



Genome-Wide Analysis of the *G2-Like* Transcription Factor Genes and Their Expression in Different Senescence Stages of Tobacco (*Nicotiana tabacum* L.)

Mingyue Qin^{1†}, Binghui Zhang^{2†}, Gang Gu², Jiazheng Yuan³, Xuansong Yang⁴, Jiahua Yang¹ and Xiaofang Xie^{1*}

¹ College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, China, ² Institute of Tobacco Science, Fujian Provincial Tobacco Company, Fuzhou, China, ³ Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC, United States, ⁴ Fujian Key Laboratory of Crop Breeding by Design, Fujian Agriculture and Forestry University, Fuzhou, China

OPEN ACCESS

Edited by:

Diane Maria Beckles,
University of California, Davis,
United States

Reviewed by:

Jingwei Yu,
Southern University of Science
and Technology, China
Ertugrul Filiz,
Duzce University, Turkey
Jun Zheng,
Chinese Academy of Agricultural
Sciences (CAAS), China

*Correspondence:

Xiaofang Xie
xxf317@fafu.edu.cn

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Genomics,
a section of the journal
Frontiers in Genetics

Received: 05 November 2020

Accepted: 22 April 2021

Published: 31 May 2021

Citation:

Qin M, Zhang B, Gu G, Yuan J,
Yang X, Yang J and Xie X (2021)
Genome-Wide Analysis of the *G2-Like*
Transcription Factor Genes and Their
Expression in Different Senescence
Stages of Tobacco (*Nicotiana*
tabacum L.).
Front. Genet. 12:626352.
doi: 10.3389/fgene.2021.626352

The *Golden2-like* (GLK) transcription factors play important roles in regulating chloroplast growth, development, and senescence in plants. In this study, a total of 89 *NtGLK* genes (*NtGLK1–NtGLK89*) were identified in the tobacco genome and were classified into 10 subfamilies with variable numbers of exons and similar structural organizations based on the gene structure and protein motif analyses. Twelve segmental duplication pairs of *NtGLK* genes were identified in the genome. These *NtGLK* genes contain two conserved helix regions related to the HLH structure, and the sequences of the first helix region are less conserved than that of the second helix motif. *Cis*-regulatory elements of the *NtGLK* promoters were widely involved in light responsiveness, hormone treatment, and physiological stress. Moreover, a total of 206 *GLK* genes from tomato, tobacco, maize, rice, and *Arabidopsis* were retrieved and clustered into eight subgroups. Our gene expression analysis indicated that *NtGLK* genes showed differential expression patterns in tobacco leaves at five senescence stages. The expression levels of six *NtGLK* genes in group C were reduced, coinciding precisely with the increment of the degree of senescence, which might be associated with the function of leaf senescence of tobacco. Our results have revealed valuable information for further functional characterization of the *GLK* gene family in tobacco.

Keywords: gene expression, senescence, phylogenetic analysis, *G2-like* transcription factors, *Nicotiana tabacum*

INTRODUCTION

The GOLDEN2-LIKE (GLK) proteins are included in the GARP (Golden2, ARR-B and Psr1) domain superfamily of transcription factors (TFs) (Riechmann et al., 2000; Xiao et al., 2019). The GLK transcription factor was originally identified in the C4 plant of maize (*Zea mays* L.) (Hall et al., 1998). A typical GLK protein usually contains two conserved domains, namely, a Myb-DNA binding domain (DBD) and a C-terminal (GCT) box (Rossini et al., 2001).

Members of the GLK family play important roles in the formation and development of chloroplasts (Rossini et al., 2001; Fitter et al., 2002; Waters et al., 2008; Powell et al., 2012;

Jarvis and López-Juez, 2013) and have been involved in the various defense processes of organisms, including biotic and abiotic stresses (Savitch et al., 2007; Schreiber et al., 2011; Murmu et al., 2014; Han et al., 2016; Nagatoshi et al., 2016). In *Arabidopsis* (*Arabidopsis thaliana*), the *AtGLK1* and *AtGLK2* genes act redundantly to regulate chloroplast development (Fitter et al., 2002; Yasumura et al., 2005). *AtGLK1* overexpression enhances resistance to the pathogen *Fusarium graminearum* (Savitch et al., 2007; Schreiber et al., 2011) and increases susceptibility to the virulent oomycete pathogen *Hyaloperonospora arabidopsidis* (*Hpa*) (Murmu et al., 2014), whereas the *glk1 glk2* double mutant increases the resistance to the *Hpa* gene compared to that of the wild type (Murmu et al., 2014). In addition, it has been reported that GLKs can interact with ANAC092 (ORE-1) to regulate leaf senescence (Rauf et al., 2013). In tomato (*Solanum lycopersicum*), the overexpression of *GLK* increases the expressions of genes related to chloroplast development and fruit photosynthesis, and these changes result in the enhancement of the carbohydrate and carotenoid levels in ripe fruits (Powell et al., 2012). In maize, *ZmGlk1* has been considered to play important roles in the chloroplast development of mesophyll cell in C4 plant tissues (Rossini et al., 2001; Fitter et al., 2002).

The senescence of tobacco (*Nicotiana tabacum*) leaves is a positive and orderly process involving the transformation and mobilization of nutrients. As a model plant, the study of the plant senescence and internal material transport rules in tobacco has special significance in the research of plant physiology and development (Gregersen et al., 2013). During the phase of leaf senescence, leaf cells undergo orderly changes in structure, metabolism, and gene expression, along with a series of degradations, including the chlorophyll content depletion and diminishing photosynthetic capacity (Lira et al., 2017; Schippers et al., 2015; Gan and Amasino, 1997; Lim et al., 2003). Normally, leaf color change is the most intuitive phenomenon, which is due to the degradation of chlorophyll in chloroplasts (Hörtensteiner, 2006; Kräutler, 2016); therefore, the degree of etiolation has usually been used as an important criterion to assess the senescence of leaves. Because of the importance of the members of the *GLK* gene family for the development of chloroplasts (Fitter et al., 2002; Waters et al., 2008; Powell et al., 2012) and its association with leaf senescence (Rauf et al., 2013), it is important to investigate this gene family and assess the relationship between gene expression and senescence in leaf.

The *GLK* gene family has been identified and characterized in several plant species, including maize (*Zea mays* L.) (Liu et al., 2016) and tomato (Liu, 2018). However, the *GLK* gene family has not been thoroughly examined in tobacco, to the best of our knowledge. A comprehensive investigation of all the *GLK* genes with current tobacco genome sequence data, including the family members and the detailed organization of the gene sequences in tobacco, should be conducted. The objectives of this study were to analyze the *GLK* gene family including the *GLK* gene structure, chromosomal localization, and phylogenetic relationship in the tobacco genome and to reveal the expression regulation of the *GLK* gene family members at different senescence stages of tobacco leaves. The information derived from this study will be useful for further functional exploration on the *GLK* gene family in tobacco.

MATERIALS AND METHODS

Identification of *GLK* Genes in Tobacco

The published *GLK* protein sequences of maize (Liu et al., 2016), rice (*Oryza sativa*) (Rossini et al., 2001), and *A. thaliana* (Fitter et al., 2002) were used as query sequences to identify the *GLK* proteins in tobacco using the BLASTP tool and the tobacco genome sequences (Edwards et al., 2017) in the Sol Genomics Network database¹. More than 30% similarity and an *E* value less than E^{-10} were set as the parameters to define the tobacco candidate *GLK* proteins. The domains of all the candidate *GLK* proteins of tobacco were checked using the Conserved Domain Database (CDD) tool² (Lu et al., 2020). Finally, the sequences with complete *GLK* domains were retained and were renamed (*NtGLK*). Detailed information of the *NtGLK* genes, including the gene IDs, physical position, sequences of the genes and proteins, and the coding sequences (CDS), were retrieved from the Sol Genomics Network database³. The features of the *NtGLK* proteins were calculated using online ExpASY programs⁴ (Bjellqvist et al., 1993; Bjellqvist et al., 1994; Wilkins et al., 1999).

Multiple Sequence Alignment and Phylogenetic Analysis

The amino acid sequences of all the tobacco *GLK* proteins were aligned using the ClustalX1.83 tool (Thompson et al., 1997). The alignment of the *NtGLK* protein conserved domain sequences was exhibited by DNAMAN⁵ (Altschul et al., 1990). A phylogenetic tree with 1,000 bootstrap replicates was generated using the neighbor-joining method of the MEGAX software (Kumar et al., 2018). The classifications of tobacco *GLK* proteins were determined according to the topology and bootstrap values of the phylogenetic tree.

Chromosomal Location and Gene Duplication

Information on the physical position image of the *NtGLK* genes was obtained based on the MapInspect tool⁶. To investigate gene duplication, the criteria for the proportion of overlap and the similarity between the two sequences were set to be >70% (Gu et al., 2002; Yang et al., 2008). Segmental duplication and tandem duplication were defined based on the method reported by Wang et al. (2010).

Gene Structure and Conserved Motif Identification

The gene structures of the *NtGLK* genes were identified by the GSDS⁷ platform using the complete sequence of the genomic sequence and CDS for *NtGLK* downloaded from the tobacco genome (Hu et al., 2015). The conserved motifs of the *NtGLK*

¹https://solgenomics.net/organism/Nicotiana_attenuata/genome

²<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

³<https://solgenomics.net>

⁴<http://web.expasy.org/protparam/>

⁵<http://dnaman.software.informer.com/>

⁶http://www.plantbreeding.wur.nl/uk/software_mapinspect.html

⁷<http://gsds.cbi.pku.edu.cn/>

proteins were analyzed by the MEME program⁸ (Ma et al., 2014), and the optimum motif width and the maximum number of motifs were set to 5–100 and 20 residues, respectively, with the remaining parameters in default. Motif annotation was identified using the CDD tools. About 1,500 bp of DNA sequence upstream of the starting codon of the *NtGLK* genes were extracted from the tobacco genome to decipher the *cis*-elements with the online tool PlantCARE⁹ (Lescot et al., 2002) for *cis*-element prediction and the TBtools software (Chen et al., 2020) for the visualization.

***NtGLK* Expression Analysis in Different Senescence Stages of Tobacco Leaves**

Nicotiana tabacum cv. Cuibi 1 (CB-1) was used in this study. According to the morphological characteristics of leaf color, vein, and villus (Xu et al., 2017), five senescence stages (M1, M2, M3, M4, and M5) of the middle leaves (eighth to 10th, counted from the bottom to the top) were collected in the same field as samples. Three biological replicates with each replicate containing three leaves from different plant were collected. Total RNA was extracted using a total RNA isolation kit (PR2401, Biotek Corporation, China). A total of 15 RNA samples were sequenced on Illumina HiSeq 2000 performed by BioMarker Technologies¹⁰ (BioMarker, Beijing, China). The gene expression level was assessed according to the FPKM (fragments per kilobase of transcript sequence per million base pairs sequenced) value (Trapnell et al., 2010). The complementary DNA (cDNA) samples were synthesized by a SMART cDNA synthesis kit, and quantitative reverse transcriptase (qRT-PCR) reactions were conducted with the SYBR Premix Ex Taq based on the manufacturer's instruction (Takara). The primers used for qRT-PCR analysis are listed in **Supplementary Table 1**. The specific exon regions of the target genes were used for primer design. The primer pair of each target gene was also prescreened to ensure the uniqueness of the amplification product. Three biological replicates and three technical replicates were performed. The expression level of each selected *NtGLK* gene in the M1 stage was used as the control, while the relative expression level of the *NtGLK* genes in the different senescence stages was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). A *t* test was conducted to assess the expression differences from the M1 stage to the M2, M3, M4, and M5 stages. Significant difference was set as $p < 0.05$ or $p < 0.01$.

