In silico genome-wide identification and phylogenetic analysis of the WRKY transcription factor family in sweet orange (Citrus sinensis)

Eduardo Goiano da Silva, Tania Mayumi Ito, Silvia Graciele Hülse de Souza

1Laboratório de Biologia Molecular, Universidade Paranaense-UNIPAR, 87502-210, Umuarama-PR, Brazil
2Departamento de Informação e Comunicação-Informática, Instituto Federal do Paraná-IFPR, 87507-014, Umuarama-PR, Brazil

*Corresponding author: silviahulse@unipar.br

Abstract

WRKY transcription factors (TFs) play important roles in several biological processes, especially in defense against different biotic and abiotic stresses. An extensive in silico characterization of the CitsWRKY gene family was performed. In this study, 77 WRKY genes from the Citrus sinensis genome were identified. Based on sequence alignment and phylogenetic analysis, these 77 proteins were classified into three main groups, with most WRKY proteins placed in group I (36 CitsWRKY), group II (35 CitsWRKY), and group III (6 CitsWRKY). Analysis of conserved motifs showed that most identified CitsWRKY proteins carry a conserved WRKY domain. Structural analysis of the genes showed that the number of introns in CitsWRKY genes varied from zero to five. CitsWRKY genes were found to be randomly distributed on all nine chromosomes. The expansion of this gene family may have resulted from 14 tandem and five segmental duplication events. Evaluation of syntenic events suggested that some WRKY genes emerged before the divergence of C. sinensis and Arabidopsis. Expression analysis demonstrated that these genes frequently occur in developing tissues such as fruit, flowers, ovaries, meristem, and phloem (65%, 66%, and 85% in groups I, II, and III, respectively), and are stimulated by biotic or abiotic stresses (15%, 16%, and 10% in groups I, II, and III, respectively). The characterization and analysis of these genes will aid in the selection of candidate genes for future functional analyses of the WRKY family in citrus, which will expand our understanding of genetic determinants of stress tolerance.

Keywords: citrus, Expressed Sequence Tags (ESTs), gene expression, transcription factor.

Abbreviations: TFs_transcription factors, CitsWRKY_Citrus sinensis WRKY, ESTs_Expressed Sequence Tags, C2H2 zinc_zinc-finger like, USDA_United States Department of Agriculture, USA_United States of America, BLASTP_Basic Local Alignment Search Tool Protein, aa_amino acids, NJ_Neighbor-Joining, MW_molecular weight, pl_ isoelectric point, AtWRKY_Arabidopsis thaliana WRKY, STZ_Salt Tolerance Zinc Finger, RD_Responsive to Desiccation genes (RD29A and RD29B), DREB_Dehydration Responsive Element Binding (DREB2A), COR_Cold-Regulated genes (COR 6.6), TAIR_The Arabidopsis Information Resource, NCBI_National Center for Biotechnology Information, ORF_Open Reading Frame, GSDS_Gene Structure Display Server, PGDD_Plant Genome Duplication Database.

Introduction

In their natural habitat, plants are exposed to environmental changes that significantly affect their growth and development and drastically decrease productivity (Zhou et al., 2008). Among the various environmental stimuli, biotic and abiotic stresses cause damage to plants due to plague and disease, where the abiotic stresses are predominant and include water deficit, heat, and soil salinity (Muthamilarasan et al., 2015). Climatic changes due to global warming represent another serious challenge that influences the sustainability and productivity of cultivated plants (Kole et al., 2015). Therefore, in order to survive and adapt to different environmental changes, plants have developed a complex signaling network, which is responsible for the orchestrated gene expression responsible for the transmission of tolerance to adverse environmental conditions (Mickelbart et al., 2015). Gene regulation at the transcriptional level plays an important role in the control of biological processes and in responses to environmental conditions that are unfavorable to plants (Zheng et al., 2013). Transcriptional regulation and gene expression are controlled by transcription factors, which have fundamental regulatory roles in this process, in which they act by activating or suppressing a signaling cascade (Riechmann et al., 2000; Chen et al., 2012).

