CiBreed
Center for Integrated Breeding Research

Factors affecting phasing quality in a commercial layer population

N. Frioni[1], D. Cavero[2], H. Simianer[1]& M. Erbe[3]

[1] University of Goettingen, Department of Animal Sciences, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Goettingen, Germany

[2] H&N International, Cuxhaven, Germany

[3] Bavarian State Research Centre for Agriculture, Institute for Animal Breeding, Grub, Germany

WCGALP 2018. Auckland – New Zealand

Paper no.: 830

Importance of haplotype phasing

- Inferring points and number of recombination
- Imputation of low frequency variants
- Study of haplotype structure
  - genetic diversity and
  - accuracy in genomic selection
- Computational approaches with phasing software
  - sample size, marker density
  - genotype accuracy
  - relatedness in the sample

Objective

- Aim: Assessing haplotype phasing quality for a highly-related laying hen population using different phasing software
  - FImpute v2.2 (with and without pedigree information)
    (Sargolzaei et al., 2014)
  - Beagle v3.3 (without pedigree information)
    (Browning and Browning, 2007)
  - Beagle v4.1
    (Browning and Browning, 2007)

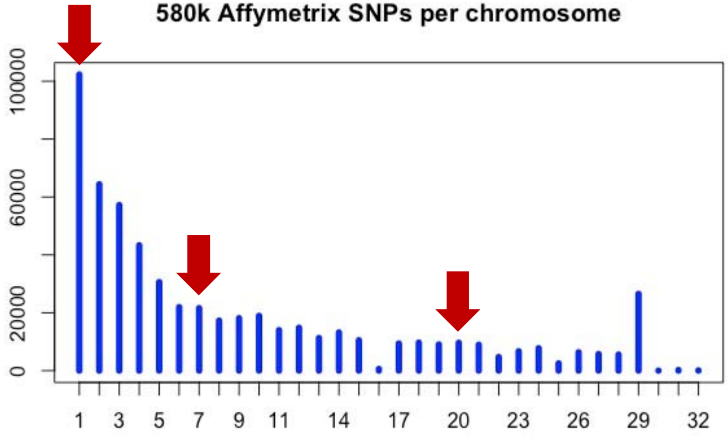
Materials

- Real SNP data from a purebred line of brown layers
- A pedigree of 53’882 individuals across 13 generations
- 918 genotyped individuals (belonging to gen. 7 to 12)
- SNPs from the 580k Affymetrix Axiom® Genome-Wide Chicken Array

Methods – Editing criteria

- Editing was done with Plink *(Purcell et al., 2007)*
  - Individuals with a call rate < 90%,
  - monomorphic SNPs and
  - SNPs not in HWE with $p<10^{-8}$

  were removed.
- 888 individuals remained in the dataset.
- We used chromosomes 1, 7 and 20.
  (77’910, 16’059 and 7’004 SNPs)



Methods - Simulation

- We performed a simulation in order to have known phases from real SNP data.

- Simulation in three steps:

  1. From the 888 individuals’ genotypes, haplotypes were derived in-silico to create a library of haplotypes.

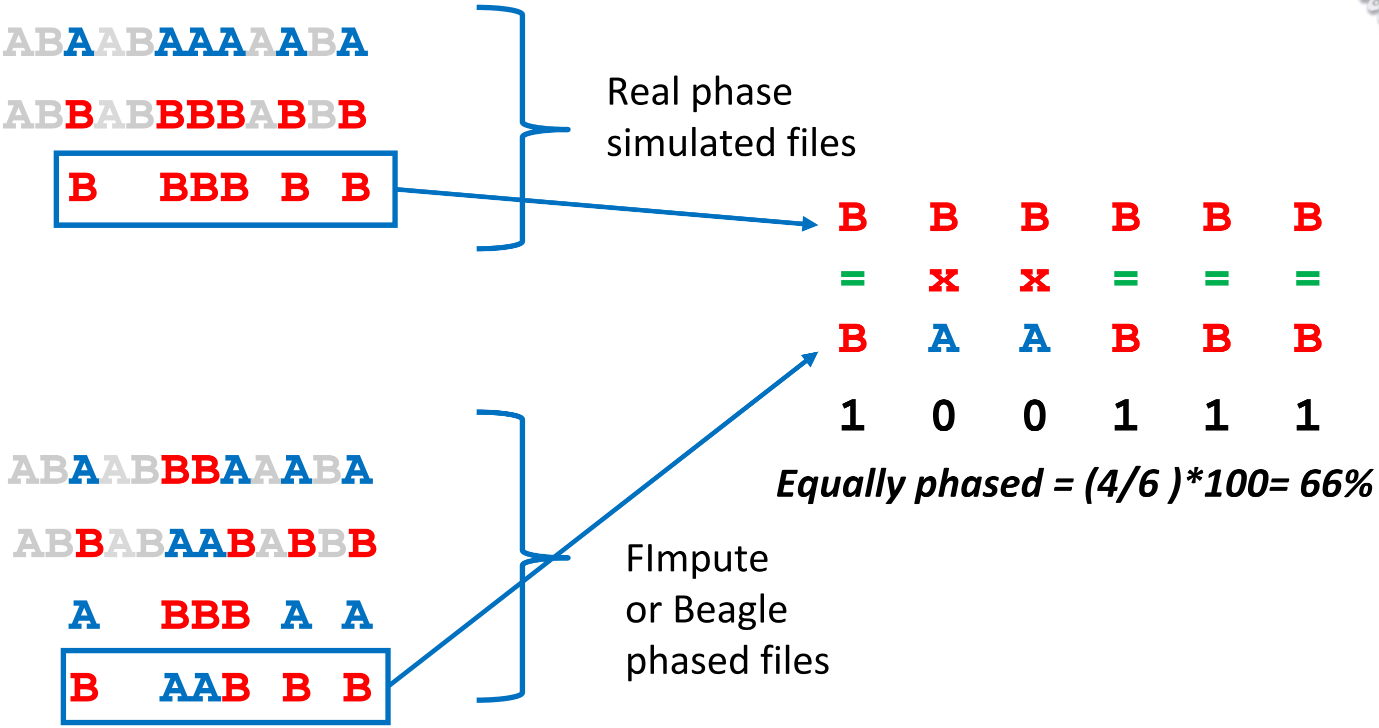
  2. 2’000 randomly sampled from the library and assigned to the founders of the pedigree.

