# Multiple Paths to Drought Resistance in Wheat: Zinc finger protein gene TaZFP13D Is Important, but Redundant

Boopalakrishnan Arul

**Abstract:** Common wheat or bread wheat (Triticum aestivum) has many methods of responding to stresses caused by drought. I investigated whether the under- or over-activation of one method of dealing with drought would lead to compensatory changes in other drought response pathways. It was found that when the gene TaZFP13D was silenced, changes in other genes typical of the “drought response” were intensified, and when the same gene was overexpressed, other genes either did not respond or responded in a direction opposite to the “drought response”. This result shows that for an important function like drought resistance, the effect of knocking out one method, whether it occurs naturally or is induced for a study, is offset by the continued and heightened effects of other pathways.

**Introduction:** Wheat, a very important crop, is subject to biotic stresses (bacterial and fungal infection, competition from weeds, and predation from pests) and abiotic stresses– of which one of the most important is drought, which increases sensitivity to salt, cold, and other abiotic stresses, which in turn intensify the effects of water stress. Reactive oxygen species (O2, O2-, H2O2, and HO) form during drought stress, and while at low levels they may be components in stress-signaling pathways, in excess they can oxidize proteins, lipids, and nucleic acids in an uncontrolled and potentially harmful ways (Cruz de Carvalho 2008). Some studies have identified genes that contribute to signaling pathways that lead to ROS scavenging, removing these molecules before they cause too much damage (Sanad et al 2020). Others have studied the effects of artificial overexpression or underexpression of genes as means of improving abiotic stress tolerance (Bouard and Houde 2022).

This study aimed to investigate if, when one pathway for dealing with reactive oxygen species varies in effectiveness, other genes belonging to other pathways are differentially expressed to compensate. In their study of C1-2i C2H2 zinc finger proteins (ZFP), Bouard and Houde found that the gene TaZFP13D was closely related to drought, high light, and H2O2 stress tolerance, which went up when TaZFP13D was upregulated and suffered when TaZFP13D was downregulated. Bouard and House argue that because TaZFP13D downregulation decreased drought stress tolerance, this gene doesn’t have a redundant function within the TaZFP family, and no other gene is able to cover for it. However, Sanad et al identify several genes involved in the creation and activation of peroxisomes, which eliminate rogue H2O2 molecules. This study therefore aims to investigate Bouard and Houde’s assumption, by detecting whether some of the genes identified by Sanad et al are differentially expressed depending on whether TaZFP13D is suppressed or not. This will help to gauge how robust wheat drought resistance mechanisms really are, and suggest avenues for improvement.

One complication in this study turned out to be the competing efforts to develop a definitive wheat reference genome. The International Wheat Genome Sequencing Consortium’s reference genome was completed in August 2018, and it is currently the favored version available for download on Ensembl, with the older TGACv1 assembly left in the archive (Guan et al 2020). While most of the existing software tools are meant for use with the IWGSC gene IDs, Sanad et al uses the older TGACv1 IDs to identify genes of interest. The significance of this will be explained further in the Methods and Results section.

**Methods:**

| Description of Dataset | Size of Dataset | NCBI SRA Run ID |
|---|---|---|
| Wild type wheat | 2.6 Gb | SRR23719836 |
| Wheat treated with empty BSMV vector | 2.1Gb | SRR23719837 |

| | | |
|---|---|---|
| Wheat treated to silence TaZFP13D (siRNA), rep 1 | 2.5G | SRR23719831 |
| Wheat treated to silence TaZFP13D (siRNA), rep 2 | 2.6G | SRR23719832 |
| Wheat treated to overexpress TaZFP13D (OEX), rep 1 | 2.5 Gb | SRR23719838 |
| Wheat treated to overexpress TaZFP13D (OEX), rep 2 | 2.5G | SRR23719839 |

Table 1: Summary of datasets from Bouard and Houde, used in this study.

For their study, Bouard and Houde used the Illumina NovaSeq 6000 to assemble 12 sets of paired end reads, each of which comes in a pair of fastq files. Six of these are from the second leaves of 21-days old wheat seedlings of plants in control conditions, and six are from similar plants in drought conditions. Within the 6 “drought” data sets, one data set is from plants that are not given any other treatment. Another dataset is treated with the Barley Stripe Mosaic Virus (BSMV) expression system, but with an empty vector, leading to no effect at all. This study will treat the wild type and empty-BSMV datasets as the two “control” reps. The remaining 4 datasets are 2 reps each of virus-induced gene overexpression and gene silencing of TaZFP13D.

