Supporting information

MS title: Biases of STRUCTURE software when exploring introduction routes of invasive species

Authors: Eric Lombaert, Thomas Guillemaud & Emeline Deleury

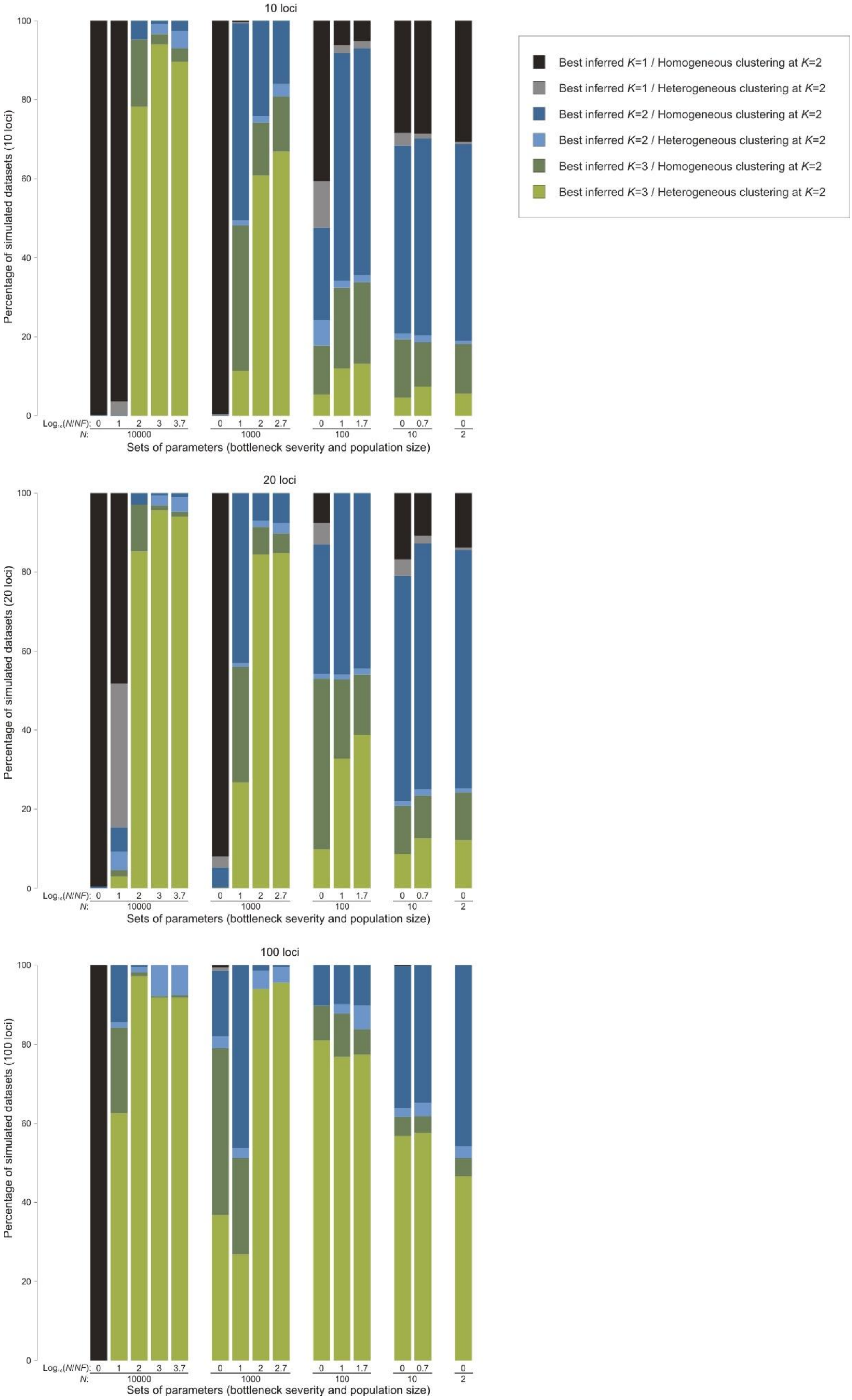

Figure S1: Proportion of best inferred values of *K* (Evanno's method) and of the absence or presence of genuine multimodality for ten STRUCTURE runs at *K*=2 for each parameter set and number of loci (i.e. 500 simulated datasets each).