Among the different TFs, the WRKY TFs are essential for plant development because they participate in the responses to several physiological processes and are also important components that regulate the plant defense response (Cai et al., 2014). WRKY family TFs present at least one highly preserved domain of 60 amino acids, which contains the preserved WRKYGQK sequence in the N-terminal region and a C2H2 zinc (zinc-finger like) binding motif(C-X4-S-C-X22.23-H-X-H) or C2HC (C-X7-C-X23-H-X-C) in its C-terminal region (Eulgem et al., 2000). The WRKY domain has a fundamental role in W-box binding to the promoter region, with a consensus sequence (C/T)TGAC(T/C), resulting in target gene expression (Eulgem et al., 2000). Based on variation in the WRKY domain and the standard “fingerlike” motif, WRKY proteins can be divided into three main groups (I, II, and III) with several subgroups (Rushton et al., 2012). Group I typically contains two WRKY domains.
and one C2H2 zinc binding motif, whereas the proteins of group II have a single WRKY domain and one C2H2 zinc binding motif WRKY. Proteins of group III have a single WRKY domain and one C2-H-C zinc binding motif (Eulgem et al., 2000).

The genes in this family participate in different pathways in response to the plant hormones jasmonate, gibberellin acid, and abscisic acid (Rushton et al., 2012). Moreover, studies have shown that WRKY TFs have regulatory roles in signaling and modulation pathways of various molecular and physiological processes like seed germination, pollen development, biomass accumulation, and flowering regulation, and are also involved in the biosynthesis of secondary metabolites (Sun et al., 2003; Rushton et al., 2012; Yu et al., 2013). The WRKY TFs have also shown defense responses to different abiotic stresses, including heat, cold, water deficiency, and salinity (Wu et al., 2009; Jiang and Deyholos, 2009; Niu et al., 2012) and different biotic stresses, such as fungal, and viral, and bacterial infection (Beyer et al., 2001; Huh et al., 2012; Pan and Jiang, 2014).

Thus, considering the fundamental role of WRKY TFs in several molecular, biochemical, and physiological processes, this gene family has been extensively characterized in different species such as Arabidopsis thaliana (Eulgem et al., 2000), Oryza sativa (Ross et al., 2007), Zea mays (Wei et al., 2012), Hordeum vulgare (Liu et al., 2014), Vitis vinifera (Wang et al., 2014), and in trees including Populus trichocarpa (He et al., 2012) and, Hevea brasiliensis (Li et al., 2014). However, few studies have identified the members of the WRKY family in oranges (Ayadi et al., 2016).

With a worldwide production of 51.8 million tons of oranges in the 2014 harvest, this fruit is economically important and is cultivated in more than 140 countries (USDA, 2015). Brazil is the major worldwide producer of orange, and together with the USA, was responsible for more than 50% of the production in 2014 (USDA, 2015). However, orange production in Brazil is subject to regular fluctuations, which can be explained by the adverse effects of climatic conditions (Sahin-Cevik and Moore, 2013; Ito et al., 2014). Therefore, considering the importance of WRKY TFs in the response to different stresses and the fact that few studies have characterized this gene family in C. sinensis, we identified and characterized WRKY TFs utilizing a computational approach. Thus, the results of study provide information on the existing genetic-molecular diversity of the WRKY family, the functional aspects of these TFs, and the potential candidate genes for posterior functional characterization in response to biotic and abiotic stresses.

Results

Identification and phylogenetic analysis of WRKY genes in C. sinensis

To identify WRKY genes in sweet oranges, protein sequences of WRKY TFs identified in A. thaliana (Eulgem et al., 2000) were used to perform a BLASTP search in the C. sinensis genome database (Phytozome v10.3: http://www.phytozome.net). In total, 82 genes that may encode proteins containing a WRKY domain were identified, five of which were excluded from subsequent analyses because they were repeated or had a partial WRKY domain. Therefore, a total of 77 genes, which encode putative WRKY proteins were identified from the C. sinensis genome and named CitsWRKY01 to CitsWRKY77 (Table S1).

The most important structural characteristic of the WRKY protein is the WRKY domain, which interacts with W-Box (CGTCA/TGAC/T), activating multiple genes related to plant defense mechanisms (Eulgem et al., 2000). A multiple alignment of the WRKY domain of approximately 60 amino acids from the 77 CitsWRKY proteins is shown in Fig. S1. From the protein sequence and number of WRKY domains, 77 CitsWRKY genes attributed to the WRKY family were classified into the three groups. A total of 36, 35, and 6 genes belonged to groups I, II, and III, respectively (Table 1). The proteins of group I possess two preserved WRKY/GQK motifs, whereas those of groups II and III present only one preserved motif, all of which contain the GQK signature (Fig. S1). The domains found in group I proteins were followed by a preserved CX4CX22-23HXH sequence, while the proteins from groups II and III present a single WRKY domain and one C2H2 and C2-H-C zinc binding motif, respectively (Fig. 1). The alignment of CitsWRKY sequences showed that all proteins contain a conserved WRKY domain and a zinc-fingerlike structure.