In this study, the first step was downloading the cDNAs of the IWGSC assembly of the wheat (Triticum aestivum) genome from Ensembl. This was necessary to create an annotated index with Kallisto, a program for quantifying abundances of transcripts in RNA sequencing data. Afterward, each of the 6 datasets in Table 1 was downloaded, and given to Kallisto for pseudoalignment and quantification according to the index created previously. A custom shell script was used to automate this process, as well as the removal of data which had already been quantified. After being combined into a single file with a shell command, the raw counts of transcripts associated with each of the target genes in Kallisto’s index were uploaded to Degust, a browser based tool for differential expression analysis. In Degust, these targets were filtered based on the fold changes in gene expression (measured in transcript abundance) across two possible pairings: control vs. siRNA, and control vs. OEX. Each of these three groups is represented by the average across the 2 provided reps. From Degust, the list of genes within each pairing for which the p-value was below 0.05 (for which the fold change was deemed significant) was downloaded. Lastly, the PANTHER classification system was used for gene ontology, to find the functions of these genes which had significant fold changes.

| Gene, chromosome locations | Name, function | Up/downregulated in drought? |
|---|---|---|
| *TaLHB1B2 (-7AS, -7BS, -7DS) | photosystem II light-harvesting complex gene (LHB1B2) | Down |
| *TaLHCA1 (-7AS, -7DS, -7BS) | Photosystem II light-harvesting complex gene 1 (LHCA 1) | Down |
| *TaPsbR (-6AL, -6BL) | Photosystem II (PSII) subunit R (PSBR) | Down |
| *TaABA1 (-2BL, -2AL, -2DL) | Zeaxanthin epoxidase (ZEP), ABA1 gene | Down |
| *TaABA4 (-3B) | (Aba)-deficient 4 (ABA4 gene) | Up |
| *TaAAO3 (-7DL, -5DL, -7AL) | Abscisic aldehyde oxidase 3 (AAO3 gene) | Down |

| | | |
|---|---|---|
| *TaPEX11A (-4AS, -4DL) | Peroxisomal biogenesis factor11 (PEX11) gene family--PEX11.A-D | Up |
| *TaPEX11B (-2AL, -2DL, -2BL) | Peroxisomal biogenesis factor11 (PEX11) gene family--PEX11.A-D | Down |
| *TaPEX11C (-4BL, -4DL) | Peroxisomal biogenesis factor11 (PEX11) gene family--PEX11.A-D | Down |
| *TaPEX11D (-5AL) | Peroxisomal biogenesis factor11 (PEX11) gene family--PEX11.A-D | Down |
| *TaBIGYIN/TaFIS1A (-4BL, -1AL) | Mitochondrial fission 1 protein A (FIS1A/BIGYIN1) | Up |
| *TaDRP3A (-3AL) | Dynamin related proteins (DRP) | Down |
| *TaDRP3B (-2DL) | Dynamin related proteins (DRP) | Up |
| *TaDRP5A (-3AL, -3B) | Dynamin related proteins (DRP) | Up |

Table 2: Genes of interest in Sanad et al.

Table 2 summarizes what genes this study is looking for in the Bouard and Houde data, and whether these genes are upregulated or downregulated (relative to control conditions of no drought) in drought conditions. Based on this, we can see what a “drought response” consists of: downregulation in photosynthesis related genes, upregulation in certain ABA synthesis, peroxisome synthesis. We can see how characteristics of these genes change when the drought conditions remain, but TaZFP13D is silenced, overexpressed, or unaffected relative to the control.

Importantly, Sanad et al uses the TGACv1 gene IDs to identify genes of interest. The next step should have been to download the cDNAs from the Ensembl archive which carried the TGACv1 IDs instead of the IWGSC version, re-do the Kallisto index and quantification steps with the new annotations, and continue on. However, PANTHER uses IWGSC annotations for gene ontology, so that would not be possible. This need not be an issue, because Sanad et al already identifies the function of its genes of interest, so I could simply see if those IDs show up among the significant fold changes in Degust and skip the PANTHER step. Here, however, there was another problem: the Ensembl archive does not actually contain the TGACv1 cDNAs, just the README, and other copies of it did not seem to be maintained either. Finally I investigated whether it would be possible to convert the old IDs to the new one with the Assembly Converter or ID History Converter, but I couldn’t get this to work either. Therefore the approach to finding Sanad et al’s genes of interest in the Bouard and Houde data relied on the descriptions given in PANTHER, which were exported as a text file and searched for key terms of interest: photo* (photosynthesis, photosystem), light* (-harvesting), ABA, abscisic, peroxisom* (-al, -e), dynamin, fission.

