia_polymorpha	Liverworts	90	18616 bp	0.01%
Medicago_truncatula	Angiosperm (Eudicots)	2853	766224 bp	0.18%
Musa_acuminata	Angiosperm (Monocots)	583	129976 bp	0.03%
Nelumbo_nucifera	Angiosperm (Eudicots)	510	104099 bp	0.01%
Nicotiana_benthamiana	Angiosperm (Eudicots)	8517	3042526 bp	0.10%
Nymphaea_colorata	Angiosperm (Basal)	2930	956910 bp	0.23%
Olea_europaea	Angiosperm (Eudicots)	5796	819001 bp	0.06%
Oryza_sativa	Angiosperm (Monocots)	2169	965901 bp	0.26%
Phalaenopsis_aequestris	Angiosperm (Monocots)	426	97744 bp	0.01%
Phaseolus_vulgaris	Angiosperm (Eudicots)	395	94142 bp	0.02%
Phoenix_dactylifera	Angiosperm (Monocots)	1819	650721 bp	0.08%
Physalis_pubescens	Angiosperm (Eudicots)	3358	3044383 bp	0.22%
Physcomitrium_patens	Moss	369	346820 bp	0.07%
Pistacia_vera	Angiosperm (Eudicots)	1578	517729 bp	0.08%
Populus_trichocarpa	Angiosperm (Eudicots)	662	319863 bp	0.08%
Prunus_persica	Angiosperm (Eudicots)	1489	388396 bp	0.17%
Punica_granatum	Angiosperm (Eudicots)	1012	234376 bp	0.07%
Salvia_hispanica	Angiosperm (Eudicots)	1256	258398 bp	0.08%
Salvia_splendens	Angiosperm (Eudicots)	1754	636994 bp	0.08%
Selaginella_moellendorffii	Lycopods	4251	1760630 bp	0.83%
Setaria_italica	Angiosperm (Monocots)	461	206496 bp	0.05%
Solanum_lycopersicum	Angiosperm (Eudicots)	5352	2279453 bp	0.28%
Solanum_melongena	Angiosperm (Eudicots)	3182	835192 bp	0.07%
Solanum_pimpinellifolium	Angiosperm (Eudicots)	8231	2653867 bp	0.33%
Solanum_stenotomum	Angiosperm (Eudicots)	5172	738175 bp	0.09%
Solanum_tuberosum	Angiosperm (Eudicots)	3589	841433 bp	0.12%
Sphagnum_fallax	Moss	1174	179329 bp	0.05%
Sphagnum_magellanicum	Moss	1825	469312 bp	0.11%
Thuja_plicata	Gymnosperm	1565	476614 bp	0.01%
Utricularia_gibba	Angiosperm (Eudicots)	988	1676754 bp	1.67%
Vanilla_planifolia	Angiosperm (Monocots)	33733	12503766 bp	0.88%
Vicia_sativa	Angiosperm (Eudicots)	18745	35170382 bp	2.13%
Vitis_vinifera	Angiosperm (Eudicots)	2303	524258 bp	0.11%
Welwitschia_mirabilis	Gymnosperm	23449	5530940 bp	0.08%
Zea_mays	Angiosperm (Monocots)	8050	4130123 bp	0.19%
Astrephomene gubernaculifera	Chlorophyte	4231	5935946 bp	5.72%
Auxenochlorella protothecoides (0710)	Chlorophyte	3826	5837978 bp	5.62%
Auxenochlorella protothecoides (UTEX 25)	Chlorophyte	0	0	0.00%
Bathycoccus prasinos	Chlorophyte	0	0	0.00%
Chlamydomonas eustigma	Chlorophyte	0	0	0.00%
Chlamydomonas incerta	Chlorophyte	51	74328 bp	0.06%
Chlamydomonas reinhardtii	Chlorophyte	69	37283 bp	0.03%
Chlamydomonas schloesseri	Chlorophyte	64	138521 bp	0.11%
Chlamydomonas sp. UWO 241	Chlorophyte	1824	338075 bp	0.16%
Chlorella desiccata (nom. nud.) (UTEX 2437)	Chlorophyte	0	0	0.00%
Chlorella desiccata (nom. nud.) (UTEX 2526)	Chlorophyte	0	0	0.00%
Chlorella ohadii	Chlorophyte	0	0	0.00%
Chlorella sorokiniana	Chlorophyte	34	228525 bp	0.38%
Chlorella variabilis	Chlorophyte	0	0	0.00%
Chloropicon primus (CCMP 1205)	Chlorophyte	0	0	0.00%
Chloropicon primus (RCC138)	Chlorophyte	0	0	0.00%
Coccomyxa sp. Obi	Chlorophyte	37	66287 bp	0.13%
Coccomyxa subellipsoidea C-169	Chlorophyte	0	0	0.00%
Dunaliella salina	Chlorophyte	359	39090 bp	0.01%
Edaphochlamys debaryana	Chlorophyte	128	74366 bp	0.05%
Haematococcus lacustris	Chlorophyte	223	35237 bp	0.01%
Helicosporidium sp. ATCC 50920	Chlorophyte	0	0	0.00%
Micractinium conductrix	Chlorophyte	265	144592 bp	0.24%
Micromonas commoda	Chlorophyte	0	0	0.00%
Micromonas pusilla CCMP1545	Chlorophyte	0	0	0.00%
Monoraphidium minutum	Chlorophyte	0	0	0.00%
Monoraphidium neglectum	Chlorophyte	0	0	0.00%
Ostreobium quekettii	Chlorophyte	183	74949 bp	0.05%
Ostreococcus lucimarinus	Chlorophyte	0	0	0.00%
Ostreococcus tauri (RCC1115)	Chlorophyte	52	82855 bp	0.56%
Ostreococcus tauri (RCC4221)	Chlorophyte	0	0	0.00%
Pedinophyceae sp. YPF-701	Chlorophyte	0	0	0.00%
Picochlorum sp. BPE23	Chlorophyte	9	20716 bp	0.14%
Picochlorum sp. BPE23	Chlorophyte	9	20716 bp	0.14%
Pycnococcus provasolii	Chlorophyte	0	0	0.00%
Raphidocelis subcapitata	Chlorophyte	0	0	0.00%
Scenedesmus sp. NREL 46B-D3	Chlorophyte	266	58527 bp	0.04%
Tetrabaena socialis	Chlorophyte	0	0	0.00%
Treubaxia sp. AI-2	Chlorophyte	0	0	0.00%
Volvox africanus	Chlorophyte	3081	8160434 bp	6.31%
Volvox reticuliferus (NIES 3786)	Chlorophyte	2087	917560 bp	0.68%
Volvox reticuliferus (NIES-3785)	Chlorophyte	3249	809246 bp	0.61%

Across the diverse landscape of plant species analyzed, the representation of Small RNA remains notably sparse.

Within the Chlorophyte lineage, a striking pattern emerges where nearly half of the species surveyed lack discernible Small RNA. However, certain exceptions stand out. Notably, *Volvox africanus* displays a Small RNA proportion of 6.31%, followed closely by *Astrephomene gubernaculifera* at 5.72% and *Auxenochlorella protothecoides (0710)* at 5.62%. The remaining species within this lineage typically exhibit Small RNA proportions ranging between 0-0.68%.

In contrast, among the lineages of Angiosperms, Ferm, Hornwort, Moss, Gymnosperms, and Lycopods, *Cucumis melo* (Angiosperm) records the highest Small RNA representation at 4.53%. Intriguingly, the Fern species *Ceratopteris richardii* stands as a unique outlier, devoid of detectable Small RNA. For the majority of species within these lineages, the proportion of Small RNA remains predominantly below the 1% threshold.

2.1.5 Simple Repeat Element Comparative and Analysis

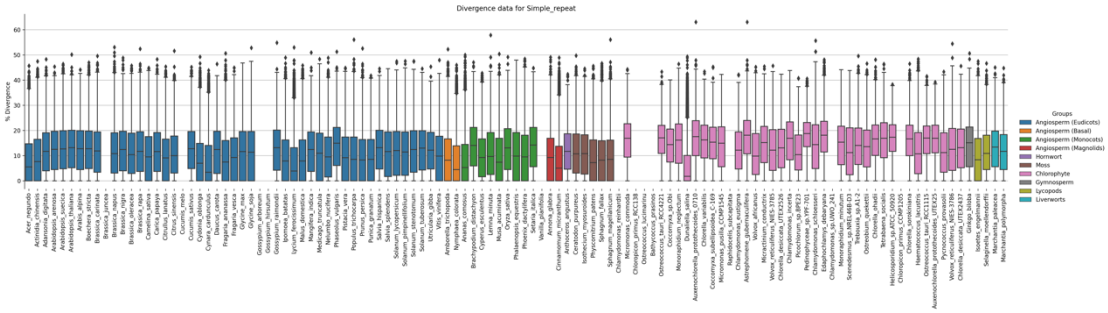
Simple repeats, often referred to as microsatellites or short tandem repeats (STRs), are short sequences of DNA motifs that are repeated consecutively multiple times. These repetitive sequences are ubiquitous across genomes and play roles in various genomic processes, including gene regulation, chromosomal organization, and genome stability. Their abundance and variability can offer insights into the evolutionary dynamics and genetic makeup of organisms.
The following table lists the number, length, and genome percentage of simple repeat elements for each of the 138 plant genomes.

