nature research

Peer Review Information

Journal: Nature Genetics
Manuscript Title: Genome assembly and genetic dissection of a prominent drought-resistant maize germplasm
Corresponding author name(s): Professor Feng Qin

Reviewer Comments & Decisions:

Decision Letter, initial version:

4th Jan 2022

Dear Professor Qin,

Your Article entitled "Genome Assembly and Genetic Dissection of a Prominent Drought-Resistant Maize Germplasm" has now been seen by 3 referees, whose comments are attached. While they find your work of potential interest, they have raised serious concerns which in our view are sufficiently important that they preclude publication of the work in Nature Genetics, at least in its present form.

While the referees find your work of some interest, they raise serious concerns about the strength of the novel conclusions that can be drawn at this stage.

Should further experimental data allow you to fully address these criticisms we would be willing to consider an appeal of our decision (unless, of course, something similar has by then been accepted at Nature Genetics or appeared elsewhere). This includes submission or publication of a portion of this work someplace else.

The required new experiments and data include, but are not limited to those detailed here. We hope you understand that until we have read the revised manuscript in its entirety we cannot promise that it will be sent back for peer review.

In the future, if you are interested in attempting to revise this manuscript for submission to Nature Genetics, you may contact us with a point-by-point response to discuss a potential appeal. Otherwise, we hope that you find our referees' comments helpful when preparing your manuscript for resubmission elsewhere.

Sincerely,

Wei Li, PhD
Senior Editor
Nature Genetics
New York, NY 10004, USA
www.nature.com/ng

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
Tian et al., Genome assembly and genetic dissection of a prominent drought-resistant maize germplasm.

The authors describe the sequencing of the drought resistant maize cultivar CIMBL55. CIMBL55 has previously been characterized and a series of loci (97) have been detected using GWAS and reported as contributing to drought tolerance. Consequently the newly available CIMBL55 maize genome has been used as scaffold to zoom in into regions previously pinpointed using GWAS and SNPs and to detect individual gene candidates that might contribute to the import and much wanted physiological characteristic.
The ms. is well written and well structured, most of the methods used and applied are sound and state of the art and the genome sequencing, assembly and analysis reported is also state of the art in currently used and applied analytical frameworks.

Criticism:
- Line 54: SV identification in maize is challenging…. I disagree. Fairly standard task these days
- Line57 ff: a couple of additional important references might be suitable
- Lines 136ff: class II genes and their potential fractionation. Speculative. Would need to be tested and evaluated by suiatbel phylogenetic analysis to test the hypothesis.
- Lines 147: what is the conclusion from the observed expression differences among syntenic and non-syntenic genes? I doubt that a GO enrichment analysis only gives new insights.
- Lines 161: the whole section on SNP and SV detection asks for a complete rewrite. Subjective observations, lots of methods that would need to go into the M&M section (along with parameters), in part adventourous approaches that are not state of the art etc.
- Line179: genetic rather generic (I guess)
- Line 207: genetic variants located downstream of ZmABF4… I'm not clear on whether thio is meant as pathway downstream of genomic position downstream
- Line 3g: I'm not convinced about the argument of the longest downstream (3UTR) sequence for the drought resistant allele. Doesn't look as if in the figure.
- Line 220: Data confirm that the ZmABF4 allele contributes to drought resistance. Well I disagree. Thie evidence given is circumstantial. A mor cautious wording would help….
- line 235: …suggesting the involvement of Rna-directed DNA CHH methylation in these regions. Again this is no hard evidence. More cautious wording
- Figures 4a, 4b and line 242 ff: The higher level methylation is not obvious for me.
- Lines 276 ff: Again overinterpretation and more cautious wording needed
- Lines 281 ff: The link of the reticulon -like protein to drought resiostance is not obvious to me. From the potential molecular function a link is not too obvious.
- Line 292 ff: the 28bp insertion in the UTR: might affect gene expression (to be experimentally

validated) but might also affect the transcript stability.
- Line 366: 4% non-syntenic genes whose expression was either undetectable or extremely low… Well, does this eventually point to false positive annotations then?
- Line 369-371. Leave out.
- Line 387: …is a homodimer of a single protein. Leave out of a single protein.

Given the potential loci/genes detected and discussed in more detail (2/3) in the ms., the sheer amount of candidate loci reported previously and the obvious complex, polygenic foundation of the physiological characteristic, I remain skeptical about the ultimate proof of drought resistance contribution of the genes reported. Presence of SNPs or SVs in the surrounding are not a convincing proof. Traditional knock-out/overexpression or genome editing approaches to test and validate the candidates reported seems to be unavoidable to demonstrate and proof the contribution and causative molecular basis for the observed, -again complex and polygenic!- physiological phenotype. Also, reporting only 2 gene candidates and some SNPs/SVs in the sourrounding doesn't appear to be sufficient to discuss the complex, multi-genic phenotype.

Reviewer #2:
Remarks to the Author:
The authors generated a relatively high quality assembly of CIMBL55, a drought tolerant genotype, and then used his to do comparative genomics with B73 and Mo17, which are relatively less drought tolerant to identify causative variants in CIMBL55 at candidate loci that were previously identified from GWAS. The authors provide a number of important validation experiments for a few loci that demonstrate the contribution of these loci to the drought tolerance that is observed in CIMBL55. However, a few concerns remain as described below. In particular are concerns of interpretations of the data that are insufficiently supported.

1. Page 3 Line 51-53 – This is a misleading statement supported by a review article rather than primary literature. There is not nearly enough data to know if the majority of causal variants contributing to phenotypic variation will be from SVs.
2. Given the sequence depth the contig N50 is a bit lower than expected. Where the reads of relatively lower quality and/or length?
3. The vast majority of the comparative analyses were done relative to B73 and Mo17. There are many other assemblies that are available in maize that could be compared to, why only these two? In particular for the variant detection that were then screened on the larger set of lines there is a big ascertainment bias that comes from only comparing CIMBL55 to B73 and Mo17 when then moving to the larger panel.
4. There is a lot of time spent on documenting findings that have already been highly studied in previous genomic studies. For example, much of the results of the syntenic vs. non-syntenic genes has been shown in Schnable et al 2011, Brohammer et al. 2018, Hufford et al. 2021 etc. Likewise for the general genome-wide methylation patterns and patterns of methylation over TE sequences, which have been well documented in many different by the Springer group and others.
5. Page 8 line 194-198 – Please make it clearer if this is looking at survival rates in just CIMBL55 and B73 or in all individuals in the population that have either of these alleles. If the latter, how many individuals in the population does this account for, in other words are there other non-CIMBL55 or B73 alleles in the population that are not considered?
6. Page 9 line 203-206 – The authors need to be careful that their hypothesis is not driving their conclusions. Just because there is an SV in LD with a previously identified SNP doesn't mean that the

SV is necessarily causative.
7. Page 10 line 259 – Page 11 line 263 – This interpretation is a stretch and not supported by much of anything.
8. I found it hard at times to know what was new in this manuscript and what was from previous studies as there are entire supplemental figures made from previous data that support previous findings.
9. Page 11 line 275-278 – not enough data to support this claim.
10. The authors themselves state on line 406-407 – "Drought resistance in plants is a complex trait that can vary based on the timing and magnitude of the stress", yet the timing and magnitude of stress is different for different experiments in this study (e.g. the stress applied in the rainout shelters).

Reviewer #3:
Remarks to the Author:
The study by Tian, Wang et al produces a genome assembly for the CIMBL55 maize genotype and uses it to further explore its unique drought resistance phenotype. Numerous candidate alleles are identified that are associated with the drought resistance in CIMBL55. Two alleles are pursued in an attempt to identify causal genetic variants. One allele is associated with ZmNac075 and the other is associated with ZmRtn16. Although there are strong associations of genetic variants around the ZmNac075 locus with drought resistance, there is no evidence provided in this study to demonstrate which are potentially causal. A much more convincing set of experiments were performed to show causality of AmRtn16, which includes a reporter assay, overexpression and CRISPR knockouts.

Major comments:

1. There is a wide range of survivability data in comparisons between figure 3e, f, k and figure 5h and i. The wildtype survival is so variable. It ranges from 10-20% in figure 3 e and f yet the OE in 3k shoots up to 50-60%. Similarly, in figure 5 the WT survival is 50% and the KOs are 10-20% yet in 5h WT is 10% and OE lines at 30-40%. This level of variability is concerning especially as the WT is always higher or lower as needed to observe the strong impact of the mutant/transgenic alleles being tested.
2. There is no evidence that DNA methylation data presented at ZmNac075 is causal for repression of gene expression. I have no doubt that two new SVs were identified and that both are DNA methylated, but no data is presented to show causality of drought resistance of any genetic variant or associated DNA methylation around this locus. This highlights the cause-and-effect issue surrounding the study of the ZmNac075 allele. For example, Figure 4 e and f could be repeated with all SNPs (or any other genetic variant in the area) and it would reveal the same result due to linkage disequilibrium. There are lots of possible genetic variations in this 20kb region and none of them have been tested in isolation from one another. This prevents assignment of causality. At this stage, there is no evidence the insertions that are methylated are causal. It is very common that insertions are methylated, but very few examples exist to date demonstrating they are causal for expression variation of nearby genes.
3. The following line in the abstract must be removed as it is inaccurate. "Differential DNA methylation analysis indicated that hypermethylation of the ZmNAC075 promoter repressed gene expression and compromised drought resistance."

