Conservation Genetics Resources
https://doi.org/10.1007/s12686-021-01201-y

TECHNICAL NOTE

Selecting a core set of nuclear SNP markers for molecular characterization of Arabica coffee (*Coffea arabica* L.) genetic resources

Dapeng Zhang[1] · **Fernando E. Vega**[1] · **William Solano**[2] · **Fuyuan Su**[3] · **Francisco Infante**[4] · **Lyndel W. Meinhardt**[1]

Received: 30 September 2020 / Accepted: 18 February 2021

Abstract

Coffee is one of the most economically important agricultural commodities in the world. Labeling accuracy and conservation efficiency are essential for coffee germplasm management and for the exchange and utilization in breeding new varieties. However, due to its homogenous genetic background, accurate identification of *Coffea arabica* germplasm has not been fully achieved. Specifically, data comparison across different laboratories and genotyping platforms has not been available. Here, we report the screening of 672 candidate SNPs using Nano-Fluidic Array genotyping. Based on call rate, Minor Allele Frequency and Linkage Disequilibrium, a set of 96 SNPs were selected for genotyping *C. arabica*. This validated panel is suitable for use in coffee germplasm conservation and crop improvement, including varietal identification, seeds and nursery accreditation, and coffee bean authentication.

Keywords Conservation genomics · Fluidigm · Germplasm · Molecular markers · Tropical agriculture

Introduction

Coffee is one of the most economically important agricultural commodities in the world and is especially important for the livelihood of 20 million coffee-farming families in Asia, Africa and Latin America (Vega et al. 2003). There are 124 species in the genus *Coffea* (Davis et al. 2006; 2011) and two species, *Coffea arabica* L. (Arabica coffee) and *Coffea canephora* Pierre ex A. Froehner (robusta coffee) account for over 99% of the commercial market (USDA-FAS 2020).

Coffea arabica is an allotetraploid ($2n=4x=44$) species and is self-compatible. The primary gene pool of this species has been identified in the highlands of southwestern Ethiopia and the Boma Plateau of South Sudan (Meyer 1965; Meyer et al. 1968; Thomas 1942). Only a small fraction of the genetic diversity has been distributed worldwide, to which the global production of Arabica coffee is based (Anthony et al. 2002; Vega 2008; Scalabrin et al. 2020). Despite this narrow genetic diversity in cultivated *C. arabica* germplasm, a large number of *C. arabica* varieties accessions are currently being maintained in various field genebanks around the world (Bramel et al. 2017).

Accurate genotype identification is essential to ensure the efficiency of coffee germplasm management and utilization. Various panels of microsatellite markers have been developed for *C. arabica* and applied on varietal identification and diversity analysis (Combes et al. 2000; Anthony et al. 2002; Cubry et al. 2008; Missio et al. 2009; Lopez-Gartner et al. 2009; Vieira et al. 2010; Razafinarivo et al. 2013; Ferrão et al. 2015; Pruvot-Woehl et al. 2020). While microsatellite markers are highly useful for coffee germplasm management, resolving genotyping results from different laboratories has not been straightforward. Moreover, the high level of genetic similarity among Arabica coffee varieties is still a main challenge for SSR-based identification of Arabica coffee germplasm. Consequently, single nucleotide polymorphism (SNP) markers have been increasingly developed and used for coffee germplasm management, because they are amendable to high throughput systems, have a universal data comparability and lower genotyping cost (Combes et al.

Dapeng Zhang
dapeng.zhang@usda.gov

[1] Sustainable Perennial Crops Laboratory, United States Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA

[2] Centro Agronómico Tropical de Investigación Y Enseñanza (CATIE), Turrialba, Costa Rica

[3] Forestry and Fruit Tree Research Institute, Wuhan Academy of Agricultural Sciences, Wuhan 430075, China

[4] El Colegio de La Frontera Sur (ECOSUR), Tapachula 30700, Chiapas, Mexico

2013; Yuyama et al. 2016; Zhou et al. 2016; Tran et al., 2016; 2018; Sousa et al. 2017; 2019; Sant'Ana et al. 2020; Gimase et al. 2020). Recently, Merot-L'Anthoene et al. (2019) reported the development of an 8.5 K SNP array for *C. Arabica and C. canephora.* This array therefore, provides ample candidate SNPs for selecting a core set SNP markers for array-based genotyping of Arabica coffee. Here, we describe a selection of 96 SNPs as a core set of markers for *Coffea arabica* germplasm identification, by screening the published candidate SNPs.

