Text S1

Testing the likelihood-based methods for estimating $N_x / N_a$

We ran coalescent simulations (Hudson 2002) to test the accuracy of our methods for estimating $N_x / N_a$. Specifically, we simulated sequence data comparable to the actual data, then calculated a point estimate and a 95% confidence interval for $N_x / N_a$ on the simulated data. These simulations assumed that $N_x = 9{,}000$, $N_a = 12{,}000$ and $\mu = 2.5 * 10^{-8}$ / bp (i.e., the "true" $N_x / N_a$ is 0.75). Each replicate involved simulating 20 X-linked and 20 autosomal loci, and each locus consisted of 5 Kb of sequenced data from a region 17 Kb long. We simulated a sample size of n = 32 for the autosomal loci and n = 16 for the X-linked loci, as well as a single outgroup (orangutan) sequence.

We considered two different scenarios for recombination, each with the same average rate of $r = 1.5 * 10^{-8}$ / bp. The first scenario assumed that the recombination rate was constant across sites. The second scenario modeled recombination hotspots, assuming hotspots of 1 Kb occurred with an average density of one hotspot per 20 Kb. We assumed that hotspots are non-overlapping and that the distance between hotspots followed a negative binomial distribution with mean 19 Kb. The relative rates of recombination in hotspot and non-hotspot regions were chosen so that 5% of the sequence contained 50% of the total recombination. We ran 100 replicates for each scenario. This took approximately one week of computing time on a standard Intel Xeon processor.

For each replicate, we calculated the point estimate and 95% confidence interval for $N_x / N_a$. We then tabulated the mean and the standard deviation of the $N_x / N_a$ estimates and determined what proportion of the time the true value of $N_x / N_a$ (0.75) was contained in the estimated 95% confidence interval. The results of these simulations are summarized below.

Both sets of simulations suggest that estimates of $N_x / N_a$ are roughly unbiased, and there was no significant difference between the distributions of estimated $N_x / N_a$ values under the two different recombination models. The coverage properties of the 95% confidence interval may be slightly non-conservative under the hotspot model, though we did not run enough trials to determine this definitively and the simulation results are consistent with null expectations.

| | Mean estimate | SD of estimate | Coverage |
|---|---|---|---|
| Uniform rec. | 0.75 | 0.09 | 0.96 |
| Hotspots | 0.76 | 0.08 | 0.92 |

To evaluate the performance of our second method, we simulated paired datasets with recombination ($r = 2.3925 * 10^{-8}$ Morgans/bp) and without recombination (Hudson 2002) for a sequence of about 5Kb within a region of about 19Kb . The datasets were generated under three different female to male mating ratios: 1:1, 4:1, 7:1, under the standard neutral model. Each dataset contained 20-X linked and 20 autosomal loci, with samples sizes of n=32 for the autosomal loci and n=16 for the X-linked loci, and including a single outgroup (orangutan) sequence. In all cases the actual mating ratio was within the range of accepted values. We conclude that the assumption of no recombination doesn't seem to affect the performance of the test.

Exploring parameters for a bottleneck/growth model to explain the observed X/A ratio

We analyzed a bottleneck model with five parameters: current effective size, bottleneck size, pre-bottleneck size, time since the end of the bottleneck, and duration of the bottleneck. We explored a range of values for each of the parameters and, using exact expressions for the expected value of nucleotide diversity (Pool and Nielsen 2007) for the X chromosome and the

autosomes, we found conditions that were consistent with a constant effective size for the autosomes of 9,900-12,100 and a ratio of expected X / A nucleotide diversity of 0.808-0.879. We considered the following ranges for a 5-parameter bottleneck model with constant size in each phase of the model:

current effective size: 10000 - 200000 individuals
ratio of bottleneck population size to current population size: 0.001 - 1
ratio of pre-bottleneck population size to current population size: 0.001 - 1
generations since the end of the bottleneck: 400 - 10000
duration of the bottleneck in generations :1 – 1000

For each parameter, we considered 20 values, totaling 3,200,000 evaluations of conditions. Of those conditions that met the specifications for long-term effective population size and the ratio of expected X / A nucleotide diversity, we simulated 1,000 datasets of 20 sequences corresponding to the autosomes using the program ms (Hudson 2002). We then generated distributions of averages of Tajima's D for each set of 20 sequences, and the 1,000 averages were sorted. The values for the percentiles were estimated from the sorted averages. In all cases the value for the 97.5 percentile was more negative than each of the observed values in our samples.

