genes

MDPI

Article

Specific LTR-Retrotransposons Show Copy Number Variations between Wild and Cultivated Sunflowers

Flavia Mascagni, Alberto Vangelisti, Tommaso Giordani, Andrea Cavallini * and Lucia Natali *

Department of Agricultural, Food, and Environmental Sciences, University of Pisa, Via del Borghetto 80, I-56124 Pisa, Italy; flavia.mascagni@unipi.it (F.M.); albertovangelisti@libero.it (A.V.); tommaso.giordani@unipi.it (T.G.)

* Correspondence: andrea.cavallini@unipi.it (A.C.); lucia.natali@unipi.it (L.N.); Tel.: +39-050-221-6663 (A.C.); +39-050-221-6665 (L.N.)

Received: 25 July 2018; Accepted: 24 August 2018; Published: 29 August 2018

Abstract: The relationship between variation of the repetitive component of the genome and domestication in plant species is not fully understood. In previous work, variations in the abundance and proximity to genes of long terminal repeats (LTR)-retrotransposons of sunflower (*Helianthus annuus* L.) were investigated by Illumina DNA sequencingtocompare cultivars and wild accessions. In this study, we annotated and characterized 22 specific retrotransposon families whose abundance varies between domesticated and wild genotypes. These families mostly belonged to the Chromovirus lineage of the Gypsy superfamily and were distributed overall chromosomes. They were also analyzed in respect to their proximity to genes. Genes close to retrotransposon were classified according to biochemical pathways, and differences between domesticated and wild genotypes are shown. These data suggest that structural variations related to retrotransposons might have occurred to produce phenotypic variation between wild and domesticated genotypes, possibly by affecting the expression of genes that lie close to inserted or deleted retrotransposons and belong to specific biochemical pathways as those involved in plant stress responses.

Keywords: *Helianthus annuus*; long terminal repeat retrotransposons; plant domestication; retrotransposon abundance; retrotransposon proximity to genes

1. Introduction

Transposable elements (TEs) are DNA sequences that are able to change their position in the chromosomes. They are classified into two classes depending on whether the transposition intermediate is RNA (Class I transposons or retrotransposons) or DNA (Class II or DNA transposons [1]. Class I elements are found in most eukaryotic lineages. In plants, the most abundant retrotransposon order is that with long terminal repeats (LTRs), two direct repeats containing promoter and RNA processing signals, flanking a region encoding a polyprotein that includes the enzymes necessary for its transposition [1]. Plant LTR-retrotransposons (LTR-REs) are classified into two main superfamilies—Copia and Gypsy [1]—which differ in the order of the enzymes within the polyprotein [2]. LTR-RE length ranges from a few hundred base pairs to over 10 kbp [2]. Copia and Gypsy superfamilies are in turn classified into different major lineages based on sequence similarity [3–5]. However, DNA sequence similarity within a lineage is minimal and limited to coding regions. When sequence similarity within a lineage extends to noncoding portions, two elements may be grouped in a single family [1].

Retrotransposons transpose by producing an RNA intermediate that is then reverse transcribed to DNA and inserted at a new genome site [2]. This transposition mechanism, which uses enzymes produced by the retrotransposon itself, a reverse transcriptase, a ribonuclease (RNAase),

a protease, and an integrase, implies the production of a new copy for each transposition event. Retrotransposon mobility in the genome is usually blocked by different epigenetic mechanisms; however, some retrotransposons under certain environmental conditions are able to escape epigenetic control by the host genome [6]. This escape can result in a huge increase in retrotransposon copy number and consequently an extremely rapid and large increase in genome size when one considers evolutionary timescales [1].

Even so, the retrotransposon component of the eukaryotic genomes is subject to rapid turnover [7,8]. While retrotransposons can increase in number in a relatively short time span, they can also be rapidly removed from the genome through the processes of unequal homologous and illegitimate recombination [9,10].

Retrotransposons proliferation and loss can lead to the creation of haplotypes with different LTR-RE numbers at specific loci [11,12]. Hence, the number of LTR-REs in a genome can also change because of random combination between LTR-RE-rich or poor haplotypes.

Retrotransposon activity produces genetic variation with important effects on the evolution of a species [13]. Transposons may insert in or near a gene, resulting in direct alteration of the coding sequence, transcription regulation modification, or altered splicing patterns [14]. Insertion in the proximity to a gene may also have consequences; since retrotransposons are epigenetically inactivated by the host, integration of an element may actually modify the epigenetic setting of the insertion site. Overall, retrotransposons are known to regulate the epigenetic setting of the genome and chromatin organization and structure. Perhaps more importantly, insertion or removal of a retrotransposon can change the expression rate or regulation of neighboring genes [15–18].

A few studies have investigated the possible role of transposable elements and other repetitive elements of genomes in the domestication of crop plants [19] and have included work on maize [20,21], rice [22], and sunflower [23].

The sunflower (*Helianthus annuus* L., Asteraceae) is one of the most important oilseed crops. The origin of the genus *Helianthus* dates back 4.75–22.7 million years [24]. It is likely the sunflower originated in Mexico and then spread through North America [25]. The first domestication of sunflower probably occurred in the eastern regions of North America. Although archaeological studies argued for an earlier cultivation in Mexico [26], molecular genetic studies have shown that modern sunflower cultivars are most genetically similar to wild accessions of the Midwestern USA [27,28]. Thus, it appears that sunflower was domesticated by Native Americans in eastern North America. The early domesticated genotypes were introduced to Europe at the beginning of the 16th century by naturalists [29–31]. A massive breeding program for high oil yield developed in Russia in the 19th century. In fact, even in North America, the first widespread cultivars were derived from materials reintroduced from Russia from this breeding program [32–34]. This implies a strongly reduced genetic variability in cultivated sunflowers in comparison to wild accessions, which colonized and adapted to multiple different environments [35].

Indeed, modern sunflower cultivars are quite different from wild accessions. They are generally single-headed, have specific oil profiles, and are dwarf. By the early 1970s, a massive increase in hybrid seed production occurred due to the availability of different heterotic groups of inbred lines as well as a system of cytoplasmic male sterility and fertility restoration derived from interspecific crosses with *Helianthus petiolaris* [36].

A number of studies have shown that genes affecting branching and other features of plant architecture, fatty acid biosynthesis, and flowering time were involved in sunflower domestication [37–42]. Baute et al. [41] analyzed the transcriptomes of wild and cultivated sunflowers and identified 137 genes associated with domestication and improvement, as indicated by their low sequence variability in domesticated genotypes compared to wild accessions. As in the previous studies, genes putatively involved in fatty acid biosynthesis, as well as in branching, were largely represented.

More recently, other authors [43] analyzing transcriptomes of wild and domesticated sunflowers have identified differential splicing divergence related to domestication, especially through intron

retention. Differential splicing has been related to genes involved in functions related to seed development. Many differential splicing patterns in cultivars probably derived from wild accessions, increasing their frequency because of selection during domestication.

The involvement of variation in the repetitive component, and especially of retrotransposon copy number, in sunflower domestication was first studied by Mascagni et al. [23]. The sunflower has a large genome of about 3.6 Gbp [42]. Its repetitive component accounts for around 80% of the genome and is mostly composed of LTR-REs [44–48], especially of the Gypsy superfamily and Chromovirus lineage. High levels of LTR-RE-related polymorphism have been found in both wild and cultivated genotypes [49].

Mobilization and consequent changes in the abundance of retrotransposons have occurred during *Helianthus* speciation, even in relatively recent times [50,51]. Sunflower LTR-REs are apparently transcribed and, although at low rates, reinserted into the genome, even in nonstressful environmental conditions [52].

