The American Journal of Human Genetics, Volume *103*

Supplemental Data

A One-Penny Imputed Genome
from Next-Generation Reference Panels

Brian L. Browning, Ying Zhou, and Sharon R. Browning

Supplemental Data

Effect of window length and overlap on imputation accuracy

Figure S1: Effect of window length and overlap on imputation accuracy

Single-threaded imputation times

Table S1: Single-threaded CPU time for imputation of chromosome 14 from 2,452 reference samples from the 1000 Genomes Project

Table S2: Single-threaded CPU time for imputation of chromosome 20 from 2,452 reference samples from the 1000 Genomes Project

Table S3: Single-threaded CPU time for imputation of chromosome 14 from 26,165 Haplotype Reference Consortium reference samples

Table S4: Single-threaded CPU time for imputation of chromosome 20 from 26,165 Haplotype Reference Consortium reference samples

Table S5: Single-threaded CPU time for imputation from 10k simulated UK-European reference samples

Table S6: Single-threaded CPU time for imputation from 100k simulated UK-European reference samples

Table S7: Single-threaded CPU time for imputation from 1M simulated UK-European reference samples

Table S8: Single-threaded CPU time for imputation from 10M simulated UK-European reference samples

Multi-threaded imputation times

Table S9: Multi-threaded wall-clock time for imputation of chromosome 14 from 2,452 reference samples from the 1000 Genomes Project

Table S10: Multi-threaded wall-clock time for imputation of chromosome 20 from 2,452 reference samples from the 1000 Genomes Project

Table S11: Multi-threaded wall-clock time for imputation of chromosome 14 from 26,165 Haplotype Reference Consortium reference samples

Table S12: Multi-threaded wall-clock time for imputation of chromosome 20 from 26,165 Haplotype Reference Consortium reference samples

Table S13: Multi-threaded wall-clock time for imputation from 10k simulated UK-European reference samples

Table S14: Multi-threaded wall-clock time for imputation from 100k simulated UK-European reference samples

Table S15: Multi-threaded wall-clock time for imputation from 1M simulated UK-European reference samples

Table S16: Multi-threaded wall-clock time for imputation from 10M simulated UK-European reference samples