A total of 130 germplasm accessions of *C. arabica* were used in the present study (Table 1). These 130 coffee accessions, representing traditional varieties, wild Arabica coffee from Ethiopia, and different type of hybrids, were selected from the international coffee genebank held at the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), in Costa Rica.

The CATIE coffee germplasm bank in Costa Rica maintains approximately 2000 coffee accessions and is party to an international treaty on the open sharing of genetic resources for the improvement of crops (Bramel et al. 2017).

DNA samples were extracted from lyophilized coffee leaves using DNeasy Plant Mini kit (Qiagen Inc., Valencia, CA, USA). A total of 672 candidate SNPs was selected from the 8.5 K SNP coffee array published by Merot-L'anthoene et al. (2019). The Fluidigm nanofluidic genotyping system (Fluidigm Corp., San Francisco, CA, USA) was used to evaluate the putative SNP markers for genotype identification. The Assay Design Group at Fluidigm Corp. designed and manufactured the putative SNP primers for competitive allele-specific PCR. An EP1 imager (Fluidigm Corp., San Francisco, CA, USA) was used to acquire fluorescent images of the endpoint reactions in the 96.96 IFC and Fluidigm Genotyping Analysis Software was used to analyze the data.

Basic genotype statistics for each marker, including call rate, minor allele frequency (MAF), and allele and genotype counts were calculated using the Quality Assurance Module from SNP Variation Suite version 8 (SVS8; Golden Helix Inc., Bozeman, Montana, USA). Quality control criteria (filters) were used to remove from further analysis any SNP having less than a 95% overall amplification and MAF < 0.05. Data filtering was further performed to remove SNPs that were in Linkage Disequilibrium (LD) with each other at $r^2 \geq 0.5$ using the same program. A final set of 96 SNPs were selected (S Table 1) and used in subsequent analysis of varietal identification, analysis of genetic diversity and population partitioning.

Using the selected panel, we calculated several population genetic statistics for each locus (Table 2) using GenAlex 6.501 (Peakall and Smouse 2006, 2012). Shannon Information Index ranged from 0.124 to 0.693. The observed and expected heterozygosity ranged from 0.008 to 0.961 and 0.052–0.500, respectively. The minor allele

Table 1 List of 130 analyzed accessions of *C. arabica* germplasm maintained in the International Coffee Collection in CATIE, Costa Rica

No	Sample ID	Sample name	Source of introduction	
			Country	Province
1	T.04278	Babaca Kaffa	Congo	Ruanda Urundi
2	T.02758	Barbuk Sudan	Sudan	Barbuk
3	T.02251	Batie	Ethiopia	Dessie
4	T.04272	Blue Mountain	Congo	Ruanda Urundi
5	T.00977	Blue Mountain	Guatemala	La Aurora
6	T.04323	Blue Mountain	Malawi	N/A
7	T.05270	Bourbon	Colombia	Caldas
8	T.02540	Bourbon Amarillo	Brasil	Campinas
9	T.04375	Bourbon Amarillo	Venezuela	Monajas
10	T.00983	Bourbon Guatemala	Guatemala	La Aurora
11	T.04258	Bourbon Mayaguez	Congo	Ruanda Urundi
12	T.03469	Bourbon Salvadoreno	El Salvador	La Libertad
13	T.02392	Carmelita Polysperma	Puerto Rico	Mayaguez
14	T.03081	Carrizal	Costa Rica	Alajuela
15	T.12841	Catimor	Mexico	Veracruz
16	T.12842	Catimor	Mexico	Veracruz
17	T.13228	Catimor	Portugal	Oreiras
18	T.16763	Catuai	Brasil	Minas Gerais
19	T.16764	Catuai	Brasil	Minas Gerais
20	T.05267	Catuai	Costa Rica	San José
21	T.05268	Catuai	Costa Rica	San José
22	T.05325	Catuai	Costa Rica	Alajuela
23	T.02542	Caturra	Brasil	Campinas
24	T.15861	Caturra	Costa Rica	San José
25	T.04078	Caturra Variegata	Colombia	Caldas
26	T.00982	Ceilan	Guatemala	La Aurora
27	T.03443	Ceilan	Puerto Rico	Mayaguez
28	T.03427	Cera	Brasil	Campinas
29	T.05199	Cioiccie I	Colombia	Caldas
30	T.05283	Cioiccie S-6	Brasil	Campinas
31	T.11950	Clon 7355	Ethiopia	N/A
32	T.12534	Clon 7357	Ethiopia	N/A
33	T.03411	Columnaris	Panamá	Chiriquí
34	T.02397	Columnaris	Puerto Rico	Mayaguez
35	T.02298	Coorg	Kenya	Ruiru
36	T.04353	Cubujuqui o San Joaquin	Costa Rica	Heredia
37	T.03165	Cumbaya	Ecuador	Quito
38	T.03635	Cumbaya	Ecuador	Quito
39	T.02727	Dalle	Kenya	Sidamo
40	T.04308	Dalle mixed	Malawi	N/A
41	T.02249	Dessie	Ethiopia	Dessie
42	T.02742	Dilla Alghe	Ethiopia	Sidamo
43	T.03129	Dos Tiempos	Cuba	Palma Soriano
44	T.03381	Dos Tiempos	Cuba	N/A

