Supporting Information

Model-based detection and analysis of introgressed Neanderthal ancestry in modern humans

Matthias Steinrücken[1,2,3], Jeffrey P. Spence[4], John A. Kamm[5], Emilia Wieczorek[6], and Yun S. Song[3,5,7]

[1]Department of Ecology and Evolution, University of Chicago
[2]Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst
[3]Department of EECS, University of California, Berkeley
[4]Computational Biology Graduate Group, University of California, Berkeley
[5]Department of Statistics, University of California, Berkeley
[6]Department of Mathematics, University of California, Berkeley
[7]Chan Zuckerberg Biohub, San Francisco

February 14, 2018

S1 Gene Ontology Analysis Results

All Gene Ontology categories with an FDR q value, as reported by GOrilla (Eden et al., 2007, 2009), less than 0.05 are reported in Tables S1, S2, S3, and S4. We note that GOrilla does not properly account for spatial correlations or uncertainty in our mean posterior introgression probabilities and so the reported q values may be anticonservative.

GO Catgory	GO ID	Description	FDR q-value
Process:	GO:0050907	detection of chemical stimulus involved in sensory perception	2.72×10^{-28}
	GO:0050911	detection of chemical stimulus involved in sensory perception of smell	6.64×10^{-28}
	GO:0009593	detection of chemical stimulus	2.6×10^{-27}
	GO:0050906	detection of stimulus involved in sensory perception	1.21×10^{-24}
	GO:0051606	detection of stimulus	4.01×10^{-17}
	GO:0007186	G-protein coupled receptor signaling pathway	4.71×10^{-13}
	GO:0007608	sensory perception of smell	3.31×10^{-6}
	GO:0007606	sensory perception of chemical stimulus	8.09×10^{-5}
Function:	GO:0004984	olfactory receptor activity	3.98×10^{-28}
	GO:0004930	G-protein coupled receptor activity	5.59×10^{-17}
	GO:0005549	odorant binding	1.3×10^{-13}
	GO:0004888	transmembrane signaling receptor activity	5.34×10^{-10}
	GO:0099600	transmembrane receptor activity	4.51×10^{-10}
	GO:0038023	signaling receptor activity	4.71×10^{-9}
	GO:0004872	receptor activity	7.19×10^{-7}
	GO:0060089	molecular transducer activity	9.93×10^{-7}
	GO:0004871	signal transducer activity	1.76×10^{-6}
	GO:0045236	CXCR chemokine receptor binding	0.0273

Table S1: Gene ontology terms associated with lack of introgression in CHB+CHS

GO Catgory	GO ID	Description	FDR q-value
Process:	GO:0050907	detection of chemical stimulus involved in sensory perception	1.11×10^{-9}
	GO:0009593	detection of chemical stimulus	7.21×10^{-9}
	GO:0050906	detection of stimulus involved in sensory perception	1.41×10^{-8}
	GO:0050911	detection of chemical stimulus involved in sensory perception of smell	1.11×10^{-7}
	GO:0051606	detection of stimulus	2.7×10^{-5}
	GO:0007608	sensory perception of smell	1.34×10^{-3}
	GO:0007606	sensory perception of chemical stimulus	3.04×10^{-3}
	GO:0018149	peptide cross-linking	0.0178
Function:	GO:0004984	olfactory receptor activity	1.33×10^{-7}
	GO:0004930	G-protein coupled receptor activity	6.29×10^{-4}
	GO:0045236	CXCR chemokine receptor binding	1.44×10^{-3}
	GO:0033038	bitter taste receptor activity	0.0143

Table S2: Gene ontology terms associated with lack of introgression in CEU

GO Catgory	GO ID	Description	FDR q-value
Process:	GO:0071493	cellular response to UV-B	1.5×10^{-3}
	GO:0030214	hyaluronan catabolic process	$7.99 \times 10-3$
	GO:0010224	response to UV-B	0.0103
	GO:0045926	negative regulation of growth	0.0355
	GO:0060337	type I interferon signaling pathway	0.0314
	GO:0030212	hyaluronan metabolic process	0.0288
	GO:0033141	positive regulation of peptidyl-serine phosphorylation of STAT protein	0.038
	GO:0033139	regulation of peptidyl-serine phosphorylation of STAT protein	0.0424
	GO:0071482	cellular response to light stimulus	0.0452
	GO:0061099	negative regulation of protein tyrosine kinase activity	0.0409
Function:	GO:0004415	hyaluronoglucosaminidase activity	2.37×10^{-4}
	GO:0005132	type I interferon receptor binding	4.4×10^{-4}
	GO:0015929	hexosaminidase activity	1.2×10^{-3}
	GO:0033906	hyaluronoglucuronidase activity	1.14×10^{-3}
	GO:0031433	telethonin binding	0.0177

Table S3: Gene ontology terms associated with enrichment of introgression in CHB+CHS

GO Catgory	GO ID	Description	FDR q-value
Process:	GO:0050911	detection of chemical stimulus involved in sensory perception of smell	3.32×10^{-9}
	GO:0050907	detection of chemical stimulus involved in sensory perception	1.13×10^{-7}
	GO:0009593	detection of chemical stimulus	1.14×10^{-7}
	GO:0050906	detection of stimulus involved in sensory perception	1.14×10^{-6}
	GO:0051606	detection of stimulus	2.7×10^{-5}
	GO:0007186	G-protein coupled receptor signaling pathway	2.65×10^{-3}
	GO:0006342	chromatin silencing	0.0399
Function:	GO:0005549	odorant binding	9.43×10^{-13}
	GO:0004984	olfactory receptor activity	4.97×10^{-10}
	GO:0001730	2'-5'-oligoadenylate synthetase activity	5.41×10^{-6}
	GO:0004930	G-protein coupled receptor activity	9.2×10^{-5}
	GO:0004950	chemokine receptor activity	0.0122
	GO:0001637	G-protein coupled chemoattractant receptor activity	0.0102
	GO:0015125	bile acid transmembrane transporter activity	0.0104
	GO:0070566	adenylyltransferase activity	0.033
	GO:0004715	non-membrane spanning protein tyrosine kinase activity	0.0421
Component:	GO:0000786	nucleosome	2.49×10^{-7}
	GO:0044815	DNA packaging complex	7.36×10^{-7}
	GO:0032993	protein-DNA complex	2.49×10^{-4}
	GO:0017101	aminoacyl-tRNA synthetase multienzyme complex	5.27×10^{-3}
	GO:0045095	keratin filament	0.0287

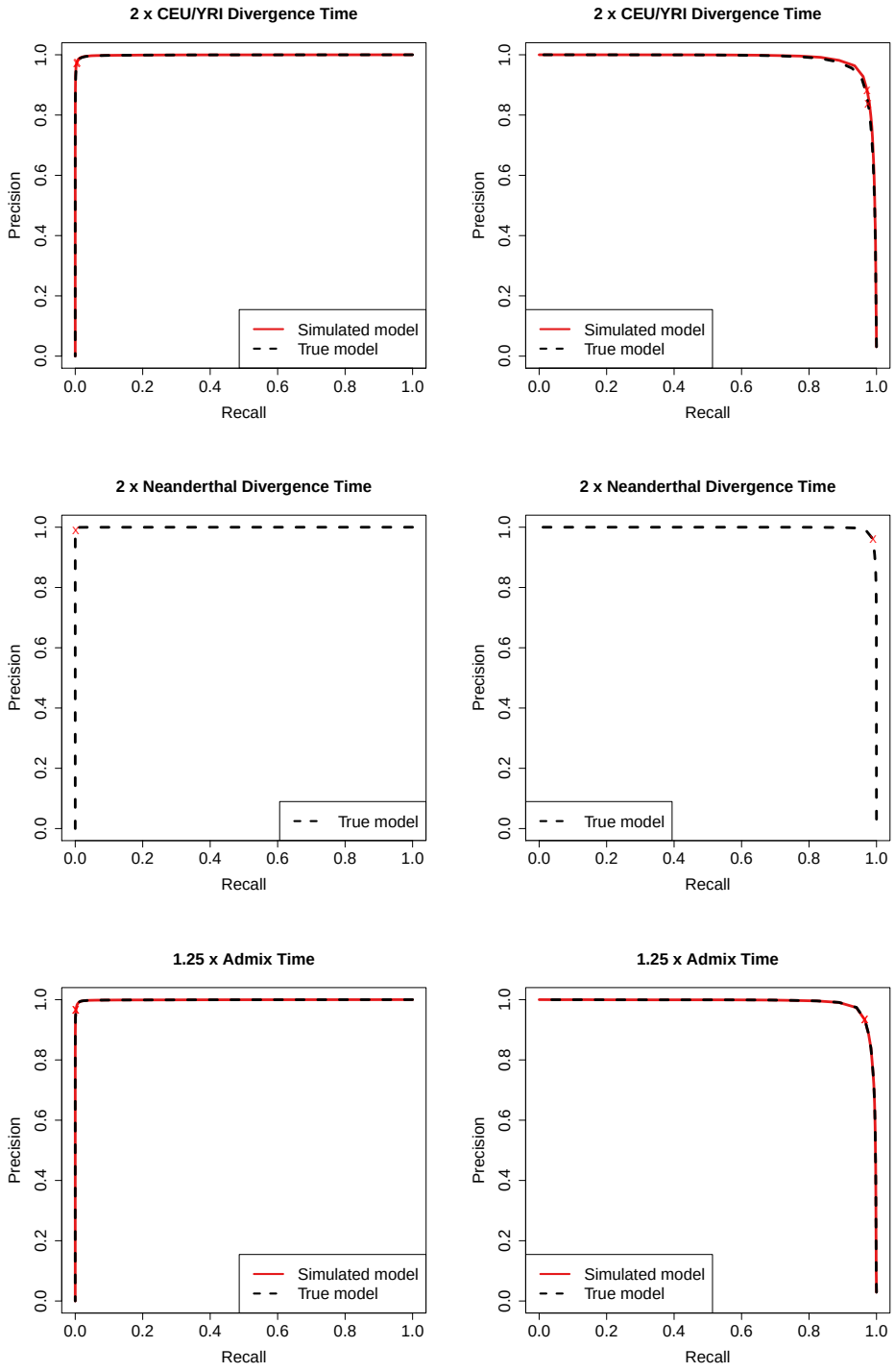
Table S4: Gene ontology terms associated with enrichment of introgression in CEU

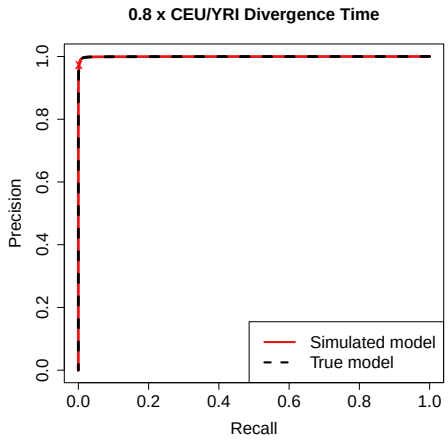
S2 Simulation Study

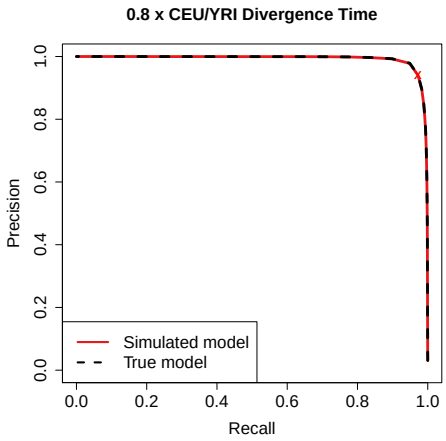
Here we present some results of the simulation study we performed to assess the accuracy of our method.

