Annals of Forest Science (2013) 70:631–635
DOI 10.1007/s13595-013-0302-z

ORIGINAL PAPER

RaBoT: a rarefaction-by-bootstrap method to compare genome-wide levels of genetic diversity

Ivan Scotti · William Montaigne · Klára Cseke · Stéphane Traissac

Received: 10 September 2012 / Accepted: 22 May 2013 / Published online: 7 June 2013

Abstract
• ***Context*** No efficient method is available to compare multi-locus estimates of diversity while taking into account inter-locus and inter-population stochastic variance. The advent of genome scan approaches makes the development of such tests absolutely necessary.
• ***Aims*** We developed a method to compare genome-wide diversity estimates while taking into account—and factoring out—variation in census size and making use of inter-locus variance to assess significance of differences in diversity levels.
• ***Methods*** An approach based on rarefaction with bootstrap re-sampling (RaBoT) was implemented into a test of multi-locus comparison of diversity coded in R. The properties of the test were studied by applying it to simulated populations with varying diversity levels and varying differences in diversity levels. The test was then applied to empirical data from disturbed and undisturbed populations of *Virola michelii* (Myristicaceae) genotyped at 693 amplified fragment length polymorphism (AFLP) markers.
• ***Results*** RaBoT was found to be rather conservative, with large numbers of false negatives when the diversity in the compared populations was similar, and false positives mostly associated to comparisons of populations with extremely high levels of diversity. When applied to empirical data, RaBoT detected higher genetic diversity in a post-disturbance than in an undisturbed population and lower genetic diversity in a seedling than in the corresponding adult population, but it also revealed differences in diversity between subgroups within the disturbed and undisturbed plots.
• ***Conclusion*** RaBoT is a sensitive method to compare multi-locus levels of diversity that can be applied both at the genotype level for dominant markers (e.g. AFLP) and at the allele level for biallelic codominant markers (e.g. single-nucleotide polymorphisms).

Keywords Population genetics · Statistical testing · Genome-level diversity · Genome scan · Diversity comparison

Handling Editor: Bruno Fady

Contribution of the co-authors IS designed the experiment, ran data analyses and wrote the paper.
WM collected data, ran data analyses and wrote the paper.
KC collected data and ran data analyses.
ST ran data analyses and wrote the paper.

Electronic supplementary material The online version of this article (doi:10.1007/s13595-013-0302-z) contains supplementary material, which is available to authorized users.

I. Scotti (✉)
Unité Mixte de Recherche "Ecologie des Forêts de Guyane" (EcoFoG), INRA—Institut National de la Recherche Agronomique, BP 709, 97387 Kourou Cedex, French Guiana, France
e-mail: ivan.Scotti@ecofog.gf

W. Montaigne
Unité Mixte de Recherche "Ecologie des Forêts de Guyane" (EcoFoG), Université des Antilles et de la Guyane, Kourou, French Guiana, France

K. Cseke
Department of Forest Tree Breeding, Experiment Station and Arboretum, Forest Research Institute, Sárvár, Hungary

S. Traissac
Unité Mixte de Recherche "Ecologie des Forêts de Guyane" (EcoFoG), AgroParisTech, Kourou, French Guiana, France

1 Introduction

The comparison of levels of (genetic) diversity in samples with different sizes is generally performed using rarefaction methods (Hurlbert 1971; Petit et al. 1998; Kalinowski 2004), which apply only to (allelic) richness. On the other hand, statistical tests of differences in diversity levels for single loci have been based on the computation of confidence intervals on diversity estimates (Grundmann et al. 2001) and non-parametric tests have been used to compare multi-locus

diversity levels for Nei's (1978) genetic diversity (Tessier and Bernatchez 1999; Jin et al. 2000) or for allelic richness with rarefaction (Kalinowski 2004). Glaubitz et al. (2003a, b) used a permutation approach to compare average multi-locus diversity estimates in pairs of forest tree populations before and after the application of sylvicultural practices. Their method, however, did not exploit stochastic variation among loci because it was based on average effects.