Table 1 (continued)

No	Sample ID	Sample name	Source of introduction	
			Country	Province
45	T.04570	E-147	Ethiopia	Illubabor
46	T.04573	E-150	Ethiopia	Illubabor
47	T.04578	E-155	Ethiopia	Illubabor
48	T.04667	E-160	Ethiopia	Kaffa Jimma
49	T.04681	E-167	Ethiopia	Kaffa Jimma
50	T.04688	E-174	Ethiopia	Kaffa Jimma
51	T.04692	E-178	Ethiopia	Kaffa Jimma
52	T.04539	E-293	Ethiopia	Kaffa Jimma
53	T.04479	E-301	Ethiopia	Kaffa Jimma
54	T.04583	E-325	Ethiopia	Illubabor
55	T.04599	E-341	Ethiopia	Kaffa Jimma
56	T.04610	E-352	Ethiopia	Kaffa Jimma
57	T.02741	Erecta	Kenya	N/A
58	T.02395	Erecta	Puerto Rico	Mayaguez
59	T.11723	Garnica	Mexico	Veracruz
60	T.12854	Garnica	Mexico	Veracruz
61	T.12851	Garnica	Mexico	Veracruz
62	T.02917	Geisha	Congo	N/A
63	T.03214	Geisha	Tanzania	N/A
64	T.01993	Goiaba	Brasil	Campinas
65	T.04250	Goiaba	Colombia	Caldas
66	T.00971	Guadeloupe	Guatemala	La Aurora
67	T.00989	Guadeloupe	El Salvador	La Libertad
68	T.04269	Guatemala 26	Congo	Ruanda Urundi
69	T.12837	Hibrido 055	Mexico	Veracruz
70	T.12846	Hibrido 089	Mexico	Veracruz
71	T.04317	I-60	Malawi	N/A
72	T.14723	Icatu	Brasil	Campinas
73	T.04259	Jackson 2	Congo	Ruanda Urundi
74	T.02919	Jackson 2	Congo	Ruanda Urundi
75	T.02731	Jimma Galla Sidamo	Kenya	N/A
76	T.04290	Jimma Kaffa	Congo	Ruanda Urundi
77	T.02246	Jimma-1	Ethiopia	Jimma
78	T.02254	Jimma-6	Ethiopia	Jimma
79	T.03215	K-7	Kenya	N/A
80	T.04273	Kabare	Congo	Ruanda Urundi
81	T.04268	Kent	Congo	Ruanda Urundi
82	T.04271	Kent	Congo	Ruanda Urundi
83	T.02137	Kona	Guatemala	La Aurora
84	T.03417	Kona	Guatemala	N/A
85	T.02676	Laurina	Camerun	Dschang
86	T.02299	Laurina	Costa Rica	Cartago
87	T.03491	Lejeune 08	Ethiopia	Bada Buna
88	T.03507	Lejeune 12	Ethiopia	Bada Buna
89	T.02257	Lekemti	Ethiopia	Lekemti
90	T.04281	Lignee-M	Congo	Ruanda Urundi
91	T.04007	Loulo	Ethiopia	Sidamo
92	T.04080	Maragogipe	Colombia	Caldas
93	T.04253	Maragogipe	Colombia	Caldas
94	T.00978	Maragogipe	Guatemala	La Aurora
95	T.04264	Mibirizi	Congo	Ruanda Urundi
96	T.02702	Mibirizi	Congo	N/A
97	T.00976	Moca	Guatemala	La Aurora
98	T.04310	Mocha	Malawi	N/A