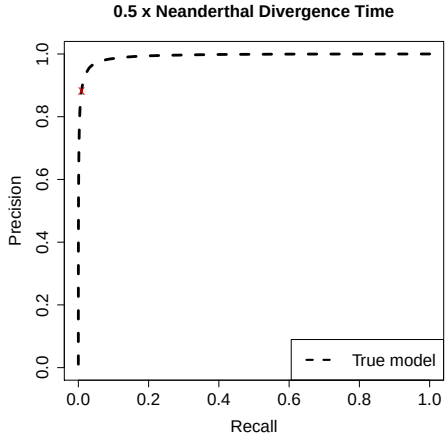
S2.1 ROC and Precision-Recall Curves from Simulated Data

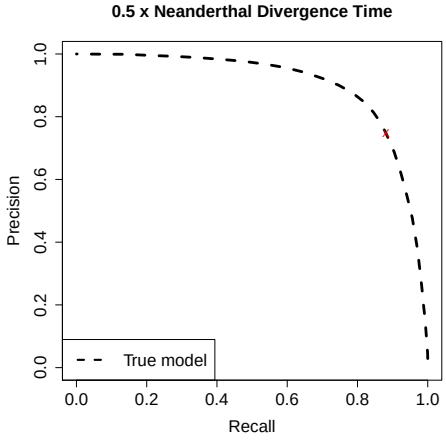
ROC and Precision-recall curves for data simulated under different models, and analyzed using the same model as simulated and the "true" model.

