MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3845–3852

NEWS AND VIEWS

COMMENT

Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics!

GABRIELE GERLACH, ALEXANDER JUETERBOCK, PHILIPP KRAEMER, JANA DEPPERMANN and PETER HARMAND

Institute of Biology and Environmental Sciences, Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany

G_{ST}-values and its relatives (F_{ST}) belong to the most used parameters to define genetic differences between populations. Originally, they were developed for allozymes with very low number of alleles. Using highly polymorphic microsatellite markers it was often puzzling that G_{ST}-values were very low but statistically significant. In their papers, Jost (2008) and Hedrick (2005) explained that G_{ST}-values do not show genetic differentiation, and Jost suggested calculating D-values instead. Theoretical mathematical considerations are often difficult to follow; therefore, we chose an applied approach comparing two artificial populations with different number of alleles at equal frequencies and known genetic divergence. Our results show that even for more than one allele per population G_{ST}-values do not calculate population differentiation correctly; in contrast, D-values do reflect the genetic differentiation indicating that data based on G_{ST}-values need to be re-evaluated. In our approach, statistical evaluations remained similar. We provide information about the impact of different sample sizes on D-values in relation to number of alleles and genetic divergence.

Keywords: D, genetic differentiation, G_{ST}, population genetics

Received 8 July 2009; revision received 15 June 2010; accepted 26 June 2010

Recently, Jost (2008) argued that G_{ST} and its relatives are inappropriate measures of genetic differentiation between populations. He showed convincingly that when using highly polymorphic microsatellite markers the G_{ST}-value cannot reach its maximum value of 1. Even when populations share no alleles at all, G_{ST}-values remain low. These low G_{ST}-values can be misinterpreted as low genetic differentiation between populations leading to the false assumption of high connectivity. However, G_{ST} remains an appropriate indicator for calculating migration rates for populations that are consistent with the finite island model (discussed in Jost 2009; Ryman & Leimar 2009). To calculate 'real' genetic differences between populations, he suggested a different method using the differentiation index D and the bias-corrected estimator D_{est}. These indices are based on the effective number of alleles resulting in a more meaningful perception of differentiation (see also Heller & Siegismund 2009; Jost 2009; Ryman & Leimar 2009). The intent of our analysis is to calculate D_{est} and G_{ST_est}-values for assessing the differentiation between populations for which the 'true' divergence and number of present alleles are known. This would enable us to find out whether D_{est} or/and G_{ST_est}-values are better estimators of population differentiation for populations of different sample sizes and markers with different number of alleles.

Correspondence: Gabriele Gerlach, Fax: +49 441 798 3937;
E-mail: gabriele.gerlach@uni-oldenburg.de

Genetic diversity values

Based on artificial and randomized data sets, we calculated G_{ST}, G_{ST_est}, D, D_{est}, H_T, H_S, H_{T_est} and H_{S_est} using our newly developed package '*DEMEtics*' in R (R Development Core Team, 2009); we tested statistical significance using the null hypothesis of zero differentiation. D-values were based on equation 11 (Jost 2008):

$$D = [(H_T - H_S)/(1 - H_S)]\,[n/(n-1)].$$

The bias-corrected D_{est}-values were based on equation 12 (Jost 2008):

$$D_{est} = [(H_{T\_est} - H_{S\_est})/(1 - H_{S\_est})]\,[n/(n-1)].$$

$G_{ST} = (H_T - H_S)/H_T$ and the bias-corrected G_{ST_est}-values were calculated according to Nei & Chesser (1983):

$$G_{ST\_est} = [(H_{T\_est} - H_{S\_est})/H_{T\_est}]\,[n/(n-1)].$$

For comparison, we also calculated G_{ST_est}-values and corresponding P values using the program FSTAT 2.93 (Goudet *et al.* 1996). However, no difference was detected between values calculated by our programme and by FSTAT 2.93.

Generating test data with theoretical allele frequencies

To compare different methods to calculate genetic diversity (G_{ST} and D), we created artificial data sets with therefore known divergence. These data consisted of two populations with 40 diploid individuals each. To show the effect of increasing number of alleles (i.e. *mutational* effects), we calculated genetic diversity for populations with N_{a_total} = 2, 4, 8, 10, 12, 16, 20, 26, 32 and 40 alleles total. All alleles had the same frequencies. For each number of alleles, we started with the assumption that populations shared no alleles (100% divergence); e.g. for N_{a_total} = 8, population A had alleles 1, 2, 3 and 4 and population B had alleles 5, 6, 7 and 8 at equal frequencies. Then we decreased divergence, simulating *migrational* gene flow by exchanging 5% of alleles in

both directions: population A obtained 5% alleles from population B (four alleles: one time alleles 5, 6, 7 and 8) and population B obtained 5% alleles from population A (four alleles: one time alleles 1, 2, 3 and 4), resulting in a divergence of 90%. In terms of frequencies for the eight alleles, this means passing from the distribution (20/80, 20/80, 20/80, 20/80, 0, 0, 0, 0) and (0, 0, 0, 0, 20/80, 20/80, 20/80, 20/80) to (19/80, …, 19/80, 1/80, …, 1/80) and (1/80, …, 1/80, 19/80, …, 19/80) in the first step and to (18/80, …, 18/80, 2/80, …, 2/80) and (2/80, …, 2/80, 18/80, …, 18/80) in the second step. With this process, we decreased genetic divergence in 10% steps from 100% to 0%. Alleles that had been introduced from the other population were not exchanged in this process. At 0% divergence, populations were identical having the same number and frequency of alleles and were considered finite populations with theoretical allele frequencies.

