

# Linkage disequilibrium in French wild cherry germplasm and worldwide sweet cherry germplasm

Uraiwan Arunyawat · Gaëlle Capdeville ·  
Véronique Decroocq · Stéphanie Mariette

Received: 29 September 2010 / Revised: 30 November 2011 / Accepted: 12 December 2011 / Published online: 8 January 2012  
© Springer-Verlag 2011

**Abstract** A basic knowledge on linkage disequilibrium (LD) is necessary in order to determine resolution of association studies. We investigated the extent and patterns of LD in a self-incompatible species (*Prunus avium* L.), in 3 groups (wild cherry, sweet cherry landraces and sweet cherry modern varieties), using a set of 35 microsatellite markers and the gametophytic self-incompatibility locus. Since population

structure might create spurious LD, we thus used the information provided by a structure analysis published in a previous study to perform the LD analysis. In the current study, we detected a greater LD extent in sweet cherry than in wild cherry, which is plausibly due to the bottleneck associated with domestication and breeding. Higher LD values in sweet cherry sub-groups may be explained by smaller sample sizes. We also showed that the remaining structure in the groups of sweet cherry, in particular landraces, is responsible for a part of the LD extent. Intra-group relatedness may also account for extensive LD in two sub-groups. These results demonstrate, if ever necessary, the importance of controlling the genetic structure and relatedness when estimating LD. Moreover, LD decays very rapidly with genetic linkage distance in both wild and sweet cherries, which seems promising for future association studies.

---

Communicated by S. González-Martínez

**Electronic supplementary material** The online version of this article (doi:10.1007/s11295-011-0460-9) contains supplementary material, which is available to authorized users.

---

U. Arunyawat · G. Capdeville · S. Mariette  
INRA, Unité de Recherche sur les Espèces Fruitières,  
71 Avenue Edouard Bourlaux, BP 81,  
33140 Villenave d’Ormon, France

U. Arunyawat · V. Decroocq  
INRA, UMR 1332 de Biologie du Fruit et de Pathologie,  
71 Avenue Edouard Bourlaux, BP 81,  
33140 Villenave d’Ormon, France

U. Arunyawat  
EGCB Research Unit, Department of Genetics, Faculty of Science,  
Kasetsart University,  
Bangkok 10900, Thailand

U. Arunyawat  
Center for Advanced Studies in Tropical Natural Resources,  
National Research University, Kasetsart University,  
Bangkok 10900, Thailand

*Present Address:*

S. Mariette (✉)  
INRA, UMR 1332 de Biologie du Fruit et de Pathologie,  
71 Avenue Edouard Bourlaux, BP 81,  
33140 Villenave d’Ormon, France  
e-mail: stephanie.mariette@bordeaux.inra.fr

**Keywords** Linkage disequilibrium · Population structure · Wild cherry · Sweet cherry · Association studies

## Introduction

A detailed understanding of the behaviour of linkage disequilibrium (LD) is important for association studies to choose appropriate experimental design and determine spacing of markers for the mapping of functional genes of interest (Flint-Garcia et al. 2003; Rafalski and Morgante 2004; Mueller 2004). The resolution of association mapping generally correlates with the pattern of LD extent. Indeed, fine-scale mapping is needed in the case of limited LD, whereas the extensive LD allows underlying traits to be detected more readily using a small number of markers (Mackay and Powell 2007). As stated in review papers from

Flint-Garcia et al. (2003), Gupta et al. (2005) and Rafalski and Morgante (2004), several factors influence LD in populations. Recombination is the unique factor that diminishes intrachromosomal LD, whereas other factors, i.e. mutation, migration, genetic drift and selection, all create LD. Moreover, the reproduction system can also influence levels of LD, autogamous species showing higher LD extent than allogamous species.

Several LD studies have been previously conducted in plants, both in natural and domesticated populations. In natural populations, the model plant *Arabidopsis thaliana* has been evaluated for LD levels and showed a decrease of LD within approximately 1 cM or 250 kb (Nordborg et al. 2002). In wild tomato populations, Arunyawat et al. (2007) showed that intragenic LD decays very rapidly with physical distance, suggesting high recombination rates and effective population sizes in the two investigated species. First investigations of forest tree natural populations have also been carried out. A number of recent studies in coniferous species (Brown et al. 2004; Kumar et al. 2004; Krutovsky and Neale 2005; Heuertz et al. 2006; Gonzalez-Martinez et al. 2006) and *Populus* (Ingvarsson 2005) described a rapid decay of LD, most probably explained by large population sizes (Neale and Savolainen 2004; Savolainen and Pyhäjärvi 2007). However, conifers would tend to exhibit lower recombination rates than angiosperms (Jaramillo-Correa et al. 2010).

In addition, the studies of LD in domesticated plants are numerous (Kraakman et al. 2004 for barley; Barnaud et al. 2006 for grapevine; Remington et al. 2001, Tenaillon et al. 2001, Jung et al. 2004, Stich et al. 2005 for maize; Garris et al. 2003 for rice; Hamblin et al. 2004 for sorghum; Zhu et al. 2003 for soybean; Jannoo et al. 1999 for sugarcane; Simko et al. 2006 for potato; Maccaferri et al. 2005 for wheat). Extensive LD is expected in these species because they have undergone bottleneck, domestication and modern breeding during their evolutionary history. A slow LD decay against distance is indeed observed in several species (10 cM for sugarcane, 100 kb for rice), but a rapid decay of LD (within 1 kb) was found in maize. These contrasted patterns of LD found in domesticated species may be explained by different reproduction systems (for instance, rice is a selfing species, whereas maize is an outcrossing species) and/or complex evolutionary histories (admixture). Different sampling strategies and important population structure may account for different LD estimations among studies. Similarly, studies that have investigated LD at the gene level or at the whole genome level and with different molecular tools have also revealed contrasted results. Remington et al. (2001) showed that the level of genome-wide LD found using SSR markers was much higher than that found with SNPs in maize. As argued by these authors, this may be due to a higher frequency of mutations in SSRs than in SNPs that arose during

the development of regional maize subpopulations after domestication.

Though the number of LD studies progressively increases in domesticated plants, several main issues still require further investigation. First, increasing the number of studied genera would allow inter-genera comparisons. Second, domestication and modern breeding effects need to be clarified, especially to take into account its impact before establishing association populations.

*Prunus avium* L. is a diploid fruit tree ( $2n=16$ ) that consists of wild cherry and its domesticated form (sweet cherry). Wild cherry is distributed throughout Europe and has been a fresh food supply for human consumption for thousands of years (Zohary and Hopf 2000). It is thought that human probably picked wild cherry in forests long before its cultivation. Several pieces of evidence reported that sweet cherry was present in Europe in Roman and early mediaeval times (Šoštarić and Küster 2001; Rösch 2008). The domestication history of sweet cherry is still unclear, though it is most likely that several events of domestication happened for sweet cherry or/and intense gene flow occurred, explaining the actual structure observed between wild and sweet cherries (Mariette et al. 2010). Morphologically, wild and sweet cherries are very similar in the fruit forms, the most obvious difference being probably the size of the fruits (Zohary and Hopf 2000). Conventional quantitative trait loci (QTL) studies based on linkage mapping have focused on the identification of functional genes of interest, e.g. genes responsible for fruit size in cherry (Clarke et al. 2009; Zhang et al. 2010). However, this linkage mapping approach has limitations such as being cost- and time-consuming and poor in QTL resolution. Association mapping may be a promising alternative approach to detect genes underlying traits in cherry, using germplasm structure and the extent of LD as fundamental information for designing the association application (Mackay and Powell 2007; Slatkin 2008). To our knowledge, information on LD in cherry, and more generally in fruit species and in their wild relatives, is still poorly investigated. It is, therefore, necessary to dissect the LD properties in cherry in order to facilitate appropriate association analysis in these species.

In this study, we employed population genetics analyses, using 35 microsatellite markers and the gametophytic self-incompatibility locus, to explore the pattern and magnitude of LD in wild and sweet cherries, all over the 8 linkage groups (LGs). Specifically, we focused on three main objectives: (1) to estimate the differences of LD extent between wild and sweet cherries, (2) to emphasise the effect of sample size, population structure and relatedness on LD estimates and (3) to evaluate its potential implication on future association studies based on LD patterns.

## Materials and methods

### Plant materials

A total of 212 French wild cherry individuals were collected at the INRA Orleans collection. Our sample was designed in order to cover most French regions, but did not represent the whole species diversity.

A total of 142 sweet cherry landraces (called landraces hereafter) and 66 sweet cherry modern varieties (called modern varieties hereafter) were obtained from 3 collection centres, including the INRA Bordeaux Prunus Genetic Resources Centre, the INRA Bordeaux sweet cherry breeding collection and the Interprofessional Technical Centre for Fruits and Legumes (CTIFL) collection. The landrace sampling was designed in order to represent European sweet cherry germplasm (e.g. France, Belgium, Czech Republic, Germany, Hungary, Iran, Italy, Romania and Spain); however, a majority of the samples originate from France. Some landraces are of unknown origin. The group of modern sweet cherry varieties selected for this study was defined in order to represent major breeding programmes worldwide (e.g. Australia, Canada, the Czech Republic, France, Italy and USA).

### Selection of marker loci and genotyping

We used 35 microsatellite markers, including 21 dinucleotide and 11 other nucleotide repeats and 3 markers for which the number of repeats is not known. Most of the genotypic data used in the current analysis were already available from our previous study (Mariette et al. 2010) and 10 additional SSR markers were further added while assessing LD. The gametophytic self-incompatibility locus, described by Sonneveld et al. (2006) and Vaughan et al. (2006), was included in this study. The alleles of the S-RNase and S-pollen (SFB) genes were amplified using primers developed by Sonneveld et al. (2006) and Vaughan et al. (2006). Allele sizes were revealed on a sequencing machine and allele numbering was deduced following Vaughan et al. (2006). Information obtained for this locus was statistically analysed for LD estimation, similarly to the microsatellite markers. The marker information is given in Table 1. Briefly, for genotyping, genomic DNA was isolated from fresh leaves using DNeasy® 96 plant kit from QIAGEN (Hilden, Germany). Multiplex PCR was performed with the QIAGEN Type It Microsatellite PCR Kit® (QIAGEN, Hilden, Germany). Marker genotypes were determined on the ABI 3730 sequencer, and the Genemapper software (Applied Biosystems, Foster City, CA, USA) was used to estimate allele sizes.