SPECIES	LINEAGE	Simple repeats (number of elements)	Simple repeats (lenth occupied)	Simple repeats (percentage of sequence)
Acer_negundo	Angiosperm (Eudicots)	216119	7980535 bp	1.80%
Acer_yangbiense	Angiosperm (Eudicots)	239501	9299191 bp	1.40%
Actinidia_chinensis	Angiosperm (Eudicots)	156921	5882583 bp	1.06%
Adansonia_digitata	Angiosperm (Eudicots)	226504	8931992 bp	1.30%
Adiantum_capillus	Fern	1150482	62486243 bp	1.30%
Amborella_trichopoda	Angiosperm (Basal)	168689	10254994 bp	1.45%
Ananas_comosus	Angiosperm (Monocots)	170938	7886524 bp	2.06%
Annona_glabra	Angiosperm (Magnolids)	183931	6681055 bp	0.65%
Anthoceros_angustus	Hornwort	18880	1004233 bp	0.84%
Arabidopsis_arenosa	Angiosperm (Eudicots)	44312	1740104 bp	1.16%
Arabidopsis_suecica	Angiosperm (Eudicots)	78273	3122193 bp	1.15%
Arabidopsis_thaliana	Angiosperm (Eudicots)	34625	1364872 bp	1.14%
Arabis_alpina	Angiosperm (Eudicots)	46948	1929752 bp	0.62%
Boechera_stricta	Angiosperm (Eudicots)	51062	2123258 bp	1.11%
Brachypodium_distachyon	Angiosperm (Monocots)	44534	1868055 bp	0.69%
Brassica_carinata	Angiosperm (Eudicots)	199749	11572338 bp	1.06%
Brassica_juncea	Angiosperm (Eudicots)	166310	6920826 bp	0.74%
Brassica_napus	Angiosperm (Eudicots)	183943	7918357 bp	0.79%
Brassica_nigra	Angiosperm (Eudicots)	85814	4050821 bp	0.76%
Brassica_oleracea	Angiosperm (Eudicots)	91722	3838788 bp	0.79%
Brassica_rapa	Angiosperm (Eudicots)	74227	3079625 bp	0.87%
Camelina_sativa	Angiosperm (Eudicots)	186422	7029626 bp	1.10%
Camellia_sinensis	Angiosperm (Eudicots)	630211	25595728 bp	0.82%
Capsicum_annuum	Angiosperm (Eudicots)	279674	12699613 bp	0.40%
Carica_papaya	Angiosperm (Eudicots)	102294	4132927 bp	1.12%
Ceratodon_purpureus	Moss	55652	2711214 bp	0.75%
Ceratopteris_richardii	Fern	1025552	417886560 bp	5.60%
Chara_braunii	Algae (Charales)	785394	41815766 bp	2.39%
Chlamydomonas_reinhardtii	Algae (Chlorophyta)	134153	8856305 bp	7.97%
Cinnamomum_micranthum	Angiosperm (Magnolids)	239595	8282682 bp	1.13%
Citrullus_lanatus	Angiosperm (Eudicots)	145137	5442063 bp	1.51%
Citrus_sinensis	Angiosperm (Eudicots)	85262	3255718 bp	1.09%
Cucumis_melo	Angiosperm (Eudicots)	130413	5631800 bp	1.28%
Cucumis_sativus	Angiosperm (Eudicots)	91557	3703318 bp	1.63%
Cycas_panzhihuaensis	Gymnosperm	964483	168403281 bp	1.61%
Cydonia_oblonga	Angiosperm (Eudicots)	132844	4997358 bp	1.02%
Cynara_cardunculus	Angiosperm (Eudicots)	153792	6553502 bp	0.90%
Cyperus_esculentus	Angiosperm (Monocots)	123005	5825357 bp	1.96%
Datura_stramonium	Angiosperm (Eudicots)	164338	8307676 bp	0.42%
Daucus_carota	Angiosperm (Eudicots)	73290	3309077 bp	0.78%
Dendrobium_officinale	Angiosperm (Monocots)	227153	14108215 bp	1.15%
Fragaria_ananassa	Angiosperm (Eudicots)	187697	7576361 bp	0.94%
Fragaria_vesca	Angiosperm (Eudicots)	55480	2191992 bp	1.02%
Ginkgo_biloba	Gymnosperm	134293	6085465 bp	0.23%
Glycine_max	Angiosperm (Eudicots)	239175	10744244 bp	1.10%
Glycine_soja	Angiosperm (Eudicots)	244417	11057772 bp	1.09%
Gnetum_montanum	Gymnosperm	101453	4563683 bp	0.21%
Gossypium_arboreum	Angiosperm (Eudicots)	161236	7161254 bp	0.44%
Gossypium_hirsutum	Angiosperm (Eudicots)	253783	11778002 bp	0.51%
Gossypium_raimondii	Angiosperm (Eudicots)	119992	5265445 bp	0.70%
Ipomoea_batatas	Angiosperm (Eudicots)	285929	11855713 bp	1.42%
Isoetes_engelmannii	Lycopods	100994	4776477 bp	0.75%
Isoetes_taiwanensis	Lycopods	135930	8977573 bp	0.54%
Isothecium_myosuroides	Moss	39166	1765016 bp	0.45%
Lactuca_sativa	Angiosperm (Eudicots)	455619	25837336 bp	1.00%
Lemna_minuta	Angiosperm (Monocots)	62815	3948166 bp	1.10%
Lycium_ferocissimum	Angiosperm (Eudicots)	205579	8843951 bp	0.73%
Lycopodium_clavatum	Lycopods	290222	16998745 bp	0.74%
Malus_domestica	Angiosperm (Eudicots)	119727	4811980 bp	0.68%
Mangifera_indica	Angiosperm (Eudicots)	90264	4656129 bp	1.19%
Marchantia_paleacea	Liverworts	36023	1463820 bp	0.58%
Marchantia_polymorpha	Liverworts	51777	2017629 bp	0.89%
Medicago_truncatula	Angiosperm (Eudicots)	103286	4299562 bp	1.00%
Musa_acuminata	Angiosperm (Monocots)	100246	4217653 bp	0.89%
Nelumbo_nucifera	Angiosperm (Eudicots)	168601	7235483 bp	0.90%
Nicotiana_benthamiana	Angiosperm (Eudicots)	290146	14545708 bp	0.48%
Nymphaea_colorata	Angiosperm (Basal)	118284	5016483 bp	1.23%
Olea_europaea	Angiosperm (Eudicots)	240977	9091452 bp	0.69%
Oryza_sativa	Angiosperm (Monocots)	96355	4409375 bp	1.18%
Phalaenopsis_equestris	Angiosperm (Monocots)	180283	11034902 bp	1.04%
Phaseolus_vulgaris	Angiosperm (Eudicots)	110406	5484400 bp	1.05%
Phoenix_dactylifera	Angiosperm (Monocots)	148440	6927681 bp	0.90%
Physalis_pubescens	Angiosperm (Eudicots)	111989	5085878 bp	0.37%
Physcomitrium_patens	Moss	173701	6861714 bp	1.45%
Pistacia_vera	Angiosperm (Eudicots)	153767	6014269 bp	0.90%
Populus_trichocarpa	Angiosperm (Eudicots)	113002	4309088 bp	1.10%
Prunus_persica	Angiosperm (Eudicots)	74760	2781607 bp	1.22%
Punica_granatum	Angiosperm (Eudicots)	92832	3396816 bp	1.06%
Salvia_hispanica	Angiosperm (Eudicots)	68783	2975443 bp	0.93%
Salvia_splendens	Angiosperm (Eudicots)	134023	5889180 bp	0.73%
Selaginella_moellendorffii	Lycopods	35754	1665109 bp	0.78%
Setaria_italica	Angiosperm (Monocots)	60686	2875102 bp	0.71%
Solanum_lycopersicum	Angiosperm (Eudicots)	106107	6220521 bp	0.75%
Solanum_melongena	Angiosperm (Eudicots)	122660	6336681 bp	0.54%
Solanum_pimpinellifolium	Angiosperm (Eudicots)	112890	6479954 bp	0.80%
Solanum_stenotomum	Angiosperm (Eudicots)	106195	5236532 bp	0.62%
Solanum_tuberosum	Angiosperm (Eudicots)	94486	6114333 bp	0.87%
Sphagnum_fallax	Moss	111726	4117712 bp	1.04%
Sphagnum_magellanicum	Moss	114686	4377903 bp	1.00%
Thuja_plicata	Gymnosperm	690845	49267313 bp	0.54%
Utricularia_gibba	Angiosperm (Eudicots)	18972	741826 bp	0.74%
Vanilla_planifolia	Angiosperm (Monocots)	402897	259247438 bp	18.30%
Vicia_sativa	Angiosperm (Eudicots)	198155	11652949 bp	0.70%
Vitis_vinifera	Angiosperm (Eudicots)	135155	5232798 bp	1.08%
Welwitschia_mirabilis	Gymnosperm	712709	59627004 bp	0.87%
Zea_mays	Angiosperm (Monocots)	124313	6583005 bp	0.30%
Astrephomene gubernaculifera	Chlorophyte	83244	4906531 bp	4.72%
Auxenochlorella protothecoides (0710)	Chlorophyte	80659	4339408 bp	4.18%
Auxenochlorella protothecoides (UTEX 25)	Chlorophyte	6692	309359 bp	1.46%
Bathycoccus prasinos	Chlorophyte	14792	561766 bp	3.73%
Chlamydomonas eustigma	Chlorophyte	10416	461491 bp	0.69%
Chlamydomonas incerta	Chlorophyte	145558	9162993 bp	7.09%
Chlamydomonas reinhardtii	Chlorophyte	128360	7568749 bp	6.81%
Chlamydomonas schloesseri	Chlorophyte	165410	9643577 bp	7.41%
Chlamydomonas sp. UWO 241	Chlorophyte	325473	19742756 bp	9.33%
Chlorella desiccata (nom. nud.) (UTEX 2437)	Chlorophyte	6670	254164 bp	1.23%
Chlorella desiccata (nom. nud.) (UTEX 2526)	Chlorophyte	6856	257987 bp	1.20%
Chlorella ohadii	Chlorophyte	47643	2464557 bp	4.32%
Chlorella sorokiniana	Chlorophyte	54937	3057484 bp	5.13%
Chlorella variabilis	Chlorophyte	58420	3038107 bp	6.58%
Chloropicon primus (CCMP 1205)	Chlorophyte	15298	581738 bp	3.34%
Chloropicon primus (RCC138)	Chlorophyte	15307	585947 bp	3.33%
Coccomyxa sp. Obi	Chlorophyte	7042	330049 bp	0.65%
Coccomyxa subellipsoidea C-169	Chlorophyte	7164	349292 bp	0.72%
Dunaliella salina	Chlorophyte	138242	8687227 bp	2.53%
Edaphochlamys debaryana	Chlorophyte	113393	6534007 bp	4.60%
Haematococcus lacustris	Chlorophyte	221539	24135783 bp	7.80%
Helicosporidium sp. ATCC 50920	Chlorophyte	2569	114252 bp	0.92%
Micractinium conductrix	Chlorophyte	91518	4987179 bp	8.17%
Micromonas commoda	Chlorophyte	13332	652938 bp	3.09%
Micromonas pusilla CCMP1545	Chlorophyte	35056	1807403 bp	8.23%
Monoraphidium minutum	Chlorophyte	130603	7981439 bp	11.71%
Monoraphidium neglectum	Chlorophyte	69185	4174548 bp	5.99%
Ostreobium quekettii	Chlorophyte	16001	789417 bp	0.52%
Ostreococcus lucimarinu	Chlorophyte	9907	613178 bp	4.64%
Ostreococcus tauri (RCC1115)	Chlorophyte	8176	488356 bp	3.31%
Ostreococcus tauri (RCC4221)	Chlorophyte	8012	470460 bp	3.61%
Pedinophyceae sp. YPF-701	Chlorophyte	14713	815279 bp	2.92%
Picochlorum sp. BPE23	Chlorophyte	2262	82192 bp	0.55%
Picochlorum sp. BPE23	Chlorophyte	2262	82192 bp	0.55%
Pycnococcus provasolii	Chlorophyte	19498	821222 bp	3.62%
Raphidocelis subcapitata	Chlorophyte	86247	5308881 bp	10.38%
Scenedesmus sp. NREL 46B-D3	Chlorophyte	137213	9654960 bp	6.36%
Tetrabaena socialis	Chlorophyte	77584	4335814 bp	3.19%
Treboxia sp. A1-2	Chlorophyte	11383	633824 bp	1.20%
Volvox africanus	Chlorophyte	75462	4157060 bp	3.21%
Volvox reticuliferus (NIES 3786)	Chlorophyte	67181	3742659 bp	2.79%
Volvox reticuliferus (NIES-3785)	Chlorophyte	65641	3343448 bp	2.51%

Across the diverse array of plant species examined, the presence of Simple Repeats is universal. Within the lineages of Angiosperms, Ferm, Hornwort, Moss, Gymnosperms, and Lycopods, the proportion of the genome occupied by Simple Repeats remains predominantly below the 2% mark. However, some exceptions capture attention. Notably, *Vanilla planifolia* (Angiosperm) stands out with a substantial representation of Simple Repeats, accounting for 18.3% of its genome. In contrast, *Ginkgo biloba* (Gymnosperm) presents the most modest representation at a mere 0.23%.

Interestingly, the scenario shifts when examining the Chlorophyte lineage. Here, in contrast to the other lineages, the proportion of Simple Repeats tends to be more pronounced. Within this lineage, several species exhibit notable percentages of Simple Repeats, such as *Monoraphidium minutum* at 11.71%, *Raphidocelis subcapitata* at 10.38%, *Chlamydomonas sp. UWO 241* at 9.33%, and *Micromonas pusilla CCMP1545* at 8.23%. The least representation within this lineage is observed in *Ostreobium quekettii* at 0.52%.

The presented boxplots delineate the %Divergence metrics for these elements.



Across the lineages of Angiosperms, Hornwort, Gymnosperms, Lycopods, and Liverworts, a notably stable pattern emerges in the boxplot. The overall median %Divergence for Simple Repeats in these lineages tends to be low, ranging approximately from 5 to 15. *Ginkgo biloba*, representing the Gymnosperms, stands out with the highest median value among these lineages. Furthermore, the IQR values for these plants remain consistently within the range of 15-20, indicating a relatively uniform distribution of %Divergence.

Contrastingly, the Chlorophyte lineage exhibits heightened variability in Simple Repeats. The median %Divergence values for these species oscillate between 10 and 20, with IQR typically falling within the 10-15 range, reflecting a broader spectrum of %Divergence values compared to other lineages.