Recent genome scan methods generally rely on biallelic markers such as amplified fragment length polymorphisms (AFLPs) or single-nucleotide polymorphisms (SNPs). These markers are ideal for genome-wide comparisons of genetic diversity with or without rarefaction, but the only measure of diversity to which rarefaction is currently applied is allelic richness, which has limited scope for biallelic markers. Here, we present a method that permits to statistically test genome-wide differences in diversity by a combination of comparisons of estimates obtained from large numbers of loci and bootstrap-based rarefaction, called "Rarefaction-by-Bootstrap Test" (RaBoT). This new tool makes use of multi-marker, biallelic genetic data, obtained from high-throughput genome scan techniques (such as SNP or AFLP markers), to statistically test genome-wide differences in diversity between populations of different census sizes or between samples of different sample sizes (that may or may not be derived from populations of different census size). RaBoT is an improvement relative to currently available methods for the comparison of diversity levels because (a) it takes advantage of rarefaction to compare diversity in populations and samples of different sizes, (b) it applies a bootstrap method to empirically test differences in diversity at the single-locus level and (c) it combines information on diversity differences from large numbers of loci in a composite statistical test, rather than using averages to summarise differences between populations. RaBoT's rationale, robustness and power are first described here, and then the method is applied to compare diversity levels between tree populations growing in disturbed and undisturbed plots.

2 Materials and methods

2.1 Development of the RaBoT algorithm

The test is inspired by rarefaction-based comparisons of species and allele richness (Hurlbert 1971; Petit et al. 1998) but it takes full advantage of data sets containing large numbers of biallelic markers (such as those produced in AFLP- and SNP-based genome scans) and provides a P value associated with the observed data under the null hypothesis of no genome-wide difference in diversity. It can apply to any transformation of the true diversity of order 2 (Jost 2006), such as the Gini–Simpson index (Jost 2006; Simpson 1949) or Nei's genetic diversity (Nei 1978), which are mathematically equivalent. To construct the test, we proceeded as follows. Census (sample) sizes were recorded as N_{small} and N_{large} for the smaller and larger population being compared, respectively. The observed diversity in the smaller population (H_{small}) was computed for each genetic marker. In the larger population, S subsamples were randomly drawn with replacement with sample sizes equal to N_{small}. H was computed for each subsample and for each marker, and the median $m_{H,large}$ was computed from the set of S values obtained for each marker. Finally, for each marker, $\Delta H = H_{small} - m_{H,large}$ was computed. The genome-wide distribution of ΔH (i.e. the distribution of ΔH values drawn from all markers) was inspected. If there was no real difference in diversity between the small and the large population, beyond the effect of population sampling stochasticity, then H_{small} values should be drawn, for each locus, from the same distribution as the random subsamples from the large population and the same locus. Under this hypothesis, the observed H_{small} value for each marker is a random outcome from the same distribution as the corresponding H values for the random subsamples. As a consequence, individual marker ΔH values should have the same probability of being positive or negative, and at the genome level, there should be equal numbers of positive and negative ΔH values. Departures from this expectation were tested by a Chi-square test and were taken as suggestive of real, genome-wide differences in diversity. An R (R Development Core Team 2008) script was used to run RaBoT (see Supplementary Methods 1).

Sensitivity and robustness of the RaBoT method were tested by a simulation approach. Twenty-five populations with varying effective size (N from 10 to 200) and varying diversity ($\theta = N\mu$ from 0.1 to 2, where μ=mutation rate) were generated using EASYPOP (Balloux 2001). Populations were forward-simulated over 10,000 generations to reach mutation–drift equilibrium, starting with a monomorphic population. One hundred unlinked loci were simulated in each population. Pairwise diversity comparisons were then performed using RaBoT, with 100 replicates per locus for all comparisons.

2.2 Application to a real case

RaBoT was subsequently applied to empirical AFLP data from disturbed and undisturbed forest tree populations of *Virola michelii* (see Supplementary Materials 1 for details on the species). Because AFLPs are dominant, it is not possible to identify all alleles and therefore a bootstrap sampling at the allele level was not feasible. RaBoT was applied instead to the presence/absence of AFLP fragments, that is, to AFLP "phenotypes".

