



8th INTERNATIONAL MEETING ON PLANT BREEDING

AI for Scientific Breakthrough: Applications and Innovations

1st to 3th October | Piracicaba – SP, Brazil



PARENTAGE ANALYSIS IN AUTOPOLYPLOIDS: INSIGHTS FROM SIMULATED AND EMPIRICAL DATA IN TETRAPLOID POTATO (*Solanum tuberosum*)



Innocent Vulou Unzimai^{1*}; Carla Cristina da Silva¹; Thiago Mendes²; Moctar Kante²; Guilherme da Silva Pereira¹

¹Federal University of Viçosa, Brazil unzimai.vulou@ufv.br

²International Potato Center, Peru

1. Introduction

- ❑ Mislabeling and pedigree errors in outbreeding clonally propagated crops like potato are inevitable due to complex work operations during parental crossing and offspring propagation (Fig. 1).
- ❑ An increasing loss of offspring-parental identity occur especially when big genetic crossing blocks are set up. Literature points to error rates of 5%-10% in many breeding programs, with pedigree errors having an adverse impact on genetic data analysis.



Fig.1. An illustration of genetic crosses and their associated labelings and taggings

2. Goal

To obtain proportions of true and false offspring-parental assignments using dosage-sensitive single nucleotide polymorphisms (SNPs) molecular marker data in tetraploid potato (*Solanum tuberosum*)

3. Breeding Population and Methods

3.1. Simulated data

- ❑ Using *MultiPolyPop* R package, we simulated a partial diallel pedigree population from crosses between 14 females and 4 males (full-sib family size = 35) based on *PedigreeSim* software, together with sequence read depths and inferred dosages for 1,630 SNPs based on *updog* R package.
- ❑ By combining an increasing sample of markers (10, 30, 60, and 90 SNPs) with varying levels of missing data (0%, 10%, and 20%), we have obtained 1,000 files per combination (Fig. 2).

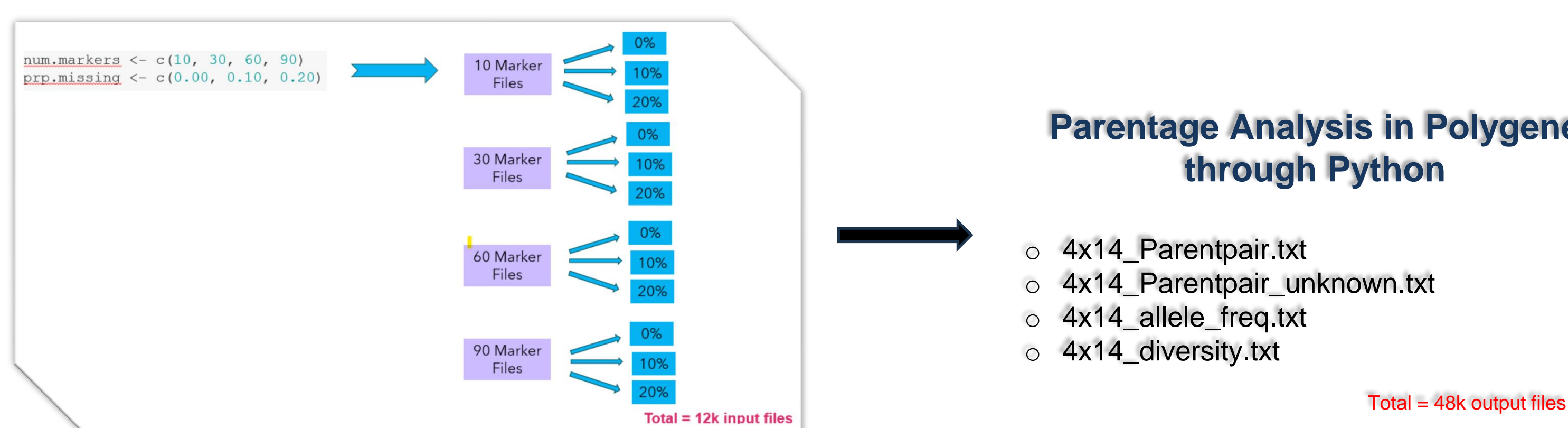


Fig.2. An illustration of sampling of varied number of markers at different proportions of missing data as input files and their associated output files after parentage analysis

3.2. Empirical data

- ❑ Empirical genotyping data for a potato population from CIP-Peru generated from crosses between 13 females and 4 males (N = 359 individuals) was done using DArTag technology with 2,315 SNPs (filtered at 10% missing data).
- ❑ Parentage analysis was conducted in *Polygene* v.1.7 software using a maximum likelihood-based polysomic inheritance model to obtain the best possible parental assignment combinations for each offspring (simulated or empirical data) (Fig. 3).

