RESEARCH RESULTS FOR THE YEAR 2015: GENOMIC PROFILING AND DEVELOPMENT OF A COMPREHENSIVE CATALOGUE OF PLUM GERMPLASM USING GENOTYPING-BY-SEQUENCING (GBS)

Chris Dardick[1], Tetyana Zhebentyayeva[2,3], Chris Saski[2,3], Ralph Scorza[1], Ann Callahan[1], Michael Rovelandro[4], and Ted DeJong[5]

[1] USDA Appalachian Fruit Research Laboratory, Kearneysville, WV 25430, USA
[2] Clemson University Genomics & Computational Biology Laboratory, Clemson, SC 29634
[3] Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634
[4] UMR BFP1332 - INRA-Universite Bordeaux II, Villenave d'Ornon, France
[5] Department of Plant Sciences, University of California, Davis, CA 95616, USA

ABSTRACT

Over last few decades the US prune industry has successfully maintained strong and vibrant export markets in Europe, Canada and Japan. However, in recent years international competition has been steadily eroding the share of US world prune exports. A long-term strategy to retain market dominance is to genetically improve California prune cultivars using recent advancements in fruit tree genetics to maintain superior fruit quality, productivity and sustainability of prune orchards. Moreover, superior and genetically distinct CA prune germplasm would enable refined labeling and packaging strategies that would help the US industry distinguish its high-quality prunes and prune products from its competitors. Likewise, targeted improvements of health promoting compounds such as antioxidants and sugars may also help the US to keep a preeminent position in a challenging world marketplace. Establishing molecular marker profiles using high-throughput genotyping by sequencing will support germplasm analysis and search for markers associated with health-promoting compounds that can be then implemented in marker-assisted breeding. Molecular marker profiles can be employed for cultivar authentication and labeling the US prunes on the international market. Also, breeding such new and superior varieties requires understanding of the origins of the major US cultivars, in particular French type, and their genetic relationships with worldwide prune germplasm.

INTRODUCTION

Currently most international sellers of dried plums market them as the 'd'Agen' variety (syn. d'Ente). Traditional cultivar identification based on morphological and phenological characteristics describes the d'Ente plum (Agen prunes) as a medium-sized fruit, rather rounded, orange to purplish red, with juicy flesh, fine and tender, very sweet, green, yellow. The d'Ente plum is self-fertile (it is also a good pollinator); it is ripe in late August. The tree tolerates all soils, except those too rich in clay. However, based on this description it is not always possible to discriminate cultivars with similar pomological characters but having a different genetic background. The d'Ente plum trees were

introduced to Europe from Middle East by crusaders and were named for the district 'Agen' where they were planted nearly 800 years ago. Over the centuries, a number of varieties with distinct characters were propagated under the name d'Agen and subsequently imported into the US (California) in 1856. Many of these d'Agen types, including several numbered clones of d'Ente and petite d'Agen, are thought to be clonal material (bud mutations) of an original d'Agen type however, hybridization with other related cultivars within this group of cultivars cannot be ruled out. The most commonly grown cultivar in the US, Improved French, is considered as open pollinated Agen prune seedlings released by Luther Burbank around 1900. This variety may have been subsequently improved over the last century through selection of clonally propagated materials as well as through self-pollination.

OBJECTIVES

The objective of this study was to determine the genetic relationship of the main industrial US cultivar 'Improved French' to other commercial germplasm that is used worldwide. The d'Ente (Agen) prunes and Improved French are being analyzed in a set of cultivars from different morphological groups of plums maintained at germplasm repository at INRA, Bordeaux representing one of the oldest, largest in diversity and best characterized plum germplasm sources. In addition, wild *Prunus* relatives were also be included in the study in attempt to determine the genetic relations of hexaploid *Prunus domestica* and its potential wild progenitors, diploid *Prunus cerasifera* and *Prunus spinosa*. This information is of great importance in choosing parents in the breeding program.

PROCEDURE

The objectives were accomplished using a new technique called Genotype-By-Sequencing (GBS). GBS is method that leverages the power of next generation DNA sequencing technologies to assess the genetic relationships among large numbers of individuals. This strategy reduces the costs of whole genome next generation sequencing using a technique that limits the genome sequencing to a smaller number of informative regions. In this way, the entire genome of each individual is not sequenced by rather a large number of snippets that typically include the regions containing genes. The platform is scalable to 48 or 96 individuals per sequencing run. Here, DNA from 192 individual plum samples was extracted and used for GBS analysis. A very general outline of the procedure is given below:

1) Obtain tissue samples from orchards and germplasm repositories.
2) Extract DNA.
3) Generate "barcoded" DNA libraries for sequencing. Barcodes are introduced by adding of 4-8 nucleotides to DNA fragments to discriminate all sequences derived form individual accession)
4) Send libraries to reputable service provider (Illumina HiSeq instrument).
5) Assign sequences to individual accessions using barcodes

6) Perform trimming and data quality control steps.
7) Assemble sequences from each sample to the reference genome (peach).
8) Identify sequence variations from each assembly (single nucleotide polymorphisms or SNPs).
9) Compare SNP profiles of all samples and generate relationship tree (i.e. Dendrogram) and PCA plot based on Principal Component Analysis.

