

ACCURATE GENOMIC PREDICTION OF *Coffea canephora* IN MULTIPLE ENVIRONMENTS USING WHOLE-GENOME STATISTICAL MODELS

SUPPLEMENTARY MATERIAL

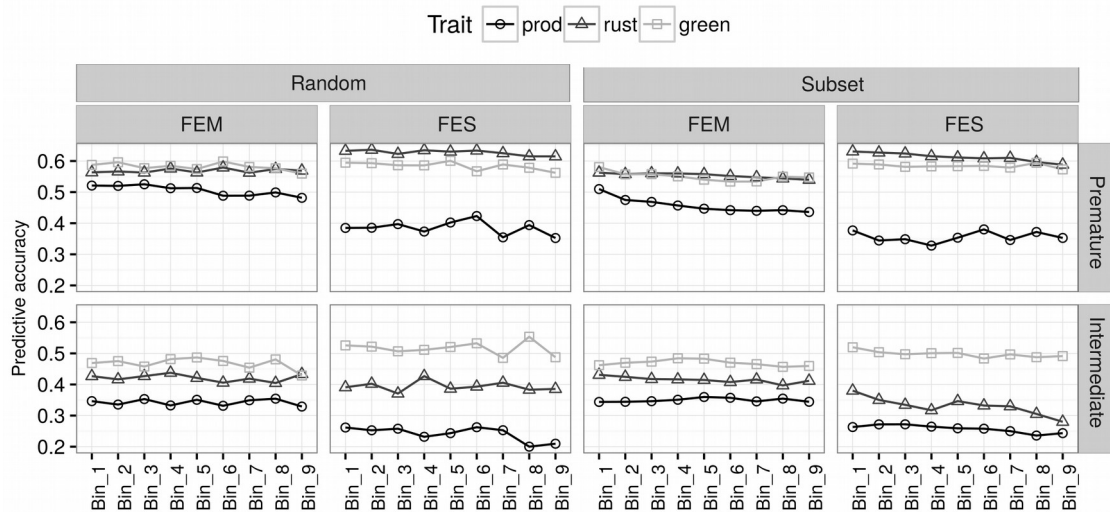


Figure S1: Mean predictive ability using different marker densities. Two *Coffea canephora* populations (Intermediate and Premature) evaluated at two locations (FEM and FES) for production (kilograms of mature coffee fruit in the cherries stage), incidence of coffee leaf rust (scaling score) and yield of green beans (grams in samples of 2 kilograms of coffee fruit in the cherries stage). Right column: SNPs were sampled across the genome. Left column: SNPs were selected based on MAF and call rate. From Bin1 to Bin9, in the Intermediate population, the number of SNPs are: 35427, 20450, 13690, 10189, 7989, 6577, 5559, 4780 and 4240. For the Premature population, the number of SNPs are: 40767, 21433, 13969, 10283, 8019, 6587, 5560, 4780 and 4240.

Table S1: Variance and covariance (VCOV) structures examined for the random effects in Model 1 for *Coffea canephora* phenotypic analysis

Model	Num. Parameters^a	Description
Ident	1	Identical variation
Diag	K	Heterogeneous variations
CompSym	$K+1$	Compound symmetry with heterogeneous variance
Uns	$K(K+1)/2$	Unstructured model

^a The number of parameters for the models follows from the sum of the parameters for the component matrices minus the number of identification constraints. K is the number of harvests.

Table S2: Analysis of variance (ANOVA) *p*-values and other statistics for candidate factors affecting the accuracy of genomic predictions in *Coffea canephora*. Candidate factors include the choice of genomic selection model, trait, population and location, and pairwise combinations of these factors.

Source of variation	df	Sum Sq	Mean Sq	F value	<i>p</i>-value
Model ^a	11	1.73	0.16	4.06	<0.0001
Trait ^b	2	30.63	15.31	395.51	<0.0001
Pop ^c	1	42.24	42.24	1090.85	<0.0001
Loc ^d	1	0.34	0.34	8.73	0.0031
Model x Pop	11	0.57	0.05	1.33	0.2021
Model x Trait	22	0.54	0.02	0.63	0.9031
Model x Loc	11	0.13	0.01	0.31	0.9840
Trait x Pop	2	2.72	1.36	35.09	<0.0001
Trait x Loc	2	8.20	4.10	105.86	<0.0001
Pop x Loc	1	0.10	0.10	2.61	0.1064
Residuals	4255	164.75	0.04		

^a Categorical variable for choice of genomic selection method (bayesA, bayesB, bayesC, bayesLASSO, bayesRR, bayesR, bayesVS, bsmlm, rblup, lasso, pls, and RForest). The fixedMLR model was not considered in the analysis. (Tukey's *post hoc* test classified all methods in the same group except the LASSO.)

