Discovery of *UPSTREAM of FLOWERING LOCUS C* (*UFC*) and *FLOWERING LOCUS C EXPRESSOR* (*FLX*) in Gladiolus ×Hybridus, G. Dalenii

Jaser A. Aljaser
University of Minnesota
Neil Anderson (ander044@umn.edu)
University of Minnesota
Andrzej Noyszewski
University of Minnesota

Research Article

Keywords: Flowering, Geophyte, Iridaceae, FLC, FRI

Posted Date: November 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1044399/v1

Abstract

Background.

Gladiolus is a geophytic floricultural crop cultivated for cut flower and garden ornamental uses. Monocotyledonous flower crops have few, if any, flowering genes identified. Ornamental geophytes such as gladiolus, lily, tulip and daffodil are examples of floral crops that are currently being investigated to understand the flowering pathway. While the flower genes and environmental / hormonal factors leading to flowering are established in *Arabidopsis*, the lack of identified flowering genes in economically important ornamental geophytic crops, such as gladiolus, is critical to further genetic research. Thus, the importance of such an ornamental crop that relies on flowers (flowering) for economic purposes encourages researchers to discover the flowering genes to breed vigorous, flowering cultivars. The understanding of the flowering mechanisms in the flowering pathway is also of paramount importance.

Results.

Herein we show the discovery of *UPSTREAM OF FLOWERING LOCUS C* (*UFC*) and *FLOWERING LOCUS C EXPRESSOR* (*FLX*) genes in *Gladiolus ×hybridus* and *G. dalenii*. The *UFC* gene is adjacent to *FLOWERING LOCUS C* (*FLC*) which is a floral repressor in many temperate species. The *FLX* gene upregulates *FRIGIDA* (*FRI*) which upregulates *FLC* expression. Discovery of both genes is a step forward in finding the *FLC* gene in gladiolus, provided they are linked. Seventeen gladiolus genotypes, consisting of early flowering and commercial cultivars, were discovered to possess the *UFC* gene, consisting of four exons in two allelic forms. The sequenced *UFC* gene, when translated into its amino acid sequence and set in pair-alignment to other species, has up to 57% in amino acid identity to *Musa acuminata*. The *UFC* protein ranges in identity with pair-alignment to other monocot species, also with 57% amino acid identity to *M. acuminata*. The *FLX* gene in gladiolus has 3/5 (60%) exons in common relative to *Ananas comosus*, i.e. lacking 2 exons and a partially complete gene sequence; the pair-alignment of the three exons shows an overall ~65% identity of *FLX* to *A. comosus*. The *UFC* protein consists of a conserved domain, DUF966, which is higher in identity (86%) and pair-alignment with *Elaeis guineensis*.

Conclusions.

The two newly-discovered genes in gladiolus, *UFC* and *FLX*, provide insight to further our understanding of the flowering mechanism, flowering pathway genes, and vernalization response in ornamental geophytes. This knowledge will be valuable for gladiolus breeders and geneticists to finding the *FLC* gene, identify segregating seedlings for both *UFC* and *FLX*, and aid in marker assisted selection for flowering gene expression.

Background

In angiosperms or flowering plants, “Florogenesis” is the transitioning process in a plant’s apical meristem from vegetative tissue to reproductive organs or flowers [1]. This transition is governed by flowering genes in which expression is influenced by factors such as vernalization, photoperiod, gibberellins, an autonomous pathway and ambient temperature [2]. In *Arabidopsis*, several flowering genes have been discovered which are involved in flowering and act as floral integrators. These flowering genes include *FT*, *SOC1*, *CO*, *VRN1*, PPD, FCA, *FLD*, and *FLK* [3–4]. These floral integrator genes specifically upregulate flowering by promoting transition from vegetative to flowering or repressing floral repressor genes. Repressor genes act as repressors of floral integrator genes and upregulate the expression of repressors, including genes such as *FLC*, *FRI*, *FLX*, *VRN2*, and SVP [3, 5–6].

FLOWERING LOCUS C (*FLC)* and *FLC-like* are floral repressors found in many dicotyledon plants, such as *Malus* [7], *Rosa* [8], *Coffea* [9] and *Brassica* [10]. The *FLC* gene is regulated by temperature changes throughout the year, both in annuals and perennials. In summer, *FLC* expression is upregulated through *FRIGIDA* (*FRI*) by binding the *FLC* promoter through the DNA-binding protein *SUPPRESSOR OF FRIGIDA4 SUF4* [11]. In addition, *FRI* expression is upregulated by the *FLOWERING LOCUS C EXPRESSOR* (*FLX*); both *SUF4* and *FLX* are in the *FRI*-specific pathway [12]. In winter, *FLC* is down-regulated through a process of vernalization as prolonged exposure of low temperature in winter in the meristem gradually reduces the expression of *FLC* [13]. In addition to the vernalization pathway, the autonomous pathway reduces the expression of *FLC* both in the meristem and leaves [13]. Gradual reduction of *FLC* allows *FLOWERING LOCUS T* (*FT*) to be expressed in the leaves and transported through phloem to reach to the meristematic tissue to stimulate the MADS box genes, thereby inducing flowering in *Arabidopsis* [14].

In wheat and barley, the flowering pathway is regulated by photoperiod, vernalization and the circadian clock [15]. *Vernalization* gene-2 (*VRN2*) is a dominant repressor or inhibitor of flowering in winter wheat (*Triticum aestivum;* a monocot grass) that is down-regulated by vernalization (a cold period; winter) [15]. Subsequently, the floral integrator leads to flowering in winter wheat while spring wheat doesn’t require vernalization due to a non-functional *VRN2* gene. However, vernalization is a facultative stimulus for earlier flowering in spring wheat [16]. In contract, maize (*Zea mays*) and rice (*Oryza sativa*) rely on plant age to build up sufficient energy requirements in order to transition to flowering through epigenetic action of *miR172* [17]. In monocot geophytes (defined as herbaceous perennial plants with underground storage organs, e.g. bulbs, corms, tubers, etc., that promote winter survival), such as *Gladiolus*, *Lilium*, *Tulipa*, *Narcissus* and *Crocus*, the flowering pathway is poorly understood. Factors of plant growth influencing flowering in commercial geophytes for commercial production are well known [18] and include photoperiod, light intensity, the autonomous pathway, gibberellins, ambient and cool temperatures (vernalization) [1]. A clearly delineated genetic pathway for monocots (ornamental and otherwise) is still in the early stages of the discovery and characterization for many floricultural crops, such as gladiolus. In contrast, the *Arabidopsis* model is readily applicable to temperate dicotyledon plants [19] and may be only partially useful for monocots. Only a few flowering genes have been discovered in monocot ornamental geophytes [1], such as *FT-like* in *Allium cepa* [20–21], *FT* in *Narcissus* [22], *NLF* in *Narcissus* [23], *LFY* in *Allium sativum* [24–25] and in *Lilium* [26]. Recently, many flowering genes have been discovered in *Lilium* × *formolongi* (*FT*, *CO-like*, *AP2*, *GA1*, *SOC1*) and/or proposed for *L. formosanum*

(*VER1, VER2*) [27–28]. The discovery of flowering genes in geophytes serve as valuable resources to model flowering pathway(s) therein.

Geophytes such as *Gladiolus, Lilium*, *Tulipa*, *Narcissus* and *Crocus* are floricultural crops with ornamental value wherein flowering is essential to maintain the marketing value for these crops. *Gladiolus* × *hybridus* Rodigas, commonly known as gladiolus(-i), is commercially cultivated as a cut flower and as garden or landscape plants. Gladioli are geophytic plants with underground modified stem structures known as corms, producing cormels as a means of vegetative propagation [29]. Flower initiation and development are crucial steps for its success as a cut flower. Therefore, understanding the flowering pathway is vital for genetics and breeding to improve the floral market value.

Gladiolus has a genome size of 1100 Mbp, although it is unclear whether this is for haploid or diploid and the species is unknown [30]. The genome weight for gladiolus was recently measured in *G. communis* as 0.67-0.68 pg for monoploid G.s. (1Cx, pg) and in *G. italicus* at 0.61 pg for monoploid G.s. (1Cx, pg) [31]. The limited knowledge of the gladiolus genome is also reflected in lack of knowing gladiolus flowering genes, there are no flowering genes discovered in gladiolus except for the gibberellin receptor gene *GID1a* in gladiolus [31]. However, the relationship of *GID1a* with flowering has not been established. In *Arabidopsis*, gibberellin binds to the gibberellin receptor forming *GID1* complex that binds to *DELLA*, causing its degradation, thereby enabling *SOC1* and *LFY* to be upregulated, leading to flowering [32–33].

Understanding the flowering pathway and gene expression is important for efficient selective breeding of gladiolus for rapid generation cycling (RGC) or early flowering types that flower in ≤1 year from seed [43]. An important flowering gene is *FLC*, a major flowering repressor found in *Arabidopsis* and many dicot species; *FLC* plays vital role in the control of flower initiation [13]. Gladiolus is a monocot genus (Iridaceae) with both summer and winter flowering species; *FLC* has not been identified herein. It had been hypothesized that there is no *FLC* gene in monocot species until the *FLC* homologue was discovered in some cereal crops, such as *Triticum aestivum* [34], *Hordeum vulgare* [35] and *Brachypodium distachyon* [36]. These *FLC* homologue studies did not discover *FRI* genes, which upregulate *FLC* expression in *Arabidopsis thaliana* [11]. A hypothesis to test would be that some monocots do not possess the flowering repressor *FLC* gene and rely on alternative gene(s) to acts as a repressor(s) *miR172* through epigenetic in plant age-dependent of *Zea* and *Oryza* [17]. Additionally it remains unknown whether there is *FLC*-dependent pathway in all monocots.