Comparison of G_{ST} and D in finite populations with theoretical allele frequencies

We then calculated the genetic difference between these two finite populations of 40 individuals each based on G_{ST}- and D-values; results are shown in Fig. 1a,b. The maximum possible G_{ST} fell rapidly as the number of alleles per population increased (see Fig. 1a). In contrast, D-values always equalled 1 when populations shared no alleles (100% divergence) independent of the number of alleles (see Fig. 1b). G_{ST}-values only equalled 1 when population A had allele 1 and population B had allele 2 and did not share any alleles. The maximum G_{ST}-value already dropped to 0.328 when 2 different alleles occurred in each population (four alleles total) even when populations did not share any alleles (100% divergence; Fig. 1a). By using four different unique alleles in each population (eight alleles total; 100% divergence), the maximum G_{ST}-value was 0.137 (Fig. 1a). For 5–20 different alleles in each population, the maximum G_{ST}-value ranged from 0.105 to 0.019 at 100% divergence. These calculations support Jost's conclusion that G_{ST} is not an appropriate metric for genetic difference while D reflects the known divergence between populations much better. G_{ST} ranked the divergence between populations in the finite island model correctly (Fig. 1a) when the number of alleles remained the same but divergence decreased (simulating *migrational* effects). For *mutational* effects (increasing number of alleles), our results confirm Jost (2009) that the ranking is a problem when calculating G_{ST}-values. With increasing number of alleles, G_{ST}-values can decrease even when the real divergence between populations increases. At equal allele frequencies, D-values ranked genetic distances correctly. The main advantage of D over G_{ST} is that it is based on the complete partitioning of heterozygosity into independent within- and between-group components, whereas G_{ST} confounds the two components.

Limitations of D and G_{ST} based on different H_S and H_T

While in the aforementioned calculations, allele frequencies in both populations were maintained equal, we next evaluated how G_{ST}- and D-values behave in populations with different H_S and H_T combinations caused by different allele frequencies. The results are shown in Fig. 2a,b.

The triangle gives the domain of definition of G_{ST} and D as a function of H_S and H_T. Because H_S is always less than or equal to H_T, points below the triangle are not possible. Points above the triangle can only be generated for more than two populations.

Each line within the triangle shows G_{ST}- (Fig. 2a) and D-values (Fig. 2b) ranging from 0 to 1 in steps of 0.1. Diamonds reflect G_{ST}- (Fig. 2a) and D-values (Fig. 2b) for a given number of alleles (N_{a_total}) and decreasing divergence between populations from the left ($D = 1$, divergence = 100%) to the right side ($D = 0$, divergence = 0%). Figure 2a shows that with increasing allele numbers the possible range of G_{ST}-values becomes rapidly smaller. With 16 alleles, G_{ST}-values were constrained to < 0.1, even when the two populations were completely different. $G_{ST} = 1$ can only be reached when $H_S = 0$ which is realized when each of both populations bears a single allele. In contrast, D-values continued to range from 0 to 1 independent of the number of alleles present. In this respect, D-values provide a clearer metric for population differentiation than G_{ST}-values.

Different H_S and H_T values can result in the same D-values as well as in the same G_{ST}-values. Table 1 gives an example for populations with two alleles (columns 1–2) and also three alleles (columns 3–5) at different frequencies. The two populations in the first column are more differentiated than those in the second column; and populations in column 3 are more differentiated than those

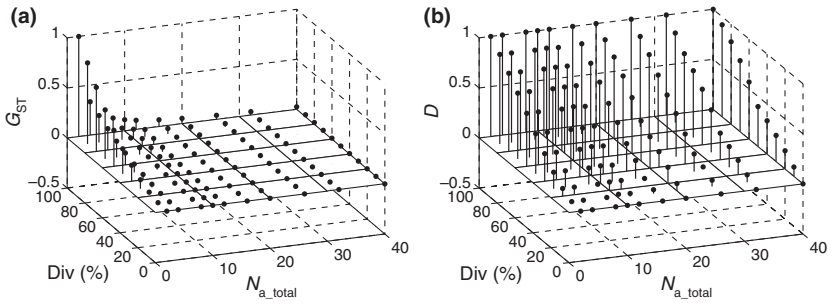

Fig. 1 (a,b) Genetic differences between two populations analysed by (a) G_{ST}- and (b) D-values. N_{a_total} = total number of alleles in two populations. Div = divergence [%]. For further explanation see text.

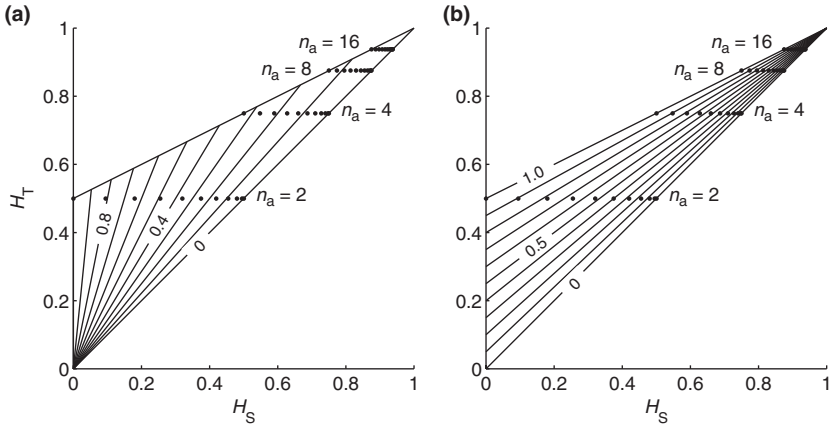

Fig. 2 (a,b) The triangle gives the domain of definition of G_{ST} and D as a function of H_S and H_T. Each line within the triangle shows (a) G_{ST}- and (b) D-values ranging from 0 to 1 in steps of 0.1. Diamonds reflect G_{ST}- and D-values for a given number of alleles (N_{a_total}) and decreasing divergence between populations from the left (H_S = 0; 100% divergence) to the right side (H_S = 1; 0% divergence).

Table 1 D- and G_{ST}-values for populations with alleles with different frequencies (F_a)

Allele	F_a [%]	F_a [%]	F_a [%]	F_a [%]	F_a [%]
Population 1					
1	62.9	75.4	10.0	15.0	30.0
2	37.1	24.6	4.4	9.2	7.8
3			85.6	75.8	62.2
Population 2					
1	1.6	17.5	10.0	25.0	33.0
2	98.4	82.5	59.2	53.5	55.2
3			30.8	21.5	14.8
D	0.5	0.5	0.5	0.5	0.5
G_{ST}	0.43	0.34	0.27	0.20	0.17

shown in column 4 and 5; however, all give the same value D = 0.5. In all these cases, the differences were reflected in different G_{ST}-values. This observation is an example of ranking problems with D. Similarly, one can construct examples with different allele frequencies giving the same G_{ST}-value but different D-values. These examples elucidate that there is not one encompassing metric to describe genetic differences between populations. However, the advantage of the D-metric is that it reflects genetic divergence between populations better than G_{ST}, because the latter is strongly limited in range with increasing allele frequency.