Minor comments:

1. Change all forms of "bisulfate" to "bisulfite"
2. Table s8 is missing bisulfite conversion rates.
3. The analysis described in 223-263 is worthy of pursuit, but these questions/results have been explored before numerous times between maize genotypes and in other plant species.
4. Manuscript requires proofreading for typos.

Although we cannot publish your paper, it may be appropriate for another journal in the Nature Portfolio. If you wish to explore the journals and transfer your manuscript please use our [redacted] manuscript transfer portal. If you transfer to Nature journals or the Communications journals, you will not have to re-supply manuscript metadata and files. This link can only be used once and remains active until used.

All Nature Portfolio journals are editorially independent, and the decision on your manuscript will be taken by their editors. For more information, please see our manuscript transfer FAQ page.

Note that any decision to opt in to In Review at the original journal is not sent to the receiving journal on transfer. You can opt in to In Review at receiving journals that support this service by choosing to modify your manuscript on transfer. In Review is available for primary research manuscript types only.

Author Rebuttal, Appeal

Response to Reviewer #1:

Remarks to the Author:

Tian et al., Genome assembly and genetic dissection of a prominent drought-resistant maize germplasm. The authors describe the sequencing of the drought resistant maize cultivar CIMBL55. CIMBL55 has previously been characterized and a series of loci (97) have been detected using GWAS and reported as contributing to drought tolerance. Consequently, the newly available CIMBL55 maize genome has been used as scaffold to zoom in into regions previously pinpointed using GWAS and SNPs and to detect individual gene candidates that might contribute to the import and much wanted physiological characteristic.

The ms. is well written and well structured, most of the methods used and applied are sound and state of the art and the genome sequencing, assembly and analysis reported is also state of the art in currently used and applied analytical frameworks.

The authors thank this reviewer for evaluating the manuscript and providing the thoughtful comments and constructive suggestions.

Criticism:

- Line 54: SV identification in maize is challenging…. I disagree. Fairly standard task these days

Response: As for this comment, we would like to respectfully clarify that structure variations (SVs) consist of many kinds, such as Inversions (INVs), Translocations (TRAs), Duplication (DUPs), Copy Number Variations (CNVs) and Presence and Absence Variations (PAVs), etc. Thus, accurate and complete SV identification between two complex genomes still remains challenging. To this end, multiple tools were developed to identify SVs, such as SyRi, Smartie-sv, CuteSV, Pbsv and Sniffles, etc, based on either the assembled genome sequences or long reads alignment. So far, there is no single toolkit competent for accurate and complete SV identification, especially for a complex genome. We revised the description as follows, in Line 57-60, Page 3.

"Given that > 85% of genetic content consists of complex repetitive sequences[17], accurate and complete SV identification in maize genome is challenging as it relies heavily on the existence of high-quality genome sequences and identification methods[18]."

- Line57 ff: a couple of additional important references might be suitable

Response: Thanks for the suggestion. Additional references regarding maize genome assemblies were added, in Line 63-66, Page 3-4.

"Recently, the germplasms of 25 founder lines of a nested association mapping (NAM) population[21], SK[22] (small kernel), A188[23], K0326Y[24] and four European flint maize lines[25] were *de novo* assembled, thus, providing a pan-genome view for maize."

- Lines 136ff: class II genes and their potential fractionation. Speculative. Would need to be tested and evaluated by suitable phylogenetic analysis to test the hypothesis.

Response: Thanks for the suggestion. To test and evaluate the potential fractionation of Class II genes in the CIMBL55 genome, "We construct two sub-genomes of CIMBL55, referring to the sorghum genome. Sorghum diverged from the maize lineage just prior to the tetraploid event, and has been commonly used as a reference for maize sub-genome analysis and fractionation bias estimation[20,33]. The CIMBL55 sub-genome organization and fractionation bias was found to be consistent with B73, supporting the evolution conservation within *Zea mays* (**Fig. S4a, b)**. Then, we found that 61.5% Class II genes were pair-retained gene in the two sub-genomes of CIMBL55, indicating that Class II genes remained relative conserved during maize evolution (**Fig. S4c**)." We appended this description in Line 155-163, Page 7. The new prepared Fig. S4a-c was appended.

- Lines 147: what is the conclusion from the observed expression differences among syntenic and non-syntenic genes? I doubt that a GO enrichment analysis only gives new insights.

Response: As for the observation that syntenic genes had a higher transcript abundance than non-syntenic genes based on RNA-seq data obtained from leaf samples, it indicated "that genomic

rearrangement likely impaired gene expression". We added the inference in the sentence, in Line 153, Page 7.

We agree that the GO enrichment analysis just gave an indication of the functional relatedness of class II genes. "To gain insights into it, we further investigated the gene synteny of several gene families involved in ABA-signaling and stress-response, including genes encoding 13 ABA receptor (PYLs)[35], 13 SnRK2s protein kinase[36], 103 protein phosphatases (PP2Cs)[37], and 55 Dehydration Responsive Element Binding proteins/C-repeat Binding Factor (DREB)[38], 130 NAM, ATAF, and CUC (NACs)[39], and 128 bZIPs[39] transcription factors (**Table S5**). Among them, 72% belonged to Class II genes, which was remarkably higher than 29% Class II genes on the whole genome level (**Fig. S5a-b, Fig. 2a**). These data indicate that genes involved in environmental response and adaptation tended to be retained in duplication during maize evolution." We appended Fig. S5a-b, and added the description in this manuscript, in Line 164-173, Page 7.

- Lines 161: the whole section on SNP and SV detection asks for a complete rewrite. Subjective observations, lots of methods that would need to go into the M&M section (along with parameters), in part adventurous approaches that are not state of the art etc.

Response: Thanks for the constructive suggestions. We rewrote this session and performed the genetic variation identification on a pan-genome level (as suggested by Reviewer #2), including 30 high-quality assembled maize genomes. Four kinds of state-of-art programs were applied to the identification of SVs in 30 maize germplasms, to achieve a maximum number of SVs identification. These strategies avoided subjective observations and adventurous approaches. The description was as follows, in Line 175-198, Page 7-8.

"In order to identify the genetic elements contributing to the prominent drought resistance of CIMBL55, we explored the genetic variants on a pan-genome level, including the newly released genome information of 25 NAM founder lines[21] and another four maize accessions, SK[22], Mo17[20], K0326Y[24], A188[23] (all of them assembled with contig N50 > 5.0 Mb). To identify a maximum number of genetic variants, the assembled genome sequence 30 maize genomes (and also the corrected long-reads sequences) were aligned and compared with B73_v5[21], through four programs, SyRi[40], Smartie-sv[41], CuteSV[42] and Sniffles[43]. As a result of this analysis, 17,581,014 SVs, including 16,311,409 Insertions/Deletions (InDels), 239,611 Duplications (DUPs), 496,685 Inversions (INVs), 499,973 Translocations (TRAs), 33,336 Copy Number Variations (CNVs) were obtained non-redundantly (**Fig. 3a, Table S6**). SyRi (based on the assembled genome sequence comparisons) showed the effectiveness in discerning various types of variations, while Smartie-sv (based on the assembled contig alignments) and CuteSV (based on the long-read alignments) performed better in identifying InDels (> 50-bp) than other programs (**Fig. 3a**). To gain more insight into the genomic variations in CIMBL55, we also called the genetic variants identified from B73 and Mo17, using CIMBL55 as a reference. As a result, 841,911 DNA variants (>20-bp) were subsequently identified (**Table S7**). We genotyped all of these genetic variants in the association panel consisting of 368 maize germplasms, of which the drought tolerance was previously analyzed regarding the seedling survival rate after drought[13,14]. Thus, a total of 544,853 SVs were successfully genotyped on the population scale by Paragraph[44] based on the 20× genomic DNA re-sequencing data of all these germplasms[22], with MAF (minor allele frequency) > 0.05, and missing rate < 0.6. "

- Line179: genetic rather generic (I guess)

Response: Sorry for the type-error, and should be genetic rather generic, in the last submission. In this reversion, we rewrote this session, in response to the previous suggestion.

- Line 207: genetic variants located downstream of ZmABF4… I'm not clear on whether this meant as pathway downstream of genomic position downstream

Response: In the previous submission, it meant genomic position downstream of *ZmABF4*, but not pathway downstream of *ZmABF4*.

- Line 3g: I'm not convinced about the argument of the longest downstream (3'-UTR) sequence for the drought resistant allele. Doesn't look as if in the figure.

Response: Thanks for the comment. We thoroughly re-analyzed all the genetic variants of *ZmABF4* and their association with plant drought-resistance. However, it was true that, so far, we cannot conclude that the variant in the 3'-UTR of $ZmABF4^{CIMBL55}$ was the most obvious causal variant. There were several significant SNPs and SVs identified within the gene locus associated with drought resistance, and their LD pattern was not clear. The data suggested that $ZmABF4^{CIMBL55}$ was probably the drought resistant allele, but the causal variant demands further investigations. The newly obtained results were appended Fig. S6a and Fig. 3d-e, and we rewrote this part in Result, Line 219-234, Page 9. It wrote as follows:

"The SNP (S1474) and SV (S3205, 42-bp) locating in the second intron of *ZmABF4* (*Zm00001d031790*) were found to be significantly associated with the gene expression and plant drought resistance (**Fig. S6a**). *ZmABF4* encodes a bZIP transcription factor, an ortholog of *Arabidopsis* ABA-responsive factor 4, which has been characterized as a major regulator of ABA- and drought-inducible gene expression[45,46]. The germplasms harboring the allele of $ZmABF4^{CIMBL55}$ genotype had a significantly higher level of gene expression and drought resistance, compared with those carrying the $ZmABF4^{B73}$ allele (**Fig. 3d-e).** Transgenic maize overexpressing the coding sequence of $ZmABF4^{B73}$ were generated to determine if increased gene expression conferred a drought-resistant phenotype **(Fig. S6b)**. Results showed that the

ZmABF4 transgenic lines had a remarkably enhanced survival rate after drought exposure compared to non-transgenic wild-type (WT) plants and reduced rates of leaf water loss in response to dehydration treatments. It indicated that *ZmABF4* positively regulated maize drought resistant and $ZmABF4^{CIMBL55}$ was probably the favorable allele (**Fig. 3f-h**). Further investigation is needed to determine which genetic variant(s) is causal for the higher expression level of $ZmABF4^{CIMBL5}$ (**Fig. S6d**)"

- Line 220: Data confirm that the *ZmABF4* allele contributes to drought resistance. Well I disagree. The evidence given is circumstantial. A more cautious wording would help….