In a previous study [23], a library of 123 LTR-retrotransposon sequence families of sunflower was produced assembling a set of 454 sequence reads of the HA412-HO line using RepeatExplorer (https://galaxy-elixir.cerit-sc.cz), a repetitive sequence online clustering tool [53]. Each cluster represents an individual family of repetitive elements, which show large sequence similarity and share a common progenitor [53]. Mascagni et al. [23] identified clusters belonging to the Gypsy and the Copia superfamilies (85 and 38 sequence families, respectively). The lineage (indicated as family in that work) of each cluster was also identified. Different clusters belonging to the same LTR-RE lineage can be defined as different LTR-RE families of that lineage. Mascagni et al. [23] showed changes in the abundance of certain lineages of Gypsy and Copia LTR-REs between cultivated and wild genotypes of sunflower. Moreover, they found differences in LTR-RE number lying proximal to gene coding sequences among the same genotypes.

Here, we extend the previous study [23], performing a new comparative analysis of LTR-REs between wild and domesticated genotypes of *H. annuus* at the family level in order to identify the involvement of specific LTR-RE families in retrotransposon-related structural variations and how they define cultivars in comparison to wild plants. Moreover, an analysis of the chromosomal localization of these LTR-RE families and of their supposed association to gene coding sequences was conducted, allowing us to hypothesize on the role of such structural variations in the domestication of the sunflower.

2. Materials and Methods

2.1. Plant Genotypes and Illumina Sequences Used in the Analyses

The sunflower cultivars and wild accessions used in this study were the same used by Mascagni et al. [23] (Table 1). We selected 7 wild accessions of *H. annuus* from different regions of North America, and 8 cultivars randomly selected from different countries in which sunflower seeds are massively produced, one cultivar per country. Wild accessions and cultivars were obtained from the United State Department of Agriculture, Agricultural Research Service (USDA-ARS), National Genetic Resources Program, USA. Further data on the genotypes can be found at the US National Plant Germplasm System webpage (http://www.ars-grin.gov/npgs/searchgrin.html) and in previous studies of ours [23,54].

Raw Illumina paired-end sequences from DNA isolated from leaves of single individuals of each genotype were available at the Sequence Read Archive (SRA) of NCBI (BioProject number PRJNA302358). Illumina reads were preprocessed [23] to remove Illumina adapters, then quality-trimmed with default settings, and the lengths of reads were defined at 90 nt.

Table 1. Sunflower genotypes used in this study. For each genotype, the United State Department of Agriculture (USDA) identification code, the area of cultivation for domesticated genotypes, and the number of reads sequenced by the Illumina technique are reported. Reads were trimmed at 90 nt and used in analyses as single ends for measuring LTR-retrotransposons (LTR-REs) abundance. For analyzing the proximity between LTR-REs and genes, paired ends were used and no specific length was defined.

Type	Name	Id Code	Area of Cultivation	Raw Reads	Trimmed Reads (as Single Ends, 90 nt)	Trimmed Reads (as Paired Ends)
Domesticated	Hata	Ames 22503	Argentina	32,100,390	31,085,284	31,624,960
	Dussol	Ames 22499	France	25,678,406	24,988,640	25,375,912
	Argentario	Ames 1842	Italy	10,759,866	10,134,402	10,566,652
	Karlik	Ames 3454	Spain	23,499,752	22,938,364	23,087,458
	Zelenka	Ames 22530	Russia	9,048,276	8,824,270	8,858,154
	Roman "A"	PI531386	Romania	19,408,888	18,621,244	19,095,974
	HOPI	PI369359	USA	15,768,198	15,254,502	15,437,790
	Seneca	PI369360	USA	13,911,506	13,334,732	13,667,436
Wild	Arizona (AZ)	Ames14400	-	14,641,510	14,013,588	14,357,666
	Colorado (CO)	PI586840	-	23,335,576	21,965,694	22,916,284
	Illinois (IL)	PI 435540	-	18,577,580	17,366,768	18,145,470
	Kentucky (KY)	PI 435613	-	14,853,802	13,845,748	14,580,828
	Mississippi (MS)	PI 435608	-	22,921,544	21,376,594	22,226,864
	North Dakota (ND)	PI586811	-	51,681,332	47,906,352	49,574,892
	Washington (WA)	PI 531018	-	6,996,658	6,479,624	6,724,410

2.2. Long Terminal Repeats-Retrotransposon Redundancy Estimation

A reference library of 11,546 contigs, belonging to 123 LTR-REs families and representative of all sunflower LTR-Res, was available [23]. This was obtained by graph-based clustering of sequences of the highly inbred sunflower line HA412-HO using RepeatExplorer [53].

This library was used as reference for mapping Illumina reads of each genotype. Mapping was carried out using an updated version of CLC-BIO Genomic Workbench (version 9.5.3, CLC-BIO, Aahrus, Denmark), with the following parameters: mismatch cost = 1, deletion cost = 1, insertion cost = 1, similarity = 0.9, and length fraction = 0.9.

Using this tool, those reads that match multiple distinct sequences were distributed randomly and, hence, the number of reads that matched to a single sequence simply gave an indication of its abundance. However, if all sequences of a sequence family (i.e., sequences that shared sufficient similarity to form a cluster) were taken together, the total number of mapped reads for that cluster (compared to the total number of all genomic reads) indicated the effective abundance of that family. Abundance values were reported as total number of mapped reads per million reads used for mapping.

The occurrence of retrotransposon abundance variation among wild and domesticated genotypes was estimated by a principal component analysis (PCA) and a permutational multivariate analysis of variance (PERMANOVA) [55]. For each family, the abundance data on 15 genotypes were used to build a Euclidean distance matrix. PCA was performed by implementation of R package FactoMineR version 1.26 [56]; PERMANOVA used R package vegan version 2.0-10 [57]. An in-house R script was used for performing statistical tests for all the families. Differences between wild and cultivated genotypes were considered significant when $p \leq 0.01$.

2.3. Retrotransposon Distribution along the Sunflower (HanXRQInbred Line) Genome

Using RepeatMasker (http://www.repeatmasker.org), each of the 17 linkage groups (LGs) of the only currently available sunflower genome sequence—the HanXRQ inbred line [42]—were compared with the datasets of Gypsy or Copia families, which showed significant differences in abundance between wild and cultivated genotypes. In addition, the analysis was also performed against a putative sunflower centromeric sequence—HAG002P01 [58]—separately under default parameters but -div 20. All LGs were then subdivided into 3-Mbp-long regions using an in-house Perl script. The number of masked bases was then counted for each 3 Mbp fragment using another in-house Perl script.

2.4. Analysis of Proximity of Long Terminal Repeats-Retrotransposons to Genes

For each genotype, a set of Illumina paired-end reads (trimmed for quality and adapters but not for specific length) was mapped onto a library containing a subset of LTR-RE families, assembled by an online clustering tool, RepeatExplorer (https://galaxy-elixir.cerit-sc.cz/), which showed differentiating abundance variation between cultivated and wild genotypes. They were also similarly mapped to a set of genes representing the whole sunflower transcriptome [59].

Mapping was performed using a Burrows–Wheeler Aligner (BWA) version 0.7.10-r789 [60] with the following parameters: aln -t 4 -l 12 -n 4 -k 2 -o 3 -e 3 -M 2 -O 6 -E 3. The resulting paired-end mappings were resolved with the "sampe" module of BWA, and the output was converted into a "bam" file using SAMtools version 0.1.19 [61]. SAMtools was used to extract the reads mapping in pairs with the function "view", option -F 12.

For each genotype, all read pairs where one read mapped onto a LTR-RE family and the other onto a gene sequence were selected, and the reads relating to the gene sequences were collected. Then, the corresponding gene sequences were retrieved from the HanXRQ genome annotation database (https://www.heliagene.org/HanXRQ-SUNRISE/), and Blast2GO [62] was used to identify the corresponding Kyoto encyclopedia of genes and genomes (KEGG) pathways (https://www.genome.jp/kegg/). Significant differences in the number of identified KEGG terms between cultivars and wild accessions were assessed by PERMANOVA, as described above.