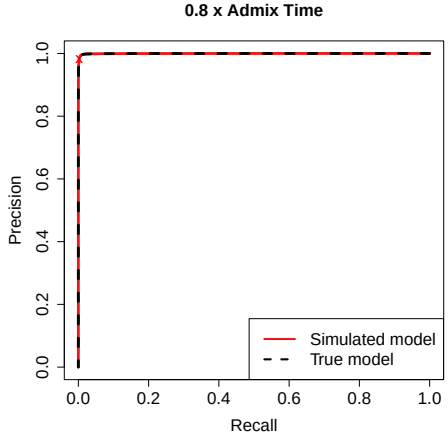


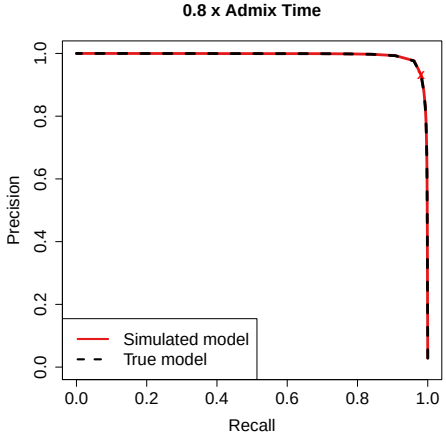

0.8 x CEU/YRI Divergence Time
Precision
Recall
Simulated model
True model

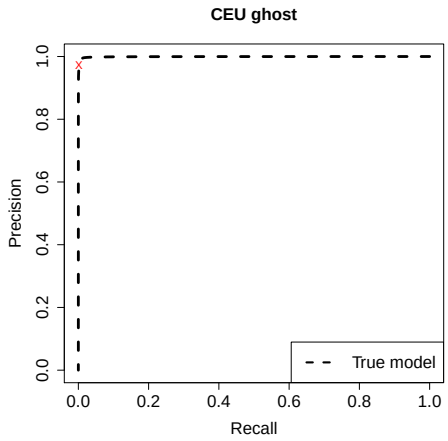

0.8 x CEU/YRI Divergence Time
Precision
Recall
Simulated model
True model

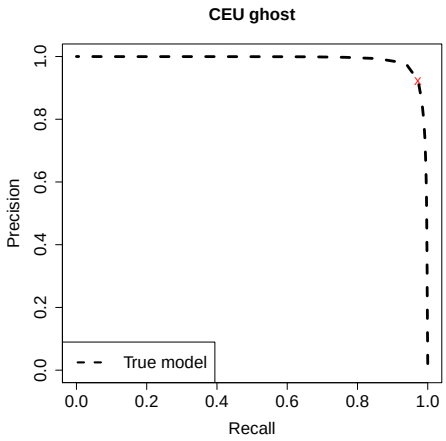

0.5 x Neanderthal Divergence Time
Precision
Recall
True model

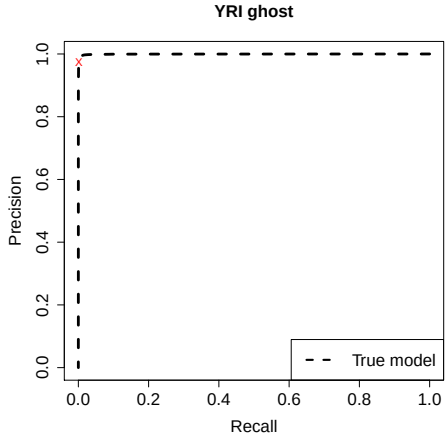

0.5 x Neanderthal Divergence Time
Precision
Recall
True model

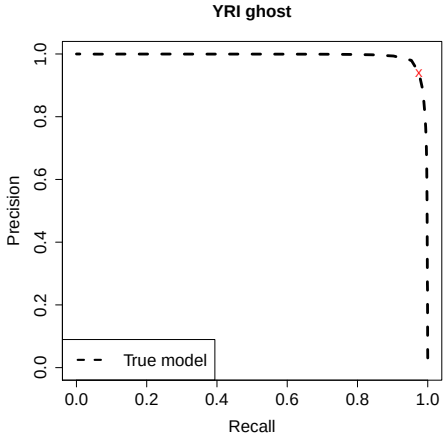

0.8 x Admix Time
Precision
Recall
Simulated model
True model

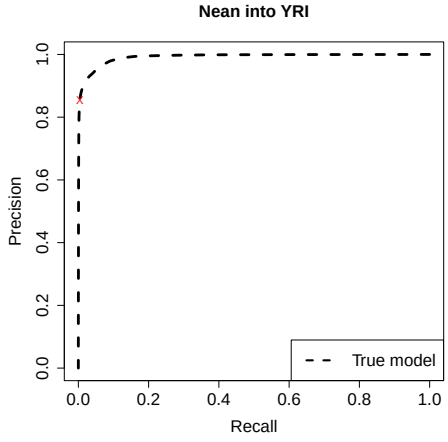

0.8 x Admix Time
Precision
Recall
Simulated model
True model

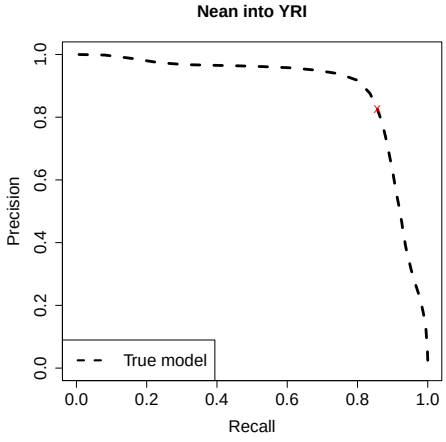

CEU ghost
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0

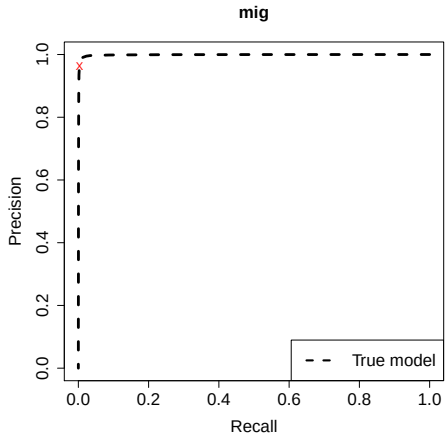

CEU ghost
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0

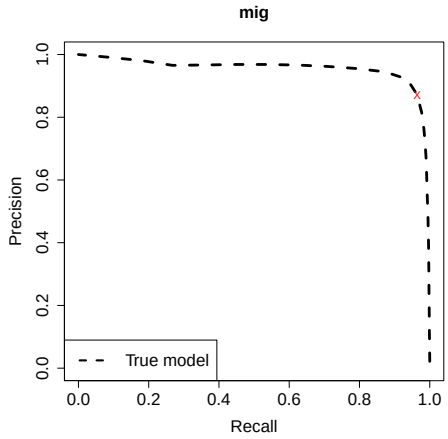

YRI ghost
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0

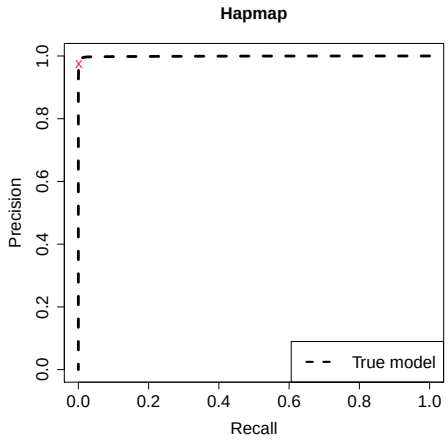

YRI ghost
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0

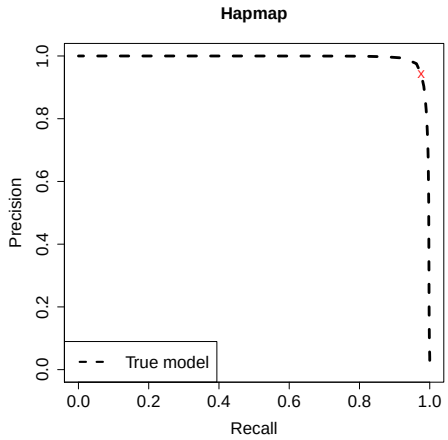

Nean into YRI
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0


Nean into YRI
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0


mig
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0

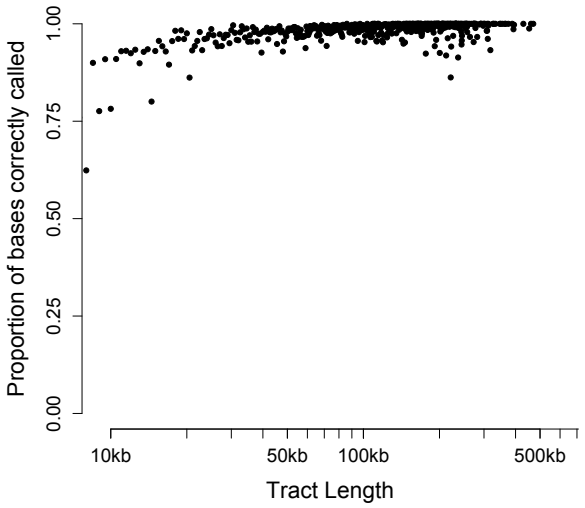

mig
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0


Hapmap
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0


Hapmap
Precision
Recall
True model
0.0
0.2
0.4
0.6
0.8
1.0

S2.2 Power as a function of tract length

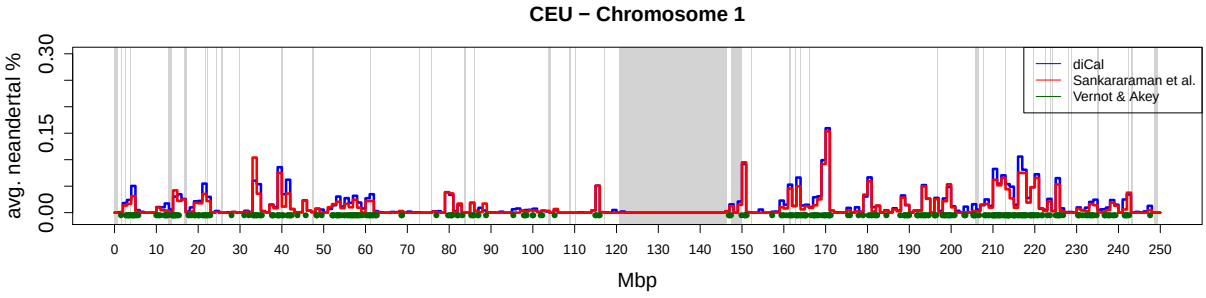
Here we use the marginal posterior obtained from analyzing the data simulated under the true "model", using the "true" model in the analysis. We categorized the introgressed fragments by their length, and plot the percentage of correctly called bases for tracts of a certain length:

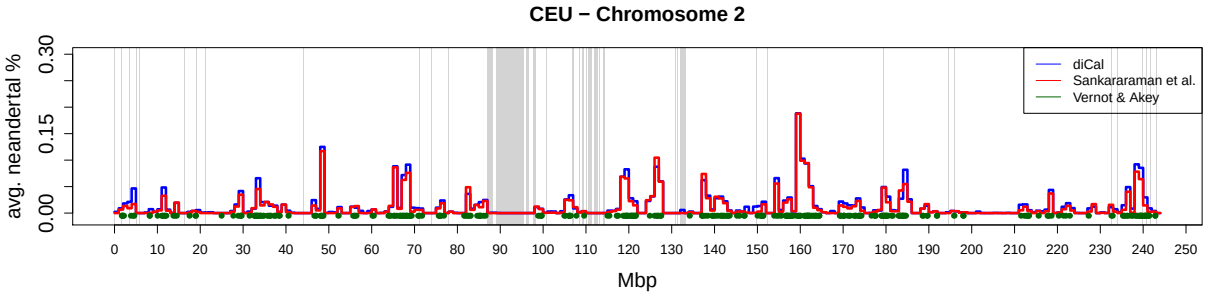


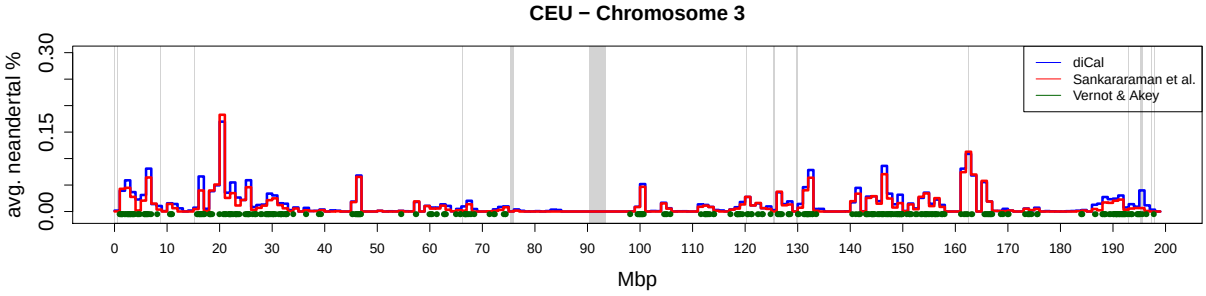
S3 Fine-scale Population Average Introgression

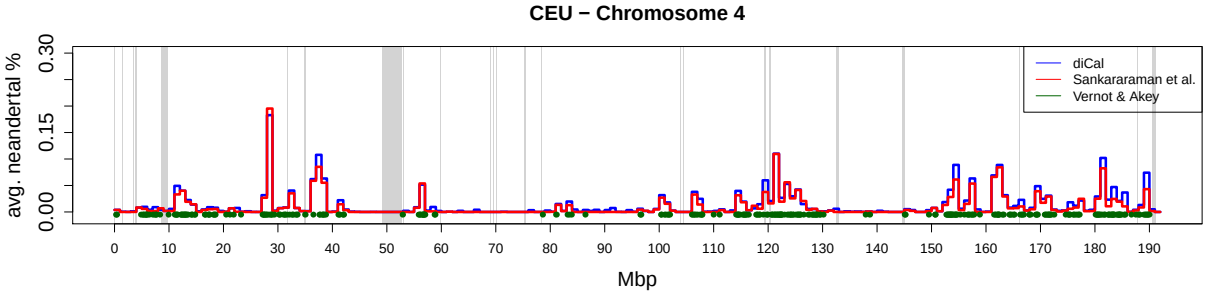
S3.1 CEU

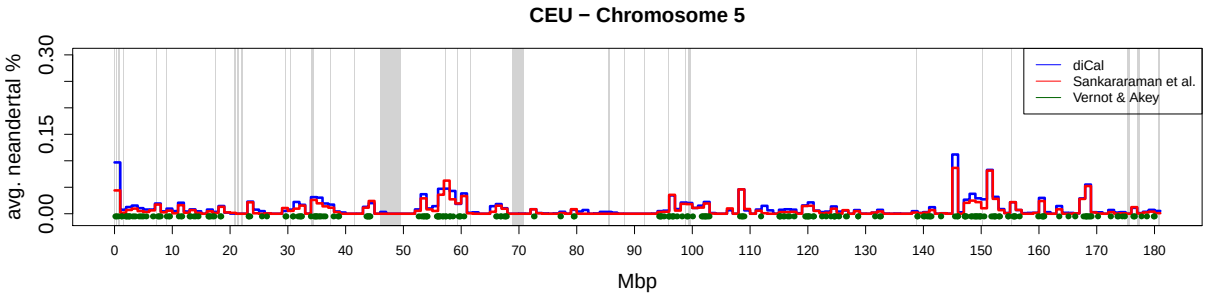
Skyline plot of the amount of Neanderthal introgression in the CEU population on the different chromosomes, averaged over all individuals in 1 Mbp windows. The results from *diCal* are indicated in blue, and the results from Sankararaman et al. (2014) indicated in red. The regions reported as introgressed by Vernot and Akey (2014) are indicated in green. The gray bars denote the regions were no calls were made in the 1000 genomes dataset, which include the centromeres.