RESULTS

Characterization of *GLK* Genes in Tobacco

A BLASTP search was performed using 117 known *GLK* protein sequences as the query sequences against the tobacco genome database to analyze the *GLK* genes in tobacco, including 59 from maize (Liu et al., 2016), 54 from tomato (Liu et al., 2016), two from rice (Rossini et al., 2001), and two from *A. thaliana*

(Fitter et al., 2002). A total of 89 *GLK* genes were identified in tobacco, and these genes were named *NtGLK1* through *NtGLK89*. Detailed information of these genes and their corresponding proteins are listed in **Table 1**, including the gene ID, gene location, number of exons, protein length (amino acids, aa), molecular mass (MS), and *pI*. The *NtGLK* genes showed a wide range of amino acid sequence lengths and molecular weights. The amino acid residues of *NtGLK* proteins were oscillating from 114 aa (*NtGLK31*) to 779 aa (*NtGLK41*), with an average of 408 amino acids, whereas the molecular mass was from 13,358.46 Da (*NtGLK31*) to 85,665.25 Da (*NtGLK41*) and the theoretical isoelectric points (*pI*) range from 4.81 (*NtGLK49*) to 9.91 (*NtGLK31*).

Structure and Phylogenetic Tree of Tobacco *GLK* Gene Members

To explore the evolutionary relationships among the tobacco *GLK* genes, an unrooted phylogenetic tree was generated using the 89 tobacco *GLK* protein sequences (**Figure 1A**); moreover, the gene structures for each *GLK* gene were analyzed (**Figure 1B**). According to the results of the gene structure and the bootstrap values (>50%) of the phylogenetic tree, the tobacco *GLK* gene family was grouped into 10 subfamilies (I to X). However, there were two genes (*NtGLK82* and *NtGLK88*) that could not be clustered into any of the 10 subfamilies because of their low bootstrap values (<50%). Among these subfamilies, subfamily I (containing 30 members) was the largest group and represented more than 30% of the total *NtGLK* members. In contrast, subfamilies III and VIII only contained two members each.

Structure analysis of the *NtGLK* genes (**Figure 1B**) showed that the number of introns in the subfamilies ranged from 0 to 11. Among them, five genes (*NtGLK12*, *NtGLK73*, *NtGLK7*, *NtGLK80*, and *NtGLK45*) clustered into subfamily I, which did not contain any intron, and two genes (*NtGLK4* and *NtGLK63*) contained only one intron. Most of the *NtGLK* genes that were clustered into the same phylogenetic groups showed similar exon/intron structures, including the intron numbers and exon length. Variations in the intron number might be one of the key factors that resulted in the diversity of the gene structure and function in the course of evolution.

Motif Analysis of *NtGLK* Proteins

The conserved motifs of the 89 *NtGLK* proteins within each subfamily were identified and analyzed using the online MEME tool (**Figure 2**). A total of 20 motifs were identified; the detailed conserved sequences of each motif are shown in **Supplementary Table 2**. With the CDD tool, seven putative motifs were functionally annotated, which were defined as Myb-SHAQKYF for motif 1, components of the conserved *GLK* domain for motifs 2 and 11, Myb-CC-LHEQLE for motif 3, and REC superfamily for motifs 4, 5, and 8. However, no functional annotation was assigned for the remaining 13 putative motifs. The motif of Myb-SHAQKYF is highly conserved in a number of myb-related genes. It was reported that the conserved motif of SHAQKYF could bind to the I-box with a DNA-binding domain located in the carboxy terminal domain and acted as a transcriptional activator

⁸<http://alternate.meme-suite.org/tools/meme>

⁹<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>

¹⁰<http://www.biomarker.com.cn/>

TABLE 1 | The *GLK* gene family in *Nicotiana tabacum* L.

Gene name	Gene ID	Chr./scaffold	Exon	Protein length (aa)	Molecular weight (Da)	pI
<i>NtGLK1</i>	Nitab4.5_0000016g0360.1	Nt04	12	588	67,407.59	6.82
<i>NtGLK2</i>	Nitab4.5_0000058g0230.1	Nt04	6	391	44,736.55	8.72
<i>NtGLK3</i>	Nitab4.5_0000061g0100.1	Nitab4.5_0000061	7	464	51,107.17	5.92
<i>NtGLK4</i>	Nitab4.5_0000102g0300.1	Nt17	2	334	36,628.15	5.95
<i>NtGLK5</i>	Nitab4.5_0000109g0060.1	Nt04	6	415	47,021.29	9.26
<i>NtGLK6</i>	Nitab4.5_0000109g0160.1	Nt04	6	281	31,575.53	8.56
<i>NtGLK7</i>	Nitab4.5_0000123g0580.1	Nt24	1	300	32,867.68	6.45
<i>NtGLK8</i>	Nitab4.5_0000147g0130.1	Nitab4.5_0000147	6	316	34,387.78	5.95
<i>NtGLK9</i>	Nitab4.5_0000388g0040.1	Nt24	7	362	40,301.64	8.58
<i>NtGLK10</i>	Nitab4.5_0000440g0010.1	Nt03	5	376	42,086.47	6.78
<i>NtGLK11</i>	Nitab4.5_0000444g0200.1	Nt24	5	312	35,385.97	8.99
<i>NtGLK12</i>	Nitab4.5_0000463g0190.1	Nt22	1	311	34,694.66	5.95
<i>NtGLK13</i>	Nitab4.5_0000476g0270.1	Nitab4.5_0000476	7	473	52,022.22	5.8
<i>NtGLK14</i>	Nitab4.5_0000543g0010.1	Nt23	6	255	28,440.86	6.67
<i>NtGLK15</i>	Nitab4.5_0000570g0280.1	Nitab4.5_0000570	5	510	57,147.73	6.45
<i>NtGLK16</i>	Nitab4.5_0000605g0080.1	Nt17	6	419	48,594.05	8.53
<i>NtGLK17</i>	Nitab4.5_0000621g0130.1	Nt02	7	484	52,492.13	6.11
<i>NtGLK18</i>	Nitab4.5_0000629g0030.1	Nt07	6	655	72,665.22	5.8
<i>NtGLK19</i>	Nitab4.5_0000672g0080.1	Nitab4.5_0000672	5	380	41,716.93	6.61
<i>NtGLK20</i>	Nitab4.5_0000676g0220.1	Nt15	6	297	32,499.06	7.58
<i>NtGLK21</i>	Nitab4.5_0000736g0030.1	Nitab4.5_0000736	5	415	47,115.15	6.25
<i>NtGLK22</i>	Nitab4.5_0000850g0070.1	Nt20	6	247	26,811.33	5.61
<i>NtGLK23</i>	Nitab4.5_0000916g0020.1	Nt12	4	582	63,275.73	5.99
<i>NtGLK24</i>	Nitab4.5_0000969g0020.1	Nt06	6	390	44,622.39	8.44
<i>NtGLK25</i>	Nitab4.5_0001083g0040.1	Nt04	3	558	61,358.02	8.95
<i>NtGLK26</i>	Nitab4.5_0001088g0210.1	Nt11	6	427	47,155.25	7.79
<i>NtGLK27</i>	Nitab4.5_0001094g0080.1	Nt09	5	667	73,495.12	7.27
<i>NtGLK28</i>	Nitab4.5_0001097g0090.1	Nt13	6	678	74,931.11	5.97
<i>NtGLK29</i>	Nitab4.5_0001164g0090.1	Nitab4.5_0001164	4	420	46,197.33	6.6
<i>NtGLK30</i>	Nitab4.5_0001318g0010.1	Nt05	5	545	60,889.33	6.17
<i>NtGLK31</i>	Nitab4.5_0001409g0040.1	Nitab4.5_0001409	4	114	13,358.46	9.91
<i>NtGLK32</i>	Nitab4.5_0001416g0010.1	Nt07	6	291	32,167.27	6.01
<i>NtGLK33</i>	Nitab4.5_0001534g0020.1	Nitab4.5_0001534	7	297	32,419.29	6.46
<i>NtGLK34</i>	Nitab4.5_0001632g0010.1	Nt08	6	405	45,548.54	9.02
<i>NtGLK35</i>	Nitab4.5_0001663g0280.1	Nt23	5	276	30,025.02	6.86
<i>NtGLK36</i>	Nitab4.5_0001777g0040.1	Nt16	6	323	37,180.25	7.66
<i>NtGLK37</i>	Nitab4.5_0001933g0010.1	Nitab4.5_0001933	3	338	37,900.5	6.14
<i>NtGLK38</i>	Nitab4.5_0002024g0040.1	Nt02	6	296	33,617.42	9.18
<i>NtGLK39</i>	Nitab4.5_0002055g0150.1	Nt23	11	563	62,891.85	6.06
<i>NtGLK40</i>	Nitab4.5_0002076g0010.1	Nt14	5	278	32,753.05	9.76
<i>NtGLK41</i>	Nitab4.5_0002117g0090.1	Nitab4.5_0002117	7	779	85,665.25	6.48
<i>NtGLK42</i>	Nitab4.5_0002389g0010.1	Nt22	7	464	51,049.8	5.01
<i>NtGLK43</i>	Nitab4.5_0002462g0050.1	Nitab4.5_0002462	5	398	44,288.65	9.23
<i>NtGLK44</i>	Nitab4.5_0002465g0020.1	Nt13	6	430	47,599.6	7.09
<i>NtGLK45</i>	Nitab4.5_0002606g0020.1	Nt05	1	304	33,054.03	6.54
<i>NtGLK46</i>	Nitab4.5_0002803g0020.1	Nitab4.5_0002803	5	654	71,278.23	6.3
<i>NtGLK47</i>	Nitab4.5_0002948g0070.1	Nitab4.5_0002948	8	296	32,331.24	6.26
<i>NtGLK48</i>	Nitab4.5_0003100g0090.1	Nitab4.5_0003100	11	570	63,380.74	6.24
<i>NtGLK49</i>	Nitab4.5_0003295g0220.1	Nt22	4	303	32,814.65	4.81
<i>NtGLK50</i>	Nitab4.5_0003484g0070.1	Nitab4.5_0003484	5	493	55,301.42	6.26
<i>NtGLK51</i>	Nitab4.5_0003610g0020.1	Nitab4.5_0003610	7	440	49,557.81	6.47
<i>NtGLK52</i>	Nitab4.5_0003711g0030.1	Nt21	5	283	33,229.33	9.11
<i>NtGLK53</i>	Nitab4.5_0003726g0010.1	Nitab4.5_0003726	5	398	44,328.48	8.75
<i>NtGLK54</i>	Nitab4.5_0003836g0060.1	Nitab4.5_0003836	5	301	33,712.65	9.77

