Theor Appl Genet (2006) 112: 708–716
DOI 10.1007/s00122-005-0174-1

ORIGINAL PAPER

Adeline Barnaud · Thierry Lacombe · Agnès Doligez

Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L

Received: 5 April 2005 / Accepted: 30 November 2005 / Published online: 10 January 2006

Abstract We present here the first study of linkage disequilibrium (LD) in cultivated grapevine, *Vitis vinifera* L. *subsp. vinifera (sativa)*, an outcrossing highly heterozygous perennial species. Our goal was to characterize the amount and pattern of LD at the scale of a few centiMorgans (cM) between 38 microsatellite loci located on five linkage groups, in order to assess its origin and potential applications. We used a core collection of 141 cultivars representing the diversity of the cultivated compartment. LD was evaluated with both independence tests and multilocus r^2, both on raw genotypic and reconstructed haplotypic data. Significant genotypic LD was found only within linkage groups, extending up to 16.8 cM. It appeared not to be influenced by the weak structure of the sample and seemed to be mainly of haplotypic origin. Significant haplotypic LD was found over 30 cM. Both genotypic and haplotypic r^2 values declined to around 0.1 within 5–10 cM, suggesting a rather narrow genetic base of the cultivated compartment and limited recombination since domestication events. These first results open up a few application opportunities for association mapping of QTLs and marker assisted selection.

Introduction

The amount, extent and distribution of linkage disequilibrium (LD) has been studied in humans (Kruglyak 1999; Jorde 2000; Mohlke et al. 2001), animals (Farnir et al. 2000; MacRae et al. 2002; Tenesa et al. 2003; Nsengimana et al. 2004), and plants (Jannoo et al. 1999; Remington et al. 2001; Tenaillon et al. 2001; Nordborg et al. 2002; Garris et al. 2003; Zhu et al. 2003; Brown et al. 2004; Hamblin et al. 2004; Jung et al. 2004; Kraakman et al. 2004; Kumar et al. 2004; Simko 2004; Maccaferri et al. 2005; Stich et al. 2005). It is a topic of major interest for designing association-mapping experiments (Kruglyak 1999) and potentially for inferring species history (Farnir et al. 2000; Jorde 2000; Remington et al. 2001). The extent of true (due to physical linkage) and spurious (due to demographic history) LD depends on demographic history (reproductive system, bottlenecks, migration, population admixture), genomic history (recombination, mutation) and selection. For domesticated plants, it can be assumed that LD will be important since the evolutive forces which can generate LD (genetic drift, small effective population size, selection, admixture) are common along the history of domestication and breeding. However, contrasted results were found among species and/or studies. While some authors found extensive LD, as for example over 10 cM in sugarcane (Jannoo et al. 1999) or durum wheat (Maccaferri et al. 2005), on the contrary some studies in maize revealed a rapid decline of LD within 100–200 bp (Tenaillon et al. 2001; Remington et al. 2001).

Vitis vinifera L. is an outcrossing heterozygote perennial species, with high diversity (Sefc et al. 2000; Aradhya et al. 2003). In this species, we do not a priori expect a small or large extent of LD, because several evolutionary processes with opposite effects on LD, occurring before, during and after domestication, shaped present grape diversity. The progenitors of cultivated grapevine are presumed to be dioecious, and thus to exhibit a low level of LD because of obligatory outcrossing. Domestication involved selection for hermaphroditic flowers, which is expected to increase the selfing rate and thus LD. It probably also induced LD increase through associated bottlenecks. The vegetative

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-005-0174-1 and is accessible for authorized users.

Communicated by A. Charcosset

A. Barnaud · T. Lacombe · A. Doligez (✉)
INRA, UMR DGPC Equipe Génétique Vigne, 2 place Viala,
34060 Montpellier Cedex 1, France
E-mail: doligez@ensam.inra.fr
Tel.: +33-4-99612503
Fax: +33-4-99612064

propagation of interesting genotypes should have maintained existing LD through decreased recombination, but also increased the role of mutation in disrupting LD. Extensive human-driven migration probably induced episodes of recombination leading to decreased LD.