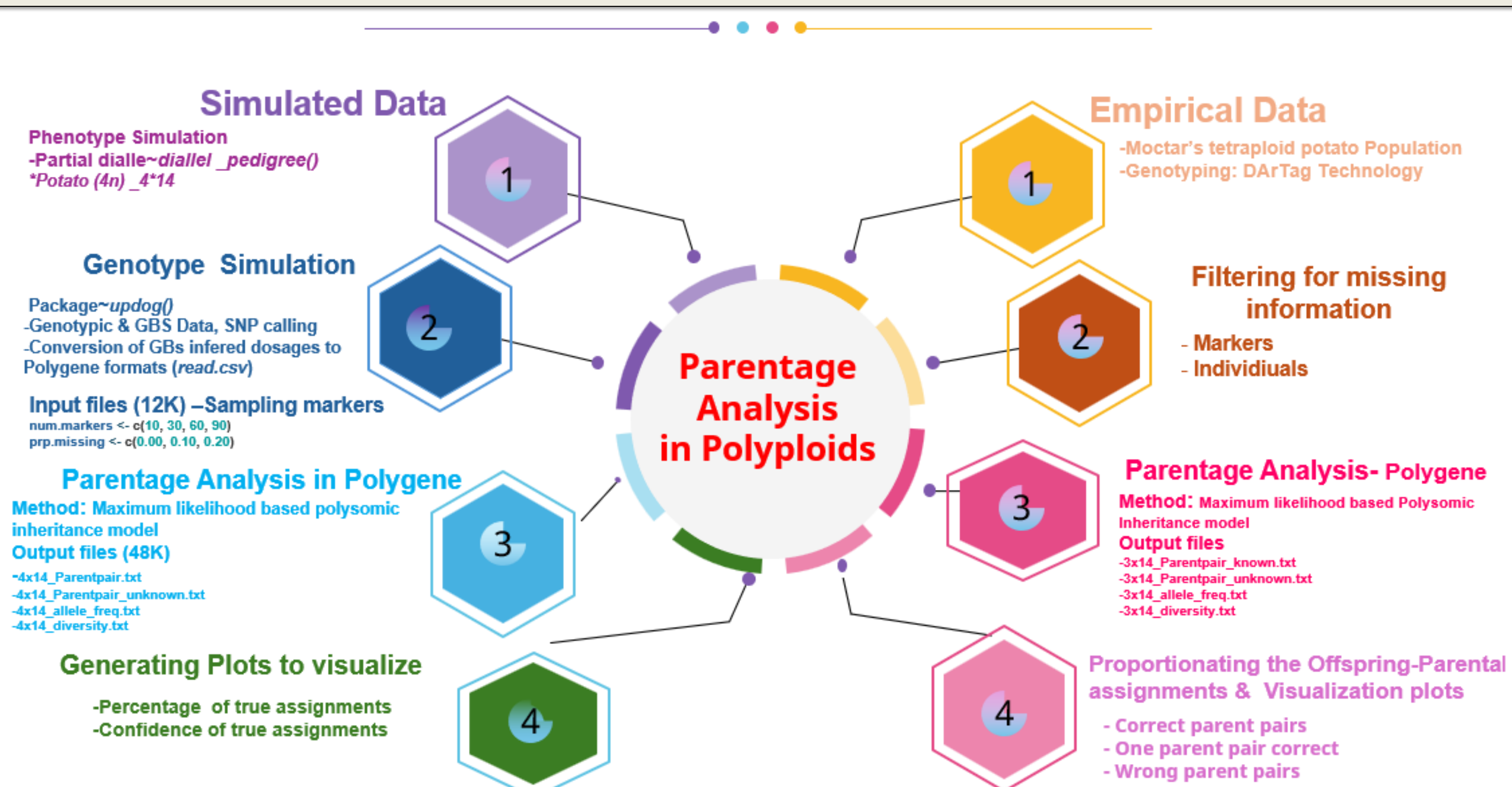


Fig.3. A nutshell of Parentage Analysis overall workflow in Autopolyploid

4. Results

- ❑ Simulated data results have shown that the percentage of true assignments was higher for crosses when parent sex is known (Fig.4) compared to when sex is unknown (Fig.5).
- ❑ In the former case, 100% true assignments were achieved at 90 markers with 0% and 10% missing data, while 90% true assignments were recorded for samples with 20% missing data.