3. Results

3.1. Some Long Terminal Repeats-Retrotransposon Families Show Significant Differences in Abundance between Wild and Cultivated Genotypes

An estimate of structural variation relating to the mobilization of LTR-REs can be determined by the increase or decrease of coverage by certain elements in the different genotypes [63]. In our study, the coverage of each family was determined in seven wild accessions of *H. annuus* from different regions of North America, and eight cultivars were randomly selected from different countries in which sunflower seeds are massively produced, one cultivar per country. In doing so, we were attempting to get a representative sample of diversity in the domesticated and wild gene pools of *H. annuus*. In fact, large genetic diversity among these genotypes has already been assessed using retrotransposon-based molecular markers [49].

The abundance of each of the 123 LTR-RE families contained in the reference library [23] in each accession was measured by counting the total number of reads (per million) that mapped onto the contigs belonging to such families. This method, which assumed that Illumina reads were sampled with uniform biases (among genotypes of the same species) for particular sequence types, if any, has been used in many species [64–67], including sunflower [23,46,54,68]. This method was previously validated by slot blot hybridization for two *Helianthus* genotypes [54].

Principal component analysis (PCA) of the intraspecific relative abundance of each of the 123 LTR-RE families was performed. No significant abundance variation was observed for the majority of LTR-RE families (101 out of 123 families). For 22 out of 123 families, wild and cultivated genotypes showed significant differences ($p \leq 0.01$; Figure 1). The abundance values (in millions of mapped reads per million) for each of the 22 LTR-RE families in the 15 selected genotypes are reported in Table S1.

In eight out of 22 LTR-RE families (seven of the Gypsy-Chromovirus lineage and one of the Copia-TAR/Tork lineage), the mean abundance was higher in cultivars than in wild accessions; the opposite trend was observed in the other 14 families (eight Gypsy-Chromovirus, three Gypsy-Athila, one Copia-Maximus/SIRE, one Copia-Angela, and one Copia-AleII; Figure 2). The percentage of LTR-RE families in which variation was higher in some lineages (such as Chromovirus and Athila of the Gypsy superfamily) than in other is shown in Figure 3.

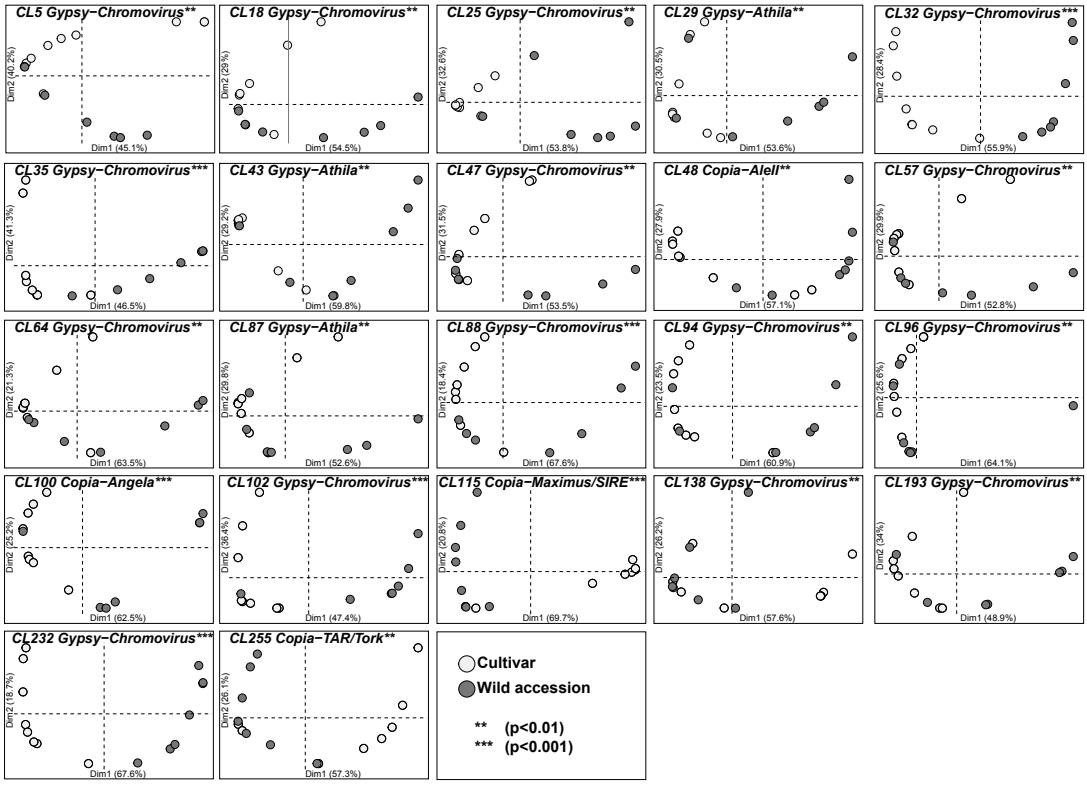

Figure 1. Principal component analysis (PCA) plots of abundance values of 22 LTR-RE families in domesticated (white dots) and wild genotypes (grey dots) of *Heliantus annuus*. The percentage of variation accounted by each axis is shown. Asterisks indicate permutational multivariate analysis of variance (PERMANOVA) significance between cultivars and wild accessions: *** $p < 0.001$; ** $p < 0.01$.

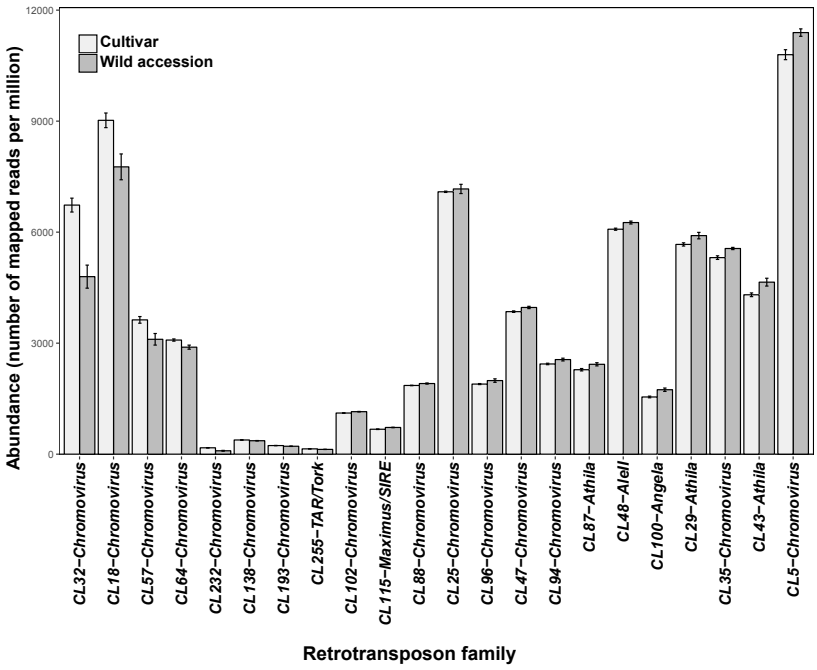

Figure 2. Mean number (± standard error) of mapped reads (x million) of each of the 22 LTR-RE families in eight cultivars and seven wild accessions of *H. annuus*. All families were differentially abundant by PERMANOVA, at $p < 0.01$.

