Genome-wide Identification and Characterization of G-type lectin in Fragaria Vesca

Lijing Ma  
DISTAL, Alma Mater Studiorum - University of Bologna

Francesca Negrini (✉ francesca.negrini6@unibo.it)  
DISTAL, Alma Mater Studiorum - University of Bologna

Zeraye Mehari Haile  
DISTAL, Alma Mater Studiorum - University of Bologna

Elena Baraldi  
DISTAL, Alma Mater Studiorum - University of Bologna

Research Article

Keywords: bulb type lectin, diploid strawberry, GNA, Ser/Thr kinase domain

DOI: https://doi.org/10.21203/rs.3.rs-780793/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background: Lectins make up a large and diverse group of proteins in plants. G-type lectins are important type of lectins involved in plant development and defense process. However, studies about G-type lectins are limited to lectin receptor kinases.

Results: In this study, genome-wide identification was carried out on G-type lectin gene family in *Fragaria vesca*. A total of 133 genes were found belonging to this family and they were classified into four groups: G-type lectin receptor kinases, G-type lectin kinases, G-type lectin receptor proteins and G-type lectin proteins, according to their domain organizations. Their chromosome localization, phylogenetic and evolutionary relationship were also analyzed. The results showed that tandem and dispersed duplication occurred frequently, which led to the expansion of G-type lectin gene family in *F. vesca* and may have increased the types of domain arrangement. The expression profile of G-type lectin genes at different developmental stages of *F. vesca* and under various biotic/abiotic stresses was inferred from the available databases. G-type lectin genes are actively expressed during *F. vesca* development and respond to multiple biotic/abiotic stresses. Additionally, to comprehend the functions of G-type lectins, we predicted strawberry genes that may co-express with these G-type lectin genes.

Conclusions: G-type lectin gene family is a large gene family in *F. vesca*. Domain organization and expression analysis imply their functions under biotic/abiotic stresses.

1. Background

Lectins are proteins containing one or more lectin domains that could specifically and reversibly bind carbohydrate or glycan structures [1]. They consist of a large and diverse class of proteins and exist in all kingdoms of life [2]. Plant lectins can be classified into 12 distinct subfamilies based on their conserved carbohydrate-binding domains: *Agaricus bisporus* agglutinin family, Amaranthins, homologs of class V Chitinases (CRA), Cyanovirin, *Euonymus europaeus* lectin (EUL), *Galanthus nivalis* agglutinin (GNA), Hevein, Jacalin-related lectin (JRL), legume Lectin, Lysin Motif (LysM), *Nicotiana tabacum* agglutinin (Nictaba), and Ricin B lectin family [3]. Such abundance and diverse carbohydrate-binding ability imply an important function of these proteins in plants. Many lectins have already been shown to be involved in plant biotic and abiotic stress regulations [4–10]. In particular, due to the capability of carbohydrate recognition and binding, lectin receptor-like kinases (LecRKs), are candidate proteins for pathogen-/damage-associated molecular pattern (P/DAMP) recognition. LecRKs usually consist of Ser/Thr kinase domain and lectin domains from CRA, GNA, Jacalin, Nictaba, LysM, and legume lectin family [11].

GNA-related lectins, also known as G-type lectin or G-lectin, contribute to a large part of the whole plant lectins that have an affinity for mannose or mannose complex [12–14]. Since the first GNA-related lectin was isolated in the bulbs of *Galanthus nivalis*, this type of lectin is also named bulb-type lectin or B-lectin,
and the domain is named B-lectin or GNA domain (*Galanthus nivalis* agglutinin-related lectin domain) [15].

Besides GNA domain, G-type lectins also contain other domains, such as S-locus glycoprotein domain (SLG), PAN/Apple domain (PAN), transmembrane domain (TM), and protein kinase domain (PK) [16]. Concerning their role, G-type lectins are predicted to have important functions in plant development and resistance. A large group of G-type lectins has shown insecticidal properties, particularly against aphids of wheat, maize, potato, and sugarcane, by affecting their development and fecundity [17–20]. *CaMBL1* and *CaGLP1* are pepper G-type lectin genes involved in signaling and plant cell death that were shown to play a role in defense against *Xanthomonas campestris pv vesicatoria* [8, 14]. Similarly, G-type LecRK gene of *Arabidopsis* takes part in defense signaling by recognizing lipopolysaccharides of *Xanthomonas* and *Pseudomonas* [21]. Lipopolysaccharides are well-described PAMPs that trigger plant innate immunity [22, 23]. The transfer of a G-type LecRK gene *Pi-d2* to rice conferred race-specific resistance to *Magnaporthe grisea* [4], and knocking down of *OslecRK* gene, also a G-type LecRK, reduced the resistance of rice plant to *X. oryzae pv. oryzae* and brown planthopper [24]. In strawberry, *FaMBL1* was found involved in the resistance of unripe fruits to *Colletotrichum acutatum* [25].

Besides their role in resistance to biotic stress, G-type lectins play a role in plant adaption to abiotic stress. *OsSIK2* enhanced rice tolerance to salt and drought stresses, also delayed dark-induced leaf senescence [26]. Transgenic *Arabidopsis* plant expressing *GsSRK* exhibited enhanced salt tolerance and higher yields under salt stress [6]. Both *OsSIK2* and *GsSRK* could be induced by abscisic acid, salt, and drought stresses [6, 26].

Interestingly, G-type lectins also have potential medical applications. Some G-type lectins could recognize some of the high-mannose N-glycans exposed at the surface of gp120 of HIV-1 [27, 28], acting as inhibitors of the entry of HIV-1 into CD4+ T-lymphocytes. Besides, there are also G-type lectins that could specifically bind altered hypermannosylation N-glycans on the surface of cancer cells and cause programmed cell death of tumor cells [29].

Functional analysis and genome-wide studies of G-type lectin have been performed in different plants, such as *Arabidopsis* [30–32], soybean [30], rice [31], tomato [32], mulberry [33], and cucumber [34]. Strawberry is a good model plant for the study of *Rosaceae* plants and the study of G-type lectin family in this species can provide information for other *Rosaceae* plants as well. Recently, with the updated genome annotation and comprehensive gene expression atlas of *F. vesca* (https://www.rosaceae.org/species/fragaria_vesca genome_v4.0.a2) [35], reliable data are available for genome-wide analysis of G-type lectin genes in strawberry. Moreover, most studies on plant G-type lectins focused on the G-type LecRKs, lacking insights on potential biological functions of G-type lectins without kinase domain. In this study, using the newly released *F. vesca* genome annotation (v4.0.a2), we identified the woodland strawberry lectin gene family members and characterized their genomic organization and phylogenic relationship. To get insights into their functions, we further analyzed the variation in domain composition and their expression profile at different stages.
2. Results

2.1. *F. vesca* G-lectin genes identification and characterization

2.1.1. G-lectin genes identification, classification, and domain organization

*F. vesca* protein sequences containing GNA domain were fished out by BLASTp from the *F. vesca* protein database included in the Genome Database for Rosaceae (GDR) (https://www.rosaceae.org/) [36] website. The first BLASTp was carried out using the amino acid sequence (from 70 to 208 aa) of the GNA domain (domain ID: IPR001480) encoded by FvH4_3g18380 gene. This gene is the homolog of *F. x ananassa* FaMBL1, that encodes for a protein containing GNA and PAN domains and was reported as overexpressed in white strawberry fruit in response to anthracnose disease [25]. The search allowed to retrieve 77 different protein sequences. To find out more proteins and reduce redundancy, 20 of these sequences with relatively low similarity were chosen and used for a second BLASTp search, leading to a total of 133 proteins with GNA domains found in *F. vesca* (Supplemental file S1). Among these, 102 proteins containing PK and TM domain were classified into G-LecRK; 23 proteins lacking both domains were classified into G-LecP, and finally 4 proteins lacking PK but retaining the TM domain, were grouped G-LecRP. In addition, four genes (FvH4_3g03241; FvH4_3g03300; FvH4_3g15980; FvH4_6g44240) missing the TM domain but containing both GNA and PK domain were found, and they were classified into G-LecK.