In *Arabidopsis, FLC* is located between two flanking genes, *UPSTREAM OF FLOWERING LOCUS C* (*UFC*; located 4.7 Kb upstream of *FLC*) and *DOWNSTREAM OF FLOWERING LOCUS C* (*DFC*; found 6.9 Kb downstream of *FLC*) [37]. The *UFC* gene expression is repressed by vernalization, independent of *FLC* repression due to vernalization [37]. Thus, both *FLC* and *UFC* are repressed by vernalization, yet are not dependent on each other for expression; the suppression is through chromatin modification in an epigenetic manner [37]. The *VRN1* gene is expressed with vernalization and acts as a floral integrator whereas *UFC* is repressed and required by *VRN1* expression dependently [38]. The potential role for *UFC* in flowering has yet to be discovered and it may not involve flowering at all since vernalization only

represses *UFC* in seeds while *DFC* is repressed by vernalization of the plant [38]. Insertion of the *NPTII* gene between the *UFC* and *FLC* region confirmed *NPTII* response to cold as the whole cluster region of *FLC* responded to cold [37]. In the *UFC* protein of *A. thaliana*, a conserved domain, DUF966, has 92 amino acids, although its function is still unknown [39]. This lack of knowledge in DUF966 function creates a challenge to identify the function of *UFC* protein. However, a recent study shows the role of *UFC* in *A. thaliana*, with the gene *SOK2*, which appears to have a role in embryogenesis, root initiation, growth and branching of the primary and lateral roots [39]. The conserved domain DUF966 is reported to be present in the *OsDSR* gene family of *Oryza sativa* [40]. The promoter for genes that contain DUF966 have a defense-stress response to pathogens, salicylic acid, jasmonate, drought or salinity [41]. The *ZmAuxRP1* gene which promotes the biosynthesis of indole-3-acetic acid (IAA) to increase resistance against pathogens in *Zea mays* [41]. Thus, all the genes that contain DUF966 vary in function but all of them are triggered by environmental or stress stimuli to overcome an undesirable change influencing plant growth and development.

The *FLX* gene encodes a putative leucine zipper domain which is required for *FRI*-mediated activation of *FLC* in *Arabidopsis* [42] (Fig. 1). Up-regulating *FLC* occurs in winter annual *Arabidopsis* [33], while late flowering phenotypes exhibit strong expression of *FLX* which indicates a role of *FLX* in the suppression of flowering [42]. Several genes have been discovered in the *FLX* gene family, e.g. *FLX-LIKE1* (*FLL1*), *FLX-LIKE2* (*FLL2*), *FLX-LIKE3* (*FLL3*), *FLX-LIKE4* (*FLL4*) [11, 33]. *FLX* and *FLL4* are the most crucial genes in flowering time control in *Arabidopsis* [33].

In order to test whether *FLC* is present in gladiolus, the adjacent gene (*UFC*) will also be probed, along with FLX which is part of the *FLC*-dependent mechanism. Therefore, the objective of this study is to identify whether *UFC* and *FLX* genes occur in the genetically diverse gladiolus germplasm of the University of Minnesota Gladiolus Breeding Program. The null hypotheses tested are: H_o1 = There is no difference among gladiolus genotypes in the existence of the *UFC* gene; H_o2 = There is no difference among gladiolus genotypes in the presence of *FLX*.

Results

The investigated genotypes of *Gladiolus* resulted from the gladiolus breeding program for the selection of rapid generation cycling, which includes 14 breeding genotypes and 3 commercial cultivars (‘Vista’, ‘Glamini®’ and ‘Carolina Primrose’; Table 1). Rapid Generation Cycling (RGC) in gladiolus is the ability of flowering in the first year from seed as an annualized perennial [43–45]. Classically, seed-propagated gladiolus have 3-5 years as juvenile, non-flowering (vegetative) seedlings before a phase change into flowering (reproductive) adults [44]. The University of Minnesota Gladiolus breeding program developed such gladiolus genotypes with a reduced juvenility period and phase change to flowering in ≤1 year from seed [43–45].

The designed probe for the *UFC* gene in *Gladiolus* (RAPiD genomics® LLC; Gainesville, FL; http://rapid-genomics.com/home/) resulted in a total of 433 sequences; 161/433 sequences had read hits of the *UFC*

gene with various percentages of coverage. Of these, 34 selected sequences were chosen for this study, based on the largest length with two sequences per genotype due to presence of two alleles per genotype. These sequences represent the genomic sequence of *UFC* in gladiolus. The sequences were analyzed for gene prediction using the HMM-based gene structure prediction of FGENESH using *A. thaliana* (Generic) as the specific gene-finding parameter since the gene prediction is optimized for *A. thaliana*. Results confirmed the presence of *UFC* exons of a protein and coding sequence. The *UFC* coding sequences of each gladiolus genotype were analyzed in Genoeious® by pair-alignment with its genomic sequence to determine each exon. After the pair-alignment, the Coding sequence was then translated and aligned in multi-alignment process using MUSCLE alignment in the neighbor joining clustering method and CLUSTALW sequencing scheme with *UFC* proteins of other monocot and dicot species: *Ananas comosus*, *Musa acuminata*, *Elaeis guineensis*, *Asparagus officinalis*, *Arabidopsis thaliana* and *Glycine max*. The *UFC* gene in *Gladiolus* × *hybridus* was assigned to *GhUFC* as a label while genotype 14, *G. dalenii* 'Carolina Primrose', is assigned to *GdUFC*.

There are two alleles of the *UFC* gene found in gladiolus, designated as A and B. Thus, the genes are designated as *GhUFC-A* and *GhUFC-B*. The median number of amino acids of *GhUFC-A* protein is 420 amino acids across all gladiolus genotypes, with some genotypes having less than 420 amino acids. One genotype has 451 amino acids which could be due to an insertion, while *GdUFC-A* also has 420 amino acids. The second allele, *GhUFC-B* protein has a range of 375 to 410 amino acids, *GdUFC-B* has 286 amino acids with an incomplete protein, missing many amino acids and a stop codon (Figs. 2, 3). Gladiolus genotypes 1 (*G.* × *hybridus* 21213; Table 1) and 15 (*Gladiolus* × *hybridus* 'Beatrice'; Table 1) were selected for pair-alignment with the species of comparison (Table 2). Similarities of identity protein sequences and the percentage of *GhUFC-A* and *GhUFC-B* occur at range of ~30–57% across all species (Table 2). The intron-exon organization of *GhUFC-A* in *Gladiolus* × *hybridus* 'Beatrice' (genotype 15; Table 1) is similar to *Elaeis guineensis* and *Asparagus officinalis* in term of exons splicing (Fig. 4) while *GhUFC-B* in *G.* × *hybridus* 21213 (genotype 1; Table 1) has some similarity with *Ananas comosus* exons splicing. The remaining genotypes fall into these two configurations of the exon; the configuration shows the location of the conserved domain for *UFC* gene DUF966, which is found in *Arabidopsis* and other selected species of comparison. The DUF966 domain has the 92 and 93 amino acids of *Ananas comosus*. The multi alignment of the *UFC* protein conserved domain is overall conserved across species, although it is polymorphic (Fig. 5). With the high identity matching in *Gladiolus* genotypes 1 and 15 exhibit a range of ~65% to ~86% across all investigated species for the DUF966 domain of the *UFC* protein (Table 3). The *GhUFC-A* allele has a high identity across gladiolus genotypes (Fig. 6), with the polymorphic exception of *G.* × *hybridus* 2231 (genotype 3; Table 1) and *G.* × *hybridus* 3931 (genotype 7). *GhUFC-B* is also conserved and identical in sequence with *G.* × *hybridus* 20732 (genotype 12), due to missing amino acids.

The *FLX* gene was identified in *Gladiolus* in 12/17 genotypes; 11 *Gladiolus* × *hybridus* genotypes have *GhFLX* whereas *GdFLX* is identified in genotype 14, *G. dalenii* 'Carolina Primrose' (Table 4). The range of amino acid proteins are from 146 to 254 amino acids missing the stop codon. Three genotype sequences in *G.* × *hybridus* 2231, 3923, and 'Glamini'® (genotypes 3, 6 and 16, respectively; Table 1) have the longest amino acid chain. *FLX* is present in many species; in *Arabidopsis* it belongs to *FLX* gene family, *FLX*,

FLOWERING LOCUS C EXPRESSOR-LIKE 1 (*FLL1*), (*FLL2*), (*FLL3*) and (*FLL4*) [33]. Based on the pair-alignment, *GhFLX* and *GdFLX* matches *FLL1* with as high as 50% in amino acid identity (Table 4). The similarities of sequences in identity of gladiolus genotypes range from ~26% to ~65% across all investigated species of the entire *FLX* protein; the highest identity is in *Ananas comosus* match with ~65%. The multi-alignment for all *FLX* indicates that the tested gladiolus genotypes with the longest amino acid sequences lack exons (Fig. 7). A pair-alignment test with *Ananas comosus* – *FLX* reveals that *GhFLX G.* ×*hybridus* 'Glamini'® (genotype 16; Table 1) lacks two exons and a stop codon (Fig. 8).

Discussion

The presence of a putative *UFC* gene in gladiolus is confirmed with two alleles, *GhUFC-A* and *GhUFC-B*. It is highly possible that allelic number is due to tetraploidy in cultivated gladioli ($2n = 4\times = 60$) [46–48]; ploidy levels of *G. dalenii* have not been reported [46]. The cultivated *Gladiolus* ×*hybridus* are interspecific hybrids [49]. Therefore, the presence of different alleles would be expected in the diverse array of genotypes included in this study: 13 are from University of Minnesota gladiolus breeding program (interspecific hybrids) and four are commercial gladiolus cultivars with unknown ancestry and relatedness (Table 1). The *GhUFC-A* gene has ~50% identity with *Musa acuminata*, *Elaeis guineensis* and *Asparagus officinalis* (Table 2). The splicing of *Elaeis guineensis* and *Asparagus officinalis* exons is similar to *GhUFC-A* (Fig. 4). *GhUFC-B* gene splicing in the first 4 exons is similar to *Ananas comosus, Arabidopsis thaliana* and *Glycine max UFC* gene splicing (Fig. 4). These divergences in splicing of the *UFC* gene support *UFC* presence in gladiolus with two alleles [46]. Further tests should be done to identify whether or not *UFC* is also present in diploid gladiolus species, such as *G. murielae*, *G. tristis* and *G. carneus* since these three species are diploids [50].