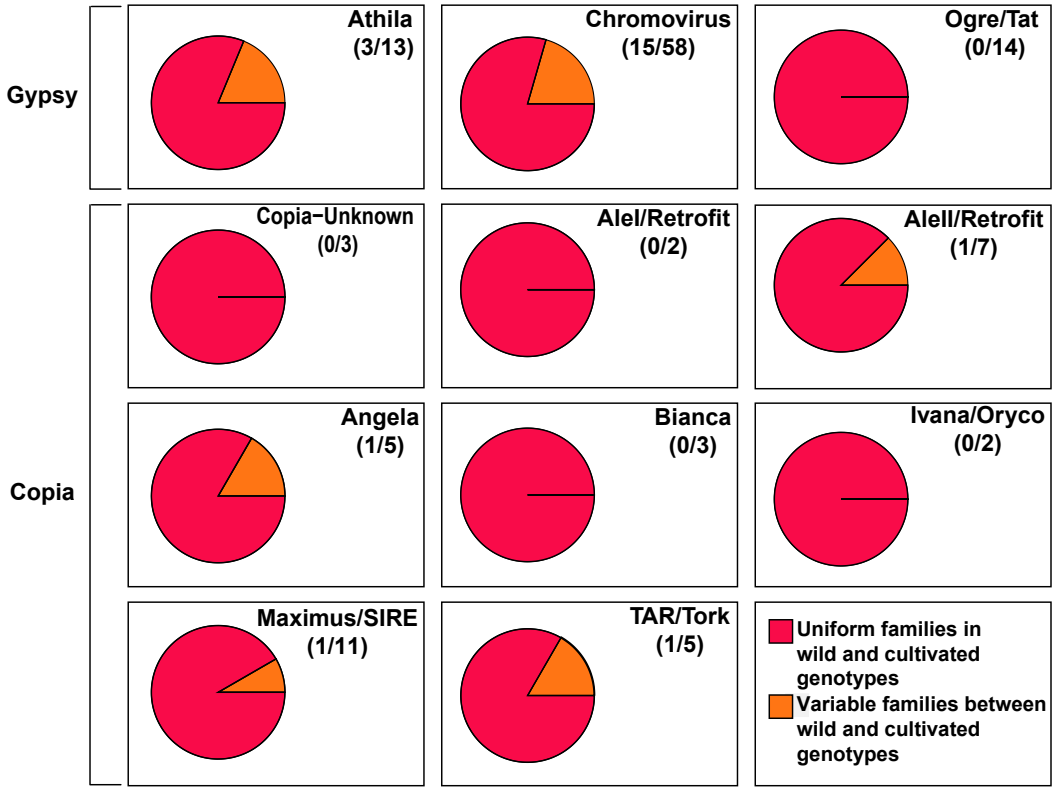

Figure 3. Percentage of long terminal repeats-retrotransposons (LTR-RE) families showing significant differences in abundance between wild and cultivated genotypes, according to the lineage to which such families belong. For each lineage, the total number of families in the sunflower genome is reported in parentheses.

3.2. Chromosomal Localization of Long Terminal Repeats-Retrotransposons Families

The 22 LTR-RE families showing significant variation in abundance between wild and cultivated genotypes were mapped to the only currently available genome sequence of *H. annuus* (HanXRQ inbred line [42]) using contigs from each family, though keeping the Gypsy and Copia superfamilies separated. To structurally describe the 17 linkage groups of the HanXRQ sequence, masking was also performed using a sequence that was previously described as interspersed, but with a prevalent centromeric localization by FISH [58], in order to identify putative centromeres.

The localization of 18 Gypsy and four Copia families is reported in Figure 4. No preferential localization was observed for Copia LTR-RE families, which were interspersed in the genome. While Gypsy families were also interspersed in the genome, similarly to Copia families, masking data suggested a preferential localization of Gypsy LTR-REs around putative centromeres in certain LGs (e.g., LGs I, IV, V, XI, XII, XV). This was consistent with the vast majority of Gypsy families that belong to the Chromovirus lineage, which has been shown to be involved in centromere structure [69].

3.3. Proximity of Retrotransposons to Genes

To evaluate the potential impact of LTR-RE insertions on overall gene function, we analyzed the association between sequences belonging to the 22 LTR-RE families showing significant differences in abundance between wild and cultivated genotypes and protein encoding genes in the genome.

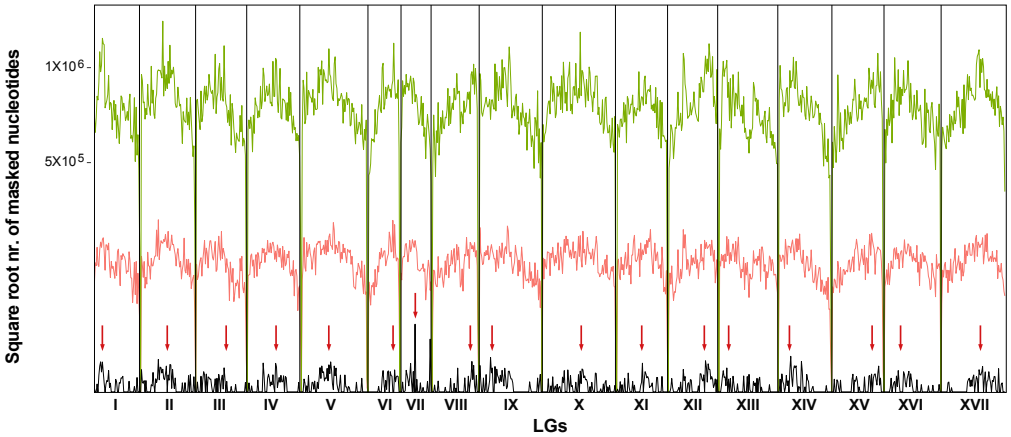

Figure 4. Distribution of the Gypsy (in green) and Copia (in red) LTR-RE families along the 17 linkage groups (LGs) of the sunflower genome (line HanXRQ [42]). The distribution of a putative centromeric sequence ([58], in black) is also reported. Red arrows indicate the most probable centromere position of each LG, corresponding to the peaks of highest frequency of the putative centromeric sequence. The space of each LG is proportional to its length in nucleotides.

The proximity of LTR-REs to genes in the wild accessions and in the cultivars of *H. annuus* was estimated by mapping Illumina paired-end reads to both the library of 22 differently abundant LTR-RE families and to a set of sequences representing the whole sunflower transcriptome [59]. The Illumina paired-end reads of which one mapped onto an LTR-RE and the other onto a gene (hereafter called gene-RE pairs) were retained from every accession (Table S2). Because of the relatively low number of genomic reads and to counter random variation, it was decided two groups of sequence would be created: gene-RE pairs from all wild genotypes and gene-RE pairs from all cultivated genotypes. In fact, because of the relatively low coverage of genomic reads used in this analysis for each genotype, any differences between single genotypes could have been determined by the stochasticity in read packages used for mapping. On the contrary, the effect of stochastic variation was greatly reduced by pooling all gene-RE pairs of all the wild genotypes and those of all the cultivars. The number of gene-RE pairs (per million reads) for each RE family in cultivars and wild accessions is reported in Table 2. On average, families of the Gypsy superfamily showed a higher number of gene-RE pairs per million reads in wild than in cultivated genotypes; for some families of the Chromovirus and Athila lineages, such differences were statistically significant (Table 2). By comparison, the number of gene-RE pairs per million reads did not change when considering LTR-RE families of the Copia superfamily in the wild and cultivated groupings.

The list of genes lying in proximity to elements belonging to LTR-RE families that showed different frequencies between cultivated and wild genotypes is reported in Table S3. A few of these genes have already been reported as being important during sunflower domestication [41] based on their sequence conservation in domesticated genotypes. Genes in cultivars included sequences for an iron-regulated protein 3, a protein of the EamA-like transporter family, an *O*-glycosyl hydrolase, a putative myosin, and an ATP synthase, subunit β. In wild accessions, sequences included genes for a carbon–sulfur lyase, a protein of the AMP-dependent synthetase and ligase family, a *P*-glycoprotein, a protein of the RING/U-box superfamily, and an RNA polymerase II transcription mediator.

In order to evaluate how such gene products might be involved in biochemical processes such that LTR-REs insertion could induce phenotypic variation between wild and cultivated genotypes, we identified the KEGG biochemical pathways of the genes lying near these LTR-REs (Figure 5). Comparing the percentages of KEGG terms between cultivated and wild genotypes, significant differences were observed for three biochemical pathways, oxidative phosphorylation, sulfur metabolism, and cysteine and methionine metabolism (Figure 5).