Across all lineages, the whisker lengths display a common lower bound of 0. However, the upper bounds span a wider range, typically between 35 and 45. Many plants across these lineages manifest numerous outliers beyond the upper limit. An exception is *Dunaliella salina* (Chlorophyte), characterized by whisker lengths ranging from 0 to 25. This species presents a notably low median value of around 3, coupled with a narrow IQR of approximately 10, and remarkably exhibits the highest count of outliers.

2.1.6 Low Complexity Element Comparative and Analysis

Low Complexity Elements (LCEs) represent regions of DNA or protein sequences that are composed of a limited set of repetitive units or motifs. These regions typically contain a high content of a few specific nucleotides or amino acids, making them less information-rich than other genomic regions. LCEs are often associated with certain functional roles, such as protein binding, but can also be involved in structural or regulatory functions. For instance, LCEs are prominently found in specialized regions such as the chromosomal centromeres, where they are essential for ensuring proper chromosome segregation during cell division. Similarly, in the telomeres, LCEs are instrumental in preserving chromosomal integrity, stability, and regulating cellular aging processes. Thus, while LCEs may be less information-rich, their functional significance in maintaining genomic stability and facilitating essential cellular processes is undeniable.

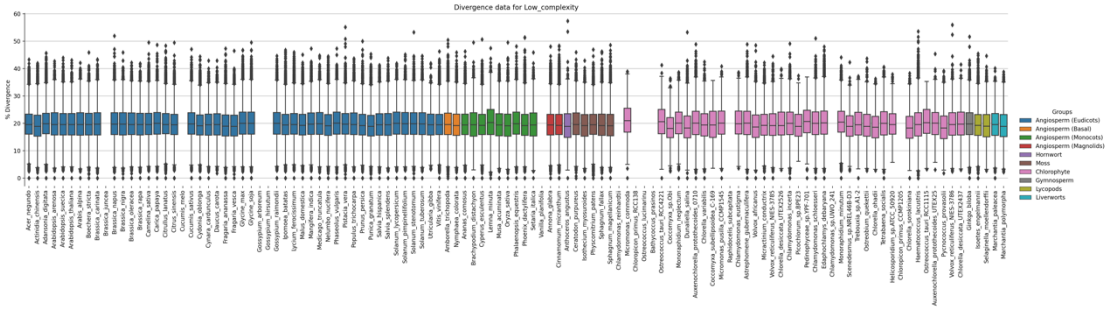
The following table lists the number, length, and genome percentage of low-complexity elements for each of the 138 plant genomes.

SPECIES	LINEAGE	Low complexity (number of elements)	Low complexity (both occupied)	Low complexity (percentage of sequence)
Acer negundo	Angiosperm (Eudicots)	29096	1443782 bp	0.33%
Acer yangbiense	Angiosperm (Eudicots)	33164	1654735 bp	0.25%
Actinidia chinensis	Angiosperm (Eudicots)	27519	1363484 bp	0.25%
Adansonia digitata	Angiosperm (Eudicots)	45945	2303348 bp	0.34%
Adiantum capillus	Fern	29042	1593944 bp	0.03%
Amborella trichopoda	Angiosperm (Basal)	30027	1834483 bp	0.28%
Ananas comosus	Angiosperm (Monocots)	22722	1266050 bp	0.34%
Annona glabra	Angiosperm (Magnolids)	32778	1585697 bp	0.15%
Anthoceros angustus	Hornwort	2234	154969 bp	0.13%
Arabidopsis arenosa	Angiosperm (Eudicots)	10525	506979 bp	0.34%
Arabidopsis suecica	Angiosperm (Eudicots)	18542	890142 bp	0.33%
Arabidopsis thaliana	Angiosperm (Eudicots)	8612	415816 bp	0.35%
Arabis alpina	Angiosperm (Eudicots)	10805	530163 bp	0.17%
Bouchea viraea	Angiosperm (Eudicots)	11372	554131 bp	0.29%
Brachypodium distachyon	Angiosperm (Monocots)	6186	307026 bp	0.11%
Brassica carinata	Angiosperm (Eudicots)	38312	1914413 bp	0.18%
Brassica juncea	Angiosperm (Eudicots)	34432	1727540 bp	0.19%
Brassica napus	Angiosperm (Eudicots)	37240	1860405 bp	0.19%
Brassica nigra	Angiosperm (Eudicots)	18758	944873 bp	0.18%
Brassica oleracea	Angiosperm (Eudicots)	18148	896214 bp	0.18%
Brassica rapa	Angiosperm (Eudicots)	15664	770303 bp	0.22%
Camelina sativa	Angiosperm (Eudicots)	35479	1692440 bp	0.26%
Camellia sinensis	Angiosperm (Eudicots)	110978	5619497 bp	0.18%
Capsicum annuum	Angiosperm (Eudicots)	46867	2434615 bp	0.08%
Carica papaya	Angiosperm (Eudicots)	16261	838966 bp	0.23%
Ceratodon purpureus	Moss	10532	598337 bp	0.17%
Ceratopteris richardii	Fern	57126	4379841 bp	0.06%
Chara braunii	Algae (Charales)	74079	4976757 bp	0.28%
Chlamydomonas reinhardtii	Algae (Chlorophyta)	12464	777651 bp	0.70%
Cinnamomum micranthum	Angiosperm (Magnolids)	33790	1669883 bp	0.23%
Citrullus lanatus	Angiosperm (Eudicots)	27916	1369144 bp	0.38%
Citrus sinensis	Angiosperm (Eudicots)	15537	722604 bp	0.24%
Cucumis melo	Angiosperm (Eudicots)	26459	1327797 bp	0.30%
Cucumis sativus	Angiosperm (Eudicots)	23138	1203817 bp	0.53%
Cycas panzhihuaensis	Gymnosperm	95411	7645820 bp	0.07%
Cydonia oblonga	Angiosperm (Eudicots)	23301	1145437 bp	0.23%
Cynara cardunculus	Angiosperm (Eudicots)	14204	704849 bp	0.10%
Cyperus esculentus	Angiosperm (Monocots)	18405	1048732 bp	0.33%
Datura stramonium	Angiosperm (Eudicots)	31218	1576834 bp	0.08%
Daucus carota	Angiosperm (Eudicots)	13016	690301 bp	0.16%
Dendrobium officinale	Angiosperm (Monocots)	36279	2019827 bp	0.17%
Fragaria ananassa	Angiosperm (Eudicots)	31209	1512971 bp	0.19%
Fragaria vesca	Angiosperm (Eudicots)	10085	486618 bp	0.23%
Ginkgo biloba	Gymnosperm	22843	1116762 bp	0.04%
Glycine max	Angiosperm (Eudicots)	50545	2553129 bp	0.26%
Glycine soja	Angiosperm (Eudicots)	51206	2605117 bp	0.26%
Gnetum montanum	Gymnosperm	19176	930392 bp	0.04%
Gossypium arboreum	Angiosperm (Eudicots)	39168	2091152 bp	0.13%
Gossypium hirsutum	Angiosperm (Eudicots)	69641	3654799 bp	0.16%
Gossypium raimondii	Angiosperm (Eudicots)	30138	1578107 bp	0.21%
Ipomoea batatas	Angiosperm (Eudicots)	38444	1839508 bp	0.22%
Isoetes engelmannii	Lycopods	12577	625776 bp	0.10%
Isoetes taiwanensis	Lycopods	15697	1469482 bp	0.09%
Isothecium misserendeo	Moss	6229	315292 bp	0.08%
Lactuca sativa	Angiosperm (Eudicots)	38227	1902664 bp	0.07%
Lemna minuta	Angiosperm (Monocots)	14819	1430489 bp	0.40%
Lycium ferocissimum	Angiosperm (Eudicots)	27201	1311042 bp	0.11%
Lycopodium clavatum	Lycopods	21457	1332617 bp	0.06%
Malus domestica	Angiosperm (Eudicots)	21568	1054880 bp	0.15%
Mangifera indica	Angiosperm (Eudicots)	17136	834930 bp	0.21%
Marchantia paleacea	Liverworts	7083	369642 bp	0.15%
Marchantia polymorpha	Liverworts	7871	413987 bp	0.18%
Medicago truncatula	Angiosperm (Eudicots)	20761	1018098 bp	0.23%
Musa acuminata	Angiosperm (Monocots)	13480	670246 bp	0.14%
Nelumbo nucifera	Angiosperm (Eudicots)	27169	1339974 bp	0.17%
Nicotiana benthamiana	Angiosperm (Eudicots)	47264	2666152 bp	0.09%
Nymphaea colorata	Angiosperm (Basal)	15511	787985 bp	0.19%
Olea europaea	Angiosperm (Eudicots)	40175	1890948 bp	0.14%
Oryza sativa	Angiosperm (Monocots)	10894	559336 bp	0.15%
Phalaenopsis equestris	Angiosperm (Monocots)	30558	2032333 bp	0.19%
Phaseolus vulgaris	Angiosperm (Eudicots)	27255	1438272 bp	0.20%
Phoenix dactylifera	Angiosperm (Monocots)	24911	1354650 bp	0.18%
Physalis pubescens	Angiosperm (Eudicots)	21224	1117068 bp	0.08%
Physcomitrium patens	Moss	24180	1237004 bp	0.26%
Pistacia vera	Angiosperm (Eudicots)	20432	1319880 bp	0.20%
Populus trichocarpa	Angiosperm (Eudicots)	21203	1028395 bp	0.26%
Prunus persica	Angiosperm (Eudicots)	13531	648600 bp	0.29%
Punica granatum	Angiosperm (Eudicots)	13112	625598 bp	0.19%
Salvia hispanica	Angiosperm (Eudicots)	13814	675430 bp	0.21%
Salvia splendens	Angiosperm (Eudicots)	23747	1171435 bp	0.15%
Selaginella moellendorffii	Lycopods	5438	260555 bp	0.12%
Setaria italica	Angiosperm (Monocots)	7174	382837 bp	0.09%
Solanum lycopersicum	Angiosperm (Eudicots)	20666	1048419 bp	0.13%
Solanum melongena	Angiosperm (Eudicots)	21787	1097269 bp	0.09%
Solanum pimpinellifolium	Angiosperm (Eudicots)	22137	1111207 bp	0.14%
Solanum stenotomum	Angiosperm (Eudicots)	21033	1242554 bp	0.15%
Solanum tuberosum	Angiosperm (Eudicots)	19511	1331915 bp	0.19%
Sphagnum fallax	Moss	15691	789487 bp	0.20%
Sphagnum magellanicum	Moss	15938	808253 bp	0.18%
Thuja plicata	Gymnosperm	95629	5819301 bp	0.06%
Utricularia gibba	Angiosperm (Eudicots)	3481	168057 bp	0.17%
Vanilla planifolia	Angiosperm (Monocots)	30847	12783742 bp	9.03%
Vicia sativa	Angiosperm (Eudicots)	35955	1901712 bp	0.12%
Vitis vinifera	Angiosperm (Eudicots)	29767	1467813 bp	0.30%
Welwitschia mirabilis	Gymnosperm	157626	8221848 bp	0.12%
Zea mays	Angiosperm (Monocots)	15437	842385 bp	0.04%
Astrephomene gubernaculifera	Chlorophyte	15674	934658 bp	0.90%
Asunacchlorella protothecoides (0710)	Chlorophyte	15727	938262 bp	0.90%
Auxenochlorella protothecoides (UTEX 25)	Chlorophyte	1299	62505 bp	0.29%
Bathycoccus prasinos	Chlorophyte	2275	124116 bp	0.82%
Chlamydomonas eustigma	Chlorophyte	668	36822 bp	0.06%
Chlamydomonas incerta	Chlorophyte	14150	854044 bp	0.66%
Chlamydomonas reinhardtii	Chlorophyte	12630	787657 bp	0.71%
Chlamydomonas schlosseri	Chlorophyte	13124	825705 bp	0.63%
Chlamydomonas sp. UWO 241	Chlorophyte	26674	1911827 bp	0.90%
Chlorella desiccata (nom. nud.) (UTEX 2417)	Chlorophyte	770	35896 bp	0.17%
Chlorella desiccata (nom. nud.) (UTEX 2526)	Chlorophyte	862	39514 bp	0.18%
Chlorella ohadii	Chlorophyte	2068	101976 bp	0.18%
Chlorella sorokiniana	Chlorophyte	3002	146743 bp	0.25%
Chlorella variabilis	Chlorophyte	4555	246826 bp	0.53%
Chlorogonium prinus (CCMP 1205)	Chlorophyte	2206	114244 bp	0.66%
Chlorogonium prinus (RCC 138)	Chlorophyte	2200	114398 bp	0.65%
Coccomyxa sp. Obi	Chlorophyte	792	40082 bp	0.08%
Coccomyxa subellipsoidea C-169	Chlorophyte	391	19363 bp	0.04%
Dunaliella salina	Chlorophyte	4384	271682 bp	0.08%
Edaphochlamys debaryana	Chlorophyte	28527	1684350 bp	1.18%
Haematococcus lacustris	Chlorophyte	23502	1617176 bp	0.52%
Halosarpheium sp. ATCC 50920	Chlorophyte	443	22289 bp	0.18%
Micractinium conductrix	Chlorophyte	7921	434516 bp	0.71%
Micromonas commoda	Chlorophyte	907	49195 bp	0.23%
Micromonas pusilla CCMP1545	Chlorophyte	1144	59240 bp	0.27%
Monoraphidium minutum	Chlorophyte	12955	715459 bp	1.05%
Monoraphidium neglectum	Chlorophyte	3794	190185 bp	0.28%
Ostreobium quekett	Chlorophyte	2398	121807 bp	0.08%
Ostreococcus lucimarinus	Chlorophyte	181	9784 bp	0.07%
Ostreococcus tauri (RCC1113)	Chlorophyte	224	12140 bp	0.08%
Ostreococcus tauri (RCC221)	Chlorophyte	237	12623 bp	0.18%
Pedinophyceae sp. YPF-701	Chlorophyte	1320	74930 bp	0.27%
Picochlorum sp. BPE23	Chlorophyte	275	14126 bp	0.09%
Picochlorum sp. BPE23	Chlorophyte	275	14126 bp	0.09%
Pyrenocystis pronuncoli	Chlorophyte	1327	69257 bp	0.31%
Raphidocelis subcapitata	Chlorophyte	8914	504934 bp	0.99%
Scenedesmus sp. NREL 46B-D3	Chlorophyte	1246	66513 bp	0.04%
Tetrabaena socialis	Chlorophyte	10422	587550 bp	0.42%
Trebouxia sp. A1-2	Chlorophyte	506	29021 bp	0.05%
Volvox africanus	Chlorophyte	4703	285006 bp	0.22%
Vulvax reticuliferus (NIES 3786)	Chlorophyte	4007	267552 bp	0.20%
Vulvax reticuliferus (NIES-3785)	Chlorophyte	4115	238713 bp	0.18%