Comparison of G_{ST_est}- and D_{est}-values in randomly sampled populations

Effect of number of alleles and sample size

We evaluated the bias correction in G_{ST_est}- and D_{est}-values based on calculations for randomly sampled populations. We used the artificial data set described above and drew 30 random subsamples of alleles (with replacement) from each of these two populations. We then calculated the mean G_{ST}-, G_{ST_est}-, D- and D_{est}-values. These randomly generated subsamples consisted either of 10, 20, 40 or 60 individuals per population. This approach should simulate the field situation of sampling different numbers of individuals from two populations and calculating genetic difference between them. Different from the natural situation, in our artificial samples, the genetic divergence and number of alleles in the original populations were known. Therefore, we could compare how accurately the calculated G_{ST_est}- and D_{est}-values responded to a sampling regime of populations with a different number of alleles as well as to differences in sample size. It is also important to notice the standard deviation of G_{ST_est}- and D_{est}-values correlated to sample size. By calculating D_{est}-values according to equation 12 (Jost 2008), the bias resulting from differences in allele frequencies generated by random sampling should be corrected. The same calculations were performed for G_{ST} and the bias-corrected G_{ST_est}-values (see Table 2).

For both, D_{est} and G_{ST_est}, we calculated P values and confidence intervals to find out whether they differed significantly from a random distribution based on bootstrapping (500 iterations).

Considering P values in almost all cases, D_{est}-values were statistically significant when G_{ST_est}-values were too. We can conclude that G_{ST_est}-values do not reflect genetic differentiation between populations when there are more than two alleles present, but at least the calculated statistical significance remains meaningful. Therefore, old data sets need to be reanalysed to find out how different populations were, but it is probably that the conclusions based on statistical significance are the same.

By using the G_{ST_est} method, a great error seemed to have been to ignore statistically significant G_{ST_est}-values simply because the values themselves were so low. For instance, G_{ST_est} was very low (0.005, Table 2, line 142) when 60 individuals were sampled per population with a total number of alleles n = 32 and a 'real' divergence of 40%, but this small number was still significantly different from the bootstrap distribution (P = 0.009). The corresponding D_{est}-value showed a similar level of significance but a much larger value to reflect the difference between populations (D_{est} = 0.268, P = 0.008, Table 2). D_{est}-values are clearly a better indicator for population structure than G_{ST_est}.

Table 2 Mean values of genetic difference and heterozygosity between two different populations based on 30 random samples drawn from populations with alleles of equal frequency; div: genetic divergence between finite populations (see text for detailed information); sample size: number of individuals sampled per population; N_a total number of alleles in both populations; H_{S_est} mean heterozygosity in each population with bias correction, H_{T_est} mean total heterozygosity with bias correction, G_{ST} biased mean G_{ST}-value, G_{ST_est} bias-corrected mean value; SD standard deviation of G_{ST_est}; CI confidence interval of G_{ST_est} (percentile method); P statistical significance based on bootstraps with 500 repeats; D mean D-value according to equation 11, Jost (2008); D_{est} bias-corrected mean D_{est}-values (equation 12 Jost 2008); SD standard deviation of D_{est}-values; CI confidence interval (percentile method); P statistical significance of D_{est}-values based on bootstraps with 500 repeats (bold values indicate statistical significance)