Response: Thanks for the suggestions. In revising, we toned down the description and made more cautious expressions, as it wrote: "Further investigation is needed to determine which genetic variant(s) is causal for the higher expression level of $ZmABF4^{CIMBL5}$ (**Fig. S6d**)" in Line 233-234, Page 9.

- line 235: …suggesting the involvement of RNA-directed DNA CHH methylation in these regions. Again this is no hard evidence. More cautious wording.

Response: We revised the expression in a more cautious way, in Line 246-249, Page 10, as follows.

"Interestingly, 24-nt small RNA (sRNA) preferentially mapped to gene boundaries and within DNA-TE regions, resembling the pattern of mCHH, indicating a possible involvement of RNA-directed DNA CHH methylations in these regions (**Fig. S7d**)."

Additionally, in revising the manuscript, we provided the experimental data of the whole genome DNA methylation analysis of the *zmdrd1* mutants which probably providing supporting evidence for this assumption. The whole genome BS-seq analysis of two *zmdrd1* mutants revealed that significant mCHH reduction compared with wildtype (Fig. 4g-k), suggesting that RNA-directed DNA CHH methylation was possibly involved in hypermethylation of the insertional regions. Actually, DNA sequence analysis revealed the two insertional fragments upstream *ZmNAC075* consisting of a number of DNA-TEs, which were representatives of Cluster 4 insertions. These experimental data supported the assumption. More exact description of the results was described afterwards, in response to the related concerns.

- Figures 4a, 4b and line 242: The higher-level methylation is not obvious for me.

Response: As for this comment, we plotted the methylation profiles for each cluster in the appended Fig. 4c. We revised the description in Line 256-261, Page 10. We hope the results described in this revision is obvious.

"Insertions belonging to clusters 1-3 exhibited considerably high levels of mCG, mCHG and mCHH; whereas those belonging to cluster 5 displayed generally low levels of DNA methylation (**Fig. 4b**). Importantly, we found that insertions of cluster 4 had remarkably higher levels of mCG, mCHG and mCHH within the insertional fragments compared to those in the flanking sequences (**Fig. 4b-c**)."

- Lines 276: Again over interpretation and more cautious wording needed

Response: Because the other reviewers also raised the same concern, to address it, in revising we generated and analyzed two CRISPR-targeted gene knock-out lines of a maize *DRD1* gene. We appended the newly prepared Fig. 4g-i, and rewrote this paragraph, in Line 277-292, Page 11.

"To unravel the effect of the hypermethylation of two SVs on the *ZmNAC075*^{B73} expression, we analyzed two CRISPR-targeted gene knock-out (KO) lines of a maize *DRD1* gene (*zmdrd1-KO#1* and *zmdrd1-KO#2*) in a LH244 genetic background which carries the *ZmNAC075*^{B73} allele (**Fig. S9**). *Arabidopsis DRD1* (*Defective in RNA-directed DNA methylation 1*) was reported to facilitate RNA-directed *de novo* DNA methylation which might function to epigenetically regulate the expression adjacent genes [53,54]. The whole genome BS-seq analysis revealed that the mCHH levels of both S-9041 and S-1425 in the *zmdrd1* mutants were significantly diminished compared with the wildtype (LH244), while the mCG and mCHG levels were not changed (**Fig. 4g-k**). Concomitantly, the expression of *ZmNAC075*^{B73} was revealed to be significantly increased, indicating that the CHH hypermethylation of the two newly identified SVs upstream *ZmNAC075*^{B73} caused the suppression of gene expression (**Fig. 4l**). Taken together with the positive correlation of levels of the gene expression with drought resistance (**Fig. 4e, f**), these data indicated that the two insertions, upstream of *ZmNAC075*^{B73}, might be the causal variants."

- Lines 281: The link of the reticulon-like protein to drought resistance is not obvious to me. From the potential molecular function, a link is not too obvious.

Response: *ZmRtn16* encodes a reticulon-like protein colocalized with endoplasmic reticulum. The biological function of the homologous proteins was considered to be responsible for the interacting protein endomembrane trafficking and subcellular localization. Therefore, the biological significance of *ZmRtn16* may be reflected as the consequence of mis-localization of its interacting proteins. We found that ZmRtn16 interacted with two subunits of maize vacuolar-

type H^+-ATPase, ZmVHA-A and ZmVHA-E3. In the *zmrtn16* knock-out plants, the tonoplast localizations of ZmVHA-A and ZmVHA-E3 were defective. Therefore, we proposed that *ZmRtn16* function in plant drought resistance by facilitating ZmVHA-A and ZmVHA-E3 tonoplast localization that was essential for the V-H^+-ATPase activity. The reduced V-H^+-ATPase activity in *zmrtn16* supported the assumption. In revising, we further compared the proton levels in the wildtype, *ZmRtn16-OE* and *ZmRtn16-KO* plants. As a result, "In relative comparison to WT plants, the vacuolar proton levels were determined to be higher (with lower pH values) in *ZmRtn16-OE* plants, whereas they were lower in *ZmRtn16-KO* plants (**Fig. 6e, Fig. S11g**). Then, we obtained the CRISPR-targeted *zmvha-E3-KO* plants (**Fig. S12**), probably owing to the potential functional redundancy with *ZmVHA-E1*. However, we were unable to obtain the mutant of *zmvha-A*, mostly likely due to its essential functional roles *in planta*. Similar drought sensitivity phenotype was observed on the *zmvha-E3-KO* plants, providing direct evidence that defects in V-H^+-ATPase due to the absence of the ZmVHA-E3 subunit on tonoplasts compromised plant drought resistance (**Fig. 6f, g**)." We appended the description in Line 368-376, Page 14 and the newly prepared Fig. 6e-g. The experimental procedure for vacuolar pH measurements were described in Methods, Line 859-869, Page 30.

In addition, we also added the related publications on V-H^+-ATPase functions in plant stress response and resistance in Discussion, in Line 400-415, Page 15-16.

"These two types of proton pumps are assumed to work synergistically using different energy resources, ensuring that plants are able to maintain the appropriate proton gradient for vacuole transport even when cells are subject to environmental stress conditions[59]. We previously reported that genetic variation in *ZmVPP1*, which encodes a V- H^+-PPase, contributes to drought resistance in maize[14]. In the present study, we demonstrated that ZmRtn16 facilitates the tonoplast localization of the A and E3 subunits of V-H^+-ATPase, which ensured V-H^+-ATPase

activity and played a positive role in drought stress resistance. Taken together, it highlighted the role of vacuole proton-pumps in maize drought resistance. Vacuolar acidification was found to be necessary for rapid stomatal closure in response to ABA, since an *Arabidopsis* double mutant of *vha-a2 vha-a3*, lacking two subunits of V-H^+-ATPase V0 complex showed delayed stomatal closure, and similar phenotype was also observed on the V-H^+-PPase mutant (*vhp1*)[60]. Mutation in VHA-C, the C subunit of V-H^+-ATPase V1 complex, rendered severe salt sensitivity in *Arabidopsis*[61]. The Os*VHA-A* RNA-interfering rice plants showed expanded stomatal aperture and compromised salt tolerance[62]."

- Line 292: the 28bp insertion in the UTR: might affect gene expression (to be experimentally validated) but might also affect the transcript stability.

Response: As for this suggestion, we compared the mRNA stability in maize leaf protoplasts when the coding sequence of *ZmRtn16*^{B73} was respectively fused with the 3'-UTR of *ZmRtn16*^{B73} and *ZmRtn16*^{CIMBL55}. "The mRNA decay of *ZmRtn16*^{CIMBL55} was found to be slower than *ZmRtn16*^{B73} in the presence of Triptolide, an inhibitor of transcription initiation, indicating that the 3'-UTR of *ZmRtn16*^{CIMBL55} might be more favorable for the mRNA stability then that of *ZmRtn16*^{B73} (**Fig. S10a**). Interestingly, an RNA motif ($^{G}/_{U}{}^{U}/_{C}$GTGA) potentially recognized by two predicted maize RNA-binding proteins was annotated (http://cisbp-rna.ccbr.utoronto.ca) within the 28-bp insertion of the 3'-UTR of *ZmRtn16*^{B73}, which was likely the cause for the observed difference (**Fig. S10b**)." We appended the description in Line 321-328, Page 12-13, and the newly prepared Fig. S10a, b. The detailed experimental procedure was appended in Online Methods, Line 850-858, Page 29-30.

-Line 366: 4% non-syntenic genes whose expression was either undetectable or extremely low… Well, does this eventually point to false positive annotations then?