Our aim in this study was to perform the first genome-wide study of LD pattern in cultivated *V. vinifera* L., in order to know more about grapevine history and evaluate opportunities and conditions for LD mapping of QTLs. We assessed LD in a core collection of 141 cultivars representing the diversity of the largest cultivated grapevine collection worldwide. We worked with 38 SSR loci on five linkage groups. Most studies of LD in animals or plants have been performed on haplotypic data. Grapevine being heterozygous, with most parentages still unknown, we did not have direct access to haplotypes. Therefore we studied LD using first the genotypic data with unknown phase and then reconstructed haplotypic data inferred under the assumption of coalescence. The potential origin and applications of observed LD are discussed.

Material and methods

Plant material

In order to study LD in a sample of cultivars as unrelated as possible, we designed a core collection containing the maximum possible genetic diversity with the minimum repetitiveness, from a large germplasm collection of grapevine. This collection is maintained at the INRA experimental station of Vassal (Hérault, France), and preserves 2,300 identified cultivars of *V. vinifera* L. Five hundred and twenty-nine cultivars were discarded at this stage, because no sufficient agro-morphological data were available for them or they presented less interest for the purpose of this study (mutants or clones). We used 50 agro-morphological traits selected out of 167 (list available from the authors upon request), for being the most discriminant ones with the largest amount of data available. These traits were all measured at the experimental station of Vassal. We used the M Strategy (Schoen and Brown 1993) extended to qualitative and quantitative variables, implemented in MSTRAT (Gouesnard et al. 2001), which allows to maximize diversity (as measured by Nei index, Nei 1978) for any given sample size. Correspondence analysis (CA) performed with GENETIX 4.04 software (Belkhir et al. 2002) did not show any structure and no strong correlation was found (data not shown). The size of the core collection was chosen near the *X*-value of the inflexion point of the plot of maximized total Nei diversity against core collection size. Once the size of the core collection was set, we performed 50 independent sampling runs and chose the most often drawn individuals to be included in the final sample.

Genotyping

Thirty-eight simple sequence repeat (SSR) markers with known map position (Fig. 1) were used in this study. The mapping population was a full-sib progeny of 139 individuals described in Doligez et al. (2002). The map consisted of 432 AFLP and SSR loci spanning 1,322 cM Kosambi on 19 linkage groups, with an average intermarker distance of 3.2 cM (unpublished results). The 38 SSR markers used to study LD were chosen in the five most densely covered genome regions, so as to be separated by distances of a few cM at most.

All the individuals of the core collection were genotyped as described in Adam-Blondon et al. (2004) with slight modifications. We used one additional fluorochrom (PET); the PCR reaction volume was 20 μl; 1.6 μM of labelled and 6.4 μM of unlabelled primers were used. Electrophoresis was carried out on an ABI PRISM 3100 Genetic Analyser (Applied Biosystems). We performed double reading. Null allele frequencies were assumed to be negligible.

Allele frequencies, observed and expected heterozygosity (Nei 1978) were estimated using GENETIX. For all subsequent analyses, rare alleles (with frequencies $<5\%$ in the total sample) were replaced by missing data to avoid biased estimations of LD.

Structure

Since population structure tends to create spurious LD between unlinked markers (Nei and Li 1973; Pritchard and Przeworski 2001) and our aim was to study LD patterns not due to structure, we performed three complementary analyses on genotypic data to assess the structure of our sample before proceeding with LD analysis. First, we performed a Principal Component Analysis (PCA) with Statistica 7.1 (StatSoft France 2005) to provide a synthetic representation of sample diversity. This PCA was based on the covariance of the *G* matrix of allelic doses g_{ij}, where $g_{ij} = -1, 0, 1$ if the genotype of individual *i* contains 0, 1, 2 times allele *j*, respectively. The distribution of cultivars along the first axes was visually inspected to detect any grouping within the core collection.

Second, cultivated grapevine germplasm can be divided in three a priori groups according to their use: table (T), wine (W), wine or table (WT) (Aradhya et al. 2003). To test for the absence of differentiation between these three groups, we used Wright's *F*st index (Weir and Cockerham 1984). The empirical distribution of *F*st under the null hypothesis of no differentiation was obtained with GENETIX using 10,000 permutations.