A phylogenetic tree was constructed based on 77 CitsWRKY proteins, utilizing the NJ method with sequences of C. sinensis amino acids (Fig. 2). The phylogenetic classification of WRKY family proteins from A. thaliana enabled the classification of 77 WRKY proteins from citrus into three main groups (I, II, and III), and seven subgroups (la, Ib, Iia, Iib, IIc, IIId, and IIIa). Of the 36 CitsWRKY proteins from group I, 24 belong to subgroup Ia, and 12 belong to subgroup Ib. Of the 37 proteins from group II, four belong to subgroup Iia, 11 to Iib, 13 to IIc, and seven to IIId. Six CitsWRKY proteins belonging to group IIIa were found. Notably, no WRKY proteins belonging to groups Ile and IIlb were found in the C. sinensis genome (Table 1; Fig. 2).

Protein properties and subcellular location of CitsWRKY proteins

In silico analyses were performed to identify and analyze the physicochemical characteristics of WRKY proteins identified in C. sinensis (Table S1). Of 77 analyzed proteins, CitsWRKY28 presented the smallest proteins with 106 amino acids (aa), whereas the largest proteins were CitsWRKY24 and CitsWRKY99 with 721 amino acids (aa). The molecular weight (MW) of proteins also varied according to protein size, from 12662.46 g/mol (CitsWRKY28) to 77798.72 g/mol (CitsWRKY169). The isoelectric point (pI) of CitsWRKY proteins ranged from 4.86 (CitsWRKY30) to 9.93 (CitsWRKY49). Most of the WRKY proteins were found to be located within the cell nucleus. The high variation in protein characteristics indicates the high complexity of WRKY genes.

Analyses of conserved motifs

Overall, regions other than the DNA-binding domain in TFs contain functionally important conserved domains. Ten distinct motifs were identified and found to be distributed among the main groups of WRKY genes (Fig. 3). The consensus sequences identified in different motifs are shown in Table 3. Motifs 1 and 3 are present in most WRKY genes and consist of a WRKY domain. As expected, most members of this family included in the phylogenetic tree presented common motifs, suggesting functional similarities among WRKY proteins within the same group and subgroup (Fig. 3; Table 2). Motifs 4 and 9 was present in members of group Ia,
Table 1. Comparison of the number of members in each WRKY family group of Arabidopsis thaliana, Oryza sativa, Vitis vinifera, Zea mays, Populus trichocarpa, and Citrus sinensis.

<table>
<thead>
<tr>
<th>Specie</th>
<th>Arabidopsis thaliana</th>
<th>Oryza sativa</th>
<th>Zea mays</th>
<th>Vitis vinifera</th>
<th>Populus trichocarpa</th>
<th>Citrus sinensis</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subgroup Ia</td>
<td>13</td>
<td>34</td>
<td>43</td>
<td>12</td>
<td>50</td>
<td>36</td>
</tr>
<tr>
<td>Subgroup Ib</td>
<td>5</td>
<td>14</td>
<td>22</td>
<td>7</td>
<td>35</td>
<td>24</td>
</tr>
<tr>
<td>GROUP II</td>
<td>45</td>
<td>30</td>
<td>78</td>
<td>41</td>
<td>44</td>
<td>35</td>
</tr>
<tr>
<td>Subgroup IIa</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Subgroup IIb</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Subgroup IIc</td>
<td>18</td>
<td>7</td>
<td>29</td>
<td>16</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Subgroup IIId</td>
<td>7</td>
<td>11</td>
<td>14</td>
<td>7</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Subgroup IIe</td>
<td>9</td>
<td>0</td>
<td>17</td>
<td>6</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>GROUP III</td>
<td>14</td>
<td>36</td>
<td>31</td>
<td>6</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Subgroup IIIa</td>
<td>8</td>
<td>11</td>
<td>-</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Subgroup IIIb</td>
<td>6</td>
<td>25</td>
<td>-</td>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>72</td>
<td>100</td>
<td>152</td>
<td>59</td>
<td>104</td>
<td>77</td>
</tr>
</tbody>
</table>

Genome size (Mb)
- He et al., 2012.
- Ling et al., 2011.
- Wei et al., 2012.
- He et al., 2012.

