Heredity (2018) 120:310–328
https://doi.org/10.1038/s41437-017-0006-5

the geneticssociety

ARTICLE

Evolutionary conservation and functional divergence of the *LFK* gene family play important roles in the photoperiodic flowering pathway of land plants

Ling Liu[1,2] · Yuanqi Wu[1,2] · Zhengqiao Liao[1,2] · Jing Xiong[1,2] · Fengkai Wu[1,2] · Jie Xu[1,2] · Hai Lan[1,2] · Qiling Tang[1,2] · Shufeng Zhou[1,2] · Yaxi Liu[3] · Yanli Lu[1,2]

Received: 21 January 2017 / Revised: 15 August 2017 / Accepted: 22 August 2017 / Published online: 11 December 2017

Abstract

ZEITLUPE (ZTL), LOV KELCH PROTEIN 2 (LKP2), and FLAVIN-BINDING KELCH REPEAT F-BOX 1 (FKF1)—blue-light photoreceptors—play important roles in regulating the circadian clock and photoperiodic flowering pathway in plants. In this study, phylogenetic analysis revealed that the LOV (Light, Oxygen, or Voltage) and Kelch repeat-containing F-box (*LFK*) gene family can be classified into two clades, *ZTL/LKP2* and *FKF1*, with clear differentiation between monocots and dicots within each clade. The *LFK* family genes underwent strong purifying selection; however, signatures of positive selection to adapt to local conditions still existed in 18 specific codons. In 87 diverse maize inbred lines, significant differences were identified ($P \leq 0.01$) for days to female flowering between the haplotypes consisting of eight positive selection sites at *ZmFKF1b* corresponding to tropical and temperate maize groups of the phylogenetic tree, indicating a key role of *ZmFKF1b* in maize adaptive evolution. In addition, positive coevolution was detected in the domains of the *LFK* family for long-term cooperation to targets. The Type-I and Type-II functional divergence analysis revealed subfunctionalization or neofunctionalization of the *LFKs*, and the *ZTL* subfamily is most likely to maintain the ancestral function of *LFKs*. Over 50% of critical amino acid sites involved in the functional divergence were identified in the Kelch repeat domain, resulting in the distinction of substrates for ubiquitination and degradation. These results suggest that evolutionary conservation contributes to the maintenance of critical physiological functions, whereas functional divergence after duplication helps to generate diverse molecular regulation mechanisms.

Introduction

The circadian clock regulates a diverse range of biological processes in plant growth and development, and promotes plant fitness by ensuring the coordination of endogenous rhythms and environmental cues (Greenham and McClung 2015; Shor and Green 2016). In plants, the onset of flowering is a significant transition from vegetative to reproductive phase, and the timing is strictly monitored by the circadian clock to ensure that plants flower at the right time (Greenham and McClung 2015; Song, et al. 2015). The circadian system is a crucial component of the plant photoperiodic flowering pathway, and the circadian clock-regulated components play major roles in triggering the day-length-specific expression of *FLOWERING LOCUS T* (*FT*) gene, which encodes florigen in leaves and moves to the shoot apical meristem (Corbesier, et al. 2007; Song, et al. 2013).

In *Arabidopsis*, the recently identified blue-light photoreceptors of E3 ubiquitin ligase—ZEITLUPE (ZTL)/LOV KELCH PROTEIN 2 (LKP2)/FLAVIN-BINDING KELCH REPEAT F-BOX 1 (FKF1)–are crucial for the regulation of the circadian clock and flowering time (Nelson, et al. 2000;

Ling Liu and Yuanqi Wu authors are contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41437-017-0006-5) contains supplementary material, which is available to authorized users.

✉ Yanli Lu
yanli.lu82@hotmail.com

[1] Maize Research Institute, Sichuan Agricultural University, 611130 Wenjiang, Sichuan, China

[2] Key Laboratory of Biology and Genetic Improvement of Maize in Southwest Region, Ministry of Agriculture, Wenjiang, China

[3] Triticeae Research Institute, Sichuan Agricultural University, 611130 Wenjiang, Sichuan, China

Schultz, et al. 2001; Somers, et al. 2000). The FKF1/LKP2/ZTL protein family contains three characteristic domains, including the light, oxygen, or voltage (LOV) domain located near the N-terminal; the F-box domain in the central region; and the six Kelch repeat domain at the C-terminal half (Zoltowski and Imaizumi 2014). The LOV domain belongs to a subfamily of the Per-ARNT-Sim (PAS) domain superfamily (Taylor and Zhulin 1999) and generally non-covalently binds a chromophore flavin mononucleotide and functions as a blue-light-sensing module (Christie, et al. 1999; Demarsy and Fankhauser 2009). The F-box protein is a component of the SKP1–Cullin–Rbx–F-box (SCF) E3 ubiquitin ligase complex that interacts with SKP proteins (Takahashi, et al. 2004). The C-terminal of the F-box domain contains specific binding sites for substrates for ubiquitination and subsequent degradation (Patton, et al. 1998). The Kelch repeat domain forms a beta-propeller according to the galactose oxidase crystal structure and functions as a protein–protein interaction domain that binds substrates for ubiquitin-mediated protein degradation (Andrade, et al. 2001). These domain structures indicate that *FKF1/ZTL/LKP2* mediates ubiquitin-dependent protein degradation under blue-light induction, and the protein turnover mediated by FKF1/ZTL/LKP2 is an essential part of the molecular mechanism underlying circadian rhythms (Nelson, et al. 2000; Schultz, et al. 2001; Somers, et al. 2000).

In *Arabidopsis*, the day-length-specific expression of FLOWERING LOCUS T (FT) protein is directly activated by the daytime CONSTANS (CO) protein (Song, et al. 2013; Song et al. 2012), which is only stabilized during the afternoon of long days (Valverde, et al. 2004). *FKF1* regulates *CO* mRNA expression and protein stability by degrading *CYCLING DOF FACTOR 1* (*CDF1*), 2 (*CDF2*), and 3 (*CDF3*) transcripts (Imaizumi, et al. 2005; Song et al. 2012). *GIGANTEA* (*GI*) has been shown to bind the LOV domain of *FKF1*, forming a blue-light absorption protein complex that regulates *CO* expression (Imaizumi, et al. 2003; Sawa, et al. 2007). In the photoperiodic pathway, *FKF1* is crucial to transfer the flowering information. Unlike *FKF1*, *ZTL* and *LKP2* affect the circadian rhythms by specifically interacting with *TIMING OF CAB EXPRESSION 1* (*TOC1*) and *PSEUDO-RESPONSE REGULATOR 5* (*PRR5*) via their LOV domains (Baudry, et al. 2010; Kiba, et al. 2007; Más, et al. 2003) to regulate flowering time.

Molecular biology analysis of single, double, and triple mutations of *fkf1*, *ztl*, and *lkp2* in *Arabidopsis* showed that the *LFK* genes exhibited functional differences in flowering time regulation (Nelson, et al. 2000; Somers, et al. 2004; Takase, et al. 2011). A single *fkf1* mutation resulted in late flowering under long day conditions (Nelson, et al. 2000), while the *ztl* mutation promoted weak flowering under short day conditions (Somers, et al. 2004; Takase, et al. 2011); further, the *lkp2* mutation enhanced the early flowering phenotype of the *ztl* mutant under both long day and short day conditions, depending on *FKF1* (Takase, et al. 2011). In addition, the *fkf1 ztl lkp2* triple mutations further reduced the *CO* expression level and slightly delayed flowering time compared to single *fkf1* mutations (Fornara, et al. 2009), indicating that *FKF1*, *ZTL*, and *LKP2* are involved in the regulation of the *CO* repressor protein stability. Overexpression of *LKP2* or *ZTL* leads to a late-flowering phenotype similar to that caused by the *ztl fkf1 lkp2* triple mutant under long day condition via the down-regulation of *CO* and *FT* expression (Somers, et al. 2004; Takase, et al. 2011), showing opposite functions to *FKF1* in the regulation of the photoperiodic pathway. Evolutionary analysis considering several species has shown that the *LFK* genes are divided into two groups, *ZTL/LKP2* and *FKF1* (Boxall, et al. 2005; Taylor, et al. 2010), corroborating their functional difference.

Gene families arose from multiple duplications of an ancestral gene, followed by mutation and divergence (Hartwell 2011). Gene duplication is important in providing new genetic materials to biological evolution, resulting in specialized or novel functions (Panchy, et al. 2016; Zhang 2003), playing a key role in increasing speciation and adaptation to environments (Long, et al. 2003; Zhang 2003). In the formation of gene families, duplicate genes mainly arise from unequal crossing over, retroposition, chromosomal duplication, or whole-genome duplication (WGD) (Panchy, et al. 2016; Zhang 2003). The generated paralogous might not maintain the original function, but allow acquiring a novel function or loss of function (Long, et al. 2003; Magadum, et al. 2013; Panchy, et al. 2016; Zhang 2003). Four major evolutionary fates of duplicate genes exist: (1) Pseudogenization, the process by which a functional gene becomes a pseudogene, which usually occurs in the first few million years after duplication if the duplicated gene is not under any selection (Lynch and Conery 2000). Functional redundancy of duplications is avoided by producing a mutation, a major force of pseudogenization (Magadum, et al. 2013; Zhang 2003). The mutations disrupting the structure and function of one of the two duplications are not deleterious and are not removed by selection, causing an unexpressed or functionless pseudogene (Zhang 2003). (2) Conservation of gene function. Ignoring the influence of functional redundancy of duplications, duplicate genes are important to generate some basic proteins or RNA products involved in biological processes, such as rRNAs and histones (Hurst and Smith 1998). Two possible mechanisms exist for paralogous genes to maintain the same function after duplication: concerted evolution ensure the members of a family to retain sequence and function similarity even after frequent gene conversion

and/or unequal crossing over (Hurst and Smith 1998), and purifying selection (Nei, et al. 2000) against mutations that modify gene function to prevent duplicated genes from diverging. (3) Subfunctionalization, in which each daughter gene adopts part of the functions of their parental gene (Hughes 1994). After duplication, both duplicates are maintained in the genome when they differ in some aspects of their functions (Nowak, et al. 1997). One form of subfunctionalization is the division of gene expression after duplication (Force, et al. 1999). (4) Neofunctionalization of duplicated genes to determine the origin of novel gene function, which requires various numbers of amino acid substitutions (Zhang, et al. 1998). However, in many cases, the functional divergence of daughter genes from parental ones occurs such that a related function, rather than an entirely new function, is acquired after gene duplication (Zhang 2003). For evolutionary forces underlying functional divergence of duplicate genes, both positive selection and relaxation of purifying selection are necessary for functional specialization and neofunctionalization of duplicates (Zhang, et al. 2002).

Recent genomic sequence data provide substantial evidence for the abundance of duplicated genes in all organisms surveyed. In plant genomes, an average of 65% of annotated genes are known to have at least one duplicate copy (Panchy, et al. 2016). Maize has been considered as a primary model organism for domestication and crop improvement studies because of the enormous phenotypic variation and genetic diversity (Strable and Scanlon 2009). This study focused on exploring variations and molecular evolution of the *LFK* family genes, which might be involved in blue-light-dependent protein degradation in flowering time regulation in plants. A genome-wide survey of *LFK* sequences in plant genomes was performed to identify *LFKs* and explore the distribution of genes in species. The evolutionary relationships of *ZTL/LKP2/FKF1* obtained from land plants were determined by reconstructing a phylogenetic tree and conducting selection and structural coevolution analyses. *ZmFKF1a* and *ZmFKF1b, LFK* family genes in maize, were further characterized among maize inbred lines and its relatives, and genetic effect analysis on heading date and silking date of maize was performed using eight positive sites. The amino acid residues responsible for functional divergence of the *LFK* subfamilies were also evaluated by detecting the functional constraints after gene duplications. By using an integrated analysis of sequence characteristics and divergent evolution of *LFK* genes, we aimed to establish a foundation for future experimental investigations to determine the functional roles of *LFK* genes in the photoperiodic flowering pathway in plants.