### Definition of samples

It is known that population structure may create spurious LD between unlinked markers (Flint-Garcia et al. 2003). In

a previous study (Mariette et al. 2010), the structure analysis (Pritchard et al. 2000) clearly separated wild and sweet cherry groups and further separated two sub-groups in sweet cherry modern varieties and three sub-groups in sweet cherry landraces, whereas wild cherry population was defined as an unstructured group. A group of landraces was admixed. Then, the LD analysis was conducted independently for wild cherry, sweet cherry, landraces, sweet cherry modern varieties and for each cherry population defined by structure (sub-groups 1, 2 and 3 for landraces and sub-groups 1 and 2 for sweet cherry modern varieties). Note that extremely admixed landrace individuals shown in Mariette et al. (2010) were not analysed as a sub-group of landraces in the current study. Sample sizes for each group and sub-group are given in Table 2.

### Linkage disequilibrium analyses

When using SSR data from diploid individuals to estimate LD, haplotype and gamete frequencies are generally not known. Consequently, we adopted two approaches to estimate LD in the different groups identified. First, we used two different methods without haplotype reconstruction. Second, after haplotype reconstruction, LD was estimated based on the most probable haplotypes. The distribution of allele frequencies may have an effect on the extent of LD; therefore, rare alleles (less than 5% allele frequency) were excluded prior to further LD analyses. The studied loci were sorted within each group following their respective order on each LG, as shown in Table 1. The list of analysed markers in each group is based on the available genotyping data; therefore, the number of markers may be slightly different from one group to another (for more details, see Table 2).

#### *Linkage disequilibrium analysis without haplotype reconstruction*

Estimates of LD directly based on gametic frequencies cannot be achieved easily in diploid natural populations because gamete frequencies are not known. To circumvent this problem, we selected two methods that were proposed in the literature.

First, considering the pairs of loci, we tested for the presence of association between genotypes at both loci using a *G* test calculated on contingency tables, as calculated by Genpop (Raymond and Rousset 1995; Rousset 2008).

Second, Cockerham and Weir (1977) proposed a composite LD to assess LD on diploid data for which gamete frequencies are not known and for which random mating cannot be assumed. A correlation coefficient is then derived and corresponding levels of significance can be calculated (Weir 1979). Correlations and levels of significance were assessed using the modification of the correlation as proposed by Garnier-Gere and Dillmann (1992) in the LinkDos program (this correlation will be called common correlation hereafter).

**Table 1** Marker information

Marker	LG	Map position (cM)	Repeat motif	References
EMPA005	1	32.7	(CT) <sub>3</sub> CAT(CT) <sub>12</sub> T(AC) <sub>23</sub>	Clarke and Tobutt (2003)
EMPA002	1	46.3	(AG) <sub>13</sub>	Clarke and Tobutt (2003)
UDP96005	1	48.5	(AC) <sub>16</sub> TG(CT) <sub>2</sub> CA(CT) <sub>11</sub>	Cipriani et al. (1999)
CPPCT029 <sup>+</sup>	1	65.1	(CT) <sub>24</sub>	Aranzana et al. (2002)
PMS67 <sup>a</sup>	1	95.1	NA	Cantini et al. (2001)
EMPA003	1	114.8	(AC) <sub>8</sub>	Clarke and Tobutt (2003)
BPPCT028 <sup>a</sup>	1	124.1	(TC) <sub>15</sub>	Dirlewanger et al. (2002)
EMPA017	2	0	(AG) <sub>19</sub>	Clarke and Tobutt (2003)
BPPCT002 <sup>a</sup>	2	20.9	(AG) <sub>25</sub>	Dirlewanger et al. (2002)
UDP98411	2	64.7	(TC) <sub>16</sub>	Testolin et al. (2000)
MA007A <sup>a</sup>	2	67.9	(TC) <sub>4</sub> C(CT) <sub>27</sub>	Yamamoto et al. (2002)
PCHGMS1	2	68.4	(AC) <sub>12</sub> (AT) <sub>6</sub>	Sosinski et al. (2000)
UCDCH11 <sup>a</sup>	2	70.5	(CT) <sub>15</sub>	Struss et al. (2003)
UCDCH12 <sup>a</sup>	2	74.3	(CA) <sub>14</sub>	Struss et al. (2003)
UCDCH21 <sup>a</sup>	2	83.6	(CA) <sub>18</sub>	Struss et al. (2003)
PCEGA34	2	86.8	NA	Downey and Iezzoni (2000)
EMPAS12	3	38.5	(TG) <sub>10</sub> a(GA) <sub>10</sub> aa(GA) <sub>13</sub>	Vaughan and Russell (2004)
EMPAS02	3	77.0	(TTG) <sub>7</sub> ctgc(TG) <sub>10</sub> (AG) <sub>8</sub>	Vaughan and Russell (2004)
EMPAS06	4	25.9	(CT) <sub>12</sub>	Vaughan and Russell (2004)
BPPCT040	4	36.7	(GA) <sub>14</sub>	Dirlewanger et al. (2002)
EMPAS10	4	50.9	(GA) <sub>28</sub>	Vaughan and Russell (2004)
PS12A02	4	83.1	NA	Joobeur et al. (2000)
EMPAS11	5	27.9	(TC) <sub>25</sub>	Vaughan and Russell (2004)
EMPAS14	5	45.9	(TC) <sub>10</sub> ccat(TC) <sub>5</sub> ccat(TC) <sub>8</sub>	Vaughan and Russell (2004)
EMPAS01	6	28.0	(GA) <sub>9</sub> (GA) <sub>11</sub>	Vaughan and Russell (2004)
UDP96001	6	30.5	(CA) <sub>17</sub>	Cipriani et al. (1999)
EMPA004	6	48.7	(GA) <sub>4</sub> AA(GA) <sub>4</sub> AA(GA) <sub>15</sub>	Clarke and Tobutt (2003)
UDP98021	6	72.4	(GA) <sub>22</sub> (CA) <sub>11</sub>	Testolin et al. (2000)
UDP98412	6	76.9	(AG) <sub>28</sub>	Testolin et al. (2000)
S-locus <sup>b</sup>	6	86.1	–	References therein
AMPA121 <sup>a</sup>	6	86.6	(GA) <sub>9</sub>	Hagen et al. (2004)
CPPCT021 <sup>a</sup>	6	91.9	(CT) <sub>10</sub>	Aranzana et al. (2002)
UCDCH14	7	54.8	(CT) <sub>18</sub>	Struss et al. (2003)
EMPA026	8	0	Complex (CT)	Clarke et al. (2009)
EMPA018	8	33.5	(GA) <sub>18</sub>	Clarke and Tobutt (2003)
UDP98409	8	49.8	(AG) <sub>19</sub>	Cipriani et al. (1999)

Map position derived from the inter-specific cross of *P. avium* (Napoleon) × *Prunus nipponica* (Clarke et al., 2009), except those rendered in italics that are from the cross of *Prunus dulcis* (Texas) × *Prunus persica* (Earlygold) (Dirlewanger et al. 2004)

<sup>a</sup>Markers added in the present study

<sup>b</sup>Gametophytic self-incompatibility locus

The analyses were performed using web versions of Genpop (<http://genepop.curtin.edu.au/>) and LinkDos (<http://genepop.curtin.edu.au/linkdos.html>).

#### Linkage disequilibrium analysis after haplotype reconstruction

We also used haplotype reconstruction in order to assess LD from haplotypic data. For this purpose, we reconstructed probable haplotypes within each LG using a Bayesian approach as implemented in the PHASE 2.1 version software (Stephens and Donnelly 2003). LD was

estimated by using squared allele frequency correlations ( $r^2$ ) and standardised disequilibrium coefficients ( $D'$ ). While  $r^2$  reflects both recombinational and mutational history,  $D'$  only reflects recombinational history.  $r^2$  depends on differences in allele frequencies at two sites and is generally favoured for potential allele–trait associations (Flint-Garcia et al. 2003). Estimation of LD ( $r^2$  and  $D'$  estimators) based on haplotypic data was performed using the TASSEL 2.1 version software (Bradbury et al. 2007). The significance level of LD ( $p$  value < 0.01) was assessed by 10,000 permutation tests as implemented in TASSEL version 2.1.



**Table 2** Number of markers and sample size for each cherry group and sub-group

Group/sub-group	Sample size	Number of markers
Wild cherry	212	29
Sweet cherry	208	30
Sweet cherry landraces	142	30
Sweet cherry modern varieties	66	30
Sweet cherry landraces sub-group 1	28	29
Sweet cherry landraces sub-group 2	34	35
Sweet cherry landraces sub-group 3	44	30
Sweet cherry modern varieties sub-group 1	43	35
Sweet cherry modern varieties sub-group 2	23	30

For genotypic and haplotypic LD estimations, heat maps of probabilities were then realised using the heat map function in the R statistical package (<http://www.r-project.org/>).

Consecutively, correlations among estimations were assessed both for all pairwise LD values and for mean LD values.

For comparing LD significance among groups, we used an experiment-wise first-type error rate of 5%. Because multiple comparisons were performed, we applied Bonferroni's correction.

#### Effect of the sample size and relatedness on LD in groups and sub-groups

##### *Effect of the sample size*

Sample size can affect LD estimations significantly (Teare et al. 2002). We thus bootstrapped sub-samples of the same data to assess the sample size effect on the extent of LD (Teare et al. 2002). Using the R software, we bootstrapped sets of 23, 43 and 66 samples in the sweet cherry modern varieties group and sets of 28, 34, 44 and 142 samples in the landraces group.  $D'$  and  $r^2$  were estimated over 1,000 replicates for each sample size using the LD pipeline from the TASSEL 2.1 software.

##### *Effect of relatedness*

The existence of relatedness within groups can result in increased LD and should thus be controlled in association studies (Slatkin 2008). We estimated relatedness within groups and sub-groups by using a pairwise kinship matrix as assessed by the TASSEL 2.1 software. Correlations were calculated between the mean kinship and mean LD levels, as measured by  $D'$  and  $r^2$  estimators.

#### Estimation of LD decay

In general, LD is negatively correlated with genetic distance. To examine the decline of LD, we computed mean LD values

at 10-cM intervals. We then tested for LD decay using a correlation between LD and pairwise distance ( $\log^{10}$  transformed; McRae et al. 2002). In addition, we plotted the LD estimates obtained with TASSEL against genetic distances (in centimorgans) between marker pairs. We performed the non-linear least squares (*nls*) function in the R statistical package (<http://www.r-project.org/>) to obtain the expected decay of LD by fitting the observed data in the following formula (Heifetz et al. 2005), which is modified from the study of Sved (1971). Assuming drift–recombination equilibrium:

$$LD = 1/(1 + 4bd) + e$$

where LD is the observed LD between marker pairs in relation to the genetic distance,  $d$  denotes the genetic distance between marker pair,  $b$  is the coefficient parameter of decay calculated using the *nls* function and  $e$  refers to the model residual obtained.