(Continued)

TABLE 1 | Continued

Gene name	Gene ID	Chr./scaffold	Exon	Protein length (aa)	Molecular weight (Da)	pI
<i>NtGLK55</i>	Nitab4.5_0003856g0030.1	Nt18	5	327	37,450.18	8.2
<i>NtGLK56</i>	Nitab4.5_0003889g0040.1	Nitab4.5_0003889	6	663	73,098.64	7.25
<i>NtGLK57</i>	Nitab4.5_0004143g0010.1	Nitab4.5_0004143	6	297	33,751.65	9.11
<i>NtGLK58</i>	Nitab4.5_0004327g0030.1	Nitab4.5_0004327	11	563	62,486.11	6.3
<i>NtGLK59</i>	Nitab4.5_0004550g0020.1	Nt19	6	601	68,120.64	5.4
<i>NtGLK60</i>	Nitab4.5_0004560g0010.1	Nitab4.5_0004560	6	563	62,408.53	5.33
<i>NtGLK61</i>	Nitab4.5_0004658g0020.1	Nitab4.5_0004658	11	568	63,109.41	6.38
<i>NtGLK62</i>	Nitab4.5_0004835g0040.1	Nt08	5	261	29,209.91	9.43
<i>NtGLK63</i>	Nitab4.5_0004892g0040.1	Nitab4.5_0004892	2	335	36,619.21	6.26
<i>NtGLK64</i>	Nitab4.5_0004991g0030.1	Nt05	8	478	54,910.12	5.77
<i>NtGLK65</i>	Nitab4.5_0005180g0030.1	Nitab4.5_0005180	6	670	73,859.9	5.97
<i>NtGLK66</i>	Nitab4.5_0005194g0010.1	Nitab4.5_0005194	5	398	43,657.14	6.98
<i>NtGLK67</i>	Nitab4.5_0005233g0020.1	Nitab4.5_0005233	6	573	63,383.99	5.68
<i>NtGLK68</i>	Nitab4.5_0006031g0030.1	Nitab4.5_0006031	5	195	21,865.34	6.5
<i>NtGLK69</i>	Nitab4.5_0006133g0010.1	Nitab4.5_0006133	7	334	36,245.31	6.27
<i>NtGLK70</i>	Nitab4.5_0006324g0050.1	Nitab4.5_0006324	6	339	46,485.81	9.13
<i>NtGLK71</i>	Nitab4.5_0006629g0020.1	Nitab4.5_0006629	7	443	48,598	5.68
<i>NtGLK72</i>	Nitab4.5_0006900g0010.1	Nitab4.5_0006900	6	294	33,016.07	9.11
<i>NtGLK73</i>	Nitab4.5_0006916g0030.1	Nitab4.5_0006916	1	311	34,696.8	6.42
<i>NtGLK74</i>	Nitab4.5_0006963g0020.1	Nt20	3	244	26,200.36	6.59
<i>NtGLK75</i>	Nitab4.5_0006980g0060.1	Nitab4.5_0006980	7	455	49,707.14	5.06
<i>NtGLK76</i>	Nitab4.5_0007123g0020.1	Nitab4.5_0007123	7	440	50,669.27	5.73
<i>NtGLK77</i>	Nitab4.5_0007572g0030.1	Nitab4.5_0007572	6	332	37,365.71	9.66
<i>NtGLK78</i>	Nitab4.5_0007848g0010.1	Nitab4.5_0007848	7	351	39,045.15	7.74
<i>NtGLK79</i>	Nitab4.5_0008054g0010.1	Nitab4.5_0008054	7	333	36,183.32	6.22
<i>NtGLK80</i>	Nitab4.5_0008332g0020.1	Nitab4.5_0008332	1	300	32,941.71	6.45
<i>NtGLK81</i>	Nitab4.5_0008336g0010.1	Nitab4.5_0008336	4	421	46,640.87	6.54
<i>NtGLK82</i>	Nitab4.5_0008908g0020.1	Nitab4.5_0008908	7	273	31,074.39	9.15
<i>NtGLK83</i>	Nitab4.5_0009217g0030.1	Nitab4.5_0009217	5	294	33,105.58	9.47
<i>NtGLK84</i>	Nitab4.5_0010430g0010.1	Nitab4.5_0010430	6	601	67,954.72	5.69
<i>NtGLK85</i>	Nitab4.5_0010689g0010.1	Nitab4.5_0010689	6	488	53,357.22	6.29
<i>NtGLK86</i>	Nitab4.5_0011083g0010.1	Nitab4.5_0011083	7	407	46,360.82	7.64
<i>NtGLK87</i>	Nitab4.5_0012578g0010.1	Nitab4.5_0012578	5	342	38,135.65	6.07
<i>NtGLK88</i>	Nitab4.5_0012878g0020.1	Nitab4.5_0012878	5	186	21,113.57	6.31
<i>NtGLK89</i>	Nitab4.5_0014621g0010.1	Nitab4.5_0014621	7	466	51,325.56	6.26

in yeast and plants (Rose et al., 2000). The motif of Myb-CC-LHEQLE was found toward the C-terminus of Myb-CC-type transcription factors, the member of the protein family reported to be involved in phosphate starvation signaling both in vascular plants and in unicellular algae (Rubio et al., 2001). The functions of the REC domains are annotated as phosphorylation-mediated switches within response regulators (RRs), and some also transfer phosphoryl groups in multistep phosphorelays.

The majority of the *NtGLK* proteins contained motifs 1 and 2 (Figure 2). In addition, the *NtGLK* protein members grouped into the same subfamily contained similar motif components and spatial distributions. Moreover, most of the members of groups III, IV, V, VI, and VII possessed a Myb-CC-LHEQLE domain (motif 3: a type of Myb-like domain), which appeared to respond to various abiotic stresses and played diverse roles in plant development (Murmu et al., 2014). This result suggested that the *NtGLK* proteins in the same subfamily might have similar functions. In addition, the specificity within a subfamily

was also identified. For instance, motif 15 is only possessed by subfamily II and motif 16 only appeared in subfamily I (Figure 2). To further decipher the similarity among the tobacco *GLK* domains, 89 tobacco *GLK* domain sequences were aligned using the DNAMAN 8¹¹ platform (Figure 3). Our results showed that the *GLKs* contained two regions of a putative DNA-binding domain with an HLH structure, which was also identified in maize (Liu et al., 2016). In this putative domain, the first helix had initial sequences of PELHRR and the second helix contained the conserved NI/VASHLQ. In addition, although the sequences had particularly conserved L and H, some variants were found for the *GLK* gene members of tobacco in this motif. The second helix region contained a highly conserved sequence of VK/VASHLQ (Liu et al., 2016), which was also similar to the *GLK* members in tomato (Liu, 2018).

¹¹<http://dnaman.software.informer.com/>

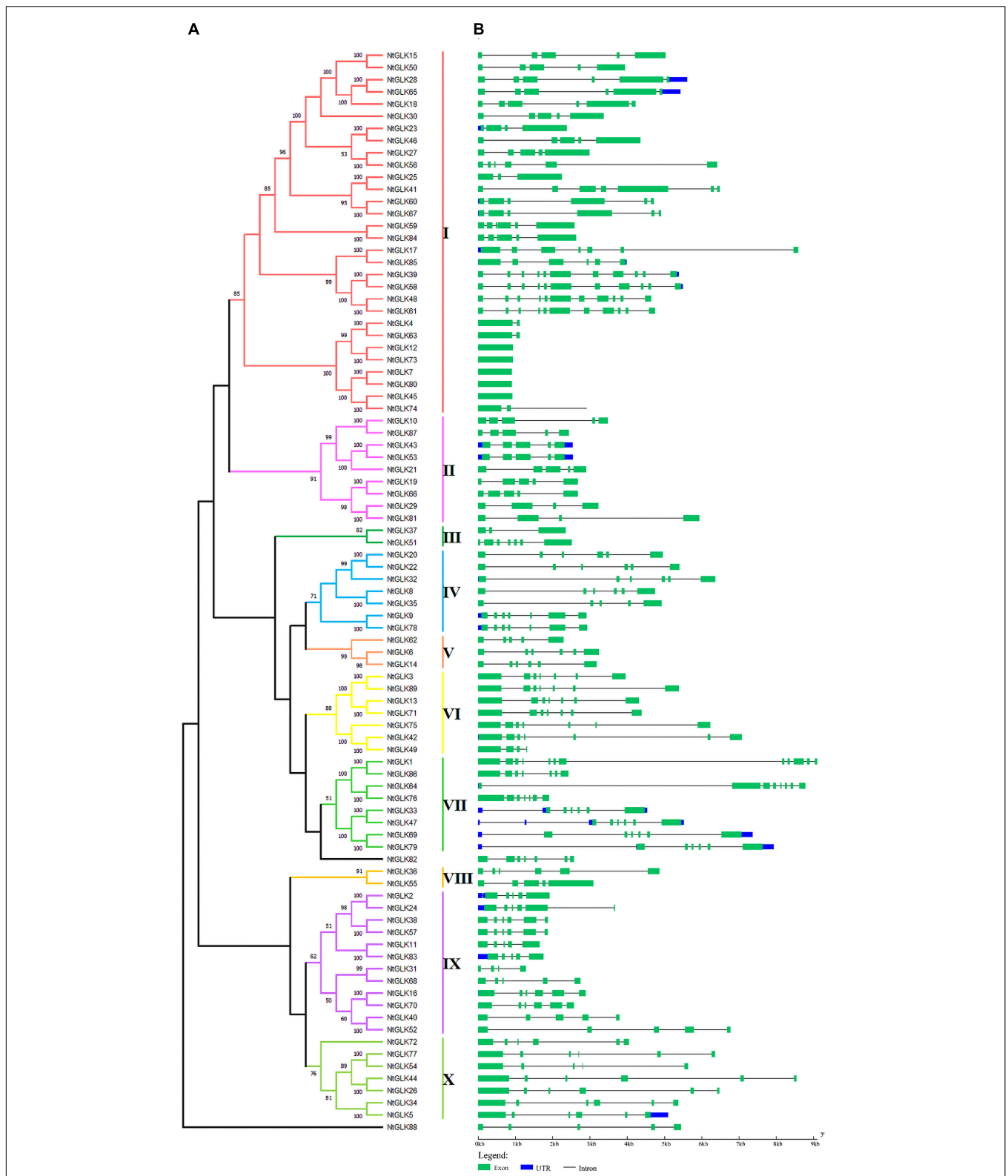


FIGURE 1 | Phylogenetic tree and gene structures of the *NtGLK* gene family. **(A)** A neighbor-joining (NJ) phylogenetic tree was generated by MEGAX based on the *NtGLK* protein sequences. The different subfamilies are distinguished by different colors. **(B)** The exon-intron structures for the *NtGLK* genes were obtained using the online software GSDS. The horizontal black lines and the green boxes represent introns and exon, respectively, and the lengths of the exons and introns can be estimated using the scale.