Materials and methods

Plant materials and phenotyping

Flowering-related traits in three environments were investigated with a collection of 87 diverse maize inbred lines (Supplementary Table S1), selected from different heterotic groups−stiff stalk, non-stiff stalk, and tropical or subtropical group. The panel was obtained from the International Maize and Wheat Improvement Center (CIMMYT), Mexico; the Sichuan Agricultural University, China; and the United States Department of Agriculture, USA. The plant materials were planted in Beijing (BJ), China (Shunyi, E 116°65′, N 40°13′) under natural long day conditions (May to August, day length > 14 h) and Hainan (HN), China (Sanya, E 109° 31′, N 18°14′) under short day conditions (November to March, day length < 11.5 h) in 2009 and in Sichuan (SC), China (Wenjiang, E 103°81′, N 30°97′) under natural long day conditions (April to July, day length > 13.5 h) in 2011. Two replicates were planted in each location of the two long-day and one short-day growing-season environments. Flowering-related traits were investigated and measured as days to male flowering (DMF: number of days from sowing to when 50% of the plants tassel) and days to female flowering (DFF: number of days from sowing to when 50% of the plants silk). Photoperiod sensitive index for DFF and DMF was calculated and evaluated using the growing degree days under long-day (BJ and SC) and short-day (HN) conditions.

Identification of LFK family genes across genomes

The known amino acid sequences of *Arabidopsis AtZTL/AtLKP2/AtFKF1* (AT5G57360/AT2G18915/AT1G68050) (Nelson, et al. 2000) were used as queries for the BLASTP search, and all LFK homolog and paralog proteins and coding sequences (CDSs) were identified and downloaded from the NCBI (http://www.ncbi.nlm.nih.gov/) and Phytozome v10.1 (http://phytozome.jgi.doe.gov/pz/portal.html) genome databases with default parameters. Their reliability was verified by detecting the protein-conserved domains by using the Pfam (http://pfam.xfam.org/), Conserved Domains and Protein Classification within NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/docs/cdd_search.html), and SMART (http://smart.embl-heidelberg.de/) databases. Sequences containing the three functional domains—LOV, F-box, and Kelch repeat—were considered to be *LFK* family members (Supplementary Table S2).

Cloning of LFK family members among maize germplasms

To achieve thorough molecular characterization of *LFKs*, we further cloned the two paralogous maize *ZmFKF1s* (*ZmFKF1a*, GRMZM2G107945 and *ZmFKF1b*, GRMZM2G106363). The genetic diversity and evolution pattern of *ZmFKF1a* among maize germplasms were analyzed in 87 maize inbred lines and those of *ZmFKF1b* were analyzed in 87 maize inbred lines and 33 teosinte accessions (Supplementary Table S1). The teosinte accessions included one each from *Zea nicaraguensis*, *Zea luxurians*, and *Zea diploperennis*; 26 from *Zea parviglumis* (diploid); and four from *Zea perennis* (tetraploid). Fresh leaves of five-leaf-old plants were harvested to extract DNA by using the CTAB method. The full-length clones of *ZmFKF1a* and *ZmFKF1b* containing the 5′ UTR and 3′ UTR were amplified using specific primers (Supplementary Table S3 and Figure S1). The corresponding CDSs of each accession were obtained by calling from full-length clones based on the maize B73 reference sequence. The CDSs were then translated to proteins by using DNAMAN 6.0 (Lynnon BioSoft) software package to verify their correctness. Polymerase chain reaction (PCR) was performed under the cycling conditions stated in the Phanta® Max Super-Fidelity DNA Polymerase instructions manual (Vazyme Biotech Co., Ltd.). PCR products were purified using the E.Z.N.A.® Gel Extraction Kit (Omega Bio-tek®) and cloned using the pEASY®-T1 Cloning Vector (TransGen Biotech Co., Ltd.) following manufacturer's protocol. Positive clones were selected for sequencing by using the Sanger method at the Beijing Genome Institute (Shenzhen, China).

Sequence alignment

Multiple protein and DNA sequences were aligned using MUSCLE 3.6 (Edgar 2004) and manually improved in the BioEdit 7.0 (Tom Hall) software package. PAL2NAL 14 (Suyama, et al. 2006) was used to build codon-based alignments of CDSs based on the protein alignments. An efficient and accurate phylogenetic analysis was performed by trimming poorly aligned regions in the two ends of multiple sequence alignments by using ClustalX 2.0 (Chenna, et al. 2003).

Characterization of ZmFKF1s

The software package DnaSP 5.10.1 (Librado and Rozas 2009) was used to calculate nucleotide diversity and genetic relationships of *ZmFKF1s* sequences among maize and teosinte germplasms. Two measures of nucleotide diversity were applied: the average number of nucleotide differences per site between two DNA sequences in all possible pairs in the sample population (π) (Nei and Li 1979) and the population mutation parameter estimate (Watterson's θ) (Watterson 1975). The number of polymorphic sites, insertion/deletion (InDel) changes, and the level of linkage disequilibrium (LD) were extracted and measured using TASSEL 5.0 (Bradbury, et al. 2007). LD decay was measured by averaging r^2 values over a distance of 100 bp and plotting the values against distance. LD plot was generated in Haploview 4.2 (Barrett, et al. 2005).

Reconstruction of phylogenetic trees of LFK gene family and ZmFKF1s

Phylogenetic analyses for the *LFK* gene family were conducted based on 171 proteins from 67 plant species and 151 CDSs for 59 species, as well as for *ZmFKF1a* and *ZmFKF1b* from 87 maize inbred lines and *ZmFKF1b* from 33 teosinte accessions by using neighbor joining (NJ), maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) methods. The tree of *LFK* gene family was constructed by using the most ancient species, bryophyte *Marchantia polymorpha*, as the outgroup (accession number: BAO66508.1). ProtTest 3.0 (Darriba, et al. 2011) and jmodeltest 2.1.6 (Posada 2008) were used to determine the best-fit evolutionary model before constructing the phylogenetic trees.

NJ trees were computed in MEGA 5.0 (Tamura et al. 2011) with 5,000 bootstrap replicates and a gamma distribution estimator (6 gamma distribution). Pairwise deletion was adopted for handling alignment gaps and missing data. MP analysis was performed using PAUP* 4.0b10 (Swofford 2002). Heuristic searches were conducted using tree bisection-reconnection branch swapping and random order of taxon addition. Heuristic searches and bootstrap support for nodes were estimated using 100 and 1,000 replicates, respectively. ML trees were constructed using PhyML 3.1 (Guindon et al. 2009) with 500 bootstraps. BI analysis was performed using Mrbayes 3.2 (Ronquist and JP 2003). Each of the four Markov Chain Monte Carlo chains (one cold and three heated) was generated by 10,000,000 iterations with a sampling frequency of 1000 iterations. The first 25% of trees from all runs were discarded as burn-in and excluded from the analysis, and the remaining trees were used to construct the majority rule consensus trees. The statistical confidence in nodes was evaluated using posterior probabilities.

Selection pressure and genetic effect analysis for positive selection loci

The possible selection forced on the *LFKs* was determined using the codon substitution models implemented in the CODEML program within the PAML 4.8 package (Yang

2007) to estimate the ratio of non-synonymous to synonymous substitution rate d_N/d_S (ω). A ω ratio of >1, =1, and <1 were indicative of positive, neutral, or purifying selection on the protein, respectively. The ω estimate was conducted for the *LFK* gene family and three duplications (*ZTL*, *FKF1*, and *LKP2*) in different plant species, as well as for *ZmFKF1a* and *ZmFKF1b* in maize and teosinte based on the codon sequence alignment across all sites and all lineages in each tested group. In addition, to further validate the selection pattern existing in the *ZmFKF1s*, we calculated Tajima's *D* and Fu and Li's *D* statistic as neutrality tests based on DNA sequence alignment and nucleotide variations of *ZmFKF1a* and *ZmFKF1b*.

Given that positive selection very rarely affects all sites over prolonged periods, it is essential to detect positive selection for some sites. We used site-specific models, which allow the ratio to vary among amino acid sites by using the variable NS sites models, including M0 (assumes the same ω for all sites), M3 (assumes a general discrete distribution), M1a (neutral model, ω values fit into two site classes: one for conserved sites with values between 0 and 1, and the other for neutral sites fixed at $\omega = 1$), M2a (positive selection model, similar to M1a, but with an extra class of ω of > 1), M7 (neutral model, ω values fit into a beta distribution, in the range from 0 to 1), and M8 (positive selection model, similar to M7, but with an extra class of ω of > 1), forming three paired comparisons: M0 vs. M3, M1a vs. M2a, and M7 vs. M8 to test for the presence of sites under positive selection. Codon frequencies were estimated by the F3 × 4 method. Likelihood ratio tests were obtained to determine whether allowing codons to evolve under positive selection yielded a significantly better fit to the data between two paired comparisons each. Twice the log-likelihood difference ($2\Delta \ln L$) between the two models approximately obeyed the χ^2 distribution, and χ^2 test could thus be performed using the difference in the numbers of free parameters between the two models as the χ^2 test degrees of freedom (*df*). The Bayes Empirical Bayes estimate from each selection model was then used to calculate the posterior probabilities of a specific codon site and identify those that were most likely under positive selection; M3 was an exception, since only the results of Naive Empirical Bayes analysis were available.

The genetic effects of the positive selection sites on flowering-related traits and their relation with phylogenetic tree were further investigated in maize germplasm. Three methods–general linear model (GLM), GLM with population structure (PCA) incorporated, and mixed linear model (MLM) incorporating both population structure and kinship (K)–were used in TASSEL 5.0 (Bradbury, et al. 2007). Haplotypes analysis was then conducted using the identified positive selection sites with minor allele frequency (MAF) ≥ 0.05, corresponding to the clustering groups with tropical and temperate maize germplasm. The effects were evaluated using the lme4 package in R.

Protein 3D structure prediction

To further investigate the possible effect of the identified amino acid sites under positive selection, we predicted the three-dimensional structure of ZmFKF1b protein by using I-TASSER 5.0 (Yang and Zhang 2015), with a de novo search for the possible folds based on multiple threading alignment approaches. The predicted structure models were finally confirmed by matching with the known proteins in the function databases, and residues under positive selection were further mapped onto the obtained 3D structures by using Chimera 1.10 (Pettersen, et al. 2004).