The *UFC* gene is responsive to vernalization by lowering expression alongside *FLC* and *DFC* in *Arabidopsis thaliana*, as all these genes are in the cluster of vernalization stimulus region [37]. *FLC* is a floral repressor, the overexpression of *FLC* results in a delay in flowering [51], while overexpression of *UFC* does not result in the altering flowering time [37]. Thus, *UFC* is adjacent to *FLC*, both are repressed by vernalization, yet *UFC* does not show any influence in flowering time. This was observed herein since the genotypes in this study include both RGC-1, which are early flowering gladiolus able to reach flowering in the first year from seed and the classical later-flowering gladiolus which requires 3-5+ years to flower from seed. The multi-alignment of *UFC* protein in RGC-1 genotypes does not show any difference from non-RGC-1 genotypes. Thus, the *UFC* gene most likely isn't involved in flowering, at least directly which was proven in a *UFC* study in *A. thaliana* [38]. The main differences (Fig. 6) represent the differences between alleles of *UFC-A* and *UFC-B*, regardless of the gladiolus genotypes tested herein (Fig. 6).

During the winter cold period, vernalization suppress both *FLC* and *UFC* expression [37], which allows floral gene integrators to promote flowering. The hypothesis would be that, after vernalization and flowering, the *UFC* protein involvement is in embryogenesis and root initiation such that growth and branching occur in the spring season (since it doesn't occur in the winter season). This could explain how

FLC and *UFC* are both negatively responsive to vernalization stimuli in the cluster genes area, while upstream of *UFC* is not responsive to vernalization [37].

The identification of *FLX* in gladiolus raises the question whether gladiolus follows the *Arabidopsis* dicot model of the flowering pathway. In the winter annual, *A. thaliana*, flowering is promoted after vernalization, which suppresses the floral suppressor *FLC* that is upregulated by *FRI* through activation of *FRI* complex of (*FRI*, *FRL1*, *FRL2*, *FES1* and *SUF4*) proteins in addition to *FLX* protein. *FLX* was proven to provide transcriptional activity for the *FRI* complex [11]. A loss of function of *FLX* in *A. thaliana* resulted in early flowering phenotypes [42], which indicates the clear role of *FLX* in flowering. The role of *FLX* in gladiolus has not been tested, particularly in RGC-1 genotypes and pedigrees; thus, *FLX* upregulation of *FRI* in gladiolus would be a rational approach. However, *FRI* was not detected in gladiolus, using the primer design of *A. thaliana FRI* (At4g00650) because the *FRI* gene has not been previously detected in any monocotyledon species. Thus, the primer used is from *A. thaliana*, the test did not detect *FRI* gene in all of the 17 gladiolus genotypes without finding a single match [45]. Our results with gladiolus provide additional data in support of these previous results for *FRI* in monocots. In addition, *VRN2*, the repressor of flowering in cereals and *A. thaliana* was not detected in gladiolus, using the primer design of *Triticum monococum*, *T. durum* and *Hordeum vulgare* ([45]; Appendices A1, A2 and A3). This is not conclusive evidence as the genetic similarities between *A. thaliana* and *Gladiolus* are low, given that *GhUFC-A* is ~32% and *GhFLX* is 50% identical to *A. thaliana* genes. Therefore, there could be an *FRI* gene in gladiolus but this would require better primer design to locate the gene because the presence of *GhFLX* might indicate in the presence of other flowering repressor genes as *FRI* protein upregulates *FLX* in *A. thaliana* and is part of the flowering pathway [11]. In addition, *Musa acuminata*, *Elaeis guineensis* and *Ananas comosus* are all monocots and tropical species which have *FLX* and *SUF4* genes as part of the *FRI* complex [51]. This indicates the presence of some of the *FRI* complex components while a lack of identification of *FRI* gene itself creates divergent possibilities: a) either there are *FRI* and *FLC* genes in these species or b) a lack of these genes and the presence of *SUF4* and *FLX* genes have other unknown flowering pathway purposes. Since *GhFLX* and *GdFLX* have similarities to *FLL1*, reaching up to 50% identity in amino acids, *FLX* gene is part of the gene family, *FLL1-FLL4* [11, 33]. While the role of *FLL1* in flowering pathway is not proven, *FLX* and *FLL4* are the most crucial genes in control of flowering time in *Arabidopsis* [33]. Additionally, the relationship between *UFC* and *FLX* indicates that a mutation in *FLX* influences *UFC* expression, e.g. the *flx* mutant in *A. thaliana* [42].

The next step in this research would be to identify the *UFC* gene in diploid gladiolus species to determine if the allele is similar to *GhUFC-A*, *GhUFC-B* or a third different allele. Use of diploids would simplify the study to determine the function of *UFC* protein in gladiolus by silencing and knocking out the gene. Locating the physical location of the *UFC* gene in *Gladiolus* will help in testing if there are other *UFC* genes in gladiolus as part of a *UFC* gene family, since the first discovered *UFC* gene (At5g10150) in *A. thaliana* is located in the cluster genes *UFC*, *FLC* and *DFC* on chromosome 5 [37]. *UFC* is also designated *SOK2* and the other *UFC* genes are grouped in *SOK* gene family such as *SOK1* (At1g05577), *SOK3* (At2g28150), *SOK4* (At3g46110) and *SOK5* (At5g59790) [39].

Identifying the *FLX* gene in Eurasian species of *Gladiolus*, particularly *G. italicus*, *G. imbricatus* and *G. communis*, would be informative since these winter-hardy, perennial species grow in temperate habitats that require vernalization to break corm dormancy in the winter season [29, 52–53]. Conversely, identifying *FLX* in subtropical gladiolus species, such as *G. crassifolius*, *G. laxiflorus* and *G. atropurpureus* [54], would allow comparison of *FLX* among these different habitats to further support the influence in *FLX* in the flowering pathway. Furthermore, the use of transgene silencing of *FLX* in gladiolus would determine whether or not *FLX* influences the production of a rapid flowering phenotype gladiolus (RGC-1), as was reported in the loss of *flx* function in *A. thaliana* [42]. In conclusion, the discovery of *UFC* and *FLX* genes in gladiolus provides insight into understanding flowering and vernalization responses in ornamental, monocot geophytes.

Conclusions

Rapid generation cycling is a powerful tool that can be implemented to reduce the juvenility period in perennial crops such as gladioli and are being applied in the breeding program ideotype. Although it is possible to annualize a perennial crop through genetically modifying the flowering pathway by overexpression a positive flowering regulator or inserting blocker of flowering suppressor, such biotechnological methods require regulatory approval. Therefore, conventional breeding methods for early flowering are widely accepted and do not require regulation for cultivar release.

The search for *FLC* and its regulatory genes in gladiolus is a step to uncover the flowering pathway in geophytes. To uncover if *FLC* is present in Gladiolus, we searched for linked genes with *FLC.* In *Arabidopsis*, *FLC* is adjacent to two genes, *UPSTREAM OF FLOWERING LOCUS C* (*UFC*) and *DOWNSTREAM OF FLOWERING LOCUS C* (*DFC)*, both of which are downregulated by vernalization. The discovery of *UFC* in gladiolus as well *FLX* (which upregulates FRI) is crucial to establish the flowering pathway. These may be early indicators of the presence of *FLC* homologue in gladiolus. Discovery of both genes are important to understand the flowering mechanism and genes in the flowering pathway to aid in breeding and selection of early flowering gladioli from seed or corms (Appendix A4).

Methods

Germplasm

The 17 gladiolus genotypes used in this study (Table 1) were chosen to represent a range of diversity within cultivated gladioli (*Gladiolus* × *hybridus*, *G. dalenii*) which includes nine genotypes of Rapid Generation Cycling-1 (RGC-1; ones that flower in ≤1 year from seed; [43-45]) and eight genotypes Non-RGC genotypes (that require 2 to 5 years to flower from seed). Fourteen of these genotypes are interspecific parents and hybrids created by the University of Minnesota Gladiolus Breeding Program, while three additional genotypes are commercial cultivars. One genotype ‘Carolina Primrose’ is derived from the species *G. dalenii* (Table 1). All gladiolus pedigrees used in this experiment are published [43-45] and commercial cultivars ‘Beatrice’ (an open-pollinated seedlings of unknown origin, occurring in a private

garden, Brookfield, Vermont, in 2003). 'Beatrice' was selected for its winter hardiness, surviving in USDA Z3). 'Glamini'® a series of shorter in height than tall summer gladiolus, bred by Dutch breeders, bloom early, has a range of flowering colors [55]. 'Carolina Primrose' is an heirloom gladiolus, bred in 1908, yellow color flowers, collected at an old homesite in North Carolina; it is a cultivar bred from *G. primulinus* [56].

Greenhouse Environment

Mature gladiolus corms (competent to flower) were planted into 1679.776 cm^2 square, deep pots (Belden Plastics, St. Paul. MN) in week 23 (2017) and grown for 18 weeks. Containers were filled with SS#8-F2-RSi potting soil, "SunGrow" (Sun Gro Horticulture, Agawam. MA). The corms were grown in a long day photoperiod (0800 – 1600 HR supplied by 400-W high-pressure sodium lamps + 2200 to 0200 HR night interruption, >150 µmol m^{-2} sec^{-1}) at a minimum setpoint of 18° C (day/night), 70-80% relative humidity, with irrigation accomplished using constant liquid feed (CLF) of 125 ppm N from water-soluble 20N–4.4P–16.6K (Scotts, Marysville, OH) and deionized water on weekends. Standard fungicide drenches and insecticides were applied either monthly or as needed, respectively.