$r^2$  is generally preferred to quantify the extent of “useful” LD for association studies. The threshold suggested by Kruglyak (1999) was  $r^2=0.1$ . We chose to take a threshold of 0.5 for  $D'$  since the observed  $D'$  values are generally at least five times higher than  $r^2$  in our study (see the results part of Table 3).

## Results

#### LD estimation: with and without haplotype reconstruction and comparison of estimators

Before comparing LD estimators among biological groups, we compared estimations with and without haplotype reconstruction and we compared estimations with the three different estimators (common correlation on genotypic data,  $r^2$  and  $D'$  on haplotypic data).

LD estimated with the common correlation estimator based on unphased data was generally lower than LD estimated with  $D'$  based on haplotypic data (data not shown). Moreover, LD estimated with  $D'$  was higher than LD estimated with  $r^2$  (Tables 3 and 4).

Mean LD estimations were significantly correlated (Pearson's correlation between  $r^2$  and  $D'=0.908$ ,  $p$  value<0.001; Pearson's correlation between  $r^2$  and the common correlation=0.850,  $p$  value<0.001; Pearson's correlation between  $D'$  and the common correlation=0.863,  $p$  value<0.01). Estimations were also significantly correlated at the 0.001 level within each group and each sub-group, except for landrace sub-group 2, while comparing  $D'$  with the common correlation.

As for the significance of disequilibrium, trends were stable among the three methods, shown in Fig. 1a–i for the comparisons between  $p$  value obtained with haplotypic data (Tassel analysis) and  $p$  value obtained with genotypic data

**Table 3** LD decay with distance for wild and sweet cherry groups

Distance range	Wild cherry	Sweet cherry	Sweet cherry landraces	Sweet cherry modern varieties
0–10 cM	0.070	0.081	0.072	0.150
	0.459	0.639	0.583	0.772
	(57.14%)	(100%)	(85.71%)	(92.86%)
10–20 cM	0.012	0.034	0.040	0.104
	0.197	0.376	0.415	0.615
	(40%)	(92.31%)	(92.31%)	(76.92%)
20–30 cM	0.017	0.040	0.036	0.073
	0.215	0.333	0.320	0.505
	(50%)	(100%)	(71.43%)	(57.14%)
30–40 cM	0.013	0.033	0.048	0.037
	0.236	0.334	0.406	0.352
	(80%)	(100%)	(100%)	(80%)
40–50 cM	0.013	0.033	0.031	0.092
	0.191	0.332	0.347	0.526
	(50%)	(75%)	(75%)	(50%)
50–60 cM	0.013	0.021	0.045	0.031
	0.257	0.313	0.409	0.366
	(100%)	(75%)	(100%)	(75%)
>60 cM	0.014	0.015	0.020	0.033
	0.173	0.269	0.318	0.389
	(23%)	(58.33%)	(41.67%)	(16.67%)
Unlinked	0.004	0.009	0.012	0.018
	0.127	0.195	0.219	0.255
	(0.63%)	(42.27%)	(32.07%)	(5.54%)
Mean	0.007	0.014	0.017	0.028
	0.144	0.226	0.250	0.300
	(8.73%)	(49%)	(39.16%)	(14.53%)
Correlation	-0.454	-0.668	-0.584	-0.604
	( $p < 0.001$ )	( $p < 0.0001$ )	( $p < 0.0001$ )	( $p < 0.0001$ )
	-0.613	-0.744	-0.595	-0.588
	( $p < 0.0001$ )	( $p < 0.0001$ )	( $p < 0.0001$ )	( $p < 0.0001$ )

Means of  $r^2$  (first line of the cell) and  $D'$  (second line of the cell) against genetic linkage distance intervals and percentage of significant LD values for each interval (in parenthesis) are given for each group. The correlation between distance ( $\log^{10}$  transformed) and LD values ( $r^2$  and  $D'$ ) is given at the bottom of the table

(LinkDos analysis) and in Supplementary Fig. 1a–i for the comparisons between  $p$  value obtained with haplotypic data (Tassel analysis) and  $p$  value obtained with genotypic data ( $G$  test on contingency tables).

Since congruent results were obtained between the haplotypic and genotypic analyses, we decided to report our results obtained using the  $D'$  and  $r^2$  estimators based on reconstructed phase data.

#### Magnitude of LD: comparison among groups

##### *Effect of domestication: wild versus sweet cherry*

LD values estimated in sweet cherry were higher than LD values estimated in wild cherry, showing significant LD levels for 50% and 8.73% of all locus pairs in sweet cherry and in wild cherry, respectively (Table 3).

##### *Effect of breeding: sweet cherry landraces versus sweet cherry modern varieties*

Mean LD values against genetic distance range were high for both landraces and modern varieties (Table 4). In landraces, 39.16% of all marker pairs exhibited significant levels of LD estimate, while 14.53% of all marker pairs showed significant LD in modern varieties.

##### *Effect of sample size*

LD decreased with sample size for both bootstrapped samples and observed data (Fig. 2a, b), except for the modern varieties sub-group 1, for which a slight increase was observed (sample size=43; Fig. 2b).

The LD decline depending on sample size partly explained the difference in LD in between groups and sub-

**Table 4** LD decay with distance for landraces and modern sweet cherry varieties

Distance range	Landrace 1	Landrace 2	Landrace 3	Modern 1	Modern 2
0–10 cM	0.193	0.131	0.102	0.210	0.157
	0.724	0.631	0.565	0.756	0.697
	(100%)	(80%)	(56.25%)	(85%)	(71.43%)
10–20 cM	0.154	0.075	0.044	0.171	0.111
	0.691	0.491	0.403	0.662	0.622
	(58.33%)	(47.62%)	(26.67%)	(71.43%)	(38.46%)
20–30 cM	0.122	0.067	0.055	0.095	0.049
	0.531	0.388	0.362	0.525	0.481
	(57.14%)	(37.5%)	(42.86%)	(62.5%)	(14.29%)
30–40 cM	0.111	0.083	0.024	0.052	0.071
	0.559	0.559	0.283	0.433	0.519
	(75%)	(66.67%)	(0%)	(33.33%)	(20%)
40–50 cM	0.174	0.072	0.044	0.157	0.055
	0.618	0.487	0.342	0.556	0.652
	(50%)	(35.71%)	(18.75%)	(53.85%)	(0%)
50–60 cM	0.118	0.072	0.039	0.062	0.063
	0.526	0.499	0.337	0.456	0.694
	(66.67%)	(62.5%)	(0%)	(50%)	(0%)
>60 cM	0.125	0.056	0.023	0.048	0.031
	0.642	0.489	0.312	0.394	0.629
	(16.67%)	(16.67%)	(0%)	(16.67%)	(0%)
Unlinked	0.045	0.031	0.014	0.021	0.029
	0.359	0.366	0.236	0.270	0.362
	(4.67%)	(4.16%)	(0%)	(0.43%)	(0.29%)
Mean	0.061	0.040	0.021	0.039	0.038
	0.401	0.393	0.267	0.320	0.403
	(12.70%)	(12.12%)	(5.03%)	(9.80%)	(4.43%)
Correlation	-0.255	-0.383	-0.559	-0.457	-0.581
	( $p=0.056$ )	( $p<0.001$ )	( $p<0.0001$ )	( $p<0.0001$ )	( $p<0.0001$ )
	-0.204	-0.300	-0.556	-0.469	-0.132
	( $p=0.128$ )	( $p<0.01$ )	( $p<0.0001$ )	( $p<0.0001$ )	( $p=0.303$ )

Means of  $r^2$  (first line of the cell) and  $D'$  (second line of the cell) against genetic linkage distance intervals and percentage of significant LD values for each interval (in parenthesis) are given for each group. The correlation between distance ( $\log^{10}$  transformed) and LD values is given at the bottom of the table

groups. LD estimated with  $r^2$  increased more for bootstrapped samples than for observed data, except for sub-group 1 of landraces (sample size=28) and for modern varieties sub-group 1 (sample size=43), for which the contrary was observed (Fig. 2a, b; Table 5). Similar LD increases were observed between observed samples and bootstrapped samples for LD estimated with  $D'$  (Fig. 2a, b; Table 5).

In conclusion, for sub-group 1 of landraces (sample size=28) and for sub-group 1 of sweet cherry modern varieties (sample size=43), the reduction of sample size was not the only possible explanation for the observed LD increase.

#### Effect of population structure and relatedness

We found a significant impact of remaining population structure on LD estimates in sweet cherry groups. Significant LD between unlinked markers indicates the presence of

genetic structure. Thus, in this section, we analysed inter-chromosomal LD. Unlinked marker pairs showed a higher percentage of significant LD in sweet cherry, in landraces and in modern varieties than in wild cherry (Tables 3 and 4; compare Fig. 1a with b–d). When we analysed LD within sub-groups, limited significant LD was observed between unlinked markers in the three sub-groups of landraces as defined by structure analysis in comparison with the pooled data obtained for landraces (Tables 3 and 4). A similar pattern was also shown in the two sub-groups of modern varieties where the significance of LD between unlinked markers was lower in sub-groups than in the pooled modern varieties samples (Tables 3 and 4; compare Fig. 1h–i to d). Nevertheless, the impact of population structure on LD extent was smaller in modern varieties than in landraces, which may be explained by the less complex population structure existing in modern varieties. In particular, landraces

**Fig. 1** Comparison of LD  $p$  values obtained with the Tassel software on haplotypic data (*upper right of the matrix*) and with the LinkDos software on genotypic data (*lower left of the matrix*). *Black squares* represent a  $p$  value between 0 and 0.0001, *dark grey squares* represent a  $p$  value between 0.0001 and 0.001, *light grey squares* represent a  $p$  value between 0.001 and 0.01 and *white squares* represent a  $p$  value between 0.01 and 1. **a** Wild cherry, **b** sweet cherry, **c** sweet cherry landraces, **d** sweet cherry modern varieties, **e** sweet cherry landraces sub-group 1, **f** sweet cherry landraces sub-group 2, **g** sweet cherry landraces sub-group 3, **h** sweet cherry modern varieties sub-group 1, **i** sweet cherry modern varieties sub-group 2

