Review of **HMMploidy: inference of ploidy levels from short-read sequencing data**

This paper is an interesting application of a Hidden Markov Model to both inferring ploidy level and detecting changes in ploidy level. The authors make a convincing case for why this is an interesting problem with potential applications in both agriculture and medicine. The new method appears to be more accurate at inferring ploidy levels than existing alternatives particularly at low sequencing coverage. The issue of whether the method is good at detecting changes in ploidy level does not appear to be explored.

My understanding of the model is that the HMM part of the model is used to model the changes in ploidy level. Perhaps I am missing something obvious, but the authors don't seem to exploit this feature in their simulations (i.e. they all have constant ploidy level). So my guess is that the superior performance of the method compared to the existing approaches is mostly coming from having a better model for the genotype likelihood and the error process rather than it being an HMM. With this is mind it would be good to move the discussion of this part of the model into the main text rather than the Supplementary Material.

Minor points

Keywords: check spelling of poliplody

Line 37 'the evolution' -> 'evolution'

The paragraph at the top of page 3 has a few typos/grammatical issues. E.g 'reference data at known ploidy set…', incorporate -> incorporates

Line 74, by diallelic do you mean that you only see at most two states at a particular site across the sample of genomes under consideration (e.g A/G or C/T) regardless of how many copies of the site there are? Or is diallelic with respect to a sequencing read, i.e. there are only two variants of a read?

Line 83. Is O_{m,n} a sequencing read or just a single site?

line 92. Is the population frequency F_n assumed to be known or is it also something that needs to be estimated? How is this done?

Line 92

It seems odd to have the genotype likelihood relegated to the supplementary material when it is a very important component of the model. It isn't a long section, so I'd suggest moving it to the main text.

line 128 how are the alpha and beta parameters for the Poisson Gamma distribution selected/estimated?

Line 155 I find the description of the simulation a bit confusing, you say that ploidy chosen from 1 to 5 is constant along the genome. I thought the point of the HMM was to be able to detect changes in ploidy level?

Line 245

"allows to overcome" -> "allows the method to overcome"

Line 248, I don't understand the sentence starting "On the former point…"

Line 256 missing full stop

Line 264 tuntimes -> runtimes