The American Journal of Human Genetics

Supplemental Data

Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent

Sharon R. Browning and Brian L. Browning

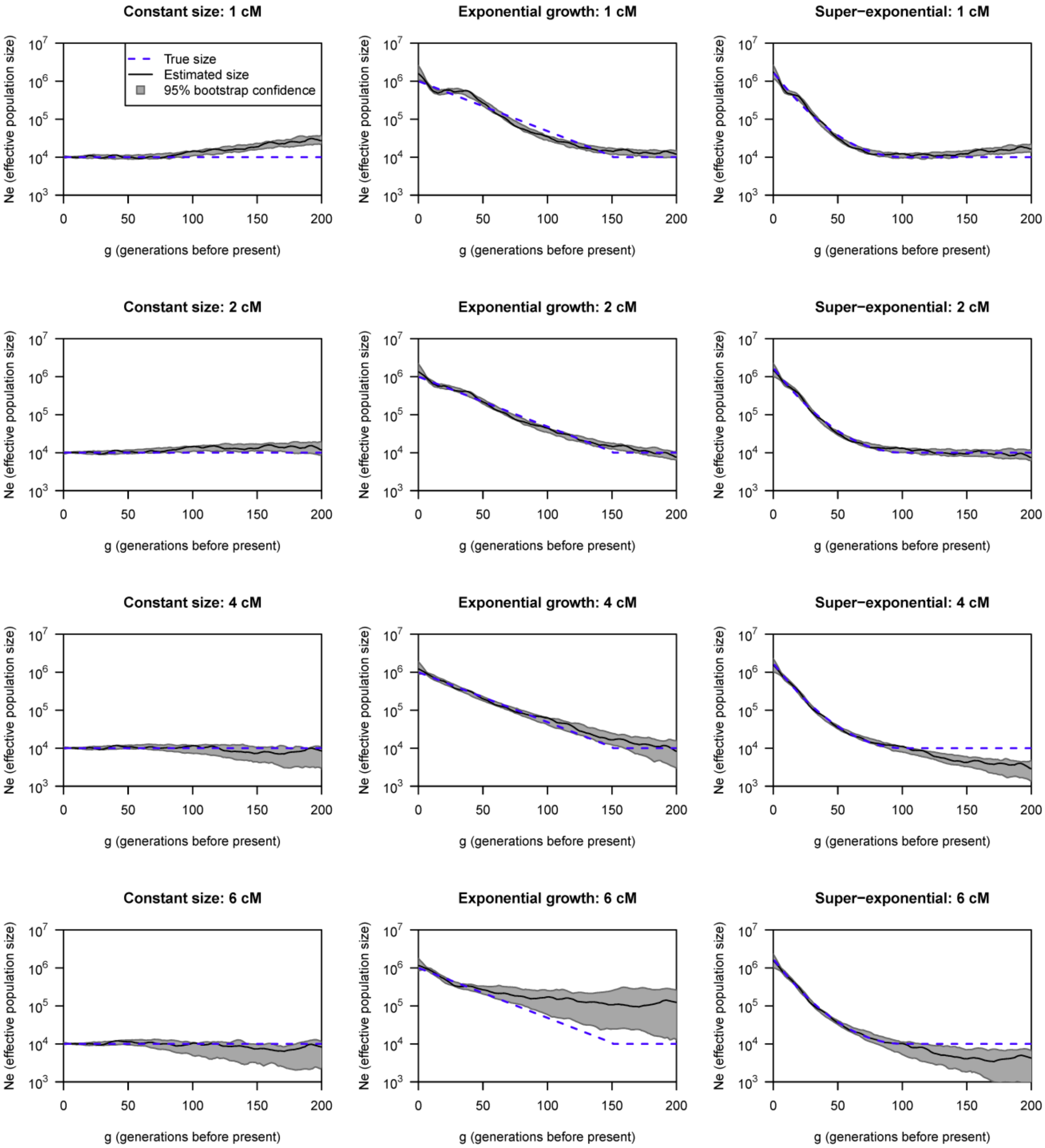

Figure S1: Estimated effective population size using IBD segments inferred from simulated sequence data with IBDseq. Each row has a different threshold on inferred IBD length (1, 2, 4 and 6 cM), while each column has a different population scenario (constant size, exponential growth and super-exponential growth). The sample size was 1000 individuals for each scenario. The blue dashed line in each plot shows the true effective population size, the black line is the estimated effective population size, and the gray regions are bootstrap 95% confidence intervals. The y-axes (effective population size) are plotted on a log scale.