Table S1. Mutation rates at 40 loci assuming a 15-million year human-orangutan divergence time.

| | μ (bp/year) |
|---|---|
| 10qMB119 | 8.497E-10 |
| 10qMB128 | 1.277E-09 |
| 12qMB46 | 1.103E-09 |
| 13qMB107 | 1.001E-09 |
| 13qMB108 | 1.055E-09 |
| 16pMB17 | 9.249E-10 |
| 18pMB7 | 1.065E-09 |
| 18qMB73 | 1.268E-09 |
| 19qMB35 | 1.219E-09 |
| 1pMB4 | 1.261E-09 |
| 20pMB7 | 9.470E-10 |
| 4qMB105 | 1.254E-09 |
| 4qMB181 | 1.056E-09 |
| 5pMB10 | 1.278E-09 |
| 5pMB4 | 1.359E-09 |
| 5qMB128 | 7.994E-10 |
| 6pMB14 | 9.859E-10 |
| 6qMB164 | 1.248E-09 |
| 7pMB8 | 1.128E-09 |
| 8pMB5 | 1.790E-09 |
| *mean* | 1.14E-09 |
| | |
| XpMB13 | 6.291E-10 |
| XpMB22 | 8.800E-10 |
| XpMB3 | 9.647E-10 |
| XpMB33 | 7.369E-10 |
| XpMB35 | 9.212E-10 |
| XpMB39 | 9.824E-10 |
| XpMB6 | 8.784E-10 |
| XpMB9 | 8.237E-10 |
| XqMB120 | 8.925E-10 |
| XqMB124 | 6.963E-10 |
| XqMB136 | 8.024E-10 |
| XqMB139 | 6.734E-10 |
| XqMB140 | 9.768E-10 |
| XqMB141 | 9.003E-10 |
| XqMB143 | 8.537E-10 |
| XqMB145 | 1.184E-09 |
| XqMB146 | 1.047E-09 |
| XqMB148 | 7.298E-10 |
| XqMB149 | 7.136E-10 |
| XqMB150 | 9.184E-10 |
| *mean* | 8.60E-10 |

Table S2. Polymorphism[a] and divergence for autosomal and X-linked loci

| | Population | Sample Size | Segregating Sites | $\pi$ (%) | $\pi$ /D[b] (%) |
|---|---|---|---|---|---|
| Autosomes | | | | | |
| | Mandenka | 14 | 374 | 0.122 | 0.036 |
| | Biaka | 14 | 383 | 0.122 | 0.036 |
| | San | 9 | 293 | 0.120 | 0.035 |
| | Han | 16 | 242 | 0.081 | 0.024 |
| | Basque | 16 | 250 | 0.090 | 0.027 |
| | Melanesians | 14 | 243 | 0.081 | 0.025 |
| X chromosome | | | | | |
| | Mandenka | 14 | 220 | 0.098 | 0.038 |
| | Biaka | 14 | 219 | 0.095 | 0.036 |
| | San | 9 | 167 | 0.085 | 0.033 |
| | Han | 16 | 133 | 0.058 | 0.022 |
| | Basque | 16 | 162 | 0.071 | 0.027 |
| | Melanesians | 14 | 128 | 0.067 | 0.025 |

[a] After subsampling to standardize the number of autosomes and X chromosomes

[b] Human-Orangutan divergence

Table S3. Breeding Sex Ratio[a]

| Population | MLE | 95%CI |
|---|---|---|
| Basque | 14.0 | (1.8 - ∞) |
| Han | 2.5 | (0.4 - ∞) |
| Melanesians | 6.0 | (1.0 - ∞) |
| Biaka | 2.9 | (0.6 - ∞) |
| Mandenka | 4.3 | (0.8 - ∞) |
| San | 1.8 | (0.2 - ∞) |

[a] Based on human - orangutan divergence

Table S4. Migration rates[a] and simulation results[b] for two-deme migration model.

| Model | $m_{12}$[c] males | $m_{12}$ females | $m_{21}$ males | $m_{21}$ females | X/A[d] deme 1 | X/A[e] deme 2 |
|---|---|---|---|---|---|---|
| Asymmetric | | 0.45 | 0.45 | | 0.750 | 0.650 |
| " | | 0.15 | 0.15 | | 0.800 | 0.640 |
| Symmetric (Females) | | 0.45 | | 0.45 | 0.750 | [e] |
| " | | 0.15 | | 0.15 | 0.754 | [e] |
| Symmetric (Males) | 0.45 | | 0.45 | | 0.744 | [e] |
| " | 0.15 | | 0.15 | | 0.754 | [e] |
| Symmetric (both sexes) | 0.45 | 0.45 | 0.45 | 0.45 | 0.758 | [e] |
| " | 0.15 | 0.15 | 0.15 | 0.15 | 0.742 | [e] |

[a] Number of migrants per generation

[b] Based on an effective population size of 5,000 for each sex.