  3. Founders’ haplotypes were dropped through the pedigree assuming random crossing-over events but no mutations.

Methods - Calculations

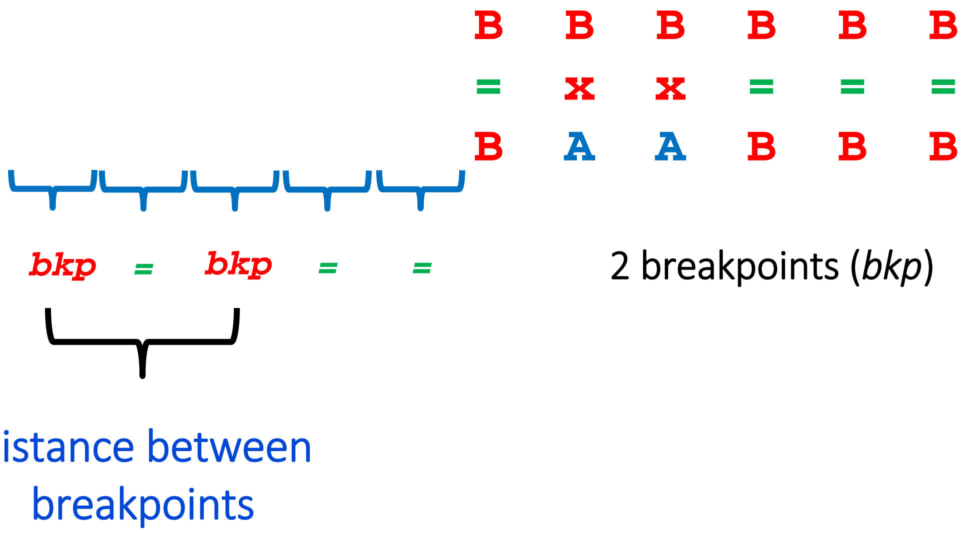
- Phasing with four approaches:
  - FImpute 2.2 (with pedigree information)
  - FImpute 2.2 (without pedigree information)
  - Beagle v 3.3
  - Beagle v 4.1
- Comparison of true simulated and in-silico phased haplotypes:
  - Proportion of equally phased heterozygous SNPs
  - Number of breakpoints (change of phase)
  - Segment size between breakpoints

Methods – Calculation of equally phased SNPs



Equally phased = (4/6)*100= 66%

Methods – Breakpoint estimation



B B B B B B

= x x = = =

B A A B B B

bkp = *bkp* = =

2 breakpoints (*bkp*)

