Compare knowledge about relatedness between genotypes and results of our population genetics analysis.

Prof. Dr. Asis Hallab

December 19, 2024

1 Introduction and data

We have knowledge about certain Untwist lines to be identical or closely related to public lines, both of which were analyzed in this population genetics assay. The knowledge on genetic similarity is extracted from supplemental table 1 of the current manuscript (December 2024).

This knowledge can be summarized as follows:

- UNT27 likely CAM174
- UNT29 likely CAM25
- UNT32 likely CAM67
- UNT34 likely CAM134
- UNT38 = CAM57
- UNT50 = CAM147
- UNT51 = CAM150
- UNT52 = CAM156
- UNT53 = CAM159
- UNT54 likely CAM165
- UNT55 maybe CAM167
- UNT56 = CAM232
- UNT57 = CAM278

Note that this list has been shared first in 2023.

2 Manual verification of correctness of the population genetics results

Based on the ADMIXTURE and hierarchical clustering results as visualized in figure

`./results/all_public_and_all_untwist_SNP_filtered_admixture_k8_barplot_IBS_hclust.pdf` (Figure 1 in the document "FOR REVIEW Manuscript Figures (V7).pdf") we can manually check the above prior knowledge (see section `Introduction and data`).

We list the above assumptions of genetic identity that are *not* reproduced by our population genetics analysis. Thus, all *not* listed below are confirmed.

- "UNT27 likely CAM174" cannot be confirmed. Both lines are positioned in different part of the dendrogram and distinct genetic markup (ADMIXTURE)
- "UNT32 likely CAM67" cannot be confirmed. Both lines are positioned in different sections of the dendrogram. However, they share approximately half of their genetic markup (population ancestry) as seen in the ADMIXTURE results (light green part of the respective bars).
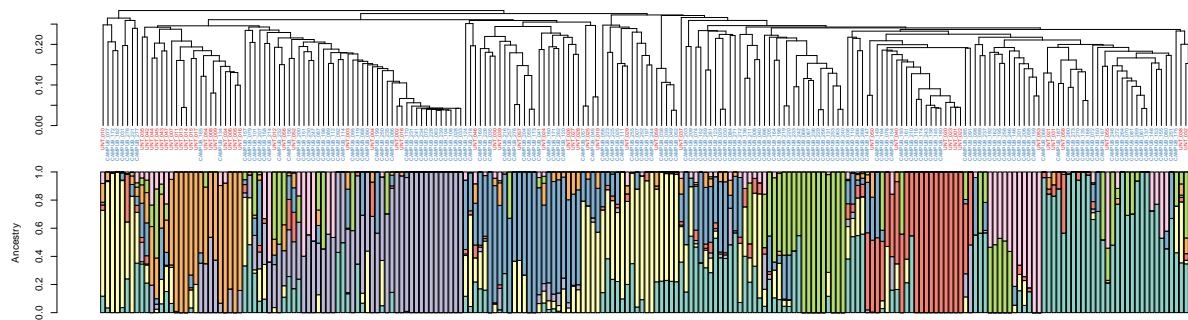
Figure 1: Dendrogram and ADMIXTURE results

- "UNT38 = CAM57" cannot be confirmed. Both lines are positioned in different parts of the dendrogram. None the less, they *do* have some genetic similarity (light yellow parts of their genetic ancestry in the admixture plot in the left panel).
- The line UNT51 is *not* in our dataset.

3 Conclusion

The a-priori knowledge about genetic similarity between Camelina lines studied in this population genetics assay can largely be confirmed (75 %), with three exceptions. The line "UNT51" was not part of the population genetics analysis due to missing data. Furthermore the suspected ("likely") identity of "UNT27" likely being "CAM174" could not be confirmed, neither by the identity by state based clustering nor the ADMIXTURE analysis. Finally, the expected ("=") identity of lines "UNT38" and "CAM57" is not confirmed by the clustering, but the ADMIXTURE analysis shows some genetic similarity. A similar result is obtained for the expected genetic similarity of UNT32 and CAM67, where the phylogenetic clustering does not indicate close relatedness. However the ADMIXTURE analysis does show that both lines share approximately 50% of genetic ancestry.