The dynamics of *V. michelii* regeneration was studied at the Paracou experimental site (French Guiana; 5°18′N, 52°55′W; http://www.ecofog.gf/spip.php?article174) (see Supplementary

Materials 1 for details). The study was carried out in two contiguous 6.25 ha plots (plot 9 (P9) and plot 11 (P11), Supplementary Fig. 1), separated by a 50-m-wide buffer zone. Plot 11 was a control plot without any logging, while plot 9 underwent experimental logging between 1986 and 1988 (see Supplementary Materials 1 for details). To test whether RaBoT would also detect local variations in diversity that are independent of the hypothesised effect of perturbation, each of the two plots was subdivided into two subplots (named P9E and PW9 for the disturbed plot, and P11E and P11W for the control plot) whose juvenile *V. michelii* populations were compared to each other, to the adult population and to the other plot's subplots. Details of population sampling, AFLP analyses and data scoring are provided in Supplementary methods 2.

RaBoT requires independent markers. Therefore, as a minimal requirement, markers from natural samples should be in genotypic equilibrium. Genotypic disequilibrium was checked for all marker pairs in the sapling dataset by testing significance of two-way contingency tables (with Bonferroni correction) as in Raymond and Rousset (1995). To define a marker set with minimum genotypic disequilibrium for subsequent analyses, we proceeded as follows. After testing all marker pairs for linkage disequilibrium, markers were sorted based on the number of pairs showing genotypic disequilibrium in which they appeared (from highest to lowest). Markers were iteratively removed from the data set one by one and the number of pairs with genotypic disequilibrium was recomputed as above. The largest marker set with fewer than 5 % of pairs in genotypic disequilibrium was retained for further diversity analyses. These iterations were performed through an R (R Development Core Team 2008) script, which is presented as Supplementary Methods 3.

3 Results

The results of simulations used to evaluate the properties of the RaBoT method are visually summarised in Fig. 1. RaBoT had intermediate type I error rates for diversity levels up to $\theta=1$ (two false positives out of 40 comparisons (5 %) at the 5 % significance level) and high type I error rates for populations with high diversity ($\theta=2$, four false positives out of ten comparisons (40 %) at the 5 % significance level). Power was quite high for populations with large differences in diversity (21 false negatives out of 150 comparisons (14 %) at the 5 % significance level for pairs including one population with $\theta=0.1$–0.2 and one population with $\theta=0.5$–2.0) but was low for comparisons of populations with smaller (non-zero) differences in diversity (74 false negatives out of 100 comparisons (74 %) at the 5 % significance level). Population size did not influence these results (Fig. 1).

Diversity was significantly lower in saplings in the control than in the post-disturbance plot: in all comparisons between P11 (control) and P9 (disturbed) subplots, diversity was significantly lower in the control than in the disturbed subplots. On the other hand, diversity in the adult population was always significantly higher than in all seedling subpopulations, no matter whether they grew in the control or in the disturbed subplots. Significant differences in diversity were also found between same-plot subplots (Supplementary Table 1 and Supplementary Fig. 2). Visual inspection of the distribution of the size of differences in diversity (Supplementary Figure 2) showed that, for comparisons between seedling populations, the values were roughly centred around zero, whereas for comparisons between the adult population and all seedling populations, the values were skewed toward strongly positive differences (i.e. the adults were largely more diverse than the seedlings).

4 Discussion

To take advantage of high-throughput, multi-locus genetic data, we have developed the RaBoT method, which takes into account diversity differences at large numbers of biallelic loci simultaneously. The RaBoT method showed average conservativeness (low type I error rate) and high power (low type II error rates) for comparisons of populations with large differences in diversity. Power was conversely low in comparisons of populations with similar levels of diversity, as one would expect, and false positive results were quite frequent for comparisons between populations with extremely high diversity levels ($\theta=2$). Nevertheless, such levels of diversity seem quite extreme for AFLP or SNP variation: mutation rates underlying them are likely to be the same as point mutation rates, which are close to $\mu=10^{-8}$ in eukaryotes (Drake et al. 1998), and therefore $\theta=N_e\mu=2$ requires very large and unlikely effective population sizes. Overall, a significant RaBoT result should therefore be considered as suggestive of large differences in diversity. The method proved globally robust to differences in population size, as false positives and false negatives were not particularly associated to the comparisons involving the smallest or the largest populations (Fig. 1).