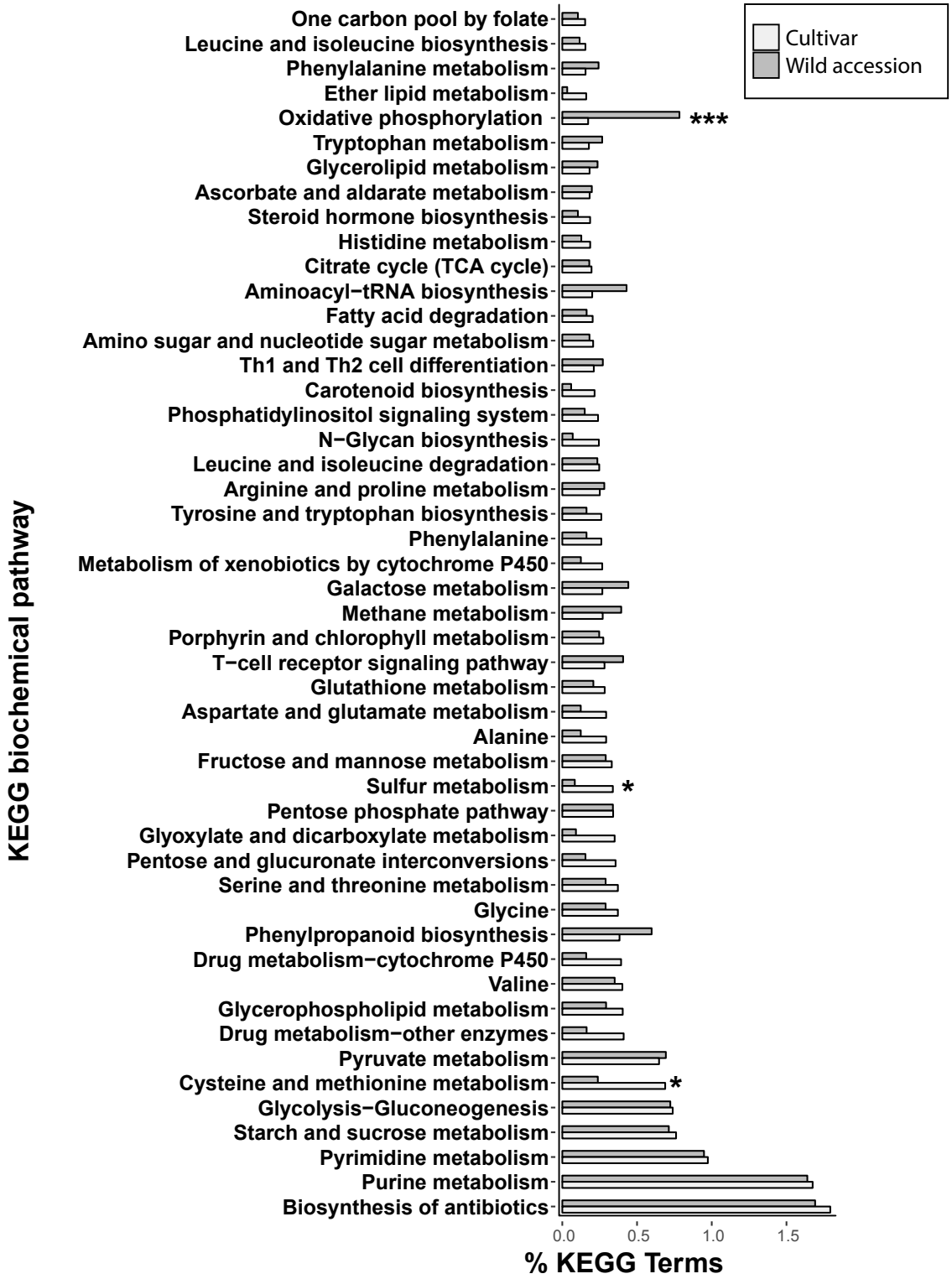

Figure 5. Percentages of Kyoto encyclopedia of genes and genomes (KEGG) biochemical pathway terms associated with gene-RE pairs in cultivars and wild accessions of *H. annuus*. Asterisks indicate PERMANOVA significance between cultivars and wild accessions: *** $p < 0.001$; * $p < 0.05$.

Table 2. Mean number (per million reads) of paired reads where one read maps to a LTR-RE family and the other to a gene sequence. For each family, the lineage and the superfamily are reported. Statistical significance of differences between cultivars and wild accessions was assessed by PERMANOVA (*** $p < 0.001$; * $p < 0.05$).

Superfamily	Lineage	Family	Mean nr. of Gene-RE Mapping Paired Reads per Million Reads		
			Cultivars	Wild accessions	PERMANOVA
Gypsy	Chromovirus	CL5	2.95	4.05	
Gypsy	Chromovirus	CL18	1.61	1.64	
Gypsy	Chromovirus	CL25	4.37	6.53	*
Gypsy	Chromovirus	CL32	1.07	0.80	
Gypsy	Chromovirus	CL35	0.80	0.90	
Gypsy	Chromovirus	CL47	0.84	1.57	***
Gypsy	Chromovirus	CL57	0.73	1.04	*
Gypsy	Chromovirus	CL64	0.47	0.65	
Gypsy	Chromovirus	CL88	0.47	0.81	*
Gypsy	Chromovirus	CL94	0.25	0.32	
Gypsy	Chromovirus	CL96	0.76	0.82	
Gypsy	Chromovirus	CL102	0.14	0.18	
Gypsy	Chromovirus	CL138	0.17	0.14	
Gypsy	Chromovirus	CL193	0.03	0.05	
Gypsy	Chromovirus	CL232	0.03	0.01	
Gypsy	Athila	CL29	1.16	1.34	
Gypsy	Athila	CL43	1.31	1.90	*
Gypsy	Athila	CL87	0.42	0.64	*
Mean Gypsy			0.98	1.38	
Copia	AleII	CL48	1.08	1.08	
Copia	Maximus/SIRE	CL115	0.12	0.21	
Copia	Angela	CL100	0.27	0.29	
Copia	TAR/Tork	CL255	0.18	0.17	
Mean Copia			0.41	0.44	

4. Discussion

It is commonly accepted that only a few loci are involved in the process of domestication of a species from its wild progenitor [70–73]. This could imply that sequence divergence between domesticated and wild genotypes might be more pronounced in those few loci that, for example, might provide characters that are favorable in cultivation but are neutral (or even negatively selected) in the wild. In these loci, extensive molecular divergence can be observed [74]. In addition, reduction in population size during artificial selection does contribute to making domesticated and wild genotypes more divergent [75]. Based on this hypothesis, 122 genes involved in sunflower domestication and 15 genes involved in sunflower cultivar improvement were identified [41].

Although coding sequence variations (and selection of specific alleles) would have played a major role in the process of domestication, other genetic mechanisms are also likely to have played a substantial part in this process, for example, differential regulation of gene expression including processes such as alternative splicing [76]. Recently, Smith et al. [43] showed an association of alternative splicing of some genes and domestication in sunflower.

Considering that phenotypic changes may arise from changes in the regulation pattern of genes, which are often derived from variation in neighboring noncoding, cis-regulatory sequences [20,73], variations in the repetitive component related to retrotransposon insertions/deletions could have had a primary role in determining the phenotypes that were selected by humans during plant domestication.