Fig 1. Predicted domains in the Citrus sinensis WRKY proteins. The conserved domains were carried out by MEME using the protein sequences of CitsWRKYS. This software was used to create the logo representations of the WRKY domains (A and B) and the zinc finger motif (C). Each column shows the consensus of each amino acids at the corresponding motif position. Protein colors are based on the biochemical properties of the various amino acids.

Table 2. Conserved motifs of WRKY proteins.

<table>
<thead>
<tr>
<th>ID</th>
<th>MOTIF</th>
<th>SEQUENCES</th>
<th>E-VALUES</th>
<th>WIDTH</th>
<th>SITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LDDGYWRKYGQKVVRKNSYTKCT</td>
<td>3.3e-1059</td>
<td>29</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CVVRKVERCSRTMVTYYEGHNIIPW</td>
<td>6.1e-910</td>
<td>29</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DIILDGVRKYGQKVVRKNSYTKCT</td>
<td>1.8e-844</td>
<td>29</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CVKVEKKVERSDKQITEICYK</td>
<td>3.6e-350</td>
<td>21</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ITKERTIREPRVQVTSEQV</td>
<td>1.0e-258</td>
<td>21</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DMVNFDEDDVEEVEGNCNIVFDDEDEPYAKRKIEAYVEG</td>
<td>3.7e-226</td>
<td>41</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>GTNAPPKQLTRVRDGNNMN</td>
<td>3.3e-156</td>
<td>21</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>SNCRCCSKKRRVVRKVRIVMAISNL</td>
<td>1.3e-140</td>
<td>29</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>LLPIRSFCFLPSSLPTTLESPVFLSNVAPQPSPTTGHI</td>
<td>3.0e-138</td>
<td>41</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>PROGLGFSAEDTITEVNSCSSDEERRLSGTPNIVEAASKENVSNGKNEIVSFD</td>
<td>6.3e-141</td>
<td>80</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

*Protein colors are based on the biochemical properties of the various amino acids.

*WRKYGQK motif is indicated by black amino acids.
Fig 2. Phylogenetic analysis of 77 *CitsWRKY* proteins. Amino acid sequences were aligned by ClustalOmega and the phylogenetic tree was built utilizing MEGA 6.0.6 software and the Neighbor-Joining method. The three groups are classified as I, II, and III and subgroups are classified as Ia, Ib, IIA, IIB, IIC, IID, and IIIa.

Fig 3. Phylogenetic tree and schematic representations of 10 preserved motifs of *CitsWRKY* family members, designed by Mega 6.06 software. MEME software was used to represent the motifs.
Fig 4. Intron/exon structure of *Citrus sinensis* WRKY genes. The exons and introns are indicated by yellow rectangles and black lines, respectively. The untranslated regions (UTRs) are indicated by blue lines.

Fig 5. Chromosomal location of WRKY genes in the *Citrus sinensis* genome (*Cits*WRKY). The chromosomes are represented by vertical bars and the numbers to the left indicate the positions of WRKY genes (MB). The red bars and green lines indicate tandem and segmental duplications, respectively. Nine chromosomes (Chr1-Chr9) and their respective WRKY genes are mapped.
Fig 6. Synteny map showing the orthologous gene positions of WRKY genes between *Citrus sinensis* and *Arabidopsis thaliana*. *C. sinensis* (Cs01 to Cs9) and *A. thaliana* chromosome (At01 to At05) maps were based on orthologous sequences and demonstrate highly conserved synteny. Each block represents an individual chromosome and the lines connecting two chromosomal regions denote syntenic regions of genomes.

Fig 7. *In silico* expression of WRKY family genes in different *Citrus sinensis* tissues.
whereas motifs 8 and 10 were found only in individuals of groups Ic and Ib, respectively (Fig 3; Table 2). These similarities in motif distribution may be related to the similar functions of WRKY proteins within the same subgroup.