99	T.04294	Mocha de Tahiti	Congo	Ruanda Urundi
100	T.02394	Mocha Java	Puerto Rico	Mayaguez
101	T.03382	Mundo Novo	Brasil	Campinas
102	T.16742	Mundo Novo	Brasil	Minas Gerais
103	T.16762	Mundo Novo	Brasil	Minas Gerais
104	T.17548	Mundo Novo	Brasil	Minas Gerais
105	T.04060	Murta	Isla Reunión	N/A
106	T.02147	Murta	Guatemala	La Aurora
107	T.02942	Pacas	El Salvador	La Libertad
108	T.15895	Pacas	El Salvador	La Libertad
109	T.03635	Cumbaya	Ecuador	Quito
110	T.00975	Padang	Guatemala	La Aurora
111	T.00992	Padang	El Salvador	La Libertad
112	T.00972	Philippine	Guatemala	La Aurora
113	T.04196	Pluma Hidalgo	Mexico	Veracruz
114	T.00981	Preanger	Guatemala	La Aurora
115	T.00986	Purpurascens	Guatemala	La Aurora
116	T.00987	Purpurascens	El Salvador	La Libertad
117	T.02744	Rume Sudan	Kenya	N/A
118	T.02724	Rume Sudan	Tanzania	N/A
119	T.00980	Sumatra	Guatemala	La Aurora
120	T.02139	Surinam	Guatemala	La Aurora
121	T.00990	Surinam	El Salvador	La Libertad
122	T.04076	Typica Amarillo	Colombia	Caldas
123	T.00996	Typica Rojo	El Salvador	La Libertad
124	T.04077	Variegata	Colombia	Caldas
125	T.04252	Variegata	Colombia	Caldas
126	T.02594	Villalobos	Costa Rica	Alajuela
127	T.10589	Villalobos	Costa Rica	Cartago
128	T.03025	Villasarchi	Costa Rica	Alajuela
129	T.04286	Wondo Sidamo	Congo	Ruanda Urundi
130	T.04292	Wush Wush Kaffa	Congo	Ruanda Urundi

frequency ranged from 0.051–0.494. The probability of identity among siblings (PID$_{\text{SIB}}$) ranged from 0.375 to 0.899. All tested individual coffee varieties were accurately differentiated and the accumulated PID$_{\text{SIB}}$ for the 96 SNPs was 2.79×10^{-16}, demonstrating a high statistical rigor for coffee genotype identification. To our knowledge, this is the first validated set of SNP markers for array-based genotyping of Arabica coffee, and therefore this core

Table 2 Information Index, observed heterozygosity (Ho), expected heterozygosity (He), minor allele frequency (MAF) and probability of identity among siblings (PID-SIB) for 96 SNPs computed across 130 *C. arabica* coffee accessions