Across the spectrum of plant lineages examined, the presence of Low Complexity Elements is pervasive, with their representation being notably consistent across different species. Despite their ubiquitous presence, the proportion occupied by LCEs in these genomes remains relatively modest. The highest proportion observed is in *Edaphochlamys debaryana* (Chlorophyte), where LCEs account for 1.18% of the sequence.

In stark contrast, all the rest of the plant species, spanning various lineages, exhibit LCE proportions below the 1% threshold. Notably, *Adiantum_capillus* (Fern) showcases the most minimal representation, with LCEs accounting for a mere 0.03% of its genome. It is noteworthy to mention that several species, including *Ginkgo_biloba* and *Gnetum_montanum* (both Gymnosperms), as well as *Scenedesmus sp. NREL 46B-D3* (Chlorophyte), demonstrate a consistent low representation of LCEs, each registering at 0.04%.

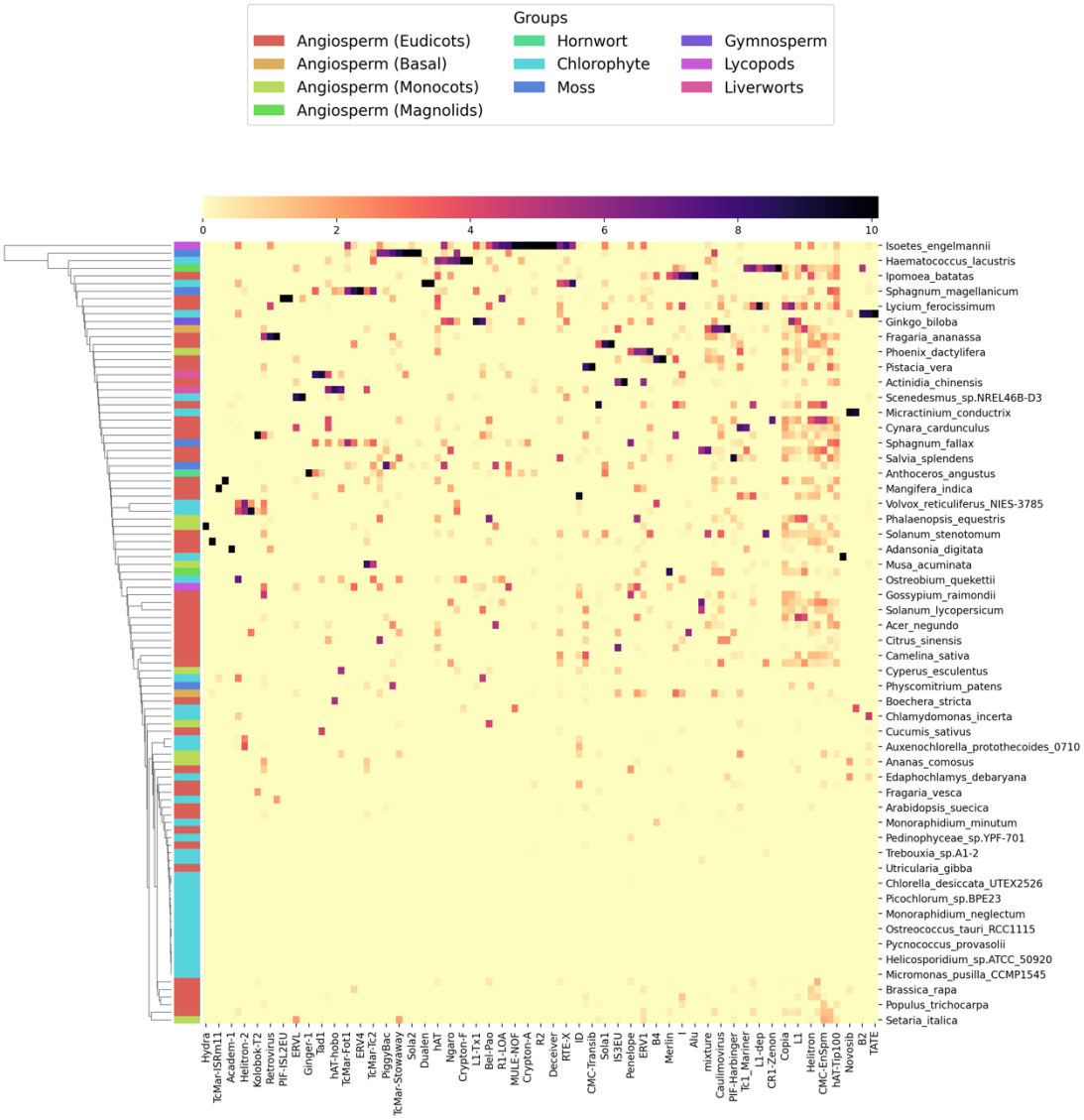
The presented boxplots delineate the %Divergence metrics for these elements.



Across all examined lineages, a striking consistency emerges in the %Divergence values associated with Low Complexity elements. Regardless of the lineage, the median %Divergence hovers around the 20 mark, showcasing an interesting uniformity without any deviations. This uniform median suggests that, despite the evolutionary differences and varied genomic contexts across these lineages, there might be consistent underlying processes or constraints governing the evolution or retention of low-complexity elements.

Additionally, the IQR values, ranging approximately from 16 to 23, further highlight this stability. Such a narrow IQR indicates that the majority of plants within these lineages share a similar %Divergence distribution, with minimal variability. The whisker lengths, spanning from 5 to 35, suggest that while there is some range in the data, most plants tend to fall within this specified range. However, the presence of outliers at both ends of the whiskers across the majority of these lineages might indicate specific evolutionary pressures or events affecting a subset of plants, leading to extreme %Divergence values.

2.2 Heatmap Representation



The heatmap delineates the abundance distribution of transposon elements across seven distinct plant lineages. The horizontal axis maps approximately fifty different transposon elements, while the vertical axis spans seven plant lineages: Angiosperms, Liverworts, Hornwort, Moss, Gymnosperms, Lycopods, and Chlorophyte.

Angiosperms: Within the Angiosperms, particularly among the Eudicots, elements such as Copia, L1, Helitron, CMC-EnSpm, and hAT-Tip100 consistently exhibit elevated abundance levels. In contrast, within the Basal Angiosperms, the element 'mixture' demonstrates exceptionally high abundance. Among the Monocots, while elements like Penelope and ERV1 manifest notably high abundance in certain species, others maintain a relatively consistent low overall TE abundance. Within the Magnolids, the element Merlin stands out with its remarkably high abundance.

Liverworts: Notably, Tad1 and hAT-hobo display heightened abundance levels. Such an increase might signify lineage-specific amplification events, which can contribute to genetic diversification or response to environmental pressures within Liverworts.

Moss: Elements like PiggyBac, TcMar-Stowaway, and Sola2 exhibit notably high abundances. Their elevated presence suggests lineage-specific amplification dynamics, potentially influencing the genomic architecture and evolutionary trajectories within Moss.

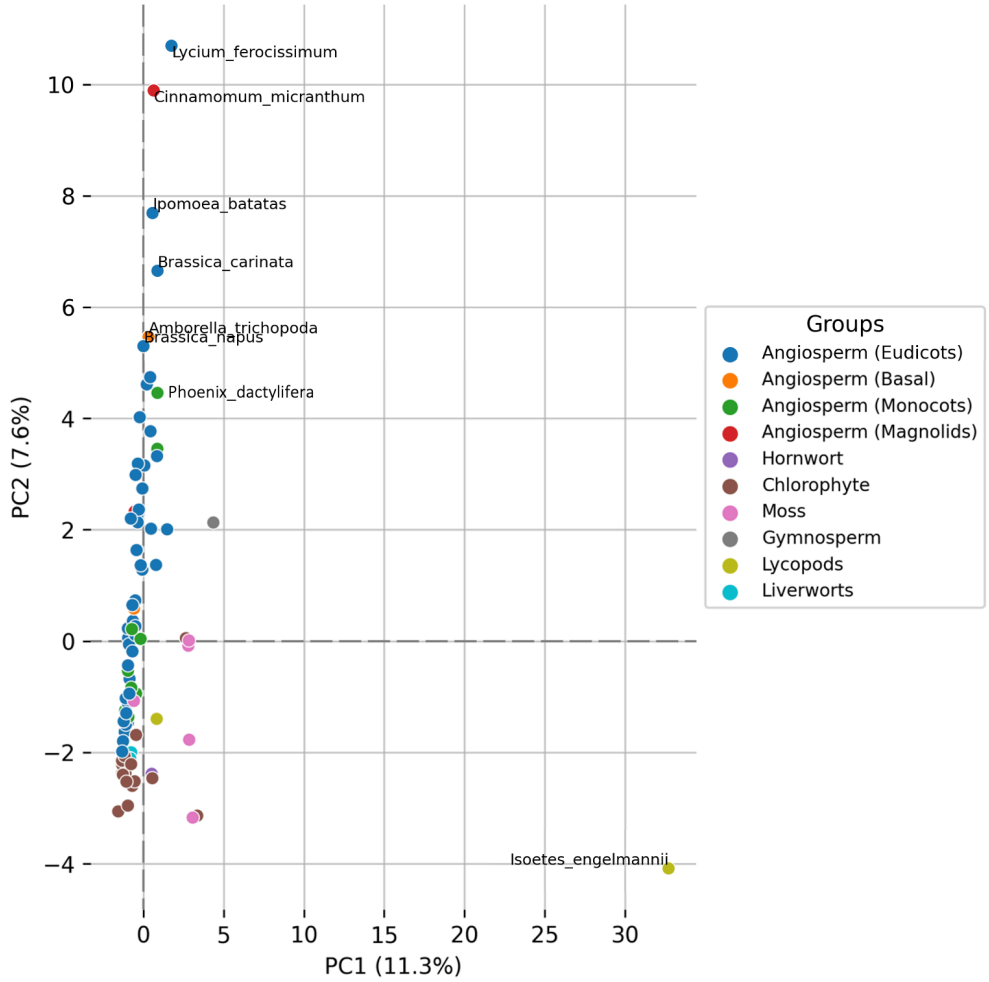
Gymnosperms: Within this lineage, Ngaro, Copia, and L1 display elevated abundance levels. LI-Tx1, in particular, showcases a dominant prevalence. The distinct transposon abundance profile in Gymnosperms, differing from patterns observed in Angiosperms, underscores unique lineage-specific dynamics, possibly reflecting evolutionary pressures or adaptive responses specific to Gymnosperms.