^b Trait in three for trait (production, incidence of coffee leaf rust, and yield of green beans).

^c Pop in two recurrent selection populations (Intermediate and Premature).

^d Loc in two locations (FEM and FES).

Table S3: Goodness-of-fit values for the genetic and residual matrices, factored by harvest. Two *Coffea canephora* populations (Intermediate and Premature) were evaluated in two locations (FEM and FES) for production (kilograms of mature coffee fruit in the cherries stage), incidence of coffee leaf rust (scaling score) and yield of green beans (grams in samples of 2 kilograms of coffee fruit in the cherries stage).

Intermediate								
Location	G _h	R _h	Production		Rust		Green	
			AIC	BIC	AIC	BIC	AIC	BIC
FEM	Ident	Ident	9709.4	9751.4	9709.4	9751.4	10816.0	10849.7
	Diag	Ident	9679.8	9737.5	9679.8	9737.5	10746.8	10795.0
	CompSym	Ident	9695.6	9742.7	9695.6	9742.7	10805.6	10844.1
	Uns	Ident	9639.7	9728.8	9639.7	9728.8	10750.6	10827.7
	Ident	Diag	9648.9	9706.6	9648.9	9706.6	10511.5	10559.7
	Diag	Diag	9627.0	9700.4	9627.0	9700.4	10516.1	10578.7
	CompSym	Diag	9632.6	9695.5	9632.6	9695.5	10484.3	10537.2
	Uns	Diag	9587.0	9691.9	9587.0	9691.9	10492.8	10584.2
FES	Ident	Ident	10296.6	10338.5	5218.2	5260.2	9661.9	9695.7
	Diag	Ident	10281.3	10339.0	5220.7	5278.5	9646.1	9694.4
	CompSym	Ident	10224.0	10271.2	5094.0	5141.3	9570.2	9608.8
	Uns	Ident	10141.2	10230.4	5077.4	5166.6	9536.6	9613.8
	Ident	Diag	10275.7	10333.5	5190.8	5248.5	9595.7	9643.9
	Diag	Diag	10256.3	10329.7	5192.3	5265.8	9595.9	9658.6
	CompSym	Diag	10203.6	10266.5	5064.0	5127.1	9498.2	9551.3
	Uns	Diag	10114.3	10219.3	5048.4	5153.4	9483.9	9575.6
Premature								
FEM	Ident	Ident	8279.9	8320.7	3318.7	3359.6	9470.3	9503.1
	Diag	Ident	8284.3	8340.5	3195.2	3251.5	9413.4	9460.2
	CompSym	Ident	8257.3	8303.3	3319.2	3365.2	9447.9	9485.4
	Uns	Ident	8259.6	8346.5	3185.6	3272.6	9389.3	9464.3
	Ident	Diag	8239.4	8295.6	3236.0	3292.3	9312.6	9359.5
	Diag	Diag	8243.8	8315.3	3134.5	3206.1	9309.2	9370.2
	CompSym	Diag	8215.5	8276.8	3236.2	3297.6	9267.4	9318.9
	Uns	Diag	8219.0	8321.2	3124.0	3226.3	9268.1	9357.1
FES	Ident	Ident	8661.8	8702.5	4509.7	4550.3	8338.5	8371.0
	Diag	Ident	8641.0	8696.9	4484.0	4539.9	8297.5	8344.0
	CompSym	Ident	8586.9	8632.6	4360.4	4406.1	8186.1	8223.2
	Uns	Ident	8525.2	8611.5	4305.1	4391.4	8110.6	8185.0
	Ident	Diag	8644.8	8700.7	4437.3	4493.1	8324.4	8370.9
	Diag	Diag	8629.0	8700.1	4414.8	4485.9	8291.3	8351.7
	CompSym	Diag	8569.6	8630.5	4286.3	4347.3	8154.3	8205.4
	Uns	Diag	8513.2	8614.8	4233.4	4334.9	8103.0	8191.4

Ident: Identical variation; Diag: Heterogeneous variations; CompSym: compound symmetry with heterogeneous variance; Uns: unstructured model. Bold numbers represent the smallest AIC and BIC values, indicating the best fitted phenotypic model.