The work reports a fist-year data for 2-year proposal submitted in 2014 and approved by CDPB to support a study executed in cooperation between AFRI USDA and Clemson University Computational Biology & Genomics Laboratory. It should be noted that Clemson University donated additional bioinformatic resources to this project in the form personnel time and computational time on CUGI servers.

RESULTS AND CONCLUSIONS

The following specific accomplishments are reported for 2015:

- 192 plum accessions in total were genotyped by sequencing. Samples represented main pomological groups of plum from USDA ARS National Germplasm Repository, Davis CA, Ted DeJong, University of CA, Davis, Ralph Scorza, USDA ARS Appalachian Fruit Research Station, Kearneysville WV, and the French National Institute for Agricultural Research (INRA), Bordeaux, France (Table1). Three accessions of Improved French from different repositories were included into analysis for verification their identity and genetic relations within European germplasm including 13 original d'Agen prunes from France. List of the 192 cultivars assigned to pomological groups is given in Table 2.
- GBS data was processed and analyzed using bioinformatic pipeline shown on Fig. A1 in Appendix with technical details. Summary of data processing for individual accessions are in Table 2.
- A total of 84,923 SNP markers were used for study genetic relations of d'Agen prunes and Improved French within European germplasm. Clustering of d'Agen varieties on dendrogram was in agreement with their separation from other pomological groups (greengages, damsons and mirabelles) on the Principal Component Analysis plot.
- Potential synonymous cultivars as well as mislabeled cultivars and sampling mistakes were pointed out based on established molecular profiles.

Plum2015 plant material

Table1. Summary of plant material in pum 2015 dataset

pomological group	number cultivars
Prune (d'Agen group)	27
Improved French	3
Greengage	28
Damson and Mirabelle	25
Prunus domestica	107
diploid *Prunus* species	2
Total	192

Dendrogram

SNP profiles for 187 samples were compared and used to generate a complete dendrogram on Fig.1. Plum accessions were separated into 6-7 groups. Noticeable, d'Agen prunes including Improved French from USA and France were clustered into distinct group on dendrogram separately from other groups of European plums.