The 1 cM threshold results in overestimation of effective size at high numbers of generations in the past due to incomplete power to detect the shorter segments. With the 6 cM threshold, there is little information in the IBD about effective population size more than 50 generations in the past, so the estimates tend to drift away from the true values. Similarly, there is little information about effective population size more than 100 generations in the past with the 4 cM threshold.

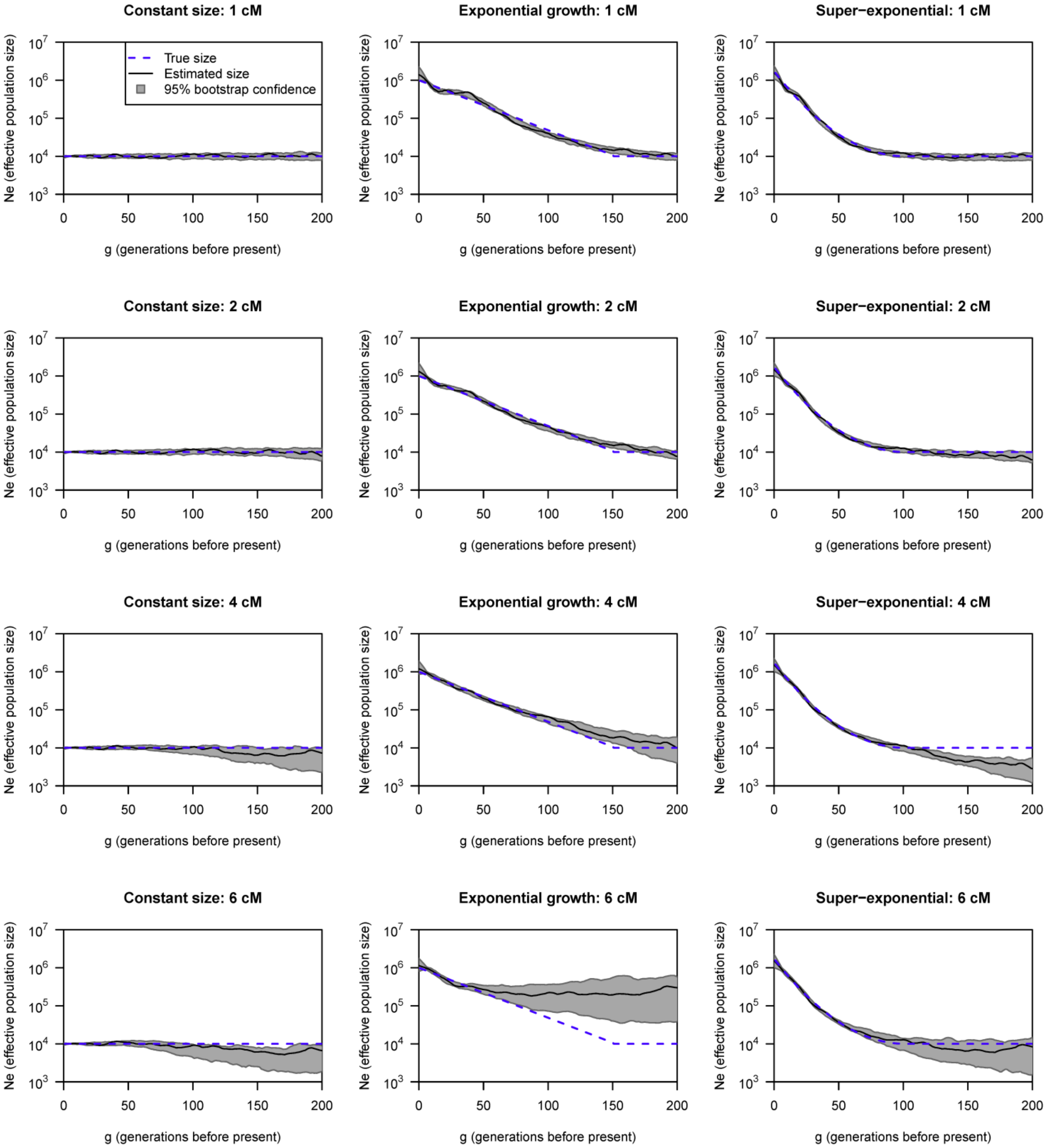

Figure S2. Estimated effective population size for the three simulated populations using actual IBD segments. Each row has a different IBD length threshold (1, 2, 4 and 6 cM), while each column is a different population scenario (constant size, exponential growth and super-exponential growth). The sample