**Genome structure and gene location**

Introns and exons are known to play important roles in the evolution of several gene families. Therefore, in order to obtain knowledge on the structural diversity of WRKY genes in citrus, the exon/intron structure in each member of CitsWRKY family in citrus was analyzed (Fig. 4). Based on the predicted structures, three genes, CitsWRKY17, CitsWRKY37, and CitsWRKY46 did not contain any introns. Although the number of introns varied from one (CitsWRKY15, CitsWRKY26, CitsWRKY28, CitsWRKY41, CitsWRKY43, CitsWRKY56, CitsWRKY57, CitsWRKY67, CitsWRKY76) to five (CitsWRKY93, CitsWRKY13, CitsWRKY14, CitsWRKY61), in general, the members with the highest numbers belonged to the same group, with the same intron/exon numbers, position, and size of preserved introns observed within subgroups (Fig. 4). For example, the structure of WRKY genes in group II seemed to be more variable, where the intron number ranged from 3 to 5. Conversely, members of group III only contained 1 or 2 introns. Most genes with introns were grouped together in the same group suggesting evolutionary conservation in C. sinensis.

Physical mapping of 77 CitsWRKY genes revealed a random distribution among nine C. sinensis chromosomes. Among the three groups, members of group I and II are present on nine chromosomes, whereas members of group III are found only in chromosomes 5 and 9 (Fig. 5). Among all chromosomes, chromosome 2 possessed the greatest number of genes (16 CitsWRKY genes), while chromosome 8 possessed a single gene (CitsWRKY08). Three genes were found on chromosomes 1 and 9, (CitsWRKY32, CitsWRKY73 and CitsWRKY20) on chromosome 1, and CitsWRKY75, CitsWRKY29 and CitsWRKY19 on chromosome 9). Two genes were found on chromosome 3 (CitsWRKY55 and CitsWRKY56) (Fig. 5). Expansion analysis of the WRKY gene family in the C. sinensis genome was examined. As shown in Fig. 5, 14 clusters of duplicated CitsWRKY genes (marked by red bars) in tandem were identified on chromosomes 2, 3, 4, 5, 6, and 7, while five segmental duplication events (marked by green lines) occurred on chromosomes 2, 4, 5, and 7.

A substantial number of Arabidopsis WRKY genes were systematically investigated (Eulgem et al., 2000) and an analysis of syntenies among WRKY genes of Arabidopsis and sweet orange was performed in this study (Fig. 6). A total of 37 pairs of syntenic relations were identified, including 19 AtWRKY genes and 18 CitsWRKY genes of Arabidopsis and sweet orange, respectively. Four AtWRKY genes (AtWRKY27, AtWRKY41, and AtWRKY36) and six CitsWRKY genes (CitsWRKY01, CitsWRKY25, CitsWRKY44, CitsWRKY60, CitsWRKY69, and CitsWRKY71) were found to be associated with at least two synteny events in which clear losses and duplications could be observed. Moreover, no synteny was observed in 66 CitsWRKY genes, suggesting that these orthologous genes were formed following the divergence of Arabidopsis and C. sinensis.

**In silico analysis of gene expression**

Bioinformatic analysis was performed using EST data. WRKY TFs display standard levels of special and temporal expression in different tissues, organs, and development stages, and in response to biotic and abiotic stresses. Expression analysis showed that these genes frequently occur in developing tissues such as fruit, flowers, ovaries, meristem, and phloem (65%, 66%, and 85% in groups I, II, and III, respectively). In addition, they are stimulated by biotic and abiotic stresses, such as pathogen exposure (15%, 16%, and 10% in groups I, II and III, respectively) (Fig. 7).

**Discussion**

In this study, 77 WRKY genes were identified in the C. sinensis genome, 26 genes were more than Ayadi et al. (2016) (51 genes WRKY). The number of WRKY genes identified in sweet orange was similar to that in A. thaliana (72), higher than that in V. vinifera (59), and lower than that in O. sativa (96), Zea mays (119) and Populus trichocarpa (104). This broad distribution of WRKY genes in different species of the vegetal kingdom indicates that these genes are highly conserved and suggest that they have important functions linked to development and stress responses (Chen et al., 2010; Chen et al., 2012). Generally, the distribution of WRKY genes in monocotyledons was almost two-fold of that in dicotyledons, possibly due to a greater rate of duplication events involving these genes during evolution. Moreover, the number of WRKY genes in citrus is not associated with genome size. Although C. sinensis has a much larger genome (319 Mb) (https://www.citrusgenomedb.org) than A. thaliana (145 Mb), the structure and phylogenetic arrangements were very similar for this gene family. In fact, irregular gene distribution of the WRKY family among different species of plants could have evolved from genome losses or duplication events of a common ancestor, which contributed to the expansion or limitation of these genes among different species.