Response: We were sorry for the inaccurate expression in the last submission. The sentence should be “The remaining 4% were non-syntenic genes whose expression was generally lower than synteny genes.” Actually, in this revision, we removed this discussion, to avoid overload of the manuscript with long discussions.

- Line 369-371. Leave out.

Response: As suggested, we removed this discussion.

- Line 387: …is a homodimer of a single protein. Leave out of a single protein.

Response: Thanks for your suggestion. We deleted it.

Given the potential loci/genes detected and discussed in more detail (2/3) in the ms., the sheer amount of candidate loci reported previously and the obvious complex, polygenic foundation of the physiological characteristic, I remain skeptical about the ultimate proof of drought resistance contribution of the genes reported. Presence of SNPs or SVs in the surrounding are not a convincing proof. Traditional knock-out/overexpression or genome editing approaches to test and validate the candidates reported seems to be unavoidable to demonstrate and proof the contribution and causative molecular basis for the observed, -again complex and polygenic!- physiological phenotype. Also, reporting only 2 gene candidates and some SNPs/SVs in the surrounding doesn’t appear to be sufficient to discuss the complex, multi-genic phenotype.

Response: As for these comments, we would like to respectfully express our viewpoint. It is because of the complexity and importance of the trait, the work is of great value. It reported about a prominent drought resistant maize germplasm, the high-quality genome assembly, thorough genetic variants identification, and dissection genetic loci potentially contributing to the

trait. Moreover, in the work, we endeavored to reveal the genetic elements contribute to the trait on the epi-genomics level. It is true that verifying the candidate causal genes requires traditional knock-out/overexpression or genome editing approaches. However, it heavily relies on the candidate locus/gene information, especially that of high confidence. Based on the revealed candidate gene information, we performed gene overexpression or/and knock-out/CRISPR validation for three genes, *ZmABF4*, *ZmRtn16*, and *ZmVHA-E3* (encoding the ZmRtn16 substrate protein), and verified their functions in maize drought resistance. Other candidate loci unraveled in this work will undoubtfully provide valuable information for future studies, either causal variant validation or gene functional study for drought-resistance improvement. Therefore, we think this work covers a wide range of investigations, including germplasm characterization, state-of-art genomic and association analyses, gene functional dissection through transgenic approaches. In addition, the finding of ZmRtn16 facilitating V-H^+-ATPase activity, together with our previous report of ZmVPP1, a V-H^+-PPase, contributing to drought resistance (Wang et al., 2016), highlighting the roles of vacuolar proton-pump in maize drought resistance.

Response to Reviewer #2:

Remarks to the Author:

The authors generated a relatively high-quality assembly of CIMBL55, a drought tolerant genotype, and then used his to do comparative genomics with B73 and Mo17, which are relatively less drought tolerant to identify causative variants in CIMBL55 at candidate loci that were previously identified from GWAS. The authors provide a number of important validation experiments for a few loci that demonstrate the contribution of these loci to the drought tolerance that is observed in CIMBL55. However, a few concerns remain as described below. In particular are concerns of interpretations of the data that are insufficiently supported.

We thank the reviewer for reading and evaluating the work. We tried to improve the work according to the thoughtful comments and advices.

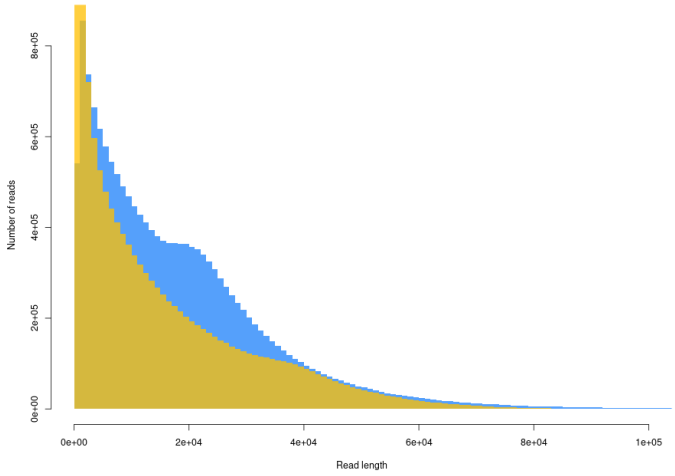
1. Page 3 Line 51-53 – This is a misleading statement supported by a review article rather than primary literature. There is not nearly enough data to know if the majority of causal variants contributing to phenotypic variation will be from SVs.

Response: Thanks for the comment. The previous description was indeed inaccurate. This review article (Liang et al., 2021) reported that ~30% QTLs, cloned in rice, were caused by the variants residing in regulatory regions rather than coding regions, which was comparable (23.9%) with the report (Wei et al., 2021). By contrast, in maize, up to May 2020, among 69 cloned QTLs and validated by transgenic approaches, 64% of them were found to be structural variations (SVs) residing in the non-coding regions (Liang et al., 2021). It is true that currently there is insufficient data to draw the conclusion. We are sorry of the inaccuracy in the previous submission, and we revised the sentence in this version. For clarity, we cited both articles and revised the description in Line 54-57, Page 3.

"In maize, the majority of (64%) the identified causal genetic variants contributing to phenotypic variation were found to be structural variations (SVs) residing in the non-coding regions, contrasting with 23.9% in rice[15,16], inferring the importance of SV identification in maize genomes."

2. Given the sequence depth the contig N50 is a bit lower than expected. Where the reads of relatively lower quality and/or length?

Response: As for this comment, we carefully compared the reads length and quality of CIMBL55 with those of B73_v5. We found that the length of CIMBL55 raw reads was not shorter, but probably they were of a lower quality than those of B73_v5 (as shown in the following graph and Table). This may lead to the contig N50 was a bit lower than expected.



The length distribution of PacBio reads of CIMBL55 (blue) and B73 (yellow).

Table PacBio CLRs quality comparison of CIMBL55 and B73_v5 using SequelTools (Hufnagel et al., 2020) and LongQC (Fukasawa et al., 2020).

	Platform	**bases**	**Reads number**	**Quality PSR/ZOR**	**Estimated non-sense read fraction**
CIMBL55	Sequel II	287G	15,752,243	0.84/0.83	0.25
B73_v5	Sequel	199G	12,770,682	0.98/0.69	0.084

PSR is the polymerase-to-subread ratio and is calculated as follows: total bases from the longest subreads per CLR divided by the total bases from subreads. When PSR is close to one the DNA template is mostly the reads of interest, whereas a total failure of library preparation would result in no reads of

interest and a PSR of zero. ZOR is the ZMW-occupancy-ratio and is calculated as the number of CLRs with subreads divided by the number of subreads. When ZOR is zero there are no DNA templates in ZMWs. When ZOR is above one, then there is more than one DNA template per ZMW on average; ideally, this value is exactly one. Third generation sequencing dataset can contain a certain amount of "non-sense reads", which we define as unique reads that cannot be mapped onto sequences of any other molecules in the same library.

In revising, we appended 60× PacBio RSII sequencing data of CIMBL55, which had been obtained before the 100× PacBio Sequel II sequencing. After combining the two datasets together, we found the assembly CIMBL55 was improved. The contig N50 was increased from 10.4 Mb to 14.5 Mb. The LAI (LTR Assembly Index) of this assemble was 28.5 and the completed BUSCOS were 98.0%. We revised Results (Line 91-93, Page 4-5) and Methods (in Line 599-603, Page 21), and we appended 60× PacBio RSII sequencing data in Data Availability: NCBI project of PRJNA765111 (Line 563-569, Page 19-20). The revisions were as follows.

"including 60× RSII and 100× Sequel II of PacBio single-molecule real-time (SMRT) sequencing, 160× BioNano Single molecule optical mapping, and 35× Hi-C (High-throughput Chromosome conformation capture) sequencing."

"For the genome assemble, the PacBio raw data of 60× RSII and 100× Sequel II were corrected and assembled by NextDenovo v2.3.1 (https://github.com/Nextomics/ NextDenovo) and set the correct_option with (minimap2_options_raw = -x ava-pb -t 30) and assemble_option with (minimap2_options_cns = -x ava-pb -t 30 -k17 -w17)."

"All data mentioned in this paper is available at NCBI under the project of PRJNA765111 including the CIMBL55 genome sequence and gene annotations, the raw data used for the assembly (PacBio, illumina, Bionano, Hi-C), as well as BS-seq data for CIMBL55, B73 and Mo17, wild type (LH244) and *zmdrd1* mutants. We deposited customized scripts in the following GitHub repository: https://github.com/ttian627/CIMBL55_genome_assembly"

3. The vast majority of the comparative analyses were done relative to B73 and Mo17. There are many other assemblies that are available in maize that could be compared to, why only these two? In particular for the variant detection that were then screened on the larger set of lines there is a big ascertainment bias that comes from only comparing CIMBL55 to B73 and Mo17 when then moving to the larger panel.

Response: Thanks for the constructive suggestions. During the accomplishment of this work, the genomes sequence of 25 NAM (Nested Association Mapping) funder lines were assembled and released (Hufford et a., 2021), providing more genetic information for maize genomics and genetics analyses. In order to improve our analysis and in response to this suggestion, we analyzed maize genetic variants at a pan-genome level, which includes 25 NAM funders and another four newly released maize genomes, SK (small kernel, a tropical and subtropical maize germplasm), K0326 (quality protein maize, a South African germplasm), A188 (excellent regeneration competence, a northwestern dent germplasm), Mo17 (belonging to a different heterotic group of B73, a Lancaster germplasm), which were all assembled at a high quality of contig N50 >5.0 Mb. Several other maize genomes, such as W22 (Springer et al., 2018) and four European flint maize lines (Haberer et al., 2020), were assembled based on Illumina short reads (contig N50 < 100 kb). This difference may hinder SV identification comparably, so that we did not include them in the current study. We conducted the analysis again and rewrote this part in Results, Line 175-198, Page 7-8, and prepared the new Fig. 3a to substitute the original Fig. 3a-b. The description was as follows.