Line	Div [%]	Sample size	N_a	H_{S_est}	H_{T_est}	G_{ST}	G_{ST_est}	SD	P	(−CI) perc	(+CI) perc	D	D_{est}	SD	P	(−CI) perc	(+CI) perc
1	100	10	2	0	0.5	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
2	80	10	2	0.187	0.499	0.636	**0.621**	0.17	**0.002**	**0.354**	**0.870**	0.760	**0.756**	0.12	**0.002**	**0.499**	**0.949**
3	60	10	2	0.29	0.496	0.432	**0.412**	0.19	**0.002**	**0.162**	**0.708**	0.573	**0.562**	0.17	**0.002**	**0.256**	**0.829**
4	40	10	2	0.397	0.492	0.213	**0.188**	0.12	**0.027**	**0.002**	**0.464**	0.322	**0.297**	0.15	**0.025**	**0.023**	**0.618**
5	20	10	2	0.481	0.497	0.054	0.029	0.06	0.421	−0.047	0.205	0.094	0.050	0.11	0.423	−0.086	0.317
6	0	10	2	0.486	0.495	0.041	0.015	0.06	0.443	−0.048	0.185	0.071	0.025	0.10	0.452	−0.086	0.283
7	100	20	2	0	0.5	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
8	80	20	2	0.182	0.499	0.64	**0.633**	0.12	**0.002**	**0.432**	**0.824**	0.770	**0.769**	0.09	**0.002**	**0.595**	**0.908**
9	60	20	2	0.341	0.498	0.325	**0.313**	0.09	**0.002**	**0.130**	**0.529**	0.478	**0.471**	0.10	**0.002**	**0.242**	**0.691**
10	40	20	2	0.425	0.492	0.147	**0.135**	0.07	**0.014**	**0.015**	**0.309**	0.243	**0.228**	0.11	**0.012**	**0.038**	**0.461**
11	20	20	2	0.477	0.496	0.05	0.037	0.03	0.124	−0.022	0.155	0.091	0.070	0.06	0.117	−0.035	0.266
12	0	20	2	0.498	0.499	0.013	0.001	0.02	0.486	−0.025	0.079	0.026	0.001	0.03	0.474	−0.049	0.143
13	100	40	2	0	0.5	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
14	80	40	2	0.19	0.499	0.622	**0.619**	0.08	**0.002**	**0.473**	**0.763**	0.761	**0.760**	0.07	**0.002**	**0.637**	**0.866**
15	60	40	2	0.319	0.498	0.364	**0.358**	0.07	**0.002**	**0.218**	**0.511**	0.525	**0.522**	0.08	**0.002**	**0.358**	**0.674**
16	40	40	2	0.433	0.497	0.135	**0.129**	0.07	**0.002**	**0.041**	**0.245**	0.229	**0.222**	0.10	**0.003**	**0.083**	**0.384**
17	20	40	2	0.48	0.499	0.042	0.036	0.04	0.125	−0.004	0.108	0.079	0.068	0.07	0.121	−0.006	0.190
18	0	40	2	0.497	0.498	0.006	0.000	0.00	0.416	−0.013	0.041	0.012	0.000	0.01	0.422	−0.024	0.074
19	100	60	2	0	0.5	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
20	80	60	2	0.181	0.5	0.639	**0.637**	0.07	**0.002**	**0.517**	**0.754**	0.776	**0.776**	0.06	**0.002**	**0.678**	**0.862**
21	60	60	2	0.322	0.499	0.358	**0.355**	0.07	**0.002**	**0.240**	**0.481**	0.522	**0.520**	0.08	**0.002**	**0.390**	**0.647**
22	40	60	2	0.425	0.499	0.151	**0.147**	0.05	**0.002**	**0.067**	**0.243**	0.257	**0.252**	0.07	**0.002**	**0.129**	**0.387**
23	20	60	2	0.474	0.499	0.053	**0.049**	0.04	**0.017**	**0.007**	**0.114**	0.099	**0.092**	0.06	**0.017**	**0.015**	**0.198**
24	0	60	2	0.499	0.498	0.003	−0.001	0.00	0.561	−0.008	0.022	0.006	−0.002	0.01	0.555	−0.017	0.044
25	100	10	4	0.498	0.746	0.351	**0.328**	0.02	**0.002**	**0.289**	**0.419**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
26	80	10	4	0.584	0.738	0.229	**0.205**	0.06	**0.002**	**0.101**	**0.326**	0.757	**0.737**	0.15	**0.002**	**0.474**	**0.934**
27	60	10	4	0.654	0.738	0.134	**0.109**	0.05	**0.018**	**0.017**	**0.231**	0.520	**0.468**	0.18	**0.017**	**0.151**	**0.766**
28	40	10	4	0.699	0.742	0.08	0.054	0.05	0.142	−0.022	0.166	0.339	0.251	0.21	0.136	−0.057	0.586
29	20	10	4	0.73	0.74	0.037	0.011	0.03	0.365	−0.041	0.103	0.175	0.053	0.14	0.376	−0.197	0.401
30	0	10	4	0.734	0.736	0.026	0.000	0.02	0.521	−0.043	0.083	0.123	−0.009	0.13	0.524	−0.227	0.328
31	100	20	4	0.494	0.746	0.346	**0.335**	0.01	**0.002**	**0.312**	**0.385**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
32	80	20	4	0.589	0.746	0.22	**0.208**	0.03	**0.002**	**0.133**	**0.289**	0.775	**0.766**	0.08	**0.002**	**0.582**	**0.910**
33	60	20	4	0.65	0.748	0.141	**0.129**	0.05	**0.002**	**0.062**	**0.208**	0.570	**0.548**	0.15	**0.002**	**0.327**	**0.750**
34	40	20	4	0.702	0.745	0.07	**0.057**	0.03	**0.012**	**0.004**	**0.126**	0.326	**0.284**	0.12	**0.010**	**0.057**	**0.527**
35	20	20	4	0.73	0.746	0.033	0.020	0.02	0.155	−0.016	0.077	0.168	0.109	0.09	0.124	−0.071	0.350
36	0	20	4	0.746	0.746	0.011	−0.001	0.01	0.539	−0.022	0.039	0.062	−0.009	0.06	0.536	−0.126	0.188
37	100	40	4	0.5	0.749	0.337	**0.331**	0.00	**0.002**	**0.322**	**0.353**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
38	80	40	4	0.59	0.749	0.217	**0.211**	0.03	**0.002**	**0.161**	**0.262**	0.778	**0.774**	0.07	**0.002**	**0.655**	**0.876**
39	60	40	4	0.658	0.746	0.124	**0.118**	0.02	**0.002**	**0.070**	**0.171**	0.528	**0.517**	0.06	**0.002**	**0.351**	**0.670**
40	40	40	4	0.711	0.747	0.054	**0.005**	0.02	**0.003**	**0.014**	**0.092**	0.271	**0.248**	0.08	**0.004**	**0.090**	**0.422**
41	20	40	4	0.738	0.747	0.018	0.012	0.01	0.149	−0.007	0.041	0.100	0.067	0.06	0.145	−0.035	0.212
42	0	40	4	0.747	0.747	0.005	−0.001	0.00	0.519	−0.011	0.019	0.029	−0.007	0.02	0.531	−0.066	0.102
43	100	60	4	0.497	0.748	0.339	**0.335**	0.01	**0.002**	**0.327**	**0.352**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
44	80	60	4	0.59	0.749	0.215	**0.211**	0.03	**0.002**	**0.170**	**0.253**	0.776	**0.773**	0.06	**0.002**	**0.677**	**0.857**
45	60	60	4	0.657	0.749	0.126	**0.122**	0.02	**0.002**	**0.083**	**0.166**	0.542	**0.535**	0.07	**0.002**	**0.402**	**0.662**
46	40	60	4	0.707	0.748	0.058	**0.054**	0.02	**0.002**	**0.025**	**0.089**	0.289	**0.275**	0.07	**0.002**	**0.143**	**0.417**
47	20	60	4	0.739	0.748	0.016	0.011	0.01	0.070	−0.003	0.033	0.087	0.066	0.04	0.066	−0.016	0.175
48	0	60	4	0.747	0.748	0.005	0.001	0.00	0.422	−0.007	0.016	0.030	0.006	0.02	0.421	−0.041	0.087
49	100	10	8	0.74	0.866	0.166	**0.141**	0.02	**0.002**	**0.107**	**0.196**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**

Table 2 (Continued)