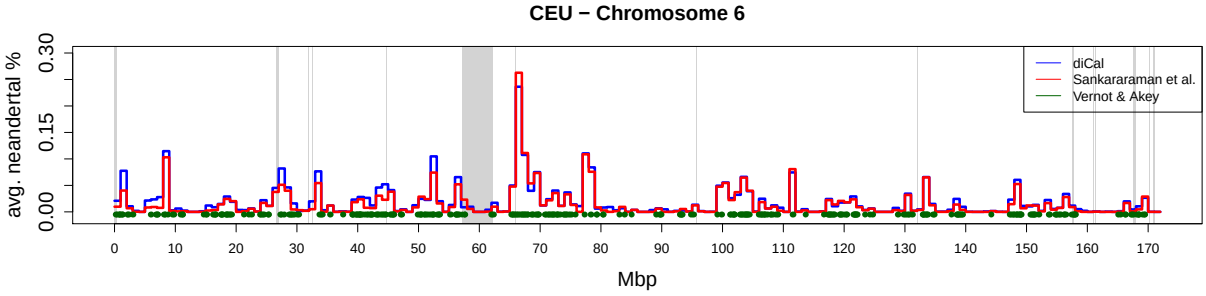


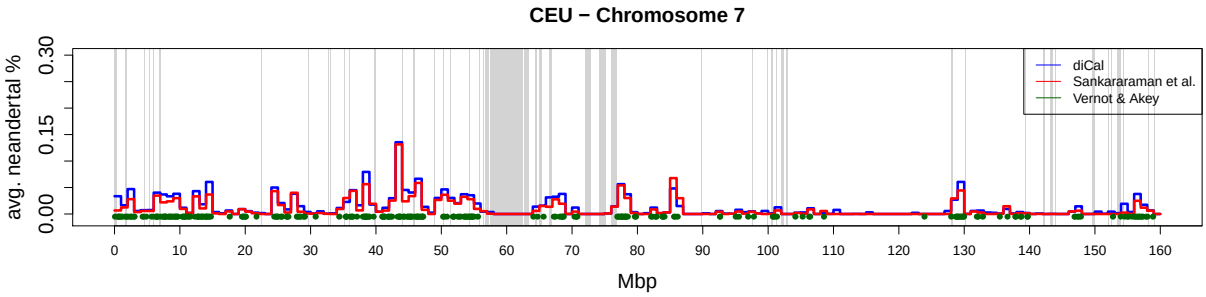


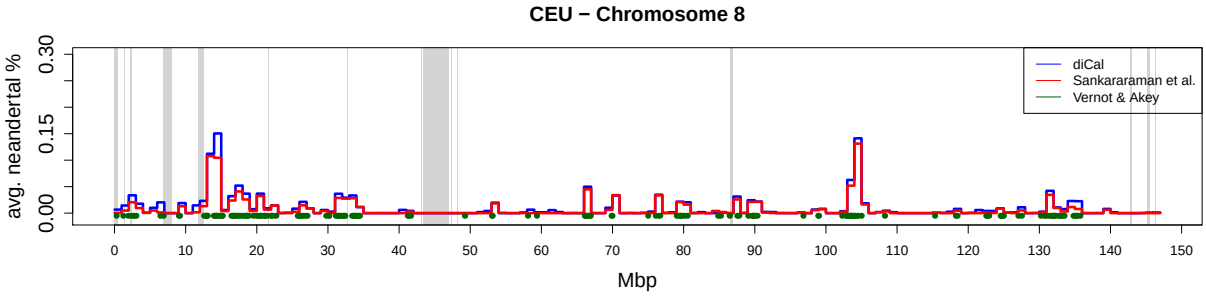


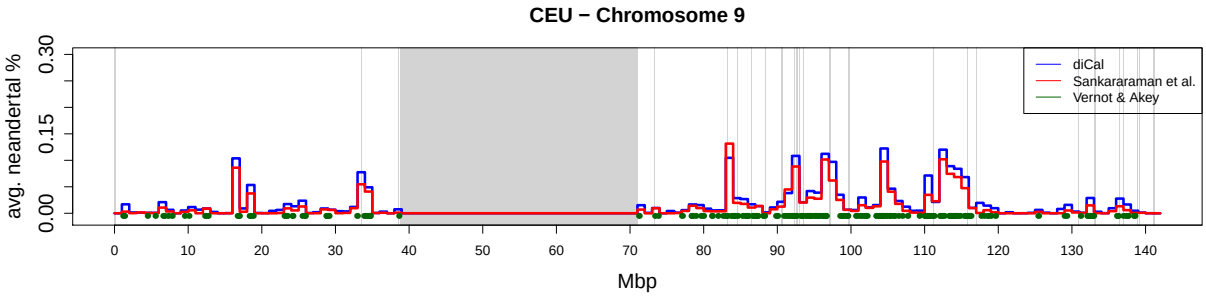


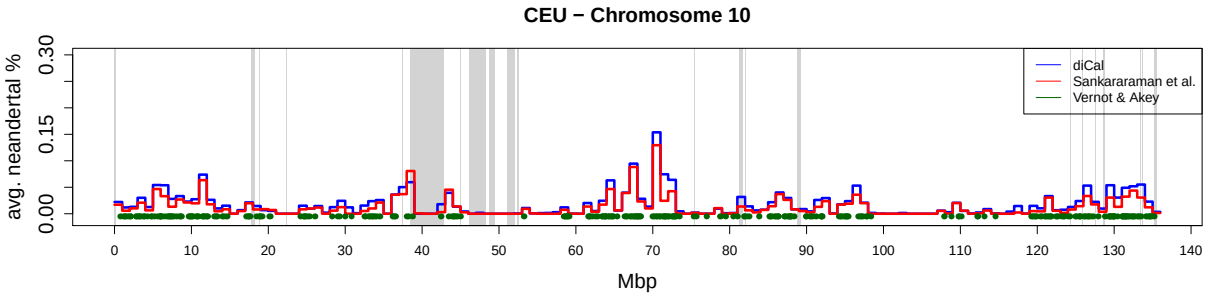

CEU − Chromosome 5
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180
Mbp
diCal
Sankararaman et al.
Vernot & Akey

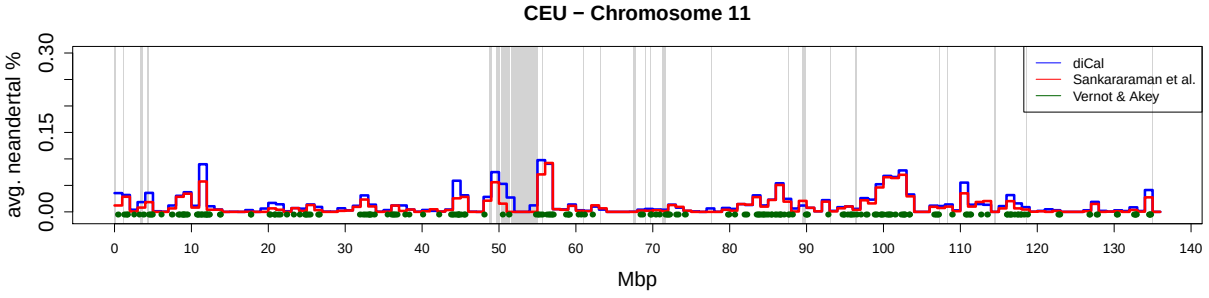

CEU − Chromosome 6
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170
Mbp
diCal
Sankararaman et al.
Vernot & Akey

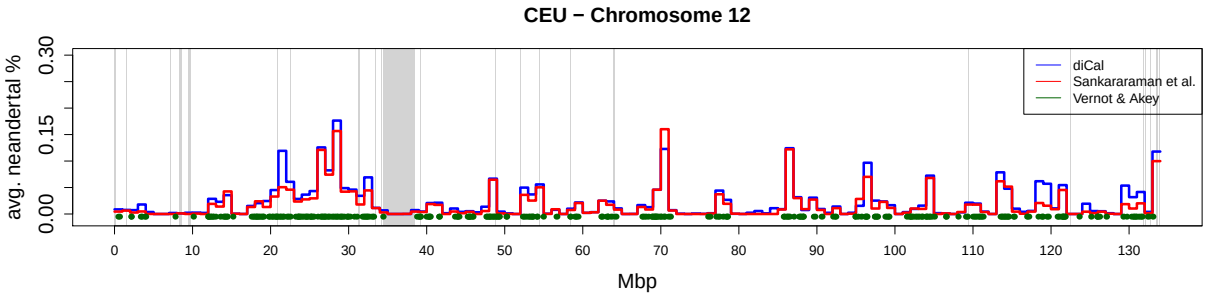

CEU − Chromosome 7
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160
Mbp
diCal
Sankararaman et al.
Vernot & Akey

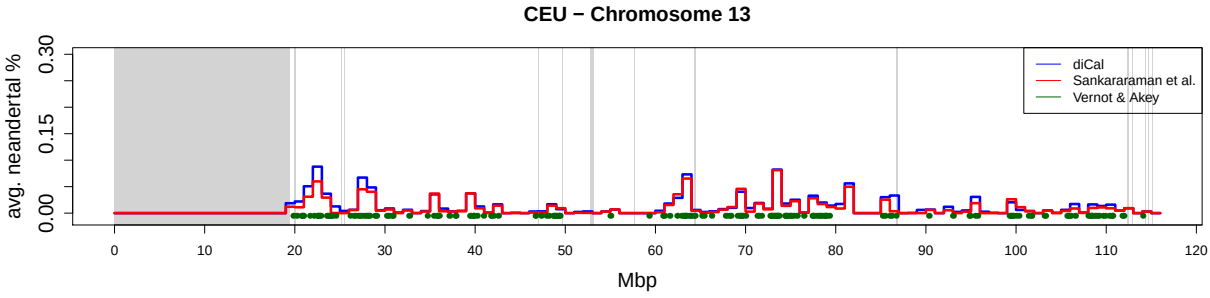

CEU − Chromosome 8
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
Mbp
diCal
Sankararaman et al.
Vernot & Akey

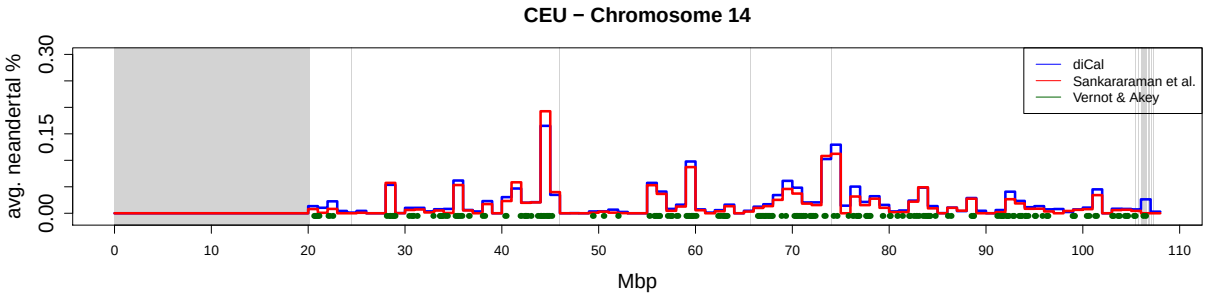

CEU − Chromosome 9
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
Mbp
diCal
Sankararaman et al.
Vernot & Akey