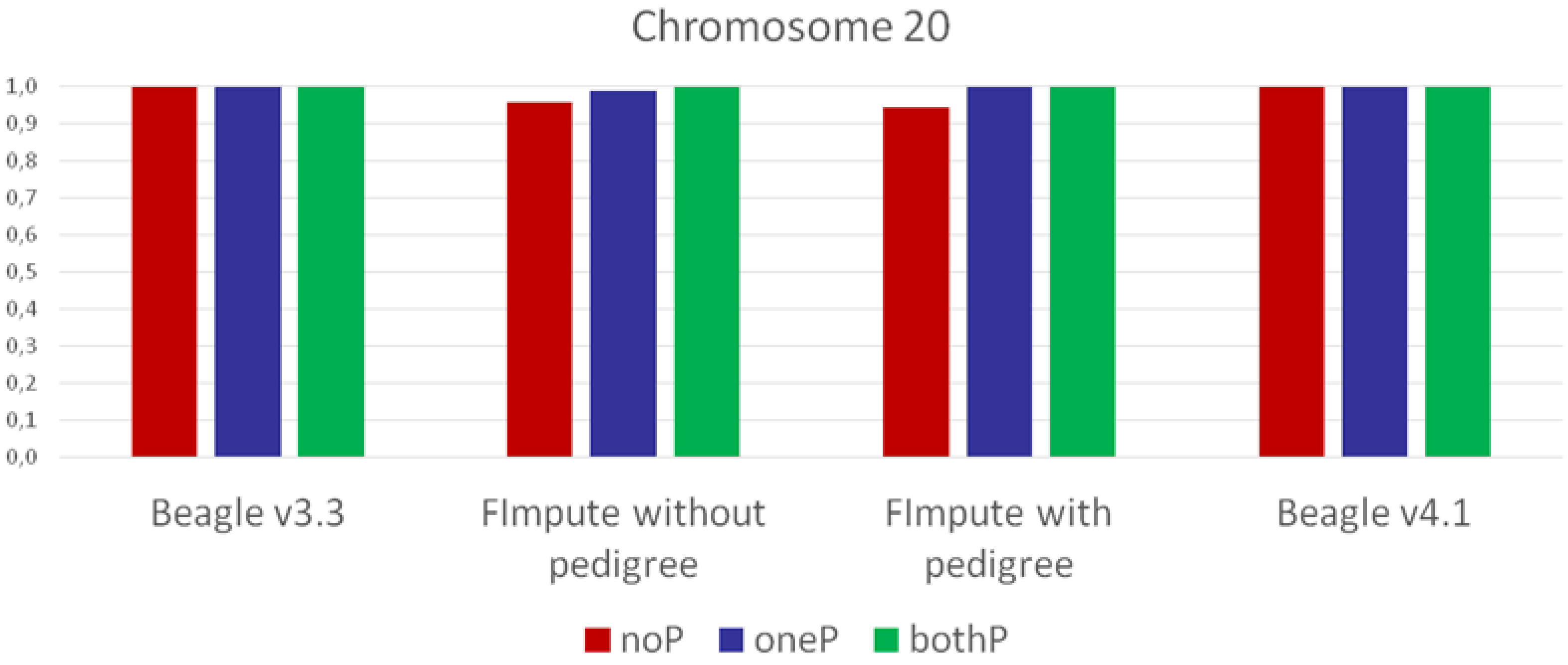
Distance between breakpoints

Methods – Definition of subsets

- Statistics were obtained for three subsets in order to analyze the effect of genotyped parents:
  - ***noP***: 231 individuals with no genotyped parents
  - ***oneP***: 606 individuals with one genotyped parent
  - ***bothP***: 51 individuals with both parents genotyped