Table S4: Mean predictive ability of thirteen methods applied to genomic prediction. Two *Coffea canephora* populations (Intermediate and Premature) evaluated in two locations (FEM and FES) for production (kilograms of mature coffee fruit in the cherries stage), incidence of coffee leaf rust (scaling score) and yield of green beans (grams in samples of 2 kilograms of coffee fruit in the cherries stage). Predictive abilities were assessed using a Replicated Training-Testing evaluation. In each replication, 80% of the individuals were assigned randomly for training data set, while the remaining 20% were assigned for testing data set (TST).

Model	Intermediate					
	FEM			FES		
	Production	Rust	Green	Production	Rust	Green
bayesA	0.3591	0.4241	0.4634	0.2686	0.3938	0.5252
bayesB	0.3539	0.4299	0.4656	0.2687	0.3936	0.5250
bayesC	0.3531	0.4332	0.4685	0.2673	0.3936	0.5290
bayesLASSO	0.3509	0.4350	0.4663	0.2623	0.3903	0.5280
bayesRR	0.3568	0.4325	0.4674	0.2655	0.3962	0.5283
bayesR	0.3465	0.4244	0.4119	0.2493	0.3856	0.5356
bayesVS	0.3470	0.4029	0.4493	0.2165	0.3499	0.4151
bslmm	0.3442	0.4195	0.4594	0.2451	0.3600	0.5046
rrblup	0.3442	0.4280	0.4653	0.2562	0.3911	0.5222
lasso	0.3013	0.4308	0.4160	0.2937	0.2445	0.4454
pls	0.3701	0.4047	0.4522	0.2891	0.3926	0.5151
RForest	0.3844	0.3970	0.4073	0.1763	0.3399	0.5104
fixedMLR	0.1219	0.0298	-0.019	0.0056	-0.017	-0.018
	Premature					
bayesA	0.5253	0.5637	0.5792	0.3796	0.6348	0.5980
bayesB	0.5248	0.5661	0.5807	0.3787	0.6341	0.5980
bayesC	0.5245	0.5663	0.5802	0.3793	0.6357	0.6003
bayesLASSO	0.5229	0.5706	0.5779	0.3723	0.6383	0.6012
bayesRR	0.5252	0.5687	0.5804	0.3792	0.6352	0.5993
bayesR	0.5371	0.5777	0.5895	0.4050	0.5980	0.5863
bayesVS	0.5067	0.5625	0.5754	0.3973	0.6371	0.5264
bslmm	0.5037	0.5515	0.5564	0.4251	0.6287	0.57
rrblup	0.5181	0.5649	0.5872	0.3822	0.6313	0.5923
lasso	0.4693	0.4953	0.4679	0.3679	0.5831	0.5717
pls	0.5184	0.5496	0.5870	0.3867	0.6305	0.5908
RForest	0.5537	0.5770	0.5646	0.4309	0.6907	0.5849
fixedMLR	0.0050	0.0025	0.0428	0.0321	0.0861	0.0193

Table S5: Mean value of the mean squared prediction error (MSPE) of thirteen methods applied to genomic prediction. Two *Coffea canephora* populations (Intermediate and Premature) evaluated in two locations (FEM and FES) for production (kilograms of mature coffee fruit in the cherries stage), incidence of coffee leaf rust (scaling score) and yield of green beans (grams in samples of 2 kilograms of coffee fruit in the cherries stage). Predictive abilities were assessed using a Replicated Training-Testing evaluation. In each replication, 80% of the individuals were assigned randomly for training data set, while the remaining 20% were assigned for testing data set (TST).