Besides GNA domain, most of G-type lectins of *F. vesca* also contain other domains like S-locus glycoprotein domain (SLG), PAN/Apple domain (PAN), and Epidermal Growth Factor domain (EGF) (Fig. 1). SLG is involved in self-incompatibility reaction during flower fertilization [37] and the PAN domain is believed to mediate protein-protein and protein-carbohydrate interactions [38]. In some cases, G-type lectins have an EGF domain which may take part in the formation of disulfide bonds [16]. Multiple arrangements of these domains lead to various G-type lectins in *F. vesca* (Fig. 1).

2.1.2. Phylogenetic tree and nomenclature of G-type lectin genes

To highlight evolutionary differences, a phylogenetic tree of all 133 proteins was generated (Fig. 2). G-lectin genes are classified into six clades (I to VI). FvH4_1g03780 does not fall in any of these clades and it is designated singleton. All proteins in clade I have the same domain arrangement, GNA/PAN/TM/PK. In clade III, IV, and V, most of proteins are G-LecRKs with domain arrangement as GNA/SLG/PAN/TM/PK; while proteins of clade VI show the biggest diversity of domain arrangements, in total ten types of domain arrangements exist in this clade.
To make it convenient to refer to the *F. vesca* G-lectin genes, we propose a nomenclature based on the similarity shown in the phylogenetic tree (Table 1), where the genes included in each clade are named following a sequential numbering. In the name, letters “Fve” indicate the gene is from the organism *F. vesca* [39]; “GLRK”, “GLRP”, “GLP” and “GLK” represent G-LecRK, G-LecRP, G-LecP, and G-LecK, respectively. Since FvH4_1g03780 did not fall in any of the clades, it was named *FveGLRK7.1*, to distinguish it from genes in the six clades. These FveG-Lectins’ names will also be used herein the following sections.
Table 1
Proposed nomenclature for G-type lectin genes in *Fragaria vesca*