Line	Div [%]	Sample size	N_a	H_{S_est}	H_{T_est}	G_{ST}	G_{ST_est}	SD	P	(−CI) perc	(+CI) perc	D	D_{est}	SD	P	(−CI) perc	(+CI) perc
50	80	10	8	0.777	0.86	0.118	**0.093**	0.03	**0.004**	**0.038**	**0.163**	0.798	**1.000**	0.13	**0.003**	**0.501**	**0.936**
51	60	10	8	0.812	0.863	0.081	**0.055**	0.03	**0.041**	**0.004**	**0.124**	0.623	**0.525**	0.22	**0.036**	**0.187**	**0.800**
52	40	10	8	0.841	0.866	0.051	0.025	0.03	0.173	−0.019	0.087	0.444	0.266	0.23	0.156	−0.101	0.611
53	20	10	8	0.856	0.862	0.029	0.004	0.02	0.421	−0.033	0.060	0.284	0.031	0.19	0.429	−0.326	0.417
54	0	10	8	0.862	0.865	0.026	0.001	0.01	0.480	−0.033	0.053	0.263	−0.008	0.20	0.491	−0.358	0.376
55	100	20	8	0.745	0.871	0.154	**0.142**	0.01	**0.002**	**0.124**	**0.170**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
56	80	20	8	0.787	0.87	0.106	**0.093**	0.02	**0.002**	**0.060**	**0.131**	0.808	**0.788**	0.07	**0.002**	**0.617**	**0.921**
57	60	20	8	0.826	0.869	0.06	**0.048**	0.02	**0.003**	**0.016**	**0.086**	0.549	**0.491**	0.13	**0.003**	**0.253**	**0.708**
58	40	20	8	0.842	0.866	0.039	**0.026**	0.01	**0.028**	−0.002	0.063	0.380	**0.291**	0.11	**0.020**	**0.037**	**0.551**
59	20	20	8	0.862	0.871	0.02	0.008	0.01	0.272	−0.013	0.037	0.229	0.098	0.13	0.249	−0.131	0.358
60	0	20	8	0.865	0.867	0.014	0.001	0.01	0.422	−0.017	0.028	0.156	0.010	0.08	0.434	−0.197	0.268
61	100	40	8	0.749	0.873	0.148	**0.141**	0.00	**0.002**	**0.133**	**0.155**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
62	80	40	8	0.792	0.874	0.099	**0.093**	0.01	**0.002**	**0.071**	**0.116**	0.802	**0.793**	0.04	**0.002**	**0.675**	**0.892**
63	60	40	8	0.827	0.872	0.057	**0.051**	0.01	**0.002**	**0.029**	**0.076**	0.553	**0.525**	0.06	**0.002**	**0.357**	**0.680**
64	40	40	8	0.849	0.871	0.031	**0.025**	0.01	**0.004**	**0.007**	**0.047**	0.335	**0.287**	0.09	**0.004**	**0.113**	**0.468**
65	20	40	8	0.867	0.873	0.012	0.006	0.01	0.211	−0.005	0.022	0.150	0.078	0.08	0.207	−0.057	0.246
66	0	40	8	0.873	0.873	0.005	−0.001	0.00	0.598	−0.009	0.011	0.067	−0.016	0.04	0.603	−0.122	0.130
67	100	60	8	0.747	0.873	0.147	**0.143**	0.00	**0.002**	0.137	0.153	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
68	80	60	8	0.792	0.873	0.096	**0.092**	0.01	**0.002**	0.074	0.111	0.789	**0.782**	0.04	**0.002**	**0.686**	**0.866**
69	60	60	8	0.829	0.873	0.054	**0.050**	0.01	**0.002**	0.032	0.069	0.536	**0.517**	0.07	**0.002**	**0.382**	**0.645**
70	40	60	8	0.852	0.873	0.028	**0.024**	0.01	**0.002**	0.010	0.040	0.315	**0.282**	0.06	**0.002**	**0.144**	**0.426**
71	20	60	8	0.866	0.873	0.011	**0.007**	0.00	**0.045**	−0.002	0.019	0.142	**0.094**	0.06	**0.041**	−0.014	0.222
72	0	60	8	0.873	0.873	0.004	0.000	0.00	0.531	−0.006	0.009	0.054	−0.002	0.04	0.540	−0.076	0.104
73	100	10	12	0.816	0.903	0.118	**0.093**	0.01	**0.002**	**0.063**	**0.141**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
74	80	10	12	0.848	0.903	0.084	**0.058**	0.01	**0.006**	**0.021**	**0.107**	0.822	**0.764**	0.11	**0.003**	**0.503**	**0.937**
75	60	10	12	0.867	0.905	0.064	0.038	0.02	0.056	**0.001**	**0.089**	0.686	**0.550**	0.23	**0.037**	**0.212**	**0.803**
76	40	10	12	0.889	0.904	0.039	0.014	0.01	0.203	−0.020	0.060	0.490	0.233	0.21	0.181	−0.159	0.580
77	20	10	12	0.895	0.903	0.031	0.006	0.02	0.406	−0.026	0.053	0.400	0.063	0.29	0.414	−0.339	0.438
78	0	10	12	0.897	0.906	0.032	0.006	0.01	0.359	−0.025	0.052	0.419	0.099	0.26	0.360	−0.306	0.477
79	100	20	12	0.825	0.91	0.104	**0.092**	0.01	**0.002**	**0.076**	**0.115**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
80	80	20	12	0.855	0.91	0.072	**0.059**	0.01	**0.002**	**0.036**	**0.087**	0.813	**0.783**	0.08	**0.002**	**0.611**	**0.915**
81	60	20	12	0.88	0.912	0.047	**0.034**	0.01	**0.026**	**0.011**	**0.062**	0.618	**0.537**	0.19	**0.023**	**0.296**	**0.745**
82	40	20	12	0.896	0.909	0.025	0.012	0.01	0.198	−0.006	0.037	0.366	0.209	0.22	0.204	−0.046	0.468
83	20	20	12	0.902	0.91	0.02	0.008	0.01	0.259	−0.010	0.032	0.313	0.134	0.17	0.228	−0.130	0.411
84	0	20	12	0.912	0.912	0.011	−0.002	0.00	0.579	−0.016	0.018	0.191	−0.041	0.09	0.615	−0.283	0.235
85	100	40	12	0.828	0.913	0.098	**0.092**	0.00	**0.002**	**0.084**	**0.103**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
86	80	40	12	0.86	0.914	0.064	**0.058**	0.01	**0.002**	**0.042**	**0.074**	0.793	**0.777**	0.05	**0.002**	**0.654**	**0.880**
87	60	40	12	0.884	0.914	0.037	**0.031**	0.01	**0.002**	**0.017**	**0.047**	0.549	**0.504**	0.09	**0.002**	**0.328**	**0.662**
88	40	40	12	0.898	0.913	0.023	**0.016**	0.01	**0.007**	**0.004**	**0.032**	0.371	**0.298**	0.10	**0.006**	**0.121**	**0.481**
89	20	40	12	0.909	0.913	0.01	0.004	0.00	0.216	−0.005	0.015	0.179	0.071	0.08	0.218	−0.086	0.252
90	0	40	12	0.912	0.913	0.006	0.000	0.00	0.535	−0.007	0.010	0.118	−0.003	0.07	0.543	−0.140	0.168
91	100	60	12	0.83	0.914	0.096	**0.091**	0.00	**0.002**	**0.086**	**0.099**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
92	80	60	12	0.862	0.914	0.061	**0.057**	0.01	**0.002**	**0.045**	**0.070**	0.780	**0.768**	0.05	**0.002**	**0.669**	**0.857**
93	60	60	12	0.883	0.914	0.037	**0.033**	0.01	**0.002**	**0.021**	**0.046**	0.553	**0.524**	0.08	**0.002**	**0.383**	**0.654**
94	40	60	12	0.899	0.914	0.02	**0.016**	0.01	**0.002**	**0.006**	**0.027**	0.341	**0.290**	0.08	**0.003**	**0.148**	**0.438**
95	20	60	12	0.909	0.914	0.009	0.005	0.00	0.077	−0.002	0.013	0.165	0.093	0.06	0.075	−0.029	0.236
96	0	60	12	0.913	0.914	0.004	0.000	0.00	0.539	−0.005	0.007	0.080	−0.004	0.05	0.544	−0.100	0.118
97	100	10	20	0.882	0.936	0.08	**0.054**	0.01	**0.003**	**0.030**	**0.092**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
98	80	10	20	0.905	0.939	0.058	**0.033**	0.01	**0.020**	**0.005**	**0.070**	0.860	**0.772**	0.13	**0.005**	**0.519**	**0.915**
99	60	10	20	0.918	0.937	0.043	0.017	0.01	0.111	−0.009	0.053	0.697	0.468	0.21	0.069	**0.108**	**0.727**
100	40	10	20	0.921	0.936	0.038	0.013	0.01	0.209	−0.014	0.049	0.634	0.320	0.31	0.178	−0.067	0.618
101	20	10	20	0.93	0.937	0.03	0.004	0.01	0.377	−0.021	0.039	0.542	0.112	0.31	0.364	−0.299	0.452
102	0	10	20	0.932	0.936	0.027	0.001	0.01	0.440	−0.024	0.036	0.494	−0.013	0.31	0.464	−0.453	0.357
103	100	20	20	0.894	0.945	0.064	**0.052**	0.00	**0.002**	**0.040**	**0.068**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
104	80	20	20	0.908	0.943	0.049	**0.036**	0.01	**0.002**	**0.021**	**0.055**	0.850	**0.807**	0.08	**0.002**	**0.633**	**0.924**
105	60	20	20	0.924	0.944	0.032	**0.019**	0.01	**0.008**	**0.004**	**0.039**	0.653	**0.531**	0.11	**0.004**	**0.277**	**0.738**