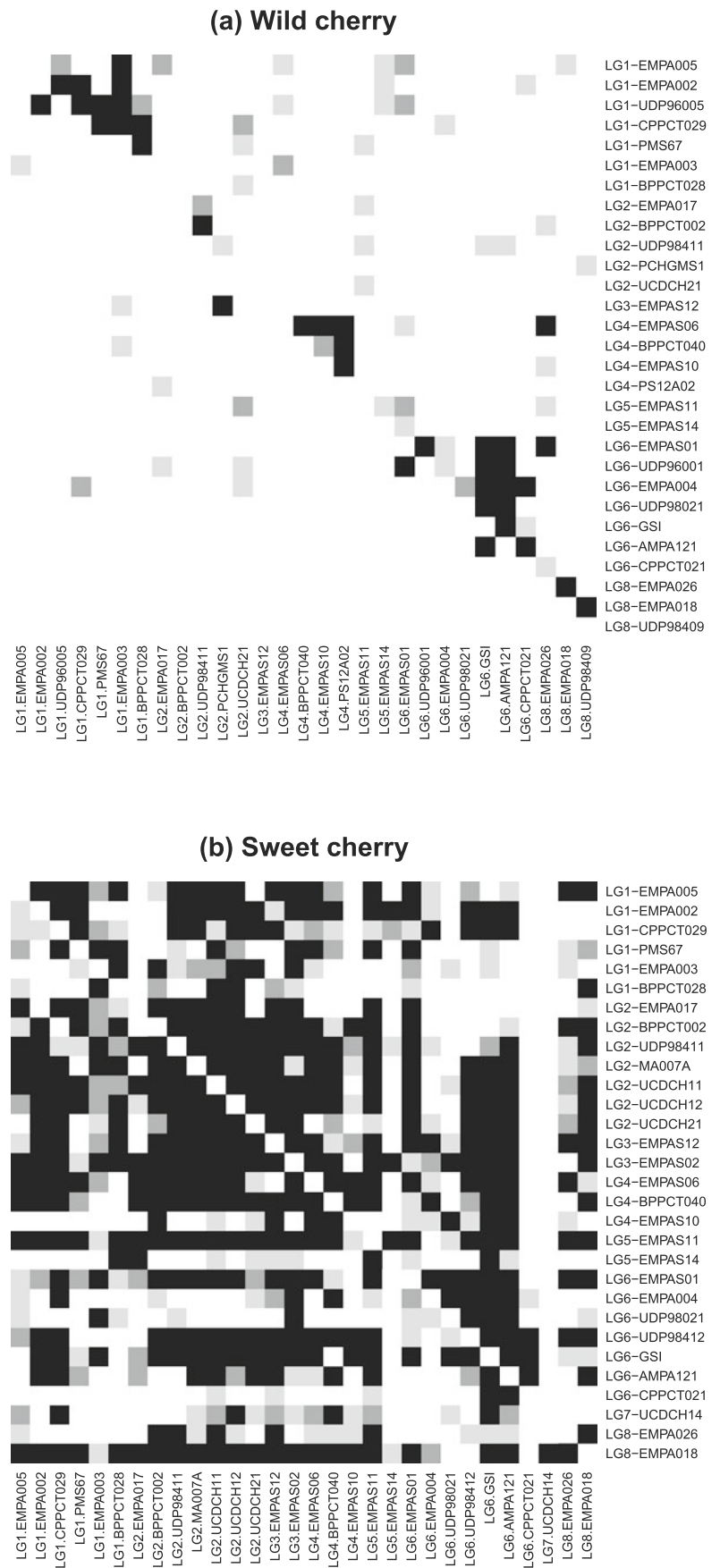
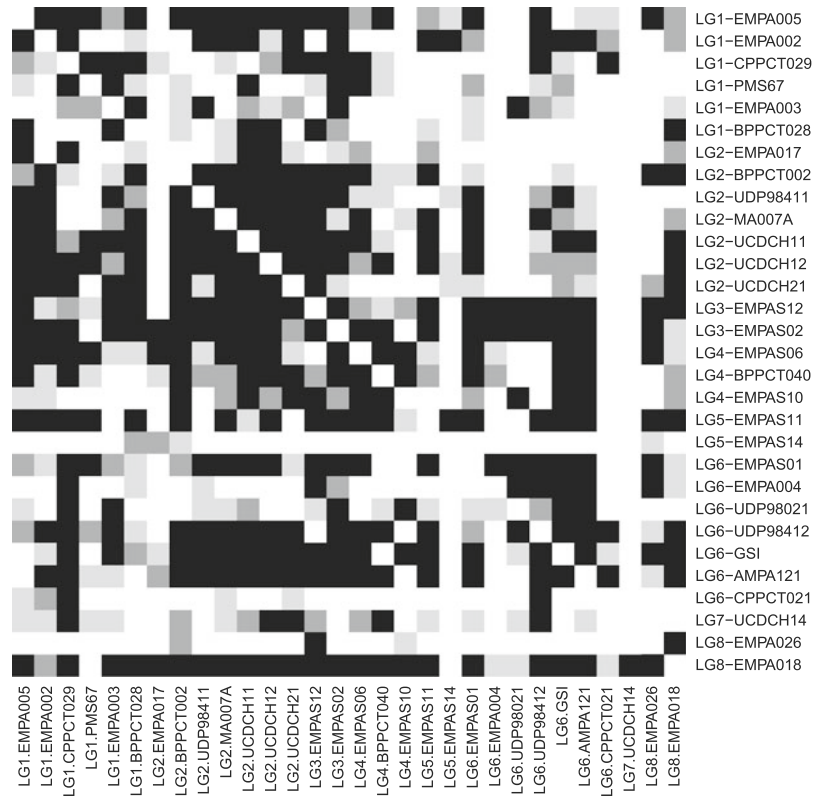




Fig. 1 (continued)

(c) Landraces



(d) Modern varieties

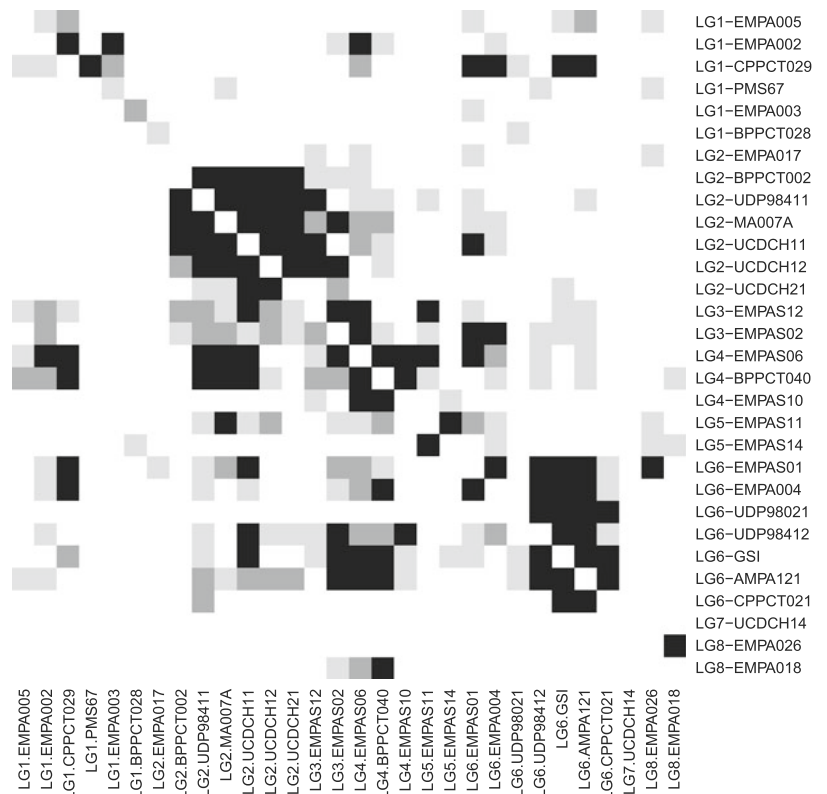


Fig. 1 (continued)