DNA extraction and probe design

Newly expanded gladiolus leaves were harvested, placed in an ice box and sent to RAPiD Genomics® LLC (Gainesville, FL; http://rapid-genomics.com/home/) for DNA extraction, probe design, sequencing and computable analysis. Probe designs for the *UFC* gene were based on banana, *Musa acuminata* subsp. *malaccensis* accession XM_009383889, from the GenBank Nucleotide Core [57] and oil palm, *Elaeis guineensis* accession XM_010920607.2 [58]. Probe design for *FLX* gene were based on oil palm, *Elaeis guineensis* accession XM_010924316.2 [59] and date palm, *Phoenix dactylifera* accession XM_008801571.2 [60]. The designed probe for *UFC* able to capture the locus in *Musa acuminata* and *Elaeis guineensis* by capturing the 2x coverage of the *UFC* exons in *Musa acuminata* and *Elaeis guineensis*, while the *FLX* probe captures the locus in *Elaeis guineensis* and *Phoenix dactylifera*. Probes are amplified in short reads of *UFC* and *FLX* genes in gladiolus. The reads are sequenced through Illumina dye sequencing technique, the raw data is demultiplexed using Illuminas BCLtofastq then assembled using MaSuRCA® software [61], creating full assembly sequences scaffolds. Afterwards, read mapping using the reference genome and blast to filter all assembled sequences for hits to the sequences provided for probes design (*UFC* and *FLX*), then count read numbers for each assembled sequence passed the filtering, accruing the final sequences for genetic analysis. Gene sequences are currently being deposited into GenBank.

Genetic Analysis

The sequence data for the *UFC* and *FLX* genes used in this study were found in the genetic sequence database under the following accession/ID numbers: *Ananas comosus* (Aco009327) *UFC* gene is from the Pineapple Genomics Database [62]; *Musa acuminata* (GSMUA_Achr5T28540_001) *UFC* from the Banana Genome Hub [63]; *Elaeis guineensis* (p5.00_sc00099_p0095) *UFC* from the Malaysian Oil Palm

Genome Programme [64]; *Asparagus officinalis* (evm.model.AsparagusV1_08.3493) *UFC* from the Asparagus Genome Project [65]; *Arabidopsis thaliana* (At5g10150) *UFC* from The Arabidopsis Information Resource (TAIR) [66]; *Glycine max* (Glyma.11G193000.1) *UFC* from the SoyBase [67]. The *FLX* protein was from the GenBank Nucleotide Core with accession numbers as follows: *Ananas comosus* (XP_020095672.1) [68], *Musa acuminata* (XP_009420070.1) [69], *Elaeis guineensis* (XP_010922618.1) [70], *Arabidopsis thaliana* – *FLX* (NP_001154541.1) [71], *Arabidopsis thaliana* – *FLL1* (NP_566492.1) [72], *Arabidopsis thaliana* – *FLL2* (NP_001320766.1) [73], *Arabidopsis thaliana* – *FLL3* (NP_564678.1) [74], *Arabidopsis thaliana* – *FLL4* (NP_001119474.1) [75] and *Glycine max* (Glyma.15g269300) *FLX* from the SoyBase [67].

Generated sequences were analyzed for gene prediction using the HMM-based gene structure prediction of FGENESH with *Arabidopsis thaliana* (Generic) as the specific gene-finding parameter. The predicted genes for *UFC* and *FLX* were analyzed in multi-alignment using Geneious© software (Biomatters Ltd, Auckland, NZ). The *UFC* protein sequences of gladiolus were analyzed for conserved domains using the Protein Homology/analogy Recognition Engine V 2.0 (Phyre2) browser [76]. Then the alignment of conserved domain was formed to compare the matching and differences in each amino acid in the sequences of gladiolus and the other comparison species. A phylogenetic tree of all *UFC* genotypes of *Gladiolus* was formed by computing the distances using the Tamura-Nei method and were in the units of the number of base substitutions per site. The tree building used the Neighbor-Joining method and a bootstrap test was performed for each tree (500 replicates).

Declarations

Ethics approval and Consent to participate

All *Gladiolus* plant material was obtained from the University of Minnesota Gladiolus Breeding Program or from commercial vendors that possessed the propagation rights for the germplasm. Fresh leaf samples were harvested from plants growing in the University of Minnesota Plant Growth Facilities.

Consent for publication

Not applicable.

Availability of data and materials

The DNA sequences produced are being deposited in the NCBI GenBank data (https://www.ncbi.nlm.nih.gov/nuccore/).

Competing interests

The authors declare that they have no competing interests.

Funding

Funding for this research was supported by grants from the Minnesota Agricultural Experiment Station (MAES21-0045) and the Minnesota Gladiolus Society which funded sequencing and greenhouse charges. The Kuwaiti Government funded a Graduate (Ph.D.) Scholarship for Jaser Aljaser to conduct this research as part of his Ph.D. dissertation.

Authors' contributions

NOA directed the entire research project. JAA and NOA designed the research. JAA performed the experimentation, while JAA and AKN analyzed the data. JAA and NOA wrote the complete manuscript with equal contributions; AKN edited the manuscript along with NOA and JAA. All authors read and approved the final draft of the manuscript.

Acknowledgements

We sincerely thank Research Scientists Michele Schermann and Rajmund Eperjesi and undergraduate students, Allison Graper, Cora Rost, Sarah Gardner, and Betty Ziskovsky for their time and work spent in growing the genotypes.

References

1. Kamenetsky R, Zaccai M, Flaishman MA. Florogenesis. In: Kamenetsky R, Okubo H., eds. Ornamental geophytes: from basic science to sustainable production. CRC Press, Boca Raton, FL; 2012. p. 197-232.
2. Srikanth A, Schmid M. Regulation of flowering time: all roads lead to Rome. Cell Mol Life Sci. 2011;68(12):2013-2037.
3. Simpson GG, Dean C. *Arabidopsis*, the Rosetta stone of flowering time? Sci. 2002;296(5566):285-289.
4. Simpson GG. The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. Cur Opin Plant Biol. 2004;7(5):570-574.
5. Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C. Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. Sci. 2002;297(5579):243-246.
6. Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH. Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. Genes and Dev. 2007;21(4):397-402.
7. Kumar G, Arya P, Gupta K, Randhawa V, Acharya V, Singh AK. Comparative phylogenetic analysis and transcriptional profiling of MADS-box gene family identified *DAM* and *FLC*-like genes in apple (*Malus* x *domestica*). Sci Rep. 2016;6:20695.
8. Guo X, Yu C, Luo L, Wan H, Zhen N, Xu T, et al. Transcriptome of the floral transition in *Rosa chinensis* 'Old Blush'. BMC Genomics 2017;18(1):199.
9. de Oliveira RR, Cesarino I, Mazzafera P, Dornelas MC. Flower development in *Coffea arabica* L.: new insights into MADS-box genes. Plt Reprod. 2014;27(2):79-94.

10. Tadege M, Sheldon CC, Helliwell CA, Stoutjesdijk P, Dennis ES, Peacock WJ. Control of flowering time by *FLC* orthologues in *Brassica napus*. Plt J. 2001;28(5):545-553.
11. Choi K, Kim J, Hwang HJ, Kim S, Park C, Kim SY et al. The *FRIGIDA* complex activates transcription of *FLC*, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. Plt Cell 2011;23(1):289-303.
12. Michaels SD, Himelblau E, Kim SY, Schomburg FM, Amasino RM. Integration of flowering signals in winter-annual *Arabidopsis*. Plt Phys. 2005;137(1):149-156.
13. Michaels SD, Amasino RM. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plt Cell 1999;11:949–956.
14. Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C. Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. Sci. 2002;297(5579):243-246.
15. Turner AS, Faure S, Zhang Y, Laurie DA. The effect of day-neutral mutations in barley and wheat on the interaction between photoperiod and vernalization. Theor Appl Genet. 2013;126(9):2267-2277.
16. Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, et al. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Sci. 2004;303(5664):1640-1644.
17. Zhu QH, Helliwell CA. Regulation of flowering time and floral patterning by *miR172*. J Exptl Bot. 2011;62(2):487-495.
18. Ehrich L. Flowering in South African Iridaceae. In: Ramawat KG, Merillon JM, eds. Bulbous plants: Biotechnology. CRC Press, Boca Raton, FL. 2013;p. 248-269.
19. Fadón E. Herrero M, Rodrigo J. Flower bud dormancy in *Prunus* species. In: Anderson JV, ed. Advances in plant dormancy. Springer, The Netherlands. 2015;p. 123-135.
20. Taylor A. Functional genomics of photoperiodic bulb initiation in onion (*Allium cepa*). Univ of Warwick, U.K. 2009; PhD Dissertation.
21. Taylor A, Massiah AJ, Thomas B. Conservation of *Arabidopsis thaliana* photoperiodic flowering time genes in onion (*Allium cepa* L.). Plt Cell Phys. 2010;51(10):1638-1647.
22. Noy-Porat T, Flaishman MA, Eshel A, Sandler-Ziv D, Kamenetsky R. Florogenesis of the Mediterranean geophyte *Narcissus tazetta* and temperature requirements for flower initiation and differentiation. Sci Hort. 2009;120(1):138-142.
23. Noy-Porat T, Kamenetsky R, Eshel A, Flaishman MA. Temporal and spatial expression patterns of the *LEAFY* homologue *NLF* during florogenesis in *Narcissus tazetta*. Plt Sci. 2010;178(2):105-113.
24. Rotem N, Shemesh E, Peretz Y, Akad F, Edelbaum O, Rabinowitch HD, et al. Reproductive development and phenotypic differences in garlic are associated with expression and splicing of *LEAFY* homologue *gaLFY*. J Exp Bot. 2007;58(5):1133-1141. doi:10.1093/jxb/erl272.
25. Rotem N, David-Schwartz R, Peretz Y, Sela I, Rabinowitch HD, Flaishman M, et al. Flower development in garlic: the ups and downs of *gaLFY* expression. Planta 2011;233(5):1063-1072.
26. Wang AJ, Tang JF, Zhao XY, Zhu LH. Isolation of *LiLFY1* and its expression in lily (*Lilium longiflorum* Thunb.). Agric Sci China 2008;7(9):1077-1083.