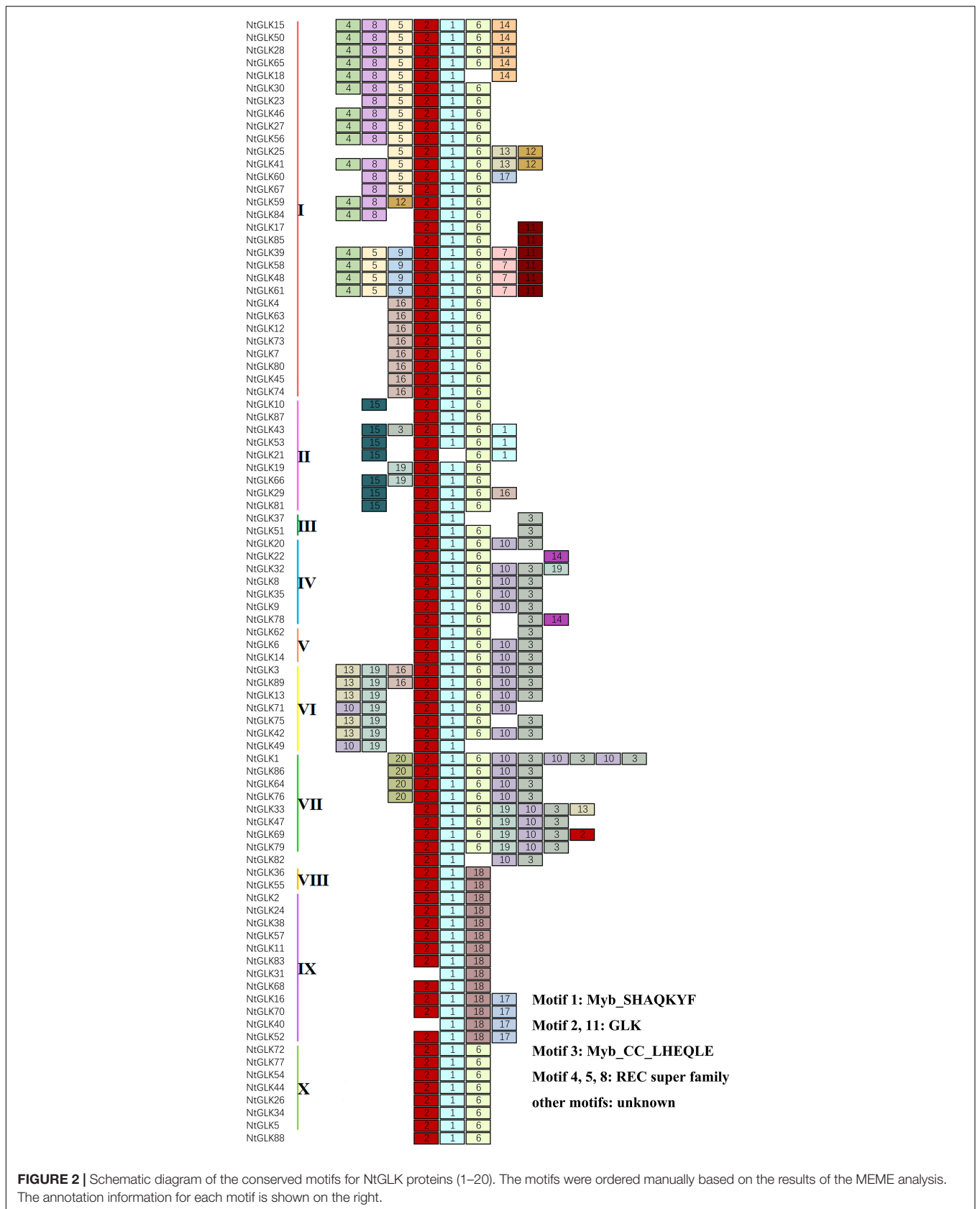


FIGURE 2 | Schematic diagram of the conserved motifs for NtGLK proteins (1–20). The motifs were ordered manually based on the results of the MEME analysis. The annotation information for each motif is shown on the right.

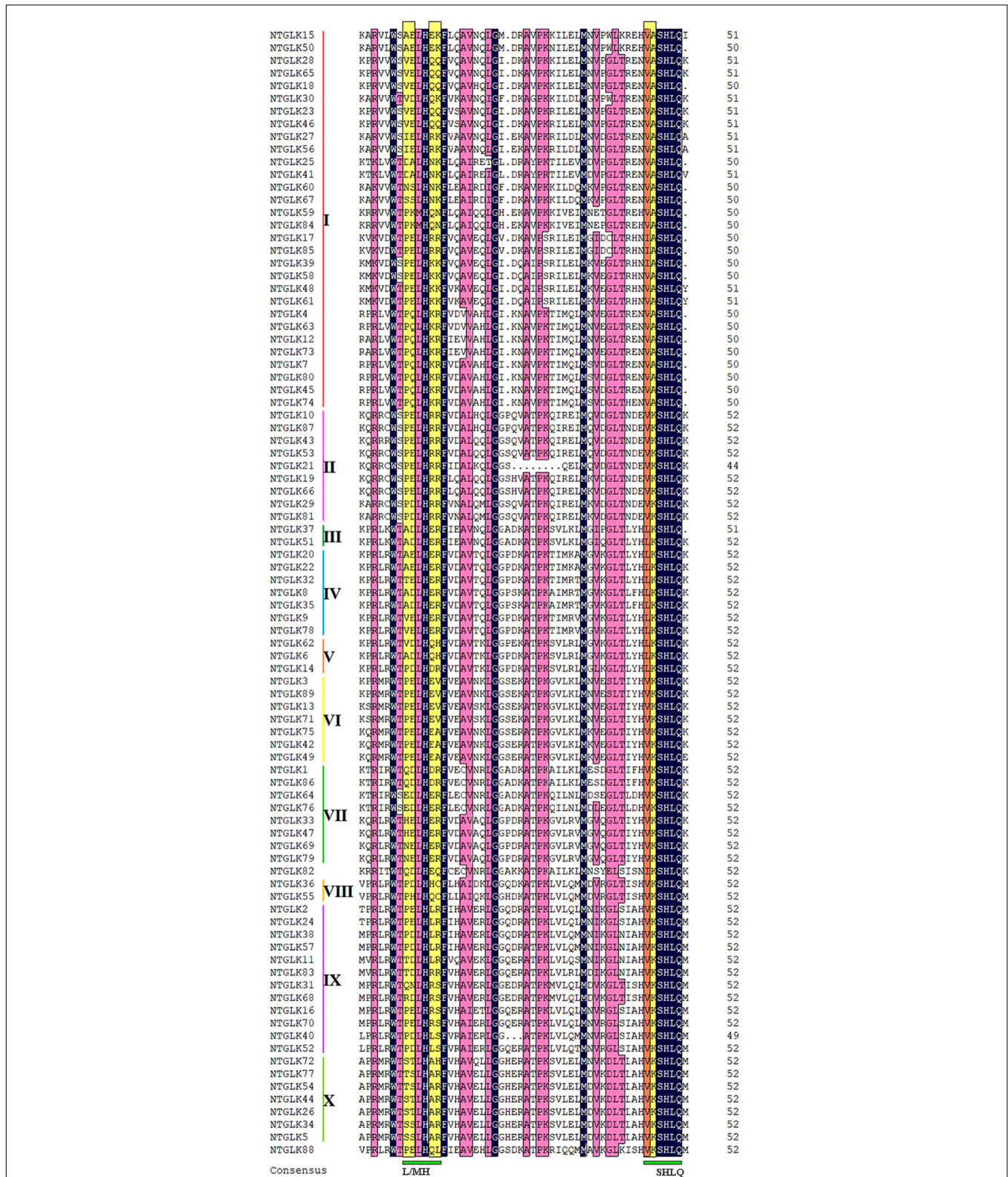


FIGURE 3 | Multiple sequence alignment of the tobacco GLK conserved domain. Identical residues are shaded in black and similar residues in pink.

Analysis of *Cis*-Regulatory Elements in the Promoter Regions of *NtGLK* Genes

Cis-acting elements in the upstream region of the *NtGLK* genes are very important in regulating gene expression in response to various stresses as well as during different developmental stages. To explore the possible expression patterns of the *NtGLK* genes under various stress conditions and senescence processes, *cis*-acting elements including components related to stress, light response, and hormone response were investigated in the promoter regions. Various potential *cis*-acting elements were identified in *NtGLK* promoter regions (**Figure 4** and **Supplementary Table 3**). Among them, the most abundant *cis*-acting elements were light-responsive elements, including GT1-motif, ACE, G-box, ATCT-motif, Box 4, TCT-motif, chs-CMA1a, GATA-motif, I-box, chs-CMA2a, GA-motif, AE-box, MBS, MRE, and TCCC-motif; Box 4 and G-box appear to be the most abundant light-responsive elements, being distributed in the promoter regions of 65 and 56 *NtGLK* genes, respectively. In terms of the hormone response-related *cis*-acting elements, a total of nine types of elements were identified, namely, TGA-element, AuxRR-core, TCA-element, ABRE, CGTCA-motif, TGACG-motif, P-box, GARE-motif, and TATC-box; among them, ABRE was the most abundant *cis*-acting hormone-responsive element in the promoter regions of 89 *NtGLK* genes, which was involved in abscisic acid (ABA) responsiveness. This abundance of hormone-responsive elements indicated that *NtGLK* genes appeared to play important roles in tobacco hormone signal transduction and senescence. In addition, a total of two kinds of *cis*-acting elements involved in various stresses were found, including long terminal repeat (LTR) in low-temperature responsiveness and TC-rich repeats in defense and stress responsiveness (**Supplementary Table 3**). Notably, many *cis*-elements have two or more copies in the 1.5-kb upstream region within the same *NtGLK* gene, which appear to enhance their binding effects to their corresponding *trans*-acting factors. Furthermore, the *NtGLK* genes in the same phylogenetic clade only showed moderate consistency in their distributions of the *cis*-elements, reflecting the complex evolutionary relationship of the diverged *NtGLK* genes, especially in promoter regions.