In previous experiments [23], significant differences in abundance of certain lineages of Gypsy and Copia LTR-REs were measured between cultivated and wild genotypes of sunflower. The present study extends those findings at the retrotransposon family level, showing that 22 of 123 LTR-RE families were significantly different in abundance between cultivated and wild genotypes. In addition, within a lineage, some families were significantly more abundant in cultivars, while others were more

abundant in wild accessions. The level of variation might be related to there being relatively few genotypes due to initial selection by early European explorers and the breeding program in Russia [27]. If it is assumed that the differences in LTR-RE abundance did not result in favorable traits of selected and bred genotypes, then the smaller or higher LTR-RE abundance of a specific family in cultivars than in wild accessions might be a consequence of genetic drift. However, if such changes in abundance of certain families resulted in favorable traits of those genotypes, it might be deduced that low or high abundance of certain LTR-RE families were unconsciously selected by the first breeders.

Long terminal repeats-retrotransposons families that showed different abundances between wild and cultivated plants were mapped to the sequenced inbred line of sunflower [42]. LTR-RE families were not confined to specific chromosomal regions notoriously filled with repetitive elements; rather, they were interspersed over all chromosomes. As such, structural variations related to these LTR-REs also occurred in gene-containing chromosomal regions.

Retrotransposon insertions or deletions may affect the phenotype of the host, especially when they occur in the proximity to genes whose expression rate they influence [77,78]. Retrotransposons mobilization may also affect the plant phenotype by modifying the epigenetic setting of the locus. This is because integration of a retrotransposon is generally accompanied by methylation of the insertion region, with consequent inactivation of proximal genes [17].

We assessed the occurrence of insertions/deletions in proximity to genes of the 22 LTR-REs belonging to families whose abundance changed between wild and cultivated sunflowers. Paired reads where one mapped onto an LTR-RE and the other onto a gene were identified. This indicated the proximity of LTR-RE insertions to genes they might influence. However, we could not exclude the possibility that in some cases, these gene sequences might have been unfunctional (e.g., portions of genes, pseudogenes).

For some LTR-RE families, the level of proximity to genes was significantly different between wild accessions and cultivars, most notably with families belonging to Chromovirus and Athila lineages of the Gypsy superfamily.

Interestingly, some of the genes indicated by Baute et al. [41] to be involved in sunflower domestication and improvement as determined by their sequence conservation, also showed differences in proximity to LTR-REs between cultivated and wild genotypes. Such differences could have contributed to the large phenotypic differences between wild and domesticated sunflowers. Analysis of the biochemical processes of genes with differentiating proximity to LTR-REs between wild and cultivated genotypes identified at least three important KEGG pathways: oxidative phosphorylation, sulfur metabolism, and cysteine and methionine metabolism.

Oxidative phosphorylation is a fundamental process in energy metabolism whereby cells oxidize nutrients, releasing energy for ATP production (see for example Reference [79]).

The sulfur metabolism and related cysteine and methionine metabolism pathways also play an important role in plants as they are involved in the production of reduced sulfur compounds for the biosynthesis of S-containing amino acids. These pathways are related to oxidative phosphorylation as sulfur compound production starts from the activation of sulfate with ATP to form adenylyl sulfate [80]. Among a large range of functions, sulfur-containing defense compounds are crucial for the survival of plants under abiotic and biotic stresses [81]. During drought stress, sulfur compounds have specific roles with the biosynthesis of osmolytes and osmoprotectants, such as polyamines, and the production of glutathione and its precursor cysteine are also increased [82].

A common feature of these biochemical processes is that they participate in some way in the response of plants to abiotic and biotic stress. It is known that a common trait of domesticated plants is an increasing susceptibility to environmental stresses [83]. It is possible that retrotransposon mobility might have affected this trait during sunflower domestication.

In conclusion, our study identified LTR-RE families specifically involved in structural variations between wild and cultivated sunflower genotypes. Our data suggest that such structural variations occurred in some cases near coding genes, with possible consequences on their expression and

consequently on the phenotype. In this sense, the occurrence of LTR-RE-related structural variation represents a further process that might have affected plant domestication, alongside the selection of alleles of specific genes. This study indicates what LTR-RE families and what genes should be taken into account in future studies on the importance of changes in the repetitive fraction of the genome in sunflower domestication. Resequencing the genome of some domesticated and wild sunflower genotypes will allow a precise measurement of the extent of LTR-RE-related structural variations and their localization to specific genes. Expression analysis of such genes will allow the effect of LTR-RE-related structural variation on the domesticated sunflower phenotype to be defined with more precision.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/9/9/433/s1. Table S1: Number of mapped reads per million reads of 22 LTR-RE families in eightdomesticated genotypes and seven wild accessions, Table S2: Number of gene-RE mapping paired reads per million reads of 22 LTR-RE families in eight domesticated genotypes and seven wild accessions, Table S3: List of genes mapped by gene-RE paired Illumina reads of domesticated (a) and wild (b) genotypes.

Author Contributions: Formal analysis, Investigation and Data curation, F.M. and A.V.; Writing and editing, F.M., A.V., T.G., A.C., and L.N.; Supervision, A.C. and L.N.

Funding: This research work was supported by the Department of Agriculture, Food and Environment of the University of Pisa, Italy, Project "Plantomics".

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [CrossRef] [PubMed]
2. Kumar, A.; Bennetzen, J.L. Plant retrotransposons. *Annu. Rev. Genet.* **1999**, *33*, 479–532. [CrossRef] [PubMed]
3. Wicker, T.; Keller, B. Genome-wide comparative analysis of *copia* retrotransposons in *Triticeae*, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. *Genome Res.* **2007**, *17*, 1072–1081. [CrossRef] [PubMed]
4. Llorens, C.; Futami, R.; Covelli, L.; Domínguez-Escribá, L.; Viu, J.M.; Tamarit, D.; Aguilar-Rodríguez, J.; Vicente-Ripolles, M.; Fuster, G.; Bernet, G.P.; et al. The Gypsy Database (*Gy*DB) of mobile genetic elements: Release 2.0. *Nucl. Acids Res.* **2011**, *39*, D70–D74. [CrossRef] [PubMed]
5. Natali, L.; Cossu, R.M.; Mascagni, F.; Giordani, T.; Cavallini, A. A survey of Gypsy and Copia LTR-retrotransposon superfamilies and lineages and their distinct dynamics in the *Populustrichocarpa* (L.) genome. *Tree Genet. Genomes* **2015**, *11*, 107. [CrossRef]
6. Vitte, C.; Bennetzen, J.L. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17638–17643. [CrossRef] [PubMed]
7. Ma, J.; Bennetzen, J.L. Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12404–12410. [CrossRef] [PubMed]
8. Wang, Q.; Dooner, H.K. Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17644–17649. [CrossRef] [PubMed]
9. Devos, K.M.; Brown, J.K.; Bennetzen, J.L. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **2002**, *12*, 1075–1079. [CrossRef] [PubMed]
10. Vitte, C.; Panaud, O. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice (*Oryza sativa* L.). *Mol. Biol. Evol.* **2003**, *20*, 528–540. [CrossRef] [PubMed]
11. Brunner, S.; Fengler, K.; Morgante, M.; Tingey, S.; Rafalski, A. Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* **2005**, *17*, 343–360. [CrossRef] [PubMed]
12. He, G.; Luo, X.; Tian, F.; Li, K.; Zhu, Z.; Su, W.; Qian, X.; Fu, Y.; Wang, X.; Sun, C.; et al. Haplotype variation in structure and expression of a gene cluster associated with a quantitative trait locus for improved yield in rice. *Genome Res.* **2006**, *16*, 618–626. [CrossRef] [PubMed]