Supplemental Text

Msprime script for simulating UK-European samples

Figure S1

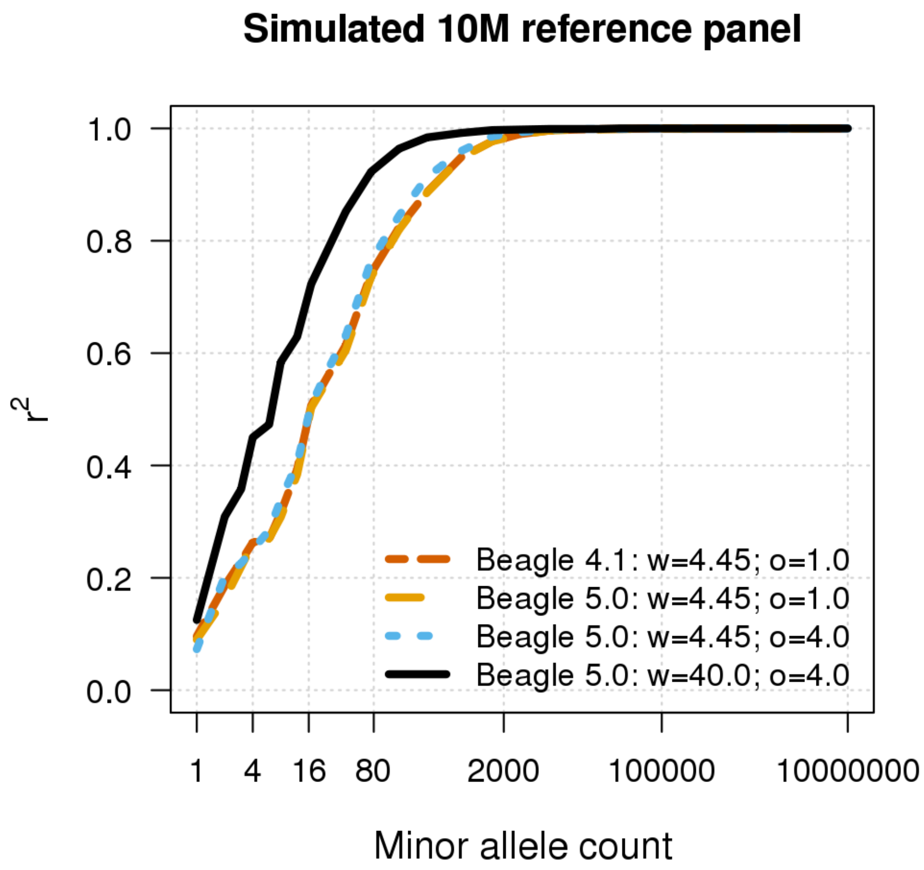

Figure S1: Effect of window length and overlap on imputation accuracy
Genotype imputation accuracy for Beagle 5.0 and Beagle 4.1 when imputing genotypes from reference panels with 10M simulated UK-European reference samples. Imputed alleles are binned according to their minor allele count in each reference panel. The squared correlation (r^2) between the true number of alleles on a haplotype (0 or 1) and the imputed posterior allele probability is reported for each minor allele count bin. The horizontal axis in each panel is on a log scale. Beagle 4.1 was run with a 4.45 cM window and 1.0 cM overlap (w=4.45; o=1.0). Beagle 5.0 was run with three different combinations of window length and overlap: 1) 4.45 cM window and 1.0 cM overlap (w=4.45; o=1.0), 2) 4.45 cM window and 4.0 cM overlap (w=4.45; o=4.0), and 3) 10.0 cM window and 4.0 cM overlap (w=10.0; o=1.0).

Table S1

Method	Target (n)	CPU seconds per target sample	Relative CPU time
Beagle 5.0	52	3.50	1.00
Beagle 4.1	52	12.70	3.62
Minimac4	52	5.56	1.59
Minimac3	52	12.39	3.53
Impute4	52	12.01	3.43

Table S1: Single-threaded CPU time for imputation of chromosome 14 from 2,452 reference samples from the 1000 Genomes Project

Single-threaded CPU time for Beagle 5.0, Beagle 4.1, Minimac4, Minimac3, and Impute4 to impute chromosome 14 (2,508,019 markers) from 2,452 reference samples from the 1000 Genomes Project into 52 target samples genotyped for 72,973 markers. Imputation analyses were run on a 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. CPU time is the sum of user and system time reported by the Unix time command. Relative CPU time is the ratio of the CPU time to the Beagle 5.0 CPU time.

Table S2

Method	Target (n)	CPU seconds per target sample	Relative CPU time
Beagle 5.0	52	2.57	1.00
Beagle 4.1	52	6.58	2.56
Minimac4	52	3.97	1.55
Minimac3	52	8.57	3.34
Impute4	52	7.99	3.11

Table S2: Single-threaded CPU time for imputation of chromosome 20 from 2,452 reference samples from the 1000 Genomes Project

Single-threaded CPU time for Beagle 5.0, Beagle 4.1, Minimac4, Minimac3, and Impute4 to impute chromosome 20 (1,718,742 markers) from 2,452 reference samples from the 1000 Genomes Project into 52 target samples genotyped for 55,885 markers. Imputation analyses were run on a 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. CPU time is the sum of user and system time reported by the Unix time command. Relative CPU time is the ratio of the CPU time to the Beagle 5.0 CPU time.

Table S3

Method	Target (n)	CPU seconds per target sample	Relative CPU time
Beagle 5.0	1000	2.15	1.00
Beagle 4.1	1000	45.80	21.29
Minimac4	1000	20.33	9.45
Minimac3	1000	21.52	10.00
Impute4	1000	54.95	25.54
Beagle 5.0	100	2.98	1.00
Beagle 4.1	100	51.39	17.22
Minimac4	100	26.87	9.00
Minimac3	100	22.06	7.39
Impute4	100	60.70	20.34
Beagle 5.0	10	9.18	1.00
Beagle 4.1	10	109.44	11.92
Minimac4	10	92.80	10.10
Minimac3	10	27.39	2.98
Impute4	10	123.19	13.41

Table S3: Single-threaded CPU time for imputation of chromosome 14 from 26,165 Haplotype Reference Consortium reference samples

Single-threaded CPU time for Beagle 5.0, Beagle 4.1, Minimac4, Minimac3, and Impute4 to impute chromosome 14 (1,236,165 markers) from 26,165 Haplotype Reference Consortium reference samples into 1000, 100, and 10 target samples genotyped for 72,886 markers. Imputation analyses were run on a 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. CPU time is the sum of user and system time reported by the Unix time command. Relative CPU time is the ratio of the CPU time to the Beagle 5.0 CPU time for the same number of target samples.