**Results:**

| | Reads processed | Reads ps-aligned | % pseudoaligned | targets |
|---|---|---|---|---|
| WT, control rep 1 | n_processed = 31,007,077 | n_pseudoaligned = 23,450,966 | 75.6% | 133,744 |
| BSMV / empty vector, control rep 2 | 24,340,261 | 18,501,518 | 76.0% | 133,744 |

| | | | | |
|---|---|---|---|---|
| siRNA13D, rep 1 | 29,014,794 | 22,199,205 | 76.5% | 133,744 |
| siRNA13D, rep 2 | 30,427,773 | 23,363,541 | 76.8% | 133,744 |
| OEX13D, rep 1 | 28,062,825 | 21,386,193 | 76.2% | 133,744 |
| OEX13D, rep 2 | 28,962,737 | 21,865,803 | 75.5% | 133,744 |

Table 3: Results of alignment and quantification of each dataset with Kallisto.

| Gene IWGSC ID, Description | Fold Change in siRNA (TaZFP13D silenced) relative to control | fold change expected in drought (Table 2) |
|---|---|---|
| TraesCS6A02G374400, Photosystem II 10 kDa polypeptide, chloroplastic | -3.02 | Down |
| TraesCS3B02G039800, PROTEIN ABA DEFICIENT 4, chloroplastic | 1.53 | Up |
| TraesCS2A02G281000, DYNAMIN-RELATED PROTEIN 3A | Transcripts:<br>TraesCS2A02G281000.3, 1.21<br>TraesCS2A02G281000.1, -3.96 | Down |

Table 4: Genes with significant fold changes between the control data sets and the siRNA data sets, according to Degust, and which correspond to the genes mentioned in Sanad et al.

| Gene, IWGSC ID, Description | Fold Change in OEX (TaZFP13D overexp.) relative to control | fold change expected in drought (Table 2) |
|---|---|---|
| TraesCS2A02G281000, DYNAMIN-RELATED PROTEIN 3A | Transcripts:<br>TraesCS2A02G281000.3, 1.26<br>TraesCS2A02G281000.1, -3.81 | Down |
| (large number of genes related to photosynthesis) | Generally “Up” | Expected “Down” |

Table 5: Genes with significant fold changes between the control data sets and the OEX data sets, according to Degust, and which correspond to the genes mentioned in Sanad et al.

Table 3 is a summary of the files given to Degust for differential expression analysis. The average is over 20 million reads per file, aligned to one of 133,744 target genes with IWGSC gene IDs.

For the control and siRNA datasets, Degust found 2545 genes in which there was a significant ($p$-value $< 0.05$), and ontologies were found for 2359 of them by PANTHER, which could not find the IDs of 54 genes in its database. For the control and OEX datasets, PANTHER took in 2514 genes from Degust, found ontologies for 2360 genes, and IDs were not found for 58 genes.

Among those genes, some could be found which were marked out as genes of interest in Sanad et al. Table 4 illustrates that when TaZFP13D is silenced and not allowed to contribute to drought resistance, then in conditions of drought the genes identified by Sanad et al experience fold changes with the same sign, in the same direction, as those fold changes characteristic of the change from control to drought conditions. For example the ABA-deficient 4 protein gene, which is upregulated as part of the drought response, experiences a further statistically significant fold change of 1.53. Photosynthesis related genes which are downregulated as part of the drought response are further downregulated when

TaZFP13 is silenced. Therefore, although Bouard and Houde contended that TaZFP13D has no redundancies within its family of proteins, outside of that family there are other pathways of drought response that become even more active, and in the expected directions, when TaZFP is silenced.