| Clade     | Gene ID       | New name   | Clade     | Gene ID       | New name   |
|-----------|---------------|------------|-----------|---------------|------------|
| Clade I   | FvH4_2g29560  | FveGLRK1.1 | Clade IV  | FvH4_6g00257  | FveGLRK4.27|
|           | FvH4_2g29545  | FveGLRK1.2 |           | FvH4_3g03322  | FveGLP4.5  |
|           | FvH4_2g29543  | FveGLRK1.3 |           | FvH4_3g03350  | FveGLRK4.28|
|           | FvH4_2g29544  | FveGLRK1.4 |           | FvH4_6g00300  | FveGLP4.6  |
|           | FvH4_2g29542  | FveGLRK1.5 |           | FvH4_3g06140  | FveGLRK4.29|
| Clade II  | FvH4_6g31370  | FveGLRK2.1 | Clade V   | FvH4_6g51830  | FveGLRK5.1 |
|           | FvH4_4g02170  | FveGLRK2.2 |           | FvH4_6g44064  | FveGLRK5.2 |
|           | FvH4_1g04840  | FveGLRK2.3 |           | FvH4_6g44101  | FveGLP5.1  |
|           | FvH4_3g15150  | FveGLP2.1  |           | FvH4_6g44100  | FveGLRK5.3 |
|           | FvH4_3g15130  | FveGLRK2.4 |           | FvH4_6g44062  | FveGLP5.2  |
|           | FvH4_3g15120  | FveGLRK2.5 |           | FvH4_6g44063  | FveGLRK5.4 |
|           | FvH4_3g15090  | FveGLRP2.1 |           | FvH4_6g44109  | FveGLRK5.5 |
|           | FvH4_3g15080  | FveGLRK2.6 |           | FvH4_6g44240  | FveGLK5.1  |
| Clade III | FvH4_3g03590  | FveGLRK3.1 |           | FvH4_6g44190  | FveGLRK5.9 |
|           | FvH4_2g24770  | FveGLP3.1  |           | FvH4_6g44107  | FveGLRK5.7 |
|           | FvH4_2g05942  | FveGLP3.2  |           | FvH4_6g44106  | FveGLRK5.8 |
|           | FvH4_3g03461  | FveGLRK3.2 |           | FvH4_6g44190  | FveGLRK5.9 |
|           | FvH4_3g03435  | FveGLP3.3  |           | FvH4_6g44242  | FveGLP5.3  |
|           | FvH4_3g03450  | FveGLRK3.3 |           | FvH4_6g44108  | FveGLRK5.10|
|           | FvH4_3g03451  | FveGLRK3.4 | Clade VI  | FvH4_6g44310  | FveGLRK6.1 |
|           | FvH4_3g03481  | FveGLRK3.5 |           | FvH4_6g26450  | FveGLRK6.2 |
|           | FvH4_3g03432  | FveGLRK3.6 |           | FvH4_6g26420  | FveGLRK6.3 |
|           | FvH4_3g03431  | FveGLRK3.7 |           | FvH4_7g14760  | FveGLRK6.4 |
|           | FvH4_3g03581  | FveGLRP3.1 | Clade VI  | FvH4_5g31680  | FveGLRK6.5 |
|           | FvH4_3g03560  | FveGLRK3.8 |           | FvH4_5g31690  | FveGLP6.1  |
|           | FvH4_5g32570  | FveGLRK3.9 |           | FvH4_2g26490  | FveGLRK6.6 |
|           | FvH4_3g03433  | FveGLRK3.10|           | FvH4_2g29070  | FveGLRK6.7 |
| Clade | Gene ID     | New name  | Clade | Gene ID     | New name  |
|-------|-------------|-----------|-------|-------------|-----------|
|       | FvH4_3g03521 | FveGLRK3.11 |       | FvH4_2g29050 | FveGLRK6.8 |
|       | FvH4_3g03520 | FveGLRK3.12 |       | FvH4_6g29840 | FveGLRK6.9 |
|       | FvH4_5g31930 | FveGLRK3.13 |       | FvH4_6g29821 | FveGLRK6.10 |
|       | FvH4_1g16211 | FveGLRK3.14 |       | FvH4_6g17930 | FveGLRP6.1 |
|       | FvH4_3g03502 | FveGLRK3.15 |       | FvH4_6g10470 | FveGLRP6.2 |
|       | FvH4_3g03501 | FveGLRK3.16 |       | FvH4_3g18370 | FveGLP6.2 |
|       | FvH4_3g03482 | FveGLRK3.17 |       | FvH4_3g18382 | FveGLP6.3 |
| Clade IV | FvH4_6g44243 | FveGLRK4.1 |       | FvH4_3g18380 | FveGLP6.4 |
|       | FvH4_6g44260 | FveGLP4.1 |       | FvH4_3g18371 | FveGLP6.5 |
|       | FvH4_6g44245 | FveGLRK4.2 |       | FvH4_3g18410 | FveGLP6.6 |
|       | FvH4_6g44244 | FveGLRK4.3 |       | FvH4_3g18383 | FveGLP6.7 |
|       | FvH4_3g03430 | FveGLRK4.4 |       | FvH4_7g30670 | FveGLRK6.11 |
|       | FvH4_3g03420 | FveGLRK4.5 |       | FvH4_6g26380 | FveGLRK6.12 |
|       | FvH4_3g21271 | FveGLP4.2 |       | FvH4_2g12390 | FveGLRK6.13 |
|       | FvH4_3g21270 | FveGLRK4.6 |       | FvH4_1g23400 | FveGLRK6.14 |
|       | FvH4_6g12332 | FveGLP4.3 |       | FvH4_1g23390 | FveGLRK6.15 |
|       | FvH4_3g21400 | FveGLRK4.7 |       | FvH4_2g14250 | FveGLRK6.16 |
|       | FvH4_3g21310 | FveGLRK4.8 |       | FvH4_1g23380 | FveGLRK6.17 |
|       | FvH4_3g15690 | FveGLRK4.9 |       | FvH4_1g23370 | FveGLRK6.18 |
|       | FvH4_3g21320 | FveGLRK4.10 |      | FvH4_7g00200 | FveGLRK6.19 |
|       | FvH4_3g15980 | FveGLK4.1 |       | FvH4_3g43403 | FveGLRK6.20 |
|       | FvH4_3g15930 | FveGLRK4.11 |      | FvH4_3g43400 | FveGLRK6.21 |
|       | FvH4_3g03310 | FveGLRK4.12 |      | FvH4_3g43402 | FveGLRK6.22 |
|       | FvH4_3g03243 | FveGLRK4.13 |      | FvH4_3g43710 | FveGLRK6.23 |
|       | FvH4_3g03301 | FveGLRK4.14 |      | FvH4_3g43401 | FveGLRK6.24 |
|       | FvH4_3g03300 | FveGLK4.2 |      | FvH4_4g33230 | FveGLRK6.25 |
|       | FvH4_3g03230 | FveGLRK4.15 |      | FvH4_6g12930 | FveGLRK6.26 |
|       | FvH4_3g03242 | FveGLRK4.16 |      | FvH4_6g12890 | FveGLRK6.27 |
| Clade   | Gene ID     | New name  | Clade   | Gene ID     | New name  |
|---------|-------------|-----------|---------|-------------|-----------|
| FvH4_6g07960 | FveGLRK4.17 |           | FvH4_5g04350 | FveGLRK6.28 |
| FvH4_3g03240  | FveGLRK4.18 |           | FvH4_5g04310 | FveGLP6.8   |
| FvH4_3g03231  | FveGLRK4.19 |           | FvH4_5g04270 | FveGLP6.9   |
| FvH4_7g19680  | FveGLRK4.20 |           | FvH4_6g12870 | FveGLRK6.29 |
| FvH4_3g03241  | FveGLK4.3  |           | FvH4_6g12920 | FveGLP6.10  |
| FvH4_3g03410  | FveGLRK4.21 |           | FvH4_6g12880 | FveGLRK6.30 |
| FvH4_3g03390  | FveGLRK4.22 |           | FvH4_2g33870 | FveGLRK6.31 |
| FvH4_3g03370  | FveGLRK4.23 |           | FvH4_2g33830 | FveGLRK6.32 |
| FvH4_6g20800  | FveGLRK4.24 |           | FvH4_2g33850 | FveGLRK6.33 |
| FvH4_3g03320  | FveGLRK4.25 |           | FvH4_2g33840 | FveGLRK6.34 |
| FvH4_3g03340  | FveGLP4.4   | singleton  | FvH4_1g03780 | FveGLRK7.1  |
| FvH4_6g00270  | FveGLRK4.26 |           |           |             |

### 2.1.3. Chromosome location and duplication of G-lectin genes

In order to visualize the chromosome location, G-lectin genes were mapped to *F. vesca* genome (Fig. 3). G-LecRKs are found distributed on all chromosomes, where the majority of G-LecRKs being on chromosomes 3 and 6, while G-LecPs are distributed on chromosomes 2, 3, 5, and 6. G-LecRPs are found only on chromosomes 3 and 6.

Tandem and dispersed duplication are two modes of single gene duplication: i) tandem duplication, which generates consecutive gene copies in the genome and is believed to originate from unequal chromosomal crossing over [40]; ii) dispersed duplication, which occurs with un-predictable and random patterns by still unclear mechanisms, resulting in two gene copies that are neither neighbors nor colinear [41]. Out of the 133 G-lectin genes, 86 of them appear as duplicated genes, with either tandem duplication (55 genes) or dispersed duplication (51 genes) or both (20 genes), showing that duplication events are common in G-lectin family of strawberry, thus leading to the expansion of G-lectin family in strawberry (Fig. 4). Genes from the same tandem duplicated cluster (Supplemental file S2) are usually close on the phylogenetic tree, dispersed duplicated genes are also close on the tree (Supplemental file S3). The duplication events of *F. vesca* G-lectin genes were visualized by chord diagram (Fig. 5). These duplications are not evenly distributed on seven chromosomes of *F. vesca*, chromosome 3 and chromosome 6 show more duplications compared with the other chromosomes. No tandem duplication of G-type lectin genes on chromosome 4 or chromosome 7 is found.
2.2. Kinase domain analysis of strawberry G-LecRK

The G-LecRKs kinase domain is a serine/threonine kinase type. Subdomains of kinases were defined according to typical patterns of conserved residues that are essential for functioning [42]. The eleven subdomains of FveG-LecRKs were analyzed by aligning the kinase subdomain motifs with the kinase domains of *Pi-d2* [4] and *OssIK2* [26] which were shown to be active and capable of phosphorylation (Supplemental file S4). In total, 21 G-LecRKs out of 102 were found with mutations in their kinase domains. The majority of mutated G-LecRKs were encoded by gene originating by tandem or dispersed duplication (Supplemental file S5). The main types of mutation are subdomain loss or amino acid substitution as summarized in Fig. 6. These occurred mainly in subdomain VIII to XI at the C-terminal of G-LecRKs, many of which were lost probably leading to inactive catalytic reaction (Supplemental file S5) whereas the subdomain III and IV of G-LecRKs in *F. vesca* appeared quite stable.