Table 2 (Continued)

Line	Div [%]	Sample size	N_a	H_{S_est}	H_{T_est}	G_{ST}	G_{ST_est}	SD	P	(−CI) perc	(+CI) perc	D	D_{est}	SD	P	(−CI) perc	(+CI) perc
106	40	20	20	0.933	0.944	0.024	0.011	0.01	0.090	−0.003	0.029	0.528	0.334	0.17	0.078	**0.058**	**0.584**
107	20	20	20	0.936	0.943	0.019	0.006	0.00	0.188	−0.008	0.024	0.431	0.187	0.15	0.169	−0.102	0.458
108	0	20	20	0.944	0.943	0.011	−0.002	0.00	0.585	−0.013	0.013	0.282	−0.075	0.14	0.609	−0.361	0.214
109	100	40	20	0.897	0.947	0.058	**0.052**	0.00	**0.002**	**0.046**	**0.060**	1.000	**1.000**	0.05	**0.002**	**1.000**	**1.000**
110	80	40	20	0.913	0.947	0.042	**0.035**	0.00	**0.002**	**0.026**	**0.046**	0.827	**0.803**	0.08	**0.002**	**0.683**	**0.897**
111	60	40	20	0.926	0.946	0.028	**0.021**	0.00	**0.002**	**0.012**	**0.032**	0.631	**0.570**	0.09	**0.002**	**0.409**	**0.717**
112	40	40	20	0.938	0.947	0.015	**0.009**	0.00	**0.009**	**0.001**	**0.019**	0.411	**0.290**	0.09	**0.006**	**0.096**	**0.480**
113	20	40	20	0.941	0.947	0.011	0.005	0.00	0.066	−0.002	0.015	0.322	0.175	0.05	0.056	−0.020	0.373
114	0	40	20	0.945	0.945	0.006	−0.001	0.00	0.583	−0.006	0.007	0.171	−0.023	0.06	0.590	−0.194	0.170
115	100	60	20	0.897	0.947	0.057	**0.053**	0.00	**0.002**	**0.049**	**0.059**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
116	80	60	20	0.913	0.948	0.04	**0.036**	0.00	**0.002**	**0.028**	**0.044**	0.823	**0.807**	0.05	**0.002**	**0.711**	**0.883**
117	60	60	20	0.928	0.948	0.024	**0.020**	0.00	**0.002**	**0.013**	**0.028**	0.593	**0.547**	0.08	**0.002**	**0.410**	**0.673**
118	40	60	20	0.938	0.948	0.014	**0.010**	0.00	**0.003**	**0.004**	**0.017**	0.400	**0.321**	0.09	**0.002**	**0.170**	**0.469**
119	20	60	20	0.943	0.948	0.009	**0.005**	0.00	**0.035**	−0.001	0.011	0.267	**0.161**	0.08	**0.031**	**0.013**	**0.318**
120	0	60	20	0.947	0.948	0.004	0.000	0.00	0.540	−0.004	0.005	0.127	−0.012	0.06	0.545	−0.135	0.134
121	100	10	32	0.918	0.954	0.06	**0.034**	0.01	**0.008**	**0.014**	**0.065**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
122	80	10	32	0.932	0.954	0.046	0.020	0.01	0.062	−0.002	0.051	0.865	**0.725**	0.17	**0.020**	**0.446**	**0.865**
123	60	10	32	0.939	0.953	0.037	0.012	0.01	0.167	−0.010	0.043	0.758	0.462	0.31	0.125	**0.116**	**0.691**
124	40	10	32	0.946	0.954	0.031	0.005	0.01	0.308	−0.016	0.034	0.677	0.195	0.45	0.277	−0.203	0.484
125	20	10	32	0.949	0.954	0.028	0.002	0.01	0.413	−0.019	0.031	0.627	0.003	0.43	0.426	−0.415	0.323
126	0	10	32	0.952	0.955	0.025	0.000	0.01	0.492	−0.021	0.028	0.597	−0.181	0.56	0.526	−0.624	0.159
127	100	20	32	0.929	0.962	0.045	**0.033**	0.00	**0.002**	**0.022**	**0.047**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
128	80	20	32	0.939	0.961	0.033	**0.021**	0.00	**0.004**	**0.009**	**0.036**	0.832	**0.755**	0.09	**0.002**	**0.563**	**0.888**
129	60	20	32	0.945	0.96	0.026	**0.014**	0.01	**0.018**	**0.001**	**0.030**	0.705	**0.547**	0.12	**0.007**	**0.296**	**0.742**
130	40	20	32	0.955	0.961	0.017	0.005	0.00	0.194	−0.006	0.019	0.541	0.228	0.18	0.165	−0.076	0.492
131	20	20	32	0.958	0.961	0.014	0.002	0.00	0.322	−0.009	0.016	0.475	0.090	0.16	0.310	−0.215	0.370
132	0	20	32	0.96	0.961	0.013	0.000	0.00	0.446	−0.010	0.014	0.438	0.002	0.22	0.437	−0.314	0.293
133	100	40	32	0.932	0.965	0.039	**0.033**	0.00	**0.002**	**0.028**	**0.039**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
134	80	40	32	0.943	0.964	0.027	**0.020**	0.00	**0.002**	**0.013**	**0.028**	0.794	**0.747**	0.07	**0.002**	**0.608**	**0.858**
135	60	40	32	0.951	0.964	0.019	**0.013**	0.00	**0.002**	**0.006**	**0.021**	0.637	**0.540**	0.08	**0.002**	**0.360**	**0.695**
136	40	40	32	0.959	0.964	0.011	0.005	0.00	0.059	−0.001	0.012	0.422	**0.230**	0.10	**0.042**	**0.031**	**0.428**
137	20	40	32	0.961	0.964	0.009	0.003	0.00	0.126	−0.003	0.010	0.369	0.151	0.09	0.106	−0.055	0.357
138	0	40	32	0.963	0.964	0.007	0.001	0.00	0.442	−0.005	0.007	0.285	0.022	0.14	0.444	−0.176	0.231
139	100	60	32	0.932	0.965	0.038	**0.034**	0.00	**0.002**	**0.030**	**0.038**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
140	80	60	32	0.945	0.965	0.024	**0.020**	0.00	**0.002**	**0.015**	**0.026**	0.782	**0.749**	0.06	**0.002**	**0.643**	**0.842**
141	60	60	32	0.952	0.965	0.017	**0.013**	0.00	**0.002**	**0.008**	**0.019**	0.613	**0.544**	0.09	**0.002**	**0.404**	**0.672**
142	40	60	32	0.96	0.965	0.009	**0.005**	0.00	**0.009**	**0.001**	**0.010**	0.397	**0.268**	0.09	**0.008**	**0.112**	**0.425**
143	20	60	32	0.963	0.965	0.006	0.002	0.00	0.126	−0.002	0.007	0.282	0.114	0.07	0.116	−0.046	0.280
144	0	60	32	0.965	0.966	0.005	0.000	0.00	0.371	−0.003	0.005	0.219	0.026	0.06	0.369	−0.127	0.192