27. Li YF, Zhang MF, Zhang M, Jia GX. Analysis of global gene expression profiles during the flowering initiation process of *Lilium* × *formolongi*. Plt Mol Biol. 2017;94(4-5):361-379.
28. Zlesak D, Anderson NO. Inheritance of non-obligate vernalization requirement for flowering in *Lilium formosanum* Wallace. Isr J Plt Sci. 2010;57(4):315-327.
29. Cohat J. *Gladiolus*. In: De Hertogh A, Le Nard M, eds. The physiology of flower bulbs. Elsevier, Amsterdam. 1993;p. 297-320.
30. Kamo KK, Krens FA, Ziv M. Biotechnology for the modification of horticultural traits in geophytes. In: Kamenetsky R, Okubo H, eds. Ornamental geophytes: from basic science to sustainable production. CRC Press, Boca Raton, FL. 2012:p. 159-195.
31. Luo X, Lu H, Yuan L, Jia Y, Wu Y, Deng Q. Cloning and expression analysis of gibberellin receptor gene in *Gladiolus hybridus*. Acta Bot. Boreali-Occid Sinica 2016;36(11):2152-2158.
32. Blázquez MA, Weigel D. Integration of floral inductive signals in *Arabidopsis*. Nature 2000;404(6780):889-892.
33. Lee J, Amasino RM. Two *FLX* family members are non-redundantly required to establish the vernalization requirement in *Arabidopsis*. Nat Comm. 2013;4:2186. doi:10.1038/ncomms3186.
34. Zhao T, Ni Z, Dai Y, Yao Y, Nie X, Sun Q. Characterization and expression of 42 MADS-box genes in wheat (*Triticum aestivum* L.). Mol Genet Genom. 2006:276(4):334.
35. Monteagudo A, Igartua E, Contreras-Moreira B, Gracia MP, Ramos J, Karsai I, et al. Fine-tuning of the flowering time control in winter barley: the importance of *HvOS2* and *HvVRN2* in non-inductive conditions. BMC Plt Biol. 2019;19(1):1-14.
36. Ruelens P, De Maagd RA, Proost S, Theißen G, Geuten K, Kaufmann K. *FLOWERING LOCUS C* in monocots and the tandem origin of angiosperm-specific MADS-box genes. Nat Comm. 2013;4(1):1-8.
37. Finnegan EJ, Sheldon CC, Jardinaud F, Peacock WJ, Dennis ES. A cluster of *Arabidopsis* genes with a coordinate response to an environmental stimulus. Cur Biol. 2004;14(10):911-916.
38. Sheldon CC, Finnegan JE, Peacock JW, Dennis ES. Mechanisms of gene repression by vernalization in *Arabidopsis*. Plt J. 2009;59(3):488-498.
39. Yoshida S, Van Der Schuren A, Van Dop M, Van Galen L, Saiga S, Adibi M, et al. A SOSEKI-based coordinate system interprets global polarity cues in *Arabidopsis*. Nat Plts. 2019;5(2):160-166. doi:10.1038/s41477-019-0363-6.
40. Chengke L, Lei T. Bioinformatics analysis of rice *DUF966* gene family. Mol Plt Br. 2017;15(12):4791-4796.
41. Ye J, Zhong T, Zhang D, Ma C, Wang L, Yao L, et al. The auxin-regulated protein *ZmAuxRP1* coordinates the balance between root growth and stalk rot disease resistance in maize. Mol Plt. 2019;12(3):360-373.
42. Andersson CR, Helliwell CA, Bagnall DJ, Hughes TP, Finnegan EJ, Peacock WJ, Dennis ES. The *FLX* gene of *Arabidopsis* is required for *FRI*-dependent activation of *FLC* expression. Plt Cell Phys. 2008;49(2):191-200.

43. Anderson NO, Carter J, Hershman A, Houseright V. Rapid generation cycling enhances selection rate of *Gladiolus* x*hybridus*. Acta Hort. 2015;1087:429-435.
44. Anderson NO. Selection tools for reducing generation time of geophytic herbaceous perennials. XIII International Symposium on Flower Bulbs and Herbaceous Perennials, Acta Hort. 2019;1237:53-66.
45. Aljaser JA. *Gladiolus* breeding for rapid generation cycling for potted production and the discovery of gladiolus genes, *UFC* and *FLX*. Univ. Minnesota, Minneapolis, MN, USA. PhD. Dissertation.
46. Bamford R. The chromosome number in *Gladiolus*. J Agr Res. 1935;51:945-950.
47. Saito K, Kusakari, K. Studies on the occurrence of polyploidy and its contribution to the flower plants breeding: IX. Cytological observations on the mechanism of decreased fertility in the summer-flowering tetraploid cultivars of *Gladiolus grandiflorus* Hort. Jap J Breed. 1972;22(2):75-82.
48. Ohri D. Cytogenetics of garden gladiolus. Cytol. 1985;50(2):213-231.
49. Benschop M, Kamenetsky R, Le Nard M, Okubo H, De Hertogh, A. The global flower bulb industry: Production, utilization, research. Hort Rev. 2010;36(1):1-115.
50. Goldblatt P, Manning J. Gladiolus in southern Africa. Fernwood Press (Pty) Ltd., Newburg, OR. 1998.
51. Amasino RM, Michaels SD. The timing of flowering. Plt Phys. 2010;154(2):516-520.
52. Cohen A, Barzilay A. Miniature gladiolus cultivars bred for winter flowering. HortSci. 1991;26(2):216-218.
53. Gonzalez A, Banon S, Fernandez JA, Franco JA, Casas JL, Ochoa J. Flowering responses of *Gladiolus tristis* (L.) after exposing corms to cold treatment. Sci Hort. 1998;74(4):279-284.
54. Goldblatt P. Gladiolus in tropical Africa: Systematics, biology and evolution. Timber Press, Portland, OR. 1996.
55. Wayside Gardens. Gladiolus Glamini® Mixed. 2020. https://www.waysidegardens.com/glamini-mixed-gladiolus/p/27192-PK-10/
56. Old House Gardens. Gladiolus: Lost forever? 2020. https://oldhousegardens.com/display/category/BackSoonGladiolus?bulb=SGL30
57. National Center for Biotechnology Information. Accession No. XM_009383889.2, PREDICTED: *Musa acuminata* subsp. *malaccensis* protein *UPSTREAM OF FLC* (LOC103970210), transcript variant X1, mRNA. 2016a. https://www.ncbi.nlm.nih.gov/nuccore/XM_009383889
58. National Center for Biotechnology Information. Accession No. XM_010920607.2, PREDICTED: protein *UPSTREAM OF FLC* [*Elaeis guineensis*]. 2017a. https://www.ncbi.nlm.nih.gov/nuccore/XM_010920607.2
59. National Center for Biotechnology Information. Accession No. XM_010924316.2, protein *FLC EXPRESSOR* [*Elaeis guineensis*]. 2017b. https://www.ncbi.nlm.nih.gov/nuccore/XM_010924316.2
60. National Center for Biotechnology Information. Accession No. XM_008801571.2, PREDICTED: protein *FLOWERING LOCUC C* EXPRESSOR-like [*Phoenix dactylifera*]. 2016b. https://www.ncbi.nlm.nih.gov/nuccore/XM_008801571.2

61. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinfo. 2013;29(21):2669-2677.
62. Xu H, Yu Q, Shi Y, Hua X, Tang H, Yang L, et al. PGD: pineapple genomics database. Hort Res. 2018;5(1):1-9. Pineapple Genomics Database http://pineapple.angiosperms.org/pineapple/html/index.html
63. Droc G, Lariviere D, Guignon V, Yahiaoui N, This D, Garsmeur O, et al. The Banana Genome Hub Database. 2013. doi:10.1093/database/bat035
64. Sanusi NSNM, Rosli R, Halim MAA, Chan KL, Nagappan J, Azizi N, et al. PalmXplore: oil palm gene database. 2018. http://palmxplore.mpob.gov.my
65. Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, et al. The asparagus genome sheds light on the origin and evolution of a young *Y* chromosome. Nat Comm. 2017;8(1):1-10. http://asparagus.uga.edu/
66. The Arabidopsis Information Resource (TAIR). The Arabidopsis Information Resource (TAIR). 2020. https://www.arabidopsis.org/servlets/TairObject?id=136005andtype=locus
67. Grant D, Nelson RT, Cannon SB, Shoemaker C. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucl Acids Res. 2010;38(suppl 1):D843-D846. https://www.soybase.org/ doi:10.1093/nar/gkp798.
68. National Center for Biotechnology Information. Accession No. XP_020095672.1, protein *FLX*-like 1 [*Ananas comosus*]. 2017c. https://www.ncbi.nlm.nih.gov/protein/XP_020095672.1
69. National Center for Biotechnology Information. Accession No. XP_009420070.1, PREDICTED: protein *FLX-like 1* [*Musa acuminata* subsp. *malaccensis*]. 2016c. https://www.ncbi.nlm.nih.gov/protein/695063148
70. National Center for Biotechnology Information. Accession No. XP_010922618.1, protein *FLC EXPRESSOR* [*Elaeis guineensis*]. 2019a. https://www.ncbi.nlm.nih.gov/protein/XP_010922618.1
71. National Center for Biotechnology Information. Accession No. NP_001154541.1, protein *FLC EXPRESSOR* [*Arabidopsis thaliana*]. 2019b. https://www.ncbi.nlm.nih.gov/protein/NP_001154541.1
72. National Center for Biotechnology Information. Accession No. NP_566492.1, structural maintenance of chromosomes domain protein [*Arabidopsis thaliana*]. 2019c. https://www.ncbi.nlm.nih.gov/protein/NP_566492.1
73. National Center for Biotechnology Information. Accession No. NP_001320766.1, sarcolemmal membrane-associated protein [*Arabidopsis thaliana*]. 2019d. https://www.ncbi.nlm.nih.gov/protein/NP_001320766.1
74. National Center for Biotechnology Information. Accession No. NP_564678.1, DNA double-strand break repair protein [*Arabidopsis thaliana*]. 2019e. https://www.ncbi.nlm.nih.gov/protein/NP_564678.1
75. National Center for Biotechnology Information. Accession No. NP_001119474.1, *FLX*-like protein [*Arabidopsis thaliana*]. 2019f. https://www.ncbi.nlm.nih.gov/protein/NP_001119474.1

76. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The *Phyre2* web portal for protein modeling, prediction and analysis. Nat Protocols 2015;10(6):845-858. http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index

Tables

Table 1. The *Gladiolus* genotypes used in this study, their codes and whether they are Rapid Generation Cycling (RGC): + is for RGC genotypes and – for Non-RGC genotypes; all gladiolus genotypes were tested for the presence of *UFC* gene, *FLX* gene and *FRI* gene.