Chromosomal Locations and Duplications of *NtGLK* Genes

Except for chromosomes 1 and 10, 40 out of the 89 *NtGLK* genes were obtained and were assigned to the 22 tobacco chromosomes (**Figure 5**). Chromosome 4 contained the largest number of *NtGLK* genes (five), and chromosomes 3, 6, 9, 11, 12, 14, 15, 16, 18, 19, and 21 possessed only one *NtGLK* gene each. A similar uneven distribution pattern of the *GLK* gene family was also found in maize (Liu et al., 2016). In addition, the potential duplication events were investigated to explore the potential mechanism for the *NtGLK* gene family. A total of 12 duplicated pairs of *NtGLK* genes were identified as segmental duplication gene pairs (**Figure 5** and **Supplementary Table 4**). However, no pairs of tandem duplicated genes were observed in this study. A biased distribution pattern was also found among the 12 segmental duplication pairs, and no pairs were

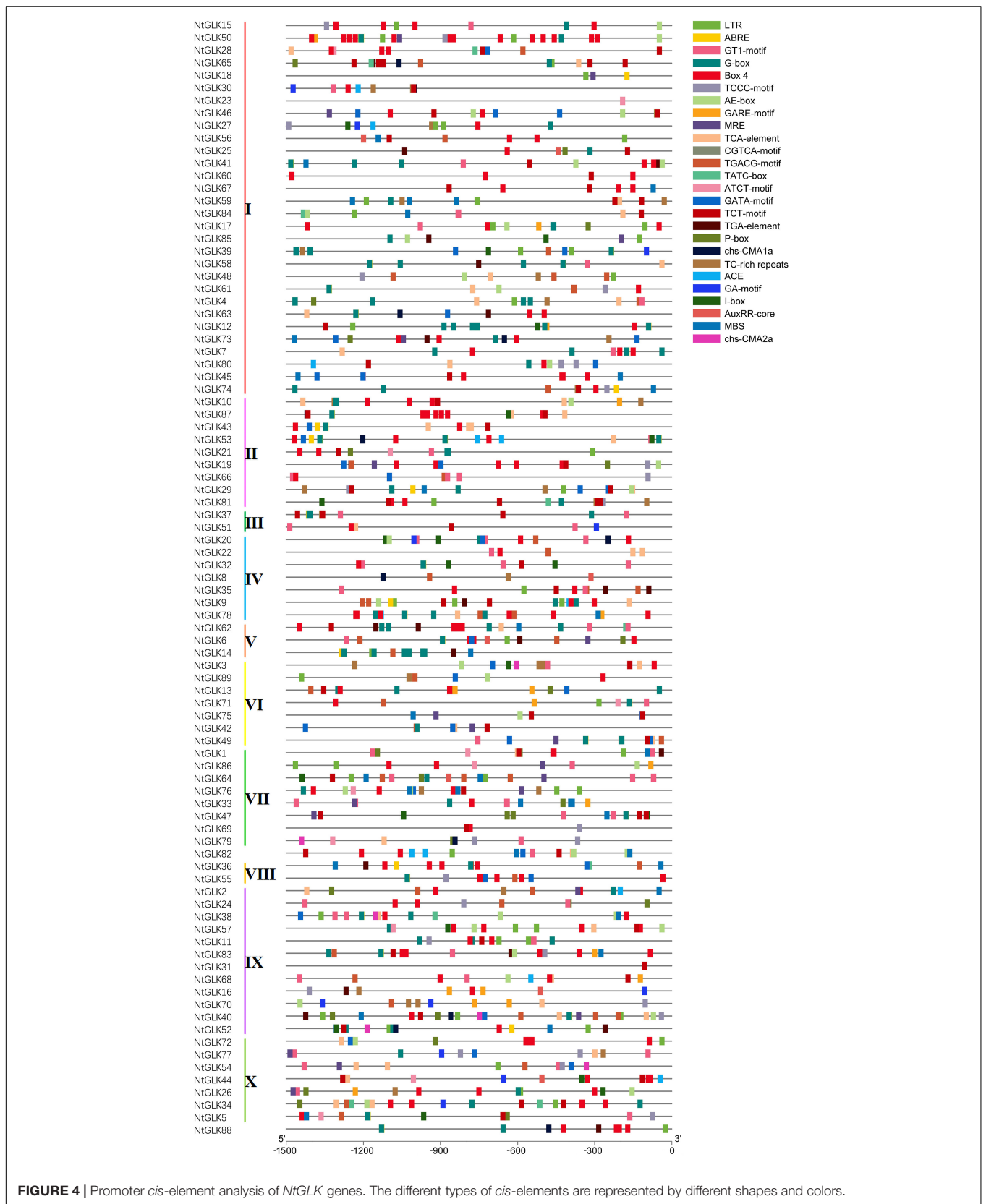
distributed on chromosomes 1, 3, 6, 10, 12, 14, 16, 17, 19, 21, and 22. These results infer that segmental duplication events may play important roles in the amplification of the tobacco *GLK* gene family.

Phylogenetic Analysis of the *NtGLK* Gene Family

To analyze the evolutionary process of the tobacco *GLK* gene family, the 89 *NtGLK* protein sequences were aligned with 59, 54, two, and two *GLK* proteins from maize (Liu et al., 2016), tomato (Liu et al., 2016), rice (Rossini et al., 2001), and *Arabidopsis* (Fitter et al., 2002), respectively. In total, 206 *GLK* proteins were clustered into eight groups (A to H; **Figure 6**). Group E was the largest subfamily, which contained 61 proteins, including 22 from tobacco, 18 from tomato, and 21 from maize. Groups B (40) and H (41) also had large numbers of *GLK* members, and these three groups represented 68.9% of the total *NtGLK* proteins. In contrast, group F was the smallest clade, which had only two *GLK* gene members. Notably, group C contained 14 *GLK* members, and among them, two genes (*At5G44190.1* and *At2G20570.2*) were confirmed to be involved in leaf senescence in *Arabidopsis* (Rauf et al., 2013). It appears that the members clustered into group C, including six members from tobacco (*NtGLK85*, *NtGLK17*, *NtGLK61*, *NtGLK48*, *NtGLK58*, and *NtGLK39*) may share similar functions. In addition, the *GLK* members from tobacco, tomato, and maize were distributed into the major subfamilies, suggesting that the *GLK* gene family existed before the separation of monocotyledons and dicotyledons.

Expression Patterns of *NtGLK* Genes in Different Senescence Stages of Tobacco Leaves

Tobacco leaves at the M1, M2, M3, M4, and M5 senescence stages were collected, which had yellowing rates of 50, 65, 75, 85, and 100%, respectively (**Figure 7**). To explore the potential functions of the *NtGLK* genes, an RNA sequencing (RNA-Seq) experiment was performed for tobacco leaves in the five senescence stages (M1–M5), and the FPKM values of the *NtGLK* genes derived from the RNA-Seq data were used to evaluate the expression levels of these *NtGLK* genes in the different senescence stages. Among them, 25 *NtGLK* genes were excluded from further heat map analysis because of low expression (FPKM < 0.5) or lack of expression in all the senescence stages. The *NtGLK* genes showed differential expressions in tobacco leaves at the different senescence stages (**Figure 8**). A total of 64 *NtGLK* genes were clustered into three groups (**Figure 8**). Among them, 20 genes (*NtGLK49*, *NtGLK33*, *NtGLK78*, *NtGLK69*, *NtGLK42*, *NtGLK9*, *NtGLK35*, *NtGLK8*, *NtGLK43*, *NtGLK81*, *NtGLK13*, *NtGLK65*, *NtGLK47*, *NtGLK75*, *NtGLK29*, *NtGLK46*, *NtGLK79*, *NtGLK48*, *NtGLK89*, and *NtGLK3*) were included in group III (**Figure 8**), which exhibited high expression levels in all the analyzed stages, hinting that these genes were essential in the five senescence stages of tobacco leaves. However, 23 *NtGLK* genes (*NtGLK54*, *NtGLK20*, *NtGLK6*, *NtGLK32*, *NtGLK66*, *NtGLK19*, *NtGLK77*, *NtGLK44*, *NtGLK27*, *NtGLK36*, *NtGLK10*, *NtGLK73*, *NtGLK88*, *NtGLK87*, *NtGLK12*, *NtGLK67*, *NtGLK2*, *NtGLK24*, *NtGLK30*,



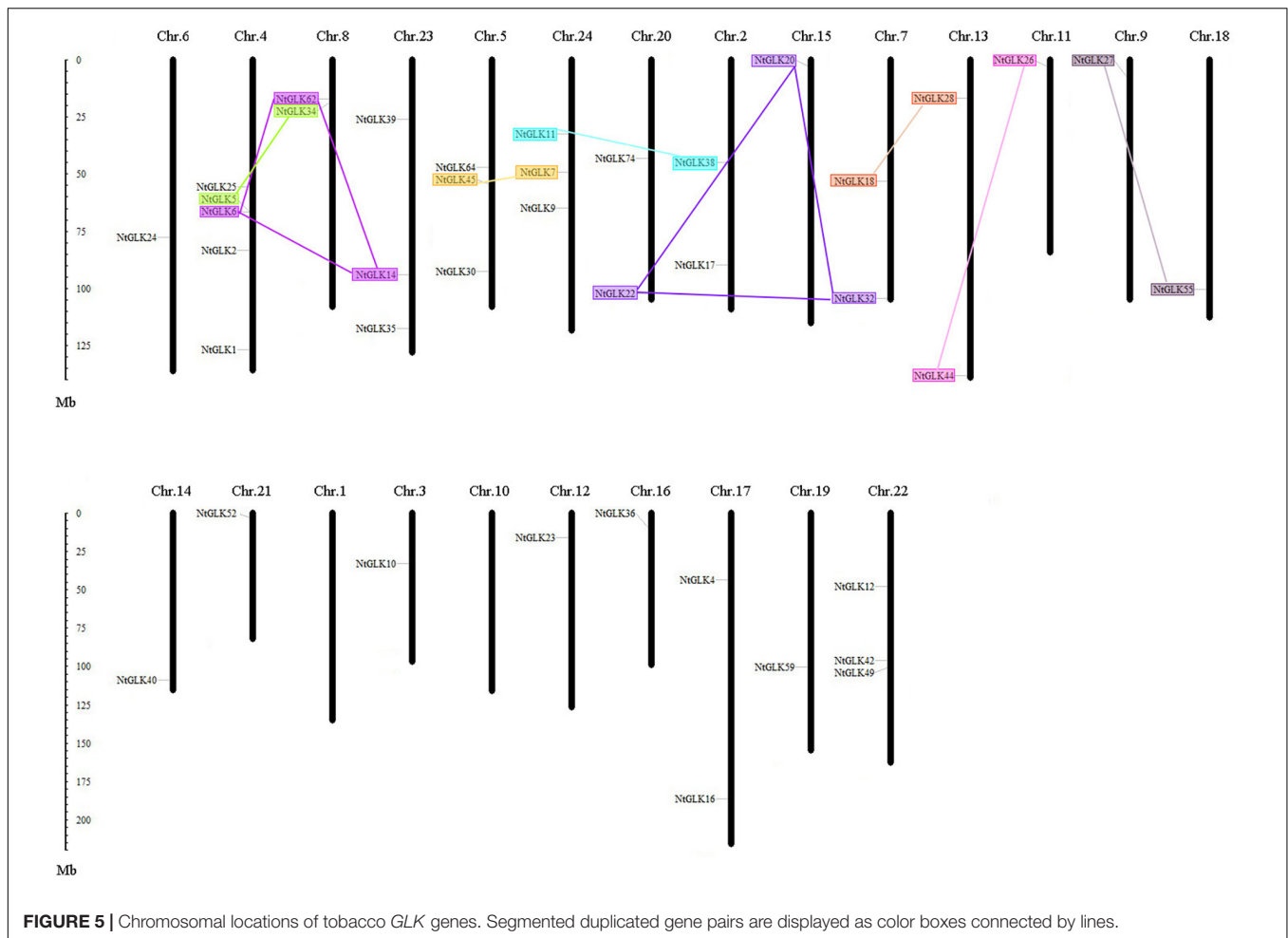


FIGURE 5 | Chromosomal locations of tobacco *GLK* genes. Segmented duplicated gene pairs are displayed as color boxes connected by lines.