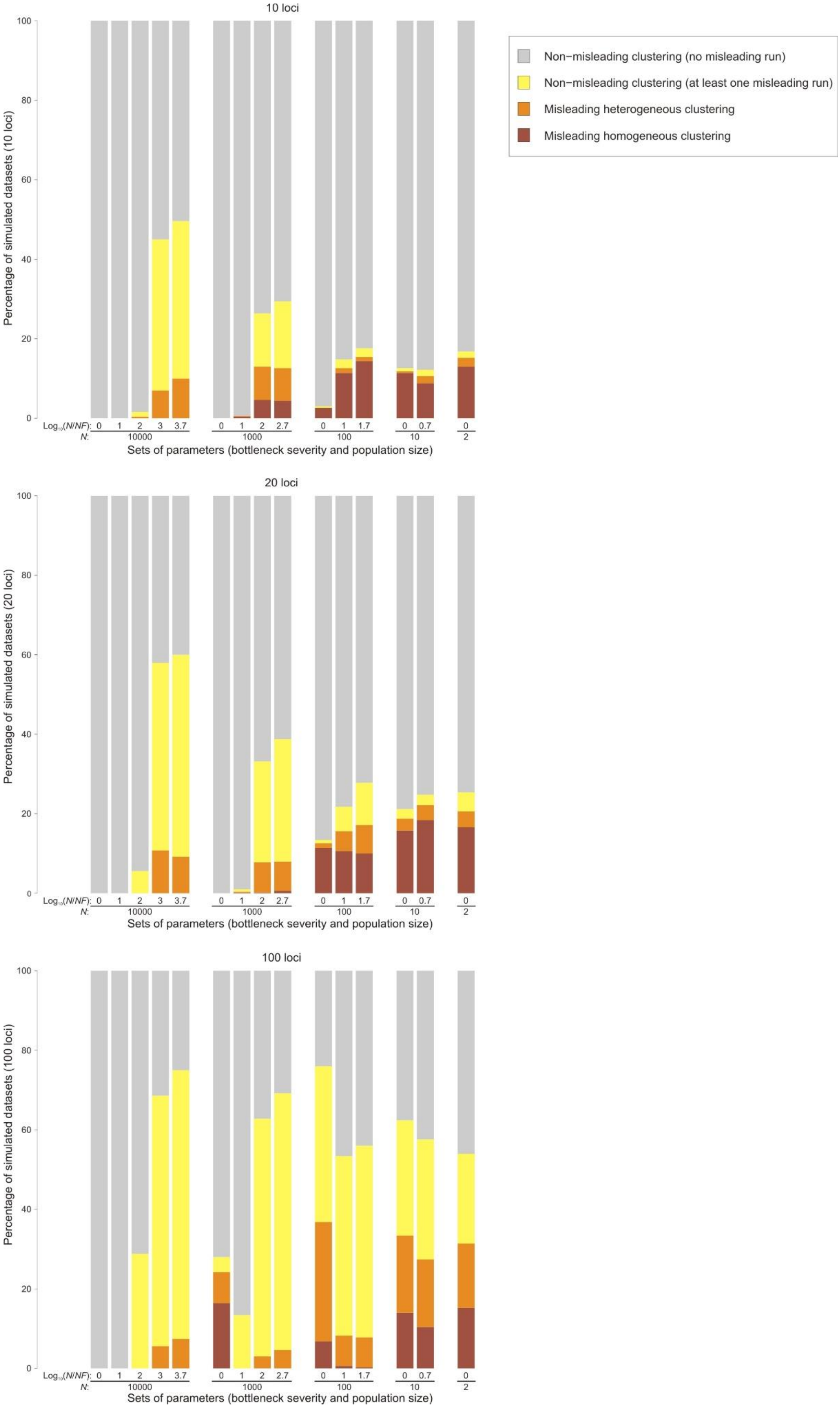

Figure S2: Proportion of datasets with or without misleading patterns over the ten STRUCTURE runs at K=2, for each set of parameters and number of loci.