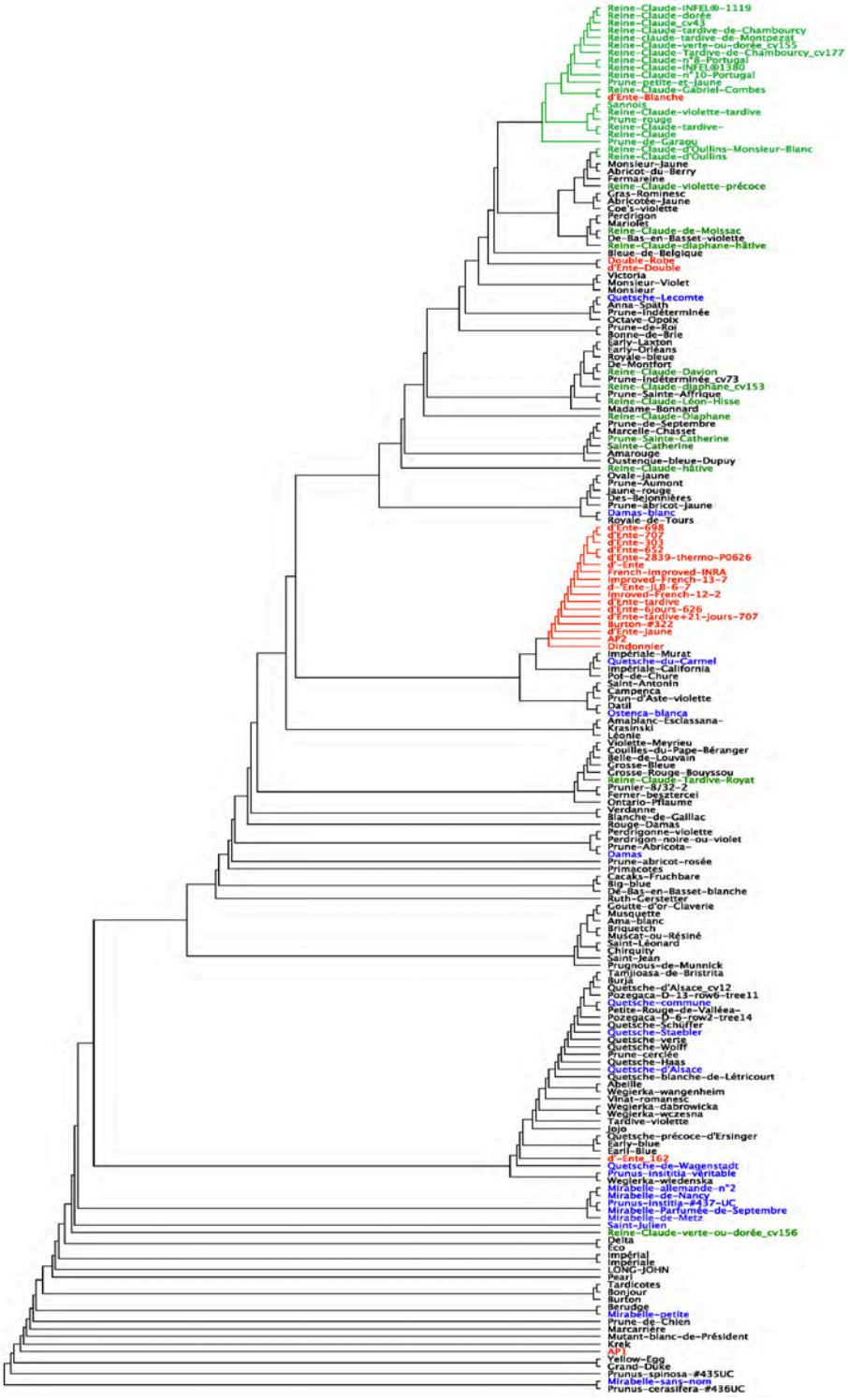

Fig.1 Dendrogram of genetic similarity 187 plum accessions estimated over 23,010 polymorphic SNP markers. Colors indicate pomological groups: red (Prune), green (Greengage), blue (Damson and Mirabelle), black hexaploid *Prunus domestica* (European plum) and diploid species *P. cerasifera* and *P. spinosa*.

Principal Component Analysis (PCA plot)

Genetic variation in plum2015 dataset are shown on PCA plot on Fig.2. The separation of clusters was not absolute though most of d'Agen prunes (red color) formed a distinct group in lower part of the plot. Greengages (green color) also created distinct group on a plot were less stratified indicating potential hybridization with other pomological groups.

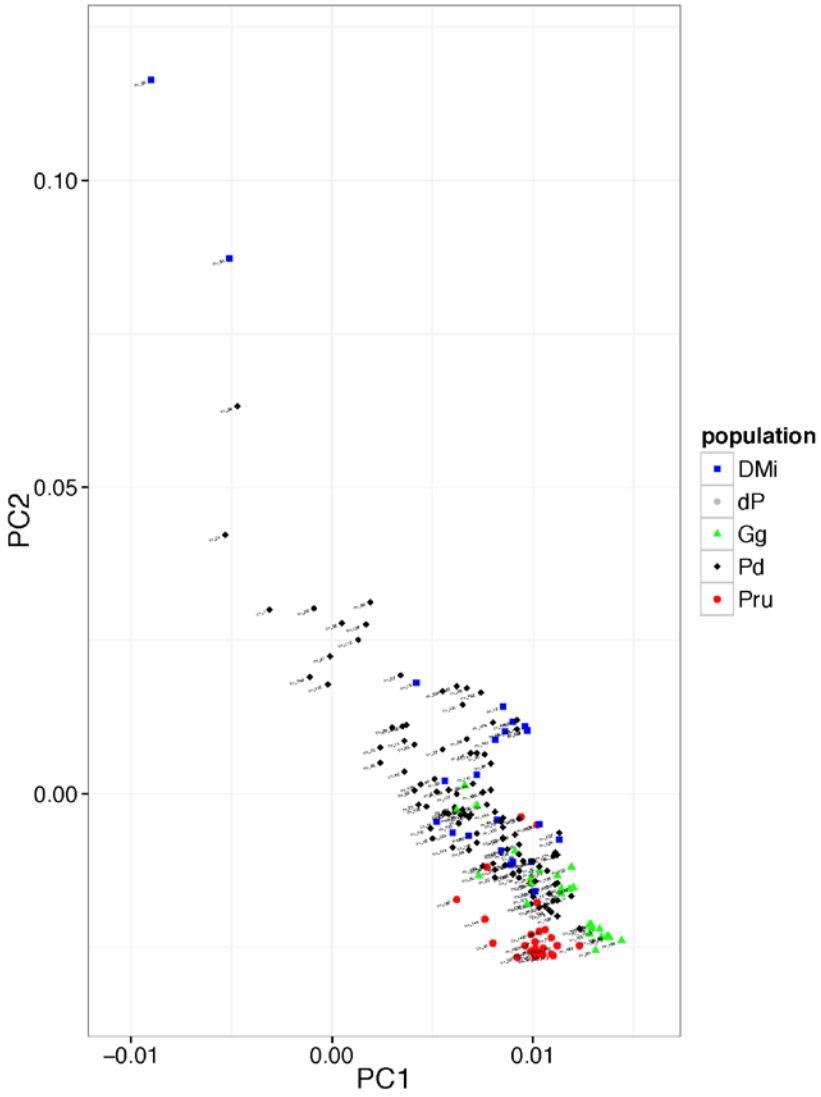

Fig. 2. Principal Component Analyses (PCA) of 187 plum accessions. Distribution of plum varieties on the two first PCA axes determined from 45,593 SNP markers. Colors and symbols represent pomological groups of plum: red (Pru, prune), green (Gg, greengage), blue (DMi, damson and mirabelle), black (Pd, *Prunus domestica* /European plum), grey (dP, diploid *P. cerasifera* and *P. spinosa*). Diploid species are not shown.