size was 1000 individuals for each scenario. The blue dashed line in each plot shows the true effective population size, the black line is the estimated effective population size, and the gray regions are bootstrap 95% confidence intervals. The y-axes (effective population size) are plotted on a log scale.

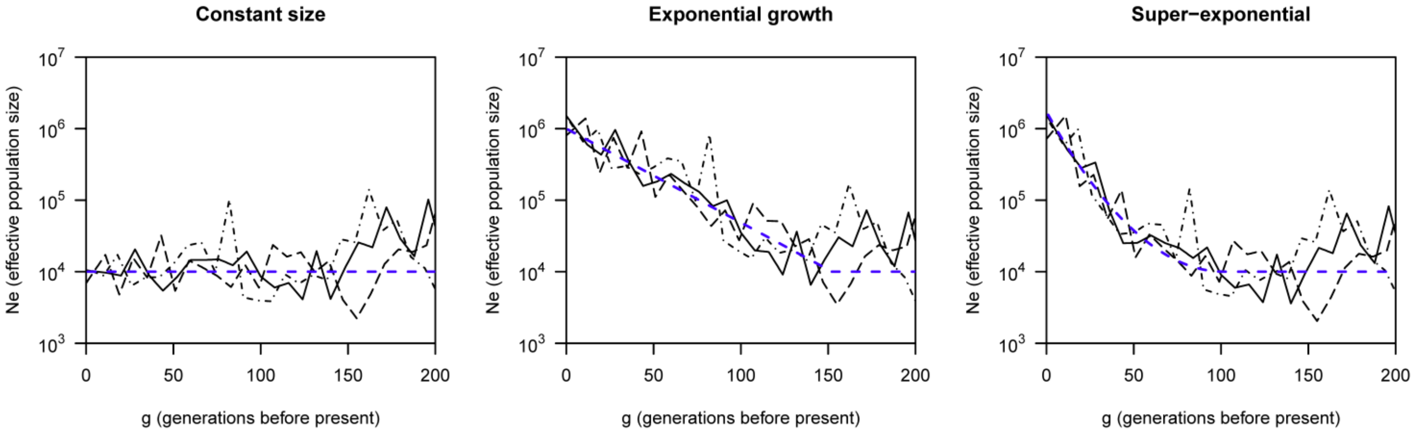

Figure S3. Estimated effective population size without averaging over multiple random starts. A 2 cM length threshold was used on the actual IBD segments. The sample size was 1000 individuals per scenario. Three random starts are shown for each simulation scenario, using a black solid line, a black dot-dash line, and a black long-dash-short-gap line. The true population size trajectory is shown with a dashed blue line. The y-axes (effective population size) are plotted on a log scale.