Table S4

Method	Target (n)	CPU seconds per target sample	Relative CPU time
Beagle 5.0	1000	1.79	1.00
Beagle 4.1	1000	40.99	22.95
Minimac4	1000	15.91	8.91
Minimac3	1000	15.59	8.73
Impute4	1000	39.84	22.31
Beagle 5.0	100	2.34	1.00
Beagle 4.1	100	44.82	19.13
Minimac4	100	20.44	8.72
Minimac3	100	16.01	6.83
Impute4	100	44.05	18.80
Beagle 5.0	10	7.24	1.00
Beagle 4.1	10	88.67	12.25
Minimac4	10	66.98	9.25
Minimac3	10	19.85	2.74
Impute4	10	87.75	12.12

Table S4: Single-threaded CPU time for imputation of chromosome 20 from 26,165 Haplotype Reference Consortium reference samples

Single-threaded CPU time for Beagle 5.0, Beagle 4.1, Minimac4, Minimac3, and Impute4 to impute chromosome 20 (862,298 markers) from 26,165 Haplotype Reference Consortium reference samples into 1000, 100, and 10 target samples genotyped for 55,013 markers. Imputation analyses were run on a 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. CPU time is the sum of user and system time reported by the Unix time command. Relative CPU time is the ratio of the CPU time to the Beagle 5.0 CPU time for the same number of target samples.

Table S5

Method	Target (n)	CPU seconds per target sample	Relative CPU time
Beagle 5.0	1000	0.12	1.00
Beagle 4.1	1000	1.81	14.86
Minimac4	1000	0.36	2.99
Minimac3	1000	0.82	6.78
Impute4	1000	0.96	7.93
Beagle 5.0	100	0.20	1.00
Beagle 4.1	100	1.99	10.08
Minimac4	100	0.47	2.38
Minimac3	100	0.85	4.32
Impute4	100	1.11	5.62
Beagle 5.0	10	0.76	1.00
Beagle 4.1	10	3.38	4.42
Minimac4	10	1.49	1.95
Minimac3	10	1.15	1.51
Impute4	10	2.55	3.34

Table S5: Single-threaded CPU time for imputation from 10k simulated UK-European reference samples

Single-threaded CPU time for Beagle 5.0, Beagle 4.1, Minimac4, Minimac3, and Impute4 to impute a 10 Mb region (88,124 markers) from 10k simulated UK-European reference samples into 1000, 100, and 10 target samples genotyped for 3,333 multi-allelic markers. Imputation analyses were run on a 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. CPU time is the sum of user and system time reported by the Unix time command. Relative CPU time is the ratio of the CPU time to the Beagle 5.0 CPU time for the same number of target samples.

Table S6

Method	Target (n)	CPU seconds per target sample	Relative CPU time
Beagle 5.0	1000	0.17	1.00
Beagle 4.1	1000	14.55	84.23
Minimac4	1000	2.10	12.14
Minimac3	1000	8.14	47.13
Impute4	1000	---	---
Beagle 5.0	100	0.29	1.00
Beagle 4.1	100	15.37	52.51
Minimac4	100	3.63	12.42
Minimac3	100	8.35	28.54
Impute4	100	---	---
Beagle 5.0	10	1.25	1.00
Beagle 4.1	10	24.77	19.76
Minimac4	10	17.41	13.89
Minimac3	10	10.45	8.33
Impute4	10	---	---

Table S6: Single-threaded CPU time for imputation from 100k simulated UK-European reference samples

Single-threaded CPU time for Beagle 5.0, Beagle 4.1, Minimac4, Minimac3, and Impute4 to impute a 10 Mb region (352,091 markers) from 100k simulated UK-European reference samples into 1000, 100, and 10 target samples genotyped for 3,333 markers. Imputation analyses were run on a 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. CPU time is the sum of user and system time reported by the Unix time command. Relative CPU time is the ratio of the CPU time to the Beagle 5.0 CPU time for the same number of target samples. A dash (---) indicates the analysis could not be performed due to software limitations.