[c] Sub-index *xy* indicates migration to deme *x* from deme *y*

[d] Mean ratio of X-linked to autosomal diversity in 10,000 simulations

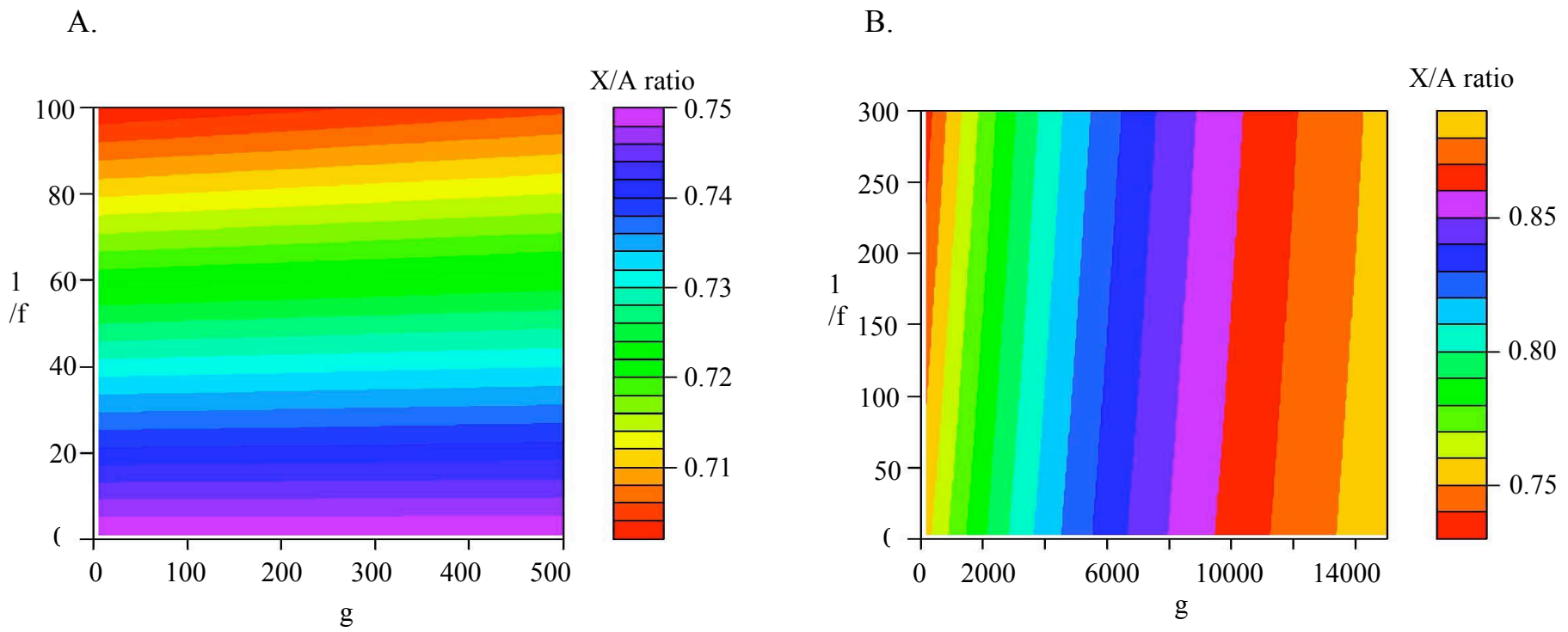
[e] Estimated expected value equal to value in deme 1

Table S5. Sample sizes (number of alleles sequenced) for each locus.

| | Locus | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10q MB119 | 10q MB128 | 12q MB46 | 13q MB107 | 13q MB108 | 16p MB17 | 18p MB7 | 18q MB73 | 19q MB35 | 1p MB4 | 20p MB7 | 4q MB105 | 4q MB181 | 5p MB10 | 5p MB4 | 5q MB128 | 6p MB14 | 6q MB164 | 7p MB8 | 8p MB5 | mean |
| Mandenka | 28 | 28 | 30 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 30 | 28 | 28 | 28 | 28 | 28 | 28 | 28.2 |
| Biaka | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28.0 |
| San | 28 | 18 | 18 | 18 | 18 | 28 | 18 | 18 | 18 | 18 | 18 | 28 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 19.5 |
| Han | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32.0 |
| Basque | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32.0 |
| Melanesians | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18.0 |

| | Xp MB13 | Xp MB22 | Xp MB33 | Xp MB35 | Xp MB39 | Xp MB3 | Xp MB6 | Xp MB9 | Xq MB120 | Xq MB124 | Xq MB136 | Xq MB139 | Xq MB140 | Xq MB141 | Xq MB143 | Xq MB145 | Xq MB146 | Xq MB148 | Xq MB149 | Xq MB150 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mandenka | 18 | 14 | 14 | 14 | 14 | 18 | 16 | 14 | 18 | 14 | 18 | 14 | 14 | 18 | 14 | 18 | 18 | 18 | 18 | 18 | 16.1 |
| Biaka | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14.0 |
| San | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9.0 |
| Han | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16.0 |
| Basque | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16.0 |
| Melanesians | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 15.0 |

Figure S1. Predicted ratio of X chromosome to autosome diversity for two bottleneck models (see Pool and Nielsen 2007). (A) Recent bottleneck. A population with initial N = 10,000 experiences a reduction in size for 100 generations followed by instantaneous growth to original size. (B) Ancient bottleneck and growth. A population with initial N = 5,750 experiences a reduction in size for 138 generations followed by instantaneous growth to its original size, and then 40-fold growth to current size. *1/f* is the strength of the bottleneck measured from current size, and *g* represents the time since the end of the bottleneck in generations.


A.
X/A ratio
0.75
0.74
0.73
0.72
0.71
100
80
60
40
20
0
1
/f
0
100
200
300
400
500
g
B.
X/A ratio
0.85
0.80
0.75
300
250
200
150
100
50
0
1
/f
0
2000
6000
10000
14000
g