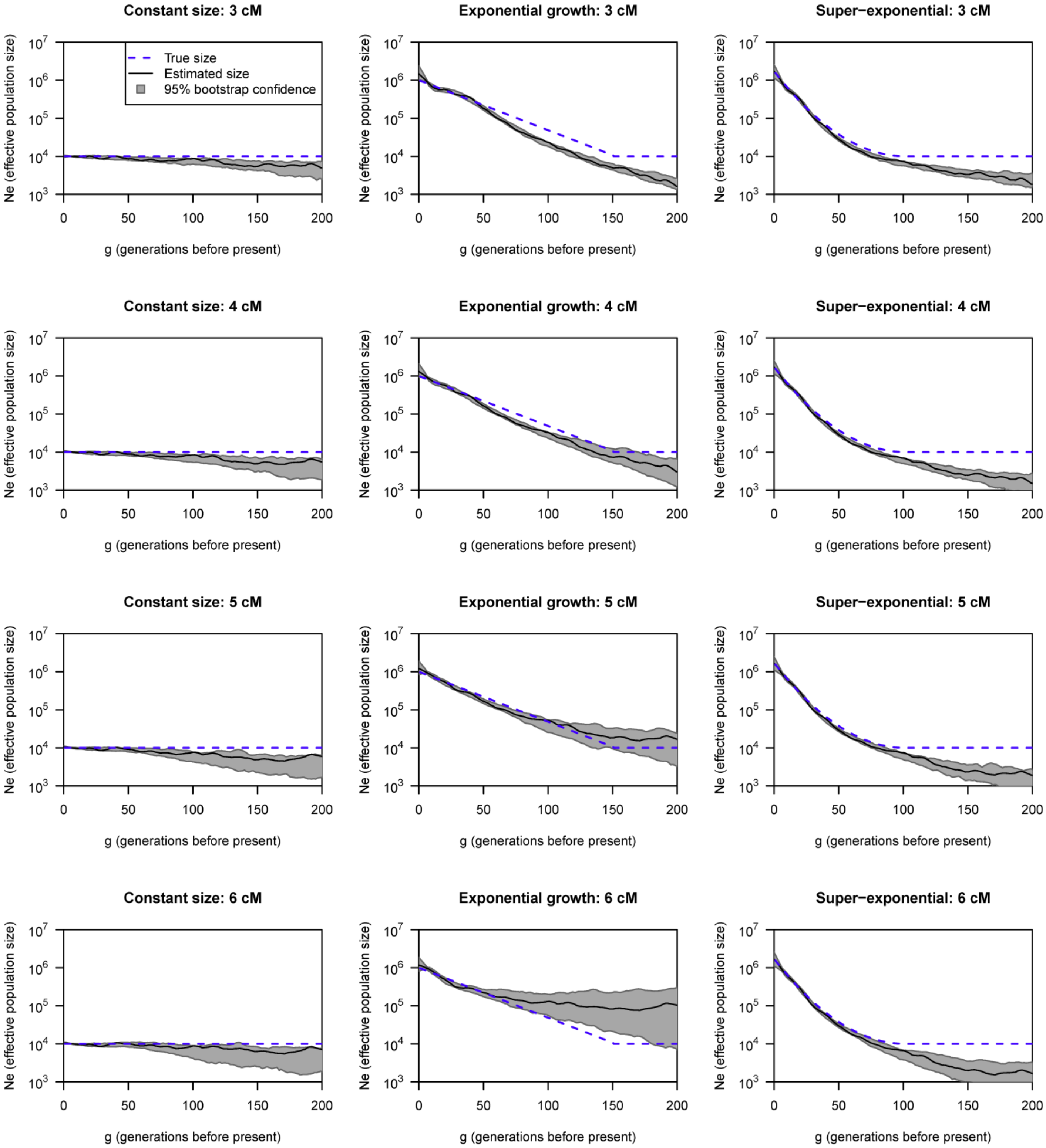

Figure S4. Effective population size estimated using IBD estimated from SNP array data. IBD segments estimated using IBDseq on SNP array data were used to estimate effective population size for the three simulated populations (constant size, exponential growth and super-exponential growth). Minimum IBD length thresholds of 3, 4, 5 and 6 cM are shown in the different rows. The sample size was 1000 individuals for each scenario. The blue dashed line in each plot shows the true effective population size, the black line is the estimated effective population size, and the gray regions are bootstrap 95% confidence intervals. The y-axes (effective population size) are plotted on a log scale.