NtGLK26, *NtGLK60*, *NtGLK15*, and *NtGLK72*) were clustered into group II and showed relatively lower expression levels in all stages. In general, the expression levels of most *NtGLK* genes exhibited a decreased trend with the increase of senescence level, such as *NtGLK89*, *NtGLK3*, *NtGLK85*, *NtGLK17*, *NtGLK58*, etc. However, inverse expression patterns were also found in some *NtGLK* genes, such as *NtGLK9*, *NtGLK35*, *NtGLK8*, and *NtGLK43*. It is worth noting that *NtGLK85*, *NtGLK17*, and *NtGLK58* were all grouped into subfamily I, and *NtGLK9*, *NtGLK35*, and *NtGLK8* were all clustered into subfamily IV (Figure 1). The results showed that the expression patterns of the *NtGLK* genes in the different senescence stages were different. The expression patterns of the *NtGLK* genes provided preliminary information for their further functional exploration.

To validate the RNA-Seq data, 10 *NtGLK* genes were selected for qRT-PCR analysis (Supplementary Table 1). Among them, variations of the transcript abundances of six genes (*NtGLK3*, *NtGLK19*, *NtGLK36*, *NtGLK69*, *NtGLK72*, and *NtGLK78*) corresponded to the increase of the senescence degrees based on the RNA-Seq data. Another four *GLK* genes (*NtGLK85*, *NtGLK17*, *NtGLK58*, and *NtGLK39*) and two *Arabidopsis* genes (*At5G44190.1* and *At2G20570.2*) were grouped together (group C; Figure 6), and these two *Arabidopsis* genes were previously

confirmed to be involved in leaf senescence (Rauf et al., 2013). The results of qRT-PCR showed that the expression levels of *NtGLK3*, *NtGLK17*, *NtGLK19*, *NtGLK39*, *NtGLK58*, *NtGLK72*, and *NtGLK85* were decreased, precisely consistent with the increase of the senescence degrees, while the expressions of *NtGLK36*, *NtGLK69*, and *NtGLK78* showed a trend of rising first and then decreasing (Figure 9). The expression patterns of these 10 genes detected by qRT-PCR showed similar trends of gene expression pattern to those detected by the RNA-Seq approach (Figure 8), indicating that the RNA-Seq data were reliable.

DISCUSSION

The function of *GLK* genes was first recognized from the analysis of a maize mutant with pale green phenotype (Langdale and Kidner, 1994) and acted as a transcription regulatory element (Hall et al., 1998). The *GLK* gene sequences have only been identified in photosynthetic eukaryotes such as green algae and higher plants, but absent in the genome of the cyanobacterium *Synechocystes*, suggesting that the function of *GLK* genes is associated with the development of chloroplasts (Rossini et al., 2001). Members of the *GLK* gene family are

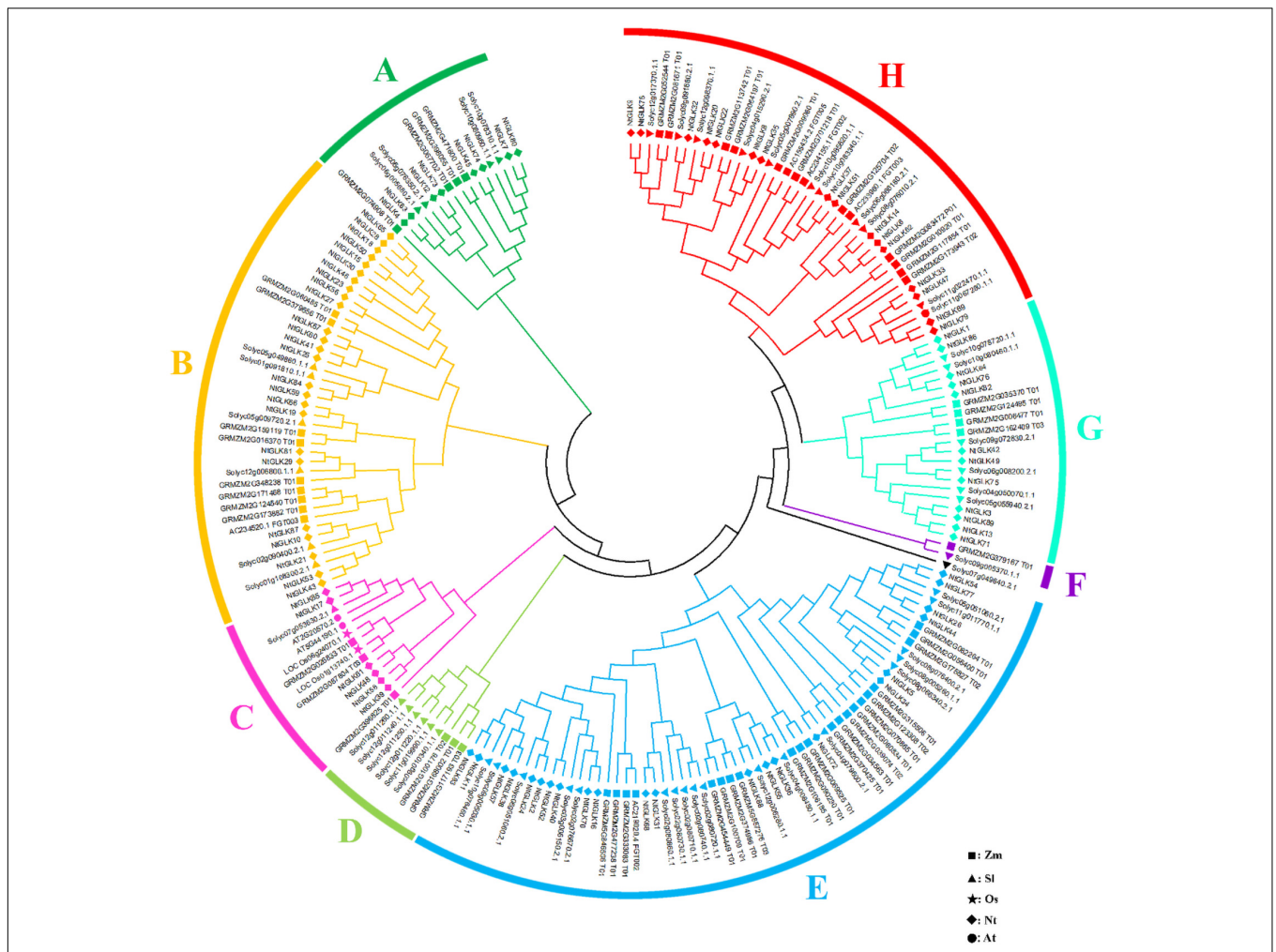
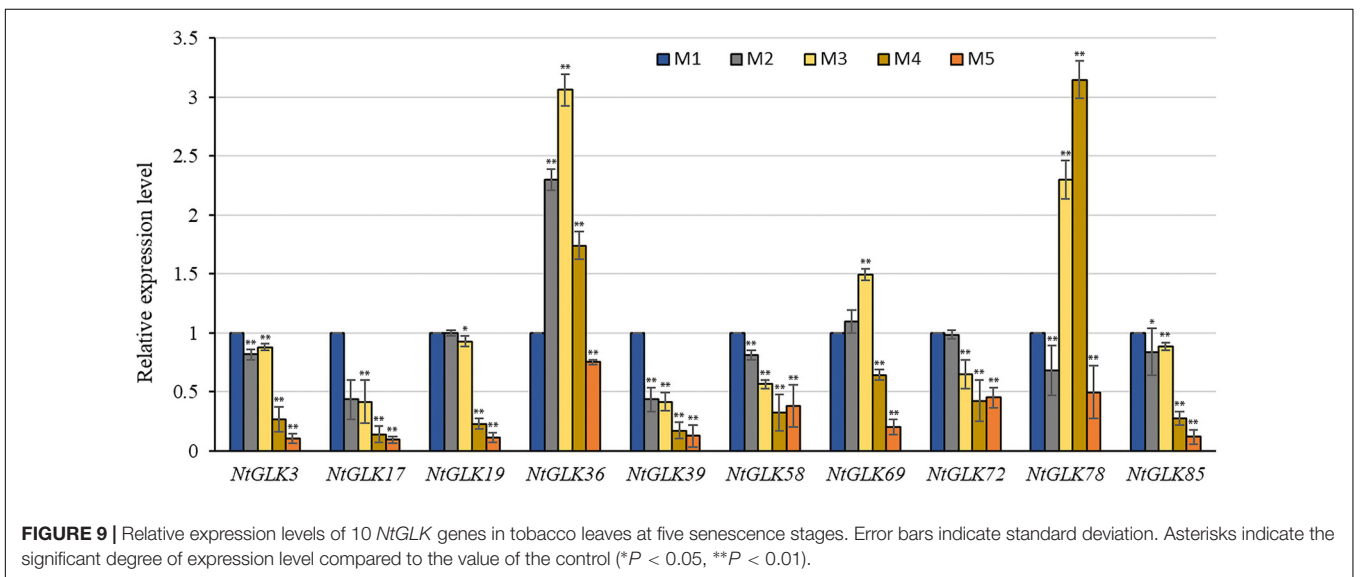
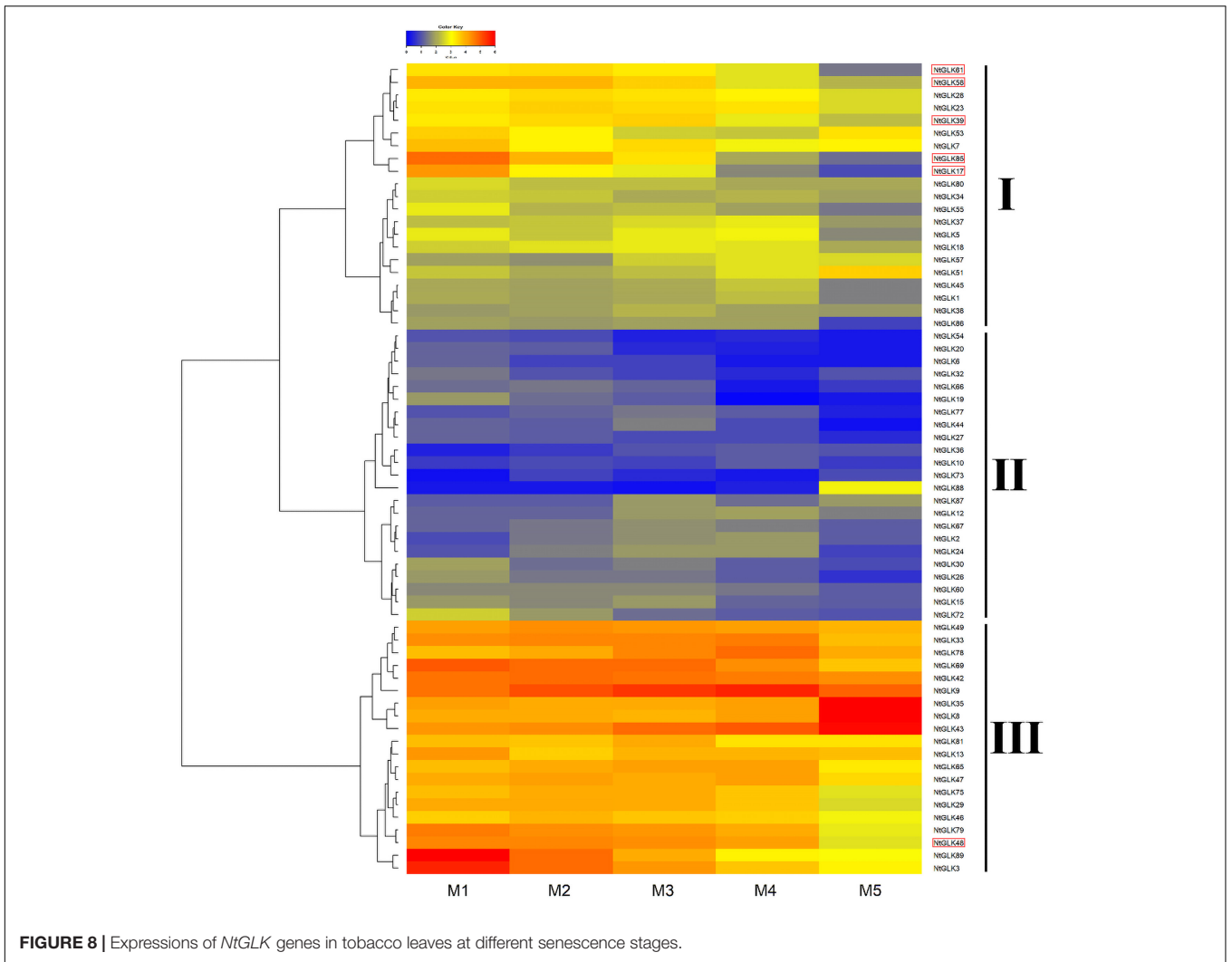