Conclusions

Analysis of the GBS data and the resulting dendrogram and PCA plot produced from reference assembly to the peach genome reveal the following conclusions:

1) The GBS strategy successfully predicted most of the known genetic relationships among plum varieties. Several known synonymous cultivars were detected as well as potentially mislabeled cultivars and technical sampling mistakes. For example, our results indicated that the *P. cerasifera* accession from INRA was mislabeled as plum Mirabelle sans nom (unknown Mirabelle).
2) On the dendrogram all d'Agen prunes but d'Ente_cv162 and d'Ente Double grouped together. This cluster also included three accessions of Improved French and other cultivars of a prune type. Positioning of d'Ente_cv162 on the dendrogram and the PCA plot may be explained by sampling/labeling mistake while d'Ente Double and synonymous cultivar Double Robe may have d'Agen prunes in their pedigree.
3) The group of the d'Agen prunes is not homogeneous. Many d'Agen plum types appear to be seedlings from self-pollination (if grouped in the vicinity of d'Ente clones) or cross-pollination with other plums (grouped in adjacent clusters). Sequencing of the progeny from self-pollination of Improved French should help in setting statistical threshold for separation F2 progeny from clonally propagated material.
4) The 2015 data gave more support for the hypothesis that *Prunus domestica* originated from hybridization between *Prunus spinosa* and *Prunus cerasifera*. Moreover, PCA analysis indicated 4 plum cultivars that could originate from hybridization of hexaploid plum with its progenitor *P. spinosa*. These are cv_34 - Prunus insititia (veritable), cv_35-Saint Julien, cv_91 - Wegierka wiedenska and cv_99 - Prune de Chien. The role of *P. cerasifera* in origin of hexaploid plum is still unclear. Alignment of GBS sequences against the chloroplast genome of peach and estimate of genetic relations within the plum 205 dataset may highlight maternal lineage of inheritance.

BUDGET NARRATIVE

Funding is still in process due to administrative issues. Clemson University and AFRI USDA have temporarily covered the costs of sequencing and labor for data analysis.