Summary statistics for the simulated datasets: mean (SD; min-max)										
Nb of loci	N	NF	Log$_{10}$(N/NF)	Proportion of population differentiation	NA1	NA2 and NA3	He1	He2 and He3	FST12 and FST13	FST23
10	10000	10000	0	0.35	10.9 (1.31; 6.6-15.2)	10.92 (1.31; 6.8-15.4)	0.83 (0.03; 0.67-0.9)	0.83 (0.03; 0.67-0.9)	0.002 (0.003; -0.007-0.019)	0.002 (0.003; -0.006-0.011)
	10000	1000	1	1	10.87 (1.19; 7.7-14.2)	9.66 (1.02; 6.7-12.9)	0.83 (0.03; 0.7-0.89)	0.81 (0.03; 0.68-0.88)	0.013 (0.006; 0.001-0.036)	0.024 (0.007; 0.009-0.05)
	10000	100	2	1	10.92 (1.26; 7.1-14.9)	4.9 (0.46; 3.4-6.5)	0.83 (0.03; 0.68-0.9)	0.65 (0.05; 0.48-0.79)	0.111 (0.021; 0.05-0.175)	0.217 (0.036; 0.118-0.325)
	10000	10	3	1	10.86 (1.26; 7.9-15.4)	1.26 (0.22; 1-4)	0.83 (0.03; 0.7-0.9)	0.14 (0.12; 0-0.65)	0.453 (0.038; 0.302-0.588)	0.901 (0.048; 0.71-1)
	10000	2	3.7	1	10.98 (1.25; 6.9-14.8)	1.02 (0.11; 1-2)	0.83 (0.03; 0.72-0.9)	0.02 (0.07; 0-0.51)	0.499 (0.022; 0.418-0.574)	0.996 (0.008; 0.951-1)
	1000	1000	0	0.994	4.16 (0.54; 2.8-6.4)	4.16 (0.54; 2.6-6.1)	0.51 (0.07; 0.25-0.7)	0.51 (0.07; 0.25-0.71)	0.024 (0.012; -0.004-0.073)	0.024 (0.013; -0.002-0.088)
	1000	100	1	1	4.15 (0.55; 2.3-5.9)	2.83 (0.38; 1.8-4.1)	0.51 (0.07; 0.26-0.69)	0.41 (0.07; 0.2-0.61)	0.12 (0.04; 0.027-0.307)	0.214 (0.06; 0.059-0.449)
	1000	10	2	1	4.18 (0.58; 2.8-6.3)	1.2 (0.24; 1-3)	0.51 (0.08; 0.28-0.7)	0.14 (0.15; 0-0.58)	0.449 (0.071; 0.198-0.677)	0.888 (0.067; 0.63-1)
	1000	2	2.7	1	4.16 (0.57; 2.6-6.1)	1.03 (0.16; 1-3)	0.51 (0.07; 0.24-0.68)	0.02 (0.08; 0-0.51)	0.499 (0.065; 0.272-0.676)	0.993 (0.018; 0.829-1)
	100	100	0	0.996	1.64 (0.27; 1-2.6)	1.65 (0.25; 1-2.7)	0.14 (0.07; 0.01-0.58)	0.14 (0.06; 0-0.5)	0.176 (0.106; -0.013-0.659)	0.177 (0.105; -0.008-0.649)
	100	10	1	0.994	1.64 (0.24; 1-2.5)	1.11 (0.22; 1-3)	0.13 (0.06; 0.01-0.31)	0.12 (0.16; 0-0.65)	0.46 (0.198; 0-0.968)	0.724 (0.292; -0.017-1)
	100	2	1.7	0.99	1.64 (0.24; 1-2.4)	1.04 (0.18; 1-3)	0.13 (0.07; 0.01-0.59)	0.03 (0.08; 0-0.5)	0.513 (0.214; 0-0.953)	0.798 (0.38; -0.017-1)
	10	10	0	0.8	1.09 (0.22; 1-3)	1.08 (0.21; 1-3)	0.08 (0.13; 0-0.66)	0.09 (0.14; 0-0.56)	0.357 (0.369; -0.014-1)	0.399 (0.378; -0.008-1)
	10	2	0.7	0.652	1.08 (0.21; 1-2)	1.02 (0.13; 1-2)	0.08 (0.14; 0-0.51)	0.03 (0.09; 0-0.51)	0.371 (0.415; -0.008-1)	0.406 (0.469; -0.008-1)
	2	2	0	0.526	1.02 (0.15; 1-3)	1.03 (0.16; 1-2)	0.02 (0.08; 0-0.54)	0.03 (0.09; 0-0.51)	0.331 (0.447; -0.008-1)	0.34 (0.445; -0.011-1)