FIGURE 6 | Phylogenetic analysis of tobacco, maize, rice, *Arabidopsis*, and tomato GLK proteins. Different GLK subfamilies are shown as different colors. The square, triangle, star, diamond, and circle represent maize, tomato, rice, tobacco, and *Arabidopsis* GLK proteins, respectively.



FIGURE 7 | Tobacco leaves with different senescence degrees.



considered as one of the most characterizable genes among the nuclear-encoded genes, which regulate chloroplast biogenesis (Fitter et al., 2002). Chloroplast converts light energy into chemical energy, and it plays important roles in the growth and development of plants (Kirchhoff, 2019). The color of leaves gradually changes during leaf senescence with the chlorophyll broken down and the photosynthetic capacity declined (Lira et al., 2017; Schippers et al., 2015; Gan and Amasino, 1997; Lim et al., 2003); therefore, the leaf senescence of tobacco can be assessed by the degree of leaf etiolation. Two *Arabidopsis* genes (*At5G44190.1* and *At2G20570.2*) were confirmed to be involved in leaf senescence (Rauf et al., 2013). Generally, genes that belong to the same subfamily share similar functions compared to those from different subfamilies. In our analysis, a total of six tobacco *GLK* genes (*NtGLK85*, *NtGLK17*, *NtGLK39*, *NtGLK48*, *NtGLK58*, and *NtGLK61*) were clustered with the *Arabidopsis* *GLK* genes (*At5G44190.1* and *At2G20570.2*) in group C (Figure 6), suggesting that these genes across different species might share structural and functional similarity. Thus, it has been postulated that these six putative *NtGLK* genes play roles in the leaf senescence of tobacco. According to the FPKM values generated from the RNA-Seq data, we found that the transcript abundance of the six *NtGLK* genes in subfamily C decreased, closely corresponding to the increase of leaf senescence level, indicating that these six *NtGLK* genes might play important roles in the process of tobacco leaf senescence. The harvest maturity of tobacco leaf is closely related to leaf senescence, and it is considered as a fundamental index to measure tobacco quality. To improve the quality of tobacco, it is very important to develop proper parameters for harvest maturity. Hence, the variations of the transcript abundance of these six *GLK* genes can be further developed as markers to evaluate the maturity of tobacco leaves.

The *NtGLK* genes are not evenly distributed in the tobacco genome. A similar uneven distribution pattern of the *GLK* gene family was also identified in maize (Liu et al., 2016). Unlike prokaryotes, it is common that functionally related genes in eukaryotes are distributed throughout genomes. The physical arrangement of genes on chromosomes is largely derived from insertions, deletions, duplications, and inversions. As a regulatory mechanism of gene expression in cells, microRNAs (miRNAs) target specific messenger RNAs (mRNAs) to regulate gene expression through the mechanism of RNA interference (RNAi). It was observed that the miRNA counts were extremely high in certain chromosomes for the 20 different species, suggesting that the higher number of miRNA genes in these chromosomes might be associated with a regulatory role in cellular functions (Atanu and Utpal, 2014). The co-localized and “operon-like” biosynthetic gene clusters have also been identified in eukaryotes with unknown mechanisms related to highly interactive domains for the regulation of co-expression within clusters of genes (Nützmann et al., 2020). Therefore, the cluster of *NtGLK* may be related to the regulation of genes with leaf senescence.

As the transcription factor, the *G2* gene is activated by the N-terminal region of a heterologous system, and the *GLK* gene family members contain the highly conserved HLH region (DNA-binding domain) and the GCT box (which functions

in dimerization) (Rossini et al., 2001). The sequence of the DNA-binding domain belongs to the GARP transcription factor family (Hosoda et al., 2002), and the two regions of the HLH DNA-binding domain are also conserved (Liu et al., 2016). These two conserved regions of HLH structure were also identified in the tobacco genome in this study; however, the conserved sequences were not totally identical in tobacco. Multiple sequence alignment demonstrated that the second helix region of the *NtGLK* genes was highly conserved (VK/VASHLQ), while a number of variants were found for the *NtGLK* genes in the first helix, except for the fairly conserved L and H, suggesting that the first helix appeared to be more important in the functional differentiation of the *GLK* gene family in tobacco. This diversity in function derived from sequence variants was also observed in maize research, where the *ZmGLK* genes were overall much conserved within the course of evolution (Liu et al., 2016). Actually, the genetic variants derived from the single nucleotide polymorphisms (SNPs) were responses for the gene function of cultivar adaptation in tropical and temperate lines, which are linked to resistance to cold and drought stresses (Liu et al., 2016). In our study, the number of variants in each subfamily seemed different among the 10 subfamilies of the *GLK* gene family in tobacco (Supplementary Table 5), and subfamilies I, IV, VII, and IX showed more sequence variants. The results suggest that sequence diversity was the important factor leading to the more diverse functions in subfamilies I, IV, VII, and IX than in the other groups.

Multigenic families are usually derived from gene duplications, and the expansion mechanism includes unequal crossing-over, various transposition events, duplication of large chromosome segments, or polyploidization events (Zhang, 2003; Passardi et al., 2004). Theoretically, duplication events can often produce two gene copies, and one or both copies can acquire novel functions under a smaller selective pressure of evolution (Van de Peer et al., 2009). It appears that the loss and the insertion of new introns are frequent events, and they play important roles in gene evolution. It was reported that the number of introns was largely reduced and less frequently gained in eukaryotes in the course of evolution (Rogozin et al., 2012). In addition, analysis of segmental duplication events in rice showed that more introns were lost than gained (Lin et al., 2006). In this study, the intron distribution within *NtGLK* genes is quite variable, and the range varied from 0 to 11 (Figure 1), inferring that the shuffling of introns has been a main configuration for the evolution of *NtGLK* genes since their origin. Moreover, a total of four duplication gene pairs, including *NtGLK6/62*, *NtGLK14/62*, *NtGLK11/38*, and *NtGLK27/55* (Supplementary Table 4), appeared to have experienced intron loss events based on our analysis. Notably, one of the transposition events, the retrotransposition of cDNA, is characterized by the loss of all introns and regulatory sequences and by a random insertion within the genome (Casacuberta and Santiago, 2003). In this study, there were five *NtGLK* genes with no introns (*NtGLK12*, *NtGLK73*, *NtGLK7*, *NtGLK80*, and *NtGLK45*), suggesting that these genes might be derived from the retrotransposition events. One of the gene pairs with no introns (*NtGLK7/NtGLK45*) met the parameters of segmental

duplication; thus, they might be due to retrotransposition, but not originated from segmental duplication.

The *GLK* genes have been derived from independent gene duplication as a group of pairs in plants including monocots, eudicots, and bryophytes, but the regulation of *GLK* gene expression appears to be different (Yasumura et al., 2005). The *GLK* genes in maize act differentially in mesophyll cells and bundle sheath for chloroplast development, while they direct the development of chloroplasts in *Arabidopsis* to be monomorphic (Rossini et al., 2001; Fitter et al., 2002). To explore the potential functions of the *NtGLK* genes, RNA-Seq (Figure 8) and qRT-PCR for tobacco leaves in five senescence stages (Figure 9) were conducted for gene expression analyses. The expression levels of some *NtGLK* genes were decreased or increased, precisely corresponding consistently to the increase of the senescence degrees. Whether or not the differential expressions of *GLK* genes lead to assessing thylakoid deformation and impaired chlorophyll biosynthesis in chloroplasts still needs to be investigated. Studies with mutant and cross-species complementation experiments have demonstrated that, although *GLK* gene functions appear to be conserved in different species, the cross-species regulatory elements cannot drive gene expression in other species, suggesting that the *GLK* functional pathway has been diverged and species-specific during land colonization (Bravo-Garcia et al., 2009). Therefore, analysis of the specific expression patterns for these *NtGLK* genes provide preliminary information for their further functional exploration.

CONCLUSION

In this study, a total of 89 *NtGLK* genes were identified in the tobacco genome. They were classified into 10 subfamilies with diverse structures. Twelve pairs of *NtGLK* genes were found to be originated from segmental duplication. Phylogenetic analysis of the *NtGLK* genes showed that the *GLK* gene family existed prior to the separation of monocotyledons and dicotyledons. The *NtGLK* genes showed differential expression patterns in tobacco

leaves at five senescence stages; among them, the expression levels of six genes (*NtGLK85*, *NtGLK17*, *NtGLK39*, *NtGLK48*, *NtGLK58*, and *NtGLK61*) were reduced, coinciding precisely with the increment of the degree of senescence, suggesting that these genes can be further developed as marker genes for maturity evaluation. Our results provide valuable information for further functional study of the *NtGLK* genes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

XX designed this research and wrote the manuscript. MQ and BZ performed the experiments and analyzed the data. GG and XY collected the plant materials and performed the experiments. JZY helped draft the manuscript. JHY participated in handling the figures and tables. All authors contributed to the article and approved the submitted version.