Table2.Plum GBS sttistics ondara processing and SNP genotyping

	ID	cultivar	Raw_reads	Mapped_reads	%mapped	SNP_SITES	MEAN_DEPTH per SNP
1	cv_1	Briquetch	2705542	1800612	66.6%	121109	63
2	cv_2	Bonne de Brie	5765192	3846158	66.7%	130639	92
3	cv_3	Bonjour	1955844	1249989	63.9%	120220	53
4	cv_4	Blanche de Gaillac	5188356	3343022	64.4%	126067	79
5	cv_5	Berudge	2984944	1888932	63.3%	114427	59
6	cv_6	Prune d'Aste violette	5996580	4076494	68.0%	131942	96
7	cv_7	D'Ente Double	6380142	4322125	67.7%	134051	97
8	cv_8	D'Ente jaune	3126714	2079045	66.5%	118794	68
9	cv_9	D'Ente 707	4382766	2947901	67.3%	125597	80
10	cv_10	Datil	2546604	1631715	64.1%	122306	67
11	cv_11	De Montfort	4586498	2939786	64.1%	125877	77
12	cv_12	Quetsche d'Alsace	4492318	2951858	65.7%	129519	90
13	cv_13	Quetsche de Wagenstadt	1763234	1110815	63.0%	118628	51
14	cv_14	Léonie	5805786	3978759	68.5%	135684	100
15	cv_15	Krek	9080518	6266601	69.0%	150225	94
16	cv_16	Jaune rouge	3075876	1983902	64.5%	131522	76
17	cv_17	Impériale Murat	5098994	3435881	67.4%	130063	82
18	cv_18	Quetsche blanche de Létricourt (thermo P1844)	640470	414119	64.7%	101036	28
19	cv_19	Mirabelle de Metz	1804850	1150702	63.8%	115520	50
20	cv_20	Mirabelle de Nancy	5555766	3704162	66.7%	133857	92
21	cv_21	Mirabelle Parfumée de Septembre	4899358	3327973	67.9%	130548	77
22	cv_22	Reine-Claude d'Oullins (Monsieur Blanc)	6249092	4221666	67.6%	136599	107
23	cv_23	Monsieur	4848236	3210534	66.2%	127263	80
24	cv_24	Reine-Claude Davion	3935272	2573000	65.4%	123445	77
25	cv_25	Reine-Claude de Moissac	5476820	3721008	67.9%	130916	107
26	cv_26	Reine-Claude diaphane hâtive	5124012	3483652	68.0%	131838	101
27	cv_27	Reine-Claude Diaphane	6567434	4415818	67.2%	128895	89
28	cv_28	Burja row2 tree1	4416244	2734451	61.9%	135832	107
29	cv_29	Reine-Claude d'Oullins	6696378	4631718	69.2%	131988	105
30	cv_30	Reine-Claude Gabriel Combes	7400342	5151238	69.6%	130107	101
31	cv_31	Reine-Claude Tardive Royat	8779388	6016328	68.5%	134155	116
32	cv_32	Reine-Claude INFEL® 1330	780016	482024	61.8%	92918	32
33	cv_33	Reine-Claude violette (tardive)	5719308	3845654	67.2%	133207	97
34	cv_34	Prunus insititia (véritable)	3900336	2486452	63.7%	121790	74
35	cv_35	Saint Julien	3151914	1953037	62.0%	125711	77
36	cv_36	Ruth Gerstetter	8344304	5467102	65.5%	141890	129
37	cv_37	Reine-Claude INFEL® 1380	4959056	3189636	64.3%	132152	91
38	cv_38	Burton	2179878	1401253	64.3%	110883	54
39	cv_39	Coe's violette	5843122	3992495	68.3%	129852	100
40	cv_40	French improved (INRA)	6049816	3887610	64.3%	137335	126
41	cv_41	D' Ente	4464736	2959137	66.3%	127122	81
42	cv_42	Prunier 8/32-2	5187104	3470433	66.9%	123320	84
43	cv_43	Reine claude	5070314	3346983	66.0%	121232	78
44	cv_44	Ferner besztercei	5386018	3617339	67.2%	126658	88
45	cv_45	Reine Claude tardive	3749646	2492158	66.5%	119692	69
46	cv_46	Reine Claude vraie	723920	451329	62.3%	94058	30
47	cv_47	Improved French 13-7	4606330	3142001	68.2%	121492	75
48	cv_48	Monsieur Jaune	5612714	3789588	67.5%	122756	87
49	cv_49	Abricotée Jaune	4061126	2683478	66.1%	123515	84
50	cv_50	D'Ente 698	5912356	3985375	67.4%	127755	90
51	cv_51	Double Robe	3513270	2342642	66.7%	116411	68
52	cv_52	D'Ente 2839 (thermo P0626)	4191530	2835192	67.6%	119396	74
53	cv_53	Prune Abricota	4351536	2977753	68.4%	123798	84
54	cv_54	Abricot du Berry	5162966	3475628	67.3%	120031	78
55	cv_55	Prune abricot rosée	4576982	3101119	67.8%	121594	77
56	cv_56	Ama blanc	2725492	1789725	65.7%	123684	73
57	cv_57	Amablanc Esclassana	4199272	2907076	69.2%	126049	86
58	cv_58	Amarouge	3006104	2044763	68.0%	120196	68
59	cv_59	D'Ente Blanche	4201802	2757258	65.6%	119839	72
60	cv_60	Quetsche blanche de Létricourt	6592478	4625490	70.2%	126137	94
61	cv_61	Damas blanc	4371720	3076038	70.4%	124102	81
62	cv_62	Damas	3514718	2378617	67.7%	116529	65
63	cv_63	Couilles du Pape Béranger	4262476	2912152	68.3%	120545	76
64	cv_64	Goutte d'or (Claverie)	6385326	4405893	69.0%	131532	96
65	cv_65	Chirquity	4253392	2776684	65.3%	129123	91
66	cv_66	Campenca	4433666	2930056	66.1%	133318	102
67	cv_67	Quetsche du Carmel	4860256	3219493	66.2%	121403	77
68	cv_68	Quetsche verte	4675838	3112019	66.6%	131446	103
69	cv_69	De Bas-en-Basset blanche	2136876	1479463	69.2%	113047	57
70	cv_70	De Bas-en-Basset violette	3101748	2038261	65.7%	124511	82

Table2.Plum GBS sttistics ondara processing and SNP genotyping (continued)