With increasing number of alleles, the standard deviation of D_{est}-values became very large for small sample sizes. When original populations were identical (0% divergence) and consisted of 32 different alleles total, D_{est}-values ranged widely and could reach 0.49 when only 10 individuals were sub-sampled per population (Table 2, line 126: mean $D_{est} = -0.181 \pm 0.56$ SD, single values ranged from −1.47 to 0.49). This result illustrates clearly how large errors can occur when sample sizes are too small in relation to a large number of alleles. Even for a sample size of 20 individuals per population (with 20 alleles, and 40% divergence between populations), mean $D_{est} = 0.228 \pm 0.178$ SD (Table 2, line 130), but single D_{est}-values ranged from −0.267 to 0.54. Here, the calculation of statistical significance is necessary to show whether D_{est}-values are significantly different from a random distribution. Independent of the method for calculating genetic differences, our example shows that such a sampling regime cannot produce reliable results. We suggest taking negative D_{est}-values as a potential warning signal that sample sizes might be too small compared to the number of alleles. So far it was also suggested to replace negative values as zero [see description of the computer program SMOGD which calculates D_{est}-values and CI (Crawford 2010)]. G_{ST_est} values could also become negative, but because the values of G_{ST_est} themselves will be much smaller, the standard deviation of G_{ST_est} will be much smaller than those of D_{est}-values.

The question remains whether P values or confidence intervals should be used to decide about statistically significant differences between populations. Statistically

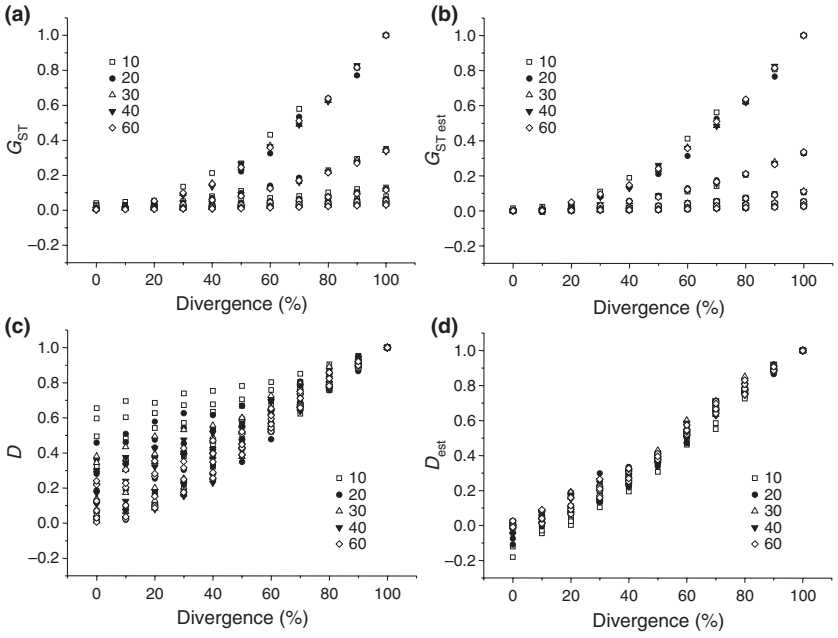

Fig. 3 (a,b,c,d) Comparison of bias corrections calculated for populations with increasing divergence, different numbers of alleles and different sample sizes. (a) Uncorrected G_{ST}- and (b) bias-corrected G_{ST_est}- (c) uncorrected D- and (d) bias-corrected D_{est}-values. Each of the five symbols occurs six times reflecting genetic differences calculated for six different numbers of alleles (N_{a_total} = 2, 4, 8, 12, 20 or 32). Because of almost equal values, symbols overlap. Box shows symbols according to different sample sizes per population (10, 20, 30, 40 and 60 individuals).

significant P values show whether genetic differences between populations are larger than generated by chance. By using very large sample sizes, P values will become significant even when differentiation between populations is small; however, considering past studies in population genetics, the real problem tends to be sample sizes that are too small rather than too large. Calculating genetic distances based on small sample sizes without considering statistical significance can create serious misinterpretation.

