Review

August 4, 2015

Xiao et al., constructed a multi-parent mapping population using 14 diverse maize lines. This population (or 10 RIL sub-populations) could be useful for future maize genetic studies. However, I am not convinced by their findings and interpretations, especially the low-frequncy argument. Theoretically, the minumium allele frequncy in their population is 1/14 (7%). Because of the unbalanced cross design (BY851 and K22 were used for multiple crosses), this number would be varied a little bit. However, in general, we are not talking about severe cases, i.e. alleles with rare allele frequency < 1%, which normally believed as the recently derived alleles (mostly deleterious) and hard to be detected by conventional approach. Given the worst scenario of rare allele frequency of <7%, it is strange to me that their joint QTL mapping (JLM) and GWAS could not detected QTLs, at least, as good as separate QTL mapping (SLM) method. SLM used bi-parental population of RILs, ideally, 50% of the alleles derived from one of the parents; there is no rare alleles in these sub-populations at all, therefore, it is unconvincing to claim that their population or method could “boost power for identifying minor-effect and low-frequency variants” (line 354). In addition, I am a little bit suspecious about the power simulation. Under scenario iv) (Figure 4, “model” used here is confusing), QTLs with opposite effects would be hard to be detected by JLM and GWAS, because their effects would be cancelled out. Why under this scenario, JLM has the highest power? For the simulation, they only talk about power issue, how about false discovery. If the boosted power was due to the increased false discovery, it seems not very useful after all.

# Minor points (line by line)

* Line 50-52: It would be great to cite more recent literature(s) or data. How about other factors for yield improvement, i.e., the contributions of nitrogen usage and plant density?
* Line 82-84: For each locus, ideally, 50% of the alleles come from B73 and 50% of the alleles come from non-B73 according to NAM design, isn't it? I agree that non-B73 alleles were derived from 25 founder lines. If that is the point, the authors should explicitly explain what is the "statistical issues" for NAM design.
* Line 127: "Three models" or three different approaches? The use of "Model" can be confused with genetic models (i.e. additive or dominant models) and/or statistical models.
* Line 129: genetic basis?
* Line 139: Did you consider the direction of effects when integrating two or more co-localized QTLs?
* Line 432: remove "each"?
* Line 627: This is not fine mapping.
* Line 652-654: "binomial distribution" is an assumpution of the data distribution, not a statistical method.
* Line 670: Broad-sense heritability should be .