Intermediate						
	FEM			FES		
Model	Production	Rust	Green	Production	Rust	Green
bayesA	0.87	0.90	0.76	0.88	0.82	0.71
bayesB	0.87	0.89	0.75	0.87	0.83	0.72
bayesC	0.87	0.89	0.75	0.87	0.83	0.71
bayesLASSO	0.88	0.90	0.76	0.87	0.84	0.72
bayesRR	0.87	0.89	0.75	0.87	0.83	0.71
bayesR	0.89	0.92	0.82	0.89	0.83	0.73
bayesVS	0.93	0.96	0.85	0.92	0.92	0.87
bslmm	0.88	0.91	0.76	0.91	0.86	0.71
rrblup	1.04	2.03	1.88	0.91	1.37	1.09
lasso	0.95	0.88	0.84	0.85	1.00	0.77
pls	3.09	3.76	2.88	1.01	1.28	1.05
RForest	0.85	0.93	0.82	0.91	0.87	0.78
fixedMLR	762.47	108.66	50.57	10 ⁴	10 ⁴	10 ³
Premature						
bayesA	0.75	0.67	0.66	0.87	0.59	0.65
bayesB	0.75	0.67	0.66	0.88	0.59	0.65
bayesC	0.75	0.66	0.66	0.88	0.59	0.65
bayesLASSO	0.76	0.66	0.67	0.89	0.59	0.67
bayesRR	0.75	0.66	0.66	0.88	0.59	0.66
bayesR	0.76	0.68	0.70	0.88	0.69	0.69
bayesVS	0.82	0.71	0.74	0.93	0.63	0.79
bslmm	0.77	0.68	0.68	0.85	0.60	0.68
rrblup	0.77	1.59	0.93	1.50	1.95	1.52
lasso	0.80	0.75	0.76	0.93	0.68	0.83
pls	0.76	1.68	0.93	1.64	1.98	1.56
RForest	0.74	0.67	0.71	0.85	0.54	0.68
fixedMLR	10 ⁴	10 ⁴	10 ⁴	10 ⁴	2794.73	10 ⁴

Table S6: Mean value of the linear regression coefficient of the observed on predicted genetic value (slope) of thirteen methods applied to genomic prediction. Two *Coffea canephora* populations (Intermediate and Premature) evaluated in two locations (FEM and FES) for production (kilograms of mature coffee fruit in the cherries stage), incidence of coffee leaf rust (scaling score) and yield of green beans (grams in samples of 2 kilograms of coffee fruit in the cherries stage). Predictive abilities were assessed using a Replicated Training-Testing evaluation. In each replication, 80% of the individuals were assigned randomly for training data set, while the remaining 20% were assigned for testing data set (TST).

Intermediate						
	FEM			FES		
Model	Production	Rust	Green	Production	Rust	Green
bayesA	1.003	1.002	1.016	0.740	1.056	1.183
bayesB	1.044	1.054	1.070	0.809	1.121	1.214
bayesC	1.114	1.102	1.114	0.865	1.171	1.248
bayesLASSO	1.197	1.240	1.230	0.903	1.275	1.396
bayesRR	1.118	1.118	1.127	0.857	1.220	1.253
bayesR	1.658	1.497	1.425	1.409	1.426	1.542
bayesVS	3.514	2.132	2.918	3.302	2.697	1.869
bslmm	1.212	1.081	1.050	0.906	1.026	1.045
rrblup	1.370	1.370	1.370	1.239	1.239	1.239
lasso	0.758	1.091	1.021	0.724	0.667	1.029
pls	0.779	0.780	0.801	0.613	0.835	0.963
RForest	1.334	1.397	1.525	0.719	1.264	1.859
fixedMLR	0.070	0.027	-0.02	0.002	0.006	-0.01
Premature						
bayesA	1.170	0.987	0.851	1.093	1.033	1.168
bayesB	1.189	1.014	0.911	1.119	1.048	1.183
bayesC	1.210	1.039	0.938	1.154	1.071	1.215
bayesLASSO	1.366	1.143	1.202	1.316	1.168	1.364
bayesRR	1.217	1.044	0.954	1.173	1.070	1.221
bayesR	1.584	1.456	1.881	1.813	1.315	1.423
bayesVS	2.027	1.495	2.375	2.510	1.454	1.580
bslmm	1.014	0.979	1.024	1.166	0.990	1.027
rrblup	1.042	1.007	0.847	1.093	0.993	1.079
lasso	1.095	0.979	0.455	0.916	1.038	1.135
pls	1.026	0.865	0.694	0.941	0.930	1.022
RForest	1.493	1.211	1.186	1.376	1.264	1.306
fixedMLR	-0.002	-0.002	0	0	0.014	0.001