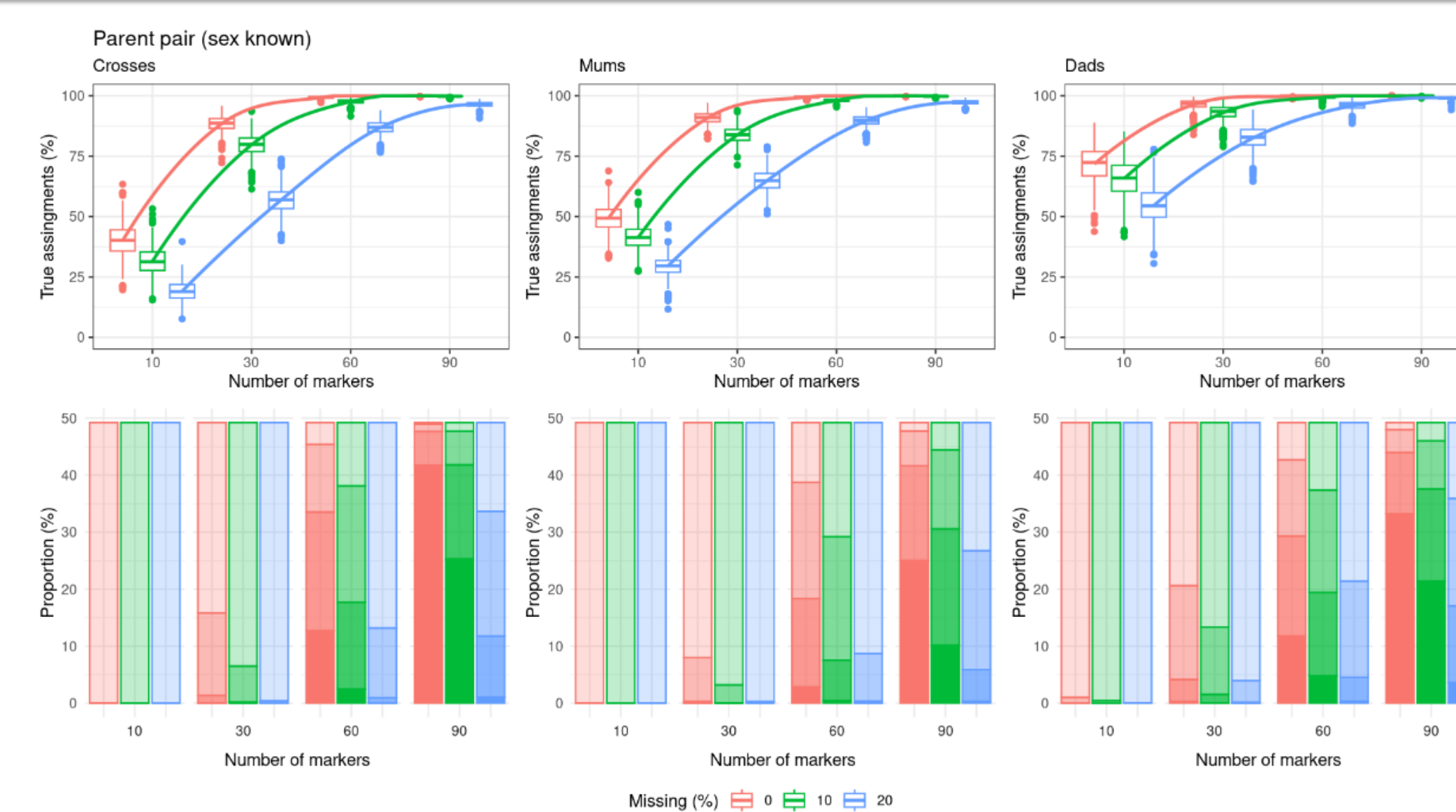


Fig.4. Percentage and confidence of true assignments in parent pair sexes-known

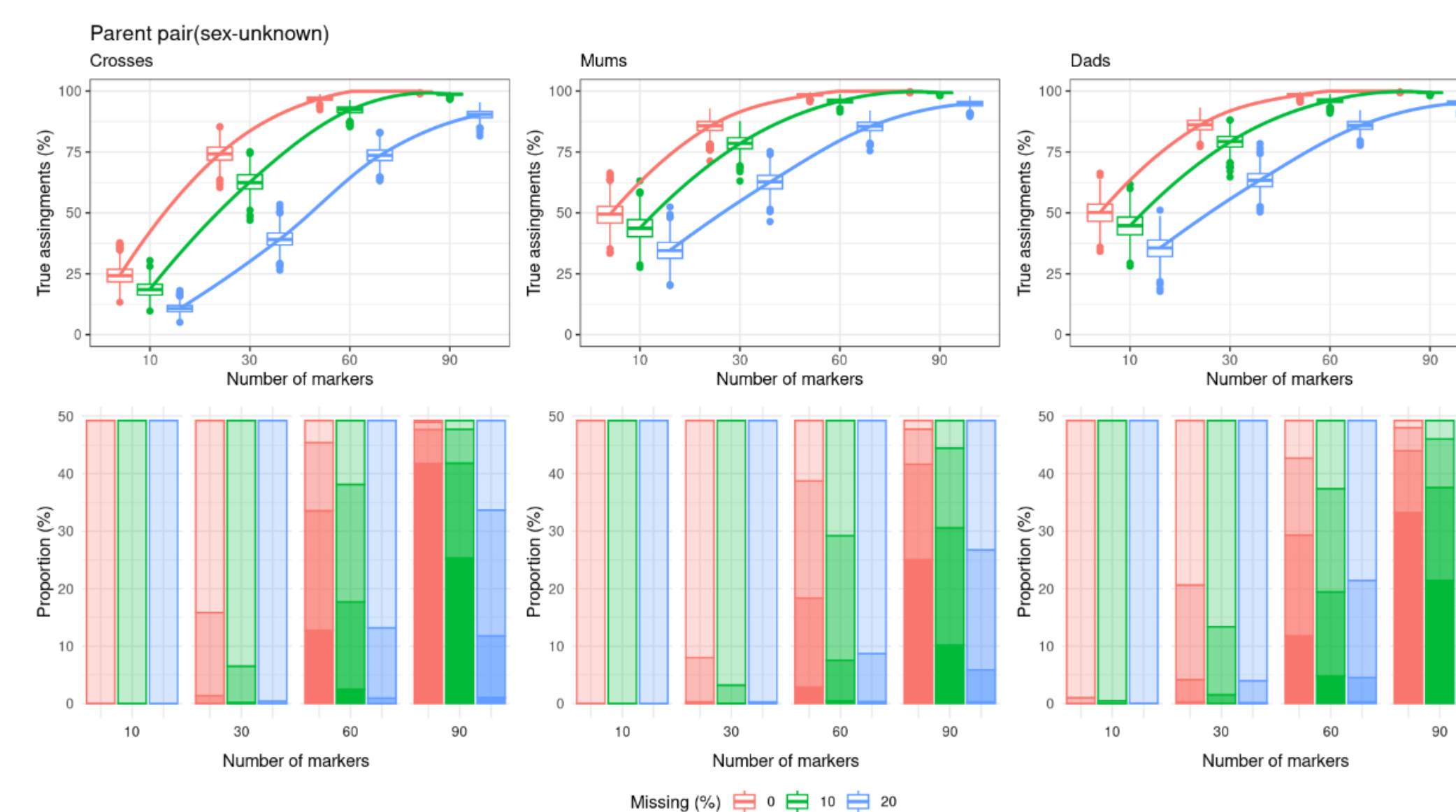


Fig.5. Percentage and confidence of true assignments in parent pair sexes-unknown

- ❑ Empirical data results show 150 (41.8%) individuals with correct assignment for both parents, 172 (47.9%) individuals with one correct parent (either female or male), and 37 (10.3%) individuals with both parents incorrectly assigned (Fig.6.)