Code	Chromosome	Position	Information Index	Ho	He	MAF	PID-SIB
Ca0079	Chr1	1224907	0.153	0.024	0.068	0.102	0.870
Ca0195	Chr1	20516023	0.448	0.038	0.276	0.156	0.562
Ca0227	Chr1	23096701	0.528	0.054	0.344	0.196	0.489
Ca0241	Chr1	23852437	0.500	0.062	0.320	0.267	0.514
Ca0360	Chr1	299951	0.693	0.123	0.500	0.449	0.375
Ca0384	Chr1	30898971	0.317	0.023	0.174	0.093	0.698
Ca0462	Chr1	33972011	0.567	0.169	0.379	0.377	0.458
Ca0468	Chr1	34172000	0.503	0.403	0.322	0.190	0.512
Ca0516	Chr1	35745650	0.657	0.141	0.465	0.363	0.395
Ca0557	Chr1	37407704	0.691	0.123	0.498	0.470	0.376
Ca1768	Chr2	10062730	0.373	0.246	0.216	0.130	0.638
Ca1899	Chr2	14994558	0.585	0.109	0.395	0.293	0.444
Ca2002	Chr2	18003365	0.684	0.157	0.491	0.476	0.380
Ca2024	Chr2	18374353	0.601	0.131	0.411	0.302	0.432
Ca2044	Chr2	18843600	0.604	0.138	0.414	0.269	0.429
Ca2073	Chr2	19829075	0.681	0.087	0.488	0.400	0.381
Ca2247	Chr2	26468579	0.687	0.125	0.494	0.494	0.378
Ca2285	Chr2	27212481	0.649	0.147	0.457	0.443	0.400
Ca2520	Chr2	39439745	0.212	0.031	0.103	0.101	0.809
Ca2606	Chr2	4316080	0.516	0.300	0.334	0.292	0.500
Ca2987	Chr3	10108964	0.689	0.914	0.496	0.433	0.377
Ca3052	Chr3	14471627	0.290	0.123	0.155	0.175	0.726
Ca3165	Chr3	23544457	0.582	0.492	0.393	0.332	0.445
Ca3177	Chr3	24190070	0.692	0.140	0.499	0.415	0.376
Ca3220	Chr3	26951766	0.350	0.038	0.198	0.126	0.663
Ca3400	Chr3	765963	0.124	0.038	0.052	0.141	0.899
Ca3457	Chr4	11198377	0.395	0.069	0.233	0.191	0.615
Ca3594	Chr4	19009387	0.545	0.031	0.359	0.312	0.475
Ca3700	Chr4	23469284	0.661	0.100	0.468	0.470	0.393
Ca3728	Chr4	23810341	0.617	0.123	0.426	0.283	0.420
Ca3794	Chr4	2625880	0.642	0.349	0.450	0.363	0.404
Ca3795	Chr4	2627257	0.494	0.219	0.314	0.206	0.520
Ca3838	Chr4	27806957	0.623	0.138	0.432	0.301	0.416
Ca3870	Chr4	3683162	0.503	0.326	0.322	0.280	0.512
Ca3970	Chr4	7540378	0.594	0.125	0.404	0.280	0.437
Ca4029	Chr4	9865731	0.688	0.362	0.495	0.469	0.378
Ca4110	Chr5	13818337	0.682	0.164	0.489	0.410	0.381
Ca4219	Chr5	20431740	0.626	0.085	0.435	0.316	0.414
Ca4223	Chr5	20525609	0.608	0.562	0.417	0.246	0.427
Ca4233	Chr5	20682692	0.617	0.615	0.426	0.232	0.420
Ca4248	Chr5	20879814	0.691	0.126	0.497	0.475	0.376
Ca4257	Chr5	21068469	0.654	0.123	0.462	0.349	0.396
Ca4373	Chr5	26397451	0.691	0.127	0.498	0.446	0.376
Ca4552	Chr6	10609357	0.216	0.080	0.106	0.056	0.805
Ca4704	Chr6	17030748	0.686	0.885	0.493	0.392	0.378
Ca4729	Chr6	18125235	0.506	0.408	0.325	0.193	0.509
Ca4820	Chr6	23297947	0.299	0.177	0.161	0.072	0.716
Ca4929	Chr6	295268	0.688	0.233	0.495	0.488	0.378
Ca4998	Chr6	32802198	0.554	0.469	0.367	0.302	0.468
Ca5014	Chr6	33600489	0.281	0.162	0.148	0.157	0.736

Table 2 (continued)

