

1 **S1 Supporting Information: Sub-experiment on genotypes diversity in fields and feral**
2 **populations.**

3

4 **Background**

5 The goal of this sub-experiment was to assess the quality of the experimental protocol. By
6 sampling plants on the border of fields, we could collect more genetic heterogeneity than inside fields.
7 Moreover, sampling 10 plants could not be representative of the genetic diversity of a whole field or
8 feral population. These questions were addressed from an extent sampling design compared to regular
9 sampling design.

10

11 **Materials and Methods**

12 In 2007, a supplementary sampling operation was performed on 5 fields and 3 feral
13 populations. 60 leaves were collected in the fields along transects (one leaf per 5 meters and an
14 average of 15 leaves per transects), and 50 leaves (except F222, 60 leaves) were collected for the feral
15 populations (see S1 Figure A for sampling design).

16 The leaves were treated with the same methodology present in the current paper. We obtained
17 genotypes from 8 SSR. Using cultivar in the analysis was not possible. We tried to assign our samples,
18 as the dataset contained almost no missing values. Less than half of the genotypes were assigned to a
19 cultivar. In consequence, the biologic unit here is the genotype. We thus examined the distribution of
20 genotypes along transects for the 5 fields and the 3 feral populations.

21 To assess the statistical significance of our current protocol (10 plants), we realized for each
22 field and for each population genotypes random sub-samplings with replacement among all the
23 available genotypes. We sub-sampled 1000 times from 1 plant to the maximum number of plants for
24 each fields and populations. These samplings (S1 Figures D and E) enabled the computation of the
25 frequency of the dominant genotype (b), the mean number of genotypes in the sampling (c), the
26 number of different genotypes (d), the frequency of assignment to the dominant genotype when at

27 least 4 leaves belonged to this genotype (e) and at least 4 leaves belonged to this genotype (f). We
28 also represented the abundance of each genotype in the total sampling (a).

29

30 **Results and discussion**

31 **Genetic diversity in fields and populations (S1 Figures B and C)**

32 These representations showed the unique genotypes of individuals, by transect then by field
33 or population. The fields are more homogeneous than the feral populations. There are 56 unique
34 genotypes in the 256 fields' leaves, against 89 unique genotypes in the 156 feral populations' leaves.
35 Generally, there is one dominant genotype by field. For 2 fields (C151 and C355), the dominant
36 genotype is the same. This dominant genotype is present in all their transects. The other genotypes,
37 often single in abundance, are present on the border or inside the field. Many genotypes seem to
38 cohabit with the dominant genotype.

39 The transects along the feral populations give similar results. If there are several leaves with
40 the same genotype, this common genotype is not restricted to a single transect. Single genotypes are
41 distributed all along the transects. Apriori, a sub-sampling on this important number of non-dominant
42 genotypes would only result in a heterogeneous group of genotypes, varying at each sampling and
43 transect. The main difference with fields is that there is no dominant genotype as in fields: only few
44 genotypes that are shared by more than one leave.

45

46 In summary, there seems to be no sampling bias when sampling plants along the border of the
47 fields rather than inside them. As for feral population, the genetic diversity seems so important that it
48 seems impossible to obtain a representative sub-sampling.

49

50 **Effect of the sub-sampling on the fields (S1 Figure D).**

51 The frequency of the dominant genotype (b) depends from its abundancy in the sample (a) but
52 also from the abundancy of the other genotypes. Thus, the more different genotypes in the sampling,

53 the harder getting a representative sub-sampling (c, d). Even when as many leaves as existing in the
54 fields C704 and C6 were sampled, with replacement, their sub-samples did not represent their
55 complete genotypic diversity.

56 Thus, if a sampled field contains one main genotype, and if this genotype is as frequent as
57 main genotypes are in our sub-experiment, taking 10 leaves from this field and decide that a minima
58 4 genotypes should be equal to this main genotype is enough to correctly assign the genotype (and
59 thus, the cultivar) to the fields. For uniform fields, setting 4 identical genotypes seems correct. But
60 with the strict criteria of a minimum 6 genotypes, taking 10 leaves is not enough to obtain a correct
61 frequency (f). A sampling of 20 leaves at least should have been consider, and even more for the fields
62 C704 and C6.

63

64 **Effect of the sub-sampling on the feral populations (S1 Figure E)**

65 The feral populations contain more different genotypes and sometimes up to 2 main genotypes
66 (F222 and F86). The frequency of the dominant genotype (b) is relatively well evaluated, but very
67 low. Evermore, the sub-sampled genetic diversity did not represent the existing genetic diversity (c
68 and d) in any of the samplings. In this case, assigning a feral population to a genotype as fields does
69 not have any meaning with fields criteria (e and f) since there are almost as many genotypes as there
70 are individuals.

71 The high genotypic diversity is probably the consequence of the possible multiple origins of
72 feral populations. Feral populations may come from dispersal events of one unique cultivar each, and
73 so contains one or two main genotypes. But in the same time, feral plants could origin from
74 reproduction inside the feral population or from seed banks established years ago.

75

76 **Conclusions on the experimental protocol**

77 We had no prior knowledge when we chose the fields for the sub-experiment in 2007. We thus
78 did not know if each field could contain one cultivar, two cultivars or could be sown with farm saved

79 seeds. Each year, more than one hundred of fields are sown in Selommès area and hundreds of feral
80 populations grow on the side of the roads. Thus, it was important to determine a number of plants
81 giving enough genetic information to characterize fields and populations while staying an
82 economically acceptable cost, giving the high number of fields and feral population to survey.
83 Moreover, more than one border was genotyped for some fields while both leaves and seeds could
84 have been genotyped.

85 Furthermore, the biologic unit here was the genotype, and not the cultivar. The genotypes
86 present in these fields probably belong to the same cultivar, which would increase the probability of
87 assignment of the field to a given cultivar.

88 For fields, this number of 10 plants, even if less satisfying if the genetic diversity of the field
89 is high, is a good compromise enabling the analysis of a maximum of borders while having reliable
90 genetic information on the main genotype if the field is homogeneous. However, this compromise is
91 not satisfying enough when working on the genetic diversity of the feral populations. The only thing
92 well estimated is the main genotype frequency, and this frequency is well assessed with only 10
93 leaves. But this main genotype frequency is so low that we cannot assign a cultivar to feral population.
94 As the genetic diversity of our feral population is so high that even 50 leaves would not have been
95 enough to correctly assess it, we assume a much dense sampling design would probably result in
96 more genetic diversity.