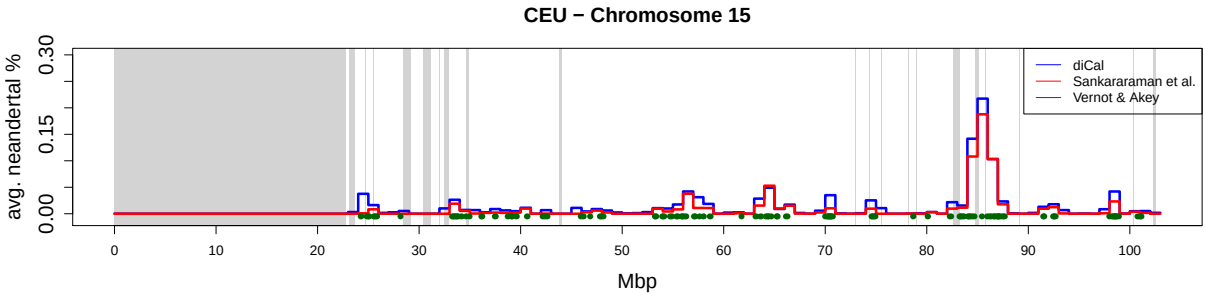

CEU − Chromosome 10
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
Mbp

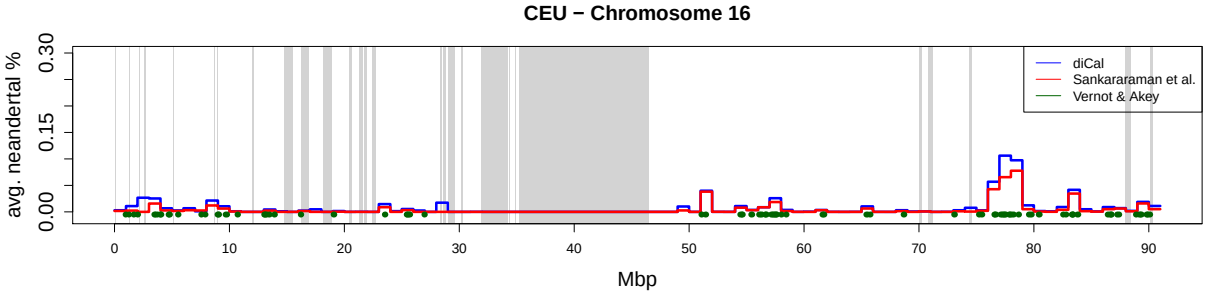

CEU − Chromosome 11
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
Mbp

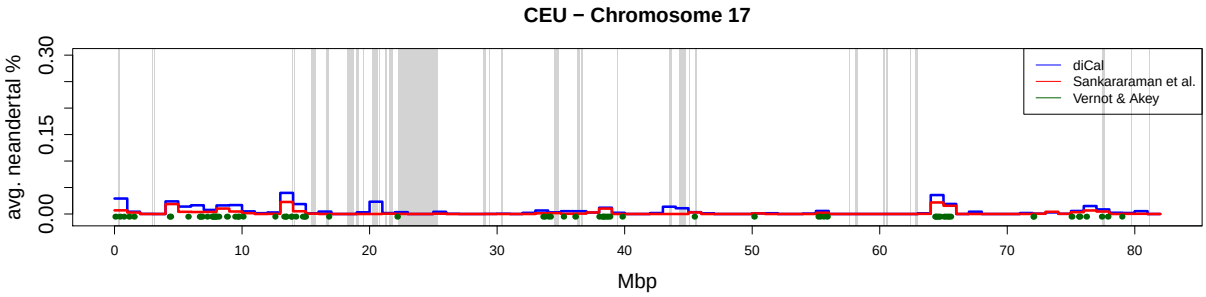

CEU − Chromosome 12
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0
10
20
30
40
50
60
70
80
90
100
110
120
130
Mbp

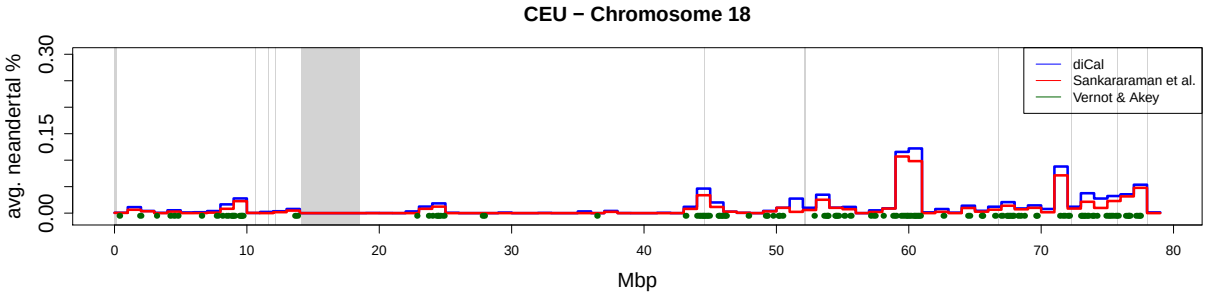

CEU − Chromosome 13
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0
10
20
30
40
50
60
70
80
90
100
110
120
Mbp

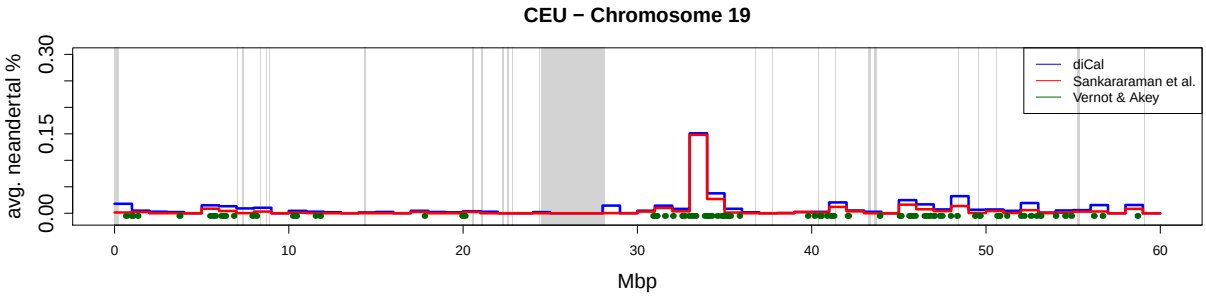

CEU − Chromosome 14
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0
10
20
30
40
50
60
70
80
90
100
110
Mbp

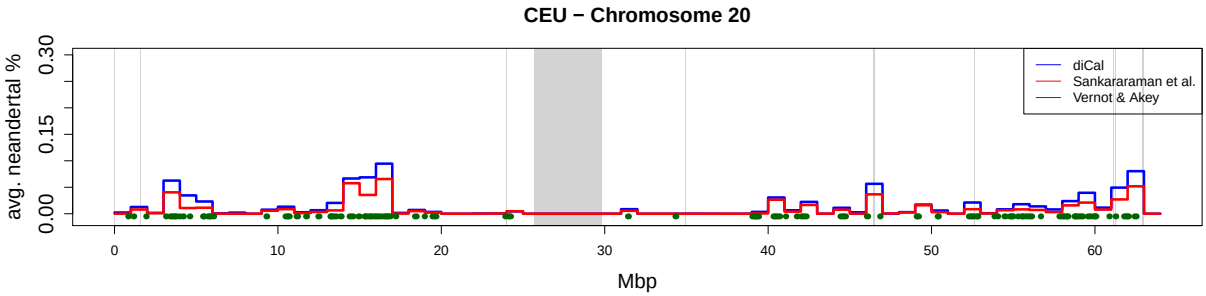

CEU − Chromosome 15
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
90
100
Mbp
diCal
Sankararaman et al.
Vernot & Akey

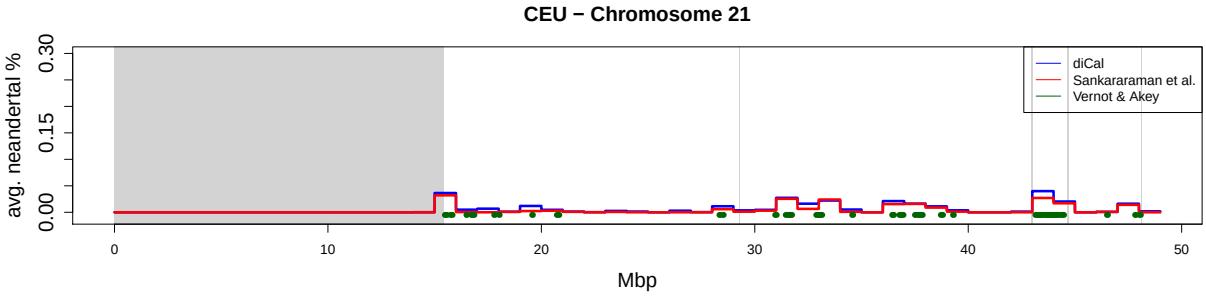

CEU − Chromosome 16
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
90
Mbp
diCal
Sankararaman et al.
Vernot & Akey

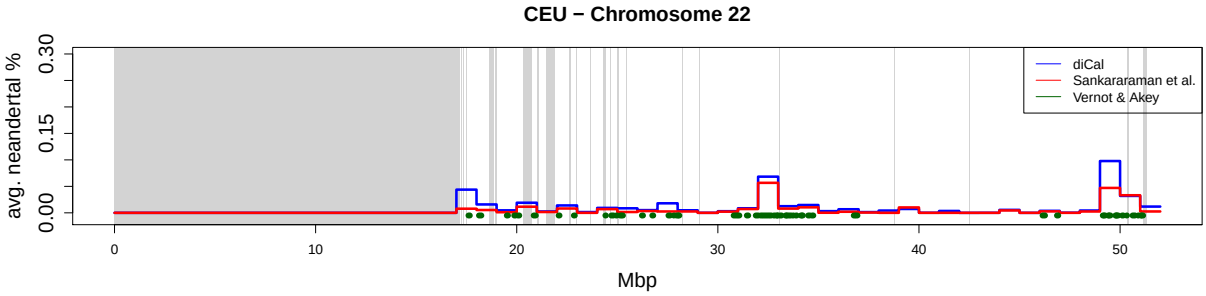

CEU − Chromosome 17
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
Mbp
diCal
Sankararaman et al.
Vernot & Akey

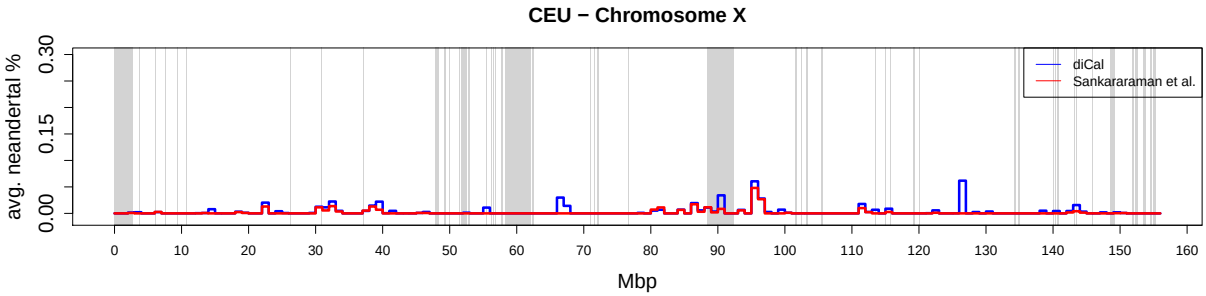

CEU − Chromosome 18
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
Mbp
diCal
Sankararaman et al.
Vernot & Akey