The application of RaBoT to the study of *V. michelii* populations illustrates how the method can be used to analyse differences in diversity levels and how comparisons between different subsets can be used to gauge the validity and significance of the results. The RaBoT analyses showed that genome-wide levels of diversity were higher in the disturbed plot than in the control plot, after controlling for population size. This suggests at first sight that regeneration processes occurring after disturbance, and in particular after the opening of canopy gaps, may influence the genetic

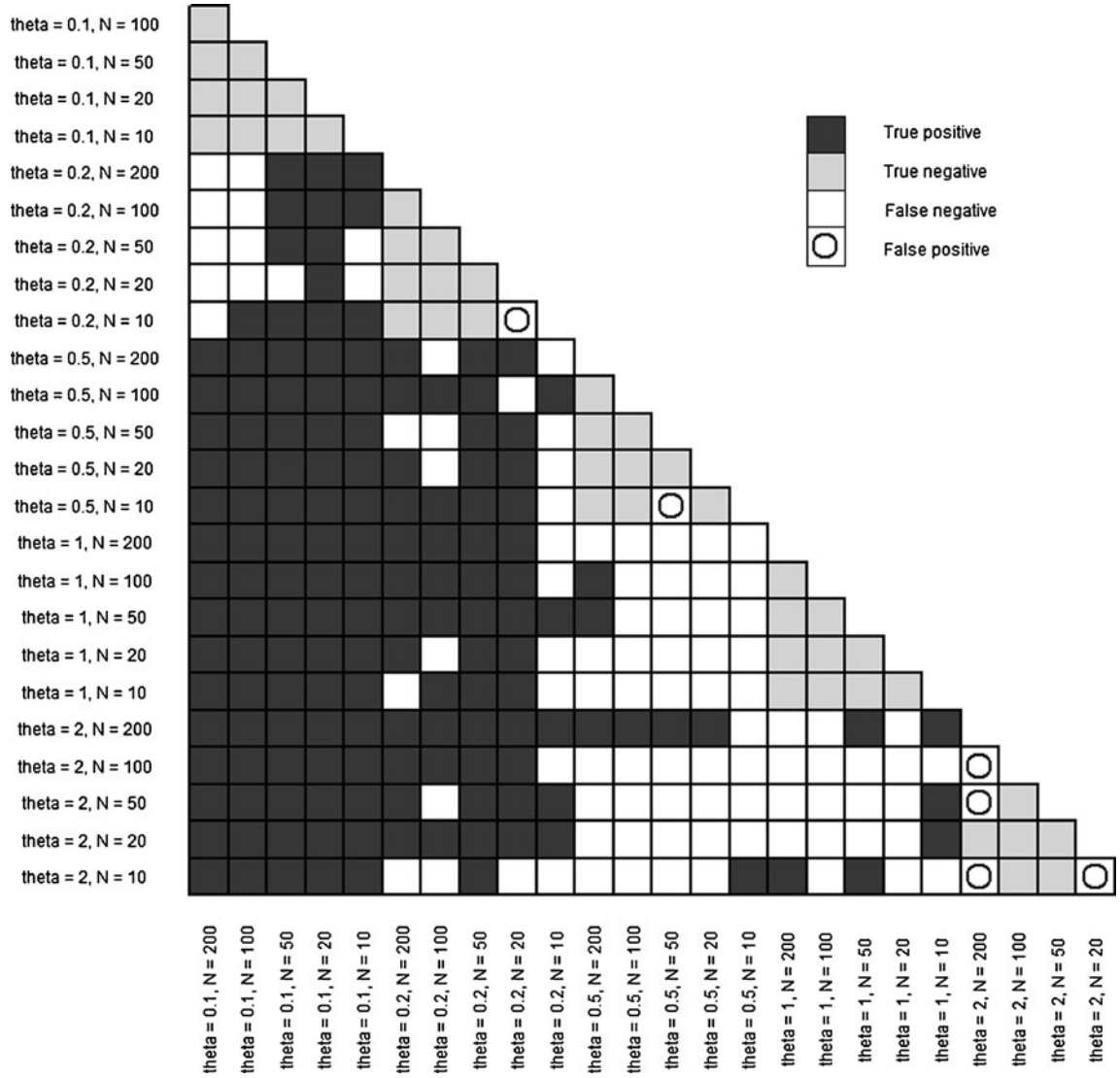

Fig. 1 Graphical representation of the results of the RaBoT comparisons of diversity for simulated populations with varying population sizes and diversity levels. *Dark gray cells*, correct positive results (significant tests for populations with different diversity levels); *light gray cells*, correct negative results (non-significant tests for populations with same diversity levels); *white cells*, false negatives; *circled cells*, false positives. Significance level=5 %