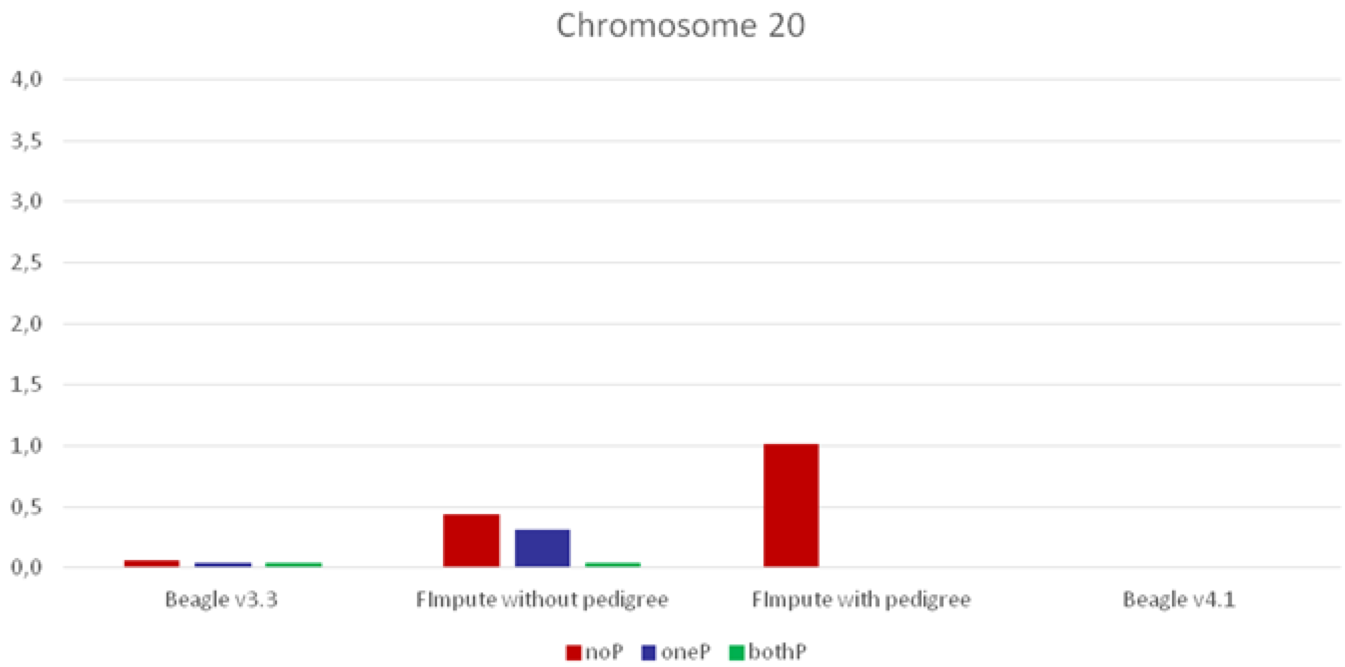
Results

Median values of proportion of correctly phased SNPs



Results

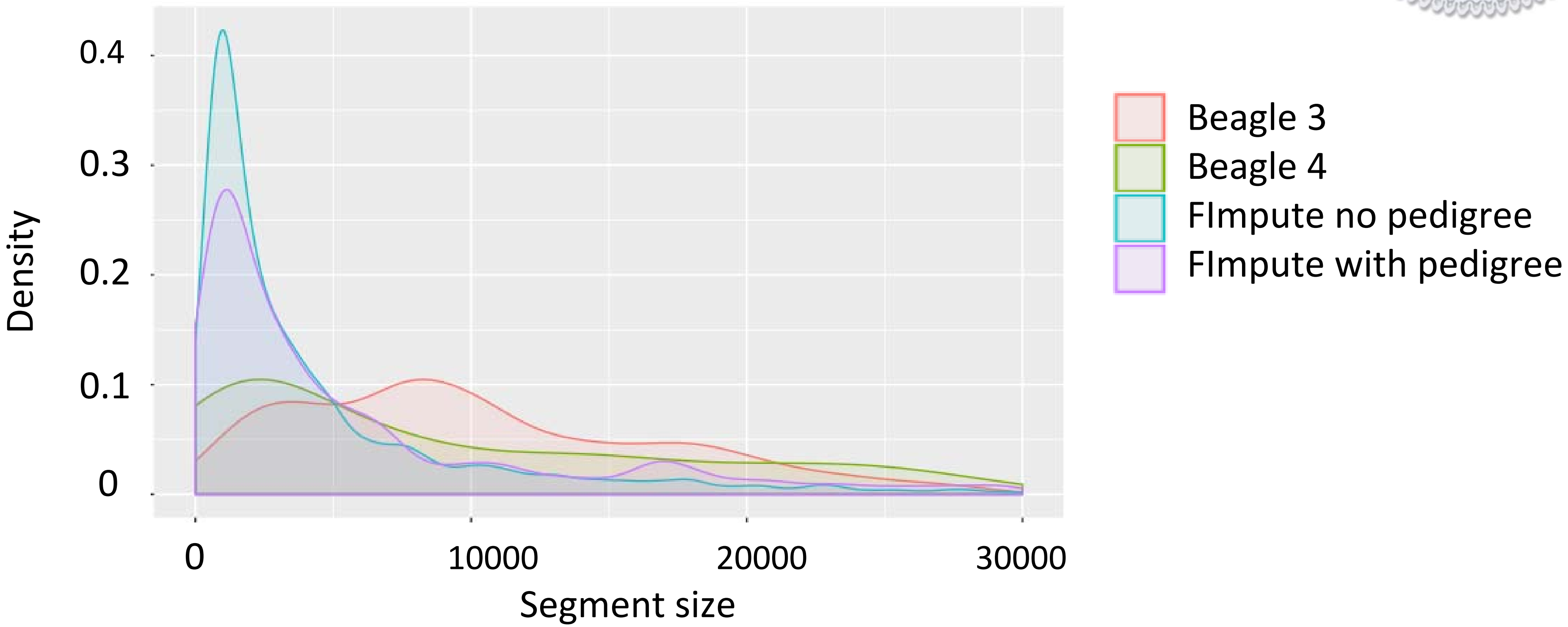
- *Number of breakpoints per 1000 SNPs*



Results

CiBreed
Center for Integrated Breeding Research

Density plots of segments (kB) between breakpoints for Chr. 7



Conclusions

- Phasing quality was in general better with Beagle v4.1
- With at least one parent genotyped, FImpute with pedigree information reached similar levels of phasing quality
- Number of breakpoints and segments'size between breakpoints varied among the alternatives studied
- FImpute recovered haplotypes with a larger amount of short inverted segments
- FImpute computation time was approx. 1/3 in relation to Beagle 4.1