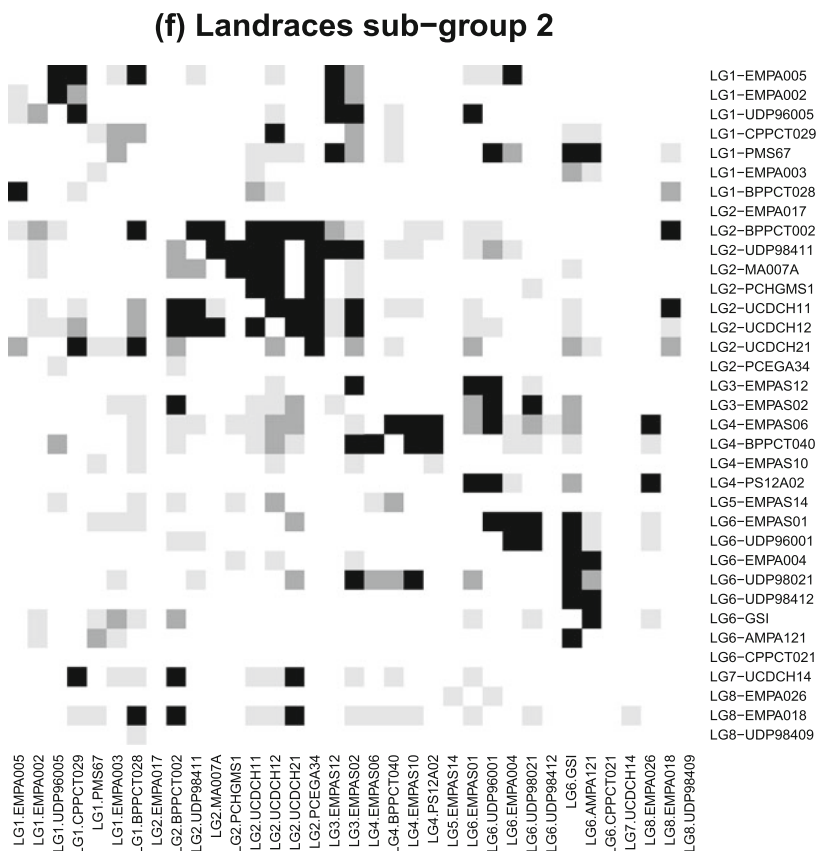
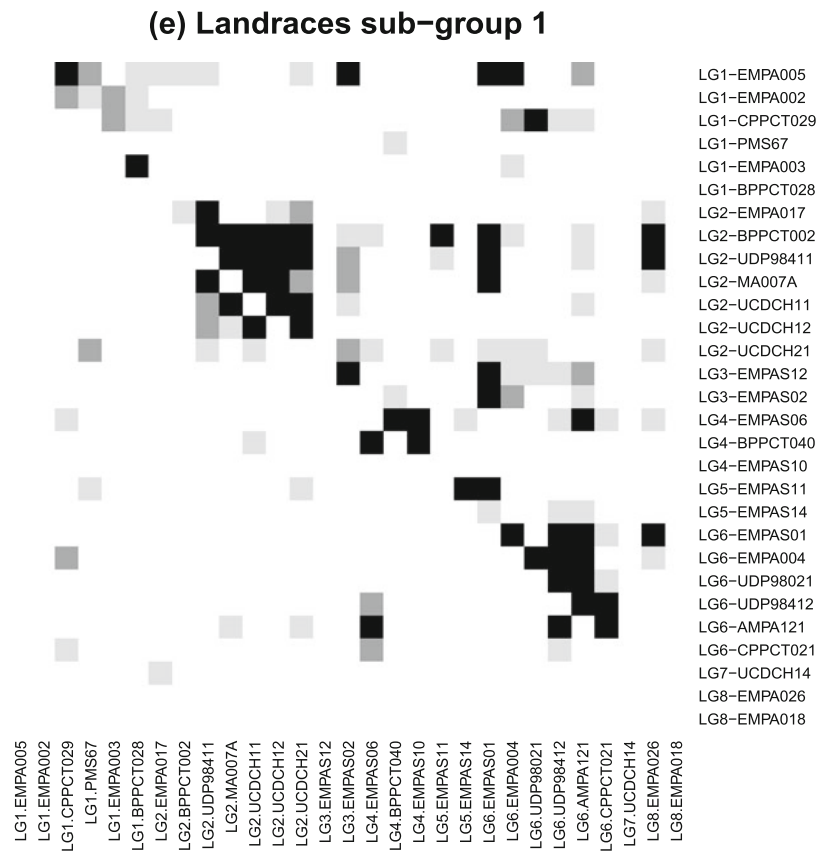
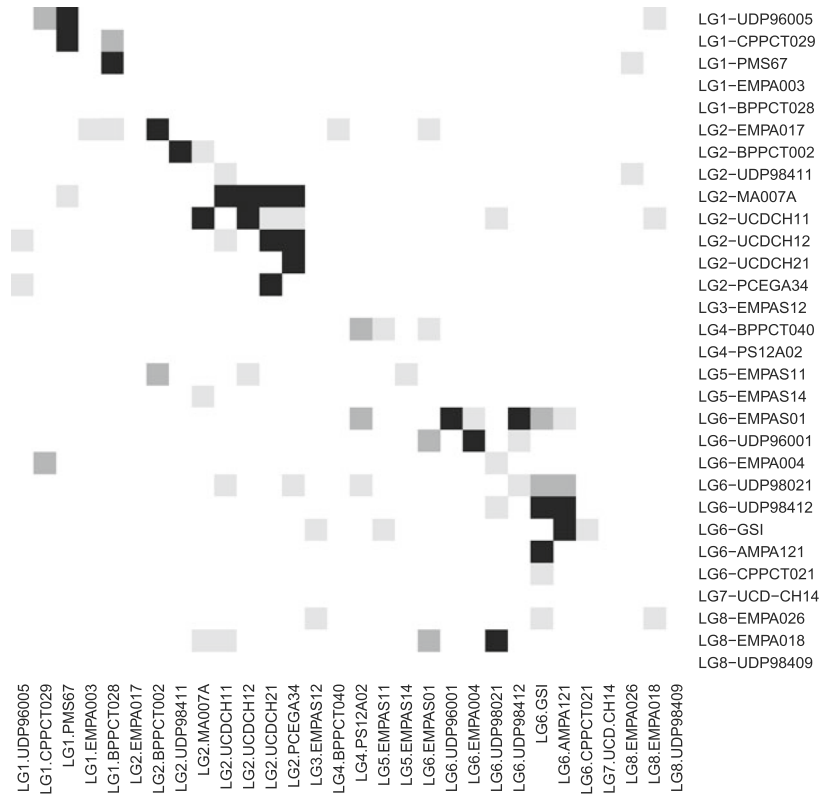


Fig. 1 (continued)

(g) Landraces sub-group 3



(h) Modern varieties sub-group 1

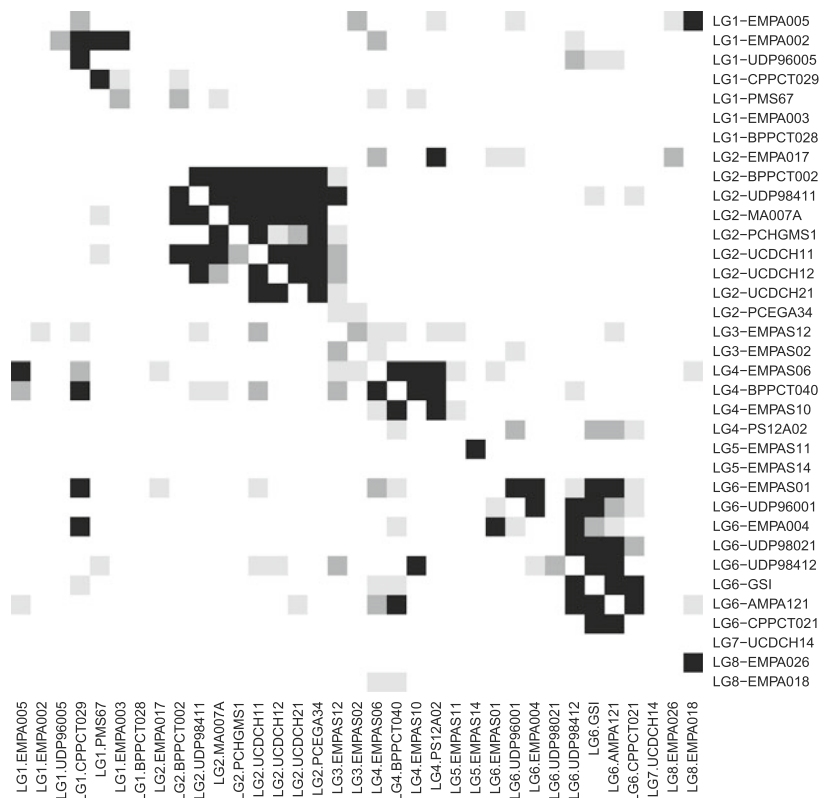
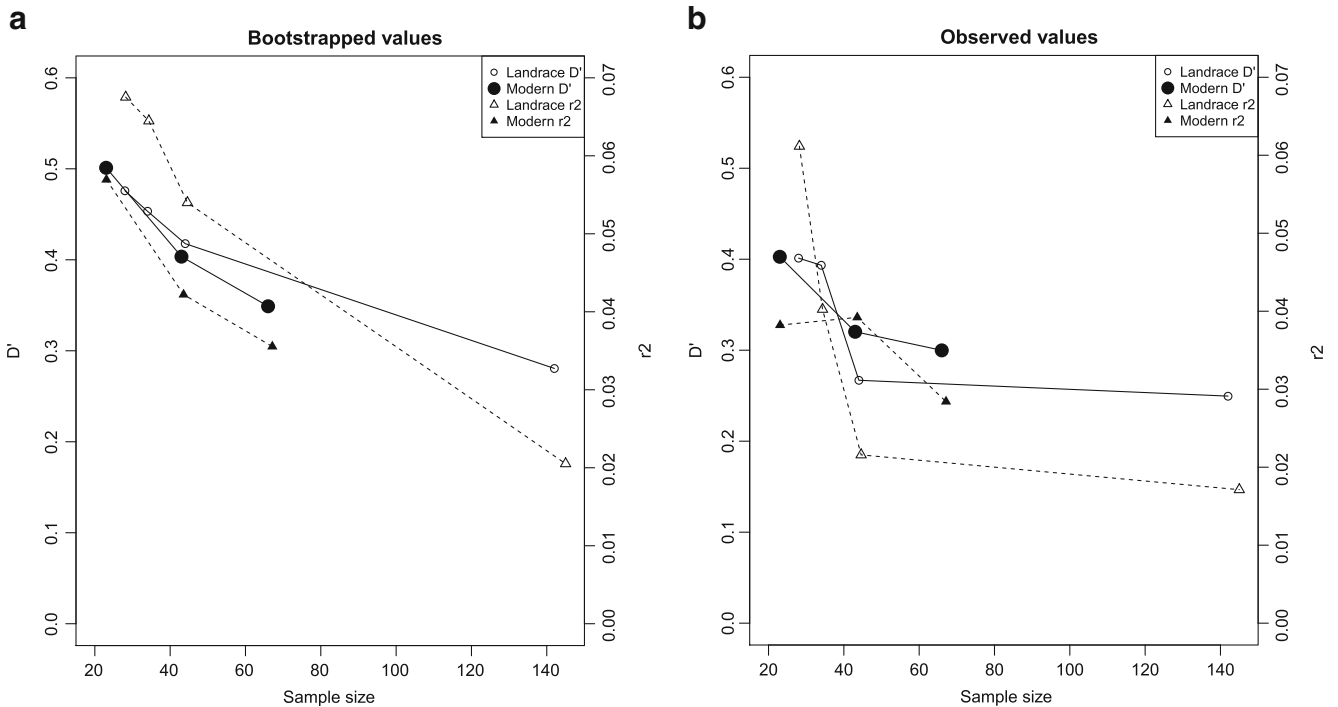
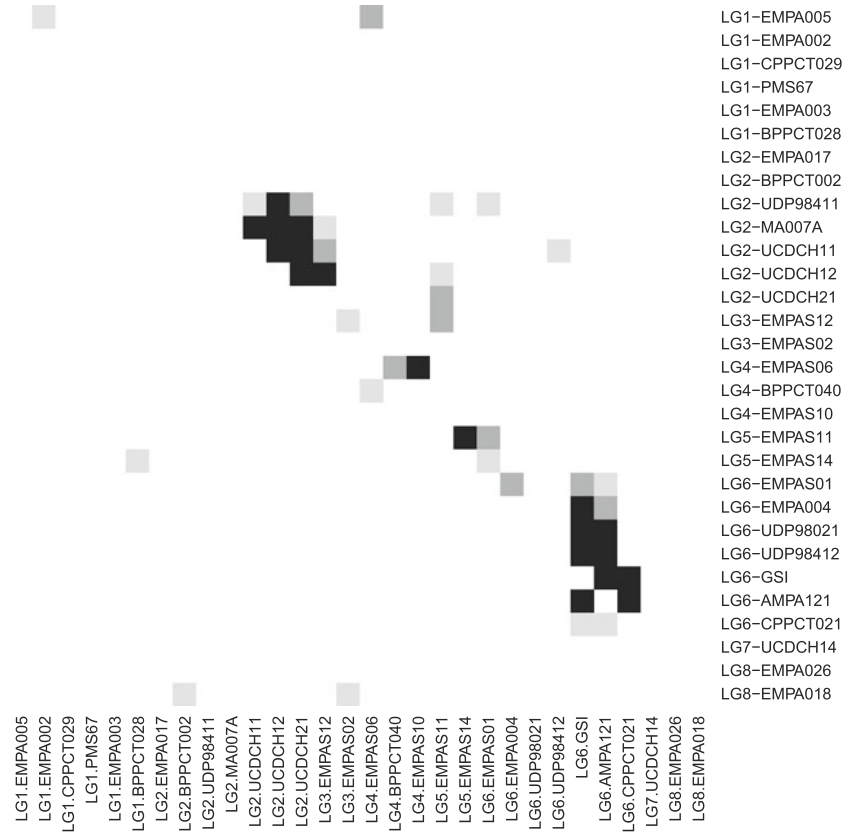