Table S7

Method	Target (n)	CPU seconds per target sample	Relative CPU time
Beagle 5.0	1000	0.36	1.00
Beagle 4.1	1000	60.94	169.23
Minimac4	1000	15.63	43.42
Minimac3	1000	132.77	368.69
Impute4	1000	---	---
Beagle 5.0	100	0.76	1.00
Beagle 4.1	100	72.18	94.76
Minimac4	100	40.87	53.66
Minimac3	100	138.01	181.19
Impute4	100	---	---
Beagle 5.0	10	4.86	1.00
Beagle 4.1	10	178.58	36.74
Minimac4	10	301.78	62.08
Minimac3	10	171.96	35.38
Impute4	10	---	---

Table S7: Single-threaded CPU time for imputation from 1M simulated UK-European reference samples

Single-threaded CPU time for Beagle 5.0, Beagle 4.1, Minimac4, Minimac3, and Impute4 to impute a 10 Mb region (1,210,904 markers) from 1M simulated UK-European reference samples into 1000, 100, and 10 target samples genotyped for 3,333 markers. Imputation analyses were run on a 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. CPU time is the sum of user and system time reported by the Unix time command. Relative CPU time is the ratio of the CPU time to the Beagle 5.0 CPU time for the same number of target samples. A dash (---) indicates the analysis could not be performed due to software limitations.

Table S8

Method	Target (n)	CPU seconds per target sample	Relative CPU time
Beagle 5.0	1000	1.37	1.00
Beagle 4.1	1000	727.91	533.16
Minimac4	1000	---	---
Minimac3	1000	---	---
Impute4	1000	---	---
Beagle 5.0	100	5.51	1.00
Beagle 4.1	100	853.80	155.01
Minimac4	100	---	---
Minimac3	100	---	---
Impute4	100	---	---
Beagle 5.0	10	44.40	1.00
Beagle 4.1	10	2060.71	46.42
Minimac4	10	---	---
Minimac3	10	---	---
Impute4	10	---	---

Table S8: Single-threaded CPU time for imputation from 10M simulated UK-European reference samples

Single-threaded CPU time for Beagle 5.0, Beagle 4.1, Minimac4, Minimac3, and Impute4 when imputing a 10 Mb region (3,145,425 markers) from 1M simulated UK-European reference samples into 1000, 100, and 10 target samples genotyped for 3,333 markers. Imputation analyses were run on a 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. CPU time is the sum of user and system time reported by the Unix time command. Relative CPU time is the ratio of the CPU time to the Beagle 5.0 CPU time for the same number of target samples. A dash (---) indicates the analysis could not be performed due to memory constraints or software limitations.

Table S9

Method	Target (n)	Wall-clock seconds per target sample	Relative wall-clock time
Beagle 5.0	52	0.62	1.00
Beagle 4.1	52	2.35	3.78
Minimac4	52	3.15	5.07
Minimac3	52	9.61	15.46

Table S9: Multi-threaded wall-clock time for imputation of chromosome 14 from 2,452 reference samples from the 1000 Genomes Project
Multi-threaded wall-clock time for Beagle 5.0, Beagle 4.1, Minimac4, and Minimac3 when using 12 CPU cores to impute chromosome 14 (2,508,019 markers) from 2,452 reference samples from the 1000 Genomes Project into 52 target samples genotyped for 72,973 markers. Imputation analyses were run on a 12-core 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. Relative wall-clock time is the ratio of the wall-clock time to the Beagle 5.0 wall-clock time.

Table S10

Method	Target (n)	Wall-clock seconds per target sample	Relative wall-clock time
Beagle 5.0	52	0.40	1.00
Beagle 4.1	52	1.43	3.63
Minimac4	52	2.17	5.49
Minimac3	52	6.40	16.19

Table S10: Multi-threaded wall-clock time for imputation of chromosome 20 from 2,452 reference samples from the 1000 Genomes Project

Multi-threaded wall-clock time for Beagle 5.0, Beagle 4.1, Minimac4, and Minimac3 when using 12 CPU cores to impute chromosome 20 (1,718,742 markers) from 2,452 reference samples from the 1000 Genomes Project into 52 target samples genotyped for 54,885 markers. Imputation analyses were run on a 12-core 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. Relative wall-clock time is the ratio of the wall-clock time to the Beagle 5.0 wall-clock time.