Genotype	Code	RGC
Gladiolus × *hybridus* 21213	1	+
Gladiolus × *hybridus* 2220	2	+
Gladiolus × *hybridus* 2231	3	+
Gladiolus × *hybridus* 2337	4	+
Gladiolus × *hybridus* 35314	5	+
Gladiolus × *hybridus* 3923	6	+
Gladiolus × *hybridus* 3931	7	+
Gladiolus × *hybridus* 74210	8	+
Gladiolus × *hybridus* 7736	9	+
Gladiolus × *hybridus* 28236	10	-
Gladiolus × *hybridus* 15531	11	-
Gladiolus × *hybridus* 20732	12	-
Gladiolus × *hybridus* 60314	13	-
Gladiolus dalenii 'Carolina Primrose'	14	-
Gladiolus × *hybridus* 'Beatrice'	15	-
Gladiolus × *hybridus* 'Glamini'®	16	-
Gladiolus × *hybridus* 'Vista	17	-

Table 2. The identity of amino acid sequences and number (%) of two *UFC* proteins (*GhUFC*-A, *GhUFC*-B) in two *Gladiolus* (genotypes 1 and 15) in relation to other species (Gene locus/ID) through pair alignment; the similarities of sequences in identity of gladiolus genotypes ranges from ~30% to 57% across all

investigated species for the whole *UFC* protein; alignment is done with MUSCLE alignment using the neighbor joining clustering method and the CLUSTALW sequencing scheme (Geneious®).

	Gladiolus ×hybridus genotype			
	Genotype 1		Genotype 15	
Species (Gene locus/ID)	***GhUFC*-A (%)**	***GhUFC-B (%)***	***GhUFC-A (%)***	***GhUFC-B (%)***
Ananas comosus (Aco009327)	189 (33.51%)	249 (53.21%)	191 (34.35%)	231 (52.98%)
Musa acuminata (GSMUA_Achr5T28540_001)	251 (54.09%)	170 (39.35%)	250 (57.74%)	148 (31.36%)
Elaeis guineensis (p5.00_sc00099_p0095)	254 (52.05%)	172 (40.86%)	250 (51.23%)	155 (39.85%)
Asparagus officinalis (evm.model.AsparagusV1_08.3493)	213 (46.61%)	148 (35.58%)	213 (50.0%)	135 (29.87%)
Arabidopsis thaliana (AT5G10150)	137 (28.54%)	129 (29.79%)	138 (31.72%)	117 (29.32%)
Glycine max (Glyma.11G193000.1)	181 (33.64%)	226 (52.19%)	188 (35.67%)	207 (52.01%)

Table 3. Identity of amino acid sequences, number (%) of *UFC* proteins in the conserved domain DUF966 in *Gladiolus* genotypes 1 and 15 in relation to other species through pair alignment. *Gladiolus* genotypes 1 and 15 exhibit a range of ~65% to ~86% across all investigated species for the DUF966 domain of *UFC* protein.

	Gladiolus ×hybridus genotype			
	Genotype 1		Genotype 15	
Species (Gene locus/ID)	***GhUFC*-A (%)**	***GhUFC-B (%)***	***GhUFC-A (%)***	***GhUFC-B (%)***
Ananas comosus (Aco009327)	70 (76.09%)	79 (84.95%)	79 (84.95%)	79 (84.95%)
Musa acuminata (GSMUA_Achr5T28540_001)	78 (84.78%)	71 (78.02%)	78 (84.78%)	71 (78.02%)
Elaeis guineensis (p5.00_sc00099_p0095)	79 (85.87%)	72 (79.12%)	79 (85.87%)	72 (79.12%)
Asparagus officinalis (evm.model.AsparagusV1_08.3493)	75 (82.42%)	70 (76.09%)	75 (82.42%)	70 (76.09%)
Arabidopsis thaliana (AT5G10150)	61 (66.30%)	61 (64.89%)	62 (67.39%)	61 (64.89%)
Glycine max (Glyma.11G193000.1)	70 (76.92%)	75 (82.42%)	70 (76.92%)	75 (81.52%)

Table 4. Number (%) of amino acid sequences of *GhFLX* protein in *Gladiolus* genotypes 3 and 6 (genotype 16 is identical to genotype 6) in relation to the other species through pair alignment; similarities of sequences in identity of gladiolus genotypes ranged from ~26% to ~65% across all investigated species of the whole *FLX* protein; alignment is done in MUSCLE, using the neighbor joining clustering method and the CLUSTALW sequencing scheme (Geneious®).

Species (Accession no.)	*Gladiolus ×hybridus* genotype	
	GhFLX Genotype 3 (%)	*GhFLX* Genotype 6 and 16 *(%)*
Ananas comosus (XP_020095672.1)	180 (64.98%)	177 (64.60%)
Musa acuminata (XP_009420070.1)	122 (48.03%)	122 (48.03%)
Elaeis guineensis (XP_010922618.1)	132 (50.00%)	131 (49.62%)
Arabidopsis thaliana – FLX (NP_001154541.1)	82 (30.48%)	82 (30.48%)
Arabidopsis thaliana – FLL1 (NP_566492.1)	135 (50.00%)	135 (50.00%)
Arabidopsis thaliana – FLL2 (NP_001320766.1)	96 (36.09%)	94 (35.34%)
Arabidopsis thaliana – FLL3 (NP_564678.1)	104 (34.67%)	104 (34.67%)
Arabidopsis thaliana – FLL4 (NP_001119474.1)	67 (26.38%)	68 (26.77%)
Glycine max (Glyma.15g269300)	156 (57.14%)	155 (56.78%)

Figures

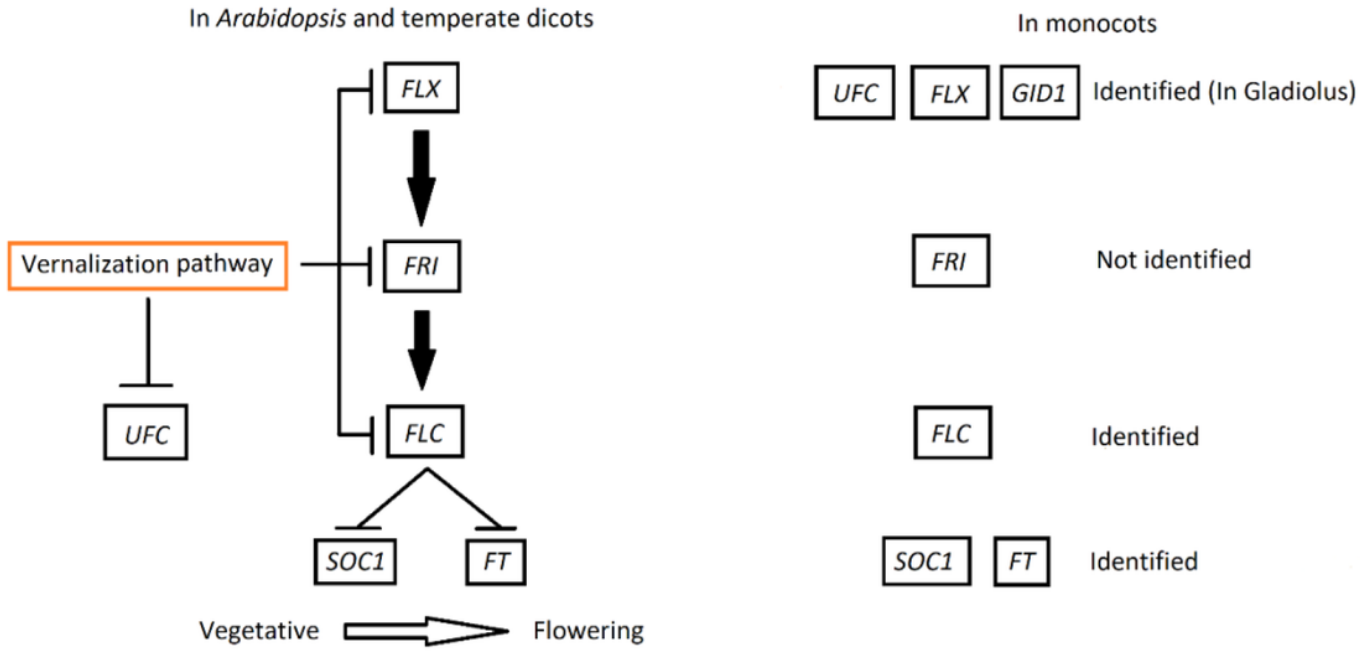

Figure 1

Model represents portion of flowering pathway regarding the role of FLX gene in flowering along with UFC gene in Arabidopsis and temperate dicots. FLX upregulates FRI which also upregulates the expression of the FLC protein and suppresses flowering by repressing the expression of the floral integrator SOC1 and FT. Vernalization pathway downregulates the expression of FLX, FRI and FLC genes allowing the floral integrators to initiate flowering in vegetative state of dicots, while the vernalization pathway also

downregulates the UFC gene [37]. In monocots, FLX, FLC, SOC1 and FT have been identified [22,35,51], while UFC and FLX have just been identified in gladiolus (in the current experiments). However, FRI was not identified either by lacking the presence of these repressor genes or monocots relying on other options of the flowering pathway genes.