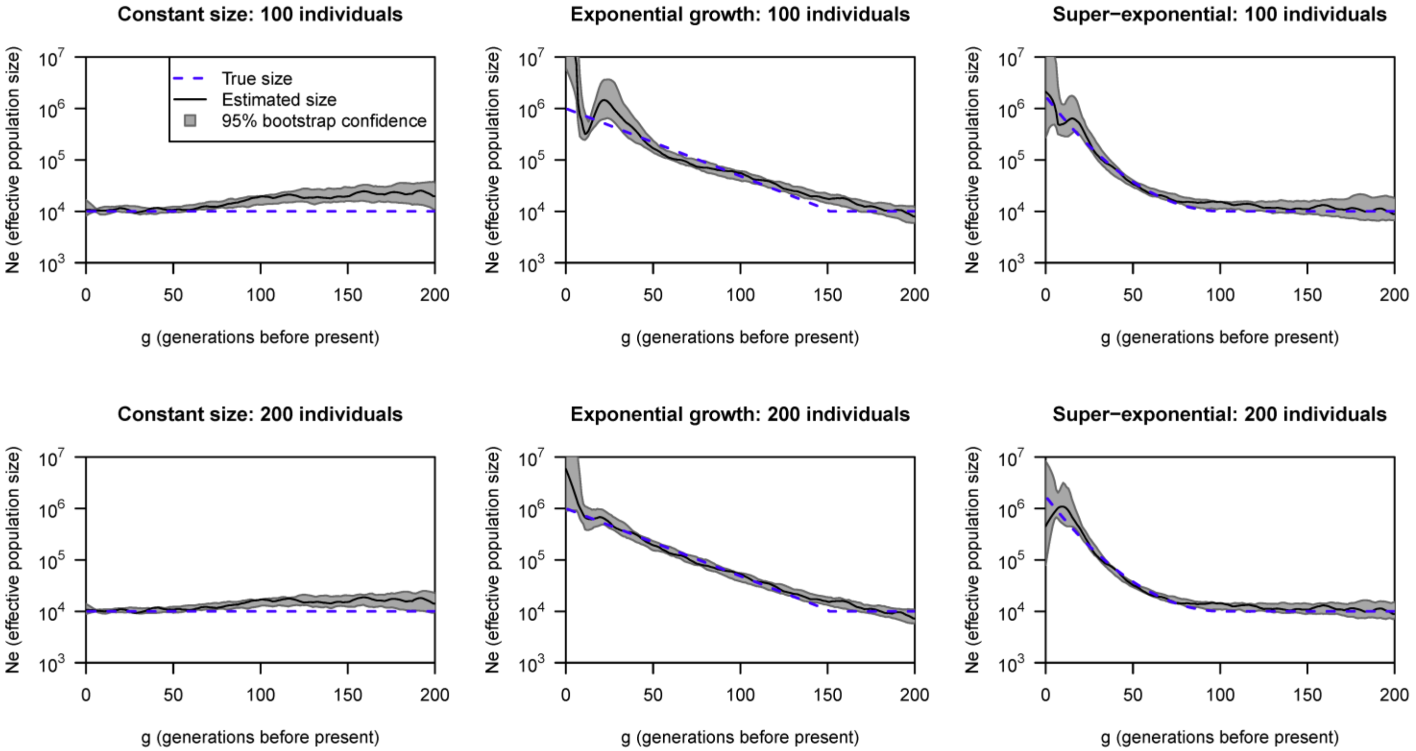

Figure S5. Effective population size estimated using inferred IBD from a small sample. Inferred IBD segments of size 2 cM and larger from 100 or 200 diploid individuals were used to estimate effective population size for the three simulated populations (constant size, exponential growth and super-exponential growth). The blue dashed line in each plot shows the true effective population size, the black line is the estimated effective population size, and the gray regions are bootstrap 95% confidence intervals. The y-axes (effective population size) are plotted on a log scale.