CEU − Chromosome 19
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
Mbp
diCal
Sankararaman et al.
Vernot & Akey


CEU − Chromosome 20
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
Mbp
diCal
Sankararaman et al.
Vernot & Akey

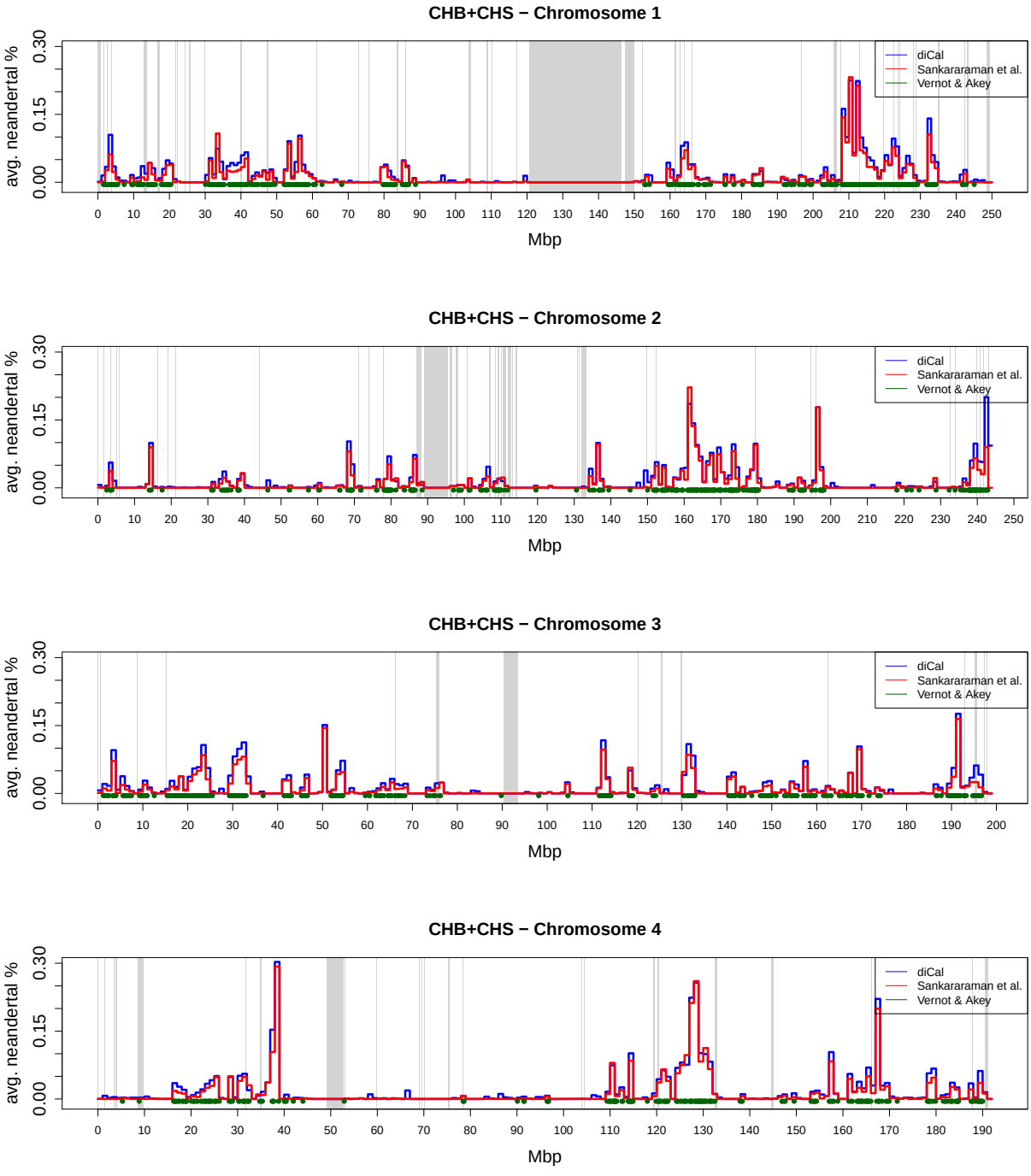

CEU − Chromosome 21
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
Mbp
diCal
Sankararaman et al.
Vernot & Akey

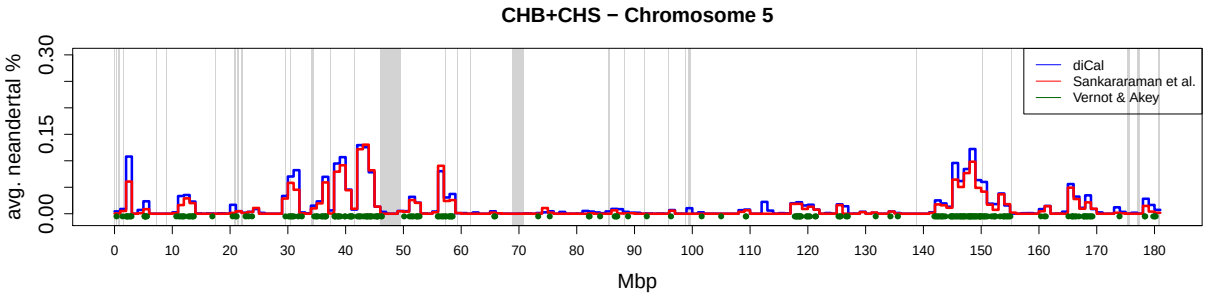

CEU − Chromosome 22
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
Mbp
diCal
Sankararaman et al.
Vernot & Akey

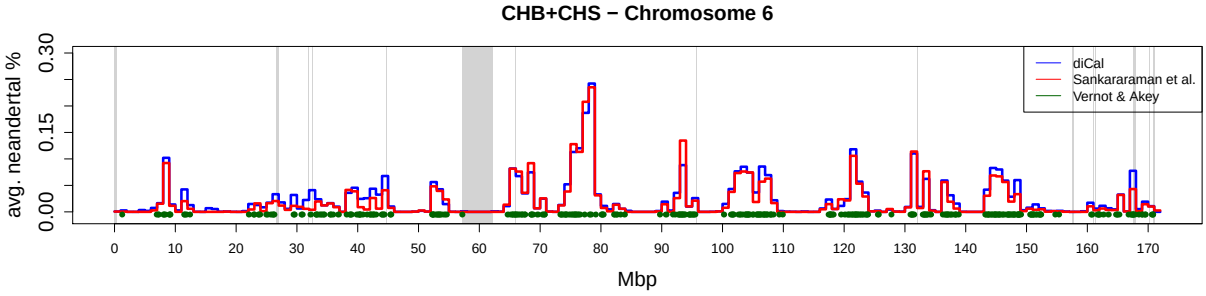

CEU − Chromosome X
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
Mbp
diCal
Sankararaman et al.

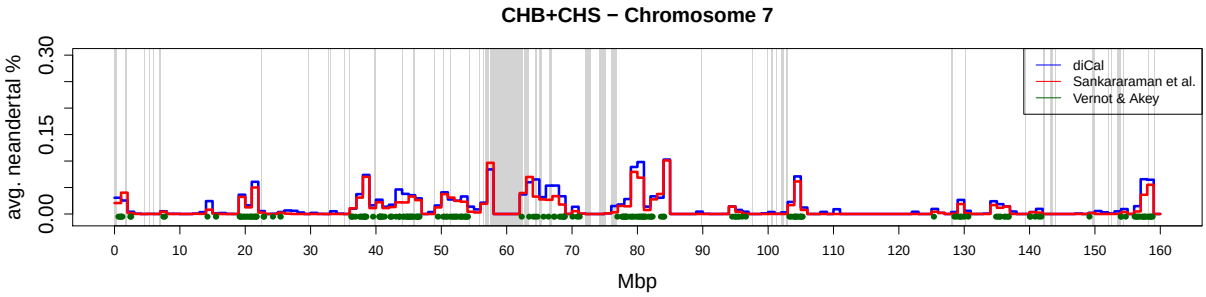
S3.2 CHB+CHS

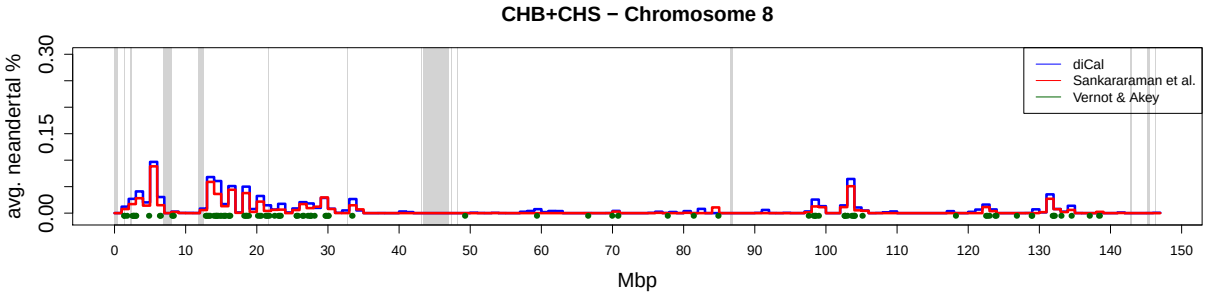
Skyline plot of the amount of Neanderthal introgression in the individuals of the CHB and CHS population on the different chromosomes, averaged over all individuals in 1 Mbp windows. The results from *diCal* are indicated in blue, and the results from Sankararaman et al. (2014) indicated in red. The regions reported as introgressed by Vernot and Akey (2014) are indicated in green. The gray bars denote the regions were no calls were made in the 1000 genomes dataset, which include the centromeres.