Coevolution in protein domains

Coevolution of domains within a single protein is necessary to guarantee molecular function to be actively coordinated among different independent domains. The coevolution between binding partners of *LFKs* was revealed by conducting a correlation analysis of the three domains of LFK proteins as Goh's method (Goh et al. 2000). The multiple sequence alignment of FKF1/ZTL/LKP2 amino acids was divided into five independent alignments: one for each domain plus two links. According to the protein structure of *LFK* domains and the amino acid position of *AtFKF1*, the LOV, F-box, and Kelch repeat domains included residues 55–164, 210–256, and 304–619, respectively. Fifteen (three genes multiplied by five segments) N × N distance matrices were generated from the multiple alignments in MEGA 5.0, where *N* is the number of analyzed sequences. The phylogenetic trees were drawn in MEGA 5.0 (Tamura et al. 2011). The coevolution of interaction partners was quantified by computing the linear Pearson product-moment correlation coefficient (*r*) for all pairwise distances in any two corresponding distance matrices as a derivation of Pearson's *r* formula (Press, et al. 1988), with $-1 \le r \le +1$. Positive or negative values of *r* would indicate a positive coevolution or anti-correlation; r values of around zero would indicate no correlation. The statistical significance of the correlation was further assessed by performing a random shuffle analysis with 1000 bootstrap replicates in R 3.1.0 (R Core Team 2014), yielding an estimate of the standard deviation of r_σ. Next, r_σ was generated from the computed bootstrap correlation coefficient r_b in the bootstrap analysis. From the resulting 1000 values of the 1000 generated sets, a *z*-score for the actual observed value *r* was calculated as:

$$z = \frac{r - \overline{r}_b}{r_\sigma}$$

where $\bar{r}_b$ is the mean value of r_b. Subsequently, the p value was obtained by following: $erfc(|z|)/\sqrt{2}$, where $erfc$ is the complement error function. Positive r values close to 0.8 are suggested to indicate a strong positive coevolution (Goh et al. 2000).

Functional divergence analysis

Type-I and Type-II functional divergences of *LFKs* were tested using Diverge 3.0 (Gu, et al. 2013). Type-I and Type-II functional divergences are the two basic types of amino acid configurations to estimate the level of functional divergence, and to predict important amino acid residues for these functional differences between member genes of a gene family (Gu 2001). Type-I functional divergence occurs shortly after gene duplication owing to site-specific evolutionary rate changes between paralogous clusters (amino acid residues are highly conserved in a gene cluster, whereas they are highly variable in another cluster), resulting in altered functional constraints between duplicate genes. Type-II functional divergence occurs in the late phase after gene duplication when evolutionary rates are consistent, resulting in changes in the biochemical properties of amino acids, but not in the altered functional constraints. The degree of functional divergence in the relative evolutionary rates at the sites between protein subfamilies was measured using θ (ranges from 0 to 1). A 0 estimate indicates that the two tested clusters have the same relative rates among sites, whereas values approaching 1 reflect increasing differences between the relative rates among sites in the two subfamilies. For Type-I functional divergence analyses, a likelihood ratio test (LRT) was constructed for testing whether the alternative hypothesis $\theta_I > 0$ can be accepted. Subsequently, the calculated fold score based on Type-II functional divergence estimate (θ_{II}) and its standard error (θ_{SE}) were used to test whether θ_{II} is significantly larger than 0 through the Z-score test (normal distribution test).

The amino acid sites that might be critically involved in the functional divergence of the *LFK* gene family were further identified by determining a posterior probability of divergence (Qk) for each site. The cut-off Qk value was determined by consecutively eliminating the highest scoring residues from the alignment until the coefficient of functional divergence reached 0. Qk values of ≥ 0.80 and significance levels of ≤ 0.01 were generally used to control false positives of residues identified as responsible for Type-I and Type-II functional divergence. Values of $Qk \geq 0.90$ were thought to indicate a high probability of different evolutionary rates or physiochemical properties of amino acid sites between two tested clusters.

Results

LFK genes were only identified in land species

In order to maximize the identification of *LFKs* and gain a comprehensive understanding of the gene distribution in different plant species, a preliminary BLASTP search was performed. In all, 171 proteins and 151 CDSs containing LOV (PAS), F-box, and Kelch repeat domains were found in 67 and 59 land plant species, respectively, including bryophytes, ferns, gymnosperms, and angiosperms (Supplementary Table S2). *LFKs* were not found in blue and green algae and photosynthetic stramenopiles. The specificity of *LFKs* in Embryophyta (land plants) suggests that *LFKs* might have originated during land plant differentiation about 360 million years ago in the Devonian period. Interestingly, *LFKs* were identified in *M. polymorpha*, but not in *Physcomitrella patens*, although both belong to Bryophyta, and could be considered to be the trailbreakers for the transition of plants from aquatic to terrestrial form and the oldest groups of existing higher plants. The differences in *LFKs* among *M. polymorpha* and *P. patens* might be attributed to the different morphologies of gametophytes and sporophytes, such as the peculiar flattened thallus and elaters of *M. polymorpha*, which are associated with photosynthesis and spore propagation.

Generally, LFK protein sequences are highly conserved and have a rigid gene structure with a similar domain order; the number of *LFK* genes varies across 67 plant species, increasing along with genome size or chromosome number of plants (Supplementary Table S2). In the JGI database, 78 *LFK* genes were identified in 38 sequenced land plant genomes. Of the 38 species, five contained only one subfamily member: two for each of *FKF1*and *ZTL* and one for *LKP2*. Twenty six species contained two gene subfamilies (*ZTL* and *FKF1*) and only seven species, all belonging to Brassicaceae (e.g., *Arabidopsis thaliana* and *Capsella rubella*), contained three subfamilies (*ZTL*, *FKF1*, and *LKP2*), (Supplementary Fig. S2).The vast majority of plant species contain two subfamily members, *ZTL* and *FKF1*: 65.67% in the 67 investigated species and 68.42% in the 38 completely sequenced species present in the JGI database (Supplementary Fig. S2), and *LKP2* seemed dispensable in most plant species. Sequence conservation of *LFKs* and gene distribution among species indicate a functional similarity among *ZTL*, *FKF1*, and *LKP2* and somewhat functional differentiation between *ZTL* and *FKF1*.

Characterization of ZmFKF1s among maize germplasms

The two *ZmFKF1s*, the 4363-bp long *ZmFKF1a* (GRMZM2G107945) and 3110-bp long *ZmFKF1b*

Table 1 Nucleotide diversity and neutrality test of *ZmFKF1s* in maize and teosinte

Gene	Taxon	N^a	Region	Size (bp)[b]	π ($\times 10^{-3}$)	θ ($\times 10^{-3}$)	K	Tajima's D	D'	F'
ZmFKF1a	Maize	87	5′-UTR	276	6.85	8.63	1.889	−0.56	−0.97	−0.9
			Exon1	226	5.20	7.03	1.175	−0.64	−2.68*	−2.35*
			Intron	1492	15.26	23.28	22.77	−1.17	−1.89	−1.90
			Exon2	1592	5.91	4.99	9.41	0.59	0.26	0.46
			3′-UTR	351	8.97	17.53	3.15	−1.52	−2.42*	−2.48*
			Total	3940	9.73	13.30	38.32	−0.92	−1.81	−1.72
ZmFKF1b	Maize	87	5′-UTR	116	6.99	15.40	0.81	−1.39	−3.81**	−3.53**
			Exon1	181	10.53	36.19	1.91	−2.21**	−6.35**	−5.68**
			Intron	690	11.38	24.17	7.85	−1.75	−6.59**	−5.52**
			Exon2	1583	6.45	23.33	10.21	−2.46**	−8.08**	−6.85**
			3'-UTR	382	14.21	23.91	5.43	−1.3	−5.42**	−4.54**
			Total	2952	8.88	24.07	26.21	−2.16*	−7.83**	−6.49**
ZmFKF1b	Teosinte	33	5′-UTR	306	15.94	30.60	4.88	−1.73	−2.86*	−2.94*
			Exon1	250	6.11	14.78	1.53	−1.94*	−2.92*	−3.07*
			Intron	659	19.01	30.66	12.53	−1.42	−1.31	−1.60
			Exon2	1591	7.86	19.82	12.50	−2.29**	−4.21**	−4.21**
			3′-UTR	481	13.92	28.69	6.70	−1.90*	−1.94	−2.28
			Total	3287	11.6	23.9	38.13	−1.97*	−3.01*	−3.15*

π: Kimura's nucleotide diversity per base pair estimated from the average heterozygosity per site

θ: Watterson's nucleotide diversity per base pair estimated from the number of segregating sites

K: average number of nucleotide differences. D': Fu and Li's D; F': Fu and Li's F

[a]The number of analyzed maize and teosinte lines

[b]The total number of sites (excluding sites with gaps/missing data)

*$P < 0.05$, **$P < 0.01$

(GRMZM2G106363) in maize B73, were further cloned to characterize sequence evolution. Multiple sequence alignment of *ZmFKF1a* across the 87 tested maize inbred lines formed a 6056-bp DNA sequence matrix (Dataset 1). A total of 182 polymorphic loci were identified over all the amplicons with MAF ≥ 0.05 (Supplementary Table S4), including 140 SNPs and 42 InDels. The 140 SNPs contained 96 (68.57%) transitions and 44 (31.43%) transversions. Comparison of polymorphism density across different gene regions suggests that introns contain higher sequence diversity than exons, whereas UTRs show the highest sequence diversity (Supplementary Table S4). Nucleotide sequence alignments of *ZmFKF1b* from the 87 maize inbred lines (Dataset 2) and 33 teosinte accessions (Dataset 3) showed the following features: (a) maize: 62 polymorphic loci were identified with MAF ≥ 0.05, with a total of 49 SNPs and 13 InDels; and (b) teosinte: 158 polymorphic loci, including 119 SNPs and 39 InDels with MAF ≥ 0.05,. The *ZmFKF1b* gene in teosinte was found to harbor a considerably higher level of nucleotide diversity, with an average of one SNP or InDel in every 26 to 80 bp. The detailed information is shown in Supplementary Table S4. Transitions accounted for 26 (53.06%) of the 49 SNPs in maize and for 71 (59.66%) of the 119 SNPs in teosinte, which were slightly higher than those for transversions (46.94% in maize and 40.34% in teosinte; Supplementary Table S4). In maize, the lowest polymorphism density was noted in exons, followed by that in introns, and the highest was in UTRs (Supplementary Table S4). In teosinte, however, the highest and lowest density regions occurred in the introns and exons, respectively. *ZmFKF1a* had richer genetic diversity than *ZmFKF1b* in maize, and considerably more variation was noted in the *ZmFKF1b* gene within teosinte population than in the maize population.