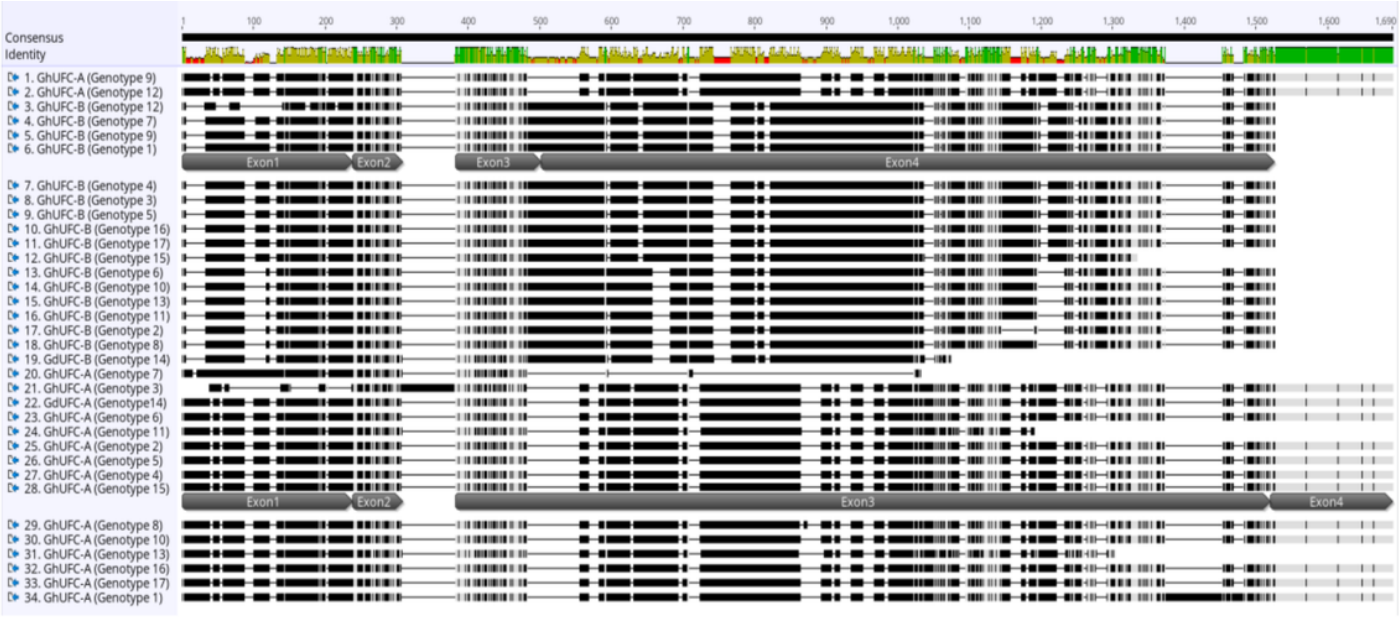

Figure 2

Multi-alignment of UFC coding sequence in Gladiolus ×hybridus (GhUFC) and Gladiolus dalenii (GdUFC), the alignment is for the 17 genotypes, each genotype has 2 alleles, allele A and allele B: GhUFC-A, GdUFC-A, GhUFC-B, GdUFC-B. Both alleles has 4 exons but allele A size is larger in coding sequence than allele B. The alignment shows insertion and missing coding sequences in some genotypes. The multi-alignment is done in MUSCLE pair-alignment using neighbor joining cluster method and CLUSTALW sequencing scheme (Geneious)®

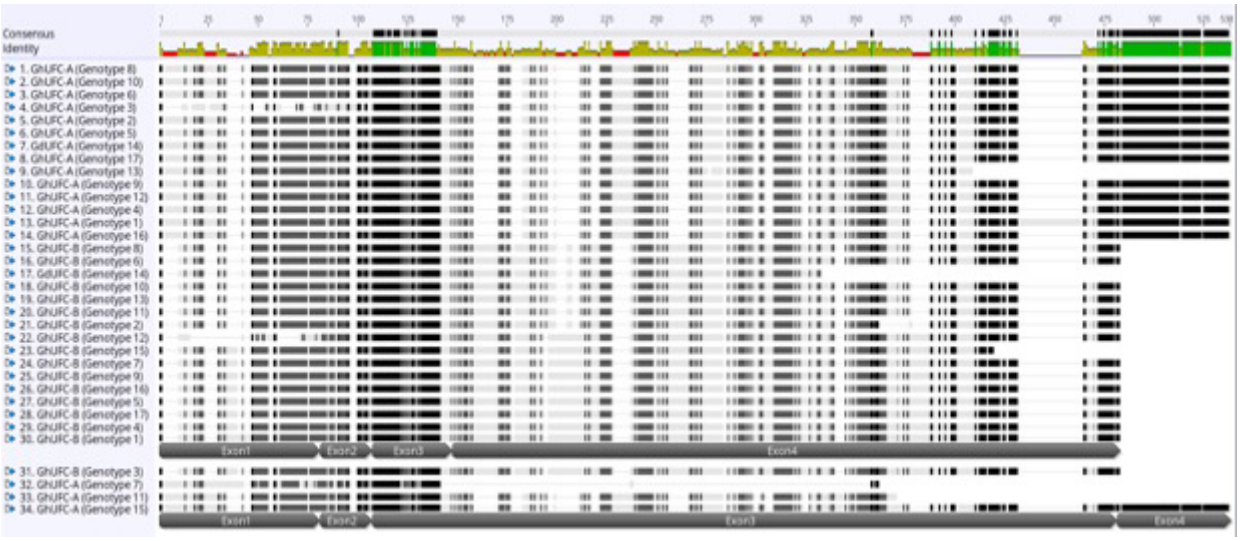

Figure 3

Multi-alignment of UFC amino acid sequence in Gladiolus ×hybridus (GhUFC) and Gladiolus dalenii (GdUFC), the alignment is for the 17 genotypes, each genotype has 2 alleles, allele A and allele B: GhUFC-A, GdUFC-A, GhUFC-B, GdUFC-B. Both alleles has 4 exons but allele A size is larger in amino acid sequence than allele B. The alignment shows insertion and missing amino acid sequences in some genotypes. The alignment identify conserved amino acid sequences (green color). The multi-alignment is done in MUSCLE pair-alignment using neighbor joining cluster method and CLUSTALW sequencing scheme (Geneious)®

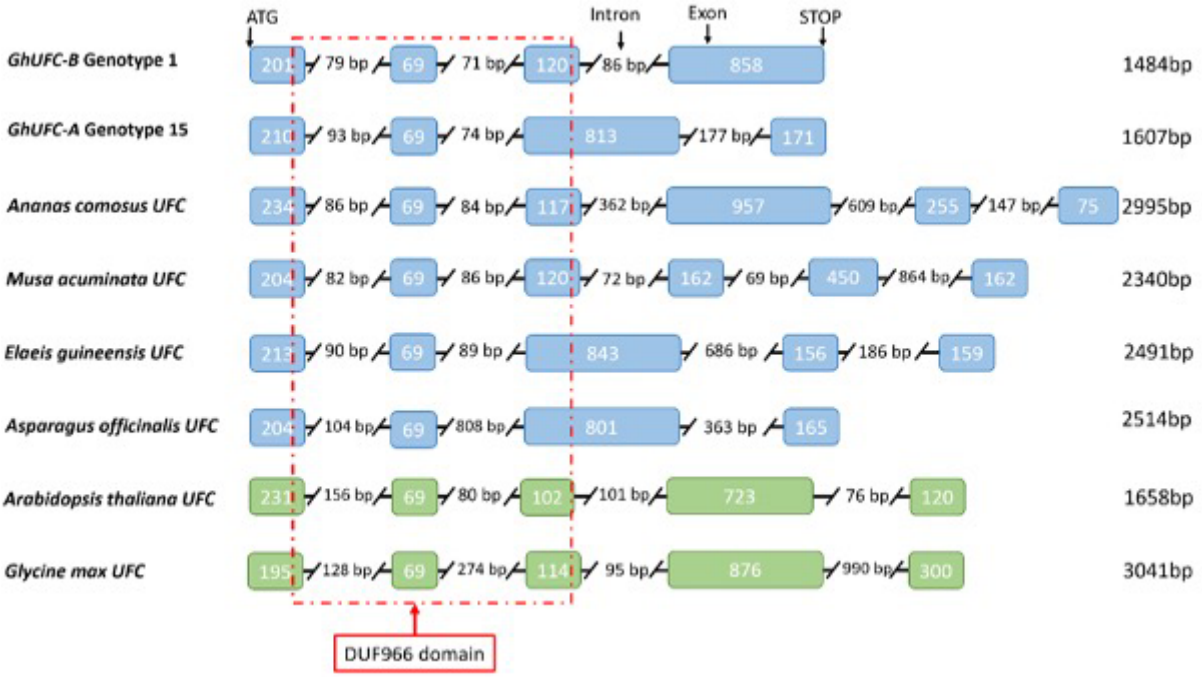

Figure 4

Intron-exon configuration of the UFC genes in Gladiolus ×hybridus of genotypes 1 and 15 in relation to several species. Monocot species are highlighted in blue: Ananas comosus, Musa acuminata, Elaeis guineensis and Asparagus officinalis. Dicot species highlighted in green for Arabidopsis thaliana and Glycine max. The intron-exon organization of GhUFC-A genotype 15 is similar to Elaeis guineensis and Asparagus officinalis in terms of exons splicing while GhUFC-B genotype 1 has some similarity with Ananas comosus exons. Sequences were aligned based on first exon sequences. Total length of the gene's coding region is listed on the right of each respected species. The red line represents the conserved domain DUF966 in relation to the location of the domain with exon configuration.