	ID	cultivar	Raw_reads	Mapped_reads	%mapped	SNP_SITES	MEAN_DEPTH per SNP
71	cv_71	Des Bejonnières	3796194	2614694	68.9%	123166	80
72	cv_72	Dindonnier	6312668	4349544	68.9%	131287	93
73	cv_73	Prune indéterminée	4834556	3150859	65.2%	124523	83
74	cv_74	Mariolet	4702864	3222675	68.5%	128377	93
75	cv_75	Monsieur Violet	7336384	5085299	69.3%	127221	93
76	cv_76	Muscat ou Résiné	4732094	3062083	64.7%	119175	70
77	cv_77	Musquette	4937584	3397255	68.8%	124752	81
78	cv_78	Ostenca blanca	2663972	1745734	65.5%	115698	63
79	cv_79	Reine-Claude dorée	4417232	2886273	65.3%	117327	72
80	cv_80	Oustenque bleue (Dupuy)	3956100	2708351	68.5%	121920	76
81	cv_81	Petite Rouge de Valléea	4859352	3325766	68.4%	118742	79
82	cv_82	Pot de Chure	4902566	3258168	66.5%	122329	76
83	cv_83	Prune petite et jaune	4934706	3356435	68.0%	128358	89
84	cv_84	Royale de Tours	6177106	4142400	67.1%	133590	99
85	cv_85	Madame Bonnard	2770836	1831572	66.1%	112000	62
86	cv_86	Krasinski	4851570	3259228	67.2%	121395	77
87	cv_87	Reine Claude	5716178	3905296	68.3%	128497	96
88	cv_88	Reine claude tardive de Montpezat	2411120	1662548	69.0%	116069	66
89	cv_89	Belle de Louvain	3469652	2200057	63.4%	124065	76
90	cv_90	Wegierka wczesna	3634154	2339327	64.4%	119591	70
91	cv_91	Wegierka wiedenska	3602032	2396791	66.5%	120582	71
92	cv_92	Mirabelle petite	2257652	1471041	65.2%	108995	53
93	cv_93	Big blue	4135410	2716552	65.7%	117579	71
94	cv_94	Early blue	3949012	2612532	66.2%	121318	68
95	cv_95	Reine Claude tardive de Chambourcy	4417078	2907227	65.8%	118343	72
96	cv_96	Quetsche Staebler	4810650	3359860	69.8%	125066	74
97	cv_97	Prugnous de Munnick	1594730	1032734	64.8%	113492	49
98	cv_98	Prune Aumont	4197594	2771037	66.0%	123202	78
99	cv_99	Prune de Chien	2443232	1551646	63.5%	115899	55
100	cv_100	Prune de Garaou	5515950	3633739	65.9%	122497	82
101	cv_101	Prune de Septembre	2983780	1945327	65.2%	113994	65
102	cv_102	Royale bleue	4930620	3371499	68.4%	132493	96
103	cv_103	Grosse Bleue	2716994	1784613	65.7%	124837	72
104	cv_104	Grosse Rouge (Bouyssou)	3334802	2209312	66.3%	120792	74
105	cv_105	Saint Jean	2682846	1727751	64.4%	119996	63
106	cv_106	Prune indéterminée	3778922	2423811	64.1%	119876	73
107	cv_107	Octave Opoix	4348388	2858326	65.7%	124645	83
108	cv_108	Saint Antonin	5813134	3787621	65.2%	128940	96
109	cv_109	Saint Léonard	2636186	1751728	66.4%	123600	68
110	cv_110	Verdanne	4761204	3203524	67.3%	130973	83
111	cv_111	Victoria	5470332	3734611	68.3%	128469	88
112	cv_112	Abeille	3913914	2661752	68.0%	124490	78
113	cv_113	Perdrigon	3626874	2392535	66.0%	130339	93
114	cv_114	Ovale jaune	4933482	3330342	67.5%	122945	79
115	cv_115	Perdrigonne violette	2862054	1834180	64.1%	120374	64
116	cv_116	Impériale California	5565344	3753075	67.4%	128884	103
117	cv_117	Quetsche commune	6819518	4663489	68.4%	127014	92
118	cv_118	D'Ente 303	4436492	2989146	67.4%	122122	77
119	cv_119	D'Ente 652	3682606	2432298	66.0%	117665	71
120	cv_120	Reine-Claude INFEL® 1119	3198878	2107308	65.9%	120129	72
121	cv_121	Mirabelle sans nom (Mirabelle unnamed)	3433042	2272310	66.2%	116096	101
122	cv_122	Marcelle Chasset	4250494	2861827	67.3%	119431	74
123	cv_123	Sannois	4769132	3168277	66.4%	133999	106
124	cv_124	Quetsche Lecomte	4632256	3026836	65.3%	129633	92
125	cv_125	Early Laxton	4884694	3291370	67.4%	130584	98
126	cv_126	Gras Rominesc	4114464	2775237	67.5%	134761	104
127	cv_127	Early Orléans	4370694	2825698	64.7%	128984	87
128	cv_128	Fermareine	3826708	2487112	65.0%	124651	83
129	cv_129	Reine-Claude violette (précoce)	5823444	3969881	68.2%	128140	94
130	cv_130	Violette Meyrieu	4091932	2675272	65.4%	118787	73
131	cv_131	Yellow Egg	5423686	3643780	67.2%	140751	74
132	cv_132	Impérial	6389882	4227014	66.2%	148327	79
133	cv_133	Grand Duke	8102440	5646573	69.7%	161930	80
134	cv_134	Rouge Damas	3134524	2085480	66.5%	123809	66
135	cv_135	Prune de Roi	5466956	3744636	68.5%	134947	89
136	cv_136	Bleue de Belgique	4631614	3109262	67.1%	123914	82
137	cv_137	Sainte Catherine	5130082	3528800	68.8%	135046	93
138	cv_138	Earli Blue	1778926	1177649	66.2%	110581	52
139	cv_139	Ontario Pflaume	3936556	2635973	67.0%	122767	77
140	cv_140	Cacaks Fruchbare	3564202	2432577	68.3%	118054	70