Code	Chromosome	Position	Information Index	Ho	He	MAF	PID-SIB
Ca5065	Chr6	36032691	0.262	0.054	0.135	0.183	0.757
Ca5170	Chr6	5930646	0.567	0.073	0.379	0.280	0.457
Ca5198	Chr6	6947717	0.395	0.223	0.233	0.178	0.615
Ca5205	Chr6	714661	0.549	0.138	0.363	0.256	0.471
Ca5273	Chr6	9797719	0.390	0.078	0.229	0.134	0.621
Ca5295	Chr7	10184095	0.691	0.146	0.498	0.452	0.376
Ca5304	Chr7	1034100	0.692	0.519	0.499	0.415	0.376
Ca5414	Chr7	15216957	0.545	0.156	0.359	0.221	0.475
Ca5511	Chr7	20925690	0.423	0.115	0.255	0.246	0.588
Ca5578	Chr7	25514814	0.523	0.372	0.340	0.306	0.494
Ca5643	Chr7	3064537	0.551	0.109	0.365	0.255	0.470
Ca5644	Chr7	3066681	0.661	0.150	0.468	0.372	0.392
Ca5659	Chr7	3389592	0.624	0.148	0.433	0.415	0.416
Ca5810	Chr7	836279	0.677	0.248	0.484	0.488	0.383
Ca5818	Chr7	862842	0.525	0.203	0.342	0.221	0.492
Ca5821	Chr7	879725	0.673	0.246	0.480	0.491	0.386
Ca5877	Chr8	11070431	0.495	0.008	0.315	0.258	0.518
Ca6081	Chr8	24472370	0.693	0.138	0.500	0.442	0.375
Ca6104	Chr8	25270044	0.513	0.008	0.331	0.245	0.503
Ca6109	Chr8	25474893	0.676	0.308	0.483	0.353	0.384
Ca6117	Chr8	25573362	0.693	0.132	0.500	0.440	0.375
Ca6219	Chr8	30215936	0.618	0.102	0.427	0.291	0.420
Ca6300	Chr8	3718478	0.545	0.125	0.359	0.256	0.475
Ca6320	Chr8	5272842	0.582	0.138	0.393	0.288	0.445
Ca6346	Chr8	5758723	0.614	0.146	0.423	0.327	0.422
Ca6385	Chr8	777297	0.687	0.132	0.494	0.396	0.378
Ca6387	Chr8	7822813	0.654	0.200	0.462	0.467	0.396
Ca6416	Chr8	9913824	0.692	0.961	0.499	0.446	0.375
Ca6474	Chr9	11926165	0.670	0.786	0.477	0.432	0.387
Ca6722	Chr9	387547	0.511	0.323	0.329	0.247	0.504
Ca6801	Chr9	6970874	0.680	0.124	0.487	0.463	0.382
Ca6811	Chr9	7340571	0.388	0.077	0.227	0.222	0.623
Ca6840	Chr9	8662102	0.654	0.240	0.461	0.430	0.397
Ca6849	Chr9	9115796	0.693	0.154	0.500	0.415	0.375
Ca0711	Chr10	12029694	0.682	0.146	0.489	0.404	0.381
Ca0726	Chr10	12775321	0.317	0.038	0.174	0.084	0.698
Ca0727	Chr10	12878978	0.318	0.023	0.175	0.076	0.696
Ca0935	Chr10	24193523	0.221	0.038	0.109	0.051	0.800
Ca0955	Chr10	24804788	0.522	0.008	0.339	0.168	0.494
Ca0999	Chr10	2654325	0.416	0.062	0.250	0.120	0.594
Ca1075	Chr10	392589	0.281	0.008	0.148	0.075	0.736
Ca1205	Chr10	9359503	0.635	0.646	0.443	0.391	0.409
Ca1320	Chr11	14483730	0.409	0.085	0.244	0.178	0.601
Ca1405	Chr11	19946378	0.337	0.039	0.189	0.182	0.676
Ca1434	Chr11	21915830	0.271	0.015	0.142	0.062	0.746
Ca1695	Chr11	544708	0.242	0.038	0.122	0.052	0.778
Mean			0.535	0.194	0.360	0.291	0.500

set of markers offer a valuable tool for supporting activities in coffee germplasm conservation and utilization.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12686-021-01201-y.

Acknowledgements This work was supported by the United States Department of Agriculture (USDA) /Agricultural Research Service (ARS). Any mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Data availability All data generated or analyzed during this study that are related to this manuscript are included in the Tables and supplementary information files.