Lycopods: R1-LOA, MULE-NOF, Crypton-A, R2, Deceiver, and RTE-X demonstrate exceptionally high abundance. Such pronounced levels might suggest potential outliers or anomalies within this lineage, warranting further investigation into their functional implications and evolutionary significance.

Hornwort: In this lineage, Tcl Mariner and L1-dep manifest notably elevated abundance. This observation hints at potential lineage-specific amplification events, highlighting the dynamic nature of transposon activity within Hornworts.

Chlorophyte: Relative to other lineages, Chlorophyte showcases reduced transposon element abundance. Nonetheless, even within this lineage, certain elements manifest significant prevalence, which might play roles in genomic stability or adaptive mechanisms specific to Chlorophyte species.

2.3 Principal Component Analysis



The PCA plot offers a multidimensional perspective on the distribution of transposon elements across diverse plant lineages. The plot is structured with the horizontal axis (PC1, representing 11.3% variance) and the vertical axis (PC2, representing 7.6% variance).

Distribution Patterns across Lineages:

Angiosperms: A majority of Angiosperms, particularly the Eudicots, are clustered around the central region of the plot, with values spanning from -2 to 6 along PC2. Notably, three Angiosperm (Eudicots) outliers extend to values of 7, 8, and 11 along PC2, hinting at unique genomic or evolutionary pressures. Basal Angiosperms occupy values around 1, 2, and 5 on PC2. The Monocots cluster closely around the central region, ranging from -1 to 5 along PC2. In contrast, a Magnolid species stands out with a distinctive position at the value of 10 along PC2.

Liverworts: Their tight clustering around (0, -2) underscores a consistent transposon profile, possibly reflecting evolutionary constraints or shared genomic characteristics within this lineage.

Moss: Moss species' spread between (2.5, -3) and (2.5, 1) suggests varied transposon dynamics. The positioning may reflect lineage-specific transposon amplification or suppression events, warranting further investigation.

Gymnosperms: Their positioning around (5, 2) indicates a unique transposon landscape distinct from Angiosperms. This might signify lineage-specific transposon evolution or genomic stabilization mechanisms in Gymnosperms.

Lycopods: The dual positioning of Lycopods is particularly intriguing. While one conforms to the expected range near (1, -1), the outlier at (33, -4) is of notable interest. This anomaly aligns with earlier heatmap observations, suggesting specific transposon elements or events driving this divergence.

Hornwort: Hornworts' predominant presence near (0, -3) suggests a shared transposon profile or evolutionary history. This clustering indicates potential conserved transposon dynamics or genomic stability within Hornworts.

Chlorophyte: Chlorophyte species, primarily clustering around (0, -3) to (0, -2), display a consistent transposon landscape. This uniformity may indicate genomic stability or shared evolutionary constraints across Chlorophyte species.

2.4 Estimating Diversity and Specificity

In the study, we embarked on an extensive re-annotation of the TE landscapes, going beyond mere identification to delve into the estimation of diversity parameters and specificity. By computing Shannon’s Entropy (Hj), we were able to accurately gauge variations and intricacies within the TE landscapes. Furthermore, the assessment of specificity, determined through the specialization index (δj index) and the divergence relative to the entire TE landscape using Kullback–Leibler divergence (Divj), provided profound insights into the behavior of these TEs.

The following two tables list the diversity and specificity across diverse plant species and of distinct transposable elements (TEs).

SPECIES	LINEAGE	DIVERSITY	SPECIFICITY
Acer_negundo	Angiosperm (Eudicots)	1.95300504	4.953885556
Actinidia_chinensis	Angiosperm (Eudicots)	1.578431023	5.328459572
Adansonia_digitata	Angiosperm (Eudicots)	1.456085704	5.450804892
Amborella_trichopoda	Angiosperm (Basal)	1.898810284	5.008080312
Ananas_comosus	Angiosperm (Monocots)	1.821733454	5.085157141
Annona_glabra	Angiosperm (Magnolids)	1.469511863	5.437378733
Anthoceros_angustus	Hornwort	2.000953275	4.90593732
Arabidopsis_arenosa	Angiosperm (Eudicots)	2.185151348	4.721739248
Arabidopsis_suecica	Angiosperm (Eudicots)	2.290236983	4.616653613
Arabidopsis_thaliana	Angiosperm (Eudicots)	2.31642118	4.590469416
Arabis_alpina	Angiosperm (Eudicots)	2.344144692	4.562745903
Astrephomene_gubernaculifera	Chlorophyte	1.955176085	4.951714511
Auxenochlorella_protothecoides_0710	Chlorophyte	2.005377781	4.901512815
Auxenochlorella_protothecoides_UTEX25	Chlorophyte	1.026717167	5.880173428
Boechera_stricta	Angiosperm (Eudicots)	2.510864562	4.396026034
Brachypodium_distachyon	Angiosperm (Monocots)	1.964108106	4.942782489
Brassica_carinata	Angiosperm (Eudicots)	2.41847129	4.488419305
Brassica_napus	Angiosperm (Eudicots)	2.352813541	4.554077055
Brassica_nigra	Angiosperm (Eudicots)	2.412288217	4.494602378
Brassica_oleracea	Angiosperm (Eudicots)	2.105145875	4.801744721
Brassica_rapa	Angiosperm (Eudicots)	2.185702726	4.72118787
Camelina_sativa	Angiosperm (Eudicots)	2.395208726	4.51168187
Carica_papaya	Angiosperm (Eudicots)	2.24989292	4.656997676
Ceratodon_purpureus	Moss	1.95870108	4.948189516
Chlamydomonas_eustigma	Chlorophyte	1.947888622	4.959001973
Chlamydomonas_incerta	Chlorophyte	1.818122544	5.088768052
Chlamydomonas_schloesseri	Chlorophyte	1.922360242	4.984530353
Chlorella_desiccata_UTEX2437	Chlorophyte	1.603679022	5.303211573
Chlorella_desiccata_UTEX2526	Chlorophyte	1.65690163	5.249988965
Chlorella_ohadii	Chlorophyte	0.656640939	6.250249657
Chlorella_sorokiniana	Chlorophyte	0.72825974	6.178630855
Chlorella_variabilis	Chlorophyte	0.902050473	6.004840123
Cinnamomum_micranthum	Angiosperm (Magnolids)	2.022123081	4.884767515
Citrullus_lanatus	Angiosperm (Eudicots)	1.888479988	5.018410608
Citrus_sinensis	Angiosperm (Eudicots)	2.218433055	4.688457541
Coccomyxa_sp.Obi	Chlorophyte	1.912649267	4.994241328
Coccomyxa_subellipsoidea_C-169	Chlorophyte	1.722229998	5.184660597
Cucumis_sativus	Angiosperm (Eudicots)	2.06838225	4.838508346
Cydonia_oblonga	Angiosperm (Eudicots)	1.793793199	5.113097397
Cynara_cardunculus	Angiosperm (Eudicots)	1.815877712	5.091012884
Cyperus_esculentus	Angiosperm (Monocots)	1.784166087	5.122724508
Daucus_carota	Angiosperm (Eudicots)	1.869897478	5.036993117
Dunaliella_salina	Chlorophyte	1.911759949	4.995130646
Edaphochlamys_debaryana	Chlorophyte	1.682697586	5.224193009
Fragaria_ananassa	Angiosperm (Eudicots)	2.396073603	4.510816993
Fragaria_vesca	Angiosperm (Eudicots)	1.998911313	4.907979282
Ginkgo_biloba	Gymnosperm	1.569356562	5.337534034
Glycine_max	Angiosperm (Eudicots)	2.335848931	4.571041665
Glycine_soja	Angiosperm (Eudicots)	2.315511953	4.591378642
Gossypium_raimondii	Angiosperm (Eudicots)	1.926727713	4.980162883
Haematococcus_lacustris	Chlorophyte	2.066997059	4.839893536
Helicosporidium_sp.ATCC_50920	Chlorophyte	1.31801285	5.588877746
Ipomoea_batatas	Angiosperm (Eudicots)	1.736471326	5.170419269
Isoetes_engelmannii	Lycopods	1.883898198	5.022992398
Isothecium_myosuroides	Moss	1.282750808	5.624139788
Lemna_minuta	Angiosperm (Monocots)	2.14271396	4.764176636
Lycium_ferocissimum	Angiosperm (Eudicots)	2.050206115	4.856684481
Malus_domestica	Angiosperm (Eudicots)	2.13696692	4.769923676
Mangifera_indica	Angiosperm (Eudicots)	1.885389006	5.021501589
Marchantia_paleacea	Liverworts	2.263303723	4.643586872
Marchantia_polymorpha	Liverworts	2.31653477	4.590355826
Medicago_truncatula	Angiosperm (Eudicots)	1.951356104	4.955534491
Micractinium_conductrix	Chlorophyte	1.299377422	5.607513174
Micromonas_commoda	Chlorophyte	0.822747326	6.084143269
Micromonas_pusilla_CCMP1545	Chlorophyte	0.934903136	5.97198746
Monoraphidium_minutum	Chlorophyte	0.941199587	5.965691008
Monoraphidium_neglectum	Chlorophyte	0.915470907	5.991419689
Musa_acuminata	Angiosperm (Monocots)	2.238802288	4.668088307
Nelumbo_nucifera	Angiosperm (Eudicots)	1.915256507	4.991634088
Nymphaea_colorata	Angiosperm (Basal)	2.00840205	4.898488545
Oryza_sativa	Angiosperm (Monocots)	1.825972678	5.080917918
Ostreobium_quekettii	Chlorophyte	1.775906381	5.130984214
Ostreococcus_tauri_RCC1115	Chlorophyte	0.960310401	5.946580194
Ostreococcus_tauri_RCC4221	Chlorophyte	0.707665831	6.199224764
Pedinophyceae_sp.YPF-701	Chlorophyte	1.20442747	5.702463125
Phalaenopsis_equestris	Angiosperm (Monocots)	1.951751363	4.955139233
Phaseolus_vulgaris	Angiosperm (Eudicots)	2.27950921	4.627381386
Phoenix_dactylifera	Angiosperm (Monocots)	2.613009928	4.293880668
Physcomitrium_patens	Moss	2.225195377	4.681695218
Picochlorum_sp.BPE23	Chlorophyte	1.932100167	4.974790429
Pistacia_vera	Angiosperm (Eudicots)	2.280412081	4.626478514
Populus_trichocarpa	Angiosperm (Eudicots)	1.840867301	5.066023294
Prunus_persica	Angiosperm (Eudicots)	2.142378369	4.764512227
Punica_granatum	Angiosperm (Eudicots)	1.847896698	5.058993898
Pycnococcus_provasolii	Chlorophyte	0.868946597	6.037943999
Salvia_hispanica	Angiosperm (Eudicots)	1.839807771	5.067082825
Salvia_splendens	Angiosperm (Eudicots)	2.174679901	4.732210694
Scenedesmus_sp.NREL46B-D3	Chlorophyte	1.892019991	5.014870605
Selaginella_moellendorffii	Lycopods	2.566286717	4.340603879
Setaria_italica	Angiosperm (Monocots)	2.146196027	4.760694568
Solanum_lycopersicum	Angiosperm (Eudicots)	1.957744071	4.949146525
Solanum_pimpinellifolium	Angiosperm (Eudicots)	1.985533403	4.921357192
Solanum_stenotomum	Angiosperm (Eudicots)	2.097490761	4.809399835
Solanum_tuberosum	Angiosperm (Eudicots)	1.93638664	4.970503955
Sphagnum_fallax	Moss	1.596805493	5.310085103
Sphagnum_magellanicum	Moss	1.69131012	5.215580476
Tetrabaena_socialis	Chlorophyte	1.527779031	5.379111565
Trebouxia_sp.A1-2	Chlorophyte	1.867433503	5.039457093
Utricularia_gibba	Angiosperm (Eudicots)	2.002433451	4.904457144
Vitis_vinifera	Angiosperm (Eudicots)	2.302595798	4.604294797
Volvox_africanus	Chlorophyte	1.970365685	4.93652491
Volvox_reticuliferus_NIES-3785	Chlorophyte	2.275028021	4.631862575
Volvox_reticuliferus_NIES-3786	Chlorophyte	2.276423256	4.630467339