Table2.Plum GBS sttistics ondara processing and SNP genotyping (continued)

	ID	cultivar	Raw_reads	Mapped_reads	%mapped	SNP_SITES	MEAN_DEPTH per SNP
141	cv_141	Vinat romanesc	2505688	1666911	66.5%	121048	66
142	cv_142	Quetsche précoce d'Ersinger	2706380	1740098	64.3%	119033	63
143	cv_143	Imroved French 12-2	4247854	2855671	67.2%	120394	70
144	cv_144	AP1	3588044	1779264	49.6%	113573	52
145	cv_145	AP2	3201318	1844617	57.6%	119344	57
146	cv_146	Wegierka dabrowicka	3995760	2641966	66.1%	124635	78
147	cv_147	Wegierka wangenheim	3811956	2556479	67.1%	120482	74
148	cv_148	Prune rouge	3454196	2324629	67.3%	116567	68
149	cv_149	Perdrigon noire ou violet	4724966	3148613	66.6%	132451	81
150	cv_150	Jojo	6374764	4225609	66.3%	136279	109
151	cv_151	d'Ente tardive +21 jours 707	15928054	10711449	67.2%	137560	112
152	cv_152	Reine Claude hâtive	4322236	2789291	64.5%	126518	78
153	cv_153	Reine Claude diaphane	9002384	6025286	66.9%	133899	95
154	cv_154	Pearl	3871354	2549750	65.9%	127470	66
155	cv_155	Reine Claude verte ou dorée	3274134	2141178	65.4%	122072	67
156	cv_156	Reine Claude verte ou dorée	6216320	4240469	68.2%	131212	83
157	cv_157	Tardive violette	4597874	3099178	67.4%	129676	76
158	cv_158	Prune d'Ente tardive	3614864	2350489	65.0%	130242	75
159	cv_159	Primacotes	6264166	4256772	68.0%	123462	78
160	cv_160	Quetsche Haas	7241428	5009031	69.2%	139317	84
161	cv_161	D 'Ente JLB 6-7	6567876	4294261	65.4%	129028	85
162	cv_162	D' Ente	5379934	3518531	65.4%	125274	72
163	cv_163	Quetsche d'Alsace	6075976	4085744	67.2%	134874	77
164	cv_164	Mirabelle allemande n°2	5086702	3234844	63.6%	127565	78
165	cv_165	Tardicotes	2853824	1928345	67.6%	121097	61
166	cv_166	Spurdente	158578	76034	47.9%	74244	8
167	cv_167	Mutant blanc de Président	5431622	3638989	67.0%	128852	70
168	cv_168	Reine claude n°10 Portugal	7522142	5084456	67.6%	141835	92
169	cv_169	LONG -JOHN	4506834	2898897	64.3%	129628	76
170	cv_170	Prune abricot jaune	3797192	2482618	65.4%	129615	80
171	cv_171	Anna Späth	3654492	2450496	67.1%	126078	75
172	cv_172	Marcarrière	3485212	2208892	63.4%	121091	60
173	cv_173	Prune Sainte Catherine	4222234	2857000	67.7%	129399	76
174	cv_174	Prune cerclée	8733454	5859338	67.1%	138828	91
175	cv_175	Prune Sainte-Affrique	4614500	2963108	64.2%	128918	79
176	cv_176	Reine-Claude Léon Hisse	2990272	2006025	67.1%	128009	75
177	cv_177	Reine-Claude Tardive de Chambourcy	6112676	3999494	65.4%	123684	82
178	cv_178	Quetsche Schüffer	5144880	3305104	64.2%	138729	87
179	cv_179	Quetsche Wolff	6712864	4487575	66.9%	139524	88
180	cv_180	Delta	6103518	4017812	65.8%	136216	88
181	cv_181	Eco	1947162	1185876	60.9%	123114	56
182	cv_182	d'Ente -6jours 626	3912674	2567727	65.6%	132108	80
183	cv_183	Impériale	4359226	2936696	67.4%	129294	81
184	cv_184	Reine claude n°8 Portugal	6021280	3939567	65.4%	138985	102
185	cv_185	Prunus-spinosa #435UC	3603942	2118179	58.8%	124233	82
186	cv_186	Prunus-institia #437 UC	2637704	1625679	61.6%	116621	61
187	cv_187	Tamjioasa de Bristrita row1 tree4	5465262	3475531	63.6%	136570	118
188	cv_188	Victory row10 tree13	258062	111754	43.3%	77123	11
189	cv_189	Burton #322	2529824	1591576	62.9%	125018	69
190	cv_190	Prunus-cerasifera #436UC	4849240	3209898	66.2%	114217	101
191	cv_191	Pozegaca D-6 row2 tree14	2722862	1658475	60.9%	129345	73
192	cv_192	Pozegaca D-13 row6 tree11	2621360	1575989	60.1%	130509	71
		in total	**243027120**	**159351260**	**66.53%**	**125,116**	**79**