The results for 200 individuals are quite good, with the three simulation scenarios being clearly distinguished from each other, although with 100 or 200 individuals we see less precision and more oscillation for the most recent generations than we see when analyzing all 1000 individuals.

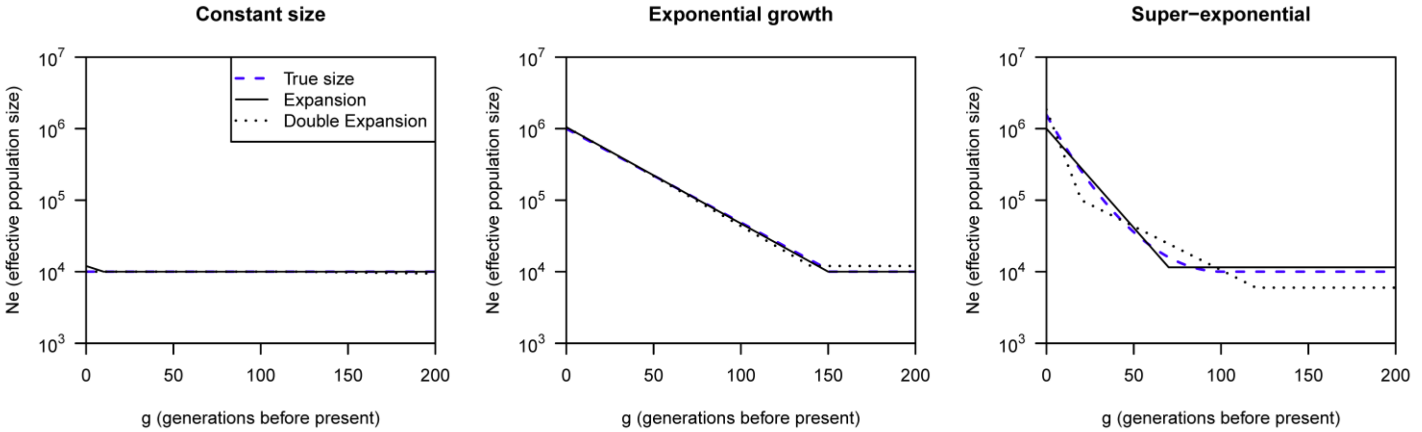

Figure S6. Effective population size estimated using the DoRIS software on actual IBD segments. Actual IBD segments of size 2 cM and larger were used. The sample size was 1000 individuals for each scenario. For each of the three scenarios, the blue dashed line represents the true effective size, the solid black line represents the estimates under DoRIS's Expansion model, while the dotted black line represents the estimates under DoRIS's Double Expansion model.

The parameter values considered for the Expansion model for the constant size scenario were: current and ancestral haploid size 1000-40,000 with increments of 1000; ancestral size 1000-40,000 with increments of 1000; generation at which growth begins 10-300 with increments of 10.

The parameter values considered for the Double Expansion model for the constant size scenario were: current haploid size 4000-40,000 with increments of 4000; size at time of change of growth rate 4000-40,000 with increments of 4000; ancestral size 4000-40,000 with increments of 4000; generation at which earlier growth begins 100-300 with increments of 20; generation at which later growth begins 20-200 with increments of 20.

The parameter values considered for the Expansion model for the exponential growth and super-exponential scenarios were: current haploid size 100,000-4,000,000 with increments of 100,000; ancestral size 1000-40,000 with increments of 1000; generation at which growth begins 10-300 with increments of 10.

The parameter values considered for the Double Expansion model for the exponential growth and super-exponential scenarios were: current haploid size 100,000-3,700,000 with increments of 400,000; size at time of change of growth rate 100,000-1,000,000 with increments of 100,000; ancestral haploid size 4000-40,000 with increments of 4000; generation at which later growth begins 20-200 with increments of 20; generation at which earlier growth begins 100-300 with increments of 20.