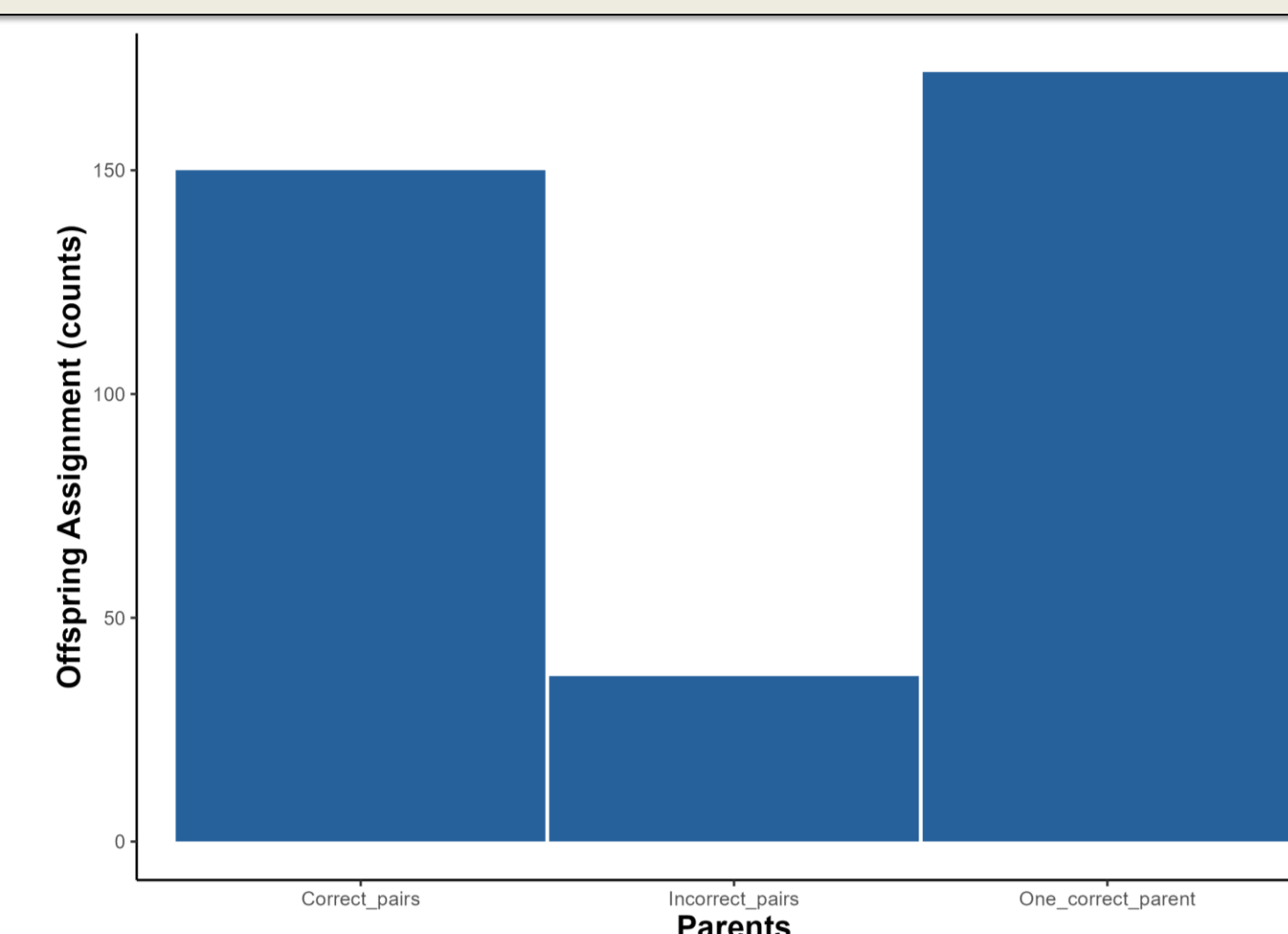


Fig.6. Proportion of the offspring-parental assignments in counts

5. Conclusion

- ❑ This study demonstrates the importance of parentage analysis to ensure higher quality control and fix pedigree issues before conducting any genetic data analysis.
- ❑ We recommend that potato breeding programs continuously improve their genetic and experimental designs through correct pedigree annotation, proper pollination control, and non-detachable labeling tags to avoid mix-ups.

6. Reference

Huang, K., Huber, G., Ritland, K., Dunn, D. W., & Li, B. (2021). Performing parentage analysis for polysomic inheritances based on allelic phenotypes. *G3*, 11(2), jkaa064.

7. Acknowledgment

We are grateful to CAPEs for sponsoring my study, to CIP-Peru for providing us with the empirical data and to UFV for the computational facilities used for conducting the analysis of this project.