2.3. Gene expression

2.3.1. G-lectin gene expression during development and under stress conditions

The expression profile of G-lectin genes was analyzed in different tissues and at different developmental stages, based on available strawberry RNA-seq datasets. As shown in the heatmap (Fig. 7), G-type Lectin genes display a wide range of transcription levels among the different tissues, some of them are highly expressed in various tissues, while others completely silenced (Fig. 7). Among the highly expressed genes, many of them belong to the G-LecRK group. Few genes appear specifically expression only in one or two tissues, as, for instance, *FveGLRP2.1* which is highly expressed in pollen and anthers, but not in the other tissues. On the contrary, *FveGLP6.4*, the homolog of a known *F. x ananassa* G-type lectin, *FaMBL1*, showed active expression in many tissues during development. In general, G-lectin genes are more expressed in the ovary wall, seedling, style, root, and leaf, while they are hardly expressed in cortex and embryo (Fig. 7) The expression profile of those *FveG-LecRK* genes identified as encoding for mutated kinase domain as reported above, was also analyzed (Fig. 8). Basically, most of these genes are barely expressed during the entire developmental stages. However, the gene *FveGLRK6.12* had a relative high expression on several tissues and developmental stages, in spite of the mutations occurring in several conserved subdomain motifs. Genes like *FveGLRK1.3, FveGLRK6.15, FveGLRK2.2*, and *FveGLRK6.21* were expressed only in one or two tissues (Fig. 8).

The log2 fold changes of G-lectin genes’ expression under stresses were visualized in a heatmap. G-lectin genes were also found differently expressed during the interaction with pathogenic fungi, *B. cinerea, P. aphanis* as well as Oomycete, *P. cactorum* (Fig. 9). Around 50 G-lectin genes were differentially expressed during the interaction with *P. cactorum* in strawberry root and most of them were upregulated. Several G-lectin genes were also found upregulated in response to *P. aphanis* infection at 8 days post-infection. Compared to these two pathogens, few G-lectin genes were transcriptionally altered upon *B. cinerea* infection (Fig. 9). Moreover, some G-lectin genes appear to be regulated by plant resistance elicitor,
benzothiadiazole, and chitosan. Interestingly, genes like FveGLP6.4 (homolog of FaMBL1) were upregulated upon B. cinera, P. cactorum, and P. aphanis infection as well as by the inducers, benzothiadiazole, and chitosan. With regard to abiotic stress, cold stress caused both upregulation and downregulation of several G-lectin genes (Fig. 9).

### 2.3.2. G-lectin gene co-expression prediction

To further get insights into the function of strawberry G-type lectins, the genes co-expressed with FveG-lectin genes were retrieved from the co-expression database (www.fv.rosaceaefruits.org) [43]. Genes with functions in different plant defense pathways were predicted to co-express with G-lectin genes (Supplemental file S6). Some G-lectin genes, for instance, FveGLRK2.3 and FveGLRK4.18, were predicted to co-express with other receptor-like kinases including cysteine-rich RLK, Leucine-rich receptor-like protein kinase, receptor-like protein 1, receptor-like protein 7. G-type lectins such as FveGLRK1.1, FveGLRK4.13, and FveGLRK3.4 were predicted to co-express with disease resistance proteins, including NB-ARC domain-containing disease resistance protein, CC-NBS-LRR class family, TIR-NBS-LRR class family. FveGLRK6.24, FveGLRK6.22, and FveGLRK2.3, were predicted to co-express with many genes to make up a big defense co-expression network. For example, FveGLRK2.3 was predicted to co-express with genes coding cysteine-rich RLK, Leucine-rich RLK, glutamate receptor 2.7, beta-glucosidase 13, and many more (Supplemental file S6).

### 2.4. Subcellular location

The subcellular localization of strawberry G-lectin proteins was predicted by CELLO and TargetP (Supplemental file S7). CELLO uses a reliable index to compare the possibility of different subcellular locations. Based on the CELLO prediction, all G-LecRKs had a higher reliable index of being located at the plasma membrane except for FveGLRK6.31 and FveGLRK4.13, which were predicted to be located on the extracellular compartment. Conversely, almost all G-LecPs were located on the extracellular compartment. TargetP localization prediction is based on the presence of a signal peptide which drives proteins into the secretory pathway. According to this prediction, most F. vesca G-lectin proteins are driven to the secretory pathway, which is consistent with the prediction of CELLO, for which most of F. vesca G-lectin genes encode for proteins that are located on the plasma membrane or extracellular compartments.

### 3. Discussion

In plants, G-type lectin is a big gene family that is believed to play roles in biotic and abiotic stresses [44, 45]. Their role in defense was also reported in strawberry. For instance, 34 G-type LecRK genes were found upregulated in F. vesca root after P. cactorum inoculation [46], and the G-type lectin gene FaMBL1 was found involved in F. x ananassa resistance against C. acutatum [25]. A study about strawberry Serine/Threonine Kinase disease resistance gene family showed that many Serine/Threonine Kinase genes belong to G-type LecRK [47], but insights about the genomic organization of G-lectin proteins in strawberry was still limited. Recently, high-quality F. vesca genome annotation provided a good chance for the genome-wide study of G-lectin genes in F. vesca.
To identify G-lectin encoding genes, we used only sequences of the GNA domain as a query rather than the whole sequence. This choice was made to avoid using the kinase domain sequence as a query, which would lead to much ambiguity in G-lectin identification. Eventually, 133 proteins were found belonging to the G-lectin family in *F. vesca* and the majority (102 out of 133) of G-lectins contained kinase domain belonging to the G-LecRK class. Four genes containing both GNA and kinase domain, but lacking TM domain, were classified into G-LecK. The lack of TM domain may lead to function alteration of these G-lectins.

TM domains are required for the plasma membrane localization of G-LecRKs [4, 48]. In rice, a single amino acid substitution (Ile144Met) in the TM domain of *Pi-d2*, a rice G-LecRK conferring resistance to *M. grisea* strain ZB15, made the plant susceptible to the strain ZB15, suggesting that the TM domain of *Pi-d2* may participate in the ligand recognition and signal transduction [4]. Indeed, the substitution did not change the plasma membrane location of *Pi-d2*, so the altered structure of the mutated TM may have lost or modified its ligand-binding function and signal transduction from the extracellular domain to the intercellular kinase catalytic domain [4]. This fact implies that the TM domain of G-lectin has a role in both membrane localization and signal transduction. However, most of G-LecPs in *F. vesca*, despite lacking TM domains, were also predicted to anchor to the plasma membrane. In this regard, a pepper G-LecP, *CaMBL1*, consisting of GNA domain and PAN domain and regulating plant defense to bacterial *X. campestris pv vesicatoria*, was reported to be located on plasma membrane [14]. Moreover, the transient expression of *CaMBL1* induces the accumulation of salicylic acid and the activation of defense-related genes, which indicates a role in defense signaling, although without TM and kinase domain [14]. These data show that despite most of the previous studies on G-lectins focused on G-LecRKs, studies on G-LecPs could also cover important functions in plant.

The kinase domain of lectin receptor kinases could interact with downstream signaling molecules and display its catalytic activity [48]. The expression analysis of G-LecRKs with mutations in their kinase subdomain reveal expression for some of these lectins in spite of a predicted loss of kinase activity. Studies should be carried out to establish the importance of each amino acid residue in the kinase domain activity so to prove a relationship between conserved motifs and G-LecRK function.