SUPERFAMILY	DIVERSITY	SPECIFICITY
Unknown	5.878525531	0.807974996
Simple_repeat	6.342301727	0.344198801
Copia	5.639299016	1.047201511
Gypsy	5.351115645	1.335384882
Low_complexity	6.226699624	0.459800903
L1	5.626460921	1.060039606
MULE-MuDR	5.44278592	1.243714607
Helitron	5.686252069	1.000248458
hAT-Ac	5.502143944	1.184356584
hAT-Tip100	5.457267804	1.229232723
PIF-Harbinger	4.824638662	1.861861866
hAT-Tag1	5.01617355	1.670326977
Caulimovirus	4.714121649	1.972378878
TcMar-Pogo	3.810336469	2.876164059
tRNA	5.471053808	1.215446719
CMC-EnSpm	5.367605857	1.31889467
LTR	4.411005586	2.275494941
TcMar-Tc1	2.800245201	3.886255326
hAT	4.522126108	2.164374419
L1-Tx1	2.605869976	4.080630552
L2	2.407940184	4.278560343
I-Jockey	2.405228369	4.281272158
rRNA	5.199422238	1.48707829
CR1	3.18365309	3.502847437
Zisupton	3.821675626	2.864824901
Merlin	3.153830191	3.532670336
DNA	3.934914043	2.751586484
ERVL	2.667848288	4.018652239
mixture	3.757429839	2.929070688
U	3.162793839	3.523706689
Pao	4.155391924	2.531108603
Ngaro	3.36652787	3.319972657
Sola2	0.945924738	5.740575789
Bel-Pao	4.075932179	2.610568348
Tc1_Mariner	3.069532083	3.616968444
RTE-BovB	4.296021338	2.390479189
ERV1	4.814212649	1.872287878
IS3EU	2.940138206	3.746362321
snRNA	3.984285356	2.702215172
5S	4.16137369	2.525126837
hAT-Charlie	3.20281877	3.483681758
I	3.446455007	3.24004552
Dada	2.923141742	3.763358785
ERVK	3.767971627	2.9185289
ID	3.263458351	3.423042176
Kolobok	0	6.686500527
Sola1	3.252358829	3.434141699
Academ-1	0.437495621	6.249004907
Satellite	4.561175626	2.125324901
Retrovirus	2.529279514	4.157221013
RTE	2.319211242	4.367289285
P	1.581090931	5.105409596
Chlamys	3.225716474	3.460784053
U-L1	4.212121378	2.474379149
SINE	3.970938539	2.715561989
Novosib	1.726083719	4.960416808
R1	3.204885645	3.481614882
LINE	1.817115643	4.869384884
CRE	2.016141873	4.670358654
Sola-1	2.689972153	3.996528374
RTE-X	3.163901845	3.522598683
Maverick	3.096356162	3.590144365
Sola-2	2.61439391	4.072106617
TcMar-Tc2	2.82348925	3.863011277
EnSpm_CACTA	2.69662165	3.989878877
Tad1	1.521723307	5.16477722
Ginger-1	0	6.686500527
PiggyBac	2.478836014	4.207664513
tRNA-L1	0	6.686500527
Helitron-2	2.119209541	4.567290986
Dualen	1.353095546	5.333404982
DIRS	3.234598809	3.451901718
hAT-hobo	1.385845432	5.300655095
TcMar-Stowaway	3.178752045	3.507748482
CR1-Zenon	0.999818247	5.68668228
Rex-Babar	0.376416213	6.310084314
hAT-hATm	0.354252442	6.332248085
Kolobok-H	2.455747989	4.230752538
tRNA-CR1	0	6.686500527
Penelope	3.420667312	3.265833215
TcMar-Fot1	2.451851554	4.234648973
tRNA-RTE	1.980615187	4.70588534
Kolobok-T2	1.291072185	5.395428342
B2	0.960518707	5.72598182
TATE	0	6.686500527
R2-Hero	0.677380947	6.009119581
TcMar-Mariner	1.915727164	4.770773363
Ginger-2	0.881290899	5.805209628
MULE-NOF	0.794102142	5.892398385
Crypton-F	1.02223936	5.664261167
L1-dep	0.897530878	5.788969649
ERV-Foamy	0	6.686500527
R1-LOA	1.850187428	4.836313099
PIF-ISL2EU	0	6.686500527
ERVL-MaLR	0.554534307	6.13196622
Retroposon	0	6.686500527
TcMar-Tigger	2.765670923	3.920829604
hAT-Blackjack	0.976234827	5.7102657
subtelo	0	6.686500527
Naiad	2.570050807	4.11644972
hAT-hAT19	0	6.686500527
Alu	0	6.686500527
B4	1.740400955	4.946099573
Crypton-A	0.793622461	5.892878066
R2	0	6.686500527
TcMar-Tc4	0	6.686500527
Deceiver	0	6.686500527
P-Fungi	0.266764988	6.419735539
TcMar-Cweed	0.699772222	5.986728305
Hydra	0	6.686500527
hAT-hATx	0	6.686500527
TcMar-ISRm11	0.319498402	6.367002125
CMC-Transib	0	6.686500527
tRNA-Core-RTE	0	6.686500527
TcMar	1.299707302	5.386793225
tRNA-Deu-RTE	1.566474269	5.120026259
ERV4	0	6.686500527
TcMar-Ant1	0	6.686500527
Crypton-H	0	6.686500527
ARTEFACT	0	6.686500527

The first table presents an analysis of diversity and specificity across over a hundred different plant species. Notably, the highest diversity is observed in *Phoenix dactylifera* (Angiosperm (Monocots)) with a value of 2.61. This species also exhibits the lowest specificity at 4.29. Conversely, *Chlorella ohadii* (Chlorophyte) demonstrates the highest specificity of 6.25 but has the lowest diversity at 0.66.

A closer examination reveals a pattern within various plant lineages. Several species within the Angiosperm lineage, such as *Arabidopsis suecica* (diversity: 2.29, specificity: 4.62) and *Brassica carinata* (diversity: 2.41, specificity: 4.49), show relatively high diversity with moderately low specificity. In contrast, species like *Anthoceros angustus* from Hornworts (diversity: 2.00, specificity: 4.91) and *Marchantia polymorpha* from Liverworts (diversity: 2.32, specificity: 4.59) present higher specificity with decent diversity. Additionally, within the Lycopods, *Selaginella moellendorffii* demonstrates a diversity of 2.57 and a specificity of 4.34.

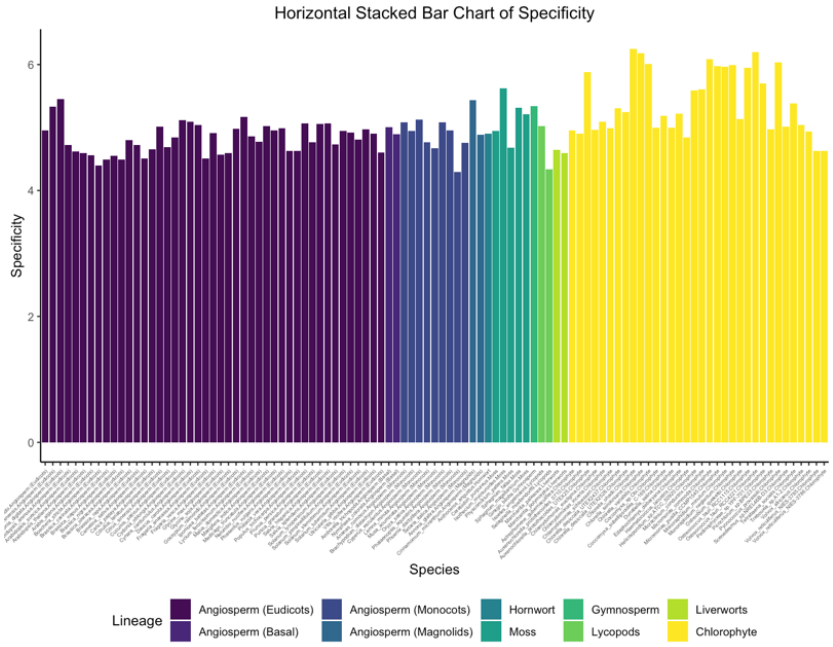
However, certain lineages display lower diversity and higher specificity. For instance, *Sphagnum fallax* from Moss has a diversity of 1.60 and a specificity of 5.31. Similarly, *Ginkgo biloba* from Gymnosperms shows a diversity of 1.60 and a specificity of 5.34. A unique observation is made with *Ostreococcus tauri RCC4221* (Chlorophyte) having a diversity of 0.71 but a remarkably high specificity of 6.20. Notably, *Volvox reticuliferus NIES-3785* from Chlorophyte exhibits a diversity of 2.28 and a specificity of 5.31.

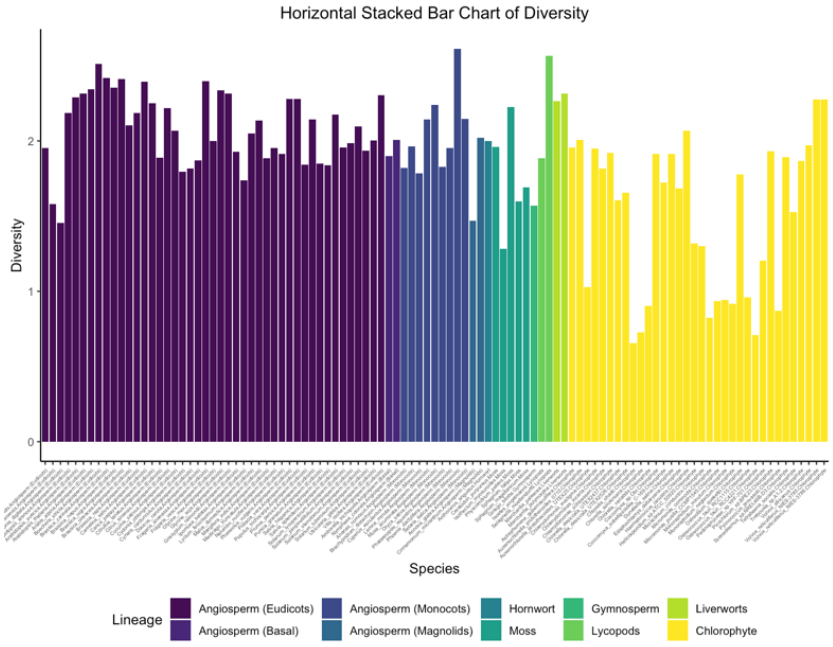
The second table offers insights into the diversity and specificity of 120 different TE superfamily elements. A striking observation is the presence of 20 TEs with a diversity value of 0, including Kolobok, Ginger-1, tRNA-L1, tRNA-CR1, TATE, ERV-Foamy, PIF-ISL2EU, subtelo, hAT-hAT19, Alu, R2, TcMar-Tc4, Deceiver, Hydra, hAT-hATx, CMC-Transib, tRNA-Core-RTEERV4, TcMar-Ant1, Crypton-H, ARTEFACT. These TEs are predominantly found in specific plant species, leading to their elevated specificity values, peaking at 6.69.