"In order to identify the genetic elements contributing to the prominent drought resistance of CIMBL55, we explored the genetic variants on a pan-genome level, including the newly released genome information of 25 NAM founder lines[21] and another four maize accessions, SK[22], Mo17[20], A188[23], K0326Y[24] (all of them assembled with contig N50 > 5.0 Mb). To identify a maximum number of genetic variants, the assembled genome sequence 30 maize genomes (and also the corrected long-reads sequences) were aligned and compared with B73_v5[21], through four programs, SyRi[40], Smartie-sv[41], CuteSV[42] and Sniffles[43]. As a result of this analysis, 17,581,014 SVs, including 16,311,409 Insertions/Deletions (InDels), 239,611 Duplications (DUPs), 496,685 Inversions (INVs), 499,973 Translocations (TRAs), 33,336 Copy Number Variations (CNVs) were obtained non-redundantly (**Fig. 3a, Table S6**). SyRi (based on the assembled genome sequence comparisons) showed the effectiveness in discerning various types of variations, while Smartie-sv (based on the assembled contig alignments) and CuteSV (based on the long-read alignments) performed better in identifying InDels (> 50-bp) than other programs (**Fig. 3a**). To gain more insight into the genomic variations in CIMBL55, we also called the genetic variants identified from B73 and Mo17, using CIMBL55 as a reference. As a result, 841,911 DNA variants (>20-bp) were subsequently identified (**Table S7**). We genotyped all of these genetic variants in the association panel consisting of 368 maize germplasms, of which the drought tolerance was previously analyzed regarding the seedling survival rate after drought[13,14]. Thus, a total of 544,853 SVs were successfully genotyped on the population scale by Paragraph[44] based on the 20× genomic DNA re-sequencing data of all these germplasms[22], with MAF (minor allele frequency) > 0.05, and missing rate < 0.6."

The detailed data analysis steps and parameters were described in Online Methods, Line 643-653, and Line 664-679, Page 22-23. The description was as follows.

"Genetic variation identification and genotyping in an association panel

As for the DNA variants identification for the 30 maize genomes, the genome sequence of B73_v5[21] was used as a reference. SyRi[40] were used to compare the two assembled genomes as previous described[23] based on the MUMmer[72] alignment output (nucmer --maxmatch -c 500 -b 500 -l 200 -g 1000; delta-filter -m -I 90 -l 1000). Smartie-sv[41] identified variants through aligned the contig sequence of the 30 maize germplasms to B37_v5 based on blasr with default parameters. CuteSV[42] and Sniffles[43] were used for the prediction of variants after alignment of the long reads of the 30 maize germplasms to B37_v5 through its supporting toolkit ngmlr, respectively. The variants were merged by BEDtools[73], parameter settings were (intersect -r -f 0.8 -v) for insertions and (window -l 100 -r 100 -v) for deletions."

"The identified SVs ≥ 50-bp obtained from the 30 maize genomes and DNA variants ≥ 20-bp were genotyped by Paragraph[44] based on 368 DNA deep re-sequencing data[22]. Finally, 544,853 variants were genotyped with a minor allele frequency (MAF) > 0.05 and a missing rate (MR) < 0.6. As for bi-directional SV identification, at first the DNA variants were identified through four programs, MUMmer Smartie-sv, Pbsv, and Sniffles. The two output files of a program (e.g., MUMmer) were obtained, based on the bi-directional analysis of either CIMBL55 as reference and B73/Mo17 as query, or CIMBL55 as query and B73/Mo17 as reference. Then, they were processed by a customized perl script (https://github.com/ttian627/CIMBL55_genome_assembly) to filter out the SNPs present only in one file. Secondly, the identified bi-directional SNPs were used as anchors. If an insertion was identified in CIMBL55 referred to B73, a deletion should be called in B73 in a comparison with CIMBL55, with the difference of their distances to the anchor SNP being less than 500 bp. Finally, 5,346 bi-directionally confirmed insertional fragments (≥20-bp) identified in CIMBL55, in comparison with B73 (also Mo17) were used for SV-related differential DNA methylation analysis."

4. There is a lot of time spent on documenting findings that have already been highly studied in previous genomic studies. For example, much of the results of the syntenic vs. non-syntenic genes has been shown in Schnable et al 2011, Brohammer et al. 2018, Hufford et al. 2021 etc. Likewise for the general genome-wide methylation patterns and patterns of methylation over TE sequences, which have been well documented in many different by the Springer group and others.

Response: As for this concern, we would like to respectfully express the point view. Due to the complexity of the maize genome, it is important and interesting to study the gene synteny between genomes. Maize is an ancient allotetraploid, consisting of two sub-genomes which came from tetraploidization of an ancient progenitor followed by diploidization and biased gene fractionation. In addition, >85% genomic contents consist of transposon elements. The current sequence comparison programs to identify SVs are still insufficient to effectively identify genomic rearrangement on a chromosome-level between two maize genomes. That's the reason why many research groups were interested in exploring the gene synteny and gene composition between maize genomes, whenever a maize genome was assembled. As the reviewer mentioned, Schnable et al. (2011) reconstructed two maize sub-genomes of B73, and proposed that fractionation events continued to occur in 33 maize accessions. Brohammer et al. (2018) compared B73 and PH207 and found less than 1% synteny genes have different fractionation pattern. Hufford et al. (2021) analyzed the core genes of 25 NAM founder genomes and identified 1,281 ancient fractionation pairs and 494 recent fractionation pairs, but their functions were unknown. As suggested by Reviewer #1, we performed sub-genome and gene fractionation analyses of the CIMBL55 genome, and checked the gene synteny of several gene families involved in ABA-signaling and stress-response. The new prepared Fig. S4a-c, Fig. S5a-b was appended, and the descriptions were added in the Results, Line 155-163 and 169-173, Page 7, as follows.

"We construct two sub-genomes of CIMBL55, referring to the sorghum genome. Sorghum diverged from the maize lineage just prior to the tetraploid event, and has been commonly used as a reference for maize sub-genome analysis and fractionation bias estimation[20,33]. The CIMBL55 sub-genome organization and fractionation bias was found to be consistent with B73, supporting the evolution conservation within *Zea mays* (**Fig. S4a, b)**. Then, it was found that 61.5% Class II genes were pair-retained genes in the two sub-genomes of CIMBL55, indicating that Class II genes remained relative conserved during maize evolution (**Fig. S4c**)."

"Among them, 72% belonged to Class II genes, which was remarkably higher than 29% Class II genes on the whole genome level (**Fig. S5a-b, Fig. 2a**). These data indicate that genes involved in environmental response and adaptation tended to be retained in duplication during maize evolution."

As for the differential DNA methylation analysis, in revising, we removed the DNA methylation analysis of the conserved region between CIMBL55 and B73/Mo17, to avoid overload of the manuscript. However, we focused on the structure variants (SVs) related differential DNA methylation, to characterize the epigenetic changes related to genetic variants and their possible effects on phenotypic alterations, which was not intensively investigated in previous researches. We developed a bi-lateral insertion identification program to systematically identified the SVs in CIMBL55 in comparison with B73, with high confidence. Then, "5,346 insertional fragments identified in CIMBL55, which could be bi-directionally confirmed (see Methods), were used for SV-related DNA methylation analysis." For clarity, we appended this sentence in Line 251-253, Page 10.

With the genome-wide BS-seq data of CIMBL55 and B73, we unraveled five methylation patterns of insertional fragments in CIMBL55. Importantly, we found that insertions in cluster 4 had a remarkably high level of mCHH than the flanking sequences (**Fig. 4b-c**). DNA-TEs, especially DTH (DNA transposon Terminal inverted repeat Harbinger), were enriched in this cluster of insertions (**Fig. 4b**). The insertion length was concentrated in the range of 100 to 1,000 bp, and were relatively close to the transcription start site (TSS) of the neighboring gene, indicating that DNA-TEs, especially DTH, proximal to genes tended to be CHH methylated (**Fig. 4b, Fig. S8a-b**). CHH methylation is asymmetric CG methylation and regarded as *de novo* methylation, which doesn't conservatively occur during DNA replication but is largely mediated by RdDM pathway (in the euchromatin regions). Thus, mCHH level might be more flexible and variable than mCG and mCHG among different genome, or in response to different conditions. Based on the SV-related differential DNA methylation analysis, we provided a case study of *ZmNAC075*. The hypermethylation on CHH were found in two insertions upstream the $ZmNAC075^{B73}$ allele. In revising, "To unravel the effect of the hypermethylation of two SVs on the $ZmNAC075^{B73}$ expression, we analyzed two CRISPR-targeted gene knock-out (KO) lines of a maize *DRD1* gene (*zmdrd1-KO#1* and *zmdrd1-KO#2*) in a LH244 genetic background which carries the $ZmNAC075^{B73}$ allele (**Fig. S9**). *Arabidopsis DRD1* (*Defective in RNA-directed DNA methylation 1*) was reported to facilitate RNA-directed *de novo* DNA methylation which might function to epigenetically regulate the expression adjacent genes[53,54]. The whole genome BS-seq analysis revealed that the mCHH levels of both S-9041 and S-1425 in the *zmdrd1* mutants were significantly diminished compared with the wildtype (LH244), while the mCG and mCHG levels were not changed (**Fig. 4g-k**). Concomitantly, the expression of $ZmNAC075^{B73}$ was revealed to be significantly increased, indicating that the CHH hypermethylation of the two newly identified SVs upstream $ZmNAC075^{B73}$ caused the suppression of gene expression (**Fig. 4l**). Taken together with the positive correlation of levels of the gene expression with drought resistance (**Fig. 4e, f**), these data indicated that the two insertions, upstream of $ZmNAC075^{B73}$, might be the causal variants." We appended this description in Line 277-292, Page 11, and newly prepared Fig. 4g-l.