Table 5 illustrates the effects of TaZFP13D overexpression on the genes of interest. A large number of photosynthesis related genes actually experienced significant positive fold changes, when what would be expected of a drought response is downregulation of photosynthesis. This indicates that TaZFP is so important to drought response that even photosynthesis can be allowed, as the stresses it might result in or exacerbate are dealt with.

**Discussion:** This study investigated whether different pathways of drought resistance in Triticum aestivum become more active when one pathway or another is silenced, or less active when another pathway is overexpressed. With the methods available, the results supported both of these assumptions. Changes made to deal with drought intensify when other methods of dealing with it are taken away, and when other methods are in turn intensified, functions which were a luxury in lean times are steadily reintroduced. Even if the TGACv1 gene annotations could not be consistently mapped to the IWGSC ones, this result is supported by the gene descriptions of the IWGSC genes discovered through PANTHER, which match those provided in Sanad et al. While a future study on this line could locate other "genes of interest" significant to newly discovered drought pathways, another approach may be to use PANTHER's own categorizing capabilities to cast a much wider net of genes related to key drought response functions.

Although fold change of gene expression could gauge the direction of changes, phenotypic measurements might have allowed more accurate assessment of the magnitude. Furthermore, it remains to be investigated why one splice variant of dynamin related protein 3A is upregulated and another is downregulated, and that these would not switch signs when TaZFP13D was overexpressed. Dynamin is responsible for endocytosis, or bringing things past the cell membrane and into the cell– it may be that this remains an essential function even in drought conditions.

References:


1. Cruz de Carvalho MH. Drought stress and reactive oxygen species: Production, scavenging and signaling. Plant Signal Behav. 2008 Mar;3(3):156-65. https://doi.org/10.4161/psb.3.3.5536.
2. Marwa N.M.E. Sanad, Arwa Abdulkreem AL-Huqail, Mohamed M. Ibrahim, Amber L. Hauvermale, Prediction of the crosstalk regulation model between the abscisic acid (ABA) signaling and peroxisome abundance during drought stress in wheat (Triticum aestivum L.), Current Plant Biology, Volume 24, 2020, https://doi.org/10.1016/j.cpb.2020.100176.
3. William Bouard, Mario Houde, The C2H2 zinc finger protein TaZFP13D increases drought stress tolerance in wheat, Plant Stress, Volume 6, 2022, https://doi.org/10.1016/j.stress.2022.100119.
4. Jiantao Guan, Diego F. Garcia, Yun Zhou, Rudi Appels, Aili Li, Long Mao, The Battle to Sequence the Bread Wheat Genome: A Tale of the Three Kingdoms, Genomics, Proteomics & Bioinformatics, Volume 18, Issue 3, 2020, https://doi.org/10.1016/j.gpb.2019.09.005.
5. Ensembl Plants release 56 - Feb 2023 https://plants.ensembl.org/Triticum_aestivum/Info/Index
6. Ensembl Plants Archive release 37 - September 2017 http://oct2017-plants.ensembl.org/Triticum_aestivum/Info/Index
7. Nicolas L Bray, Harold Pimentel, Páll Melsted and Lior Pachter, Near-optimal probabilistic RNA-seq quantification, Nature Biotechnology 34, 525–527 (2016), doi:10.1038/nbt.3519
8. David R. Powell. Degust: interactive RNA-seq analysis, DOI: 10.5281/zenodo.3258932.
9. PANTHER: Making genome-scale phylogenetics accessible to all, Paul D. Thomas, Dustin Ebert, Anushya Muruganujan, Tremayne Mushayahama, Laurent-Philippe Albou and Huaiyu Mi Protein Society. 2022;31(1):8-22. doi:10.1002/pro.4218

Common wheat or bread wheat (Triticum aestivum) has many methods of responding to stresses caused by drought. I investigated whether the under- or over-activation of one method of dealing with drought would lead to compensatory changes in other drought response pathways. It was found that when the gene TaZFP13D was silenced, changes in other genes typical of the “drought response” were intensified, and when the same gene was overexpressed, other genes either did not respond or responded in a direction opposite to the “drought response”. This result shows that for an important function like drought resistance, the effect of knocking out one method, whether it occurs naturally or is induced for a study, is offset by the continued and heightened effects of other pathways.