Table S11

Method	Target (n)	Wall-clock seconds per target sample	Relative wall-clock time
Beagle 5.0	1000	0.26	1.00
Beagle 4.1	1000	4.26	16.54
Minimac4	1000	3.02	11.71
Minimac3	1000	12.26	47.55
Beagle 5.0	100	0.64	1.00
Beagle 4.1	100	8.67	13.63
Minimac4	100	9.88	15.54
Minimac3	100	13.04	20.51
Beagle 5.0	10	4.41	1.00
Beagle 4.1	10	30.95	7.01
Minimac4	10	77.57	17.58
Minimac3	10	18.56	4.21

Table S11: Multi-threaded wall-clock time for imputation of chromosome 14 from 26,165 Haplotype Reference Consortium reference samples

Multi-threaded wall-clock time for Beagle 5.0, Beagle 4.1, Minimac4, and Minimac3 when using 12 CPU cores to impute chromosome 14 (1,236,165 markers) from 26,165 Haplotype Reference Consortium reference samples into 1000, 100, and 10 target samples genotyped for 72,886 markers. Imputation analyses were run on a 12-core 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. Relative wall-clock time is the ratio of the wall-clock time to the Beagle 5.0 wall-clock time for the same number of target samples.

Table S12

Method	Target (n)	Wall-clock seconds per target sample	Relative wall-clock time
Beagle 5.0	1000	0.20	1.00
Beagle 4.1	1000	4.75	23.24
Minimac4	1000	2.29	11.19
Minimac3	1000	8.67	42.45
Beagle 5.0	100	0.48	1.00
Beagle 4.1	100	6.76	14.15
Minimac4	100	7.11	14.90
Minimac3	100	9.38	19.65
Beagle 5.0	10	3.28	1.00
Beagle 4.1	10	24.48	7.48
Minimac4	10	54.85	16.75
Minimac3	10	13.22	4.04

Table S12: Multi-threaded wall-clock time for imputation of chromosome 20 from 26,165 Haplotype Reference Consortium reference samples

Multi-threaded wall-clock time for Beagle 5.0, Beagle 4.1, Minimac4, and Minimac3 when using 12 CPU cores to impute chromosome 20 (862,298 markers) from 26,165 Haplotype Reference Consortium reference samples into 1000, 100, and 10 target samples genotyped for 55,013 markers. Imputation analyses were run on a 12-core 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. Relative wall-clock time is the ratio of the wall-clock time to the Beagle 5.0 wall-clock time for the same number of target samples.

Table S13

Method	Target (n)	Wall-clock seconds per target sample	Relative wall-clock time
Beagle 5.0	1000	0.01	1.00
Beagle 4.1	1000	0.14	9.70
Minimac4	1000	0.07	5.15
Minimac3	1000	0.25	17.39
Beagle 5.0	100	0.05	1.00
Beagle 4.1	100	0.21	4.62
Minimac4	100	0.18	3.91
Minimac3	100	0.32	6.83
Beagle 5.0	10	0.32	1.00
Beagle 4.1	10	0.82	2.58
Minimac4	10	1.21	3.79
Minimac3	10	0.86	2.71

Table S13: Multi-threaded wall-clock time for imputation from 10k simulated UK-European reference samples

Multi-threaded wall-clock time for Beagle 5.0, Beagle 4.1, Minimac4, and Minimac3 when using 12 CPU cores to impute a 10 Mb region (88,124 markers) from 10k simulated UK-European reference samples into 1000, 100, and 10 target samples genotyped for 3,333 markers. Imputation analyses were run on a 12-core 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. Relative wall-clock time is the ratio of the wall-clock time to the Beagle 5.0 wall-clock time for the same number of target samples.

Table S14

Method	Target (n)	Wall-clock seconds per target sample	Relative wall-clock time
Beagle 5.0	1000	0.02	1.00
Beagle 4.1	1000	1.34	66.07
Minimac4	1000	0.46	22.74
Minimac3	1000	4.31	212.06
Beagle 5.0	100	0.06	1.00
Beagle 4.1	100	1.82	28.84
Minimac4	100	1.86	29.54
Minimac3	100	4.53	72.00
Beagle 5.0	10	0.46	1.00
Beagle 4.1	10	5.87	12.66
Minimac4	10	15.54	33.52
Minimac3	10	6.81	14.70

Table S14: Multi-threaded wall-clock time for imputation from 100k simulated UK-European reference samples

Multi-threaded wall-clock time for Beagle 5.0, Beagle 4.1, Minimac4, and Minimac3 when using 12 CPU cores to impute a 10 Mb region (352,091 markers) from 100k simulated UK-European reference samples into 1000, 100, and 10 target samples genotyped for 3,333 markers. Imputation analyses were run on a 12-core 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. Relative wall-clock time is the ratio of the wall-clock time to the Beagle 5.0 wall-clock time for the same number of target samples.