In contrast, the TE with the highest diversity is Simple repeats at 6.34, being ubiquitously present across all analyzed plant species. This widespread distribution translates to its notably low specificity of 0.34. Other TEs like Unknown (diversity: 5.88, specificity: 0.81), Copia (diversity: 5.64, specificity: 1.05), Gypsy (diversity: 5.35, specificity: 1.34), and Low complexity (diversity: 6.23, specificity: 0.46) also exhibit high diversity but with reduced specificity values, suggesting their prevalent presence across various plant genomes.

Visualizing Diversity and Specificity in Plant Species and TE Superfamilies

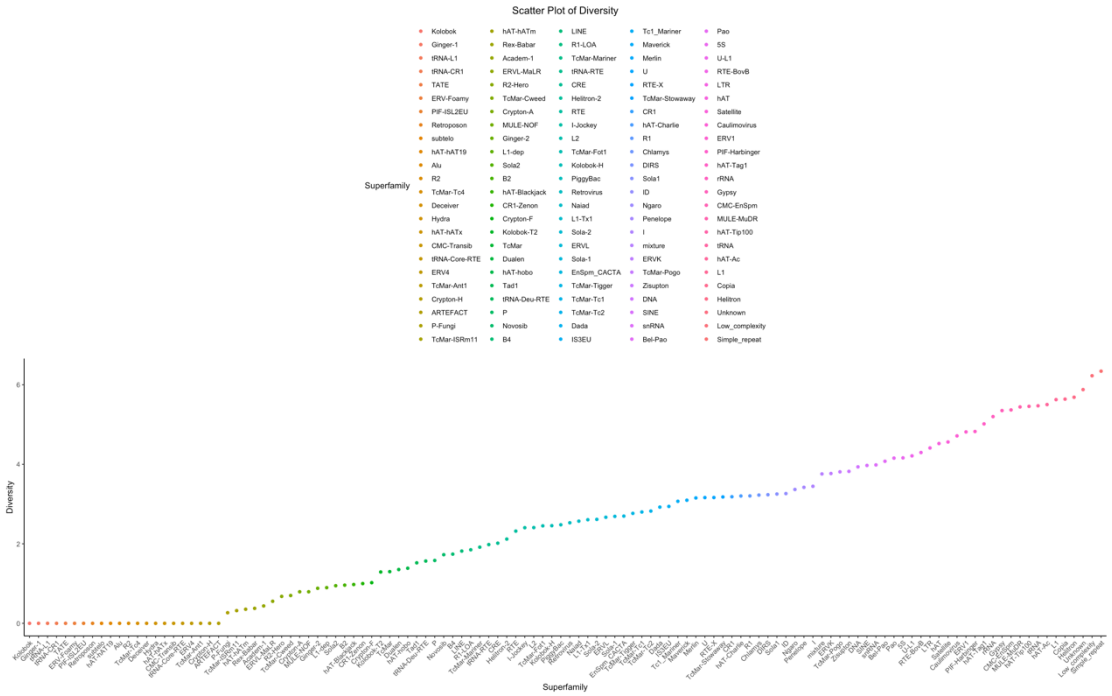
To provide a more intuitive understanding, horizontal stacked bar charts were utilized to visualize the diversity and specificity of plant species.

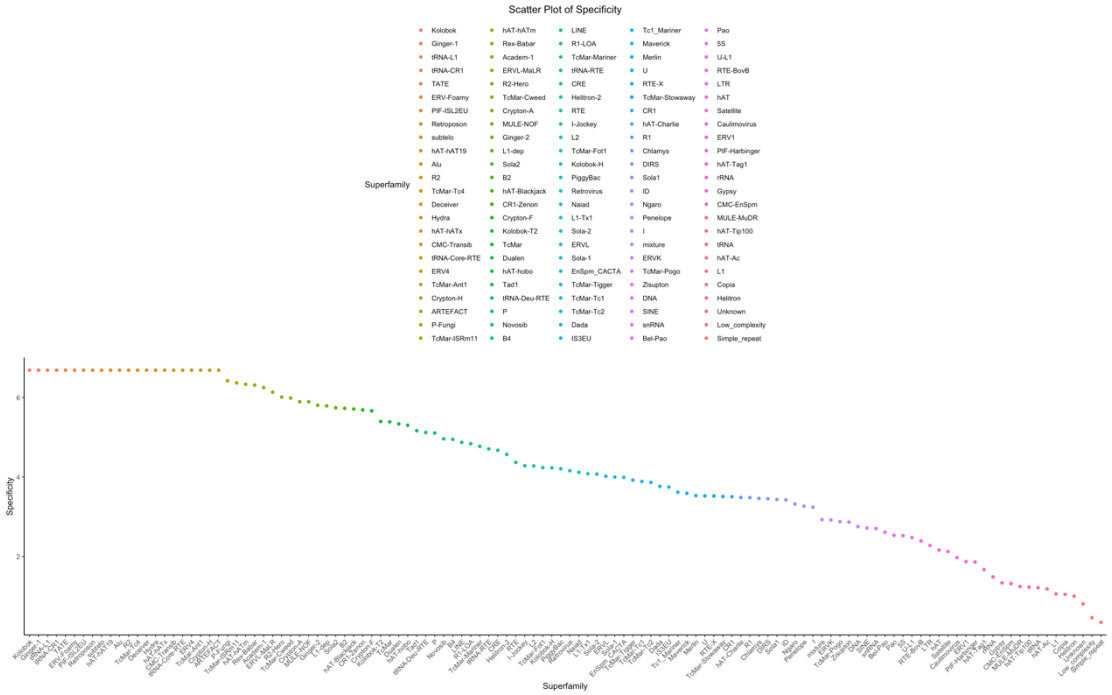




The horizontal stacked bar charts reveal distinct patterns among different plant lineages. Notably, Angiosperms exhibit significantly higher overall diversity compared to Chlorophytes. Additionally, Liverworts, Lycopods, and Hornworts tend to showcase relatively high diversity with lower specificity. In contrast, Gymnosperms demonstrate moderate levels of both diversity and specificity. Moss, on the other hand, exhibits lower diversity but higher specificity. These observations collectively highlight the nuanced diversity and specificity profiles across different plant lineages.

Subsequently, scatter plots were generated to visually represent the diversity and specificity patterns across various TE superfamilies.





As evident from the depicted figures, several TE superfamilies, including Kolobok, Ginger-1, tRNA-L1, tRNA-CR1, TATE, ERV-Foamy, PIF-ISL2EU, subtelo, hAT-hAT19, Alu, R2, TcMar-Tc4, Deceiver, Hydra, hAT-hATx, CMC-Transib, tRNA-Core-RTEERV4, TcMar-Ant1, Crypton-H, and ARTEFACT, exhibit a diversity value of 0, concurrently demonstrating the highest specificity.

On the contrary, simple repeats, low complexity, unknown, helitron, copia, L1, and others showcase remarkably high diversity, coupled with notably low specificity, implying a widespread and less constrained presence across the genome.

Discussion

1. The Unique TE Landscape Across Divergent Lineages

The intricate interplay between transposable elements and their host genomes has long been a subject of interest in evolutionary biology. The analysis combining both heatmap and PCA visualizations, provides a nuanced understanding of TE landscape variations across seven distinct plant lineages.

The heatmap delineates clear lineage-specific abundance distributions for various transposon elements. For instance, the pervasive abundance of Copia, L1, Helitron, CMC-EnSpm, and hAT-Tip100 in the majority of Angiosperms underscores the potential for active transposition events within this lineage. The last table further highlights the notably high diversity values associated with these elements. Such patterns may arise from a combination of environmental pressures and inherent genomic factors, leading to genomic plasticity and adaptability in Angiosperms.

As the heatmap indicates an increase in the content of specific transposable elements, there is a potential corresponding tendency for elevated values along the PC2 among various angiosperm species. Notably, *Lycium ferocissimum* exemplifies this trend, displaying the highest PC2 value, which aligns with its substantial content of several transposon elements, including L1 and Copia. Likewise, in the case of *Ipomoea batatas*, heightened levels of Merlin, I, and Alu elements are apparent, correlating with a noteworthy PC2 value of approximately 8. Moreover, within monocots, *Phoenix dactylifera* distinguishes itself with the highest PC2 value, showcasing a significant abundance of Penelope and ERV1 transposable elements. This observation suggests a plausible association between the increase in specific TE content in the heatmap and the resulting variations in PC2 values across diverse angiosperm taxa.

Interestingly, both Moss and Liverworts exhibit notable lineage-specific amplification dynamics. Elements like PiggyBac, TcMar-Stowaway, and Sola2 in Moss, and Tad1 and hAT-hobo in Liverworts, respectively, highlight the dynamic nature of TE evolution within these lineages. Such dynamics might be pivotal in shaping the genomic architecture and influencing adaptive responses to environmental cues.

The distinct clustering of Gymnosperms around specific PCA coordinates emphasizes their unique TE landscape, diverging from the patterns observed in Angiosperms. Such divergence might be indicative of lineage-specific evolutionary pressures or adaptive responses, underscoring the need for detailed genomic studies in Gymnosperms.

A particularly noteworthy observation in the analysis pertains to the Lycopods lineage. In the PCA plot, an outlier was identified corresponding to the Lycopods species, *Isoetes engelmannii*, with coordinates (33, -4). Upon examination, this species

exhibited a Diversity value of 1.88 and a Specificity value of 5.02, potentially reflecting specialized evolutionary dynamics or distinct genomic features within this species. The heatmap results were subsequently checked, revealing that *Isoetes engelmannii* exhibited a marked elevation in the abundances of several transposon elements, including R1-LOA, MULE-NOF, Crypton-A, R2, Deceiver, and RTE-X. Examination of the data reveals that the Diversity values for R2 and Deceiver are 0, indicative of exceptionally high specificity, implying that these transposon elements exist exclusively in *Isoetes engelmannii*. Given the significant abundance of these transposons in *Isoetes engelmannii*, it is plausible to infer that the pronounced anomaly in the PCA plot might be attributed to these exceptionally high transposon levels. The presence and abundance of these elements in *Isoetes engelmannii* warrant further investigation to elucidate their potential functional implications and evolutionary significance within the Lycopods lineage.

Furthermore, Chlorophytes showcase unique features. In the PCA plot, Chlorophytes primarily cluster around -3 to -2 along PC2. Meanwhile, in the heatmap, Chlorophytes exhibit significantly diminished TE content.

The findings align with the hypothesis that the most divergent lineages indeed harbor the most unique TE landscapes. The variations in TE abundance, distribution, and dynamics across different plant lineages reflect their respective evolutionary trajectories, adaptive potentials, and genomic architectures.

Contextualizing with Existing Literature:

Comparing our results with the findings from the study titled "PlantLTRdb: An interactive database for 195 plant species LTR-retrotransposons" (Mokhtar et al. 2023) underscores the breadth and depth of TE research in plants. The PlantLTRdb database, with its comprehensive coverage of 195 plant species, has highlighted that LTR-RT activity stands as a pivotal factor driving genome evolution. Specifically, the activation and proliferation of LTR-RTs can lead to significant expansions in genome size, with certain plant genomes having over 70% of their content attributed to these elements. Furthermore, the aforementioned study delineates that there exists a strong positive correlation between the total length of LTR-RTs and genome size across these plant species.

A critical point of intersection between our study and the findings from PlantLTRdb lies in the functional impacts of LTR-RTs on genes. The study emphasizes that LTR-RTs, when situated near or within genes, have the potential to directly modify gene function. Additionally, the research conducted a detailed statistical analysis of LTR-RT lengths and their respective superfamily, such as Copia and Gypsy elements. This information provides a more nuanced understanding of the structural variations and complexities within LTR-RTs across plant genomes.