diversity of populations of a light-responsive tree species at early life stages. Nevertheless, differences in diversity levels were found also between subplots within the disturbed and the undisturbed plot, thus preventing any firm conclusion on the effect of disturbance on diversity levels, which may stochastically vary at small spatial scales. At the same time, the corresponding extant adult population had a higher diversity than juveniles, either from the control or the post-disturbance plot. This implies that, if the next adult generation is a random representation of the current juvenile cohorts, but with the same population size as the extant adults, it may be less diverse than the current adult population. The reason why our results allow us to draw such a conclusion rests on the way the method is constructed: RaBoT compares diversity in the adult population and in random subsamples of the current juveniles, having the same sample size as the adult population size; if the current adult population were to be replaced, at constant size, by a random sample of the current juveniles, then the future adult population would be equivalent to one of the random draws obtained by RaBoT in the juvenile population. Therefore, the distribution of diversity values in juvenile subsamples represents the expected distribution of the possible diversity values in the future adult population. The most likely explanation for the observed differences in diversity between adults and juveniles is that the cohort of seedlings, growing at any given time in a given (small) portion of the forest, represents only a subset of the population of seeds that

makes up the adult population. Rare long-distance dispersal events, occurring over long time spans, may have contributed disproportionately to the genetic diversity of the adult population. Moreover, only a subset of the extant adult population may have so far contributed to regeneration, which would therefore contain only a small subset of the adults' diversity. In other words, the adult population may be the outcome of the establishment of multiple cohorts of seedlings, which may have cumulatively provided higher levels of diversity than the currently observed seedling cohort.

It has to be noted that we have applied RaBoT to AFLP "phenotypes" (presence vs absence of fragments) but that a finer analysis could be performed on alleles, and therefore on Nei's (1978) genetic diversity; RaBoT is actually suitable as-is without modification for the comparison of genetic diversity in any data set comprising multi-locus, biallelic markers (e.g. SNPs) because it is a general method to test differences in true diversity of order 2 (Jost 2006) of which Simpson's diversity (applied here) (Simpson 1949) and Nei's genetic diversity (Nei 1978) are mathematically identical particular cases. Moreover, the method can easily be extended to any other measure of diversity, as the rarefaction-by-bootstrap strategy and the chi-square test of diversity differences are general methods to obtain population subsamples and to compare whole genomes, respectively, and they are not linked to a particular kind of diversity estimator (although for different diversity statistics, power, false discovery rate and robustness may be different and must be checked again). As large SNP data sets, obtained through genotyping-by-(next-generation)-sequencing, become available, RaBoT may become a convenient tool to compare genome-wide diversity levels across populations, life stages and even species.

Acknowledgments The authors wish to thank Saint-Omer Cazal for plant material collection and DNA extractions. We wish to thank Eric Marcon, Caroline Scotti-Saintagne, Rosane G. Collevatti, Andrea Piotti and Pierre-Michel Forget for critically reading the manuscript.

Funding The research was funded by the "ECOFOR—Biodiversité et gestion forestière" program.

References

Balloux F (2001) EASYPOP (Version 1.7): a computer program for population genetics simulations. J Hered 92:301–302

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667–1686

Glaubitz JC, Murrell JC, Moran GF (2003a) Effects of native forest regeneration practices on genetic diversity in *Eucalyptus consideniana*. Theor Appl Genet 107:422–431

Glaubitz JC, Wu HX, Moran GF (2003b) Impacts of silviculture on genetic diversity in the native forest species *Eucalyptus sieberi*. Conserv Genet 4:275–287

Grundmann H, Hori S, Tanner G (2001) Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. J Clin Microbiol 39:4190–4192

Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52:577–586

Jin L, Baskett ML, Cavalli-Sforza LL, Zhivotovsky LA, Feldman MW, Rosenberg NA (2000) Microsatellite evolution in modern humans: a comparison of two data sets from the same populations. Ann Hum Genet 64:117–134

Jost L (2006) Entropy and diversity. Oikos 113:363–375

Kalinowski S (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv Genet 5:539–543

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12:844–855

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Simpson E (1949) Measurement of diversity. Nature 163:688

Tessier N, Bernatchez L (1999) Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). Mol Ecol 8:169–179