Acknowledgments

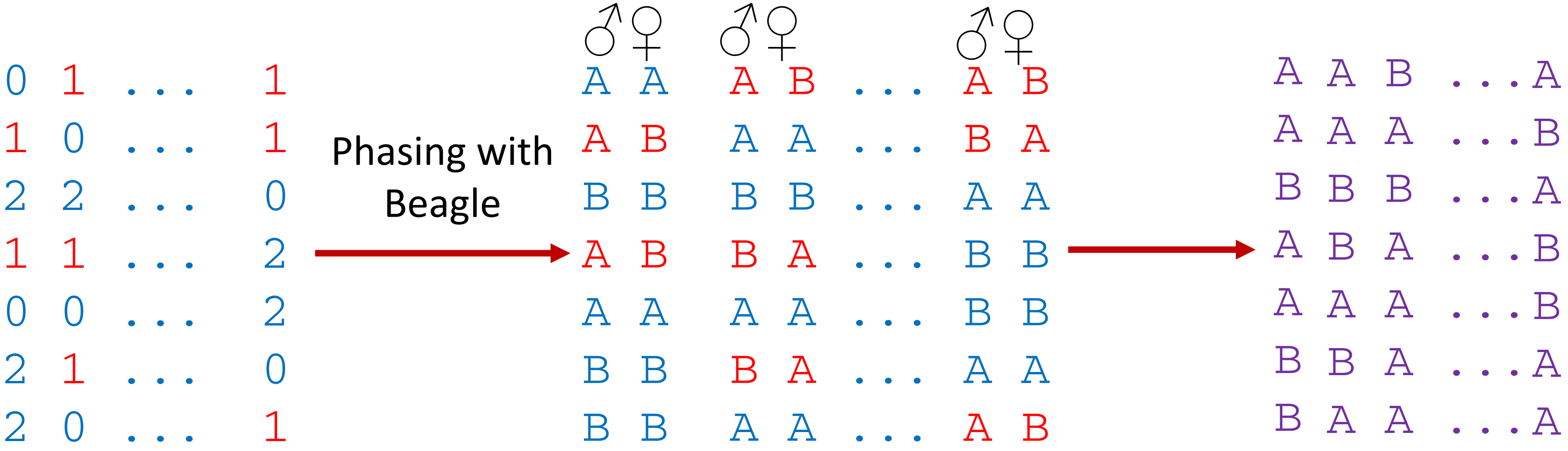
These studies were financially supported by the research training group RTG 1644 ‘Scaling Problem in Statistics’, financed by the German Research Foundation (DFG).

We thank Lohmann Tierzucht GmbH for providing the data.

Thanks for your attention!

Methods - Simulation

Real data



0	1	. . .	1
1	0	. . .	1
2	2	. . .	0
1	1	. . .	2
0	0	. . .	2
2	1	. . .	0
2	0	. . .	1

Population of 888 individuals

Phasing with Beagle

♂	♀	♂	♀		♂	♀
A	A	A	B	. . .	A	B
A	B	A	A	. . .	B	A
B	B	B	B	. . .	A	A
A	B	B	A	. . .	B	B
A	A	A	A	. . .	B	B
B	B	B	A	. . .	A	A
B	B	A	A	. . .	A	B

888 phased diplotypes (1776 haplotypes)

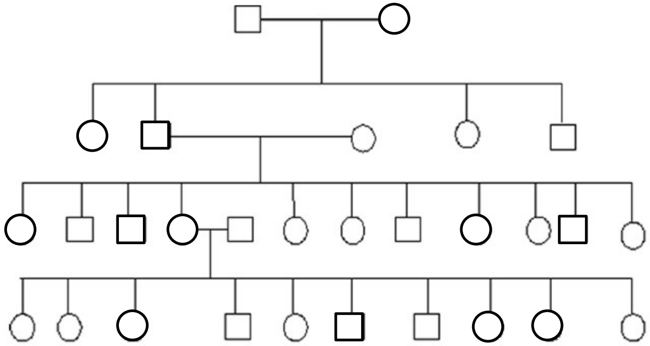
A	A	B	. . . A
A	A	A	. . . B
B	B	B	. . . A
A	B	A	. . . B
A	A	A	. . . B
B	B	A	. . . A
B	A	A	. . . A

2000 haplotypes randomly chosen for the simulation process

Methods - Simulation

A B	A A
A A	A B
B B	A A
B A	B B
A A	A B
B A	A A
A A	A A

Assign random haplotypes to founders



AA	AB	BB	AB	BB	AA	BA	BB	AB	BB
AA	AB	BB	AA	AB	AA	BB	BA	AA	AA
BB	AA	AB	BB	AA	BA	AA	AA	BA	BA
BA	BB	AA	AB	BA	AA	BB	AB	AA	AA
AA	BA	BA	BA	BB	AA	AA	BB	BA	AA
BA	AA	AA	BB	AA	AA	BA	AB	AA	AA
AA	BB	BA	AA	BA	BB	AA	BB	BB	BA

True simulated diplotypes

Drop through real pedigree

-no mutations

-random crossing overs

(Poisson-distribution recombination, uniform distribution)

0	1	. . .	2
0	1	. . .	0
2	0	. . .	1
1	2	. . .	0
0	1	. . .	0
1	0	. . .	0
0	2	. . .	1

