

# Marker imputation and linkage map construction using a skim sequenced diploid potato population

C Clot<sup>1</sup>, X Wang<sup>1</sup>, A Thérèse Navarro<sup>1</sup>, J Bucher<sup>1</sup>, RGF Visser<sup>1</sup>, R Finkers<sup>1,2</sup>, HJ van Eck<sup>1</sup>

<sup>1</sup> Plant Breeding, Wageningen University and Research, Po Box 386, 6700 AJ Wageningen, The Netherlands

<sup>2</sup> GenNovation B.V., Agro Business Park 10, 6708 PW Wageningen, The Netherlands



## Background

With the low cost of next-generation sequencing and the development of algorithm adapted to outcrossing crops [1], genotyping-by-sequencing strategies based on skim sequencing and imputation have become increasingly popular. Unfortunately low depth skim sequencing results predominantly in missing values. To overcome this problem we implemented in R an imputation pipeline based on clustering of parent specific SNPs in bins of 100 kb. We applied it to a population of 1536 full-sibs descending from heterozygous diploid potato sequenced at an average read depth 1.7x.

## Results

- We imputed 4893 female and 4735 male segregating markers and used these to construct parental linkage maps.
- Linkage maps displayed high collinearity with the physical map, except for recently discovered 5.8-Mb inversion on chr03 [2].
- Local failure of imputation coincided with low parent specific SNP density (e.g. male chr07) or extreme skewed segregation (e.g. male chr01).
- Skewed segregation was identified on eight chromosomes. Some distortion colocalized with known loci (e.g. S-locus, StSIEL) [3].
- Using the imputed markers in a single marker regression model we identified a major QTL for yield colocalizing with *StCDF1*.

## Methods

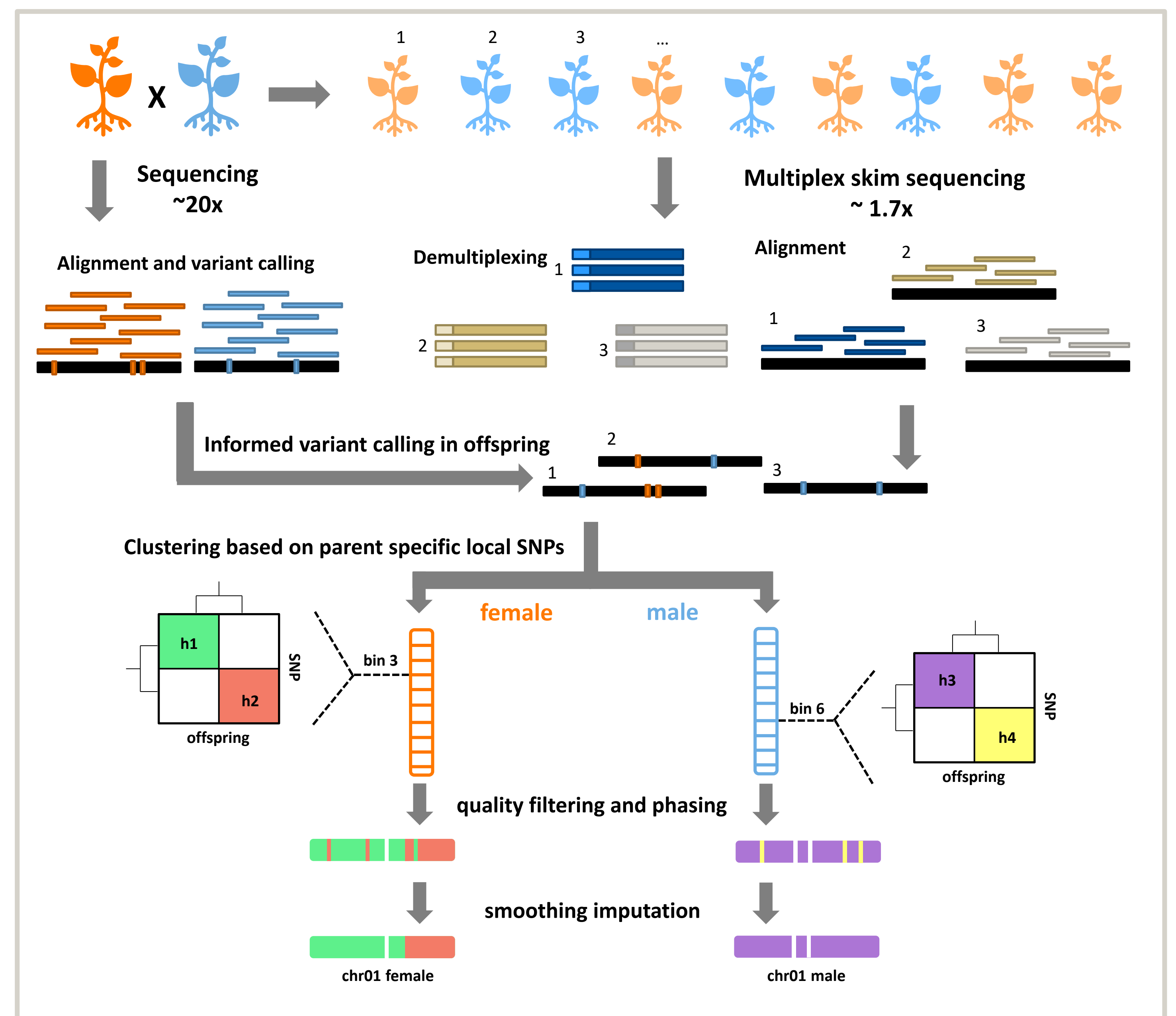


Figure 1. Imputation pipeline (modified from OutcrossSeq [1])

