Highly Informative SNP Marker Panels in Yam (*Dioscorea rotundata*) for Genomic Selection and Prediction.

Author: Paterne Angelot Agre (p.agre@cgiar.org) & Asrat Asfaw (a.amele@cgiar.org)

Source code: https://github.com/pa279/Panel-SNP-marker-for-GS-

DOI: https://doi.org/10.25502/xjtv-mq33/p

Background

Yam is a food choice in many African countries, and its cultivation is among the fastest-growing agricultural commodities in the tropics and subtropics (Asfaw et al., 2021). Its production has consistently increased in recent years, especially in sub-Saharan Africa (FAO, 2020). This trend is expected to continue with the ever-increasing food demands of a growing human population. Meeting the steadily growing market of yam as a food of choice requires rapidly implementing technological innovations that improve production. Genetic improvement technologies such as genomic selection, which utilizes genetic markers to identify plants of superior genetic values to improve production and food quality traits, are increasingly used by many crop breeding programs (Aguilar et al., 2010; Legarra et al., 2014). Such genetic technologies have shown to be highly effective in realizing genetic gains in many crop plants. To enable the yam breeding programs access to the benefits of genomic selection, we selected a panel of highly informative SNP markers well distributed across the yam genome.

Selection of highly polymorphic SNP markers

The Yam Breeding Unit has developed a set of SNP markers for genomic prediction and selection. This set of selected markers can also be used for the Genomic Prediction of Cross Performance (GPCP). We have documented the different steps in selecting highly informative SNP markers for such analysis. From 136K SNP markers generated from the whole genome re-sequencing, we filtered the original markers to only retain markers with high genetic parameter function using a customized R script. We then retained only markers with a ratio of observed to expected heterozygosity of less than 1.5, a GC content of 40 to 60% in the 101 bp window extending 50 bp on either side of the SNP, and two or fewer flanking SNPs in that same window. An annealing method based on a simulated approach was used to choose the set of SNP markers. In the second approach, the dataset was subdivided into sub-population using discriminant analysis of principal components and run against a simulated annealing algorithm developed in R to find a set of 3000 markers that maximized diversity within and between the

sub-populations. Additionally, QC/QA, and associated SNP markers to key traits previously identified in the yam breeding program were also included in the final list of the markers to be used for prediction in the yam breeding. Flanking SNPs were ignored if their quality score was 900 lower or overlapping. Finally, 3,092 markers were selected in total. The process of the SNP markers selection, including all R codes, is documented and available in GitHub page (https://github.com/pa279)

The selected SNP markers were later mapped against the reference genome and their distribution were assessed (Figure 1). List of markers including physical position, chromosome and flanking sequencing can be downloaded from https://figshare.com/account/items/22619971/edit

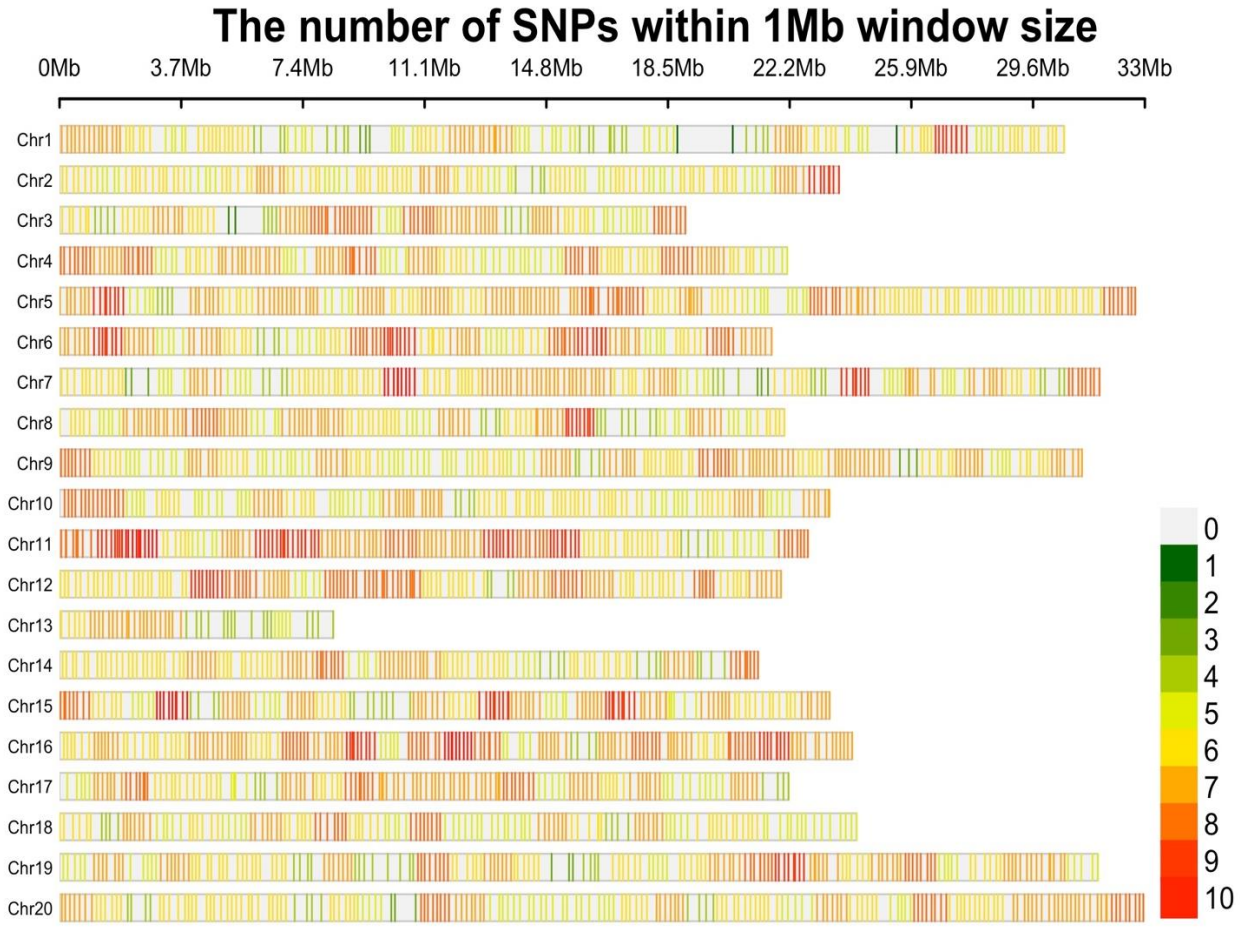

Figure 1: Distribution of selected SNP markers across the yam genome

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93: 743-752.

Asfaw, A., Aderonmu, D. S., Darkwa, K., De Koeyer, D., Agre, P., Abe, A., et al. (2021). Genetic parameters, prediction, and selection in a white Guinea yam early-generation breeding population using pedigree information. Crop Sci. 2021, 61(2)–1038-1051. doi: 10.1002/csc2.20382

FAOSTAT (2021). Food and agricultural organization of the United Nations. In: FAOSTAT statistical database, FAO (Rome). Available at: http://www.fao.org/faostat/en (Accessed 1 December 2021).

Legarra, A., O. F. Chistensen, I. Aguilar, and I. Misztal. (2014). Single step, a general approach for genomic selection. Livest. Prod. Sci. 166:54-65.