Simulated genotypes

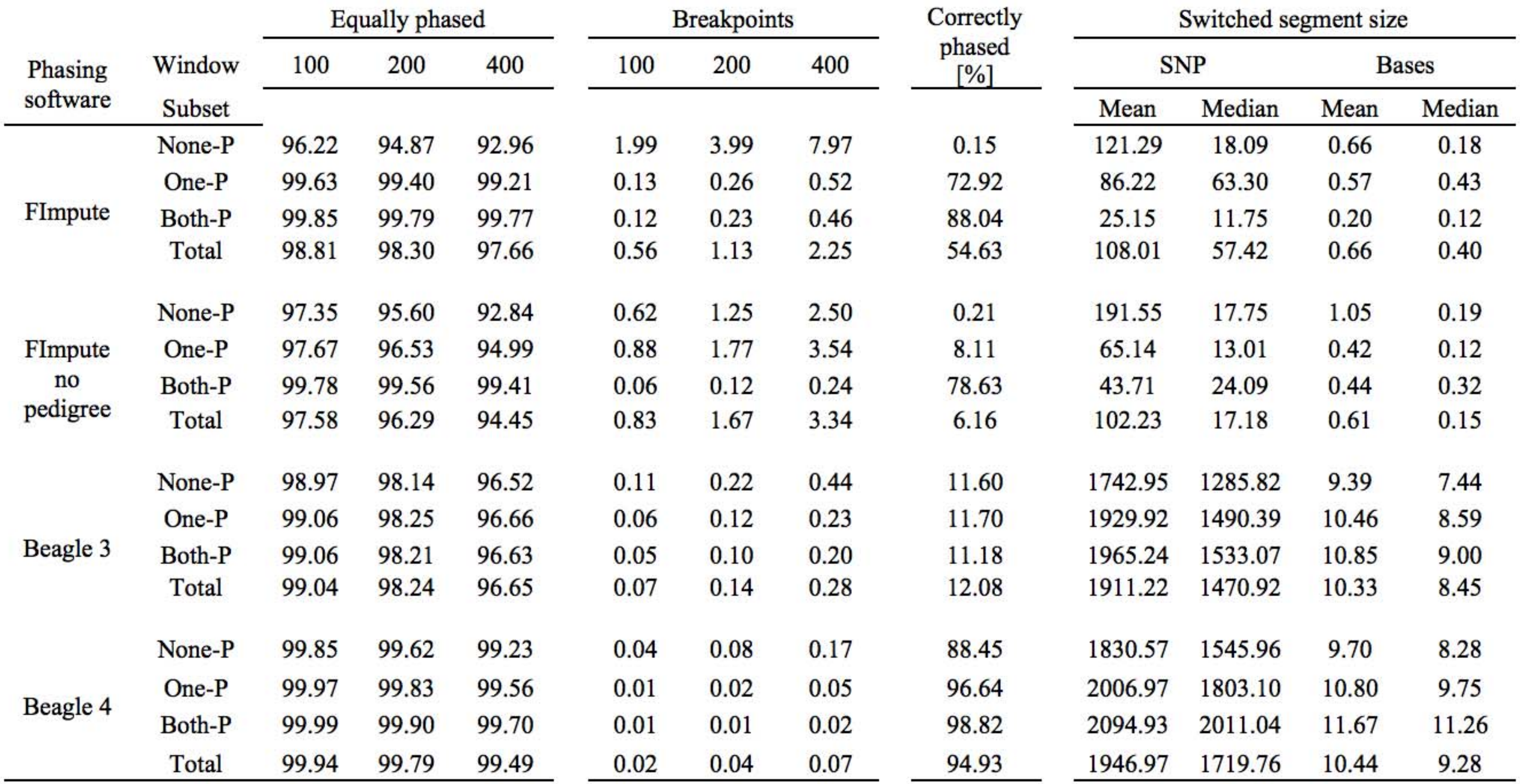
Phasing with

-FImpute v2.2(w or w/out Pedigree)