| | | | | |
|---|---|---|---|---|
| Results | missing more than 2 components | missing 2 of components | has a few scientific errors/ misconceptions<br>-or -<br>is missing 1 of components | - has written text that clearly walks reader through results<br>- any figures are well labeled and easy to understand<br>- results document processing steps from methods (i.e. reads that pass each step, quality control, etc.)<br>**- results contribute to answering biological question** |
| Discussion | missing more than 2 components | missing 2 of components | has a few scientific errors/ misconceptions<br>-or -<br>is missing 1 of components | - interpreta the meaning of results in terms of the original biological question<br>- shows a strong understanding of the science<br>- discusses potential limitations of this analysis<br>- makes a suggestion for further research |

**Final Report (45% of grade) - Due Friday, June 9th, 11:59 PM (last day of week 10)**
Your final report should be structured like a scientific journal article. It should have a title and abstract, and must be broken down into introduction, method, results and discussion sections. You are required to use at least 4 scientific journal articles in your introduction, and you should cite any bioinformatics tools you use. Use in-text (parenthetical) citations next to the information from each source, and include a reference list for all of the citations at the end. There is a length limit (maximum) of 5 pages. (52 points).

| Gene, IWGSC ID | Description | Fold Change in OEX (TaZFP13D overexp.) relative to control | Fold Change expected in drought (from Sanad et al) |
|---|---|---|---|
| TraesCS2A02G281000 | DYNAMIN-RELATED PROTEIN 3A | Transcript TraesCS2A02G281000.3, 1.26<br>TraesCS2A02G281000.1, -3.81 | Down |
| TraesCS6B02G331700 | PHOTOSYNTHETIC NDH SUBUNIT OF LUMENAL LOCATION 4, CHLOROPLASTIC | 1.54 | Not mentioned in Sanad et al, but expected “Down” |
| TraesCS7B02G006600 | PHOTOSYNTHETIC NDH SUBUNIT OF SUBCOMPLEX B 3, CHLOROPLASTIC | 1.13 | |
| TraesCS7D02G549100 | PHOTOSYSTEM II STABILITY/ASSEMBLY FACTOR HCF136, CHLOROPLASTIC | 1.03 | |

| | | | |
|---|---|---|---|
| TraesCS5D02G396200 | PHOTOSYSTEM II D1 PRECURSOR PROCESSING PROTEIN PSB27-H2, CHLOROPLASTIC | 1.33 | |
| TraesCS4A02G355600 | PHOTOSYSTEM II REACTION CENTER PSB28 PROTEIN, CHLOROPLASTIC | Transcripts: TraesCS4A02G355600.1, 1.54 TraesCS4A02G355600.3, -2.71 | |
| TraesCS5B02G210800 | PHOTOSYSTEM I CHLOROPHYLL A/B-BINDING PROTEIN 6, CHLOROPLASTIC | Transcripts: TraesCS5B02G210800.1, 1.73 TraesCS5B02G210800.2, 1.31 | |
| TraesCS3A02G443700 | PHOTOSYNTHETIC NDH SUBUNIT OF SUBCOMPLEX B 5, CHLOROPLASTIC | 0.80 | |
| TraesCS4A02G101500 TraesCS4B02G203100 | PHOTOSYNTHETIC NDH SUBUNIT OF LUMENAL LOCATION 1, CHLOROPLASTIC | Transcripts: TraesCS4A02G101500.1, 1.15 TraesCS4A02G101500.2, 1.44 TraesCS4B02G203100.2, 1.52 TraesCS4B02G203100.1, -0.03 | |
| TraesCS3B02G477300 | PHOTOSYNTHETIC NDH SUBUNIT OF SUBCOMPLEX B 5, CHLOROPLASTIC | 1.03 | |
| TraesCS7A02G560200 | PHOTOSYSTEM II STABILITY/ASSEMBLY FACTOR HCF136, CHLOROPLASTIC | 1.00 | |
| TraesCS6A02G084200 TraesCS6B02G112800 | LIGHT-HARVESTING COMPLEX-LIKE PROTEIN 3 ISOTYPE 1, CHLOROPLASTIC | TraesCS6A02G084200.1, 0.71 TraesCS6B02G112800.1, 0.59 | |

| | | | |
|---|---|---|---|
| TraesCS3A02G081300 | LIGHT-REGULATED PROTEIN, CHLOROPLASTIC | 1.16 | |

Table 5: Genes with significant fold changes between the control data sets and the OEX data sets, according to Degust, and which correspond to the genes mentioned in Sanad et al.