Genome-wide analysis of Golden2-Like transcription factor gene family in *Gossypium hirsutum*

Title: The title is not appropriate, being "genomewide" provides a notion that the given gene family has been identified in the whole cotton genomes, basically to carry out a genomewide studies, it important to include the A and D genome, being they played a critical for the evolution of the AD, which is has been documented in this manuscript. So I strongly advice the authors to include the A and D genomes, and if not included then the term "genomewide" has to change. The possible titles to adopted by the authors as shown below.

i. Analysis of the Golden-2-like transcription factors gene family in *Gossypium hirsutum*
ii. Identification of the Golden-2-like transcription factors gene family in *Gossypium hirsutum*
iii. Evaluation of the Golden-2-like transcription factors gene family in *Gossypium hirsutum*

Abstract: Basically an abstract is a summary of the entire manuscript, in this work; the authors have briefly introduced the gene in question and immediately went into results without showing the methods adopted. Moreover, the entire abstract section is too shallow, given an indication that the authors never put much effort in this work, and therefore, the authors have to completely redo the abstract section.

Introduction: The introduction section is not well written, it is so difficult for the future readers of this manuscript if published to understand the information put across. The literature cited is too old, and it is paramount of the authors to know that the gene or the TF in question has been widely studied as shown by the following authorities;

i. Chen et al 2016. GOLDEN 2-LIKE Transcription Factors of Plants. Observed the Golden2-like (GLK) transcription factors are members of the GARP family of Myb transcription factors with an established relationship to chloroplast development in the plant kingdom
ii. Ahmad et al 2019. GOLDEN2-LIKE Transcription Factors Regulate WRKY40 Expression in Response to Abscisic Acid. Shown their role in ABA.

Thus it is critical for the authors to integrate and cite recent and relevant literature. Moreover, a number of statement are not supported by any literature for instance

Line 31-33

Line 55 – 56

Line 56- 58

Language and continuity is lacking, moreover repetition such as in "line 38-39 and line 49 "*GLK* transcription factor was first identified in maize (Jenkins et al., 1926)" can be minimized and the sentence rephrased without losing the intended purpose.

Line 87: "So far, no report of *GLK* gene family in *G. hirsutum* was identified" rephrase to correct the grammar and be in line with the context.

In the entire manuscript, the authors need to differentiate between "gene" and transcription factor (TF). And they have to maintain one form throughout the document.

Materials and methods

Databases: to be corrected to database, only a single database is shown in the manuscript. Moreover, this section should be merged with "the genomewide identification of the GLK genes in *Gossypium hirsutum*" .

The section for genomewide identification…. Should be corrected to state "Identification of the GLK genes in *Gossypium hirsutum*"

A serious concern of the use of the two protein domain "PF00249.31 and PF14379.6" in which PF00249.31 Myb_DNA-binding (PF00249) - Pfam: Family and PF14379.6: Myb_CC_LHEQLE; MYB-CC type transfactor, LHEQLE motif, and yet the gene or the TF family under study is the GLK: can the authors explain the mismatch, because this raises serious question of the identification of the gene or the TF under study.

Phylogenetic tree analysis: The authors have used protein sequences obtained from *A. thaliana*, *tomato* and *G. hirsutum*. The very first questions which rings into the mind of the reader, what was the justification of using the protein sequences from these three plants to carry out the phylogenetic tree analysis, moreover, is tomato evolutionary closer to cotton compared to *Theobroma cacao*. The second serious concern is no information is provided within the manuscripts on how the protein sequences from Tomato and Arabidopsis was obtained.

This section is poorly done, and must be redone, with protein sequences obtained from amore closer relatives to cotton.

Promoter cis-elements analysis of *GLK*: Change to *Cis*-regulatory elements. Is true as pointed by the authors that through the determination of the *cis*-regulatory elements one is able to determine the function of the gene or the TF, the answer is no, so the authors need to rephrase.

Differential gene expression analysis: The authors need to simply state "RNA seq expression analysis" reason for doing this not stated not indicated, and furthermore, it is a secondary data already in the public domain, unless the data was generated by the authors, this need to be clearly shown in order to improve the validity and novelty of the submitted manuscript

Stress treatments and qRT-PCR analysis: The treatment initiated “cold, salt and drought” and yet the RNA seq used as the reference point were profiled after the plants were exposed to “low temperature, salt and drought”, the question to the authors can low temperature be equated to cold stress?

Results: This section can only be said to be correct once the materials and methods section is correctly done.

Recommendation:

Substantial revision is needed before the paper could be reviewed again, and if the authors are not able to carry out all the issues stated, I fully recommend rejection. Moreover, the manuscript is too shallow, if possible the authors could be advised to validate the key genes or the TFs as per the RNA seq. and RT-qPCR analysis.