



**HAL**  
open science

# Evaluation of the power of assignment methods on chloroplast microsatellites for provenance identification of reforested stands

Miguel Navascués

► **To cite this version:**

Miguel Navascués. Evaluation of the power of assignment methods on chloroplast microsatellites for provenance identification of reforested stands. Population Genetics and Genomics of Forest Trees: from Gene Function to Evolutionary Dynamics and Conservation, Oct 2006, Alcalá de Henares, Spain. 2006. hal-02755472

**HAL Id: hal-02755472**

**<https://hal.inrae.fr/hal-02755472v1>**

Submitted on 3 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Evaluation of the power of assignment methods on chloroplast microsatellites for provenance identification of reforested stands or seedlots



**Miguel de Navascués**

Unit of Forest Genetics, Department of Forest Systems and Resources, Centre of Forest Research (CIFOR-INIA), Carretera de La Coruña km 7.5, 28040 Madrid, Spain [m.navascues@gmail.com](mailto:m.navascues@gmail.com)

In this work, the performance of several assignment methods have been assessed when applied to chloroplast microsatellite (cpSSR) data for the identification of geographic origin of seedlots or reforestations.

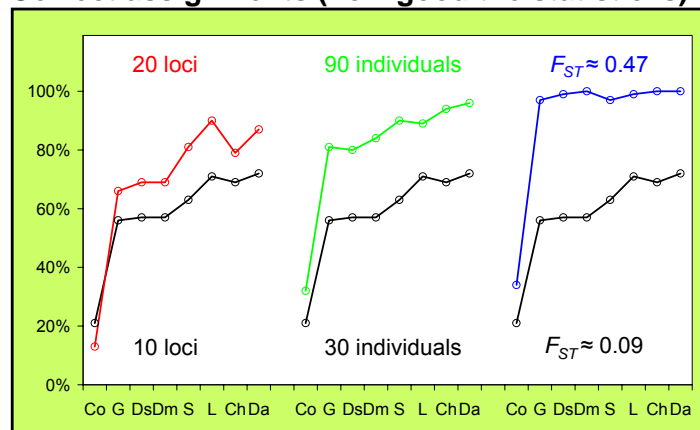
These methods share a common procedure: first, a statistic (i.e. genetic distance, likelihood ratio, etc) is calculated for every pair reference population-sample to assign. The value of this statistic gives a criterion to rank the candidate populations of origin in an order of increasing fit to the sample to assign, usually identifying one population with the best fit. In a second step, the data from each candidate populations is resampled to generate the distribution of the statistic. This distribution compared to the obtained statistic allows to estimate the probability of getting such value on the statistic for a sample belonging to the candidate population. (NB: The method of conformity follows a different procedure. Although some results on this method presented here, they need a more detailed discussion that will be done in a future work)

Artificial datasets were generated with SIMCOAL. We simulated the evolution of cpSSR loci in 10 isolated populations that diverged from a common population a number of generations in the past. For each run of the simulation, the resulted data for 10 cpSSRs was organised in 10 reference samples of 30 individuals and 10 samples for identification of their origin with 30 individuals each. The populations diverged from 1000 generations ago (approx.  $F_{ST} \approx 0.09$ ).

The power of the methods for this set of parameters was analysed by counting the proportion of correct assignment based on the population with the best fit of the statistic and by calculating the 'quality index' (Piry *et al* 2004 *J Hered*), a measure that evaluates whether the true population of origin gets the best score from the resampling procedure (results in black in the figures). Higher number of loci and sample size or the study of more divergent populations can potentially increase the power of these methods. To explore these scenarios further simulations were carried out only changing one of the parameters at a time (results represented in red/green/blue in the figures).

The best results were obtained when using Nei's  $D_A$  distance as a test statistic. The low performance of the conformity method is due to the low frequency for getting only one origin with a 'best score', but when this happened it was the correct origin 99-100% of the times for all scenarios.

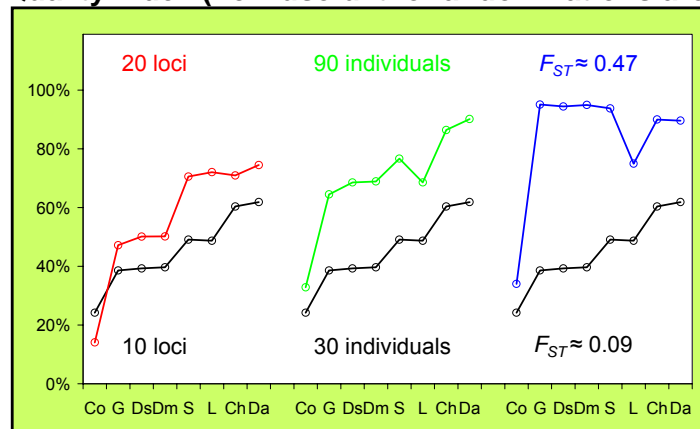
## Correct assignments (how good the statistic is)



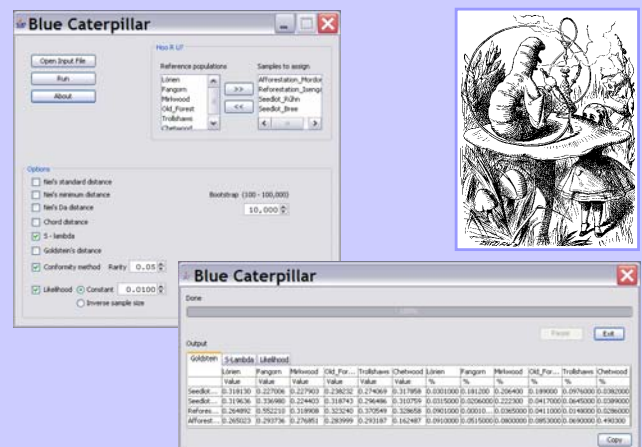
## Methods Assessed

- Conformity method **Co** (see Deguilloux *et al* 2003 *Mol Ecol*)
- $S_A$  method **S** (see Ribeiro *et al* 2002 *Ann For Sci*)
- Likelihood ratio method **L** (see Paetkau *et al* 1995 *Mol Ecol*)
- Distance methods (see Cornuet *et al* 1999 *Genetics*)
  - Goldstein **G** (Goldstein *et al* 1995 *PNAS*)
  - Nei's Standard **Ds** (Nei 1972 *Am Nat*)
  - Nei's Minimum **Dm** (Nei 1973 *PNAS*)
  - Nei's  $D_A$  **Da** (Nei *et al* 1983 *J Mol Evol*)
  - Chord distance **Ch** (Cavalli-Sforza & Edwards 1967 *Evolution*)

## Quality index (how useful the randomizations are)



## Blue Caterpillar v1.0 beta by J Navascués & M Navascués



The methods assessed in this poster can be applied to haplotypic non-recombining data with **Blue Caterpillar v1.0 beta**. More information at <http://m.navascues.googlepages.com>

## Acknowledgements

This work was funded by grants CC03-048 (INIA/Ministry of Environment) and REPROFOR, AGL2005-07440-C02-01 (Ministry of Education and Science). I am indebted to Santiago González-Martínez and Ricardo Alía for supporting me doing this project. I am also grateful to Concetta Burgarella for helpful discussion. The collaboration of Julián de Navascués on the software development was fundamental for the project.