Fig. 1 (continued)

(i) Modern varieties sub-group 2



**Fig. 2 a** Effect of sample size on the mean of  $r^2$  and  $D'$  for all locus pairs, obtained by re-sampling. Points represent sample sizes of 28, 34, 44 and 142 for landraces and sample sizes of 23, 43 and 66 for modern varieties.

**b** Effect of sample size on the observed mean of  $r^2$  and  $D'$  estimated with all locus pairs. Points represent sample sizes of 28, 34, 44 and 142 for landraces and sample sizes of 23, 43 and 66 for modern varieties

**Table 5** Bootstrapped and observed relative  $r^2$  and  $D'$  values in landraces and modern varieties sub-groups

Sub-group/ group	Observed ( $r^2$ )	Bootstrapped ( $r^2$ )	Observed ( $D'$ )	Bootstrapped ( $D'$ )
44–142	1.26	2.63	1.07	1.49
34–142	2.35	3.14	1.58	1.62
28–142	3.57	3.29	1.61	1.70
43–66	1.38	1.18	1.07	1.16
23–66	1.34	1.60	1.34	1.44

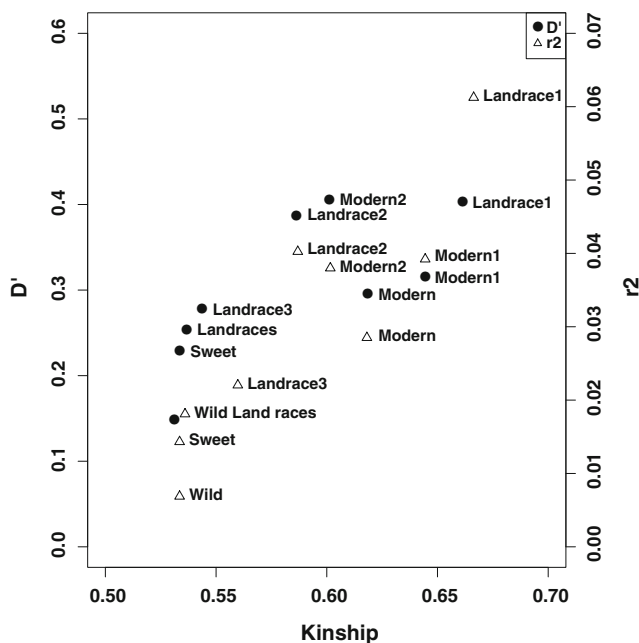
For each sample size, the LD value measured in the smallest sample size was divided by the LD value measured in the total group

showed a higher level of admixed individuals than modern varieties.

A significant correlation was observed between relatedness within groups of samples and mean LD level, both for  $r^2$  and  $D'$  (Fig. 3; Pearson’s correlation=0.901,  $p$  value<0.001 for  $r^2$  and Pearson’s correlation=0.740,  $p$  value<0.05 for  $D'$ ). The lowest kinships and the lowest mean LDs were observed for wild cherry, sweet cherries and landraces. The highest kinships and the maximum mean LDs were observed for landraces sub-group 1 and for modern varieties sub-group 1.

Pattern of LD along linkage groups

We evaluated the pattern of LD along the LGs by considering the significance of LD blocks as presented in Fig. 1a–i and Supplementary Fig. 1a–i. Significant LD estimates were scattered all over the loci. A comparison among markers at



**Fig. 3** Effect of kinship estimated using Tassel software on the observed mean of  $r^2$  and  $D'$  estimated with all locus pairs

less than 10 cM was possible for three LGs (LG1, LG2 and LG6). Interestingly, among all cherry groups compared in this study, we observed significant LD blocks over several loci on LG2 both in groups of landraces and in groups of modern varieties. However, they were absent in wild cherry. In comparison, the significant LD blocks on LG6 (particularly between the GSI locus and the AMPA121 SSR marker) appeared to be common in all cherry groups, which may be explained by the effect of natural selection at the gametophytic self-incompatibility locus on LD extent.

The decay of LD

As expected, strong LD was observed in nearby closely spaced markers, whereas lower levels of LD were found between more distant markers (Tables 3 and 4). For instance, in wild cherry, the estimation of  $r^2$  decreased from 0.070 to 0.014 for markers at short (up to 10 cM) to long (>60 cM) range of genetic distance. LD decay with distance ( $\log^{10}$  transformed) was significant for all groups, except for the landraces sub-group 1 and for the  $D'$  estimate for the modern varieties sub-group 2 (Tables 3 and 4).

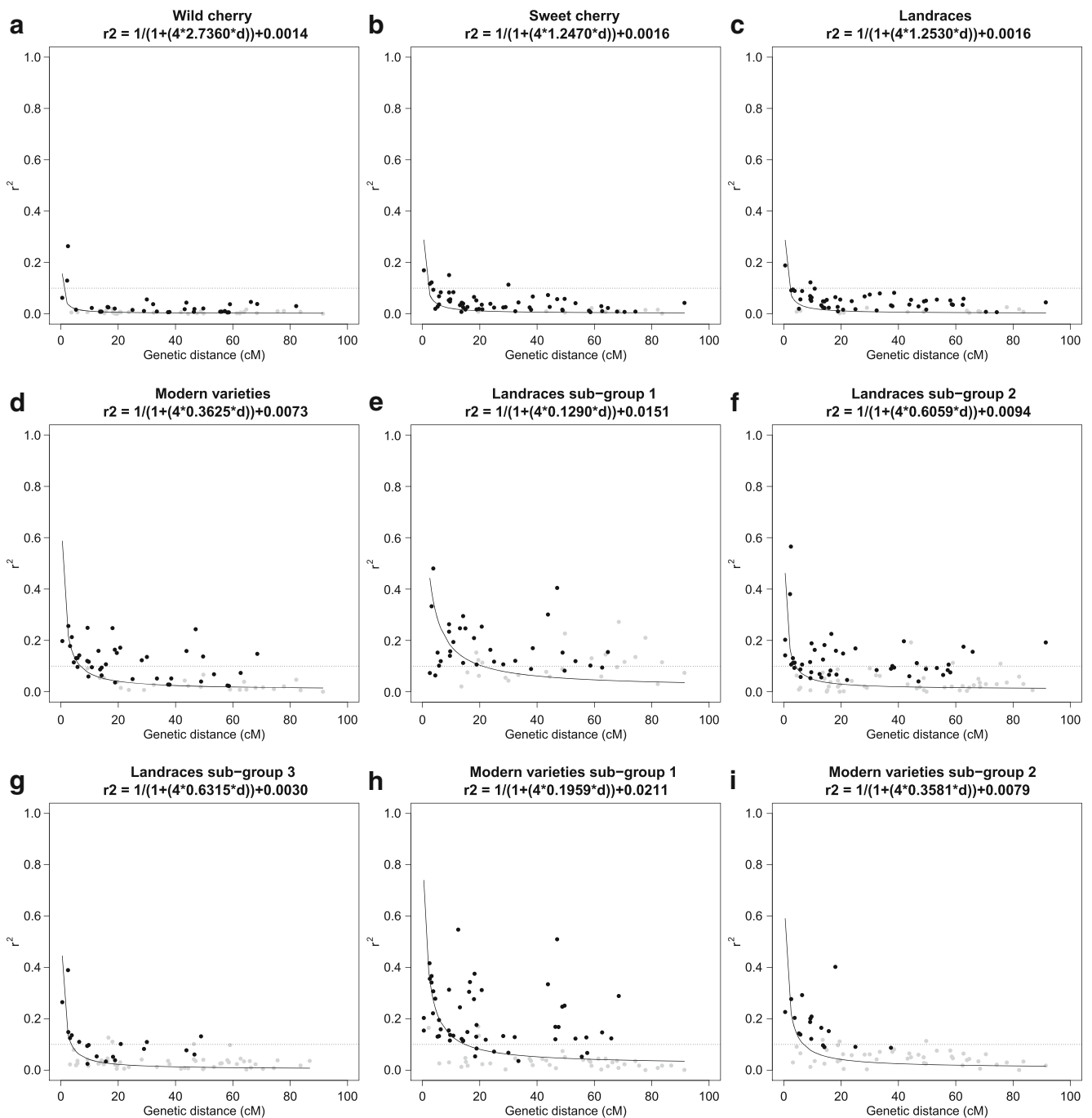
Predicted values of  $r^2$  declined much more rapidly than predicted values of  $D'$  (Fig. 4a–i and Supplementary Fig. 2a–i for  $r^2$  and  $D'$ , respectively). Predicted values of  $r^2$  generally decayed rapidly to less than 0.1 within a range inferior to 10 cM (Fig. 4a–i). There was, however, variation among groups, since the predicted LD decay was less rapid in sweet cherry groups, especially in modern varieties (Fig. 4d), in the landraces sub-group 1 (Fig. 4e) and in the modern varieties sub-group 1 (Fig. 4h).

Predicted values of  $D'$  declined more rapidly in wild cherry than in all groups and sub-groups of sweet cherry (Supplementary Fig. 2a–i). The extent of the predicted  $D'$  was particularly important in modern varieties (Supplementary Fig. 2d) and in sub-groups (Supplementary Fig. 2e–i), except in the landraces sub-group 3 (Supplementary Fig. 2g).