References

Anthony F, Combes M, Astorga C, Bertrand B, Graziosi G, Lashermes P (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theore Appl Genet 104:894–900. https://doi.org/10.1007/s00122-001-0798-8

Bramel P, Krishnan S, Horna D, Lainoff B, Montagnon C (2017) Global conservation strategy for coffee genetic resources. The Crop Trust and World Coffee Research, Bonn, Germany

Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. Mol Ecol 9:1178–1180. https://doi.org/10.1046/j.1365-294x.2000.00954-5.x

Combes M-C, Dereeper A, Severac D, Bertrand B, Lashermes P (2013) Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. New Phytol 200:251–260. https://doi.org/10.1111/nph.12371

Cubry P, Musoli P, Legnaté H, Pot D, de Bellis F, Poncet V, Anthony F, Dufour M, Leroy T (2008) Diversity in coffee assessed with SSR markers: structure of the genus Coffea and perspectives for breeding. Genome 51:50–63. https://doi.org/10.1139/G07-096

Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). Bot J Linn Soc 152:465–512. https://doi.org/10.1111/j.1095-8339.2006.00584.x

Davis AP, Tosh J, Ruch N, Fay MF (2011) Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data: implications for the size, morphology, distribution and evolutionary history of *Coffea*. Bot J Linn Soc 167:357–377. https://doi.org/10.1111/j.1095-8339.2011.01177.x

Ferrão LFV, Caixeta ET, Pena G, Pena G, Zambolim EM, Cruz CD, Zambolim L, Ferrão MAG, Sakiyama NS (2015) New EST–SSR markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. Mol Breed 35:31. https://doi.org/10.1007/s11032-015-0247-z

Gimase JM, Thagana WM, Omondi CO, Cheserek JJ, Gichimu BM, Gichuru EK, Ziyomo C, Sneller CH (2020) Genome-wide association study identify the genetic loci conferring resistance to coffee berry disease (*Colletotrichum kahawae*) in *Coffea arabica* var. Rume Sudan. Euphytica 216:86. https://doi.org/10.1007/s10681-020-02621-x

Gimase JM, Thagana WM, Omondi CO, Cheserek JJ, Gichimu BM, Gichuru EK, Ziyomo C, Sneller CH (2020) Genome-wide association study identify the genetic loci conferring resistance to coffee berry disease (*Colletotrichum kahawae*) in *Coffea arabica*var. Rume Sudan. Euphytica 216:86. https://doi.org/10.1007/s10681-020-02621-x

López-Gartner G, Cortina H, McCouch SR, Moncada MDP (2009) Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. Tree Genet Genomes 5:435–446. https://doi.org/10.1007/s11295-008-0197-2

Merot-L'Anthoene V, Tournebize R, Darracq O, Rattina V, Lepelley M, Bellanger L, Tranchant-Dubreuil C, Coulée M, Pégard M, Metairon S, Coralie F, Stoffelen P, Janssens SB, Kiwuka C, Musoli P, Sumirat U, Legnaté H, Kambale J-L, da Costa Neto JF, Revel C, de Kochko A, Descombes P, Crouzillat D, Poncet V (2019) Development and evaluation of a genome-wide coffee 8.5K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. Plant Biotechnol J 17:1418–1430. https://doi.org/10.1111/pbi.13066

Meyer FG (1965) Notes on wild *Coffea arabica* from southwestern Ethiopia, with some historical considerations. Econ Bot 19:136–151. https://doi.org/10.1007/BF02862825

Meyer FG, Fernie LM, Narasimhaswamy RL, Monaco LC, Greathead DJ (1968) FAO coffee mission to Ethiopia 1964–1965. FAO, Rome

Missio RF, Caixeta ET, Zambolim EM et al (2009) Development and validation of SSR markers for *Coffea arabica* L. Crop Breed Appl Biotechnol 9:361–371

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pruvot-Woehl S, Krishnan S, Solano W, Schilling T, Toniutti L, Bertrand B, Montagnon C (2020) Authentication of *Coffea arabica* varieties through DNA fingerprinting and its significance for the coffee sector. J AOAC Int 103:325–334. https://doi.org/10.1093/jaocint/qsz003