Over all parameter sets					5.29 (4.21; 1-15.4)	2.93 (3.17; 1-15.4)	0.45 (0.32; 0-0.9)	0.26 (0.3; 0-0.9)	0.292 (0.277; -0.014-1)	0.474 (0.428; -0.017-1)
20	10000	10000	0	0.582	10.92 (0.87; 8.85-13.9)	10.93 (0.89; 8.55-14.05)	0.83 (0.02; 0.76-0.89)	0.83 (0.02; 0.75-0.89)	0.002 (0.002; -0.005-0.013)	0.003 (0.002; -0.004-0.011)
	10000	1000	1	1	10.91 (0.9; 7.9-13.35)	9.66 (0.76; 7-11.9)	0.83 (0.02; 0.74-0.88)	0.81 (0.02; 0.72-0.87)	0.013 (0.004; 0.003-0.027)	0.024 (0.005; 0.012-0.041)
	10000	100	2	1	10.97 (0.85; 8.35-14.3)	4.93 (0.33; 3.7-5.95)	0.83 (0.02; 0.77-0.88)	0.66 (0.03; 0.55-0.73)	0.108 (0.015; 0.069-0.166)	0.213 (0.025; 0.155-0.297)
	10000	10	3	1	10.89 (0.93; 8.45-13.8)	1.27 (0.11; 1-1.65)	0.83 (0.02; 0.74-0.89)	0.09 (0.05; 0-0.52)	0.452 (0.025; 0.373-0.53)	0.901 (0.033; 0.775-0.986)
	10000	2	3.7	1	10.91 (0.89; 8.4-13.4)	1.03 (0.14; 1-2)	0.83 (0.02; 0.74-0.88)	0.04 (0.11; 0-0.52)	0.498 (0.015; 0.447-0.551)	0.996 (0.007; 0.965-1)
	1000	1000	0	1	4.15 (0.37; 3-5.45)	4.15 (0.38; 2.9-5.3)	0.51 (0.05; 0.35-0.64)	0.51 (0.05; 0.32-0.63)	0.024 (0.009; 0-0.06)	0.024 (0.009; 0.002-0.057)
	1000	100	1	1	4.16 (0.39; 3.1-5.3)	2.82 (0.26; 2-3.6)	0.51 (0.05; 0.34-0.62)	0.41 (0.05; 0.24-0.57)	0.118 (0.027; 0.047-0.227)	0.212 (0.041; 0.11-0.363)
	1000	10	2	1	4.15 (0.39; 3.2-5.35)	1.19 (0.11; 1-2)	0.51 (0.05; 0.36-0.66)	0.08 (0.08; 0-0.58)	0.456 (0.048; 0.281-0.612)	0.892 (0.044; 0.747-0.998)
	1000	2	2.7	1	4.16 (0.4; 2.65-5.4)	1.03 (0.12; 1-2)	0.51 (0.05; 0.3-0.63)	0.03 (0.09; 0-0.51)	0.506 (0.044; 0.356-0.638)	0.994 (0.01; 0.907-1)
	100	100	0	1	1.63 (0.17; 1.2-2.25)	1.63 (0.17; 1.15-2.2)	0.13 (0.04; 0.02-0.26)	0.13 (0.04; 0.01-0.25)	0.184 (0.077; 0.013-0.522)	0.185 (0.081; 0.019-0.463)
	100	10	1	1	1.64 (0.17; 1.2-2.3)	1.11 (0.15; 1-3)	0.13 (0.04; 0.02-0.29)	0.08 (0.13; 0-0.61)	0.493 (0.138; 0.064-0.83)	0.808 (0.164; 0-1)
	100	2	1.7	1	1.64 (0.17; 1.15-2.2)	1.03 (0.13; 1-3)	0.13 (0.04; 0.04-0.25)	0.04 (0.1; 0-0.52)	0.567 (0.133; 0.099-0.856)	0.965 (0.12; 0-1)
	10	10	0	0.954	1.08 (0.17; 1-3)	1.1 (0.18; 1-3)	0.09 (0.14; 0-0.64)	0.08 (0.13; 0-0.51)	0.504 (0.328; -0.011-1)	0.524 (0.331; 0-1)
	10	2	0.7	0.91	1.08 (0.13; 1-2)	1.03 (0.13; 1-2)	0.08 (0.14; 0-0.51)	0.04 (0.11; 0-0.51)	0.599 (0.369; -0.008-1)	0.646 (0.439; -0.008-1)
	2	2	0	0.77	1.03 (0.13; 1-2)	1.04 (0.14; 1-2)	0.04 (0.11; 0-0.53)	0.04 (0.11; 0-0.51)	0.532 (0.453; -0.017-1)	0.525 (0.457; -0.008-1)
Over all parameter sets					5.29 (4.18; 1-14.3)	2.93 (3.15; 1-14.05)	0.45 (0.32; 0-0.89)	0.26 (0.3; 0-0.89)	0.337 (0.286; -0.017-1)	0.527 (0.421; -0.008-1)