Except for GNA, TM, and kinase domains, G-Lectins also contain some of SLG, PAN, and EGF domains. The various domain arrangements of G-lectins create an enormous degree of protein diversity. Proteins consisting of arrangements with PAN and SLG domains have GO functions related to the recognition of pollen, protein phosphorylation, and cell recognition which make these proteins important in reproduction and in general in signal perception or/and transduction [49]. Multiple domain proteins are more species-specific compared with single-domain proteins, which are commonly shared among many plant species [49]. In *F. vesca*, more than 90% of G-type lectins were found to belong to multiple domain proteins. These species-specific domain arrangements might be a consequence of frequent duplication events followed by lineage-specific retention [50]. This is consistent with our result where a big portion of *F. vesca* G-lectin genes appear to originate from duplication and various domain arrangements. The various domain
arrangements of G-type lectins could be considered as a kind of flexible genetic mechanism to produce species-specific adaptation to changing environments [49].

Tandem and dispersed duplication significantly contribute to the expansion of the G-lectin gene family in *F. vesca*. More than half of G-type lectin genes of *F. vesca* originate from duplication events. Chromosome 3, where the highest number of G-type lectin genes is located, showed a big number of duplication events of G-type lectin genes. Conversely, no tandem duplication event on G-type lectin genes on chromosome 4 and chromosome 7 was found and these two chromosomes also contain fewer G-type lectin genes than other chromosomes. Species-specific expansion of the G-type lectin gene family was also reported in a study about lectins in soybean, rice, and *Arabidopsis* where tandem and segmental duplications have been regarded as the major mechanisms to drive lectin expansion [30]. Consistently, a study about lectin genes in cucumber also revealed that 106 out of 146 genes (76.8%) were involved in the tandem duplication events [34].

According to the transcriptome data, many G-lectin genes, no matter G-LecRKs or G-LecPs, are actively expressed on different tissues at different developmental stages of strawberries. G-lectins in *F. vesca* actively respond to pathogens, abiotic stress, and elicitors; and some G-lectin genes appear to respond to both biotic and abiotic stress. Up to now, only one G-lectin gene, *FaMBL1* (homolog of *FveGLP6.4*) was studied for its involvement in resistance against pathogens in strawberries [25]; however, the molecular mechanism underneath is not yet elucidated. *FveGLP6.4* appears to be not only expressed in several tissues of strawberry during its development but also found upregulated after challenges by *B. cinerea*, *P. aphanis*, and *P. cactorum* (Fig. 7 and Fig. 9) pathogens, implying the involvement of *FveGLP6.4* in *F. vesca* (or *FaMBL1* in *F. x ananassa*) in plant defense.

The molecular features of some G-type lectins from other plant species are better known: *Pi-d2* [4], LORE [21], *OsSIK2* [26], and *CaMBL1* [14], which could regulate plant defense responses, were proved to be located at the plasma membrane by using confocal microscopy. For *CaMBL1*, its ability of mannose affinity and the importance of GNA domain for its localization are known [14]. According to the study, *CaMBL1* has affinity toward Manα and/or Manβ and GalNAc residues, and GNA domain is essential for its binding to D-mannose. A preliminary working model of *OslecRK* was also proposed by Cheng et al. [24]. Here sensing of biotic stress first stimulates *OslecRK* expression, followed by the interaction of its kinase domain with *OsADF* (actin-depolymerizing factor) to transduce the signals. Following these events, the expression of defense-related genes (*PR1a, LOX* and *CHS*) was induced to strengthen the plant’s immune response.

To further predict the function of G-lectins in *F. vesca*, we retrieved the genes predicted to co-express. G-lectin genes could co-express with other G-lectin genes, receptor kinase, and disease resistance genes which provides clues for uncovering their function. These data need to be proven experimentally.

4. Conclusion
In conclusion, G-type lectin is a big gene family in *F. vesca* with various domain arrangements and actively response to biotic/abiotic stresses. This indicates that G-lectins consist of an important gene family that requires further studies to understand their role in-depth, especially their involvement during biotic stress. Studying mannose-binding ability and identifying downstream interacting proteins of G-type lectins is important to uncover their role in strawberry defensing.

5. Materials And Methods

5.1. Identification and characterization of G-type lectin genes

5.1.1. Identification and domain organization

To identify G-type lectin genes, BLASTp search was performed first using GNA domain of FvH4_3g18380 (homolog of *FaMBL1*) as the query in the Genome Database for Rosaceae (GDR) (https://www.rosaceae.org/) [36] website and using the database Fragaria_vesca_v4.0.a2 proteins. Results with E-value < 1E-6 were considered as candidate G-type lectin proteins. With the same setting, a second BLASTp was conducted using GNA domains of G-lectin proteins with higher variety found in the first BLASTp (FvH4_1g23370, FvH4_2g12390, FvH4_2g14250, FvH4_2g26490, FvH4_2g29050, FvH4_2g33830, FvH4_3g03230, FvH4_3g03301, FvH4_3g03410, FvH4_3g03430, FvH4_3g06140, FvH4_3g15930, FvH4_3g18370, FvH4_3g21270, FvH4_3g43440, FvH4_4g02170, FvH4_5g31680, FvH4_6g00300, FvH4_6g12870, FvH4_6g44106). The domains of each candidate gene were checked manually by InterProScan (https://www.ebi.ac.uk/interpro/search/sequence/) [51] website. The transmembrane domain was checked by using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [52].

5.1.2. Phylogeny analysis

To build a phylogenetic tree, full-length protein sequences were obtained by running Blastx on GDR website, using mRNA sequences from GDR as a query to score protein database. Protein sequences were aligned using MUSCLE mode by MEGA-X [53]. Aligned sequences were analyzed via maximum likelihood bootstrapping (ML-BS) using IQ-TREE 1.6.12 (http://www.iqtree.org/) [54]. Once the best-fit model (WAG + F + I + G4) of molecular evolution was determined for G-type lectin genes, based on the Bayesian information criterion (BIC) scores [55], ML-BS analysis was conducted with IQ-TREE 1.6.12. Statistical support for the branches was evaluated by conducting a ML-BS bootstrap analysis of 5000 replicates. The tree was annotated by iTOL (https://itol.embl.de/) [56].

5.1.3. Chromosome location

The visualization of chromosome location of G-lectin genes was accomplished through MapGene2Chromosome V2 (http://mg2c.iask.in/mg2c_v2.0/) [57]. The location coordinates of G-lectin genes on the strawberry genome (*F. vesca* v4.0.a2) were obtained from GDR website.
5.2. Expansion and evolution of G-lectin genes

Gene duplication was investigated in *F. vesca* genome. Gene tandem duplication was explored using the PTGBase plant tandem duplicated gene database (http://ocri-genomics.org/PTGBase/) [58], and dispersed duplication was investigated on the plant duplicate gene (plantDGD) database (http://pdgd.njau.edu.cn:8080/) [41]. The amount of tandem and dispersed duplication was showed by Venn diagram using R package VennDiagram [59]. The relationships of duplicated gene pairs were visualized by chord diagram using R package circlize [60].