Razafinarivo NJ, Guyot R, Davis AP, Couturon E, Hamon S, Crouzillat D, Rigoreau Dubreuil-Tranchant MC, Poncet V, Kochko AD, Rakotomalala J-J, Hamon P (2013) Genetic structure and diversity of coffee (*Coffea*) across Africa and the Indian Ocean islands revealed using microsatellites. Ann Bot 111:229–248. https://doi.org/10.1093/aob/mcs283

Sant'Ana, GC, Pereira LFP, Pot D, Ivamoto ST, Domingues DS, Ferreira RV, Pagiatto NF, da Silva BSR, Nogueira LM, Kitzberger DSG, Scholz MBS, de Oliveira FF, Sera GH, Padilha L, Labouisse JP, Guyot R, Charmetant P, Leroy T (2018) Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in Coffea arabica L. Sci Rep 8: 465. https://doi.org/10.1038/s41598-017-18800-1

Scalabrin S, Toniutti L, Di Gaspero G, Scaglione D, Magris G, Vidotto M, Pinosio S, Cattonaro F, Magni F, Jurman I, Cerutti M, Liverani FS, Navarini L, Del Terra L, Pellegrino G, Ruosi MR, Vitulo N, Valle G, Pallavicini A, Graziosi G, Klein PE, Bentley N, Murray S, Solano W, Al Hakimi A, Schilling T, Montagnon C, Morgante M, Bertrand B (2020) A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. Sci Rep 10:4642. https://doi.org/10.1038/s41598-020-61216-7

Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, Pereira TV, Sakiyama NS, de Resende Júnior MFR, Zambolim L (2017) Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. Tree Genet Genomes 13:124. https://doi.org/10.1007/s11295-017-1208-y

Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, Pereira AA, Sakiyama NS, Zambolim L, Resende MDV (2019) Early selection enabled by the implementation of genomic selection in *Coffea arabica* breeding. Front Plant Sci 9:19–34. https://doi.org/10.3389/fpls.2018.01934

Thomas AS (1942) The wild Arabica coffee on the Boma Plateau, Anglo-Egyptian Sudan. Empire J Exp Agric 10:207–212

Tran HT, Lee LS, Furtado A, Smyth H, Henry RJ (2016) Advances in genomics for the improvement of quality in coffee. J Sci Food Agric 96:3300–3312. https://doi.org/10.1002/jsfa.7692

Tran HTM, Furtado A, Vargas CAC, Smyth H, Lee LS, Henry R (2018) SNP in the *Coffea arabica* genome associated with coffee quality. Tree Genet Genomes 14:72. https://doi.org/10.1007/s11295-018-1282-9

USDA-FAS (United States Department of Agriculture – Foreign Agriculture Service) (2020) Coffee: world markets and trade. https://apps.fas.usda.gov/psdonline/circulars/coffee.pdf" https://apps.fas.usda.gov/psdonline/circulars/coffee.pdf.

Vega FE, Rosenquist E, Collins W (2003) Global project needed to tackle coffee crisis. Nature 425:343. https://doi.org/10.1038/425343a

Vega FE (2008) The rise of coffee. Am Sci 96:138–145. https://doi.org/10.1511/2008.70.3640

Vieira ESN, Von Pinho ÉVDR, Carvalho MGG, Esselink DG, Vosman B (2010) Development of microsatellite markers for identifying Brazilian *Coffea arabica* varieties. Genet Mol Biol 33:507–514. https://doi.org/10.1590/S1415-47572010005000055

Yuyama PM, Júnior OR, Ivamoto ST, Domingues DS, Carazzolle MF, Pereira GAG, Charmetant P, Leroy T, Pereira LFP (2016) Transcriptome analysis in *Coffea eugenioides*, an Arabica coffee ancestor, reveals differentially expressed genes in leaves and fruits. Mol Gen Genomics 291:323–336. https://doi.org/10.1007/s00438-015-1111-x

Zhou L, Vega FE, Tan H, Lluch AER, Meinhardt L, Fang W, Mischke S, Zhang D (2016) Developing single nucleotide polymorphism (SNP) markers for the identification of coffee germplasm. Tropical Plant Biol 9:82–95. https://doi.org/10.1007/s12042-016-9167-2