In addition, two overall estimates of total nucleotide diversity—Watterson's theta (θ_W) and π—were separately calculated for *ZmFKF1a* and *ZmFKF1b* in maize and teosinte (Table 1). The estimated nucleotide diversity in different gene regions was overall high, but unevenly distributed. For *ZmFKF1a*, the lowest nucleotide diversity ($\pi = 5.20 \times 10^{-3}$) was found in exon 1, with $\theta_W = 7.03 \times 10^{-3}$, whereas the most abundant diversity was registered in intron ($\pi = 15.26 \times 10^{-3}$, $\theta_W = 23.28 \times 10^{-3}$) in maize (Table 1). For *ZmFKF1b*, the lowest nucleotide diversity ($\pi = 6.45 \times 10^{-3}$) was found in exon 2, with $\theta_W = 23.33 \times 10^{-3}$, whereas the most abundant diversity was registered in 3′-UTR ($\pi = 14.21 \times 10^{-3}$, $\theta_W = 23.91 \times 10^{-3}$) in maize, followed by that in the intron (Table 1). In teosinte, the two

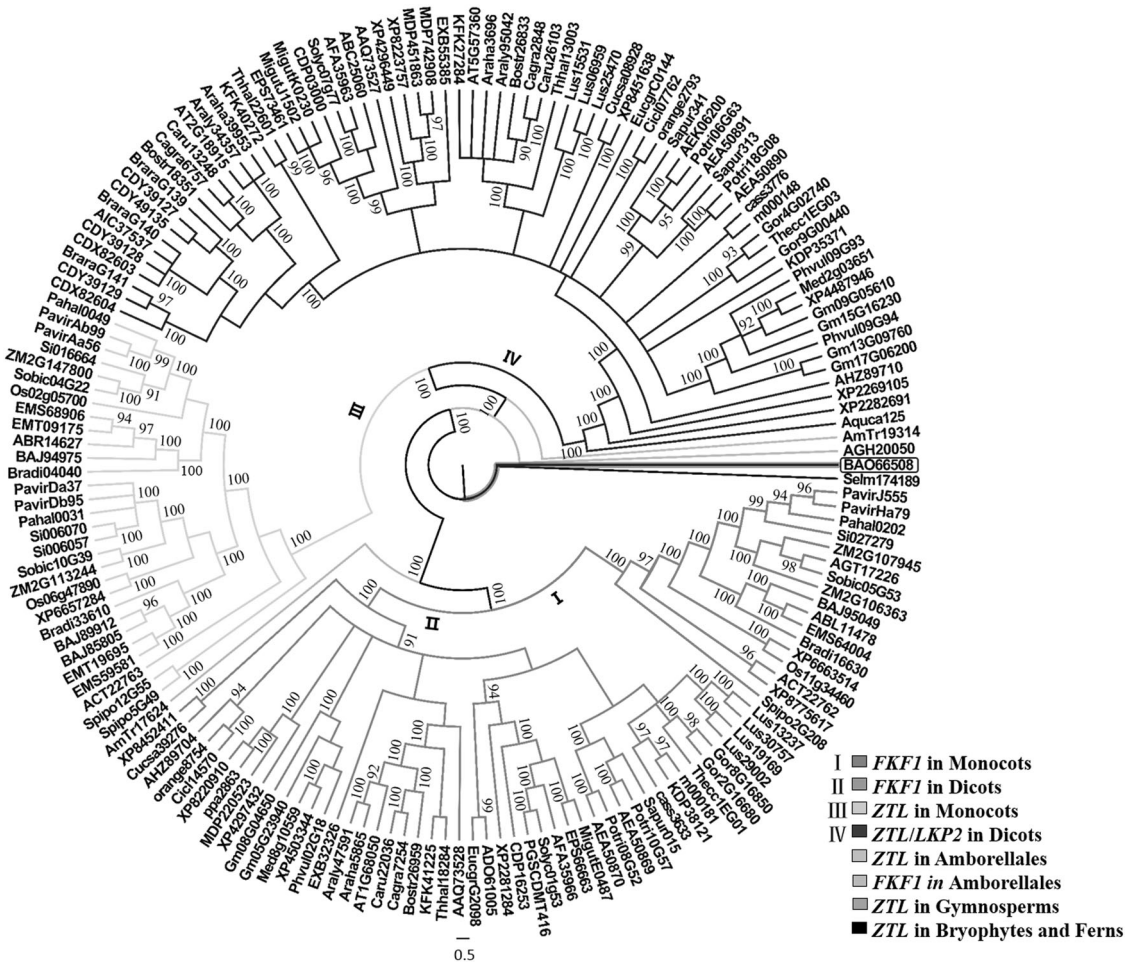

Fig. 1 A Bayesian tree inferred from the full-length protein sequences of 171 *LFKs* in 67 land plants. The phylogenetic tree was rooted using a lineage of Bryophytes (*Marchantia polymorpha*, BAO66508.1) as an outgroup. The four major clusters and four ancestral species' proteins that did not well group into clusters were labeled with colored lines. Numbers above nodes are Bayesian inference posterior probability of > 90%, and tip labels are short names of gene IDs

coding regions, exons 1 and 2, showed relatively low nucleotide diversity, and the noncoding regions, especially in intron, showed the highest diversity based on π and θ_W values. The results indicate that there are higher genetic variation in coding regions of *ZmFKF1b* than in *ZmFKF1a*.

Joint analysis of nucleotide diversity (π) and LD along with sliding window could gain deeper insight into the sequence characterizations of *ZmFKF1b* in the tested population. The overall nucleotide diversity was relatively higher in teosinte than in maize, but not evenly distributed in different regions, with encoding regions (exon 1 and exon 2) containing lower genetic variation (Supplementary Fig. S3 A; Table 1) in teosinte, which was supported by LD patterns shown in Supplementary Fig. S3B and C. The r^2 declined to 0.2 at a distance of ~100 bp in teosinte, whereas considerably slower LD decay existed in maize, with $r^2 >$ 0.2 at the distance of 2.0 kb. Several discrete LD blocks, with $r^2 > 0.6$, were detected in maize. The rapid LD decay of *ZmFKF1b* in teosinte suggests a higher recombination rate in teosinte, and the relatively slower LD decay coupled with a large proportion of synonymous mutations in maize suggests the evolutionary conservation in maize during domestication and improvement.

Phylogenetic analyses of LFK genes

Phylogenetic trees based on 171 *LFK* proteins, 151 CDSs, 87 *ZmFKF1a* DNA sequences, and 120 *ZmFKF1b* DNA sequences from 87 maize inbred lines and 33 teosinte accessions were implemented using the best ProtTest model (Jones, Taylor, and Thornton [JTT] + I + G) and the best jmodeltest model (TVM + I + G), TrNef + I + G, and TrN + G, respectively. A congruent tree topology, with high bootstrap and posterior probability support values, based on 171 *LFK* family genes from 67 plant species was constructed using the NJ, MP, ML, and BI methods (Fig. 1 and Supplementary Figs. S4, S5 and S6). The sequence similarity and topology of phylogenetic trees suggest that the

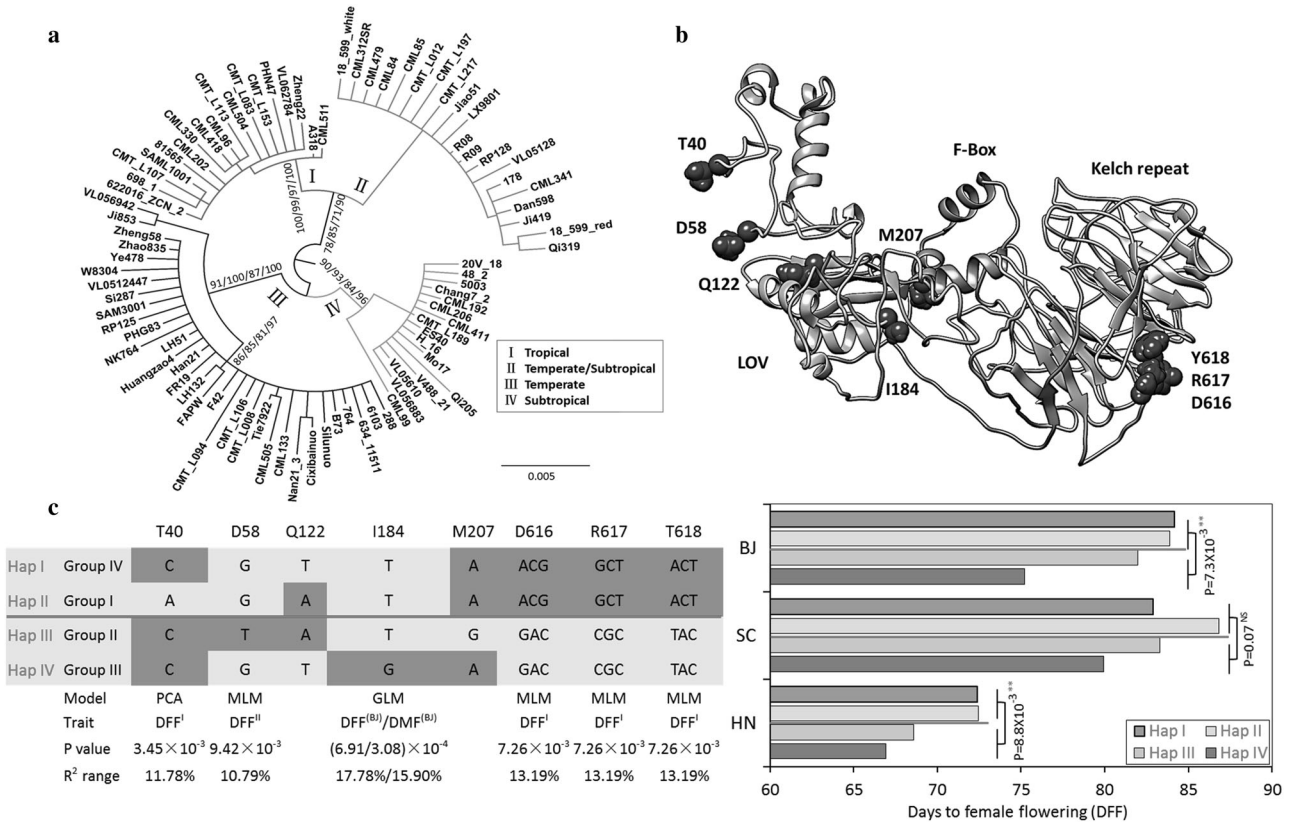

Fig. 2 Phylogenetic analysis and genetic effects of positive selection loci of *ZmFKF1b* among maize germplasms. **a** The phylogenetic tree of *ZmFKF1b* gene inferred from 87 maize inbred lines. Support values of the main branches are shown for nodes as neighbor-joining bootstrap/maximum parsimony bootstrap/maximum likelihood bootstrap/Bayesian inference posterior probability. **b** Residues under positive selection falling on the protein surface. Structures of a random coil in the N-terminal, LOV, F-box, and Kelch repeat domains are shown in red, orange, green, and light blue, respectively. Eight amino acid sites (T40, D58, Q122, I184, M207, D616, R617, and Y618) undergoing positive selection (blue balls) were mapped onto the outer surface of the protein structure. **c** Estimated effects of haplotypes consisting of eight positive selection loci on DFF among the 87 maize inbred lines. DMF, DFF, DFF ^{I}, and DFF ^{II} indicate days to male flowering and days to female flowering collected from BJ, and photoperiod sensitive index of DFF collected from long-day and short-day conditions (BJ-HN and SC-HN), respectively. *BJ* Beijing, *HN* Hainan, *SC* Sichuan

LFKs could be classified into two groups/clades, *ZTL/LKP2* and *FKF1*. Both clades showed a clear subdivision between monocots and dicots (angiosperms, Fig. 1 and Supplementary Figs. S4, S5 and S6), which is in agreement with plant taxonomy. *ZTL* showed a distant phylogenetic relationship between the group comprising bryophytes, ferns and gymnosperms and that of angiosperms (especially monocots). *M. polymorpha*, *Picea abies*, and *Selaginella moellendorffii* only had *ZTL* sequences, which clustered in the basal branching of the phylogenetic tree, suggesting an ancestral placement of the *ZTL* gene in the *LFK* family. *Amborella trichopoda* was the sole living member of the sister group to all other extant flowering plants (Albert, et al. 2013) and was considered as the most basal lineage in the clade of angiosperms (Soltis and Soltis 2013). In this study, the phylogenetic trees showed that both *AmbZTL* and *AmbFKF1* were located in the base of each group of *ZTL* and *FKF1*. The high conservation of gene sequence, confirmed by the relationships in the phylogenetic tree, suggests functional conservation of *LFK* genes. The structure of phylogenetic tree constructed using the 151 CDSs was consistent with that constructed using the *LFK* amino acid sequences (Supplementary Fig. S7), suggesting that the CDSs in the different species were also highly conservative in gene sequences and functions.