5.3. Kinase subdomain investigation of G-LecRKs

The kinase domains are divided into eleven (I-XI) smaller subdomains, according to Hanks and Hunter’s study [42]. The eleven subdomains of *F. vesca* G-LecRKs were analyzed by aligning their kinase domain protein sequences to kinase domains of known G-LecRKs, *Pi-d2* [4] and *OsSIK2* [26].

5.4. Expression analysis of G-type lectin genes

*F. vesca* G-lectin gene expression profiles were extracted from the database reported by Li and his colleagues [35]. The expression levels of the genes in different tissues: flowers, fruit of different developmental stages, seedlings, leaves, meristems, and roots were used to draw a heatmap through R package, ComplexHeatmap [61].

The transcriptome profiles of *F. x ananassa* strawberry infected by *Botrytis cinerea* [62, 63], *F. vesca* infected by *Phytophthora cactorum* [46], and by *Podosphaera aphanis* [64] were used to retrieve G-lectin gene expression. Besides this, transcriptome profile of *F. x ananassa* after cold stress [65] and preharvest application of benzothiadiazole (BTH) and chitosan [66] were also used to obtain G-lectin gene expression profile. G-lectin genes from *F. x ananassa* transcriptome datasets were converted to their *F. vesca* orthologs.

The co-expression genes of *F. vesca* G-lectin genes were retrieved from the co-expression database (www.fv.rosaceaefruits.org) [43] upon conversion of gene names from the previous genome annotations to the version 4.0.a2 [35], here used for lectin gene identification. Since different networks indicate varying correlation strengths, for G-lectin genes co-expression analysis, networks with the highest correlation were chosen: consensus100_hd_ltpm (consensus100 network of hand-dissected tissues), consensus100_lcm_ltpm (consensus100 network of laser captured tissues), and consensus100_fruit_ltpm (consensus100 network of ripening fruit tissue-only).

5.5. Subcellular localization prediction

For subcellular localization, amino acid sequences of strawberry G-lectin genes were submitted to two online predictors, TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP-1.1/index.php) [67] and “subCELlular LOcalization predictor” CELLO v.2.5 (http://cello.life.nctu.edu.tw/) [68]. Using the default setting, the predicted positions of G-lectin genes were obtained.
Abbreviations

CRA: homologs of class V Chitinases
EUL: Euonymus europaeus lectin
GNA: Galanthus nivalis agglutinin
JRL: Jacalin-related lectin
LysM: Lysin Motif
Nictaba: Nicotiana tabacum agglutinin
LecRKs: lectin receptor-like kinases
P/DAMP: pathogen-/damage-associated molecular pattern
SLG: S-locus glycoprotein domain
PAN: PAN/Apple domain
TM: transmembrane domain
PK: protein kinase domain
GDR: Genome Database for Rosaceae
EGF: Epidermal Growth Factor domain
BIC: Bayesian information criterion
BTH: benzothiadiazole

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
All data generated or analysed during this study are included in this published article and its supplementary information files.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This research was funded by Italian Ministry of Research and University (MIUR).

**Authors’ contributions**

LM, EB, and FN conceived and designed the research. LM and FN conducted bioinformatic analysis. LM wrote the paper. EB, FN and ZH revised the manuscript. EB funded this research. All authors have read and approved the manuscript.

**Acknowledgements**

LM is financially supported by China Scholarship Council.

**References**

1. Van Holle S, Van Damme EJM. Signaling through plant lectins: modulation of plant immunity and beyond. Biochem Soc Trans. 2018;46:217–33.

2. Van Holle S, Van Damme EJM. Messages from the past: New insights in plant lectin evolution. Front Plant Sci. 2019;10:36.

3. Van Damme EJM, Lannoo N, Peumans WJ. Plant Lectins. In: Kader J-C, Delseny M, editors. Advances in Botanical Research. Elsevier Masson SAS; 2008. p. 107–209.

4. Chen X, Shang J, Chen D, Lei C, Zou Y, Zhai W, et al. A B-lectin receptor kinase gene conferring rice blast resistance. Plant J. 2006;46:794–804.

5. Singh P, Chien CC, Mishra S, Tsai CH, Zimmerli L. The Arabidopsis LECTIN RECEPTOR KINASE- VI.2 is a functional protein kinase and is dispensable for basal resistance to Botrytis cinerea. Plant Signal Behav. 2013;8:e22611.

6. Sun XL, Yu QY, Tang LL, Ji W, Bai X, Cai H, et al. GsSRK, a G-type lectin S-receptor-like serine/threonine protein kinase, is a positive regulator of plant tolerance to salt stress. J Plant Physiol. 2013;170:505–15.
7. Guidarelli M, Zoli L, Orlandini A, Bertolini P, Baraldi E. The mannose-binding lectin gene FaMBL1 is involved in the resistance of unripe strawberry fruits to Colletotrichum acutatum. Mol Plant Pathol. 2014;15:832–40.

8. Kim NH, Lee DH, Choi DS, Hwang BK. The pepper GNA-related lectin and PAN domain protein gene, CaGLP1, is required for plant cell death and defense signaling during bacterial infection. Plant Sci. 2015;241:307–15.

9. Liu Y, Wu H, Chen H, Liu Y, He J, Kang H, et al. A gene cluster encoding lectin receptor kinases confers broad-spectrum and durable insect resistance in rice. Nat Biotechnol. 2015;33:301–7.

10. Van Holle S, Smagghe G, Van Damme EJM. Overexpression of Nictaba-Like Lectin Genes from Glycine max Confers Tolerance toward Pseudomonas syringae Infection, Aphid Infestation and Salt Stress in Transgenic Arabidopsis Plants. Front Plant Sci. 2016;7:1590.

11. Van Holle S, De Schutter K, Eggermont L, Tsaneva M, Dang L, Van Damme EJM. Comparative study of lectin domains in model species: New insights into evolutionary dynamics. Int J Mol Sci. 2017;18:1136.

12. Ghahremani M, Tran H, Biglou SG, O’Gallagher B, She YM, Plaxton WC. A glycoform of the secreted purple acid phosphatase AtPAP26 co-purifies with a mannose-binding lectin (AtGAL1) upregulated by phosphate-starved Arabidopsis. Plant Cell Environ. 2019;42:1139–57.

13. Hogervorst PAM, Ferry N, Gatehouse AMR, Wäckers FL, Romeis J. Direct effects of snowdrop lectin (GNA) on larvae of three aphid predators and fate of GNA after ingestion. J Insect Physiol. 2006;52:614–24.

14. Hwang IS, Hwang BK. The pepper mannose-binding lectin gene CaMBL1 is required to regulate cell death and defense responses to microbial pathogens. Plant Physiol. 2011;155:447–63.

15. Van Damme EJM, Allen AK, Peumans WJ. Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (Galanthus nivalis) bulbs. FEBS Lett. 1987;215:140–4.

16. Vaid N, Macovei A, Tuteja N. Knights in action: Lectin receptor-like kinases in plant development and stress responses. Mol Plant. 2013;6:1405–18.