Therefore, the gene synteny and SV related differential DNA methylation analyses in this work were of different purpose or strategy, compared with previous researches. The results provided important information in dissection of the genetic basis of the prominent drought resistance of CIMBL55.

5. Page 8 line 194-198 – Please make it clearer if this is looking at survival rates in just CIMBL55 and B73 or in all individuals in the population that have either of these alleles. If the latter, how many individuals in the population does this account for, in other words are there other non-CIMBL55 or B73 alleles in the population that are not considered?

Response: Thanks for the suggestion. The survival rates were calculated as the average of the germplasms carrying either CIMBL55 or non-CIMBL55 allele among the association panel consisting 368 accessions. The detailed information was list in Table S8. In this version, we changed the column graphs into violin plots, in the related statistical analyses (Fig. 3d-e; Fig. 4e-f; Fig. 5e-f), which might have a better readability. The sample size was indicated in the figures.

In the previous submission, because only the genetic variants between CIMBL55 and B73 were genotyped on the population level, there was either the CIMBL55 or B73 allele on each locus. In revising, as suggested, we identified the genetic variants through analyzing 30 high-quality maize genomes referring to B73, and genotyped all the non-redundant variants in an association population (containing 368 maize germplasms). Therefore, in this revision, there were the non-CIMBL55 alleles (identified in the germplasms other than CIMBL55) identified as drought tolerant alleles (the germplasm carrying the allele averagely had a significant higher drought survival rate than those harboring the counterpart allele). We re-analyzed the data, updated Fig. 3c, and revised the description in Line 210-218, Page 9, as follows.

"Resultantly, significant SVs associated with drought-tolerance were identified for 69 genes, which were in a strong linkage disequilibrium (LD, $r^2 > 0.5$) with the previous lead SNPs of 79 genes (**Fig. 3c)**. Notably, the superior haplotype (both the SNP and SV were the drought-resistant alleles) of 65 genes were found in CIMBL55, explaining the predominant drought-resistance of CIMBL55 among the population (**Table S8**). Moreover, among 108 genes, the superior haplotype of 11 candidate genes were identified in germplasms other than CIMBL55 (non-CIMBL55 allele), suggesting that other germplasms may contain complementary genetic resources to CIMBL55, concerning maize drought resistance (**Fig. 3b, Table S8**)."

6. Page 9 line 203-206 – The authors need to be careful that their hypothesis is not driving their conclusions. Just because there is an SV in LD with a previously identified SNP doesn't mean that the SV is necessarily causative.

Response: We agreed with the comment that an SV in LD with a previously identified SNP didn't mean that the SV was necessarily causative. However, the newly identified SVs (significantly associated with the trait) provided valuable information for the follow-up causal variation validation and allelic mining. In addition, if the significant SNP(s) could be supported by identified SV(s), it would strengthen the reliability of the locus as a candidate. Due to the out-pollination and highly-polymorphic nature of maize, a candidate gene supported by several significant variants was usually considered more reliable than the candidate gene pinpointed by a single significant variant. Thus, it was also the significance of systematical identification of SVs for the previously proposed candidate genes. For clarity, as for the significant variants of *ZmABF4*^{CIMBL55}, without experimental validation, we wrote the result in a more cautious sentence, in Line 233-234, Page 9, as follows.

"Further investigation is needed to determine which genetic variant(s) is causal for the higher expression level of *ZmABF4*^{CIMBL5}."

We also appended the description in Discussion, Line 384-391, Page 15.

"The genetic variability among 30 maize genomes was carefully explored, which were currently available and assembled with high-quality (contig N50 > 5 Mb). In total, 544,853 SVs were identified on a pan-genome level and genotyped on a maize association panel, which facilitated the SV identification linked with drought-resistant candidate genes and the analysis of SV-related DNA methylation. The newly identified SVs that were significantly associated with drought resistance may not be the actual causal variants, however, they can serve as potential targets for further validation and allelic mining."

7. Page 10 line 259 – Page 11 line 263 – This interpretation is a stretch and not supported by much of anything.

Response: According to this comment, we removed the differential DNA methylation analysis of the conserved region between CIMBL55 and B73/Mo17, to avoid overload of the manuscript with interpretations without hard evidence. In revision, we focused on the SV-related differential DNA methylation analysis, which was not intensively investigated in previous works.

8. I found it hard at times to know what was new in this manuscript and what was from previous studies as there are entire supplemental figures made from previous data that support previous findings.

Response: As for this concern, we think there were plenty of novel information and findings in this work, especially after taking the constructive suggestions and thoughtful comments from three reviewers.

In this revision, genetic variants were identified on a pan-genome level, including the most recently released 30 high-quality genome assembles, besides B73_v5. The novel and significant SVs of 69 genes were identified, from 108 previously proposed drought-resistant candidate genes. The SV-associated DNA methylation patterns were explored and classified. All these results were significant advances on a basis of previous research. It was also undoubtedly novel findings that DNA hypermethylation of two insertional fragments upstream of *ZmNAC075* suppressed the gene expression, and that *ZmABF4* and *ZmRtn16* positively functions in maize drought resistance. We hope the reviewer can find that the quality of the revised manuscript was substantially improved, and the description was more concise and clearer.

As for the novelty of supplemental figures, although the analytic methods and presentation formats were ordinary for some of them (Fig. S2, S3, S4, S7), they contained novel information based on the new genome assembly of CIMBL55. As for Fig. S1a-c, they demonstrated the pair-comparison of the drought survival rate of CIMBL55, B73 and Mo17, which had not been closely compared in previous work. Fig. S1d demonstrated the yield performance of these genotypes under field drought conditions, which had not been studied before. They were the supportive information for main Fig. 1a-d. These results were important, because they demonstrated CIMBL55 was prominent drought resistant throughout all developmental stages and gave grain yield advantages. Similarly, the other supplemental figures all contained novel information or findings either to support the results and figure in main text or completed information of the related analysis. Thus, we think they were necessary.

9. Page 11 line 275-278 – not enough data to support this claim.

Response: This concern was also raised by the other reviewers. Therefore, we constructed new experiments, appended the newly prepared Fig. 4g-l, and Fig. S9, and added the description in Results, Line 277-292, Page 11, as follows.

"To unravel the effect of the hypermethylation of two SVs on the *ZmNAC075*^{B73} expression, we analyzed two CRISPR-targeted gene knock-out (KO) lines of a maize *DRD1* gene (*zmdrd1-KO#1* and *zmdrd1-KO#2*) in a LH244 genetic background which carries the *ZmNAC075*^{B73} allele (**Fig. S9**). *Arabidopsis DRD1* (*Defective in RNA-directed DNA methylation 1*) was reported to facilitate RNA-directed *de novo* DNA methylation which might function to epigenetically regulate the expression adjacent genes[53,54]. The whole genome BS-seq analysis revealed that the mCHH levels of both S-9041 and S-1425 in the *zmdrd1* mutants were significantly diminished compared with the wildtype (LH244), while the mCG and mCHG levels were not changed (**Fig. 4g-k**). Concomitantly, the expression of *ZmNAC075*^{B73} was revealed to be significantly increased, indicating that the CHH hypermethylation of the two newly identified SVs upstream *ZmNAC075*^{B73} caused the suppression of gene expression (**Fig. 4l**). Taken together with the positive correlation of levels of the gene expression with drought resistance (**Fig. 4e, f**), these data indicated that the two insertions, upstream of *ZmNAC075*^{B73}, might be the causal variants."

10. The authors themselves state on line 406-407 – "Drought resistance in plants is a complex trait that can vary based on the timing and magnitude of the stress", yet the timing and magnitude of stress is different for different experiments in this study (e.g., the stress applied in the rainout shelters).

Response: It is true that drought resistance is a complex trait that can based on timing and magnitude of the stress, but it doesn't mean that a drought-resistant gene function must be

limited to a specific condition or growing stage. Some central molecular and physiological mechanisms are conserved in drought response and resistance. The genes involved in the central mechanisms might be effective in different timing and stress magnitude. Usually, to make research work feasible, usually studies were performed at the seedling stage and under a specific stress condition. However, if the gene was found to play a role in the stress resistance, its' function then could be intensively investigated under different developmental stages and stress conditions. That was the reason why we extended the experiment to test the gene function in field drought conditions (under rain-out shelters) to all growing stages, in addition to characterizing its' roles in the lab condition at the seedling stage. The approval of the gene function to mitigate the yield loss, required much efforts and labors. But it was of great importance. Currently, most of researches usually can focus on only one segment of a complex phenomenon. Thus, at the end of Discussion, Line 421-427, Page 427, we proposed that "Future high-throughput phenomic and metabolomic technologies will facilitate data-obtaining from different stages and tissues, which will further enhance the genetic dissection of this trait. The high- quality assembly of CIMBL55, a prominent drought resistant germplasm, undoubtedly provides a foundational resource for both gene function discovery and the genetic improvement of drought resistance."