In order to study evolutionary relations among WRKY proteins of sweet orange and known WRKYs of A. thaliana, a phylogenetic tree was constructed based on the alignment of WRKY protein sequences of sweet orange and Arabidopsis. As shown in Figs. 2 and 3, 77 CitsWRKY proteins were classified into three main groups. This is consistent with the classification of WRKY family members in A. thaliana (Eulgem et al., 2000), O. sativa (Ross et al., 2007), Z. mays (Wei et al., 2012), and V. vinifera (Wang et al., 2014).

Sequence alignment and phylogenetic analysis showed that most CitsWRKY proteins belong to groups I (36), II (35) and III (6), respectively. A total of 24, 12, 4, 11, 13, seven, and six proteins were attributed to subgroups Ia, Ib, Ia, I, Ic, IId, and IIIa, respectively. Analysis of preserved domains revealed that most WRKY genes in group Ia possess two preserved WRKY domains, with the exception of three genes (CitsWRKY07, CitsWRKY54, and CitsWRKY53). This was also reported in A. thaliana (Eulgem et al., 2000), Z. mays (Wei et al., 2012).and M. esculenta (Wei et al., 2016). This can occur because the WRKY domain in the C-terminal region of group I genes has a specific DNA-binding sequence containing a W-box, whereas the WRKY domain of the N-terminal region presents weak DNA binding (Maeho et al., 2001). The conserved sequence of the C-terminal region domain of A. thaliana AtWRKY4-Cis directly involved in DNA binding.

The N-terminal WRKY domain can participate in DNA binding or, alternatively, provide an interface for protein-protein interactions, which coincides with the function of some zinc finger-like domains (Mackay and Crossley, 1998). Thus, the WRKY domain of the N-terminal region seems to have been more variable throughout evolution, and may
therefore have been excluded from the sequence or evolved to another standard to obtain additional functions. Xie et al. (2005) reported that some WRKY genes in O. sativa from groups II and III may have evolved from group I following the loss of the WRKY domain in the N- or C-terminal region. Thus, it can be deduced that a group may contain original genes from other groups and that CitsWRKY107, CitsWRKY54, and CitsWRKY53 emerged later during the evolutionary process.

Structural diversity provides additional evidence of phylogenetic groupings, besides evolutionary information, for multigene families in which the gain or loss of an exon/intron can result from rearrangements and/or fusions of different chromosomal regions (Wang et al., 2014). The most closely related members of the same group, in general, showed the same standard exon/intron structure, such as number, position, and size of intron. This similarity may be the consequence of a series of duplication events, with the evolutionary conservation of WRKYTFs in C. sinensis considering the 77 WRKY genes in sweet orange, 37 contain two introns (Fig. 4). This was also observed in O. sativa and H. brasiliensis, in which 42 of 92, and 40 of 81 WRKY genes contain two introns, respectively (Xie et al., 2005; Li et al., 2014). This can be explained by the introns-early hypothesis, where by Rogozin et al. (2005) showed that the localization of introns may be conserved between different species if the introns already existed in a common ancestor of monocotyledon and dicotyledons. Therefore, with regard to the introns-early hypothesis, there is a significant similarity between WRKY transcription factors from O. sativa, H. brasiliensis, and C. sinensis in terms of genetic structure. Marked differences in the properties of CitsWRKY proteins, such as the number of amino acids, molecular weight, and isoelectric point of these proteins were observed and these variations can be attributed to the presence of new variants. Ten motifs, identified by the MEME algorithm, were annotated in InterProScan. Annotation of structural and functional proteins in the database showed that motifs 1, 2, and 3 are part of the DNA-binding domain, WRKYTFs. This was expected since it is a common characteristic of the WRKY family. In motif 8, a cluster of zinc finger-like motifs common to the WRKY family was found. As expected, most closely related members of this family on the phylogenetic tree had motifs in common, suggesting functional similarities among WRKY proteins within the same group and subgroup (Fig. 3).