Discussion

For cultivated species, it is generally observed that domestication and breeding have strong impact on the level of genetic diversity in populations. This statement was supported by our recent study, showing that breeding was responsible for a non-negligible bottleneck observed with all SSR markers in the sweet cherry modern varieties (Mariette et al. 2010). In addition, domestication and breeding are expected to influence the level of LD within a population. In the present study, we assessed the level of LD using SSRs markers and the self-incompatibility locus in wild and sweet cherries (*P. avium* L.).





**Fig. 4** Plots of  $r^2$  against genetic distance between markers pairs. Curves show non-linear fit of  $r^2$  on distance, following the equation given in the “Materials and methods” section. Estimated parameters of the model are provided on each figure. The dotted line indicates the threshold at  $r^2=0.1$ . **a** Wild cherry, **b** sweet cherry, **c** sweet cherry

landraces, **d** sweet cherry modern varieties, **e** sweet cherry landraces sub-group 1, **f** sweet cherry landraces sub-group 2, **g** sweet cherry landraces sub-group 3, **h** sweet cherry modern varieties sub-group 1, **i** sweet cherry modern varieties sub-group 2

#### Comparison of LD estimators

LD estimated with unphased data and with phased data was correlated among groups, and trends for LD significance were similar. We decided on presenting our results obtained using the  $r^2$  and  $D'$  estimators based on reconstructed phase data.

The comparison between these two estimators revealed that  $D'$  was always higher than  $r^2$ . This difference has already been reported, for example, in maize (Remington et al. 2001), in chicken (Heifetz et al. 2005) and in sheep (Meadows et al. 2008). Rare alleles and unobserved haplotypes are expected to inflate  $D'$  but not  $r^2$  (Meadows et al.

2008). Despite this difference between  $D'$  and  $r^2$ , the two estimators gave congruent conclusions when we analysed the patterns of LD among cherry groups. Either one or the other statistic could be thus used in the current study to depict the patterns.

However, the predicted decay of  $r^2$  against genetic linkage distance was more rapid than the predicted decay of  $D'$ . Though, again, trends were quite congruent between the two estimators (for example, a more rapid decay in wild cherry than in sweet cherry groups), this discrepancy leads to different conclusions concerning the absolute range of LD decay (around 5–10 cM in wild cherry with  $r^2$  and around 60 cM with  $D'$ ). Since  $D'$  extends over larger linkage distances, it reduces the power to identify true association. In the following sections of the “Discussion,” we thus preferred to use  $r^2$  for concluding about LD decay and its consequences for association mapping in cherries.

What affects the patterns of LD among cherry groups?

The interplay of several factors influences the patterns of observed LD within species (Flint-Garcia et al. 2003; Rafalski and Morgante 2004). Generally, LD decays more rapidly in outcrossing compared to selfing species because recombination may be less effective in selfing species (Nordborg and Tavaré 2002). Not surprisingly, we detected a rapid decline of LD in our cherry populations, especially in wild cherry, which belong to a strictly self-incompatible species. Nevertheless, the decline of LD in cherry seems to be relatively slow compared to previous studies based on sequence data in self-incompatible species. For example, LD decays to negligible levels within 150–750 bp in wild tomatoes (Arunyawat et al. 2007) and within 0.2–1.5 kbp in maize (Remington et al. 2001; Tenaillon et al. 2001). Besides, the extent of LD we observed in self-incompatible sweet cherry is comparable to the extent of LD recently published in peach that is self-compatible, 13–15 cM (Aranzana et al. 2010). However, the decay of LD with distance in this study is comparable to LD decay in SSR-based studies performed in both wild and cultivated grapevine germplasm (Barnaud et al. 2010). On the whole, direct comparisons of LD extent with other studies should be questioned because of differences in LD measurement parameters and/or molecular markers used in each study. More interestingly, the comparison between wild and domesticated material provide useful information for future association studies.

Although the decay of LD was comparable among groups, important LD differences were detected between wild and sweet cherries. Domestication and breeding are most likely one of the main factors that leads to a higher significant LD detected in sweet cherry than in its wild relative. Population bottleneck is likely a consequence of domestication that may explain the LD extent in sweet

cherry, especially for modern varieties for which a bottleneck was indeed demonstrated (Mariette et al. 2010). A significant effect of domestication on the LD extent has been demonstrated in several crop plants due to bottlenecks and genetic drift. For example, LD is more extensive in elite barley cultivars and most likely due to domestication and breeding (Caldwell et al. 2006). LD extended 12 times further in cultivated grape compared to a wild population (Barnaud et al. 2010). Cultivated sunflower exhibited somewhat greater LD extent than in wild sunflower (Liu and Burke 2006), and a similar trend was found between cultivated and wild rice (Mather et al. 2007). Pattern of diversity and LD in modern sunflower cultivars is another example that has been shaped by domestication and breeding bottlenecks (Kolkman et al. 2007). Interestingly, while estimating the LD in *Arabidopsis lyrata*, authors reported that the population bottleneck is a cause of LD extent in the region linked to self-incompatibility (Kamau et al. 2007).

Nevertheless, our conclusion about the impact of domestication and breeding should be tempered by the fact that the wild sample is of French origin only, whereas landraces and modern varieties represent larger geographical origins. If we had included material originating from the putative geographical basin of domestication (Caucasia) in this study, wild cherry may show population structure that may result in higher levels of LD (see therein our results on the impact of structure).

Using small sample sizes is expected to bias LD estimations (Teare et al. 2002). In this study, LD estimators were overestimated in small sample size groups (sub-groups of landraces and modern varieties). Following a bootstrapping approach, we showed that sample size alone may account for differences of LD among groups and sub-groups. The slower decrease of LD observed in sub-groups than in the pooled data for both landraces and modern varieties may be due, in part, to the small sample size that may reduce statistical power. We, therefore, cannot ignore the effect of small sample size that may influence the significance level of LD in the sub-groups data. Small sample sizes in our study could also explain why we observed similar extents of LD as in the self-compatible peach sub-groups as stated before (Aranzana et al. 2010), since the study in peach was performed in larger samples (224 individuals in the whole sample and 39, 91 and 94 individuals in each sub-group).

Finally, as fully detailed in the two following paragraphs, population structure, relatedness and selection also influence LD estimations in cherries.

Do population structure and relatedness play a role on LD patterns of cherries?

Sweet cherries, particularly landraces, exhibit a complex population structure, whereas wild cherry appear to be an unstructured population, at least at the studied geographical

level (Mariette et al. 2010). The presence of population structure in sweet cherry is plausibly one of the factors to explain the difference in LD significance between wild and sweet cherries, especially for unlinked markers. Additionally, the LD significance was obviously higher in sweet cherry, which is a pool of landraces and modern varieties, in comparison with the LD in unstructured wild cherry population. Finally, clearly marked LD significance was detected in landraces and in modern varieties, in comparison with the LD in sub-groups. We thus postulate that complex population structure have a significant impact on the significance of LD in our cherry material. Besides, intra-group relatedness could also explain the differences we noticed. In particular, for two groups (landrace sub-group 1 and modern sub-group 1), higher relatedness may account for higher LD values. It is worth noting that less rapid LD decay was also found in these sub-groups.

In other species, there is considerable evidence that population structure has influence on the magnitude and pattern of LD. For instance, the study of population structure at a large scale in European *Arabidopsis* accessions showed the existence of population structure among groups and in consequence was suggested to be responsible for the level of LD detected (Ostrowski et al. 2006). Furthermore, previous studies reported that complex population structure may shape patterns of genetic variation and also influence levels of LD in plant species regardless of wild or cultivated varieties, such as in maize, *Populus*, and barley (Remington et al. 2001; Ingvarsson 2005; Comadran et al. 2009). Finally, population structure and relatedness may also yield spurious associations in association studies (Pritchard et al. 2000; Helgason et al. 2005).

Consequently, the general observed pattern of LD in sweet cherry populations may be explained by genetic structure and intra-group relatedness.

#### Insights into cherry association studies: potential applications on LD mapping

The differences between LD extents in wild and sweet cherries are of interest for LD mapping application, in particular on LG2 where a significant LD extent was observed in sweet cherry but not in wild cherry. One explanation is the effect of selection on particular genes maintaining the excess of LD in that LG. It is worth noting that a recent QTL study reported that genes responsible for fruit size are located, in “Emperor Francis” cherry linkage map, on LG2 (Zhang et al. 2010). In addition, a QTL linked to fruit size was previously detected on LG2 in other *Prunus* species, e. g. in sour cherry (Wang et al. 2000) and peach (Quilot et al. 2005). One practical application would be the selection of big fruits in *Prunus* species.

Similarly, in our study, the gametophytic self-incompatible locus (S-locus) located on LG6 seems to be responsible for the occurrence of a significant LD on this LG. Indeed, several authors reported a similar selection scenario at the S-locus, for example, in *Arabidopsis* and *Brassicaceae* species (Takebayashi et al. 2003; Kamau and Charlesworth 2005; Edh et al. 2009).

The rate of LD decline influences the resolution of association analysis and inversely the density of DNA markers needed for functional variation mapping (Nordborg et al. 2002; Rafalski and Morgante 2004; Mackay and Powell 2007; Slatkin 2008). High genotypic and phenotypic variability, as well as the rapid decay of LD in our unstructured wild cherry population, provides a promising insight for future association studies with the purpose of fine-scale genotype–phenotype variation association. A high-density marker genotyping may be required for the mapping of genes of interest in wild cherry. Conversely, a significant extent of LD in some sub-groups of sweet cherry, useful for association analysis, may require fewer markers for QTL mapping. However, the complex population structure evidenced in sweet cherry, especially in landraces, might cause difficulties when using LD to map phenotypic variation. The extent of LD in modern varieties, together with a less pronounced effect of population structure, may make association mapping possible in modern varieties with fewer markers. Nevertheless, the number of markers we assayed is relatively low, and our study did not cover all the LGs. Our results need further investigation with more markers and/or sequence data.

#### Conclusion

This study provides the first detailed understanding of the magnitude and patterns of LD in cherry. Overall, the significance and extent of LD is greater in sweet cherry than in wild cherry, which is most likely explained by the presence of population structure in concert with the impact of bottleneck associated with domestication and breeding programme in sweet cherry. Our data also demonstrated the role of selection, either natural or artificial, on the occurrence of LD blocks creation. Finally, LD decays rapidly with increasing genetic linkage distance in the analysed cherries, particularly wild cherry, which seems promising for future association studies aiming at mapping phenotypic variation in cherry.