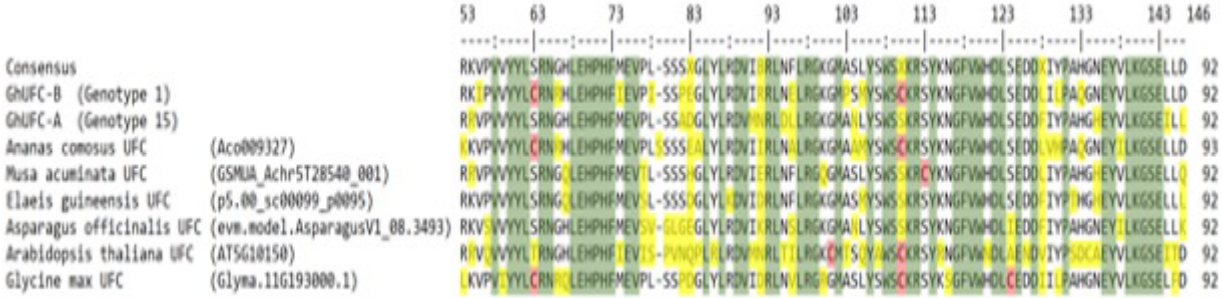

Figure 5

Alignment of the globular region containing DUF966 domain of UFC proteins from Gladiolus ×hybridus of genotypes 1 and 15, Ananas comosus, Musa acuminata, Elaeis guineensis Asparagus officinalis, Arabidopsis thaliana and Glycine max. Green coloration shows identical amino acid sequence; yellow color highlights the polymorphisms while red color shows the cytosine amino acid. The conserved domain DUF966 is 92 amino acids.

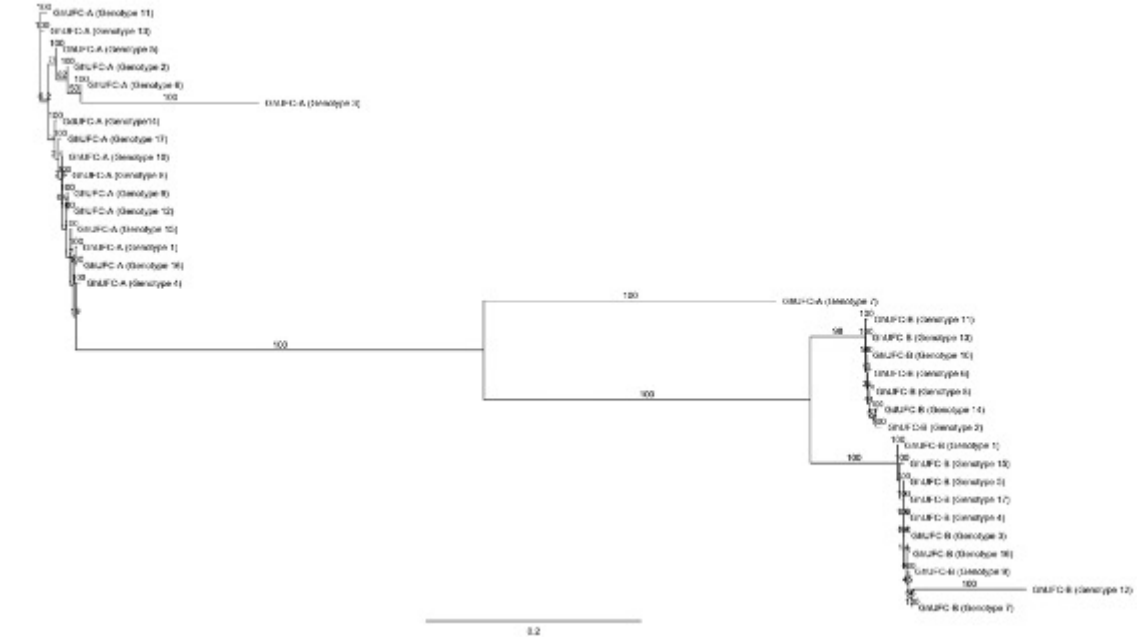

Figure 6

The phylogenetic tree of all UFC genotypes in Gladiolus; genetic distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. The tree build using the Neighbor-Joining method and the bootstrap test was performed for each tree (500 replicates) and the tree format is organized and ordered with a scale bar of 0.2 (Geneious®).

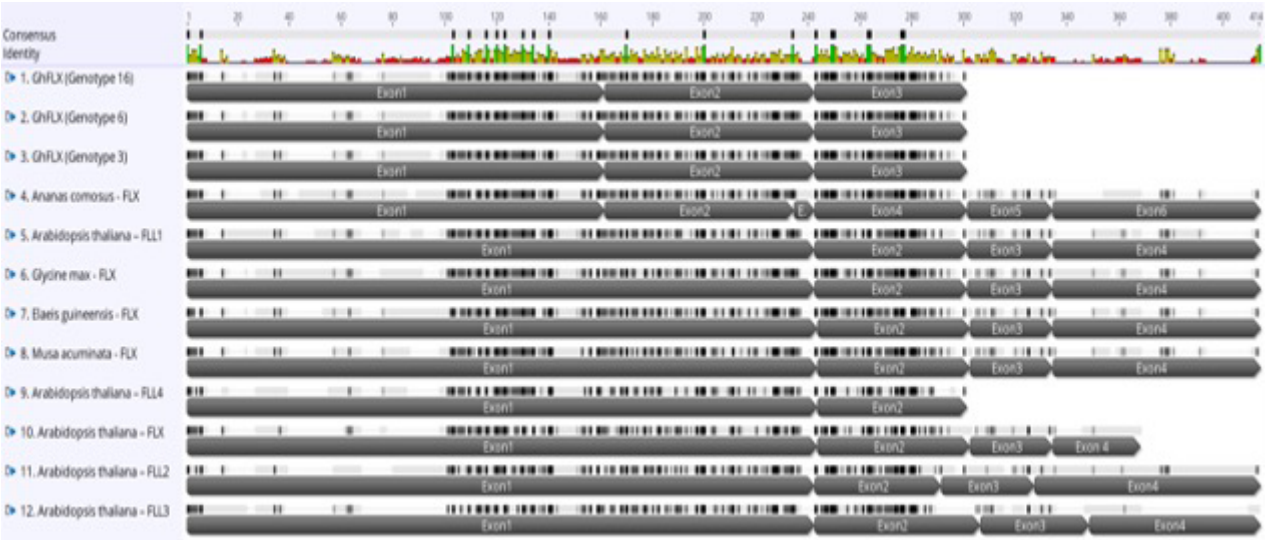

Figure 7

Multi-alignment of FLX amino acid sequence in Gladiolus ×hybridus (GhFLX) with other species; Arabidopsis thaliana, Ananas comosus, Elaeis guineensis, Musa acuminata and Glycine max. The alignment is for the 3 gladiolus genotypes (3, 6 and 16) each genotype has 3 exons. The alignment shows missing amino acid sequences in gladiolus genotypes 3, 6 and 16 as amino acid sequences does not have a stop codon. The alignment identifies conserved amino acid sequences (green color). Note Arabidopsis thaliana – FLL4 is a functional protein which has two exons only [33]. The multi-alignment is done in MUSCLE pair-alignment using neighbor joining cluster method and CLUSTALW sequencing scheme (Geneious)®.

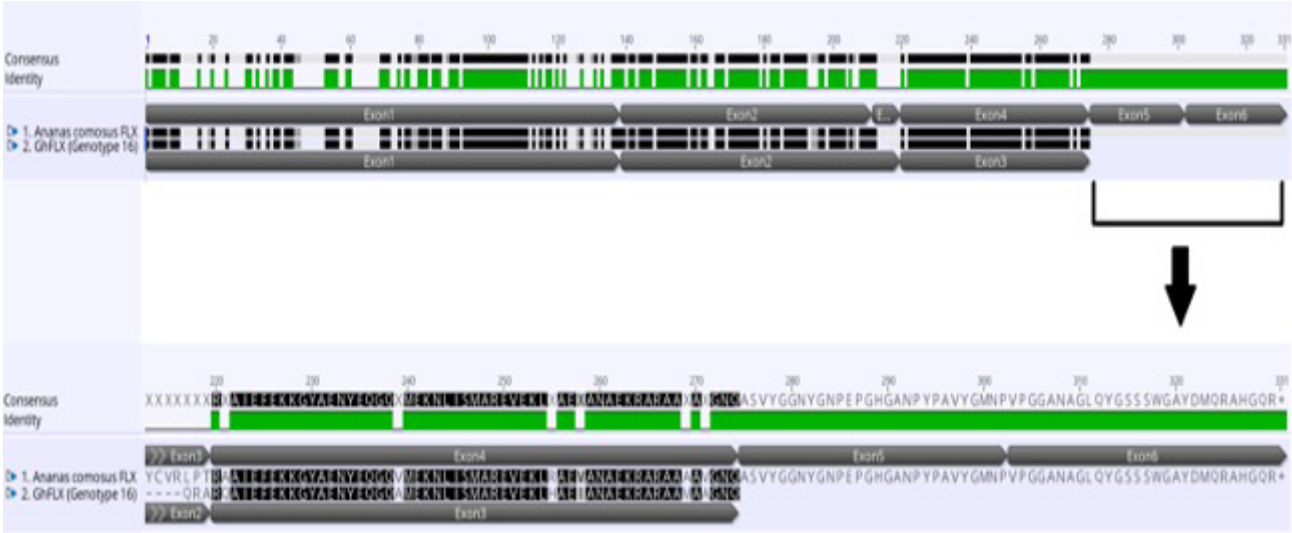

Figure 8

Pair alignment of FLX protein in Ananas comosus and Gladiolus ×hybridus genotype 16 showing the complete FLX protein in Ananas comosus with five exons while incomplete FLX protein in Gladiolus ×hybridus (GhFLX) which lacks the remaining two exons and stop codon. The alignment is done in MUSCLE pair-alignment using neighbor joining cluster method and CLUSTALW sequencing scheme (Geneious)®.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendicesA1A3.pdf
- AppendicesA2A3originalgelpics.pdf