Response to Reviewer #3:

Remarks to the Author:

The study by Tian, Wang et al produces a genome assembly for the CIMBL55 maize genotype and uses it to further explore its unique drought resistance phenotype. Numerous candidate alleles are identified that are associated with the drought resistance in CIMBL55. Two alleles are pursued in an attempt to identify causal genetic variants. One allele is associated with ZmNac075 and the other is associated with ZmRtn16. Although there are strong associations of genetic

variants around the ZmNac075 locus with drought resistance, there is no evidence provided in this study to demonstrate which are potentially causal. A much more convincing set of experiments were performed to show causality of AmRtn16, which includes a reporter assay, overexpression and CRISPR knockouts.

The authors thank this reviewer for the critical reading and evaluating the manuscript.

Major comments:

1. There is a wide range of survivability data in comparisons between figure 3e, f, k and figure 5h and i. The wildtype survival is so variable. It ranges from 10-20% in figure 3 e and f yet the OE in 3k shoots up to 50-60%. Similarly, in figure 5 the WT survival is 50% and the KOs are 10-20% yet in 5h WT is 10% and OE lines at 30-40%. This level of variability is concerning especially as the WT is always higher or lower as needed to observe the strong impact of the mutant/transgenic alleles being tested.

Response: As for this comment and concern, we would respectfully explain the experimental procedure for drought resistance assay in which the seedling survival was taken as an index. So far, it is the most rigorous and widely-adopted method to compare two genotypes (e.g., wildtype vs. transgenics) in the same soil-containing box side-by-side, in which the plants were treated by water withholding for the same time period. When the difference of leaf-wilt was most apparent between two genotypes or the plants were stressed near to death, water was resumed to recover the plants. Then, usually after 3-5 days recovery, the seedling survival rate was counted and compared. There was no absolutely fixed time point for rewatering. It was the time point to resume watering, when the difference of two genotype was most obvious. Therefore, we always focused on the difference between two genotypes in the same experiment, but we could not directly compare the survival rate across different tests. Thus, if the wildtype was compared with

a drought-resistant genotype plants (e.g., WT vs. *ZmABF4-OE* or WT vs. *ZmRtn16-OE*), usually we stressed the plants stronger (by water-withholding longer) to display the difference. So that, the survival rate of wildtype was usually around 10-20% (Fig. 3f-g, Fig. 5h). Whereas, when the wildtype was compared with drought sensitive plants (e.g., WT vs. *zmrtn16-KO* or WT vs. *zmvha-E3-KO*), if the stress treatment reducing wildtype survival rate to 10-20%, the survival rate of drought-sensitive genotype plants might be reduced to 0-10%. The difference would not be marginal and ambiguous. Therefore, in this case, we usually applied a relatively weaker stress by shortening the period of water-withholding (usually three days shorter, depending on the tolerance of the tested plants). So that, the survival rate of wildtype was about 40-60% (Fig. 5i, Fig. 6g). Based on our experience in evaluating seedling drought resistance for years, the results obtained from this assay was the most reliable and reproducible compared with other methods (such as drought-stressed in two soil-containing boxes separately, field-drought treatment for seedlings, or PEG-solution-mimicking treatment). Because in this assay two genotype plants were compared in one soil-containing-box under nearly-identical environmental conditions, including fluctuations of light and temperature, soil texture, and water potential etc. Unless the genetic difference was evident, the survival rates of two genotypes would not be consistently and significantly different in repeated assays. For clarity, we detailed the description regarding the drought-resistance assay in Methods, Line 811-820, Page 38, as follows. We hope this description can address the concerns.

"As for the drought resistance assay in lab conditions, different genotype plants were planted side by side in a cultivation box (35 × 20 × 10 cm, length × width × depth) that was filled with 1.5 kg of enriched soil (turf to vermiculite in a ratio of 1:1). A drought treatment was applied to the soil-grown plants at the three-leaf seedling stage by withholding water. Approximately 17 days water-withholding was applied to the plants for the comparison between the wildtype and *ZmABF4-OE* (or *ZmRtn16-OE*) plants, and 14 days water-withholding was applied to the plants for comparison of the wildtype and *zmrtn16-KO* (or *zmvha-E3-KO*) plants. Water was resumed

to allow plants to recover, when the difference of leaf-wilt became most obvious. The number of surviving plants was recorded around 3-5 days later."

2. There is no evidence that DNA methylation data presented at ZmNac075 is causal for repression of gene expression. I have no doubt that two new SVs were identified and that both are DNA methylated, but no data is presented to show causality of drought resistance of any genetic variant or associated DNA methylation around this locus. This highlights the cause-and-effect issue surrounding the study of the ZmNac075 allele. For example, Figure 4 e and f could be repeated with all SNPs (or any other genetic variant in the area) and it would reveal the same result due to linkage disequilibrium. There are lots of possible genetic variations in this 20kb region and none of them have been tested in isolation from one another. This prevents assignment of causality. At this stage, there is no evidence the insertions that are methylated are causal. It is very common that insertions are methylated, but very few examples exist to date demonstrating they are causal for expression variation of nearby genes.

Response: Thanks for the thoughtful comments. Because the other reviewers also raised the same concern, to address it, in revising we generated and analyzed two CRISPR-targeted gene knock-out lines of a maize *DRD1* gene. We appended the newly prepared Fig. 4g-I and Fig. S9 and the description of the obtained results, in Line 277-292, Page 11, as follows.

"To unravel the effect of the hypermethylation of two SVs on the *ZmNAC075*^{B73} expression, we analyzed two CRISPR-targeted gene knock-out (KO) lines of a maize *DRD1* gene (*zmdrd1-KO#1* and *zmdrd1-KO#2*) in a LH244 genetic background which carries the *ZmNAC075*^{B73} allele (**Fig. S9**). *Arabidopsis DRD1* (*Defective in RNA-directed DNA methylation 1*) was reported to facilitate RNA-directed *de novo* DNA methylation which might function to epigenetically regulate the expression adjacent genes[53,54]. The whole genome BS-seq analysis revealed that the mCHH levels of both S-9041 and S-1425 in the *zmdrd1* mutants were significantly diminished compared with the wildtype (LH244), while the mCG and mCHG levels were not changed (**Fig.**

4g-k). Concomitantly, the expression of *ZmNAC075*^{B73} was revealed to be significantly increased, indicating that the CHH hypermethylation of the two newly identified SVs upstream *ZmNAC075*^{B73} caused the suppression of gene expression (**Fig. 4l**). Taken together with the positive correlation of levels of the gene expression with drought resistance (**Fig. 4e, f**), these data indicated that the two insertions, upstream of *ZmNAC075*^{B73}, might be the causal variants."

However, it was regretful that we failed to obtain the gene knock-out mutant of *ZmNAC075*. As for constructing the gene knock-down mutants, currently the genetic transformable maize genotypes carry *ZmNAC075*^{B73} haplotype which was already of a low level of gene expression. The gene knock-down mutant would not be expected to give a clear compromised drought resistance compared with the wildtype plants. We also tried to generate the gene overexpression plants, but only one transgenic line with dwarfism was obtained. Thus, we think the expression level of *ZmNAC075* might be strictly modulated to achieve its proper function in plants. Anyway, mCHH hypermethylation of the two insertions suppressed the gene expression which was positively related to the drought resistance, suggesting that they are potentially the causal variants for the trait. Previously, several publications reported that the type of NAC transcription factors positively functioned in plant drought resistant (Hu et al., 2006; Mao et al., 2015; Tran et al., 2004). Taken together, it indicated that the two insertions, upstream of *ZmNAC075*^{B73}, might be the causal variants.

3. The following line in the abstract must be removed as it is inaccurate. "Differential DNA methylation analysis indicated that hypermethylation of the ZmNAC075 promoter repressed gene expression and compromised drought resistance."

Response: Through generating and analyzing the *zmdrd1-KO* plants, we think that we provided the evidence that mCHH hypermethylation of two insertions upstream of *ZmNAC075* suppressed

the gene expression, which might compromise plant drought resistance. We revised the sentence more objectively and cautiously in Abstract, as follows:

"Structure-variation related differential DNA methylation analysis indicated that repressed gene expression, which resulted from hypermethylation of a drought-sensitive allele of *ZmNAC075* upstream sequence, may compromise drought resistance."

Minor comments:

1. Change all forms of "bisulfate" to "bisulfite"

Response: Thank you for the reminder. We have corrected such errors throughout the manuscript.

2. Table S8 is missing bisulfite conversion rates.

Response: We mapped the BS-seq reads to the non-methylated chloroplast genome to assess bisulfite conversion efficiency. More than 99% of cytosine bases were successfully bisulfite converted. We add "Bisulfite conversion rate" column in Table S9, in the revision.

3. The analysis described in 223-263 is worthy of pursuit, but these questions/results have been explored before numerous times between maize genotypes and in other plant species.

Response: Thanks for the comments. Epigenetic modifications implicate gene expression regulations and trait diversities. It is true that many research efforts have been spent on the profiling and comparison of DNA or histone methylation pattern among different genotypes in different plant species. However, in this work, we intensively explored the structure variation (SV) related differential DNA methylation, which was not systematic investigated in between of

other maize genomes. Therefore, we think this information was novel and important. As suggested by Reviewer #2, to avoid overload of the manuscript with descriptive information, we removed DNA methylation analysis on the conserved regions. To make it clearer, we revised the description of this part in Results, in Line 250-266, Page 10-11, as follows.