100	10000	10000	0	0.996	10.93 (0.41; 9.9-12.3)	10.93 (0.4; 9.72-12.22)	0.83 (0.01; 0.8-0.86)	0.83 (0.01; 0.79-0.86)	0.002 (0.001; -0.001-0.006)	0.002 (0.001; -0.001-0.006)
	10000	1000	1	1	10.94 (0.4; 9.93-12.69)	9.68 (0.34; 8.67-11.13)	0.83 (0.01; 0.8-0.86)	0.81 (0.01; 0.78-0.85)	0.013 (0.002; 0.009-0.019)	0.024 (0.002; 0.018-0.033)
	10000	100	2	1	10.96 (0.39; 9.82-12.11)	4.92 (0.15; 4.43-5.45)	0.83 (0.01; 0.8-0.85)	0.65 (0.01; 0.61-0.69)	0.109 (0.007; 0.088-0.131)	0.216 (0.011; 0.189-0.252)
	10000	10	3	1	10.92 (0.42; 9.83-12.31)	1.26 (0.05; 1.13-1.43)	0.83 (0.01; 0.8-0.86)	0.08 (0.02; 0.03-0.13)	0.453 (0.011; 0.422-0.493)	0.903 (0.014; 0.86-0.936)
	10000	2	3.7	1	10.93 (0.39; 9.84-12.62)	1.03 (0.03; 1-2)	0.83 (0.01; 0.8-0.86)	0.02 (0.07; 0-0.51)	0.499 (0.007; 0.474-0.521)	0.996 (0.003; 0.981-1)
	1000	1000	0	1	4.13 (0.16; 3.45-4.55)	4.13 (0.17; 3.55-4.62)	0.5 (0.02; 0.43-0.56)	0.5 (0.02; 0.43-0.57)	0.024 (0.004; 0.013-0.04)	0.023 (0.004; 0.012-0.033)
	1000	100	1	1	4.13 (0.16; 3.69-4.7)	2.81 (0.12; 2.42-3.29)	0.51 (0.02; 0.43-0.58)	0.41 (0.02; 0.33-0.48)	0.118 (0.012; 0.082-0.153)	0.213 (0.02; 0.159-0.28)
	1000	10	2	1	4.15 (0.18; 3.68-4.62)	1.19 (0.04; 1.07-1.34)	0.51 (0.02; 0.44-0.57)	0.06 (0.01; 0.02-0.1)	0.458 (0.023; 0.387-0.531)	0.893 (0.019; 0.827-0.946)
	1000	2	2.7	1	4.14 (0.16; 3.71-4.66)	1.03 (0.03; 1-2)	0.51 (0.02; 0.45-0.56)	0.02 (0.08; 0-0.52)	0.507 (0.019; 0.431-0.562)	0.993 (0.005; 0.974-1)
	100	100	0	1	1.63 (0.08; 1.44-1.88)	1.64 (0.07; 1.4-1.88)	0.13 (0.02; 0.08-0.19)	0.13 (0.02; 0.09-0.2)	0.194 (0.033; 0.099-0.323)	0.193 (0.036; 0.105-0.31)
	100	10	1	1	1.64 (0.08; 1.43-1.9)	1.11 (0.03; 1.03-1.22)	0.13 (0.02; 0.08-0.19)	0.03 (0.01; 0-0.07)	0.516 (0.058; 0.293-0.658)	0.841 (0.044; 0.621-0.952)
	100	2	1.7	1	1.63 (0.08; 1.41-1.86)	1.03 (0.04; 1-2)	0.13 (0.02; 0.07-0.19)	0.02 (0.07; 0-0.51)	0.582 (0.054; 0.417-0.732)	0.978 (0.016; 0.903-1)
	10	10	0	1	1.08 (0.03; 1-1.18)	1.08 (0.03; 1-1.19)	0.02 (0.02; 0-0.49)	0.02 (0.01; 0-0.18)	0.676 (0.116; 0.176-0.947)	0.677 (0.117; 0.096-0.96)
	10	2	0.7	1	1.08 (0.03; 1-1.18)	1.03 (0.03; 1-2)	0.02 (0.01; 0-0.05)	0.02 (0.06; 0-0.51)	0.8 (0.099; 0.068-0.989)	0.929 (0.068; 0-1)
	2	2	0	1	1.04 (0.08; 1-2)	1.03 (0.04; 1-2)	0.02 (0.07; 0-0.51)	0.02 (0.06; 0-0.5)	0.903 (0.102; -0.002-1)	0.901 (0.107; 0.051-1)
Over all parameter sets					5.29 (4.16; 1-12.69)	2.93 (3.14; 1-12.22)	0.44 (0.33; 0-0.86)	0.24 (0.3; 0-0.86)	0.39 (0.29; -0.002-1)	0.586 (0.401; -0.001-1)