Tandem and segmental duplications may have been the main causes of gene family expansions in plants (Cannon et al., 2004). The existence of two or more genes located on the same chromosome is considered to be a tandem duplication event, whereas gene duplication among different chromosomes is considered to be a segmental duplication event (Liu et al., 2011). According to Holub (2001), tandem duplications are 200-kb regions of a chromosome that contain two or more genes with high similarity. Therefore, 14 tandem duplication events of paralogous CitsWRKY genes were identified on chromosomes 2, 3, 4, 5, 6, and 7 (Fig. 5). Five segmental duplications with high similarity were detected among paralogous CitsWRKY62 and CitsWRKY64 genes on chromosome 2, CitsWRKY08 and CitsWRKY11 on chromosome 6, CitsWRKY23, CitsWRKY24, and CitsWRKY69 on chromosomes 2 and 4, CitsWRKY18, CitsWRKY28, and CitsWRKY57 on chromosomes 2 and 7, and among CitsWRKY44, CitsWRKY45, and CitsWRKY46, and CitsWRKY47, CitsWRKY48, and CitsWRKY49 on chromosomes 5 and 7, respectively. These results suggest that tandem duplication was predominant during the evolution of the WRKY gene family in oranges; however, segmental duplication may also have contributed to the expansion of this genetic family.

During the analysis, 37 pairs of synteny relations were observed and 10 genes, four from group IIIa, including two AtWRKY (AtWRKY41 and AtWRKY46) genes and two CitsWRKY genes (CitsWRKY25 and CitsWRKY71), presented more than one synteny event (Fig. 6). This indicates that WRKY genes are preserved in group IIIa, which suggests that this group may have functions that are fundamental to plant development. These synteny events suggest that CitsWRKY genes emerged before the divergence of Arabidopsis and sweet orange.

Based on EST annotations (Fig. 7), standard gene expression of 77 WRKY genes in C. sinensis was studied. In silico analysis of ESTs revealed that 27, 32, and 37% of genes in groups I, II, and III, respectively, were expressed in fruit libraries. Wang et al. (2012) showed that the MaWRKY/TF is involved in the maturation of banana fruit. TFs of the MaHST gene interact with MaWRKY1, promoting banana fruit ripening. This suggests that WRKYTFs can interact with other genes that are responsible for this physiological process. Moreover, studies have shown that several WRKY proteins can induce plant tolerance to biotic and abiotic stresses (Guo et al., 2004). In this study, in silico analysis showed that 14%, 16%, and 10% of ESTs for groups I, II, and III, respectively, were expressed in libraries of tissues exposed to stress. These genes contain the WRKYGAK sequence preserved in their DNA-binding domain that can bind to the W-box (TTGAC), regulating target gene expression (Eulgem et al., 2000). The super-expression of WRKY genes in Arabidopsis activates the expression of other genes related to water deficit, low temperatures, and tolerance to soil salinity (Jiang and Deyholos, 2009). Super-expression of the GmWRKY54 gene from Glycine max in transgenic Arabidopsis induced WRKY expression, resulting in enhanced tolerance to salt and drought (Zhou et al., 2008). Similar results were observed for the TaWRKY2 and TaWRKY19 genes of Triticum aestivum. The super-expression of these genes induced tolerance to salt, drought, and freezing in transgenic Arabidopsis, plants, because these genes activate the expression of different target genes, such as STZ, RD20B, DREB2A, RD29A, and COR 6.6 (Niu et al., 2012). Three WRKYTFs of Carica papaya, TF 12,199, TF 807.3, and TF21.156, were regulated in response to ringspot virus infection of C. papaya (Pan and Jiang, 2014). Thus, future studies should consider CitsWRKY TFs found in these libraries as promising candidates to improve the regulation mechanisms of WRKY genes in C. sinensis under different types of stresses.

**Materials and methods**

**Identification of WRKY genes in the Citrus sinensis genome**

WRKY protein sequences from the model plant A. thaliana were downloaded from the The Arabidopsis Information Resource (TAIR) Database (https://www.arabidopsis.org/). The WRKY domain sequences of Arabidopsis were utilized as query sequences to identify WRKY protein in the Citrus sinensis genome from the Phytozome Database of C. sinensis (http://www.phytozome.net), utilizing the BLASTP algorithm. The sequences were compared to others stored in GenBank database using BlastP and BlastX programs (National Center for Biotechnology Information, NCBI, http://www.ncbi.nlm.nih.gov) to confirm their homology. The initially collected sequences, whose ORFs were
Phylogenetic analysis
Phylogenetic analysis was performed by aligning the protein sequences and the phylogenetic tree was built using MEGA software, version 6.06 utilizing neighbor-joining (NJ), p-distance substitution model, and pairwise deletion. The reliability of phylogenetic trees was tested using bootstrap with 1000 replicates.