13. Lisch, D. How important are transposons for plant evolution? *Nat. Rev. Genet.* **2013**, *14*, 49–61. [CrossRef] [PubMed]
14. Dubin, M.J.; Mittelsten Scheid, O.; Becker, C. Transposons: A blessing curse. *Curr. Opin. Plant Biol.* **2018**, *42*, 23–29. [CrossRef] [PubMed]
15. VanDriel, R.; Fransz, P.F.; Verschure, P.J. The eukaryotic genome: A system regulated at different hierarchical levels. *J. Cell Sci.* **2003**, *116*, 4067–4075. [CrossRef] [PubMed]
16. Song, X.; Sui, A.; Garen, A. Binding of mouse VL30 retrotransposon RNA to PSF protein induces genes repressed by PSF: Effects on steroidogenesis and oncogenesis. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 621–626. [CrossRef] [PubMed]
17. Hollister, J.D.; Gaut, B.S. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **2009**, *19*, 1419–1428. [CrossRef] [PubMed]
18. Hollister, J.D.; Smith, L.M.; Guo, Y.L.; Ott, F.; Weigel, D.; Gaut, B.S. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2322–2327. [CrossRef] [PubMed]
19. Vitte, C.; Fustier, M.A.; Alix, K.; Tenaillon, M.I. The bright side of transposons in crop evolution. *Brief. Funct. Genom.* **2014**, *13*, 276–295. [CrossRef] [PubMed]
20. Springer, N.M.; Ying, K.; Fu, Y.; Ji, T.; Yeh, C.T.; Jia, Y.; Wu, W.; Richmond, T.; Kitzman, J.; Rosenbaum, H.; et al. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* **2009**, *5*, e1000734. [CrossRef] [PubMed]
21. Albert, P.S.; Gao, Z.; Danilova, T.V.; Birchler, J.A. Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenet. Genome Res.* **2010**, *129*, 6–16. [CrossRef] [PubMed]
22. Naito, K.; Cho, E.; Yang, G.; Campbell, M.A.; Yano, K.; Okumoto, Y.; Tanisaka, T.; Wessler, S.R. Dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17620–17625. [CrossRef] [PubMed]
23. Mascagni, F.; Barghini, E.; Giordani, T.; Rieseberg, L.H.; Cavallini, A.; Natali, L. Repetitive DNA and plant domestication: Variation in copy number and proximity to genes of LTR-retrotransposons among wild and cultivated sunflower (*Helianthus annuus*) genotypes. *Genome Biol. Evol.* **2015**, *7*, 3368–3382. [CrossRef] [PubMed]
24. Schilling, E.E. Phylogenetic analysis of *Helianthus* (Asteraceae) based on chloroplast DNA restriction site data. *Theor. Appl. Genet.* **1997**, *94*, 925–933. [CrossRef]
25. Schilling, E.E.; Linder, C.R.; Noyes, R.D.; Rieseberg, L.H. Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. *Syst Bot.* **1998**, *23*, 177–187. [CrossRef]
26. Lentz, D.L.; Pohl, M.D.; Alvarado, J.L.; Tarighat, S.; Bye, R. Sunflower (*Helianthus annuus* L.) as a pre-Columbian domesticate in Mexico. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6232–6237. [CrossRef] [PubMed]
27. Harter, A.V.; Gardner, K.A.; Falush, D.; Lentz, D.L.; Bye, R.A.; Rieseberg, L.H. Origin of extant domesticated sunflowers in eastern North America. *Nature* **2004**, *430*, 201–205. [CrossRef] [PubMed]
28. Blackman, B.K.; Scascitelli, M.; Kane, N.C.; Luton, H.H.; Rasmussen, D.A.; Bye, R.A.; Lentz, D.L.; Rieseberg, L.H. Sunflower domestication alleles support single domestication center in eastern North America. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14360–14365. [CrossRef] [PubMed]
29. Zukovsky, P.M. *Cultivated Plants and Their Wild Relatives*; Commonwealth Agriculture Bureau: Farnham Royal, UK, 1950.
30. Meyer, R.S.; Purugganan, M.D. Evolution of crop species: Genetics of domestication and diversification. *Nat. Rev. Genet.* **2013**, *14*, 840–852. [CrossRef] [PubMed]
31. Olsen, K.M.; Wendel, J.F. A bountiful harvest: Genomic insights into crop domestication phenotypes. *Ann. Rev. Plant Biol.* **2013**, *64*, 47–70. [CrossRef] [PubMed]
32. Semelczi-Kovacs, A. Acclimatization and dissemination of the sunflower in Europe. *Acta Ethnogr. Acad. Sci. Hung.* **1975**, *24*, 47–88.
33. Korell, M.; Mosges, G.; Friedt, W. Construction of a sunflower pedigree map. *Helia* **1992**, *15*, 7–16.
34. Burke, J.M.; Tang, S.; Knapp, S.J.; Rieseberg, L.H. Genetic analysis of sunflower domestication. *Genetics* **2002**, *161*, 1257–1267. [PubMed]

35. Rogers, C.; Thompson, T.; Seiler, G.J. *Sunflower Species of the United States*; National Sunflower Association: Bismarck, ND, USA, 1982.
36. Leclercq, P. Une stérilité male cytoplasmique chez le tournesol. *Annales de l'Amelioration des Plantes* **1969**, *19*, 99–106. (In French)
37. Blackman, B.K.; Rasmussen, D.A.; Strasburg, J.L.; Raduski, A.R.; Burke, J.M.; Knapp, S.J.; Michaels, S.D.; Rieseberg, L.H. Contributions of flowering time genes to sunflower domestication and improvement. *Genetics* **2011**, *187*, 271–287. [CrossRef] [PubMed]
38. Chapman, M.A.; Burke, J.M. Evidence of selection on fatty acid biosynthetic genes during the evolution of cultivated sunflower. *Theor. Appl. Genet.* **2012**, *125*, 897–907. [CrossRef] [PubMed]
39. Mandel, J.R.; Nambeesan, S.; Bowers, J.E.; Marek, L.F.; Ebert, D.; Rieseberg, L.H.; Knapp, S.J.; Burke, J.M. Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet.* **2013**, *9*, e1003378. [CrossRef] [PubMed]
40. Mandel, J.R.; McAssey, E.V.; Nambeesan, S.; Garcia-Navarro, E.; Burke, J.M. Molecular evolution of candidate genes for crop-related traits in sunflower (*Helianthus annuus* L.). *PLoS ONE* **2014**, *9*, e99620. [CrossRef] [PubMed]
41. Baute, G.J.; Kane, N.C.; Grassa, C.; Lai, Z.; Rieseberg, L.H. Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytol.* **2015**, *206*, 830–838. [CrossRef] [PubMed]
42. Badouin, H.; Gouzy, J.; Grassa, C.J.; Murat, F.; Staton, S.E.; Cottret, L.; Lelandais-Brière, C.; Owens, G.L.; Carrère, S.; Mayjonade, B.; et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **2017**, *546*, 148–152. [CrossRef] [PubMed]
43. Smith, C.C.R.; Tittes, S.; Mendieta, J.P.; Collier-zans, E.; Rowe, H.C.; Rieseberg, L.H.; Kane, N.C. Genetics of alternative splicing evolution during sunflower domestication. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6768–6773. [CrossRef] [PubMed]
44. Santini, S.; Cavallini, A.; Natali, L.; Minelli, S.; Maggini, F.; Cionini, P.G. Ty1/*copia*-and Ty3/*gypsy*-like DNA sequences in *Helianthus* species. *Chromosoma* **2002**, *111*, 192–200. [CrossRef] [PubMed]
45. Natali, L.; Santini, S.; Giordani, T.; Minelli, S.; Maestrini, P.; Cionini, P.G.; Cavallini, A. Distribution of Ty3-*gypsy*- and Ty1-*copia*-like DNA sequences in the genus *Helianthus* and other Asteraceae. *Genome* **2006**, *49*, 64–72. [CrossRef] [PubMed]
46. Natali, L.; Cossu, R.M.; Barghini, E.; Giordani, T.; Buti, M.; Mascagni, F.; Morgante, M.; Gill, N.; Kane, N.C.; Rieseberg, L.H.; et al. The repetitive component of the sunflower genome as shown by different procedures for assembling next generation sequencing reads. *BMC Genom.* **2013**, *14*, 686. [CrossRef] [PubMed]
47. Staton, S.E.; Bakken, B.H.; Blackman, B.K.; Chapman, M.A.; Kane, N.C.; Tang, S.; Ungerer, M.C.; Knapp, S.J.; Rieseberg, L.H.; Burke, J.M. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* **2012**, *72*, 142–153. [CrossRef] [PubMed]
48. Giordani, T.; Cavallini, A.; Natali, L. The repetitive component of the sunflower genome. *Curr. Plant Biol.* **2014**, *1*, 45–54. [CrossRef]
49. Vukich, M.; Schulman, A.H.; Giordani, T.; Natali, L.; Kalendar, R.; Cavallini, A. Genetic variability in sunflower (*Helianthus annuus* L.) and in the *Helianthus* genus as assessed by retrotransposon-based molecular markers. *Theor. Appl. Genet.* **2009**, *119*, 1027–1038. [CrossRef] [PubMed]
50. Ungerer, M.C.; Strakosh, S.C.; Stimpson, K.M. Proliferation of Ty3/*Gypsy*-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. *BMC Biol.* **2009**, *7*, 40. [CrossRef] [PubMed]
51. Buti, M.; Giordani, T.; Cattonaro, F.; Cossu, R.M.; Pistelli, L.; Vukich, M.; Morgante, M.; Cavallini, A.; Natali, L. Temporal dynamics in the evolution of the sunflower genome as revealed by sequencing and annotation of three large genomic regions. *Theor. Appl. Genet.* **2011**, *123*, 779–791. [CrossRef] [PubMed]
52. Vukich, M.; Giordani, T.; Natali, L.; Cavallini, A. Copia and Gypsy retrotransposons activity in sunflower (*Helianthus annuus* L.). *BMC Plant Biol.* **2009**, *9*, 150. [CrossRef] [PubMed]
53. Novák, P.; Neumann, P.; Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform.* **2010**, *11*, 378. [CrossRef] [PubMed]
54. Mascagni, F.; Giordani, T.; Ceccarelli, M.; Cavallini, A.; Natali, L. Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L.). *BMC Genom.* **2017**, *18*, 634. [CrossRef] [PubMed]