Table S15

Method	Target (n)	Wall-clock seconds per target sample	Relative wall-clock time
Beagle 5.0	1000	0.04	1.00
Beagle 4.1	1000	6.03	155.68
Minimac4	1000	6.08	157.13
Minimac3	1000	---	---
Beagle 5.0	100	0.15	1.00
Beagle 4.1	100	10.94	72.60
Minimac4	100	32.30	214.43
Minimac3	100	---	---
Beagle 5.0	10	1.20	1.00
Beagle 4.1	10	57.99	48.41
Minimac4	10	295.33	246.56
Minimac3	10	---	---

Table S15: Multi-threaded wall-clock time for imputation from 1M simulated UK-European reference samples

Multi-threaded wall-clock time for Beagle 5.0, Beagle 4.1, Minimac4, and Minimac3 when using 12 CPU cores to impute a 10 Mb region (1,210,904 markers) from 1M simulated UK-European reference samples into 1000, 100, and 10 target samples genotyped for 3,333 markers. Imputation analyses were run on a 12-core 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. Relative wall-clock time is the ratio of the wall-clock time to the Beagle 5.0 wall-clock time for the same number of target samples. A dash (---) indicates the analysis could not be performed due to memory constraints.

Table S16

Method	Target (n)	Wall-clock seconds per target sample	Relative wall-clock time
Beagle 5.0	1000	0.16	1.00
Beagle 4.1	1000	72.83	457.99
Minimac4	1000	---	---
Minimac3	1000	---	---
Beagle 5.0	100	0.83	1.00
Beagle 4.1	100	139.65	167.54
Minimac4	100	---	---
Minimac3	100	---	---
Beagle 5.0	10	7.12	1.00
Beagle 4.1	10	662.04	93.00
Minimac4	10	---	---
Minimac3	10	---	---

Table S16: Multi-threaded wall-clock time for imputation from 10M simulated UK-European reference samples

Multi-threaded wall-clock time for Beagle 5.0, Beagle 4.1, Minimac4, and Minimac3 when using 12 CPU cores to impute a 10 Mb region (3,145,425 markers) from 10M simulated UK-European reference samples into 1000, 100, and 10 target samples genotyped for 3,333 markers. Imputation analyses were run on a 12-core 2.6 GHz Intel Xeon E5-2630v2 computer with 128 GB of memory. Relative wall-clock time is the ratio of the wall-clock time to the Beagle 5.0 wall-clock time for the same number of target samples. A dash (---) indicates the analysis could not be performed due to memory constraints.