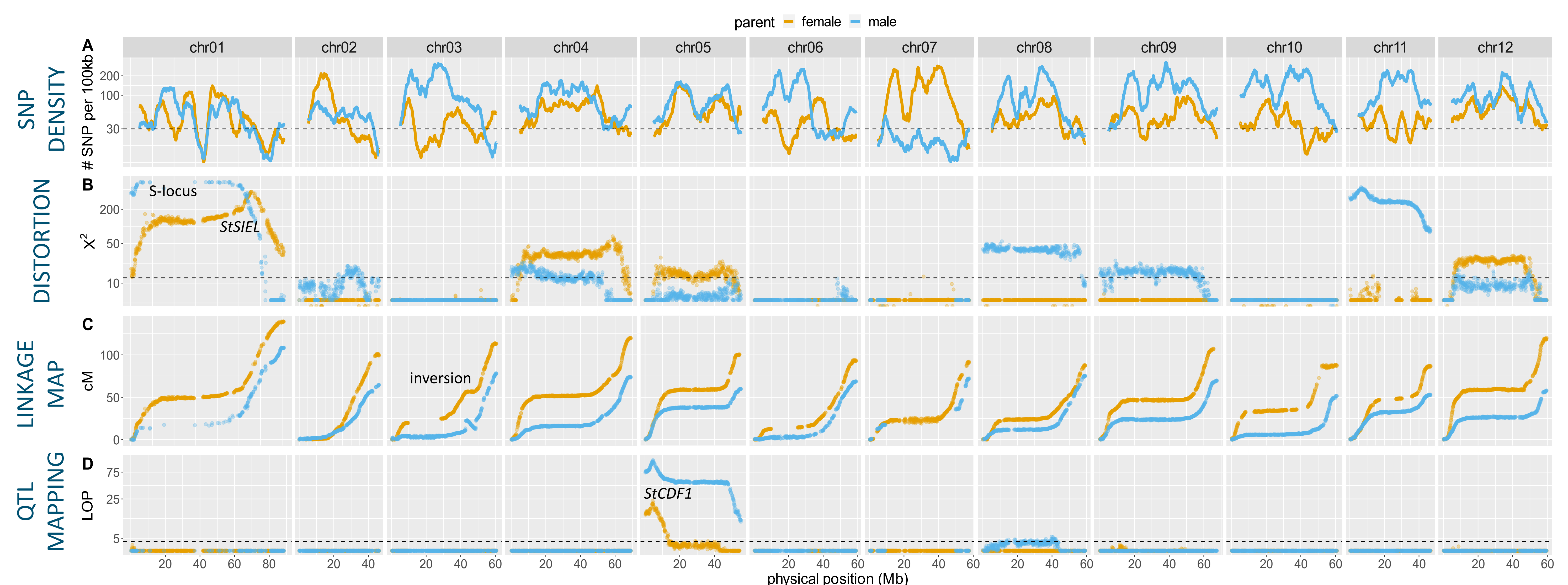


Figure 2. **A.** Number of parent specific SNPs per bins of 100kb used for imputation (moving average over 5Mbp). **B.** Deviation from Mendelian segregation ratio measured by chi-square goodness-of-fit test. **C.** Relationship between physical and genetic distance. **D.**  $-\log(p\text{-value})$  of single marker regression for yield in greenhouse (n=1534).

## Conclusion

The pipeline described in this poster facilitates the imputation of genotyping data from skim sequenced full-sib diploid populations. The resulting imputed markers can be used to construct high density linkage maps, explore transmission ratio distortion and identify loci regulating important traits.

## References

- [1] Chen, M., Fan, W., Ji, F., Hua, H., Liu, J., Yan, M., Ma, Q., Fan, J., Wang, Q., Zhang, S. and Liu, G., 2021. Genome-wide identification of agronomically important genes in outcrossing crops using OutcrossSeq. *Molecular plant*, 14(4), pp.556-570.
- [2] Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., Bao, Z., Liu, Z., Feng, S., Zhu, X. and Li, D., 2022. Genome evolution and diversity of wild and cultivated potatoes. *Nature*, pp.1-7.
- [3] Zhang, C., Wang, P., Tang, D., Yang, Z., Lu, F., Qi, J., Tawari, N.R., Shang, Y., Li, C. and Huang, S., 2019. The genetic basis of inbreeding depression in potato. *Nature Genetics*, 51(3), pp.374-378.