17. Down RE, Gatehouse AMR, Hamilton WDO, Gatehouse JA. Snowdrop lectin inhibits development and decreases fecundity of the glasshouse potato aphid (Aulacorthum solani) when administered in vitro and via transgenic plants both in laboratory and glasshouse trials. J Insect Physiol. 1996;42:1035–45.

18. Luo S, Zhangsun D, Tang K. Functional GNA expressed in Escherichia coli with high efficiency and its effect on Ceratovacuna lanigera zehntner. Appl Microbiol Biotechnol. 2005;69:184–91.

19. Wang Z, Zhang K, Sun X, Tang K, Zhang J. Enhancement of resistance to aphids by introducing the snowdrop lectin gene gna into maize plants. J Biosci. 2005;30:627–38.

20. Miao J, Wu Y, Xu W, Hu L, Yu Z, Xu Q. The Impact of Transgenic Wheat Expressing Gna (snowdrop lectin) on the Aphids Sitobion avenae, Schizaphis graminum, and Rhopalosiphum padi. Environ Entomol. 2011;40:743–8.
21. Ranf S, Gisch N, Schäffer M, Illig T, Westphal L, Knirel YA, et al. A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in Arabidopsis thaliana. Nat Immunol. 2015;16:426–33.

22. Zeidler D, Zähringer U, Gerber I, Dubey I, Hartung T, Bors W, et al. Innate immunity in Arabidopsis thaliana: Lipopolysaccharides activate nitric oxide synthase (NOS) and induce defense genes. Proc Natl Acad Sci U S A. 2004;101:15811–6.

23. Silipo A, Molinaro A, Sturiale L, Dow JM, Erbs G, Lanzetta R, et al. The elicitation of plant innate immunity by lipooligosaccharide of Xanthomonas campestris. J Biol Chem. 2005;280:33660–8.

24. Cheng X, Wu Y, Guo J, Du B, Chen R, Zhu L, et al. A rice lectin receptor-like kinase that is involved in innate immune responses also contributes to seed germination. Plant J. 2013;76:687–98.

25. Guidarelli M, Zoli L, Orlandini A, Bertolini P, Baraldi E. The mannose-binding lectin gene FaMBL1 is involved in the resistance of unripe strawberry fruits to Colletotrichum acutatum. Mol Plant Pathol. 2014;15:832–40.

26. Chen LJ, Wuriyanghan H, Zhang YQ, Duan KX, Chen HW, Li QT, et al. An S-domain receptor-like kinase, OsSlK2, confers abiotic stress tolerance and delays dark-induced leaf senescence in rice. Plant Physiol. 2013;163:1752–65.

27. Hoorelbeke B, Van Damme EJM, Rougé P, Schols D, Van Laethem K, Fouquaert E, et al. Differences in the mannose oligomer specificities of the closely related lectins from Galanthus nivalis and Zea mays strongly determine their eventual anti-HIV activity. Retrovirology. 2011;8:10.

28. Houles Astoul C, Peumans WJ, Van Damme EJM, Rougé P. Accessibility of the high-mannose glycans of glycoprotein gp120 from human immunodeficiency virus type 1 probed by in vitro interaction with mannose-binding lectins. Biochem Biophys Res Commun. 2000;274:455–60.

29. Wu L, Bao JK. Anti-tumor and anti-viral activities of Galanthus nivalis agglutinin (GNA)-related lectins. Glycoconj J. 2013;30:269–79.

30. Jiang SY, Ma Z, Ramachandran S. Evolutionary history and stress regulation of the lectin superfamily in higher plants. BMC Evol Biol. 2010;10:79.

31. Vaid N, Pandey PK, Tuteja N. Genome-wide analysis of lectin receptor-like kinase family from Arabidopsis and rice. Plant Mol Biol. 2012;80:365–88.

32. Teixeira MA, Rajewski A, He J, Castaneda OG, Litt A, Kaloshian I. Classification and phylogenetic analyses of the Arabidopsis and tomato G-type lectin receptor kinases. BMC Genomics. 2018;19:239.

33. Saeed B, Baranwal VK, Khurana P. Identification and Expression Profiling of the Lectin Gene Superfamily in Mulberry. Plant Genome. 2016;9:plantgenome2015.10.0107.

34. Dang L, Van Damme EJM. Genome-wide identification and domain organization of lectin domains in cucumber. Plant Physiol Biochem. 2016;108:165–76.

35. Li Y, Pi M, Gao Q, Liu Z, Kang C. Updated annotation of the wild strawberry Fragaria vesca V4 genome. Hortic Res. 2019;6:61.

36. Jung S, Lee T, Cheng CH, Buble K, Zheng P, Yu J, et al. 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic Acids Res. 2019;47:D1137–45.
37. Kachroo A, Schopfer CR, Nasrallah ME, Nasrallah JB. Allele-specific receptor-ligand interactions in
Brassica self-incompatibility. Science. 2001;293:1824–6.

38. Tordai H, Bánya L, Patthy L. The PAN module: The N-terminal domains of plasminogen and
hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and
with a novel domain found in numerous nematode proteins. FEBS Lett. 1999;461:63–7.

39. Jung S, Bassett C, Bielenberg DG, Cheng CH, Dardick C, Main D, et al. A standard nomenclature for
gene designation in the Rosaceae. Tree Genet Genomes. 2015;11:108.

40. Wang Y, Wang X, Paterson AH. Genome and gene duplications and gene expression divergence: A
view from plants. Ann N Y Acad Sci. 2012;1256:1–14.

41. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, et al. Gene duplication and evolution in recurring
polyploidization-diploidization cycles in plants. Genome Biol. 2019;20:38.

42. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic)
domain structure and classification. FASEB J. 1995;9:576–96.

43. Shahan R, Zawora C, Wight H, Sittmann J, Wang W, Mount SM, et al. Consensus coexpression
network analysis identifies key regulators of flower and fruit development in wild strawberry. Plant
Physiol. 2018;178:202–16.

44. Ghequire MGK, Loris R, De Mot R. MMBL proteins: From lectin to bacteriocin. Biochem Soc Trans.
2012;40:1553–9.

45. Siripipatthana P, Phaonakrop N, Roytrakul S, Senawong G, Mudalige-Jayawickrama RG, Sattayasai
N. The first trimeric Galanthus nivalis agglutinin-related lectin of Orchidaceae was found in
Dendrobium pendulum: purification, characterization, and effects of stress factors. Plant Cell Rep.
2015;34:1253–62.

46. Toljamo A, Blande D, Kärenlampi S, Kokko H. Reprogramming of strawberry (Fragaria vesca) root
transcriptome in response to Phytophthora cactorum. PLoS One. 2016;11:e0161078.

47. Martínez Zamora MG, Castagnaro AP, Díaz Ricci JC. Genetic diversity of Pto-like serine/threonine
kinase disease resistance genes in cultivated and wild strawberries. J Mol Evol. 2008;67:211–21.

48. Sun Y, Qiao Z, Muchero W, Chen J-G. Lectin Receptor-Like Kinases: The Sensor and Mediator at the
Plant Cell. Front Plant Sci. 2020;11:596301.