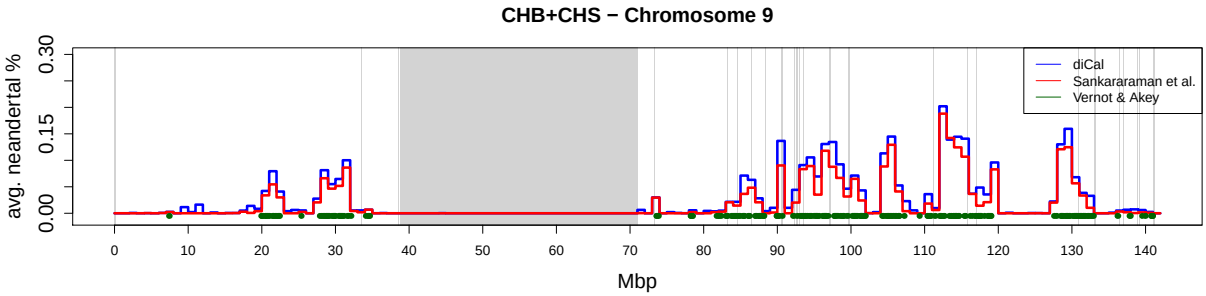


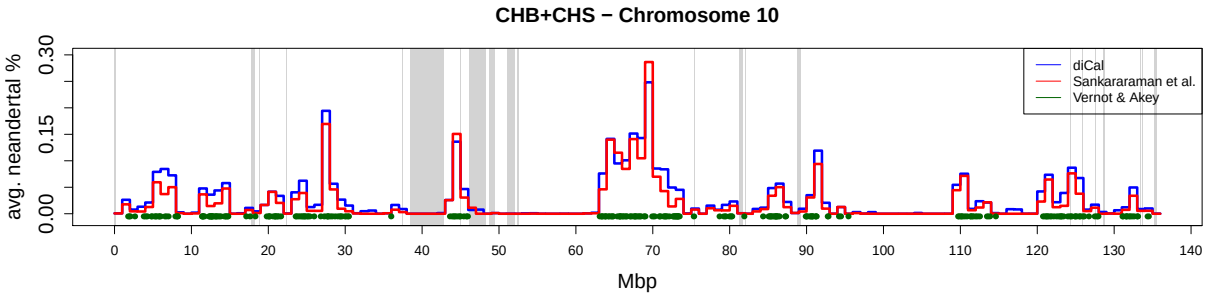

CHB+CHS − Chromosome 5
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180
Mbp
diCal
Sankararaman et al.
Vernot & Akey

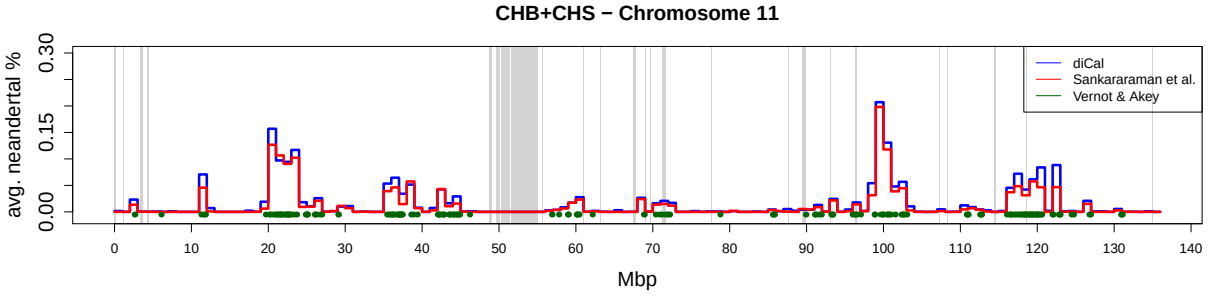

CHB+CHS − Chromosome 6
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170
Mbp
diCal
Sankararaman et al.
Vernot & Akey

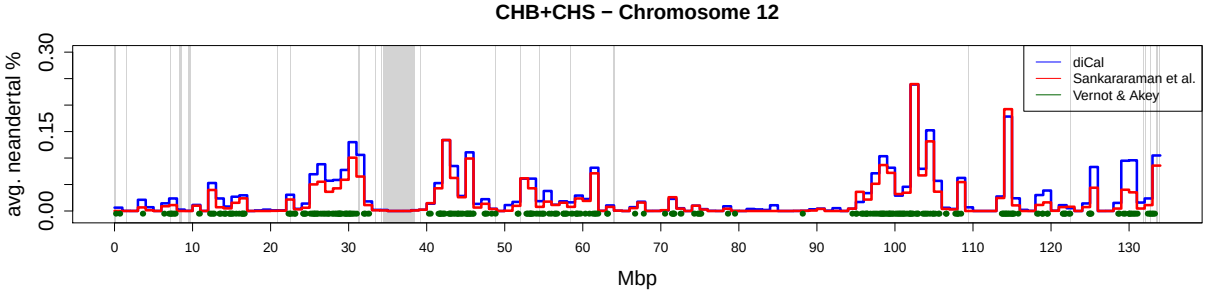

CHB+CHS − Chromosome 7
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160
Mbp
diCal
Sankararaman et al.
Vernot & Akey

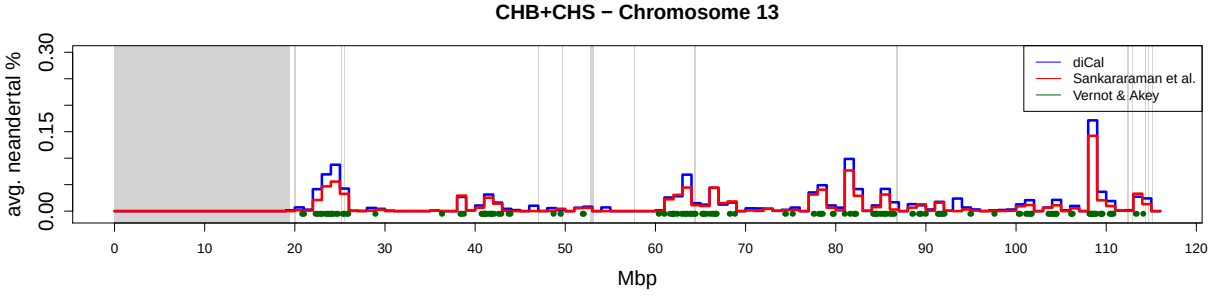

CHB+CHS − Chromosome 8
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
Mbp
diCal
Sankararaman et al.
Vernot & Akey

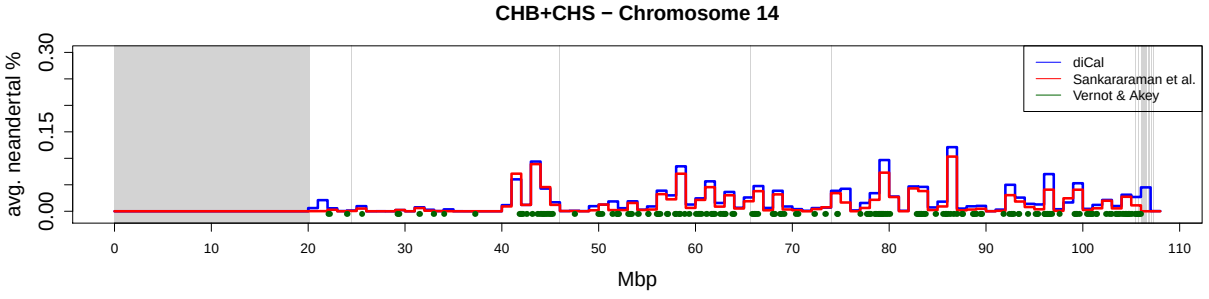

CHB+CHS − Chromosome 9
avg. neandertal %
0.00
0.15
0.30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
Mbp
diCal
Sankararaman et al.
Vernot & Akey

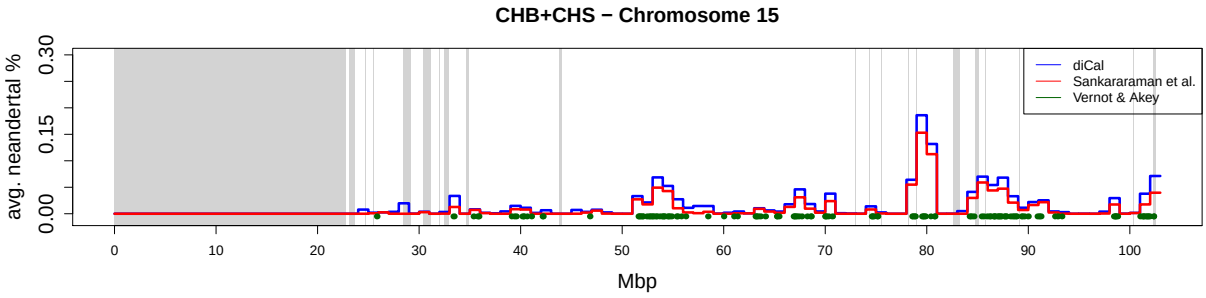

CHB+CHS − Chromosome 10
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
Mbp

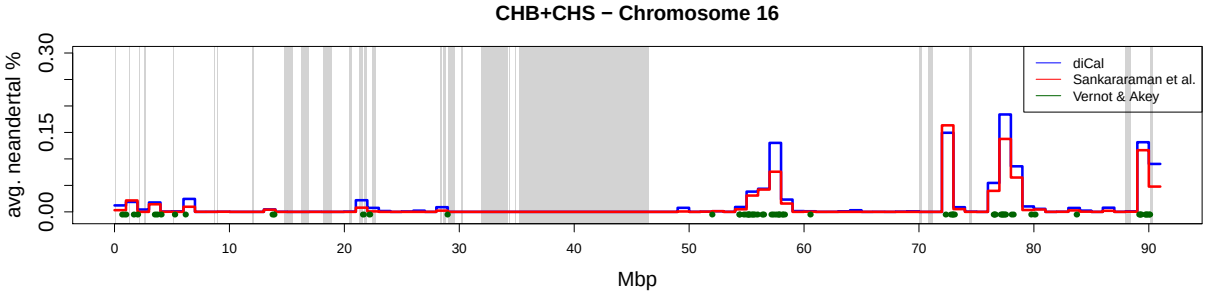

CHB+CHS − Chromosome 11
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
Mbp