The evolutionary relationships of *ZmFKF1s* among maize germplasms were determined by reconstructing phylogenetic trees inferred from the 87 DNA sequences of *ZmFKF1a* and *ZmFKF1b* by using the NJ, MP, ML, and BI methods. For two *ZmFKF1s*, the topology trees obtained from different methods were highly consistent with a slight difference of bootstrap test values for each branch node (Fig. 2a, Supplementary Fig. S8). According to the topology tree of *ZmFKF1a*, 86 maize inbred lines were clustered into five groups with CML85 considered as an outgroup. Group I mainly consisted of temperate maize inbred lines, and the other four groups mainly consisted of tropical and subtropical maize inbred lines (Supplementary Fig. S8). The

Table 2 Comparison of the estimates of selective pattern among codons in PAML

Gene	Number[a]	Seq[b]	ω	2ΔlnL[c]	Positive selection sites[d]
LFK family	151	2577	0.058	0	None
ZTL subfamily	79	2352	0.055	0	None
FKF1 subfamily	60	2376	0.07	8.34*	I 90 (0.544)
LKP2 subfamily	12	1935	0.095	7.34*	S1 (0.833), S8 (0.716), F57 (0.729), A64 (0.678), E79 (0.506), Y106 (0.878), R115 (0.529), Y371 (0.959)
$ZmFKF1a_{m}$	87	1827	0.033	16.24***	P33 (0.701), I189 (0.645), V504 (0.990)
$ZmFKF1b_{m}$	87	1863	0.316	1191.4***	T40 (1.000), D58 (1.000), Q122 (1.000), I184 (1.000), M207 (1.000), D616 (1.000), R617 (1.000), Y618 (1.000)
$ZmFKF1b_{t}$	33	1878	0.400	77.58***	V30 (0.982), H122 (0.985), I184 (0.843), L601 (0.899), D615 (0.655), T616 (1.000)
$ZmFKF1b_{m\ and\ t}$	120	1878	0.328	290.68***	T49 (1.000), H122 (1.000), S184 (1.000), I207 (0.848), A615 (1.000), T616 (1.000)

m maize, *t* teosinte, ω d_N/d_S values of the class of codons in M0

[a]The CDS number performed for each analysis

[b]Codon site number based on the multiple sequence alignment of each dataset

[c]Twice the difference in the natural logs of the likelihoods (2ΔlnL) of the two models (M7–M8) being compared $*P < 0.05$, $***P < 0.001$, under the χ^2 test

[d]Codons assigned to the class evolving under positive selection in M8 with a posterior probability in following parentheses; codons underlined have the posterior probability of > 0.90 by Bayes Empirical Bayes estimates

phylogenetic tree inferred from 87 maize *ZmFKF1b* DNA sequences showed that the different maize inbred lines could be grouped into four major clusters (Fig. 2a). Clusters I and II comprised the tropical maize collected from CIMMYT and China, Cluster III mainly contained subtropical maize, and Cluster IV mainly included temperate maize from China and the USA. The evolutionary relationship between maize and its ancestor, teosinte, was further analyzed by reconstructing the phylogenetic trees based on 87 maize inbred lines and 33 teosinte accessions (Supplementary Fig. S9). Excluding a small number of teosinte (two *Z. nicaraguensis* accessions, one *Z. diploperennis* accession and four *Z. perennial* accessions), the other 26 *Z. mays* spp. *parviglumis* accessions were clustered with maize inbred lines.

Selective pressure and genetic effects of positive selection sites

The positive selection was determined by performing multiple sequence alignments generated from CDS datasets of the *LFK* family genes, three subfamilies, and *ZmFKF1a* and *ZmFKF1b* in maize and teosinte. The results of all NS site comparisons for each CDS homolog multiple sequence alignment are shown in Table S5; the results of the M7 vs. M8 comparisons, performed using the F3 × 4 model of codon preference, are summarized in Table 2. The LRT estimate suggests that model M3 remarkably fitted better for all the eight datasets than model M0 (Supplementary Table S5), indicating a heterogeneous selection among amino acid sites. The positive selection model (M8) was significantly more suitable than the null model (M7) in six of the tested groups ($P < 0.05$), including *FKF1* subfamily, *LKP2* subfamily, *ZmFKF1a* in maize, *ZmFKF1b* in maize, *ZmFKF1b* in teosinte, and *ZmFKF1b* in maize and teosinte (Table 2, Supplementary Table S5). In all datasets, the d_N/d_S (ω) values were significantly lower than 1 ($P < 0.001$), indicating a strong purifying selection on the coding portions (Table 2 and Supplementary Table S5). For *LFK* subfamilies, the ω values obtained under the discrete model (M3) were 0.006, 0.070, and 0.224 for the *ZTL* subfamily, and 0.014, 0.119, and 0.449 for the *FKF1* subfamily (Supplementary Table S5), indicating that purifying selection existed in almost all the amino acid sites. A small part of amino acid positions in the *LKP2* subfamily protein alignment had a ω value of > 1 ($\omega_2 = 1.305$, $p_2 = 0.155$; Supplementary Table S5), indicating that these codon sites might have undergone positive selection. Further Naive Empirical Bayes analysis (only used for M3 selection model) yielded marginal support for positive selection (Y371, $P = 0.993$; Supplementary Table S5). Nevertheless, few amino acids were found to undergo positive selection with ω values of 5.460, 27.605, 50.422, and 13.887 under the M3 model in the *ZmFKF1a* in maize, *ZmFKF1b* in maize, *ZmFKF1b* in teosinte, and *ZmFKF1b* in maize and teosinte groups, respectively. Moreover, the selection

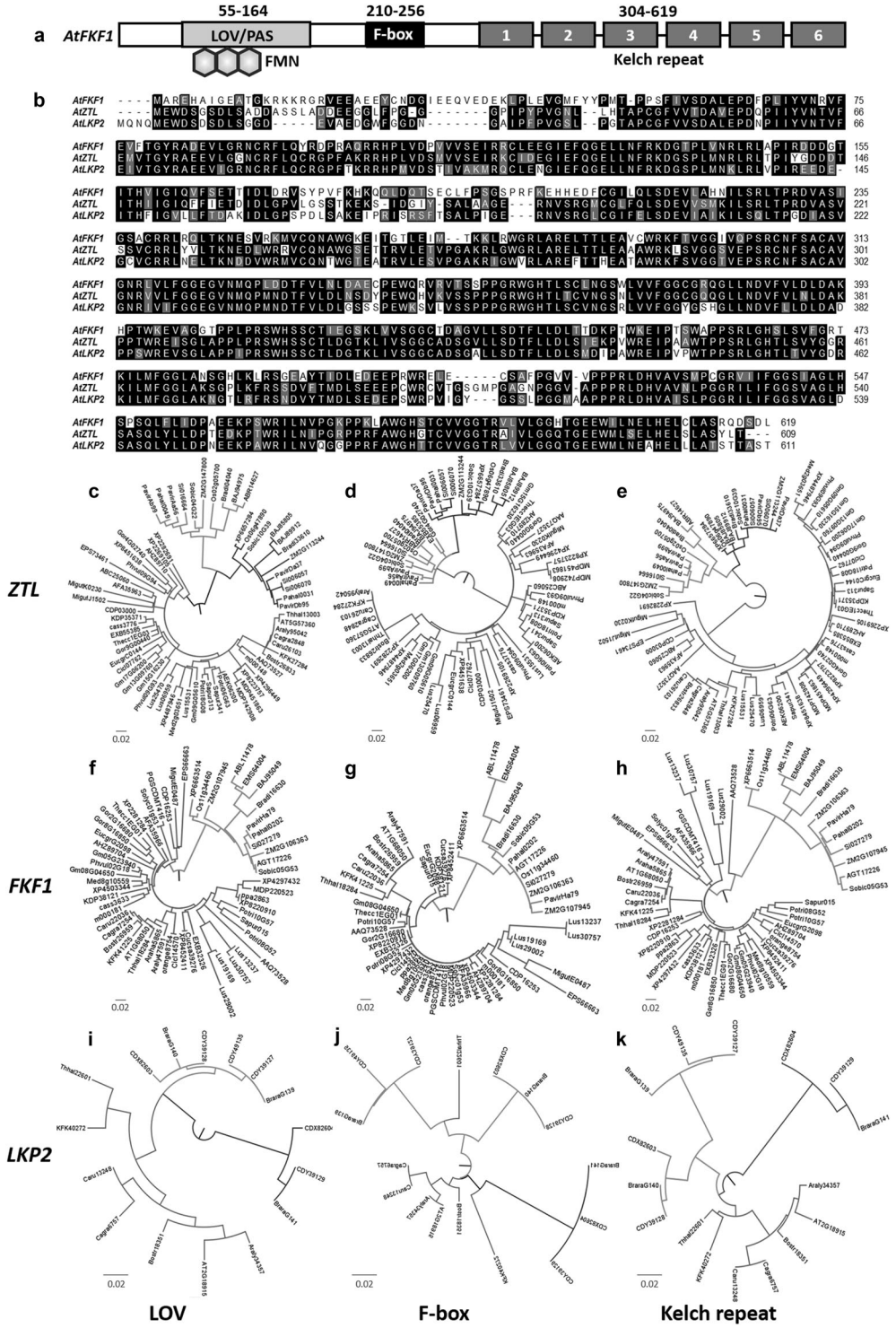

Fig. 3 Coevolution analysis of *LFK* subfamilies in the domains of LOV, F-box, and Kelch repeat. **a** Schematic representation of functional domains of *FKF1* in Arabidopsis (*At*). LOV (PAS)-domain bound to an FMN molecule functions as a blue-light-sensing domain. The LFK proteins possess one LOV domain at the N-terminus region, followed by an F-box domain and six Kelch repeat in the C-terminal region. The first and last amino acid numbers of each domain are indicated. **b** Sequence alignment of AtFKF1, AtZTL, and AtLKP2 amino acid sequences. Identical and similar amino acids are shaded in black and magenta, respectively. The numbers on the right indicate the protein length. **c–k** Phylogenetic trees based on pairwise sequence distances of LOV (PAS), F-box, and Kelch repeat domains in *ZTL*, *FKF1*, and *LKP2* subfamilies, respectively. The lines with red, blue, and green colors show three different clusters among species

pressure presents in the *ZmFKF1a* and *ZmFKF1b* genes during maize domestication and improvement was further validated by conducting Tajima's *D* and Fu and Li's tests on the different gene regions (Table 1). In *ZmFKF1a*, no significant selection signal was detected in any of the regions. However, in *ZmFKF1b*, exons 1 and 2 in both maize and teosinte showed significantly negative Tajima's *D* values, indicating population size expansion after a bottleneck or a selective sweep and/or purifying selection occurred in these regions, which is consistent with the mediate/ low level of nucleotide diversity in the two regions. Fu and Li's statistics also showed evidence for purifying selection in the entire gene sequence of maize and in the majority portion of teosinte (excluding non-significant positive selection in intron and 3′-UTR regions). These results suggest that despite the abundant nucleotide diversity of *ZmFKF1b*, its function was conserved from the ancestor (teosinte) to the domesticated (maize) lines. Both the neutrality tests of entire DNA sequence of *ZmFKF1b* in maize and teosinte and the positive selection analysis of *LFK* subfamilies showed that the *LFK* genes underwent a strong purifying selection over a long time.