Table S1: Demographic parameters (*N* and *NF*), corresponding to bottleneck severity (log$_{10}$(*N*/*NF*)), proportion of the dataset displaying significant population differentiation (Fisher's exact test) and mean summary statistics for the simulated datasets (500 datasets per number of loci and set of parameters).

Clustering codes	10 loci		20 loci		100 loci	
	Nb runs	*%*	Nb runs	*%*	Nb runs	*%*
0/100/100	5,930	*7.91*	7,700	*10.27*	11,545	*15.39*
50/50/50	26,169	*34.89*	17,449	*23.27*	5,591	*7.45*
C_{1A}/0/100 and C_{1A}/100/0	39,629	52.84	46,470	*61.96*	57,051	*76.07*
0/0/100	17,671	*23.56*	21,127	*28.17*	23,246	*30.99*
0/100/0	18,235	*24.31*	20,798	*27.73*	22,479	*29.97*
25/100/0	1,529	*2.04*	1,873	*2.50*	5,341	*7.12*
25/0/100	1,439	*1.92*	1,960	*2.61*	5,590	*7.45*
50/0/100	755	*1.01*	712	*0.95*	395	*0.53*
Other codes	3,272	4.36	3,381	*4.51*	813	*1.08*
Total	75,000	100	75,000	100	75,000	100

Table S2: Number and proportion of the main clustering codes observed in STRUCTURE runs (for each number of loci: 15 sets of parameters x 500 datasets per set x 10 runs per dataset = 75,000 STRUCTURE runs).

	10 loci		20 loci		100 loci		Total	
Cut-off values	Nb datasets	*%*	Nb datasets	*%*	Nb datasets	*%*	Nb datasets	*%*
0.1 and 0.9	1,081	*14.4*	1,595	*21.3*	3,466	*46.2*	6,142	*27.3*
0.2 and 0.8	**1,148**	***15.3***	**1,655**	***22.1***	**3,526**	***47.0***	**6,329**	***28.1***
0.3 and 0.7	1,210	*16.1*	1,713	*22.8*	3,544	*47.3*	6,467	*28.7*

Table S3: Number and proportion of datasets with at least one run (of ten) displaying a “misleading pattern” (i.e. code 0/100/100), depending on the cut-off values used to define the summarized codes of STRUCTURE patterns (over 7,500 datasets for each number of loci). The row shown in bold corresponds to the cut-off used in the article.