Msprime script for simulating UK-European samples

```
# Msprime simulation script for generating a user-specified length of
# chromosome with constant recombination rate for UK individuals with
# European ancestry.

# Population history for eur, yri, asn is based on the Gravel et al.
# 2011 analysis (see https://doi.org/10.1073/pnas.1019276108).
# Recent growth is based on our IBDNe analysis of the UK10K data
# (see https://doi.org/10.1016/j.ajhg.2015.07.012).

# Growth rates change at 150 generations ago (due to introduction
# of agriculture) and 15 generations ago (start of increased growth
# seen in IBDNe analysis).

# Fitted effective population size is 3.96e6 now, 5.64e5 at 15
# generations ago, 9.80e3 at 150 generations ago.  Hence growth rate
# .00292 until 150 generations ago, then .030 until 15 generations
# ago, then .13 until present

# Msprime generates positions of mutations as floating numbers, but
# converts to integer positions when writing the vcf file, and it
# outputs only diallelic variants. Our simulation script rescales
# distance so that distinct mutations occur at distinct positions.
# After simulation, the output positions should be rescaled by
# dividing by the scale factor, and mutations mapping to the same
# rescaled position can be converted into multi-allelic markers.

# usage: python2.7 uk_scale.py seed nhaps nbp scale > vcf.file

# for the results in the paper, we used seed=1; nhaps=20002000;
# nbp=10000000; scale=100.
```

```
import msprime, sys
from math import log
from math import exp

seed = int(sys.argv[1])
nhaps = int(sys.argv[2])
scale = int(sys.argv[4])
nbp = int(sys.argv[3])*scale

mu=1.25e-8/scale # mutation rate per bp
rho=1e-8/scale   # recombination rate per bp
genlen=25        # years per generation

N0=7310   # initial population size, and reference effective size
Thum=5920 # =148000/genlen; generations back to initial pop growth (advent of modern humans)
Naf=14474 # size of african population from t0 onwards

Tooa=2040 #51000/genlen # number of generations back to Out of Africa
Nb=1861                 # size of out of Africa population
mafb=1.5e-4      # migration rate between Africa and Out-of-Africa

Teu=920 #23000/genlen # number of generations back to Asia-Europe split
Neu=1032; Nas=554 # bottleneck population sizes after the split
mafeu=2.5e-5; mafas=7.8e-6; meuas=3.11e-5 # migration rates between Africa, Europe and Asia
ras=0.0048 # growth rates per generation in Asia
reu=0.00292

Tex=150   # generations back to accelerated population growth
rex=0.03  # accelerated growth rate
Tmod=15
rmod=0.13 # growth rate in most recent generations

# pop0 is Africa, pop1 is eur (uk), pop2 is east asn
# will generate nhaps from eur pop
samplesize = nhaps
othersize = 0
pop_config = [
     msprime.PopulationConfiguration(sample_size=othersize,initial_size=Naf*exp(rex*Tex), growth_rate= rex),
```

```
     msprime.PopulationConfiguration(sample_size=samplesize,initial_size= Neu*exp(reu*(Teu-Tex))*
exp(rex*(Tex-Tmod))* exp(rmod*Tmod), growth_rate=rmod),
     msprime.PopulationConfiguration(sample_size=othersize,initial_size= Nas*exp((Teu-Tex)*ras)*
exp(rex*Tex),growth_rate=rex)
]

mig_mat = [[0,mafeu,mafas],[mafeu,0,meuas],[mafas,meuas,0]]

# recent change in growth rate
recent_event = [
    msprime.PopulationParametersChange(time=Tmod,growth_rate=rex,population_id=1)
]

# populations stop having accelerated growth (advent of agriculture)
ag_event = [
    msprime.PopulationParametersChange(time=Tex,growth_rate=ras,population_id=2),
    msprime.PopulationParametersChange(time=Tex,growth_rate=reu,population_id=1),
    msprime.PopulationParametersChange(time=Tex,growth_rate=0.0,population_id=0)
]

# Asia and Europe merge, migration changes, population size changes, growth stops
eu_event = [
    msprime.MigrationRateChange(time=Teu,rate=0.0),
    msprime.PopulationParametersChange(time=Teu,growth_rate=0.0, population_id=2),
    msprime.MassMigration(time=Teu+0.0001,source=2,destination=1,proportion=1.0),
    msprime.PopulationParametersChange(time=Teu+0.0001,initial_size=Nb,growth_rate=0.0, population_id=1),
    msprime.MigrationRateChange(time=Teu+0.0001,rate=mafb,matrix_index=(0,1)),
    msprime.MigrationRateChange(time=Teu+0.0001,rate=mafb,matrix_index=(1,0))
]

# Out of Africa event (looking back, Africa and Europe merge)
ooa_event = [
    msprime.MigrationRateChange(time=Tooa,rate=0.0),
    msprime.MassMigration(time=Tooa+0.0001,source=1,destination=0,proportion=1.0)
]

# initial population size
```

```
init_event = [
    msprime.PopulationParametersChange(time=Thum,initial_size=N0,population_id=0)
]

# cat all the events together
events = recent_event + ag_event + eu_event + ooa_event + init_event

# run the simulation
treeseq = msprime.simulate(population_configurations=pop_config, migration_matrix=mig_mat,
demographic_events=events, length=nbp, recombination_rate=rho, mutation_rate=mu, random_seed=seed)
# print results
with sys.stdout as vcffile:
        treeseq.write_vcf(vcffile,2) # 2 is for diploid
```