Site-specific model analysis allowed the determination of positive selection acting at individual amino acid residues along the protein-coding sequences, although purifying selection was detected in 358 *LFK* gene sequences. A total of 18 positive selection sites ($\omega > 1$, $P > 0.9$) were identified under positive-selection model (M8), and their posterior probabilities are shown in Table 2. Among the identified homologs and orthologs of the *LFK* genes, only one (Y371) of the eight codons in *LKP2* subfamily was identified as evolving under positive selection, with a highly significant posterior probability ($P > 0.95$). Because of the high conservation of sequence and structure for *LFK* family members, particularly for each subfamily, the three-dimensional structure of *ZmFKF1b* was predicted based on the identified positive selection codons by using the protein sequence of maize line B73. For *ZmFKF1b* in maize, eight specific codons with the highest posterior probability ($P > 0.99$) were detected, which were scattered across the three-dimensional structure of the protein (Fig. 2b). Three (D58, Q122, and M207) of the eight amino acid sites identified as being positively selected were a part of functional domains, of which Q122 was involved in a beta-propeller fold in the LOV domain (Fig. 2b). Simultaneously, three and five codon sites in *ZmFKF1b* were identified as evolving under positive selection ($P > 0.90$) in teosinte and all maize germplasms, respectively. Although negative selection pressures were present on *LFKs* during plant evolution, positive selection signatures in specific codons were still found. The genetic effects of the positive selection sites and their relation with adaptive evolution were further investigated by determining the association between the positive selection loci and flowering-related traits in a diverse panel with 87 maize inbred lines. Six of them were significantly associated with DFF, DMF, and the photoperiod sensitivity index of DFF by using three statistical models of GLM, GLM + PCA, and MLM (PCA + K) at $P < 0.01$, each with explained phenotypic variation more than 10% (Fig. 2c). Further, four haplotypes were determined using eight positive selection variants, corresponding to the four groups of the phylogenetic tree (Figs. 2a, c). Hap I and II, corresponding to groups IV and I with tropical and subtropical maize germplasm, had larger DFF values, indicating stronger photoperiod sensitivity than Hap III and IV, corresponding to groups II and III with temperate maize germplasm in the three tested environments. Significant difference of DFF values between haplotypes (Hap I and II and Hap III and IV) were found in the two environments, Beijing and Hainan ($P < 0.01$), which is consistent with the structure of the phylogenetic tree and the characteristics of tropical and temperate maize germplasms (Figs. 2a, c). The three loci (D616, D617, and D618) at the complete linkage disequilibrium level could differentiate these haplotypes and tropical/subtropical and temperate maize inbred lines (Fig. 2c). These results indicated that the six positive selection sites in *ZmFKF1b* contribute to maize adaptive evolution through the regulation of flowering time from tropical to temperate environments during maize domestication.

Coevolution of domains among LFK proteins

Covariation of amino acid sites in protein domains occurs when two or more residues exert selective pressure on each other to maintain the structural and functional conservation of proteins. The coevolution among domain partners was investigated by inferring nine phylogenetic trees (Fig. 3) from three of the five independent alignments, including the three functional domains and two linking regions for each ortholog group of the *FKF1/ZTL/LKP2* subfamilies by using the NJ method and pairwise sequence distances. In the *ZTL* subfamily, sequences were classified into three clades within each of the LOV, F-box, and Kelch repeat domains, which were topologically similar, with a largest clade consisting of dicots, another one consisting of monocots, and its sister clade from an immediate common ancestor where duplication of *ZTL* occurred (Figs. 3c–e). In the *FKF1* subfamily, monocots and dicots were clearly divided into two distinct clades, which were roughly equidistant in each of the three domains (Figs. 3f–h). The phylogenetic relationships obtained for the three domains of *ZTL* and *FKF1* subfamilies exhibited a high degree of consensus, indicating that three domains were functionally related and had a clearly positive coevolution, especially between domains LOV and Kelch repeat. This

synchronization exposes an intensive positive coevolution of those two domains corresponding to the binding and ubiquitination functions of *FKF1*, which is an E3 ubiquitin ligase for targeted protein degradation. In the *LKP2* subfamily, a certain degree of positive coevolution among domains was revealed by phylogenetic trees (Figs. 3i–k). Considering that only cruciferous plants possess *LKP2* subfamily genes and a small sample size limited to the statistical power of evaluations, the observation of *LKP2* subfamily trees only provided clues to coevolution patterns.

The degree of coevolution among *FKF1*/*ZTL*/*LKP2* subfamilies was evaluated by calculating the correlation coefficients among different pairs of interacting partners in three subfamily groups. Significant correlation ($P < E^{-10}$), with a medium-high correlation coefficient, was found among the different pairs in each of the three subfamily groups (Table 3). For example, in the *FKF1* subfamily, the correlation coefficients were 0.78 between LOV and F-box, 0.86 between LOV and Kelch repeat, and 0.84 between F-box and Kelch repeat, with *z*-scores of 32.48, 34.91, and 35.19, respectively. Significant correlation was also detected in the domain-link pairs in *FKF1*, indicating a strong positive coevolution and relatively constant evolutionary rate among different species. In the *ZTL* subfamily, significant correlation was found between LOV and Kelch repeat domains, with an extremely high correlation coefficient of 0.93 and a z-score of 43.78. However, correlations between the pairs associated to link II were relatively weak: correlation coefficients were 0.69 between LOV domain and link II, 0.58 between F-box and link II, and 0.49 between the two links. Thus, the evidence of coevolution was not robust for link II in the *ZTL* subfamily, but coevolution and highly positive correlations were found for domain-domain pairs in the *ZTL* and *FKF1* subfamilies. In *LKP2*, correlation coefficient and z-scores among different pairs were relatively low. These results indicate that the different regions might have been subjected to different selection pressures and could have different extents of positive coevolution for the interacting pairs. Consistent coevolution evidence in domains of LOV, F-box, and Kelch repeat of the three subfamilies was revealed by both phylogenetic trees and correlation analysis.

Table 3 Correlation coefficients among domain-link pairs of *LFK* genes

Pairs	Group1 (*FKF1*)		Group2 (*ZTL*)		Group3 (*LKP2*)	
	r	Z-score	*r*	Z-score	*r*	Z-score
LOV and Link I	0.77	31.39	0.72	35.23	0.46	5.12
LOV and F-box	0.78	32.48	0.77	38.06	0.59	6.41
LOV and Link II	0.73	29.37	0.69	33.49	0.42	4.57
LOV and Kelch repeat	0.86	34.91	0.93	43.78	0.69	7.62
Link I and F-box	0.76	31.67	0.71	32.41	0.77	8.38
Link I and Link II	0.77	32.09	0.49	23.37	0.67	7.06
Link I and Kelch repeat	0.81	32.66	0.77	37.73	0.70	7.99
F-box and Link II	0.72	30.28	0.58	27.87	0.82	9.03
F-box and Kelch repeat	0.84	35.19	0.80	40.37	0.86	9.26
Link II and Kelch repeat	0.83	32.40	0.72	36.25	0.86	9.11

r Pearson's correlation coefficient

Z-score zero-mean normalization

Significant correlation was found among the different pairs at $P < 1E\text{-}10$

Functional divergence of LFK genes

Despite the high level of conservation in the *LFK* gene sequences, type-I (shifts in evolutionary rate) and type-II (altered physiochemical properties of amino acid residues) functional divergence were assessed based on the five clusters from the BI phylogenetic tree (Fig. 1): *ZTL* in dicots (Cluster A), *LKP2* in dicots (Cluster B), *ZTL* in monocots (Cluster C), *FKF1* in monocots (Cluster D), and *FKF1* in dicots (Cluster E) (Table 4). The results showed medium to large θ_I values for type-I functional divergence when clusters were compared (Table 4), with a significance level at $P < 0.001$ according to the LRT estimate (Table 4), which provided solid evidence of type-I functional divergence among the subfamilies of *LFKs* in land plants.

Interestingly, a pairwise comparison (clusters A and B) among groups that are more closely related to one another than to the others presented a higher type-I functional divergence coefficient ($\theta_I = 0.32$) than that in other pairwise clusters (A and C and D and E). The variation of functional divergence coefficient was consistent with a gene duplication evolutionary pattern: pairwise clusters (A and C and D and E) that have relatively smaller θ_I value are orthologs in monocots and dicots and probably perform the same function. Relatively larger θ_I value and more functional divergence are expected between clusters A and B, because cluster B is a paralog of cluster A. Other ortholog pairs also showed the same regular pattern (Table 4), suggesting that Type-I functional divergence played a major evolutionary role in the functional divergence of *LFKs* subfamilies among species. A strong evidence of type-II functional divergence was found for seven pairwise comparisons, except for the pairwise clusters A and C, C and D, and D and E that had non-significant θ_{II} values (Table 4). The results indicate a radical shift in amino acid properties among the different *LFK* subfamilies after gene duplication. The large θ_I value between pairwise clusters showed that the

Table 4 Statistics of the relative evolutionary rate between *LFK* subfamilies in land plants

		Cluster A	Cluster B	Cluster C	Cluster D	Cluster E
Type-I	Cluster A		$0.32 \pm 0.07(7)$	$0.22 \pm 0.05(0)$	$0.66 \pm 0.05(25)$	$0.55 \pm 0.04(22)$
	Cluster B	31.70^{***}		$0.57 \pm 0.09(5)$	$0.73 \pm 0.10(22)$	$0.52 \pm 0.08(7)$
	Cluster C	19.71^{***}	40.50^{***}		$0.49 \pm 0.07(3)$	$0.42 \pm 0.05(12)$
	Cluster D	159.73^{***}	54.36^{***}	53.90^{***}		$0.22 \pm 0.04(2)$
	Cluster E	152.96^{***}	45.14^{***}	79.80^{***}	37.20^{***}	
Type-II	Cluster A		$0.14 \pm 0.05(61)$	$0.01 \pm 0.06(0)$	$0.19 \pm 0.06(47)$	$0.23 \pm 0.06(93)$
	Cluster B	2.67^{**}		$0.13 \pm 0.05(33)$	$0.19 \pm 0.06(75)$	$0.18 \pm 0.06(78)$
	Cluster C	0.17	2.47^{**}		$0.07 \pm 0.07(1)$	$0.11 \pm 0.07(20)$
	Cluster D	3.17^{***}	3.36^{***}	1.11		$0.05 \pm 0.07(2)$
	Cluster E	3.63^{***}	3.03^{**}	1.71^{*}	0.71	

Clusters A, B, C, D, and E refer to *ZTL* in dicots cluster, *LKP2* in dicots cluster, *ZTL* in monocots cluster, *FKF1* in monocots cluster, and *FKF1* in dicots, respectively. Type-I functional divergence estimate ($\theta_I \pm$ standard error; upper right diagonal of Type-I matrix), and LRT values for significance (lower left diagonal of Type-I matrix) were estimated using DIVERGE 3.0. Type-II functional divergence estimate ($\theta_{II} \pm$ standard error) is shown in the upper right diagonal of Type-II matrix. The fold score (lower left diagonal of Type-II matrix) for functional divergence estimate was obtained as θ_{II}/θ_{SE} by using the *Z*-score test (normal distribution test). Numbers in the parentheses are amino acid sites with a posterior probability of > 0.90

$^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$

Type-I rather than type-II functional divergence was the main pattern in the functional divergence of *LFK* subfamilies.