FUNDING

This project has been supported by grants from Fujian Tobacco Company (2019350000240137), Fujian Agriculture and Forestry University Innovation Foundation (no. CXZX2019052G), and the National Natural Science Foundation of China (no. 31501085).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.626352/full#supplementary-material>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Atanu, G., and Utpal, G. (2014). MiRNA gene counts in chromosomes vary widely in a species and biogenesis of miRNA largely depends on transcription or post-transcriptional processing of coding genes. *Front. Genet.* 5:100. doi: 10.3389/fgene.2014.00100
- Bjellqvist, B., Basse, B., Olsen, E., and Celis, J. E. (1994). Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* 15, 529–539. doi: 10.1002/elps.1150150171
- Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., et al. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 14, 1023–1031. doi: 10.1002/elps.11501401163
- Bravo-Garcia, A., Yasumura, Y., and Langdale, J. A. (2009). Specialization of the Golden2-like regulatory pathway during land plant evolution. *New Phytol.* 183, 133–141. doi: 10.1111/j.1469-8137.2009.02829.x
- Casacuberta, J. M., and Santiago, N. (2003). Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene* 311, 1–11. doi: 10.1016/S0378-1119(03)00557-2
- Chen, C. T., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Edwards, K. D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A. D., Bombarely, A., et al. (2017). A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* 18:448. doi: 10.1186/s12864-017-3791-6
- Fitter, D. W., Martin, D. J., Copley, M. J., Scotland, R. W., and Langdale, J. A. (2002). *GLK* gene pairs regulate chloroplast development in diverse plant species. *Plant J.* 31, 713–727. doi: 10.1046/j.1365-313x.2002.01390.x
- Gan, S., and Amasino, R. M. (1997). Making sense of senescence (molecular genetic regulation and manipulation of leaf senescence). *Plant Physiol.* 113, 313–319. doi: 10.1104/pp.113.2.313
- Gregersen, P. L., Culetic, A., Boschian, L., and Krupinska, K. (2013). Plant senescence and crop productivity. *Plant Mol. Biol.* 82, 603–622. doi: 10.1007/s11103-013-0013-8

- Gu, Z. L., Cavalcanti, A., Chen, F. C., Bouman, P., and Li, W. H. (2002). Extent of gene duplication in the genomes of *Dro-sophila*, Nematode, and Yeast. *Mol. Biol. Evol.* 19, 256–262.
- Hall, L. N., Rossini, L., Cribb, L., and Langdale, J. A. (1998). GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. *Plant Cell* 10, 925–936. doi: 10.1105/tpc.10.6.925
- Han, X. Y., Li, P. X., Zou, L. J., Tan, W. R., Zheng, T., Zhang, D. W., et al. (2016). GOLDEN2-LIKE transcription factors coordinate the tolerance to *Cucumber mosaic virus* in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 477, 626–632. doi: 10.1016/j.bbrc.2016.06.110
- Hörtensteiner, S. (2006). Chlorophyll degradation during senescence. *Annu. Rev. Plant Biol.* 57, 55–77. doi: 10.1146/annurev.arplant.57.032905.105212
- Hosoda, K., Imamura, A., Katoh, E., Hatta, T., Tachiki, M., Yamada, H., et al. (2002). Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the *Arabidopsis* response regulators. *Plant Cell* 14, 2015–2029. doi: 10.1105/tpc.002733
- Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817
- Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell Biol.* 14, 787–802. doi: 10.1038/nrm3702
- Kirchhoff, H. (2019). Chloroplast ultrastructure in plants. *New Phytol.* 223, 565–574. doi: 10.1111/nph.15730
- Kräutler, B. (2016). Breakdown of chlorophyll in higher plants—phylobilins as abundant, yet hardly visible signs of ripening, senescence, and cell death. *Angew. Chem. Int. Ed. Engl.* 55, 4882–4907. doi: 10.1002/anie.201508928
- Kumar, S., Stecher, G., Li, M., Nkayaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Langdale, J. A., and Kidner, C. A. (1994). bundle sheath defective, a mutation that disrupts cellular differentiation in maize leaves. *Development* 120, 673–681.
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325
- Lim, P. O., Woo, H. R., and Nam, H. G. (2003). Molecular genetics of leaf senescence in *Arabidopsis*. *Trends Plant Sci.* 8, 272–278. doi: 10.1016/S1360-1385(03)00103-1
- Lin, H. N., Zhu, W., Silva, J. C., Gu, X., and Buell, C. R. (2006). Intron gain and loss in segmentally duplicated genes in rice. *Genome Biol.* 7:R41. doi: 10.1186/gb-2006-7-5-r41
- Lira, B. S., Gramagna, G., Trench, B. A., Alves, F., Silva, E. M., Silva, G., et al. (2017). Manipulation of a senescence-associated gene improves fleshy fruit yield. *Plant Physiol.* 175, 77–91. doi: 10.1104/pp.17.00452
- Liu, F., Xu, Y. J., Han, G. M., Zhou, L. Y., Ali, A., Zhu, S. W., et al. (2016). Molecular evolution and genetic variation of G2-Like transcription factor genes in maize. *PLoS One* 11:e0161763. doi: 10.1371/journal.pone.0161763
- Liu, J. F. (2018). *Bioinformatics Analysis of tomato G2-like Transcription Factor Family and Identification of Resistance-Related Genes*. Dissertation's thesis. Harbin: Northeast Agricultural University.
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lu, S. N., Wang, J. Y., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991
- Ma, W. X., Noble, W. S., and Bailey, T. L. (2014). Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat. Protoc.* 9, 1428–1450. doi: 10.1038/nprot.2014.083
- Murmu, J., Wilton, M., Allard, G., Pandeya, R., Desveaux, D., Singh, J., et al. (2014). *Arabidopsis* GOLDEN2-LIKE (GLK) transcription factors activate jasmonic acid (JA)-dependent disease susceptibility to the biotrophic pathogen *Hyaloperonospora arabidopsidis*, as well as JA-independent plant immunity against the necrotrophic pathogen *Botrytis cinerea*. *Mol. Plant Pathol.* 15, 174–184. doi: 10.1111/mp.12077
- Nagatoshi, Y., Mitsuda, N., Hayashi, M., Inoue, S., Okuma, E., Kubo, A., et al. (2016). GOLDEN 2-LIKE transcription factors for chloroplast development affect ozone tolerance through the regulation of stomatal movement. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4218–4223. doi: 10.1073/pnas.1513093113
- Nüttmann, H. W., Daniel, D., América, R. C., Jesús, E. S., Eva, W., Marco, D. S., et al. (2020). Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc. Natl. Acad. Sci. U.S.A.* 117, 13800–13809. doi: 10.1073/pnas.1920474117
- Passardi, F., Longet, D., Penel, C., and Dunand, C. (2004). The class III peroxidase multigenic family in rice and its evolution in land plants. *Phytochemistry* 65, 1879–1893. doi: 10.1016/j.phytochem.2004.06.023
- Powell, A. L., Nguyen, C. V., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, H., et al. (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. *Science* 336, 1711–1715. doi: 10.1126/science.1222218
- Rauf, M., Arif, M., Dortay, H., Matallana-Ramírez, L. P., Waters, M. T., Gil Nam, H., et al. (2013). ORE1 balances leaf senescence against maintenance by antagonizing G2-like-mediated transcription. *EMBO Rep.* 14, 382–388. doi: 10.1038/embor.2013.24
- Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110. doi: 10.1126/science.290.5499.2105
- Rogozin, I. B., Carmel, L., Csuros, M., and Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biol. Direct* 7:11. doi: 10.1186/1745-6150-7-11
- Rose, A., Meier, I., and Wienand, U. (2000). The tomato I-box binding factor LeMBY1 is a member of a novel class of myb-like proteins. *Plant J.* 20, 641–652.
- Rossini, L., Cribb, L., Martin, D. J., and Langdale, J. A. (2001). The maize *golden2* gene defines a novel class of transcriptional regulators in plants. *Plant Cell* 13, 1231–1244. doi: 10.1105/tpc.13.5.1231
- Rubio, V., Linhares, F., Solano, R., Martin, A. C., Iglesias, J., Leyva, A., et al. (2001). A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes Dev.* 15, 2122–2133.
- Savitch, L. V., Subramaniam, R., Allard, G. C., and Singh, J. (2007). The GLK1 ‘regulon’ encodes disease defense related proteins and confers resistance to *Fusarium graminearum* in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 359, 234–238. doi: 10.1016/j.bbrc.2007.05.084
- Schippers, J. H., Schmidt, R., Wagstaff, C., and Jing, H. C. (2015). Living to die and dying to live: the survival strategy behind leaf senescence. *Plant Physiol.* 169, 914–930. doi: 10.1104/pp.15.00498
- Schreiber, K. J., Nasmith, C. G., Allard, G., Singh, J., Subramaniam, R., and Desveaux, D. (2011). Found in translation: high-throughput chemical screening in *Arabidopsis thaliana* identifies small molecules that reduce Fusarium head blight disease in wheat. *Mol. Plant Microbe Interact.* 24, 640–648. doi: 10.1094/MPMI-09-10-0210
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600
- Wang, L. Q., Guo, K., Li, Y., Tu, Y. Y., Hu, H. Z., Wang, B. R., et al. (2010). Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *BMC Plant Biol.* 10:282. doi: 10.1186/1471-2229-10-282
- Waters, M. T., Moylan, E. C., and Langdale, J. A. (2008). GLK transcription factors regulate chloroplast development in a cell-autonomous manner. *Plant J.* 56, 432–444. doi: 10.1111/j.1365-313X.2008.03616.x
- Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., et al. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 112, 531–552. doi: 10.1385/1-59259-584-7:531
- Xiao, Y., You, S., Kong, W., Tang, Q., Bai, W., Cai, Y., et al. (2019). A GARP transcription factor anther dehiscence defect 1 (OsADD1) regulates rice anther dehiscence. *Plant Mol. Biol.* 101, 403–414.
- Xu, X. Y., Liao, K. F., Dai, J. R., Hu, W. Z., Cheng, S. Y., Duan, Y. M., et al. (2017). A study on different harvest maturity levels in leaf structures and

- physiological and biochemical properties of fresh tobacco leaves. *J. Yunnan Univ.* 39, 313–323.
- Yang, S. H., Zhang, X. H., Yue, J. X., Tian, D. C., and Chen, J. Q. (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* 280, 187–198.
- Yasumura, Y., Moylan, E. C., and Langdale, J. A. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *Plant Cell* 17, 1894–1907. doi: 10.1105/tpc.105.033191
- Zhang, J. Z. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298. doi: 10.1016/S0169-5347(03)00033-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Qin, Zhang, Gu, Yuan, Yang, Yang and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.