There are several methods to calculate confidence intervals considering the asymmetry of the bootstrap distribution that occurs especially when sample sizes are small [normal, percentile, basic, accelerated, studentized or bias corrected, accelerated (BCA) bootstrap] (DiCiccio & Efron 1996). To our knowledge, there is no uniformly superior and universally accepted method to calculate bootstrap confidence intervals. We used two different methods: bootstrapping estimated half of the width of the 95%-confidence interval as 1.96 * SD (normal method), centred at the D_{est} estimator. Shown in Table 2 we used a second method (similar to the program SPADE, Chao & Shen 2009; see also Chao *et al*. 2008), which calculates confidence intervals following the percentile method. This seems appropriate when distributions of bootstrap values are asymmetrical; however, the problem of asymmetry still persists for some distributions.

Table 2 shows that in most cases P values were consistent with confidence intervals based on the percentile method. Only for few cases, this last method generated statistically significant results for D_{est}-values when P values did not (Line 99, 106, 123 Table 2). For instance, at low sample sizes (n = 10) and a high number of alleles (N_{a_total} = 20 or 32), P values indicated a significant difference between populations when the 'real' divergence was 80%; confidence intervals indicated statistical significant differences when the 'real' divergence was 60%.

Bias correction of G_{ST} and D

Drawing random samples from infinite populations produces a bias in calculating G_{ST}- and D-values (see Table 2 and Fig. 3) (differences between real divergence between populations and D-values). Bias corrections for G_{ST}-values had been suggested by Nei & Chesser (1983). A similar bias correction was suggested by Jost (2008) for D-values. In Fig. 3a–d, we show the effect of bias corrections of G_{ST}- and D-values. While the bias correction was small for G_{ST}-values, D_{est}-values differed much more from D-values. We compared the relative change for G_{ST} and D introduced by the bias correction and found that the one for D is more pronounced. With a slight underestimation, D_{est} correctly showed the true divergence that existed in the populations from which samples were drawn with larger samples sizes showing a better result.

Summary and conclusions

By using artificial data sets with known divergence between two populations, we could test the accuracy of different methods of calculating genetic differences. We found that for more than two alleles, G_{ST_est}-values do not clearly reflect genetic differences while D_{est}-values do. Our simulations support the use of G_{ST} as a fixation index that increases with the decrease of allele diversity caused by genetic drift or selection. The maximum value of G_{ST} is dependent on the averaged homozygosity over populations which was more generally described by Nei (1973) and Hedrick (2005). Thus, G_{ST} is only appropriate to measure population differentiation for two alleles (see Table 2). This reflects that G_{ST} measures fixation and D differentiation.

In randomly generated data sets with different allele frequencies, bias corrections were smaller for G_{ST}- than for D-values, and D_{est}-values reflected the 'real' divergence better

than G_{ST_est}-values. However, negative D_{est}-values could occur especially when sample size was too small for the given number of alleles.

Calculations of statistical significance and confidence intervals (e.g. based on bootstrapping) are essential for estimating if genetic differences are meaningful. We could show that statistical significance did not differ when genetic difference was calculated with G_{ST_est} or D_{est}. Therefore, while prior calculations of G_{ST_est}-values do not show differentiation between populations, prior calculations of statistical significance between populations are probably to remain the same. To calculate real differences between populations, analyses which used G_{ST} to rank data sets need to be redone. We clearly want to state that D_{est}-values in combination with calculations of statistical significance will give a much better and appropriate description of genetic difference between populations.

The package '*DEMEtics*' to calculate D_{est}, G_{ST_est}-values as well as confidence intervals and *P* values can be requested from the authors or downloaded (http://cran.r-project.org/web/packages/DEMEtics/index.html).

Acknowledgements

We thank Lou Jost and Anne Chao for very helpful comments and Jelle Atema for reviewing the manuscript and improving the English.

References

Chao A, Shen T-J (2009) Program **SPADE** (Species Prediction and Diversity Estimation). Program and user's guide at http://chao.stat.nthu.edu.tw.

Chao A, Jost L, Chiang SC, Jiang Y-H, Chazdon R (2008) A two-stage probabilistic approach to multiple-community similarity indices. *Biometrics*, **64**, 1178–1186.

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

DiCiccio TJ, Efron B (1996) Bootstrap confidence intervals (with Discussion). *Statistical Science*, **11**, 189–228.

Goudet J, Raymond M, Demeeüs T, Rousset F (1996) Testing genetic differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Hedrick P (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Heller R, Siegismund R (2009) Relationship between three measures of genetic differentiation GST, DEST and G'ST: how wrong have we been? *Molecular Ecology*, **18**, 2080–2083.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Jost L (2009) D vs. GST: response to Heller and Siegismund (2009) and Ryman and Leimar (2009). *Molecular Ecology* **18**, 2088–2091.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, **70**, 3321–3323.

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**, 253–259.

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ryman N, Leimar O (2009) GST is still a useful measure of genetic differentiation – a comment on Jost's *D*. *Molecular Ecology*, **18**, 2084–2087.

Gabriele Gerlach is broadly interested in evolutionary ecology, particularly in the application of molecular markers to ecological questions. Her current attention concentrates on the impact of behavior and adaptations on the genetic structure of populations. Alexander Jueterbock, Philipp Kraemer and Jana Deppermann are graduate students working on the genetic structure, dispersal and genomics of different marine organisms in Gerlach's lab. Peter Harmand is a mathematician specialized in analysis and statistics.

---

doi: 10.1111/j.1365-294X.2010.04784.x