The amino acid sites involved in the functional divergence of the *LFK* family were investigated by calculating posterior probability of divergence (*Qk*). In all, 356 amino acid sites ranging from 2 (clusters A and C) to 172 sites (clusters B and D) were identified at $Qk \geq 0.80$ in the 10 pairwise comparisons of the Type-I functional divergence (Supplementary Table S6). Of them, 105 amino acid sites were identified at $Qk \geq 0.90$ in nine pairwise comparisons, except for the pairwise cluster A and C, which is the homolog of the *ZTL* subfamily in dicots and monocots (Table 4). Only two critical amino acid sites were identified at $Qk \geq 0.90$ for pairwise cluster D and E, the two homologs of *FKF1* subfamily. However, 25, 22, 22, and 12 critical amino acid sites at $Qk \geq 0.90$ were identified for pairwise clusters A and D, A and E, B and D, and C and E, respectively, which are paralogs of the *LFK* family, suggesting a faster evolution rate of these critical amino acid positions (Table 4, Supplementary Table S6). For Type-II functional divergence, 455 amino acid sites were identified at $Qk \geq 0.80$ in nine pairwise comparisons, of which 410 critical amino acid sites were found at $Qk \geq 0.90$ and $P < 0.01$, except for pairwise cluster A and C. (Table 4, Supplementary Table S6). Interestingly, a total of 99 critical sites were shared between Type-I and Type-II functional divergence at $Qk \geq 0.80$, and 26 and 15 of them were identified at $Qk \geq 0.90$ and $Qk \geq 0.95$, which are involved in both evolutionary rate shift and physiochemical property alternatives. The critical amino acid sites associated with Type-I and Type-II functional divergence, mainly identified in the LOV, F-box, and Kelch repeat domains, accounted for 84.55% (301/356) and 82.86% (377/455) of the total critical sites, respectively (Supplementary Table S5). In addition, out of 301 and 377 critical amino acid sites in domains involved in Type-I and Type-II functional divergence, 184 (61.13%) and 249 (66.05%) sites were found in the Kelch repeat domain, accounting for 51.69% and 54.73% of the total critical sites, respectively, which indicates that a major functional divergence occurred in this functional domain, causing differences in binding targets and degradation.

A star-like topology was obtained using Type-I functional divergence as branch length by using DIVERGE 3.0 (Supplementary Fig. S10). A smaller branch length represents the function closer to the ancient function of a gene; thus, the function of *ZTL* in dicots (Cluster A) was the closest to the ancient function of the *LFKs*. Cluster B (*LKP2* in dicots) was further distanced to the ancient function than to any of the other clusters, and might cause a function derivative, which is consistent with its late appearance in the evolutionary history of plants inferred from the phylogenetic analysis and positive selection. The functional distance analysis of ten cluster pairs for Type-I functional divergence revealed that the prioritization of clusters to alter the functional constraint after gene duplication was in the order clusters B > D > E > C > A, which might help understanding the phylogenetic footprint. This result further shows that the *ZTL* subfamily in dicots mostly maintains the ancestral function of the *LFK* family, whereas the *LKP2* subfamily that only exists in dicots shows the highest functional distance. These results are consistent with the phylogenetic analyses performed for the whole gene family

and with the positive selection found in the *LKP2* subfamily.

Discussion

Evolutionary conservation of the LFK gene family in land plants

The comparative analysis and physical clustering of gene families may provide insight into the evolutionary mechanisms among species that shaped adaptation and diversity. The 171 *LFK*s identified in this study all belong to land plants, with strong purifying selection and high conservation, similar to housekeeping genes (Shapiro and Alm 2005). *LFK* homologs had identical exon sizes and similar gene structures, but various intron sizes across species, indicating differential regulation processes among species. Lower land plants such as *S. moellendorffii, P. abies* and *M. polymorpha* usually contain a single *LFK* gene, which is consistent with the findings of a previous study (Banks, et al. 2011), and the members of the *LFK* family present in these species belong to the *ZTL* subfamily. A previous study showed that the *LFK* member of *M. polymorpha* and the *GI* ortholog could control the growth-phase transition as in angiosperms, indicating a co-opted role of the GI-FKF1 complex from gametophyte to sporophyte generation during evolution (Kubota, et al. 2014), suggesting that the *LFKs* might be involved in plant reproduction. Despite the remarkable expansion of *LFK*s found in angiosperm genomes, the family member number is limited at a relatively low level (Supplementary Fig. S2). In other words, the *LFK* gene family has a conserved number of genes across various species, without showing any explosive radiation (Supplementary Table S2). The regulatory role of *AtFKF1*/*AtZTL*/*AtLKP2* in circadian rhythms in the photoperiodic flowering pathway provides a reasonable explanation: the *LFKs* were predicted to be involved in flowering and reproduction. Therefore, the *LFKs* involved in a gene network for the photoperiodic flowering pathway are a unique system in land plants. Based on phylogenetic trees, *LFKs* were divided into two categories,*FKF1*and *ZTL/LKP2*, which is consistent with previous results. (Boxall, et al. 2005; Taylor, et al. 2010). The high accordance between phylogenetic tree and taxonomic classification of plants suggests sequence conservation over the evolutionary history. Identification of multiple *LFK* components evolving under significant purifying selection also indicates systematic conservation of the blue-light-dependent protein degradation.

Lineage-specific chromosomal duplication events and repetitive DNA might have caused the uneven distribution of *LFK*s in various plants. The comparative genomic analysis has shown that ancient polyploidy events occurred in several lineages, with almost all flowering plants reflecting one or more ancient polyploidy events, even before the divergence of monocots and eudicots (Jiao, et al. 2011). The common ancestors of Poaceae, the grass family that includes important crop species such as rice, maize, wheat, and sorghum, shared polyploidization before the major cereals diverged (Paterson, et al. 2004). Maize recently underwent tetraploidization and showed an uneven ancient gene loss in two of the subgenomes (Schnable, et al. 2011). *Arabidopsis* genome has undergone two recent WGDs, which are shared only within the Brassicaceae lineage, and one triplication event (paleo-hexaploidy), which is probably shared by all core eudicots at the base of the angiosperm radiation right after the monocot-eudicot divergence (Simillion, et al. 2002). After WGD, some duplicated genes were lost or subjected to sub- and/or neofunctionalization because of their functional redundancy (Semon and Wolfe 2007). Ancient polyploidization, followed by the subsequent gene loss and diploidization, shaped the genomes of extant plants and played a significant role in plant diversification and evolution (Jiao, et al. 2011), particularly in species radiation and adaptation and functional capacity modulation (Jaillon, et al. 2007). Therefore, the unique *LKP2* subfamily found in cruciferous (Brassicaceae) plants might be the result of an ancient polyploid event, followed by diploidization occurring within the sister lineage of the *LFK* family, after the divergence of eudicots during species radiation. The divergence of the ancient *LFK* gene family would thus be accompanied by polyploidization events in monocots and eudicots.

The genetic loci of ZmFKF1b associated with maize flowering time

Evolutionary adaptation promotes species respond to environmental change rapidly and fitly (Hoffmann and Sgrò 2011; Lascoux, et al. 2016; Thomas, et al. 2017), and the genetic basis of local adaptation needs further exploration. The quantitative trait locus (QTL) mapping (Young 1996) and association mapping (Pritchard, et al. 2000) are two powerful genomic tools to dissect the genetic basis contributing to fitness-related traits. The flowering time, which contributes to survival and reproduction, is considered to be an important fitness-related trait of local adaptation in plants, and fitness variations were identified and estimated using the two genetic tools (Buckler, et al. 2009; Fournier-Level, et al. 2011; Leinonen, et al. 2013; Olsson and Ågren 2002; Verhoeven, et al. 2008). In addition, the population genetic analyses based on high-quality polymorphisms are effective to detect migration, selection, and other evolutionary forces in creating and maintaining local adaptation (Savolainen, et al. 2013; Yeaman and Whitlock 2011). In this study, the full-length sequences of *ZmFKF1b* were

cloned and resequenced in 87 maize inbred lines and 33 teosinte accessions. Abundant genetic variations were detected in *ZmFKF1b*, of maize and teosinte with a large proportion of synonymous mutations in the coding regions. Tajima's *D* and Fu and Li' *D* and *F* tests of *ZmFKF1b* in maize and teosinte from different locations showed that the *ZmFKF1b* underwent a strong purifying selection or selective sweep or showed population expansion after a bottleneck. The phylogenetic analysis of *ZmFKF1b* indicated that the 87 maize lines can be grouped into temperate and tropical/subtropical groups based on their geographic information. Few lines were not classified well into the phylogenetic tree, because it is difficult to classify different species or groups by only using the sequence of a single gene is difficult. In addition, not all maize inbred lines can be strictly defined as tropical, temperate, and subtropical, owing to the complex genetic breeding backgrounds among maize germplasms. By conducting genetic effect analysis between the identified positive selection loci and flowering phenology of maize germplasms, we identified six loci that accounted for fitness variation. Haplotypes consisting of the eight positive selection sites, which were effective to shape the classification of tropical and temperate maize groups, were found to be significantly associated with maize flowering time. These results indicate that the six polymorphisms enriched by the expression of *ZmFKF1b* were caused by parallel evolution of maize lines from different locations, and the genetic loci associated with maize flowering time might facilitate a flexible evolutionary state to counter stressful conditions or realize ecological opportunities arising from environment change.

Coevolution of LFK domains functioning as an E3 ubiquitin ligase

Molecular coevolving domains in a protein are functionally coupled by exerting selective pressure on each other, because of their close distance in three-dimensional structures. More strikingly, many coevolving positions are located at functionally important amino acid residues to maintain or refine the functional interaction between these entangled pairs (Yeang and Haussler 2005) and might influence the evolution of gene sequences and protein structures and functions. Many studies on the coordinated changes of amino acid residues showed that sequence covariation is a powerful detector of protein–protein interactions (Ramani and Marcotte 2003), receptor ligand-binding residues (Goh et al. 2000), and the three-dimensional structure (Atchley, et al. 2000) of single proteins. In this study, the phylogenetic tree constructed for each domain and the significantly positive correlation coefficients obtained for all interacting partners suggest a positive coevolution in the domains of the *LFK* family for long-term cooperation to target binding and degradation. In the E3 ubiquitination-dependent protein degradation pathway, the LOV domain plays a dual function—it absorbs blue light and interacts with GI after blue light absorption, whereas GI binds to CDF1 (a substrate of FKF1) for ubiquitination-dependent protein degradation (Imaizumi, et al. 2003; Sawa, et al. 2007). The F-box domain is a component of the SCF complex, which is the largest family of E3 ubiquitin ligases responsible for the turnover of many key regulatory proteins. The Kelch repeat domain forms a β-propeller structure and functions as a protein–protein interacting domain that binds substrates for ubiquitin-mediated protein degradation (Andrade, et al. 2001). Considering the close liaison and collaboration between the domains of *LFKs* during the E3 ubiquitin degradation of key regulatory proteins involved in the circadian clock of plants, this positive coevolution might have been the factor that ensures the success regulation of degradation process.

Functional divergence of the conserved LFK gene family

Generally, a large portion of eukaryotic genes are organized in gene families by undergoing the genome-wide or local chromosome duplication events during evolution. The generated paralogous genes might not maintain the original function, but allow acquiring a novel function or loss of function. In the classical classification of amino acid configurations, Types I and II functional divergence are the two major types to describe functional divergence after gene family proliferation. After gene duplication, one gene copy maintains its original function in different organisms, whereas the other mostly evolves toward functional divergence, which is highly correlated with the evolutionary rate variation among individual sites (Gu 2001). This correlation complements a fundamental rule that functionally more important sites have immediate relevance and have a high evolutionary conservation (a low evolution rate) of genes or DNA sequences (Gu 2001). Our results reflected the strong difference between the relative evolutionary rates of each site in protein function, as well as provided a prediction for the overall degree of functional divergence between protein subfamilies. Further, remarkable changes were found in the site-specific rates from *LKP2* in dicots to *FKF1* in monocots, which resulted in altered functional constraints between the two clusters. The critical amino acid residues might be directly responsible for the functional divergence observed in the *LFK* gene family.