Table S1. MaCS commands used to simulate data. MaCS version 0.5d was used. "$seed" represents an integer value between 1 and 30 (one seed for each simulated chromosome).

Population	Command
Constant	macs 2000 1e8 -T -t 4e-4 -r 4e-4 -h 1e3 -s $seed
Growing	macs 2000 1e8 -T -t 4e-2 -r 4e-2 -h 1e3 -s $seed -G 122804.5 -eN 3.75e-5 0.01
Super-exponential	macs 2000 1e8 -T -t 4e-4 -r 4e-4 -h 1e3 -s $seed -eN 0.0 156.02 -eN 2.5e-05 141.17 -eN 5e-05 127.87 -eN 7.5e-05 115.93 -eN 0.0001 105.21 -eN 0.000125 95.583 -eN 0.00015 86.921 -eN 0.000175 79.123 -eN 0.0002 72.096 -eN 0.000225 65.759 -eN 0.00025 60.039 -eN 0.000275 54.872 -eN 0.0003 50.199 -eN 0.000325 45.971 -eN 0.00035 42.140 -eN 0.000375 38.668 -eN 0.0004 35.517 -eN 0.000425 32.655 -eN 0.00045 30.054 -eN 0.000475 27.688 -eN 0.0005 25.534 -eN 0.000525 23.571 -eN 0.00055 21.780 -eN 0.000575 20.146 -eN 0.0006 18.653 -eN 0.000625 17.288 -eN 0.00065 16.039 -eN 0.000675 14.895 -eN 0.0007 13.846 -eN 0.000725 12.884 -eN 0.00075 12.001 -eN 0.000775 11.190 -eN 0.0008 10.444 -eN 0.000825 9.7571 -eN 0.00085 9.1248 -eN 0.000875 8.5420 -eN 0.0009 8.0045 -eN 0.000925 7.5082 -eN 0.00095 7.0498 -eN 0.000975 6.6260 -eN 0.0010 6.2339 -eN 0.001025 5.8709 -eN 0.00105 5.5345 -eN 0.001075 5.2226 -eN 0.0011 4.9333 -eN 0.001125 4.6646 -eN 0.00115 4.4150 -eN 0.001175 4.1829 -eN 0.0012 3.9670 -eN 0.001225 3.7659 -eN 0.00125 3.5787 -eN 0.001275 3.4042 -eN 0.0013 3.2414 -eN 0.001325 3.0895 -eN 0.00135 2.9476 -eN 0.001375 2.8151 -eN 0.0014 2.6912 -eN 0.001425 2.5754 -eN 0.00145 2.4670 -eN 0.001475 2.3655 -eN 0.0015 2.2705 -eN 0.001525 2.1815 -eN 0.00155 2.0980 -eN 0.001575 2.0198 -eN 0.0016 1.9464 -eN 0.001625 1.8776 -eN 0.00165 1.8130 -eN 0.001675 1.7524 -eN 0.0017 1.6955 -eN 0.001725 1.6421 -eN 0.00175 1.5920 -eN 0.001775 1.5450 -eN 0.0018 1.5008 -eN 0.001825 1.4594 -eN 0.00185 1.4205 -eN 0.001875 1.3840 -eN 0.0019 1.3499 -eN 0.001925 1.3178 -eN 0.00195 1.2879 -eN 0.001975 1.2599 -eN 0.0020 1.2337 -eN 0.002025 1.2092 -eN 0.00205 1.1865 -eN 0.002075 1.1653 -eN 0.0021 1.1457 -eN 0.002125 1.1275 -eN 0.00215 1.1107 -eN 0.002175 1.0953 -eN 0.0022 1.0811 -eN 0.002225 1.0682 -eN 0.00225 1.0565 -eN 0.002275 1.0460 -eN 0.0023 1.0367 -eN 0.002325 1.0284 -eN 0.00235 1.0212 -eN 0.002375 1.0151 -eN 0.0024 1.0101 -eN 0.002425 1.0060 -eN 0.00245 1.0030 -eN 0.002475 1.0010 -eN 0.0025 1.0000