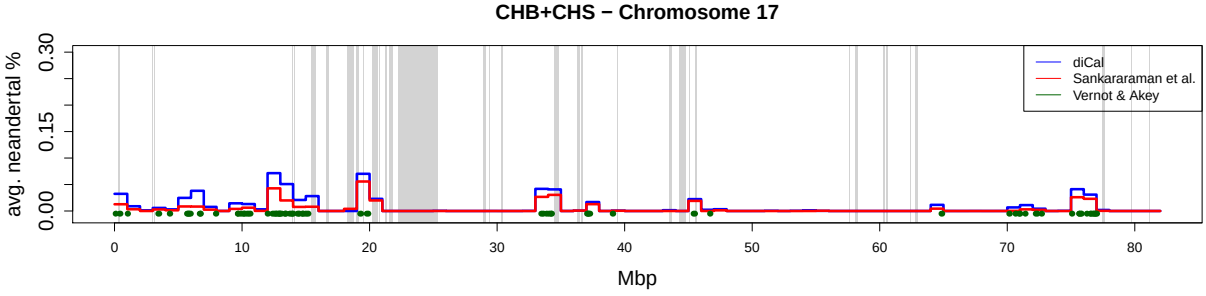

CHB+CHS − Chromosome 12
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0 10 20 30 40 50 60 70 80 90 100 110 120 130
Mbp

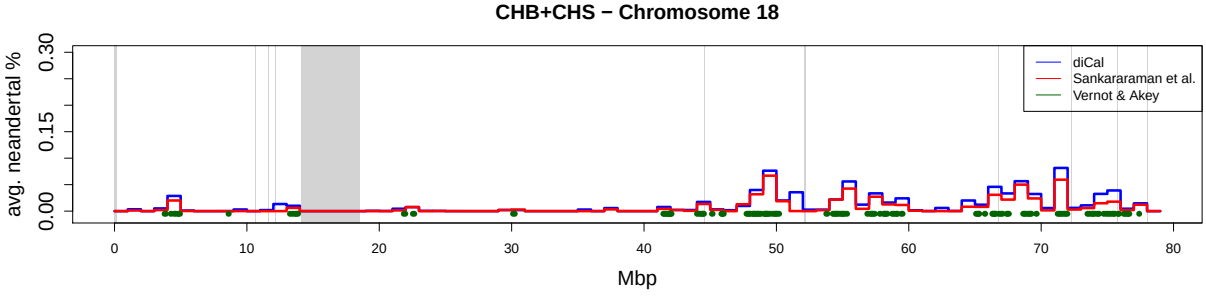

CHB+CHS − Chromosome 13
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0 10 20 30 40 50 60 70 80 90 100 110 120
Mbp

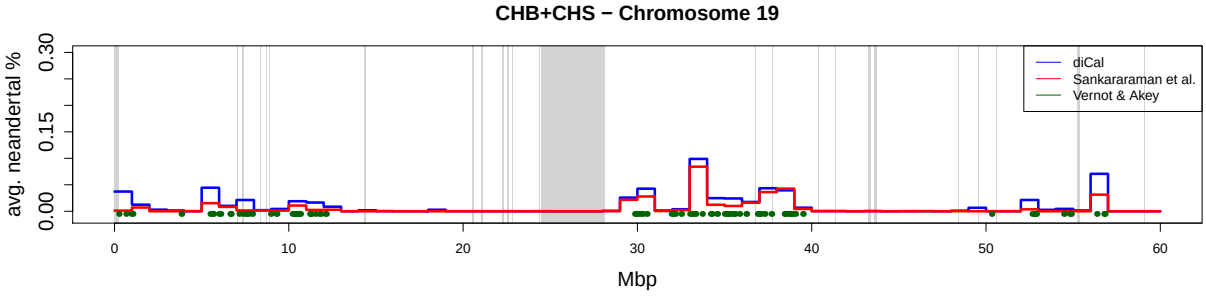

CHB+CHS − Chromosome 14
avg. neandertal %
0.00
0.15
0.30
diCal
Sankararaman et al.
Vernot & Akey
0 10 20 30 40 50 60 70 80 90 100 110
Mbp

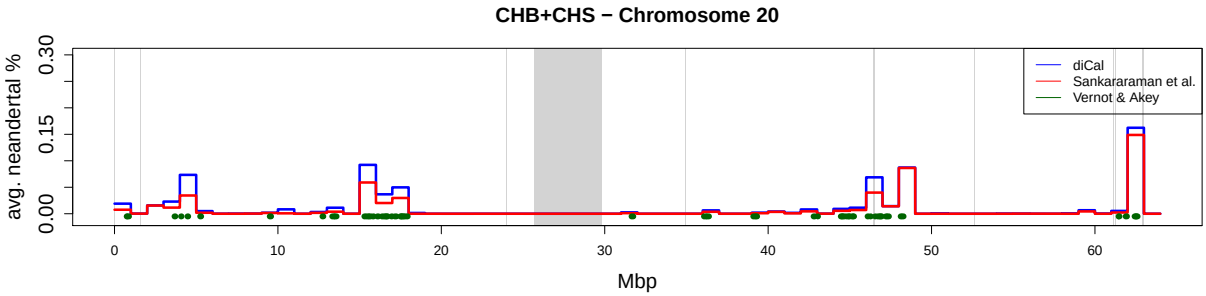

CHB+CHS − Chromosome 15
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
90
100
Mbp
diCal
Sankararaman et al.
Vernot & Akey

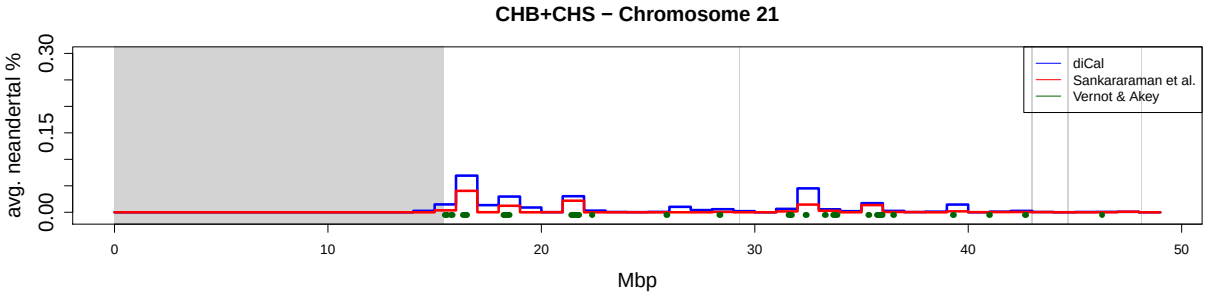

CHB+CHS − Chromosome 16
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
90
Mbp
diCal
Sankararaman et al.
Vernot & Akey

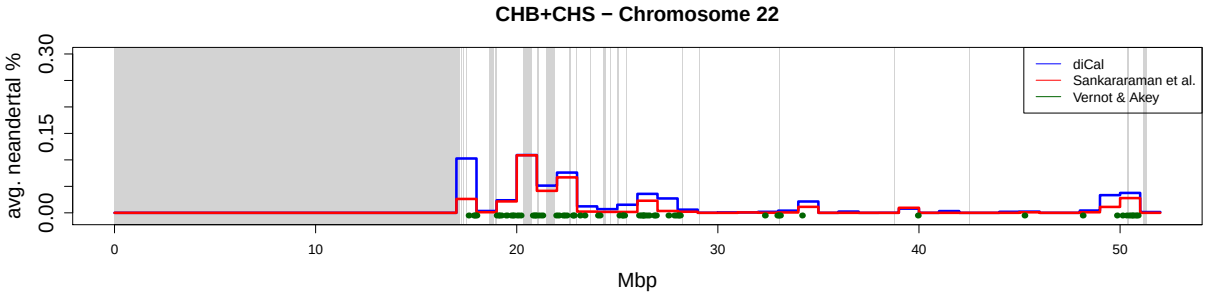

CHB+CHS − Chromosome 17
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
Mbp
diCal
Sankararaman et al.
Vernot & Akey

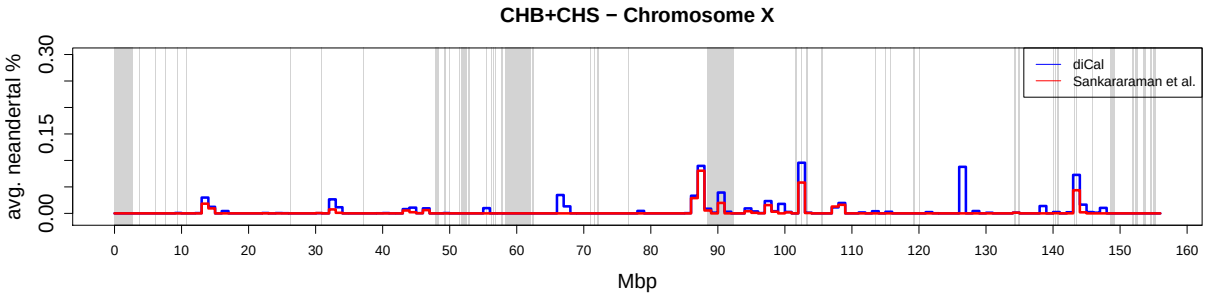

CHB+CHS − Chromosome 18
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
Mbp
diCal
Sankararaman et al.
Vernot & Akey


CHB+CHS − Chromosome 19
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
Mbp
diCal
Sankararaman et al.
Vernot & Akey


CHB+CHS − Chromosome 20
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
Mbp
diCal
Sankararaman et al.
Vernot & Akey


CHB+CHS − Chromosome 21
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
Mbp
diCal
Sankararaman et al.
Vernot & Akey


CHB+CHS − Chromosome 22
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
Mbp
diCal
Sankararaman et al.
Vernot & Akey


CHB+CHS − Chromosome X
avg. neandertal %
0.00
0.15
0.30
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
Mbp
diCal
Sankararaman et al.

References

Eden, E., Lipson, D., Yogev, S., and Yakhini, Z. (2007). Discovering motifs in ranked lists of dna sequences. *PLOS Computational Biology*, **3,**(3) 1–15.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). Gorilla: a tool for discovery and visualization of enriched go terms in ranked gene lists. *BMC Bioinformatics*, **10,**(1) 48.

Sankararaman, S., Mallick, S., Dannemann, M., Prufer, K., Kelso, J., Paabo, S., Patterson, N., and Reich, D. (2014). The genomic landscape of neanderthal ancestry in present-day humans. *Nature*, **507,**(7492) 354–357.

Vernot, B. and Akey, J. M. (2014). Resurrecting surviving neandertal lineages from modern human genomes. *Science*, **343,**(6174) 1017–1021.