Pedigree reconstruction for plant species using simulated annealing: case study of oil palm

David CROS¹, Jesus FERNÁNDEZ², Jean-Marc BOUVET¹, Léopoldo SANCHEZ³

¹ Genetic improvement and adaptation of Mediterranean and tropical plants Research Unit (AGAP), CIRAD, International campus of Baillarguet, 34398 Montpellier, France
² Departamento de Mejora Genética Animal, INIA, Ctra. Coruña Km 7.5, 28040 Madrid, Spain
³ Forest Tree Improvement, Genetics and Physiology Research Unit (AGPF), INRA, 2163 Avenue de la Pomme de Pin, 45075 Orleans, France

Introduction

The approach of Fernandez and Toro (2006) to reconstruct pedigrees from molecular data of contemporaneous individuals of animal species using a simulated annealing algorithm was extended to plant species and made more flexible. New features include the choice of the mode of sexual reproduction (hermaphroditism, monoecy or dioecy), the possibility of selfings, accounting for a predefined coancestry matrix between founders and specifying different number of individuals per generation.

Validation

The new method was validated using 16 individuals from the last generation of the Yangambi breeding population of oil palm. Their pedigree was known for 5 generations and they were genotyped with 166 SSR. According to the percentage of relationships to reconstruct and the number of SSR used, the Pearson correlation between the pedigree-based coancestries calculated on the true and on the reconstructed genealogies ranged from 0.74 to 0.99 (Fig 1A). The RMSE ranged from 0.02 to 0.12 (Fig 1B). Even when pedigree was assumed completely unknown, accurate reconstruction was achieved with 38 SSR. Using 100 SSR or more, the Pearson correlation was very high (0.98) and the RMSE very low (0.06). Unlinked SSR were more efficient than linked SSR, as they gave a RMSE between true and reconstructed genealogical coancestries 37% smaller than linked SSR, and a Pearson correlation 21% higher (Fig. 2).

Application

The new method was applied to 104 individuals from the last generation of a key breeding population (Deli). The individuals were genotyped with 160 SSR. Their pedigree was only known for the recent past and was reconstructed on 7 generations. Coancestries calculated on the reconstructed genealogy and molecular coancestries were highly correlated (>0.9) when using 80 SSR or more (Fig. 3).

Conclusion

This method gave likely pedigrees with satisfactory reliability for plant species, using a realistic number of polymorphic markers. The methodology has been implemented in the software MOL_COANC (version 3).

Fig. 1 Effect of percentage of unknown parentages and number of SSR markers on the accuracy of reconstruction of Yangambi pedigree. A. Pearson correlation and B. Root mean square error (RMSE) between true and reconstructed genealogical coancestries. Values are means over 64 replicates ± SEM.

Fig. 2 Effect of linkage between SSR on the reconstruction of pedigrees. True and reconstructed genealogies are compared by Root mean square error (RMSE) and Pearson correlation. Values are means over 3 replicates ± SD.

Fig. 3 Effect of the number of SSR markers on Pearson correlation between molecular and reconstructed genealogical coancestries in Deli population. Values are means over 8 replicates ± SEM.