"We identified differential DNA methylations in SV-related genomic regions based on the assembled CIMBL55 and B73 genome sequences. A total of 5,346 insertional fragments identified in CIMBL55, which could be bi-directionally confirmed (see Methods), were used for SV-related DNA methylation analysis. The DNA methylation patterns of the insertional sequence (including its 0.5-kb-flanking regions) were clustered into 5 groups as determined by hierarchical clustering based on the levels of mCG, mCHG, and mCHH (**Fig. 4a-c, Table S10**). Insertions belonging to clusters 1-3 exhibited considerably high levels of mCG, mCHG and mCHH; whereas those belonging to cluster 5 displayed generally low levels of DNA methylation (**Fig. 4b**). Importantly, we found that insertions of cluster 4 had remarkably higher levels of mCG, mCHG and mCHH within the insertional fragments compared to those in the flanking sequences (**Fig. 4b-c**). DNA-TEs, especially DTH (DNA transposon Terminal inverted repeat Harbinger), were enriched in this cluster of insertions (**Fig. 4b**). The insertion length was concentrated in the range of 100 to 1,000 bp, and were relatively close to the transcription start site (TSS) of the neighboring gene, indicating that DNA-TEs, especially DTH, proximal to genes tended to be CHH methylated (**Fig. 4b, Fig. S8a-b**)."

In Data Availability the information was added, Line 567-569, Page 19-20.

"We deposited customized scripts in the following GitHub repository: https://github.com/ttian627/CIMBL55_genome_assembly."

4. Manuscript requires proofreading for typos.

Response: Thank you for your reminder. We carefully proofread the revised manuscript. The revised manuscript was also sent to Dr. Dale Kalson of Journal Doctors Inc. (http://www.journaldoctors.com/about.html) for English editing and polishing.

Additional References in this Response Letter

Fukasawa, Y., Ermini, L., Wang, H., Carty, K., and Cheung, M.S. (2020). LongQC: A Quality Control Tool for Third Generation Sequencing Long Read Data. G3 (Bethesda) *10*, 1193-1196.

Hufnagel, D.E., Hufford, M.B., and Seetharam, A.S. (2020). SequelTools: a suite of tools for working with PacBio Sequel raw sequence data. BMC bioinformatics *21*, 429.

Springer, N.M., Anderson, S.N., Andorf, C.M., Ahern, K.R., Bai, F., Barad, O., Barbazuk, W.B., Bass, H.W., Baruch, K., Ben-Zvi, G., *et al.* (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. Nature genetics *50*, 1282-1288.

Decision Letter, Appeal:

30th Aug 2022

Dear Dr. Qin,

Thank you for your message of asking us to reconsider our decision on your manuscript "Genome Assembly and Genetic Dissection of a Prominent Drought-Resistant Maize Germplasm". I have now discussed the points of your letter with my colleagues, and we think that you have some valid points. We therefore invite you to upload necessary documents so that we can send your manuscript back to the original reviewers for peer review.

Please be sure that your manuscript is accompanied by a separate letter detailing the changes you have made and your response to the points raised. At this stage we will need you to upload:

1) a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument.

2) a copy of the manuscript in MS Word .docx format. Please ensure that your manuscript conforms to our Article format instructions, available
here.

3) a revised version of Reporting Summary:

https://www.nature.com/documents/nr-reporting-summary.pdf
(Here you can read about the role of the Reporting Summary in reproducible science: https://www.nature.com/news/announcement-towards-greater-reproducibility-for-life-sciences-research-in-nature-1.22062)

4) a revised version of Editorial Policy Checklist:
https://www.nature.com/documents/nr-editorial-policy-checklist.pdf

Please use the link below to be taken directly to the site and view and upload the files:

[redacted]

With kind wishes,

Wei Li, PhD
Senior Editor
Nature Genetics
New York, NY 10004, USA
www.nature.com/ng

Decision Letter, second revision:

4th Nov 2022

Dear Dr. Qin,

Thank you for submitting your revised manuscript "Genome Assembly and Genetic Dissection of a Prominent Drought-Resistant Maize Germplasm" (NG-A58901R1). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Genetics, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements soon. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics Please do not hesitate to contact me if you have any questions.

Sincerely,
Wei

Wei Li, PhD

Senior Editor
Nature Genetics
New York, NY 10004, USA
www.nature.com/ng

Reviewer #1 (Remarks to the Author):

Thanks for the very thorough revision and for the very thoughtful adressing of my points and the revision of the manuscript. All of the points I raised have been sufficiently adressed and the ms. has further ripened. An inspiring piece of work.

Reviewer #2 (Remarks to the Author):

I thank the authors for their careful consideration of the concerns raised by all three reviewers. The revised manuscript addressed my concerns raised with the previous version.

Reviewer #3 (Remarks to the Author):

Although this work is certainly publishable, in my opinion it does meet the impact criteria for publication in Nature Genetics. The assembly is high-quality and will be useful to the community, but there are over 30 high quality genome assemblies available today. The list of candidate drought related genes is a great starting point to explore their roles in this important plant trait. Unfortunately, no causality has been demonstrated at this time and even if causality has been demonstrated it's not clear why this result would be the prestige of Nature Genetics.

My review primarily focused on the DNA methylation results. The addition of the drd crispr lines does not directly address if CHH at the two SVs upstream of the NAC protein are causal. Removal of genome-wide CHH not only removes CHH of these SVs, but ALL CHH in the genome. It is equally as likely that a trans factor (another gene) has been affected by the removal of CHH that affects NAC expression.

I appreciate the authors toning down some of their conclusions and would encourage them to further remove any assignment of causality to DNA methylation regulation of the NAC at this time.

Author Response, second revision:

Point-to-point Response Letter to Reviewers' comments

Reviewer #1 (Remarks to the Author):

Thanks for the very thorough revision and for the very thoughtful addressing of my points and the revision of the manuscript. All of the points I raised have been sufficiently addressed and the ms. has further ripened. An inspiring piece of work.

Reviewer #2 (Remarks to the Author):

I thank the authors for their careful consideration of the concerns raised by all three reviewers. The revised manuscript addressed my concerns raised with the previous version.

Response: Thanks for the positive comments on our revised manuscript and all the inspiring words from both reviewers.

Reviewer #3 (Remarks to the Author):

Although this work is certainly publishable, in my opinion it does meet the impact criteria for publication in Nature Genetics. The assembly is high-quality and will be useful to the community, but there are over 30 high quality genome assemblies available today. The list of candidate drought related genes is a great starting point to explore their roles in this important plant trait. Unfortunately, no causality has been demonstrated at this time and even if causality has been demonstrated it's not clear why this result would be the prestige of Nature Genetics.

My review primarily focused on the DNA methylation results. The addition of the drd crispr lines does not directly address if CHH at the two SVs upstream of the NAC protein are causal. Removal of genome-wide CHH not only removes CHH of these SVs, but ALL CHH in the genome. It is equally as likely that a trans factor (another gene) has been affected by the removal of CHH that affects NAC expression.

I appreciate the authors toning down some of their conclusions and would encourage them to further remove any assignment of causality to DNA methylation regulation of the *NAC* at this time.

Response: We completely agree with this review's opinion. Thus, we toned down the conclusion of this part in Page 10, Line 262. It reads:

"Concomitantly, the expression of *ZmNAC075*^{B73} was revealed to be significantly increased, indicating that the CHH hypermethylation of the two newly identified SVs upstream *ZmNAC075*^{B73} may cause the suppression of gene expression (**Fig. 4l**)."

We also removed the conclusion sentence of assignment the causality to two SVs in the *ZmNAC075* promoter in the revision. The original sentence was as follows:

"Taken together with the positive correlation of levels of the gene expression with drought resistance (**Fig. 4e, f**), these data indicated that the two insertions, upstream of *ZmNAC075*^{B73}, might be the causal variants."

Final Decision Letter:

11th Jan 2023

Dear Dr. Qin,

I am delighted to say that your manuscript "Genome assembly and genetic dissection of a prominent drought-resistant maize germplasm" has been accepted for publication in an upcoming issue of Nature Genetics.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Genetics style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

You will not receive your proofs until the publishing agreement has been received through our system.

Due to the importance of these deadlines, we ask that you please let us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

Your paper will be published online after we receive your corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the Nature Press Office (press@nature.com) after sending your e-proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NG-A58901R2) and the name of the journal, which they will need when they contact our Press Office.

Before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Genetics. Our Press Office may contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Please note that Nature Genetics is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <a href="https://www.nature.com/nature-portfolio/editorial-policies/self-archiving-and-license-to-publish. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Please note that Nature Portfolio offers an immediate open access option only for papers that were first submitted after 1 January, 2021.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

If you have not already done so, we invite you to upload the step-by-step protocols used in this manuscript to the Protocols Exchange, part of our on-line web resource, natureprotocols.com. If you complete the upload by the time you receive your manuscript proofs, we can insert links in your article that lead directly to the protocol details. Your protocol will be made freely available upon publication of your paper. By participating in natureprotocols.com, you are enabling researchers to more readily reproduce or adapt the methodology you use. Natureprotocols.com is fully searchable, providing your protocols and paper with increased utility and visibility. Please submit your protocol to https://protocolexchange.researchsquare.com/. After entering your nature.com username and password you will need to enter your manuscript number (NG-A58901R2). Further information can be found at https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#protocols

Sincerely,
Wei

Wei Li, PhD
Senior Editor
Nature Genetics
New York, NY 10004, USA
www.nature.com/ng