Review of the article entitled "High-quality SNPs from genic regions highlight introgression patterns among European white oaks (*Quercus petraea* and *Q. robur*)"
Authored by Lang et al.

Lang et al produced Sanger sequences from over 800 gene fragments (including a set of genes representing broad functional categories potentially involved in species ecological preferences as well as a random set of genes) across the genic portion in 25 individuals of 3 European oak species. They set up a pipeline to clean up and characterize these gene fragments giving over 14500 polymorphisms that were used to provide various summary statistics withing and among species. The authors observed patterns of significantly higher diversity in *Q. petraea* vs *Q. robur* and a heterogeneous landscape of both diversity and divergence. The authors highlighted the usefulness of the generated data in medium scale landscape and molecular ecology projects.

Comment to authors

The manuscript is very well written, the provided data are sound and well used to support the drawn conclusions and the discussion section was well constructed. The generated resources will be valuable for the community. I particularly liked the fact that answers to questions coming up while reading the manuscript could be found in the discussion part. For example, one immediate question was the usefulness of such a data covering only 529Kb of genic regions, which corresponds to barely 0.072% of the *Q. robur* genome or 1% of the gene space length, while methods such a GBS are very popular nowadays with decreasing costs.
Even though NGS methods are now the preferred ones for genomic studies, Sanger sequences still valuable resources especially regarding their lower error rates; and as indicated by the authors, this dataset will be useful as control for future NGS sequences.

- Although the authors acknowledged a possible ascertainment bias depending on which materials their produced resources will be used on, it would be useful that they discuss the possibility that the produced resourced might be skewed towards more conserved genes in Quercus and we know that the more transferable primers are, the more conserved the targeted genomic regions are. For instance, what would be the outcome if primers specific to each species were used (pairs of primers amplifying fragments in one species and failing in the other and vice versa, hence targeting more divergent loci), in terms of calculated summary statistics). Based on the current data, could the authors specify whether some primer pairs were actually successful in only a specific species? If so, would it be possible to produce within species summary statistics for those amplicons and compare them to common ones?

- What species were used to produce the first 103.000 Sanger sequences?

- Line 740 – : I don’t quite agree with the authors. Since it is possible to multiplex hundreds of samples for methods such as RAD-seq with a reasonable cost, such methods could capture even larger genome portions than the one obtained here, to address questions such as those addressed in the manuscript. If there is a low overall differentiation as mentioned, it is even more likely that enzyme digestion produces similar generated fragments for sequencing. Of course, simulation using the published *Q. robur* genome could tell what might be the proportion of genic regions that would be sequenced if only genic regions are of interest. And to me, developing SNP arrays is another question that should be separated from methods such as RAD-seq.