Comparison of pool-seq methods in 2019 and 2023 hake papers

Dr. Ralph Tiedemann (Affiliation) served on the Panel at the (2019 meeting). As he has considerable expertise in genomics laboratory techniques, he was asked by the 2023 Panel to comment on how the pool-seq methods used in Forde et al. 2023 (MARAM_IWS_2023_Hake_BG1) compare to those used in Henriques et al. 2019 (MARAM/IWS/2019/Hake/P3). Below is his response.

The genotyping/pooing approach is essentially similar (the same data?), however the data analysis in Forde et al. 2023 is much more comprehensive and several analyses are specifically tailored towards pooled data. In its current form, I would accept that the approach can generally discern between panmixia and population structure. However, if there is structure, it may be quite difficult to infer the number of strata: Say, for the sake of the argument, that CN and WC2 hold different population which admix in WC (the in-between area). If CN and WC2 are sufficiently differentiated, they will significantly differ in their allele frequencies. However, in the admixture area WC, you cannot assign the individuals to the two respective populations of origin (because of the pooling). Instead, you would get a third set of allele frequencies which may look like a third population. This consideration might be of relevance for Merluccius capensis, where previous analyses inferred two strata, but the pool-seq supports three strata.

One might argue that in the above mentioned example WC should be intermediate between CN and WC2 in the PCA. However, as any pool is essentially only one data point, we have very few data points in the PCA which nonetheless tries to stratify the few data points along principal components. I have no experience with PCAs using so few data points (is it applicable at all?), but for me it seems possible that random fluctuations in some SNP allele frequencies among the pools would be picked up for (pseudo-)stratification. In so far, I found the site frequency distributions (in Figure 3 and Tables 1 and 2 of Forde et al.) more informative and indeed, if they are significantly skewed this could indicate population structure. Yet, I could envision that areas of admixture could differ in their spectra from both populations of origin and would then erroneously inferred as separate populations (along the lines outlined above), if the analysis is run with pooled data as here.