-Beagle v3.3

-Beagle v4.1

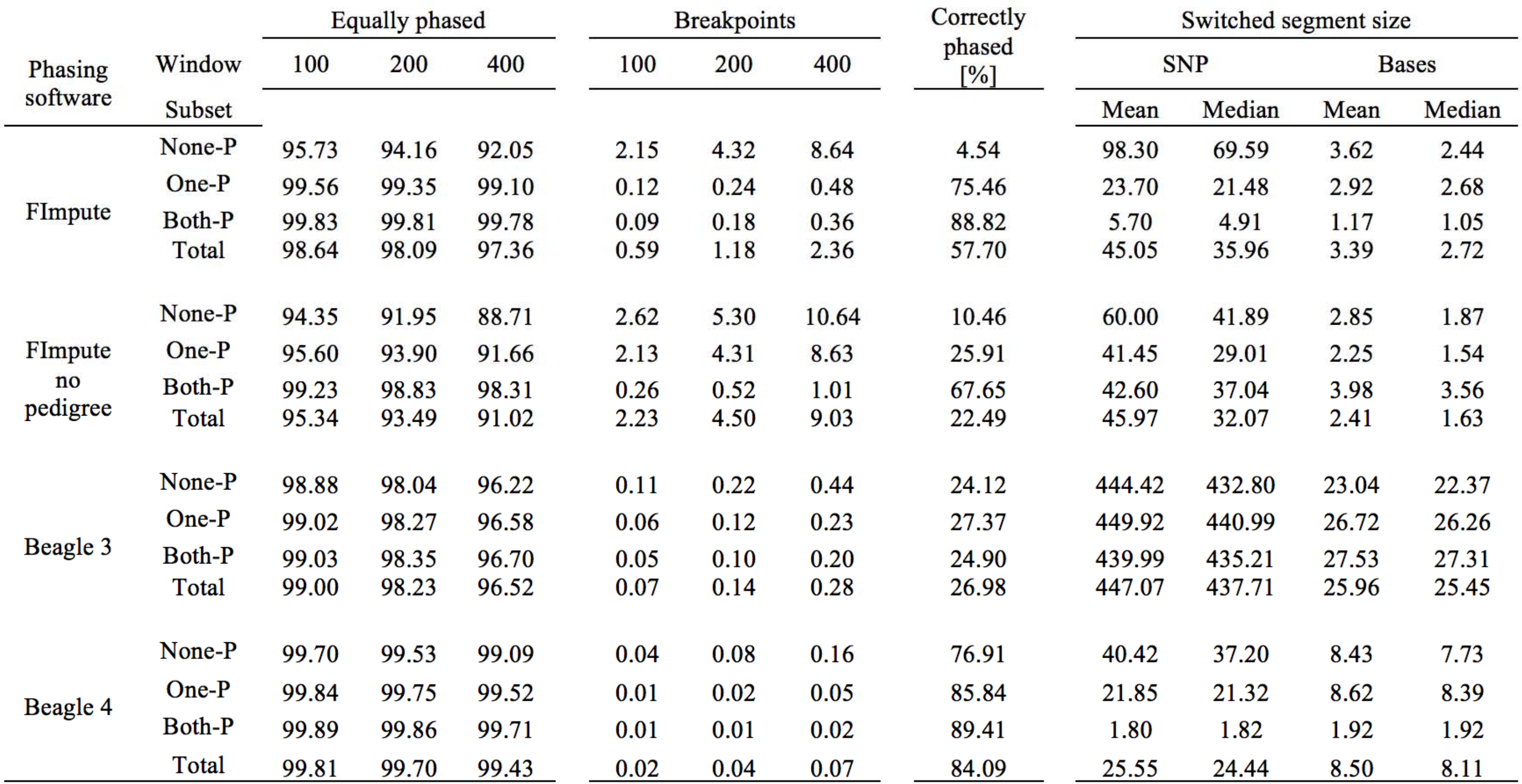
Results Chr. 1



		Equally phased			Breakpoints			Correctly phased [%]	Switched segment size			
Phasing software	Window	100	200	400	100	200	400		SNP		Bases	
	Subset								Mean	Median	Mean	Median
FImpute	None-P	96.22	94.87	92.96	1.99	3.99	7.97	0.15	121.29	18.09	0.66	0.18
	One-P	99.63	99.40	99.21	0.13	0.26	0.52	72.92	86.22	63.30	0.57	0.43
	Both-P	99.85	99.79	99.77	0.12	0.23	0.46	88.04	25.15	11.75	0.20	0.12
	Total	98.81	98.30	97.66	0.56	1.13	2.25	54.63	108.01	57.42	0.66	0.40
FImpute no pedigree	None-P	97.35	95.60	92.84	0.62	1.25	2.50	0.21	191.55	17.75	1.05	0.19
	One-P	97.67	96.53	94.99	0.88	1.77	3.54	8.11	65.14	13.01	0.42	0.12
	Both-P	99.78	99.56	99.41	0.06	0.12	0.24	78.63	43.71	24.09	0.44	0.32
	Total	97.58	96.29	94.45	0.83	1.67	3.34	6.16	102.23	17.18	0.61	0.15
Beagle 3	None-P	98.97	98.14	96.52	0.11	0.22	0.44	11.60	1742.95	1285.82	9.39	7.44
	One-P	99.06	98.25	96.66	0.06	0.12	0.23	11.70	1929.92	1490.39	10.46	8.59
	Both-P	99.06	98.21	96.63	0.05	0.10	0.20	11.18	1965.24	1533.07	10.85	9.00
	Total	99.04	98.24	96.65	0.07	0.14	0.28	12.08	1911.22	1470.92	10.33	8.45
Beagle 4	None-P	99.85	99.62	99.23	0.04	0.08	0.17	88.45	1830.57	1545.96	9.70	8.28
	One-P	99.97	99.83	99.56	0.01	0.02	0.05	96.64	2006.97	1803.10	10.80	9.75
	Both-P	99.99	99.90	99.70	0.01	0.01	0.02	98.82	2094.93	2011.04	11.67	11.26
	Total	99.94	99.79	99.49	0.02	0.04	0.07	94.93	1946.97	1719.76	10.44	9.28