The assumption of site independence indicates that, if the evolutionary innovation occurred mainly in the early stage of divergence, then the duplicated subfamilies within each cluster might simply be under purifying selection, despite their different functional constraints. Furthermore,

coevolution between sites is an important mechanism for functional divergence after gene duplication. Exploring the pattern of functional divergence among members of a gene family combined with positive selection analysis and site coevolution is helpful in better understanding the functional diversity at the genome level. Functional divergence of the *LFK* gene family was found to alter the substrates for ubiquitin-mediated protein binding and degradation during protein–protein interactions in the Kelch repeat domain. Our results illustrate the dialectical relationship between genetic diversity and functional conservation of *ZmFKF1b*, between the coevolution of domains and functional divergence of the conserved *LFKs* family, giving evolution clues of plant evolved from lower plants to higher species.

Acknowledgements This work was supported by the National Key Research and Development Program of China (2016YFD0100605.3); the National Natural Science Foundation of China (31522041 and 31561143014); the Foundation for the Author of National Excellent Doctoral Dissertation of China; the Fok Ying-Tong Education Foundation, China (20135103210002); and the Sichuan Youth Science and Technology Innovation team of China (2016TD0003).

Author contributions Y. Lu, L.L., and Y.W. conceived the experiments and designed the study. L.L., Y.W., J. Xiong, J. Xu, and F.W. performed the experiments and analyzed the data. Z.L., H.L., Q.T., S. Z. provided maize germplasm. L.L., Y. Liu, and Y. Lu drafted the manuscript, which was reviewed by all authors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

Albert VA, Barbazuk WB, Der JP et al. (2013) The Amborella genome and the evolution of flowering plants. Science 342:1241089

Andrade MA, González-Guzmán M, Serrano R et al. (2001) A combination of the F-box motif and kelch repeats defines a large Arabidopsis family of F-box proteins. Plant Mol Biol 46:603–614

Atchley WR, Wollenberg KR, Fitch WM et al. (2000) Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. Mol Biol Evol 17:164–178

Banks JA, Nishiyama T, Hasebe M et al. (2011) The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science 332:960–963

Barrett J, Fry B, Maller J et al. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Baudry A, Ito S, Song YH et al. (2010) F-box proteins FKF1 and LKP2 act in concert with ZEITLUPE to control Arabidopsis clock progression. Plant Cell 22:606–622

Boxall SF, Foster JM, Bohnert HJ et al. (2005) Conservation and divergence of circadian clock operation in a stress-inducible Crassulacean acid metabolism species reveals clock compensation against stress. Plant Physiol 137:969–982

Bradbury P, Zhang Z, Kroon D et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Buckler E, Holland J, Bradbury P et al. (2009) The genetic architecture of maize flowering time. Science 325:714–718

Chenna R, Sugawara H, Koike T et al. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497–3500

Christie JM, Salomon M, Nozue K. (1999) LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): binding sites for the chromophore flavin mononucleotide. Proc Natl Acad Sci 96(15): 8779–8783

Corbesier L, Vincent C, Jang S et al. (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science 316:1030–1033

Darriba D, Taboada G, Doallo R et al. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165

Demarsy E, Fankhauser C (2009) Higher plants use LOV to perceive blue light. Curr Opin Plant Biol 12:69–74

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Force A, Lynch M, Pickett F et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Fornara F, Panigrahi KC, Gissot L et al. (2009) Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Developmental Cell 17:75–86

Fournier-Level A, Korte A, Cooper M et al. (2011) A map of local adaptation in Arabidopsis thaliana. Science 334 (6052):86–89

Goh CS, Bogan AA, Joachimiak M et al. (2000) Co-evolution of proteins with their interaction partners. J Mol Biol 299:283-293

Greenham K, McClung CR (2015) Integrating circadian dynamics with physiological processes in plants. Nat Rev Genet 16:598–610

Gu X (2001) Maximum-likelihood approach for gene family evolution under functional divergence. Mol Biol Evol 18:453–464

Gu X, Zou Y, Su Z et al. (2013) An update of DIVERGE software for functional divergence analysis of protein family. Mol Biol Evol 30:1713–1719

Guindon S, Delsuc F, Dufayard JF (2009) Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol 537:113–137

Hartwell L, Hood L, Goldberg ML et al. (2011) Genetics: from Genes to Genomes. 4th ed. New York: McGraw Hill

Hoffmann A, Sgrò C (2011) Climate change and evolutionary adaptation. Nature 470:479–485

Hughes A. (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc London B 256:119–124

Hurst L, Smith N. (1998) The evolution of concerted evolution. Proc R Soc London B 265:121–127

Imaizumi T, Schultz TF, Harmon FG et al. (2005) FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. Science 309:293–297

Imaizumi T, Tran HG, Swartz TE et al. (2003) FKF1 is essential for photoperiodic-specific light signalling in *Arabidopsis*. Nature 426:302–306

Jaillon O, Aury JM, Noel B et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jiao Y, Wickett NJ, Ayyampalayam S et al. (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100

Kiba T, Henriques R, Sakakibara H et al. (2007) Targeted degradation of PSEUDO-RESPONSE REGULATOR5 by an SCFZTL complex regulates clock function and photomorphogenesis in *Arabidopsis thaliana*. Plant Cell 19:2516–2530

Kubota A, Kita S, Ishizaki K et al. (2014) Co-option of a photoperiodic growth-phase transition system during land plant evolution. Nat Commun 5:3668
Lascoux M, Glémin S, Savolainen O. (2016) Local Adaptation in Plants. eLS
Leinonen P, Remington D, Leppälä J et al. (2013) Genetic basis of local adaptation and flowering time variation in *Arabidopsis lyrata*. Mol Ecol 22:709–723
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452
Long M, Betran E, Thornton K et al. (2003) The origin of new genes: glimpses from the young and old. Nat Rev Genet 4:865–875
Lynch M, Conery J (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155
Más P, Kim WY, Somers DE et al. (2003) Targeted degradation of TOC1 by ZTL modulates circadian function in *Arabidopsis thaliana*. Nature 426:567–570
Magadum S, Banerjee U, Murugan P et al. (2013) Gene duplication as a major force in evolution. J Genet 92:155–161
Nei M, Li W. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci 76:5269-5273
Nei M, Rogozin I, Piontkivska H. (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. Proc Natl Acad Sci 97:10866-10871
Nelson DC, Lasswell J, Rogg LE et al. (2000) FKF1, a clock-controlled gene that regulates the transition to flowering in *Arabidopsis*. Cell 101:331–340
Nowak M, Boerlijst M, Cooke J et al. (1997) Evolution of genetic redundancy. Nature 388:167–171
Olsson K, Ågren J (2002) Latitudinal population differentiation in phenology, life history and flower morphology in the perennial herb Lythrum salicaria. J Evol Biol 15:983–996
Panchy N, Lehti-Shiu M, Shiu SH (2016) Evolution of gene duplication in plants. Plant Physiol 171:2294–2316
Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908
Patton EE, Willems AR, Tyers M (1998) Combinatorial control in ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. Trends Genet 14:236–243
Pettersen E, Goddard T, Huang C et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256
Press W, Teukolsky S, Vetterling W et al. (1988) Numerical Recipes in C. Cambridge UP, Cambridge
Pritchard J, Stephens M, Rosenberg N et al. (2000) Association mapping in structured populations. Am J Human Genet 67:170–181
Ramani AK, Marcotte EM (2003) Exploiting the co-evolution of interacting proteins to discover interaction specificity. J Mol Biol 327:273–284
Ronquist F, JP H (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574
Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. Nat Rev Genet 14:807–820
Sawa M, Nusinow DA, Kay SA et al. (2007) FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. Science 318:261–265
Schnable JC, Springer NM, Freeling M. (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci 108:4069-4074
Schultz TF, Kiyosue T, Yanovsky M (2001) A role for LKP2 in the circadian clock of Arabidopsis. The Plant Cell 13:2659–2670
Semon M, Wolfe KH (2007) Rearrangement rate following the whole-genome duplication in teleosts. Mol Biol Evol 24:860–867
Shapiro BJ, Alm EJ. (2005) Comparing patterns of natural selection across species using selective signatures. PLoS Genet pPreprint: e23. https://doi.org/10.1371/journal.pgen.0040023.eor
Shor E, Green RM (2016) The impact of domestication on the circadian clock. Trends Plant Sci 21:281–283
Simillion C, Vandepoele K, Van Montagu MC et al. (2002) The hidden duplication past of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 99:13627–13632
Soltis PS, Soltis DE. (2013) Angiosperm phylogeny: a framework for studies of genome evolution[M]//Plant Genome Diversity Volume 2. Springer Vienna: 1-11. https://link.springer.com/chapter/10.1007/978-3-7091-1160-4_1
Somers DE, Kim W-Y, Geng R (2004) The F-box protein ZEITLUPE confers dosage-dependent control on the circadian clock, photomorphogenesis, and flowering time. The Plant Cell 16:769–782
Somers DE, Schultz TF, Milnamow M (2000) ZEITLUPE encodes a novel clock-associated PAS protein from Arabidopsis. Cell 101:319–329
Song YH, Ito S, Imaizumi T (2013) Flowering time regulation: photoperiod- and temperature-sensing in leaves. Trends Plant Sci 18:575–583
Song YH, Shim JS, Kinmonth-Schultz HA et al. (2015) Photoperiodic flowering: time measurement mechanisms in leaves. Annu Rev Plant Biol 66:441–464
Song YH, Smith RW, To BJ et al. (2012) FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. Science 336:1045–1049
Strable J, Scanlon MJ (2009) Maize (*Zea mays*): a model organism for basic and applied research in plant biology. Cold Spring Harbor Protocols 2009(10):emo132. pdb
Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612
Swofford D. (2002) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates: Sunderland, MA, USA
Takahashi N, Kuroda H, Kuromori T et al. (2004) Expression and interaction analysis of Arabidopsis Skp1-related genes. Plant Cell Physiol 45:83–91
Takase T, Nishiyama Y, Tanihigashi H et al. (2011) LOV KELCH PROTEIN2 and ZEITLUPE repress Arabidopsis photoperiodic flowering under non-inductive conditions, dependent on FLAVIN-BINDING KELCH REPEAT F-BOX1. Plant J 67:608–621
Tamura K, Peterson D, Peterson N (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739
Taylor A, Massiah AJ, Thomas B (2010) Conservation of *Arabidopsis thaliana* photoperiodic flowering time genes in onion (Allium cepa L.). Plant Cell Physiol 51:1638–1647
Taylor BL, Zhulin IB (1999) PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev 63:479–506
R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/
Thomas L, Kennington W, Evans R et al. (2017) Restricted gene flow and local adaptation highlight the vulnerability of high-latitude reefs to rapid environmental change. Global Change Biol 23:2197–2205

Valverde F, Mouradov A, Soppe W et al. (2004) Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science 303:1003–1006

Verhoeven K, Poorter H, Nevo E et al. (2008) Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations. Mol Ecol 17:3416–3424

Watterson G (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 43: W174–181

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yeaman S, Whitlock M (2011) The genetic architecture of adaptation under migration–selection balance. Evolution 65:1897–1911

Yeang C-H, Haussler D. (2005) Detecting the coevolution in and among protein domains. PLoS Comput Biol Preprint: e211. https://doi.org/10.1371/journal.pcbi.0030211.eor

Young ND (1996) QTL mapping and quantitative disease resistance in plants. Ann Rev Phytopathol 34(1):479–501

Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18:292–298

Zhang J, Rosenberg H, Nei M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci 95:3708-3713

Zhang J, Zhang Y, Rosenberg H (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat Genet 30:411–415

Zoltowski BD, Imaizumi T (2014) Structure and function of the ZTL/FKF1/LKP2 group proteins in arabidopsis. Enzymes 35:213–239