Another significant dimension highlighted by the study is the insertion age of the plant species studied, which reflects the evolutionary rate associated with the uniqueness of their genomic content. This temporal aspect adds a layer of complexity to our understanding of TE dynamics, suggesting that the genomic impacts of LTR-RTs may vary depending on the evolutionary histories of individual plant lineages.

Moreover, the research states that environmental stresses can serve as triggers for TE activation. This perspective resonates with our discussion on the potential adaptive responses of plants to their environments through TE dynamics. It's conceivable that certain TE activations within specific plant lineages could be a direct response to environmental pressures, thereby contributing to the observed genomic plasticity and adaptability in these lineages.

While PlantLTRdb offers a holistic, database-driven perspective on TE dynamics, our research provides a granular analysis of lineage-specific dynamics. Together, these insights paint a comprehensive picture of TE evolution across plant taxa, emphasizing the intricate roles TEs play in plant evolution, genomic diversification, and environmental adaptation.

In conclusion, the synthesis of heatmap, PCA analyses, and insights from PlantLTRdb enhances our understanding of TE dynamics in various plant lineages. Mokhtar et al.'s study enriches this discourse, emphasizing the multifaceted impacts of LTR-RTs on genome evolution, gene function, environmental adaptation, and evolutionary rates reflected in insertion ages. This underscores the necessity for a comprehensive approach in unraveling complex interactions between TEs and plant genomes.

2. Genome Size and Its Influence on TE Landscape

The hypothesis 2 posited a correlation between genome size and the TE landscape across plant species, suggesting that larger genomes may provide more "space" for TE insertions, influencing gene density and spatial organization, ultimately affecting TE accumulation and distribution patterns.

The scatter plot reinforces this hypothesis, revealing distinct patterns across various plant lineages. Gymnosperms, characterized by extensive genomes, particularly exemplified by *Cycas panzhihuaensis*, display the largest Genome Size and the highest content of Bases Masked, exceeding 80%. Conversely, Chlorophytes, with smaller genomes, exemplified by *Helicosporidium sp. ATCC 50920*, exhibits the smallest Genome Size and the lowest Bases Masked content.

Despite lineage variations, the scatter plot consistently illustrates a positive correlation between Genome Size and Bases Masked. The grey reference line serves as a visual guide, emphasizing the clustering of data points around the line. This consistent

relationship highlights that, irrespective of lineage, larger genomes tend to correlate with higher TE content, aligning with our hypothesis.

In conclusion, the scatter plot effectively supports the hypothesis, providing visual evidence for the positive correlation between Genome Size and Bases Masked content across diverse plant lineages.

Given these observations, the comprehensive study by Pedro DLF et al. provides an essential context, emphasizing an association between genome size and TE content across diverse plant lineages. While their exhaustive analysis across 67 plant genomes reveals a general trend of larger genomes having higher occurrences of TEs, our study contributes by examining a broader range of samples. This expanded dataset further underscores the complexity and variability within plant genomes. The insights from Pedro DLF et al. regarding the pervasive nature of TE landscapes, combined with our extensive sampling, collectively emphasize the intricate interplay between genome size and TE content, highlighting its multifaceted role in shaping genomic evolution across plant species.

3. TE Activity and Variability in Rapidly Evolving Lineages

The hypothesis put forth in this research suggests that plant lineages undergoing rapid evolution, particularly those with elevated %Divergence in LTR elements, might manifest increased TE activity or mutation rates. Such a hypothesis is rooted in the premise that heightened evolutionary dynamics, as indicated by increased %Divergence in LTR elements, could be accompanied by escalated TE mobility or activity. This heightened activity might serve as a genomic response mechanism, aiding these rapidly evolving lineages in adapting to a myriad of environmental challenges.

Upon delving into the data, certain findings align with this hypothesis.

Starting with the LTR elements, a clear distinction between Chlorophytes and other plant lineages emerged. Chlorophyte species exhibited a broader range of LTR element %Divergence, predominantly clustering below 15%. This pattern suggests a distinct evolutionary trajectory within this ancestral group. Conversely, advanced plant lineages, including Angiosperms, Hornwort, Moss, Gymnosperms, Lycopods, and Liverworts, consistently demonstrated higher median %Divergence values, hovering around 15-20%. Notably, *Ginkgo biloba*, a Gymnosperm, stood out with a median value reaching 25%, underscoring elevated TE activity or mutation rates within this advanced lineage.

Similar patterns were observed for Copia and Gypsy elements. While most advanced plant lineages displayed median values between 15-20%, Chlorophytes exhibited greater variability, with median values predominantly below 15%.

Analysis of LINE and SINE elements reinforced these patterns. Advanced plant lineages consistently exhibited higher median %Divergence values, particularly within Angiosperms. In contrast, Chlorophyte species demonstrated lower median values, reinforcing the distinct evolutionary dynamics associated with these ancestral lineages. These findings align with our hypothesis, suggesting that rapidly evolving plant lineages may indeed exhibit increased TE diversity and variability.

The Unclassified Elements present an interesting pattern across various plant lineages. Angiosperms, Hornwort, Moss, Lycopods, and Liverworts consistently manifest median %Divergence values predominantly between 15 and 20. Notably, Gymnosperns deviate slightly with a median %Divergence peaking at 23. In stark contrast, the Chlorophyte lineage stands apart, reflecting a more subdued median %Divergence that spans between 5 and 15. These observations further emphasize the heightened TE activity or mutation rates in advanced plant lineages compared to their more primitive counterparts.

The analysis of Simple Repeat Elements presented a contrasting scenario. Advanced plant lineages, including Angiosperms, Hornwort, Moss, Lycopods, and Liverworts, exhibited a stable pattern with median %Divergence values ranging from 5 to 15%. In contrast, Chlorophyte species displayed higher variability, with median values fluctuating between 10 to 20%. This divergence in patterns suggests that while some TE elements may be influenced by evolutionary rates, others might be governed by different evolutionary constraints or mechanisms.

Lastly, the Low Complexity Elements showcased consistent median values across all lineages, hovering around 20%. This uniformity suggests that whether situated at centromeric regions or telomeres, the properties of these components remain nearly identical. Such consistency may arise from the similar functional or structural roles that LCEs play in plant evolution.

In summation, the findings of this study shed light on the nuanced dynamics of TEs across plant lineages. Advanced, rapidly evolving lineages, like Angiosperms and Gymnosperms, seem to harbor increased TE diversity, hinting at potential adaptive evolutionary strategies. Conversely, ancestral lineages maintain a more stable TE landscape, underscoring the delicate balance between TE activity and genomic stability. These insights further underscore the multifaceted roles of TEs in plant genome evolution and the adaptive strategies employed by different lineages.

Conclusions

The exploration of transposable elements (TEs) across a spectrum of plant lineages, from Chlorophytes to Angiosperms, offers profound insights into the evolutionary dynamics and functional implications of these genomic entities. TEs, as dynamic components, play pivotal roles in sculpting the genomic architecture, influencing gene regulation, and potentially driving evolutionary innovations.

Lineage-Specific TE Dynamics: Our study underscores the intricate and unique TE landscapes inherent to distinct plant lineages. Advanced lineages, notably Angiosperms and Gymnosperms, exhibit pronounced TE variability, potentially reflecting adaptive evolutionary strategies. In contrast, ancestral lineages like Chlorophytes maintain a more conserved TE profile, highlighting the interplay between evolutionary stability and TE activity.

Genome Size and TE Interplay: The scatter plot unequivocally confirm a positive correlation between genome size and TE content across diverse plant lineages. This correlation underscores the important role of genomic expansiveness in facilitating TE insertions.

Rapidly Evolving Lineages and TE Activity: Our findings substantiate the hypothesis that rapidly evolving lineages may harbor increased TE activity or mutation rates. Elevated %Divergence values in advanced lineages, coupled with distinct TE profiles, suggest an evolutionary response mechanism, enabling these lineages to navigate diverse environmental challenges.

Functional and Evolutionary Implications: Beyond mere genomic components, TEs hold functional significance, potentially influencing specific genomic features or evolutionary traits. Their pervasive presence across diverse lineages underscores their indispensable role in shaping plant genome architecture and function.

In essence, this research illuminates the multifaceted roles of TEs in plant evolution, genome dynamics, and adaptation. As genomic studies continue to unravel the complexities of plant genomes, TEs remain central players, necessitating continued exploration to decipher their full spectrum of functions and evolutionary impacts.

References

Bourque, G., Burns, K.H., Gehring, M. *et al.* Ten things you should know about transposable elements. *Genome Biol* **19**, 199 (2018). https://doi.org/10.1186/s13059-018-1577-z

Ramakrishnan, M., Satish, L., Sharma, A. *et al.* Transposable elements in plants: Recent advancements, tools and prospects. *Plant Mol Biol Rep* **40**, 628–645 (2022). https://doi.org/10.1007/s11105-022-01342-w

Muñoz-López M, García-Pérez JL (2010) DNA transposons: nature and applications in genomics. Curr Genom 11:115–128

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973–982. https://doi.org/10.1038/nrg2165

Brown PO, Bowerman B, Varmus HE, Bishop JM. Correct integration of retroviral DNA in vitro. Cell. 1987;49:347–56.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 1993;72:595–605.

Malik HS, Eickbush TH. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res. 2001;11:1187–97.

Mhiri, C., Borges, F. and Grandbastien, M.A., 2022. Specificities and dynamics of transposable elements in land plants. Biology, 11(4), p.488. https://doi.org/10.3390/biology11040488

Schubert I, Vu GTH. Genome stability and evolution: attempting a holistic view. Trends Plant Sci. 2016;21:749–57.

Haley, A.L. and Mueller, R.L., 2022. Transposable element diversity remains high in gigantic genomes. Journal of Molecular Evolution, 90 (5), pp.332-341. https://doi.org/10.1007/s00239-022-10063-3

Almeida, M.V., Vernaz, G., Putman, A.L. and Miska, E.A., 2022. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. Trends in Genetics, 38(6), pp.529-553 https://doi.org/10.1016/j.tig.2022.02.009

Alla Mikheenko, Andrey Prjibelski, Vladislav Saveliev, Dmitry Antipov, Alexey Gurevich,
Versatile genome assembly evaluation with QUAST-LG,
Bioinformatics (2018) 34 (13): i142-i150. doi: 10.1093/bioinformatics/bty266

Jacques Dainat, Darío Hereñú, Dr. K. D. Murray, Ed Davis, Kathryn Crouch, LucileSol, Nuno Agostinho, pascal-git, Zachary Zollman, & tayyrov. (2023). NBISweden/AGAT: AGAT-v1.2.0 (v1.2.0). Zenodo. https://doi.org/10.5281/zenodo.8178877

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences, 117(17), 9451-9457. https://doi.org/10.1073/pnas.1921046117

A.F.A. Smit, R. Hubley & P. Green. RepeatMasker at http://repeatmasker.org.

Zhang, R.-G., Li, G.-Y., Wang, X.-L., Dainat, J., Wang, Z.-X., Ou, S., & Ma, Y. (2022). TEsorter: An accurate and fast method to classify LTR-retrotransposons in plant genomes. Horticulture Research, 9, uhac017. https://doi.org/10.1093/hr/uhac017

Mokhtar, M. M., Alsamman, M., & El Allali, A. (2023). PlantLTRdb: An interactive database for 195 plant species LTR-retrotransposons. Frontiers in Plant Science, 14, Sec. Plant Systematics and Evolution. https://doi.org/10.3389/fpls.2023.1134627

Pedro, D. L. F., Amorim, T. S., Varani, A., Guyot, R., Domingues, D. S., & Paschoal, A. R. (2021). An Atlas of Plant Transposable Elements. F1000Research, 10, 1194. https://doi.org/10.12688/f1000research.74524.1

Hadley Wickham, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, and Kara Woo. (2022). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics (Version 3.4.4) [Software].
https://CRAN.R-project.org/package=ggplot2