Protein characteristics and subcellular location of WRKY genes in C. sinensis

The physical and chemical characteristics of WRKY proteins in C. sinensis were calculated (http://web.expasy.org/compute_pi) including the number of amino acids (aa), molecular weight (MW), and theoretical isoelectric point (pI). Sequences of predicted WRKY proteins were analyzed in silico in terms of their subcellular location using WoLFPSORT algorithm (http://wolfpsort.org).

Conserved motif analysis
Conserved WRKY protein motifs predicted for C. sinensis were identified using the MEME algorithm (http://meme-suite.org/). The resulting motifs were verified in the NCBI database (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and using PROSITE (http://www.expasy.org/psort) to verify their significance.

Genome structure
In order to identify the intron position and exon/intron organization of genes, full-length sequence and coding sequence corresponding to each WRKY gene were aligned through Gene Structure Display Server (GSDS) (http://gsds.cbi.pku.edu.cn/).

Chromosomal location, duplication, and synteny of WRKY genes in the C. sinensis genome

C. sinensis WRKY genes (CitsWRKY) were mapped on chromosomes of sweet orange according to their coordinates in the genome. The genes were plotted on chromosomes using MapChart software, and data on their physical location are available in NCBI Mapviewer (http://www.ncbi.nlm.nih.gov/projects/mapview/). CitsWRKY genes were blasted against each other and those presenting 80% coverage and 70% identity were considered to represent gene duplications (Lopes-Caitar et al., 2013; Fan et al., 2015). The Plant Genome Duplication Database (PGDD; http://chibba.agtec.uga.edu/duplication/) was utilized to identify orthologous WRKY genes in A. thaliana and C. sinensis. A synteny map was built using Circos software, v. 0.63 (http://circos.ca/).

In silico analysis of gene expression
Expressed Sequences Tags (ESTs) were used to detect WRKY gene expression patterns. Nucleotide sequences of predicted sweet orange WRKY genes were utilized as query sequences to perform BLASTN in the NCBI EST database (http://www.ncbi.nlm.nih.gov/nucest) of C. sinensis. All redundant sequences were eliminated. Expression in different tissues was determined by analyzing information obtained through NCBI EST.

Conclusion
In conclusion, in this study, a total of 77 WRKY genes were identified from the C. sinensis genome. Based on WRKY sequence alignment and phylogenetic analysis, these proteins were classified into three groups I, II, and III. Analysis of preserved domains showed that most identified CitsWRKY genes contain a conserved WRKY domain, as well as additional motifs in some proteins. Structural analysis of genes showed that the number of introns in CitsWRKY genes varied from zero to five. WRKY genes mapped in the C. sinensis genome were randomly distributed among nine chromosomes. Few synteny events between A. thaliana and C. sinensis were identified and multiple duplication events of CitsWRKY genes occurred following the divergence of Arabidopsis and sweet orange, which were fundamental for the expansion of the WRKY family. Expression analysis revealed that these genes frequently occur in the libraries of plant development tissues and are also stimulated by some types of biotic or abiotic stresses. During genetic evolution, characteristics that allow plants to survive under diverse environmental conditions were selected. This study provides information on the WRKY family in C. sinensis, which may be useful in the selection of candidate genes for future functional analysis of the WRKY family in citrus. This will enhance understanding of the genetic determinants of tolerance to biotic and abiotic stresses as well the different physiological processes. As members of the WRKY family are expressed in an orchestrated manner, it is possible that, through genetic engineering, the manipulation of one single gene alter the function of the whole plant, resulting in more tolerant genotypes to different stresses that affect orange cultivation.

Acknowledgments
To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the grants to TMI and to the Universidade Paraense (UNIPAR) for the research’s financial support.

References


Wei KF, Chen J, Chen YF, Wu LJ, Xie DX (2012) Molecular phylogenetic and expression analysis of the complete
YWY transcription factor family in maize. DNA Res. 19:153–164.