Results Chr. 7

CiBreed
Center for Integrated Breeding Research



Phasing software	Window	Equally phased			Breakpoints			Correctly phased [%]	Switched segment size			
		100	200	400	100	200	400		SNP		Bases	
	Subset								Mean	Median	Mean	Median
FImpute	None-P	95.73	94.16	92.05	2.15	4.32	8.64	4.54	98.30	69.59	3.62	2.44
	One-P	99.56	99.35	99.10	0.12	0.24	0.48	75.46	23.70	21.48	2.92	2.68
	Both-P	99.83	99.81	99.78	0.09	0.18	0.36	88.82	5.70	4.91	1.17	1.05
	Total	98.64	98.09	97.36	0.59	1.18	2.36	57.70	45.05	35.96	3.39	2.72
FImpute no pedigree	None-P	94.35	91.95	88.71	2.62	5.30	10.64	10.46	60.00	41.89	2.85	1.87
	One-P	95.60	93.90	91.66	2.13	4.31	8.63	25.91	41.45	29.01	2.25	1.54
	Both-P	99.23	98.83	98.31	0.26	0.52	1.01	67.65	42.60	37.04	3.98	3.56
	Total	95.34	93.49	91.02	2.23	4.50	9.03	22.49	45.97	32.07	2.41	1.63
Beagle 3	None-P	98.88	98.04	96.22	0.11	0.22	0.44	24.12	444.42	432.80	23.04	22.37
	One-P	99.02	98.27	96.58	0.06	0.12	0.23	27.37	449.92	440.99	26.72	26.26
	Both-P	99.03	98.35	96.70	0.05	0.10	0.20	24.90	439.99	435.21	27.53	27.31
	Total	99.00	98.23	96.52	0.07	0.14	0.28	26.98	447.07	437.71	25.96	25.45
Beagle 4	None-P	99.70	99.53	99.09	0.04	0.08	0.16	76.91	40.42	37.20	8.43	7.73
	One-P	99.84	99.75	99.52	0.01	0.02	0.05	85.84	21.85	21.32	8.62	8.39
	Both-P	99.89	99.86	99.71	0.01	0.01	0.02	89.41	1.80	1.82	1.92	1.92
	Total	99.81	99.70	99.43	0.02	0.04	0.07	84.09	25.55	24.44	8.50	8.11

Results Chr. 20

		Equally phased			Breakpoints			Correctly phased [%]	Switched segment size			
Phasing software	Window	100	200	400	100	200	400		SNP		Bases	
	Subset								Mean	Median	Mean	Median
FImpute	None-P	96.03	94.47	92.32	1.91	3.88	7.98	27.84	81.52	62.54	5.70	4.63
	One-P	99.57	99.40	99.20	0.11	0.22	0.44	84.25	50.16	44.68	3.51	3.22
	Both-P	99.83	99.78	99.75	0.07	0.16	0.33	93.12	11.42	9.82	0.50	0.44
	Total	98.75	98.24	97.57	0.52	1.05	2.15	70.55	70.89	57.91	4.93	4.23
FImpute no pedigree	None-P	96.17	94.45	92.20	1.51	3.03	6.06	41.91	83.80	63.17	5.95	4.69
	One-P	97.31	96.11	94.52	1.03	2.06	4.13	52.60	91.44	71.79	6.36	5.22
	Both-P	99.00	98.45	97.90	0.26	0.53	1.05	72.30	74.81	62.19	5.66	4.87
	Total	97.10	95.79	94.08	1.12	2.24	4.48	50.57	90.36	70.61	6.31	5.16
Beagle 3	None-P	99.18	98.49	97.09	0.10	0.19	0.39	64.74	316.55	306.15	22.37	21.77
	One-P	99.33	98.66	97.41	0.05	0.10	0.19	66.61	309.11	302.72	23.96	23.56
	Both-P	99.38	98.70	97.59	0.04	0.07	0.14	69.74	307.13	306.06	25.87	25.82
	Total	99.30	98.63	97.35	0.06	0.12	0.24	66.35	311.58	303.97	23.64	23.18
Beagle 4	None-P	99.73	99.52	99.15	0.04	0.07	0.15	87.68	122.67	113.62	9.69	9.26
	One-P	99.85	99.72	99.47	0.01	0.03	0.05	92.27	136.08	129.12	8.71	8.33
	Both-P	99.89	99.79	99.60	0.01	0.01	0.03	93.91	50.46	50.46	3.79	3.79
	Total	99.83	99.68	99.41	0.02	0.04	0.07	91.37	128.57	120.96	8.90	8.51