49. Kersting AR, Bomberg-Bauer E, Moore AD, Grath S. Dynamics and adaptive benefits of protein
domain emergence and arrangements during plant genome evolution. Genome Biol Evol.
2012;4:316–29.

50. Paterson AH, Freeling M, Tang H, Wang X. Insights from the comparison of plant genome sequences.
Annu Rev Plant Biol. 2010;61:349–72.

51. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: Protein
domains identifier. Nucleic Acids Res. 2005;33:116–20.

52. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology
with a hidden Markov model: Application to complete genomes. J Mol Biol. 2001;305:567–80.
53. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.

54. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.

55. Chernomor O, Von Haeseler A, Minh BQ. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst Biol. 2016;65:997–1008.

56. Letunic I, Bork P. Interactive Tree of Life (iTOl) v4: Recent updates and new developments. Nucleic Acids Res. 2019;47:256–9.

57. Chao J, Kong Y, Wang Q, Sun Y, Gong D, Lv J, et al. MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages. Yi chuan = Hered. 2015;37:91–7.

58. Yu J, Ke T, Tehrim S, Sun F, Liao B, Hua W. PTGBase: An integrated database to study tandem duplicated genes in plants. Database. 2015;2015:bav017.

59. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.

60. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. Bioinformatics. 2014;30:2811–2.

61. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9.

62. Haile ZM, Nagpala-De Guzman EG, Moretto M, Sonego P, Engelen K, Zoli L, et al. Transcriptome Profiles of Strawberry (Fragaria vesca) Fruit Interacting With Botrytis cinerea at Different Ripening Stages. Front Plant Sci. 2019;10:1131.

63. Xiong JS, Zhu HY, Bai YB, Liu H, Cheng ZM. RNA sequencing-based transcriptome analysis of mature strawberry fruit infected by necrotrophic fungal pathogen Botrytis cinerea. Physiol Mol Plant Pathol. 2018;104:77–85.

64. Jambagi S, Dunwell JM. Global Transcriptome Analysis and Identification of Differentially Expressed Genes after Infection of Fragaria vesca with Powdery Mildew (Podosphaera aphanis). Transcriptomics. 2015;3:1000106.

65. Zhang Y, Zhang Y, Lin Y, Luo Y, Wang X, Chen Q, et al. A Transcriptomic Analysis Reveals Diverse Regulatory Networks That Respond to Cold Stress in Strawberry (Fragaria × ananassa). Int J Genomics. 2019;2019:7106092.

66. Landi L, De Miccolis Angelini RM, Pollastro S, Feliziani E, Faretra F, Romanazzi G. Global transcriptome analysis and identification of differentially expressed genes in strawberry after preharvest application of benzothiadiazole and chitosan. Front Plant Sci. 2017;8:235.

67. Emanuelsson O, Nielsen H, Brunak S, Von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 2000;300:1005–16.

68. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, et al. CELLO2GO: A web server for protein subCELlular IOcalization prediction with functional gene ontology annotation. PLoS One.
## Figures

### G-LecRK

|   | GNA | SLG | PAN | TM | PK |
|---|-----|-----|-----|----|----|
| 66|     |     |     |    |    |
| 13|     |     |     |    |    |
| 12|     |     |     |    |    |
| 11|     |     |     |    |    |

### G-LecK

|   | GNA | SLG | PAN | TM | PK |
|---|-----|-----|-----|----|----|
| 4 |     |     |     |    |    |

### G-LecRP

|   | GNA | SLG | PAN | TM |
|---|-----|-----|-----|----|
| 2 |     |     |     |    |
| 1 |     |     |     |    |
| 1 |     |     |     |    |

### G-LecP

|   | GNA | SLG | PAN |
|---|-----|-----|-----|
| 2 |     |     |     |
| 8 |     |     |     |
| 3 |     |     |     |
| 10|     |     |     |

---

**Figure 1**
Domain arrangements and classification of G-type lectins in Fragaria vesca. Numbers on the left indicate genes belonging to the type of domain organization. GNA: GNA domain; SLG: S-locus glycoprotein domain; PAN: PAN/Apple domain; PK: protein kinase domain; TM: transmembrane domain. Epidermal Growth Factor domain which overlaps with PAN domain and it is not shown in this figure.

**Figure 2**

Phylogenetic tree of Fragaria vesca G-type lectin genes. Different background colors indicate the six different clades. Symbols (circle, star, triangle, rectangular) at the end of each branch indicate the domain arrangement of each gene. Circle in different colors represent different domain arrangements of G-LecRKs; stars represent domain arrangement of G-LecKs; triangle in different colors represent different domain arrangement of G-LecRPs; rectangular in different colors represent different domain arrangements of G-LecPs. Numbers on each branch show the bootstrap.
Figure 3

Chromosome localization of G-LecRKs, G-LecKs, G-LecRPs, and G-LecPs in Fragaria vesca
Figure 4

Tandem and dispersed duplication of the G-lectin genes of Fragaria vesca.
Figure 5

Tandem and dispersed duplication events on G-type lectin genes among 7 chromosomes of Fragaria vesca. The different chromosomes are indicated in different colors. Genes connected by pink lines are tandem duplicates, while those lines connected by blue lines are dispersed duplicates.
Figure 6

Mutations occurred on 11 kinase subdomains of G-LecRKs in Fragaria vesca. Blue column shows the total number (102) of G-LecRKs in F. vesca; yellow column of each subdomain shows the number of G-LecRKs with mutation by amino acid substitution; red column shows the number of G-LecRKs with mutation of subdomain missing.
Figure 7

Expression profiles of G-type lectin genes in Fragaria vesca. The expression profiles of different types of G-lectin genes are indicated in different colors: G-LecRKs in black, G-LecPs in blue, G-LecRPs in red, and G-LecKs in green. The expression levels are indicated by TPM (transcript per million reads) and the TPM value for each gene is scaled before constructing the heatmap. Yw.22D: white fruit stage of Yellow Wonder 5AF7 (receptacle tissues at 22 days post-anthesis), Rg.22D: white fruit stage of Ruegen, Yw.15D:
green fruit stage of Yellow Wonder 5AF7 (receptacle tissues at 15 days post-anthesis), Rg.15D: green fruit stage of Ruegen, Wall: ovary wall, Ghost: entire seed with its embryo removed (seed coat), REM: receptacle meristem, FM: flower meristem, SAM: shoot apical meristem

Figure 8

Expression of G-LecRKs with a mutation in their conserved kinase subdomains.
Figure 9

Differently expressed strawberry G-lectin genes after challenged by pathogens and treated with inducers and cold. Data are expressed using log2 FC; negative value represents downregulation; positive value represents upregulation; and NDE represents no different expression. WF: white fruit; RF and red: red fruit; HWL: leaves of strawberry cv. Hawaii 4; YWL: leaves of Yellow Wonder 5AF7; Bc: Botrytis cinerea; Pc:
Phytophthora cactorum; Pa: Podosphaera aphanis; BTH: benzothiadiazole; CHI: chitosan; Cold_leaf: cold stressed leaf of seedlings; h/dpi: hours/days post-inoculation; hpt: hours post-treatment.

Supplementary Files

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

- Additionalfile1.xlsx
- Additionalfile2.pdf