55. Anderson, M.J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **2001**, *26*, 32–46.
56. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **2008**, *25*, 1–18. [CrossRef]
57. Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; et al. *Package 'Vegan', Community Ecology Package. R Package*, version 2.0-10; 2013.
58. Cavallini, A.; Natali, L.; Zuccolo, A.; Giordani, T.; Jurman, I.; Ferrillo, V.; Vitacolonna, N.; Sarri, V.; Cattonaro, F.; Ceccarelli, M.; et al. Analysis of transposons and repeat composition of the sunflower (*Helianthus annuus* L.) genome. *Theor. Appl. Genet.* **2010**, *120*, 491–508. [CrossRef] [PubMed]
59. Rowe, H.C.; Rieseberg, L.H. Genome-scale transcriptional analyses of first-generation interspecific sunflower hybrids reveals broad regulatory compatibility. *BMC Genom.* **2013**, *14*, 342. [CrossRef] [PubMed]
60. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
61. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
62. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [CrossRef] [PubMed]
63. Alkan, C.; Coe, B.P.; Eichler, E.E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **2011**, *12*, 363–376. [CrossRef] [PubMed]
64. Swaminathan, K.; Varala, K.; Hudson, M.E. Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. *BMC Genom.* **2007**, *8*, 132. [CrossRef] [PubMed]
65. Tenaillon, M.I.; Hufford, M.B.; Gaut, B.S.; Ross-Ibarra, J. Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zealuxurians*. *Genome Biol. Evol.* **2011**, *3*, 219–229. [CrossRef] [PubMed]
66. Barghini, E.; Natali, L.; Cossu, R.M.; Giordani, T.; Pindo, M.; Cattonaro, F.; Scalabrin, S.; Velasco, R.; Morgante, M.; Cavallini, A. The peculiar landscape of repetitive sequences in the olive (*Olea europaea* L.) genome. *Genome Biol. Evol.* **2014**, *6*, 776–791. [CrossRef] [PubMed]
67. Barghini, E.; Natali, L.; Giordani, T.; Cossu, R.M.; Scalabrin, S.; Cattonaro, F.; Šimková, H.; Vrána, J.; Doležel, J.; Morgante, M.; et al. LTR retrotransposon dynamics in the evolution of the olive (*Olea europaea*) genome. *DNA Res.* **2015**, *22*, 91–100. [CrossRef] [PubMed]
68. Mascagni, F.; Cavallini, A.; Giordani, T.; Natali, L. Different histories of two highly variable LTR retrotransposons in sunflower species. *Gene* **2017**, *634*, 5–14. [CrossRef] [PubMed]
69. Neumann, P.; Navrátilová, A.; Koblížková, A.; Kejnovský, E.; Hřibová, E.; Hobza, R.; Widmer, A.; Doležel, J.; Macas, J. Plant centromeric retrotransposons: A structural and cytogenetic perspective. *Mob. DNA* **2011**, *2*, 4. [CrossRef] [PubMed]
70. Wang, R.L.; Stec, A.; Hey, J.; Lukens, L.; Doebley, J. The limits of selection during maize domestication. *Nature* **1999**, *398*, 236–239. [CrossRef] [PubMed]
71. Gepts, P.; Papa, R. Evolution during domestication. In *Encyclopedia of Life Sciences*; Nature Publishing Group: London, UK, 2002.
72. Olsen, K.M.; Purugganan, M.D. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* **2002**, *162*, 941–950. [PubMed]
73. Doebley, J. The genetics of maize evolution. *Annu. Rev. Genet.* **2004**, *38*, 37–59. [CrossRef] [PubMed]
74. Innan, H.; Kim, Y. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10667–10672. [CrossRef] [PubMed]
75. Eyre-Walker, A.; Gaut, R.L.; Hilton, H.; Feldman, D.L.; Gaut, B.S. Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4441–4446. [CrossRef] [PubMed]
76. Bellucci, E.; Bitocchi, E.; Ferrarini, A.; Benazzo, A.; Biagetti, E.; Klie, S.; Minio, A.; Rau, D.; Rodriguez, M.; Panziera, A.; et al. Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Plant Cell* **2014**, *26*, 1901–1912. [CrossRef] [PubMed]

77. Butelli, E.; Licciardello, C.; Zhang, Y.; Liu, J.; Mackay, S.; Bailey, P.; Reforgiato-Recupero, G.; Martin, C. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* **2012**, *24*, 1242–1255. [CrossRef] [PubMed]
78. Falchi, R.; Vendramin, E.; Zanon, L.; Scalabrin, S.; Cipriani, G.; Verde, I.; Vizzotto, G.; Morgante, M. Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *Plant J.* **2013**, *76*, 175–187. [CrossRef] [PubMed]
79. Millar, A.H.; Eubel, H.; Jansch, L.; Kruft, V.; Heazlewood, J.L.; Braun, H.P. Mitochondrial cytochrome c oxidase and succinate dehydrogenase complexes contain plant specific subunits. *Plant Mol. Biol.* **2004**, *56*, 77–90. [CrossRef] [PubMed]
80. Cooper, A.J.L. Biochemistry of sulfur-containing amino acids. *Ann. Rev. Biochem.* **1983**, *52*, 187–222. [CrossRef] [PubMed]
81. Rausch, T.; Wachter, A. Sulfur metabolism: A versatile platform for launching defence operations. *Trends Plant Sci.* **2005**, *10*, 503–509. [CrossRef] [PubMed]
82. Chan, K.X.; Wirtz, M.; Phua, S.Y.; Estavillo, G.M.; Pogson, B.J. Balancing metabolites in drought: The sulfur assimilation conundrum. *Trends Plant Sci.* **2013**, *18*, 18–29. [CrossRef] [PubMed]
83. Tanksley, S.D.; McCouch, S.R. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* **1997**, *277*, 1063–1066. [CrossRef] [PubMed]

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).