**Acknowledgements** We thank François Hubert, Claire Thiéry, Sandra Cortijo and Mohamed-Amine Belka for the assistance in acquiring genetic data. We are grateful to Emilie Balsemin, Hélène Christmann and Jacques Joly for the technical help while sampling at the INRA sweet cherry collections (Genetic Resources Centre and breeding collections). We are thankful to Frédérique Santi and Jean Dufour at the INRA wild cherry collection (breeding collection) for the sampling assistance. We are

grateful to Franck Salin for his help in acquiring data on the genotyping platform (IFR103). We warmly thank Santiago González-Martínez and three anonymous reviewers for their useful comments that helped us in improving our manuscript. This work was funded by INRA Plant Genetics and Breeding Department (*Prunus*Div genotyping and *Prunus*LD projects) to S.M., and U.A. was recipient of a CCRDT # 22000270 postdoctorate fellowship from the regional council of Aquitaine.

**Conflict of interest** The authors declare no conflict of interest.

## References

- Aranzana MJ, Garcia-Mas J, Carbo J, Arus P (2002) Development and variability analysis of microsatellite markers in peach. *Plant Breed* 121:87–92
- Aranzana MJ, El-Kadri A, Howad W, Arus P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genet* 11:69
- Arunyawat U, Stephan W, Stadler T (2007) Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. *Mol Biol Evol* 24:2310–2322
- Barnaud A, Lacombe T, Doligez A (2006) Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. *Theor Appl Genet* 112:708–716
- Barnaud A, Laucou V, This P, Lacombe T, Doligez A (2010) Linkage disequilibrium in wild French grapevine, *Vitis vinifera* L. subsp. *silvestris*. *Heredity* 104:431–437
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc Natl Acad Sci USA* 101:15255–15260
- Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172:557–567
- Cantini C, Iezzoni AF, Lamboy WF, Boritzki M, Struss D (2001) DNA fingerprinting of tetraploid cherry germplasm using simple sequence repeats. *J Am Soc Hortic Sci* 126:205–209
- Cipriani G, Lot G, Huang WG, Marrazzo MT, Peterlunger E, Testolin R (1999) AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L.) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. *Theor Appl Genet* 99:65–72
- Clarke JB, Tobutt KR (2003) Development and characterization of polymorphic microsatellites from *Prunus avium* ‘Napoleon’. *Mol Ecol Notes* 3:578–580
- Clarke JB, Sargent DJ, Boskovic RI, Belaj A, Tobutt KR (2009) A cherry map from the inter-specific cross *Prunus avium* ‘Napoleon’ x *P. nipponica* based on microsatellite, gene-specific and isoenzyme markers. *Tree Genet Genomes* 5:41–51
- Cockerham CC, Weir BS (1977) Digenic descent measures for finite populations. *Genet Res* 30:121–147
- Comadran J, Thomas WTB, van Eeuwijk FA, Ceccarelli S, Grandi S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, Romagosa I, Hackett CA, Russell JR (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. *Theor Appl Genet* 119:175–187
- Dirlwanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, Arus P, Laigret F (2002) Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor Appl Genet* 105:127–138
- Dirlwanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arus P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc Natl Acad Sci USA* 101:9891–9896
- Downey SL, Iezzoni AF (2000) Polymorphic DNA markers in black cherry (*Prunus serotina*) are identified using sequences from sweet cherry, peach, and sour cherry. *J Am Soc Hortic Sci* 125:76–80
- Edh K, Widen B, Ceplitis A (2009) Molecular population genetics of the SRK and SCR self-incompatibility genes in the wild plant species *Brassica cretica* (Brassicaceae). *Genetics* 181:985–995
- Flint-Garcia SA, Thornsberry JM, Buckler ES IV (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374
- Garnier-Gere P, Dillmann C (1992) A computer program for testing pairwise linkage disequilibria in subdivided populations. *J Hered* 83:239
- Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). *Genetics* 165:759–769
- Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172:1915–1926
- Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol Biol* 57:461–485
- Hagen LS, Chaib J, Fady B, Decroocq V, Bouchet JP, Lambert P, Audergon JM (2004) Genomic and cDNA microsatellites from apricot (*Prunus armeniaca* L.). *Mol Ecol Notes* 4:742–745
- Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH, Kresovich S (2004) Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. *Genetics* 167:471–483
- Heifetz EM, Fulton JE, O’Sullivan N, Zhao H, Dekkers JCM, Soller M (2005) Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* 171:1173–1181
- Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K (2005) An Icelandic example of the impact of population structure on association studies. *Nat Genet* 37:90–95
- Heuertz M, De Paoli E, Kallman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174:2095–2105
- Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). *Genetics* 169:945–953
- Jannoo N, Grivet L, Dookun A, D’hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. *Theor Appl Genet* 99:1053–1060
- Jaramillo-Correa JP, Verdú M, González-Martínez SC (2010) The contribution of recombination to heterozygosity differs among plant evolutionary lineages and life forms. *BMC Evol Biol* 10:22
- Joobeur T, Periam N, Vicente MC, King GJ, Arus P (2000) Development of a second generation linkage map for almond using RAPD and SSR markers. *Genome* 43:649–655
- Jung M, Ching A, Bhatramakki D, Dolan M, Tingey S, Morgante M, Rafalski A (2004) Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. *Theor Appl Genet* 109:681–689
- Kamau E, Charlesworth D (2005) Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. *Curr Biol* 15:1773–1778



- Kamau E, Charlesworth B, Charlesworth D (2007) Linkage disequilibrium and recombination rate estimates in the self-incompatibility region of *Arabidopsis lyrata*. *Genetics* 176:2357–2369
- Kolkman JM, Berry ST, Leon AJ, Slabaugh MB, Tang S, Gao W, Shintani DK, Burke JM, Knapp SJ (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics* 177:457–468
- Kraakman ATW, Niks RE, Van den Berg P, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168:435–446
- Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22:139–144
- Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics* 171:2029–2041
- Kumar S, Echt C, Wilcox PL, Richardson TE (2004) Testing for linkage disequilibrium in the New Zealand radiata pine breeding population. *Theor Appl Genet* 108:292–298
- Liu AZ, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 173:321–330
- Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range LD in a durum wheat elite collection. *Mol Breed* 15:271–289
- Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci* 12:57–63
- Mariette S, Tavaud M, Arunyawat U, Capdeville G, Millan M, Salin F (2010) Population structure and genetic bottleneck in sweet cherry estimated with SSRs and the gametophytic self-incompatibility locus. *BMC Genet* 11:77
- Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177:2223–2232
- McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J (2002) Linkage disequilibrium in domestic sheep. *Genetics* 160:1113–1122
- Meadows JRS, Chan EKF, Kijas JW (2008) Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genet* 9:61
- Mueller JC (2004) Linkage disequilibrium for different scales and applications. *Brief Bioinform* 5:355–364
- Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends Plant Sci* 9:325–330
- Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. *Trends Genet* 18:83–90
- Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* 30:190–193
- Ostrowski MF, David J, Santoni S, Mckhann H, Reboud X, Le Corre V, Camilleri C, Brunel D, Bouchez D, Faure B, Bataillon T (2006) Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. *Mol Ecol* 15:1507–1517
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Quilot B, Kervella J, Genard M, Lescourret F (2005) Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. *J Exp Bot* 56:3083–3092
- Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet* 20:103–111
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86:248–249
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES IV (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA* 98:11479–11484
- Rösch M (2008) New aspects of agriculture and diet of the early medieval period in central Europe: waterlogged plant material from sites in south-western Germany. *Veg Hist Archaeobot* 17:225–238
- Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol Ecol Resour* 8:103–106
- Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. *Curr Opin Plant Biol* 10:162–167
- Simko I, Haynes KG, Jones RW (2006) Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics* 173:2237–2245
- Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9:477–485
- Sonneveld T, Robbins TP, Tobutt KR (2006) Improved discrimination of self-incompatibility S-RNase alleles in cherry and high throughput genotyping by automated sizing of first intron polymerase chain reaction products. *Plant Breeding* 125:305–307
- Sosinski B, Gannavarapu M, Hager LD, Beck LE, King GJ, Ryder CD, Rajapakse S, Baird WV, Ballard RE, Abbott AG (2000) Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. *Theor Appl Genet* 101:421–428
- Šoštarić R, Küster H (2001) Roman plant remains from Veli Brijuni (island of Brioni), Croatia. *Veg Hist Archaeobot* 10:227–233
- Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169
- Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor Appl Genet* 111:723–730
- Struss D, Ahmad R, Southwick SM, Boritzki M (2003) Analysis of sweet cherry (*Prunus avium* L.) cultivars using SSR and AFLP markers. *J Am Soc Hortic Sci* 128:904–909
- Sved J (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol* 2:125–141
- Takebayashi N, Brewer PB, Newbigin E, Uyenoyama MK (2003) Patterns of variation within self-incompatibility loci. *Mol Biol Evol* 20:1778–1794
- Teare MD, Dunning AM, Durocher F, Rennart G, Easton DF (2002) Sampling distribution of summary linkage disequilibrium measures. *Ann Hum Genet* 66:223–233
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA* 98:9161–9166
- Testolin R, Marrazzo T, Cipriani G, Quarta R, Verde I, Dettori MT, Pancaldi M, Sansavini S (2000) Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* 43:512–520
- Vaughan SP, Russell K (2004) Characterization of novel microsatellites and development of multiplex PCR for large-scale population studies in wild cherry, *Prunus avium*. *Mol Ecol Notes* 4:429–431
- Vaughan SP, Russell K, Sargent DJ, Tobutt KR (2006) Isolation of S-locus F-box alleles in *Prunus avium* and their application in a novel method to determine self-incompatibility genotype. *Theor Appl Genet* 112:856–866
- Wang D, Karle R, Iezzoni AF (2000) QTL analysis of flower and fruit traits in sour cherry. *Theor Appl Genet* 100:535–544
- Weir BS (1979) Inferences about linkage disequilibrium. *Biometrics* 35:235–254



- Yamamoto T, Mochida K, Imai T, Shi YZ, Ogiwara I, Hayashi T (2002) Microsatellite markers in peach [*Prunus persica* (L.) Batsch] derived from an enriched genomic and cDNA libraries. *Mol Ecol Notes* 2:298–301
- Zhang GR, Sebolt AM, Sooriyapathirana SS, Wang D, Bink MCAM, Olmstead JW, Iezzoni AF (2010) Fruit size QTL analysis of an F1 population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. *Tree Genet Genomes* 6:25–36
- Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. *Genetics* 163:1123–1134
- Zohary D, Hopf M (2000) Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley, 3rd edn. Oxford University Press, New York