Appendix (technical details)

GBS library preparation and sequencing

In genotyping by sequencing experiments (GBS) in complex plant genomes, the enzyme choice for genomic selection is critical in collecting the sequencing depth of coverage necessary for dense marker distribution. Using bioinformatic tools we virtually digested a closely related peach genome to predict a number of restriction sites with most common enzymes used for GBS. In 2014, based on this '*in silico'* prediction, we selected *Pst*1 restrictase and successfully genotyped 96 plum accessions by sequencing DNA fragments in one direction. Though successful, this experiment has shown that only 40% of fragments were distributed in a range of 200-400bp, which is considered as optimal for unidirectional sequencing. Significant proportion of fragments (28%) was distributed in the range of 400-700bp and technically can be sequenced only using sequencing from both ends deemed as paired–end sequencing. Thus, in 2015 we modified previous protocol and sequenced fragment from both ends to incorporate into analyses information from large fragments (up to 700-800bp). Total genomic DNA was prepared for 192 accessions and digested with *Pst*I. Restriction fragments (200-700 bp) were selected, individually indexed and tailed with Illumina sequencing adapters. In post library treatments accessions were multiplexed at level 48x per lane on four flow cell lanes of an Illumina HiSeq2500 (Illumina) on high output mode using a single-end 2x125bp run cycle.

Demultiplexing, filtering, and coverage

Raw plum2015 data were processed using bioinformatic pipeline shown on Fig.A1. A total of 886,240,424 reads were collected from 4 lanes of Illumina sequencing. Of these, 867,727,768 (97,9%) reads were deemed good due to presence barcode and the *Pst*1 restriction site. This suggested a high quality of DNA and sequencing library preparations, as well as efficient clustering and sequencing on Illumina instrument. A total of 574,294,165 GBS reads (66.53%) were aligned to the peach reference genome (*Prunus persica* v2.0) using Burrows-Wheeler Aligner (Li and Durbin 2009). Plum accessions were genotyped using software package Stacks (Catchen et al., 2013). In average 125,116 total polymorphic sites per accession were detected with average density of 79 reads per site. Five accession 'Quetsche blanche de Létricourt' (cv_18), 'Reine-Claude INFEL® 1330' (cv_32), 'Reine Claude vraie' (cv_46), 'Spurdente' (cv_166) and 'Victory row10 tree13' (cv_188) were excluded from downstream analyses because of low density SNP reads (<30x).

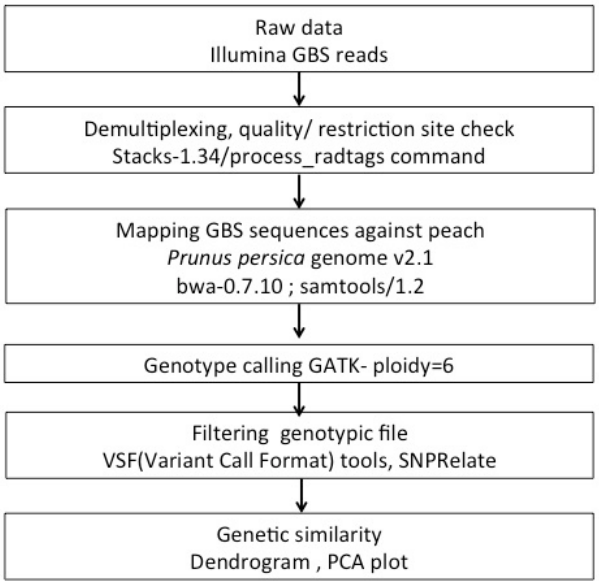

Fig.A1. Plum data processing pipeline

Dendrogram construction

In total 136,120 polymorphic variants were detected and 84,923 SNPs passed basic quality check. Of these, 51,291 markers were excluded after additional filtering at default settings (monomorphic "TRUE", MAF <5%, missing rate <10%). Retained 23,010 SNPs were used to calculate dissimilarity matrix and construct dendrogram with R package SNPRelate (Zheng et al., 2012).

Principal Component Analysis (PCA plot)

To estimate genetic separation pomological groups we also generated 45,593 SNP markers using less stringent criteria for frequency of minor alleles (<0.001%). PCA analysis was conducted using the R software packages 'SNPRelate' (Zheng et al., 2012).

References

Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W (2013) Stacks: an analysis tool set for population genomics. Molecular Ecology, 22: 3124-3140.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28: 3326–3328.