Third, we used the Bayesian model-based clustering method of Pritchard et al. (2000), implemented in the Structure 2.1 software (http://www.pritch.bsd.uchicago.edu). We used the basic model without admixture and with uncorrelated allele frequencies, with the assumed number of populations (*K*) varying from 1 to 10, five replicate runs per *K* value, a burnin period length of

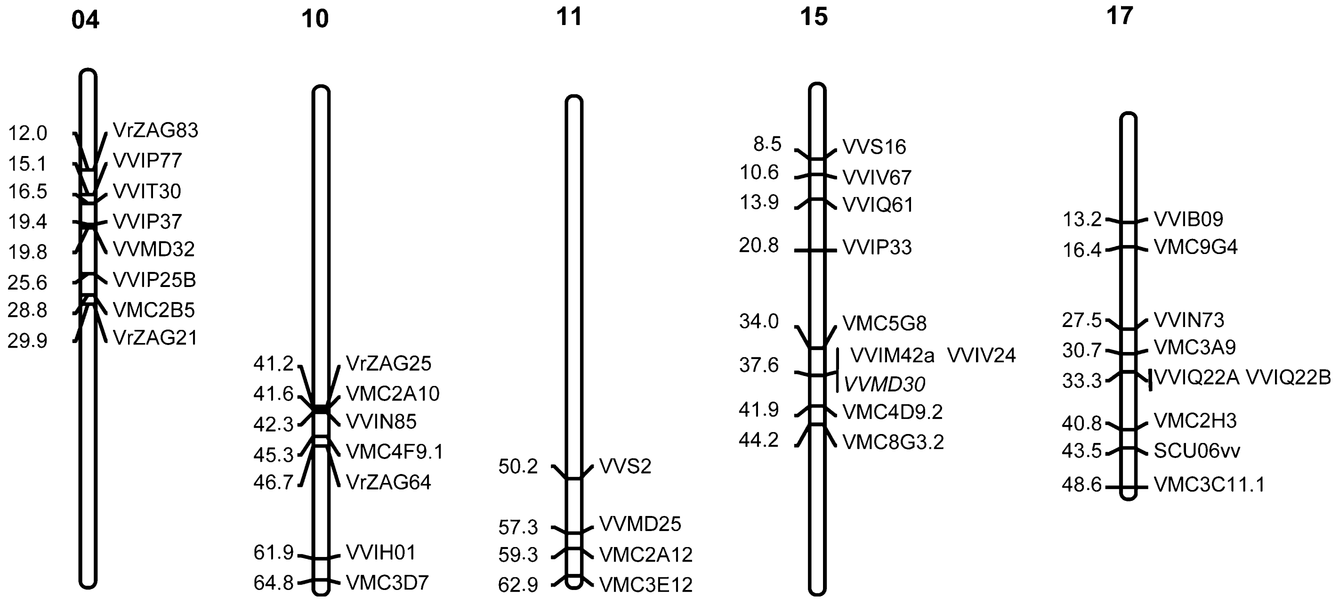


Fig. 1 Map of the SSRs used in this study for the F1 cross described in Doligez et al. (2002). Distances are in cM and linkage groups are numbered according to Riaz et al. (2004)

10^6, and a post-burnin simulation length of 1.5×10^6. This model assumes that each genotype in the sample comes purely from one of the unknown number K of differentiated ancestral populations. For each linkage group, only the locus with the fewest missing data was used, so five unlinked loci were included in this analysis.

Linkage disequilibrium

As *V. vinifera* L. is a heterozygous species, it is not possible to distinguish between the two possible double heterozygotes AB/ab and Ab/aB when parentage is unknown. Therefore in the present study, we measured and tested LD using two complementary data sets, first the raw unphased genotypic data and then haplotypic data reconstructed based on the coalescent theory. Direct analyses of unphased genotypic data presented the advantage not to rely on any assumption about the genetic history or the genotype frequencies of the sample. However, they are expected to induce some loss in power (Pritchard and Przeworski 2001) compared to analyses of haplotypic data. Therefore we also analysed LD on reconstructed haplotypes, even though this reconstruction is based on the assumption of coalescence.

For unphased genotypic data, we tested the null hypothesis of independence between all locus pairs using the Fisher's exact test implemented in GDA 1.1 software (Lewis and Zaykin 2002), which is based on shuffling of genotypic data, and yields estimates of the exact significance levels based on genotypic contingency tables. We used 10,000 permutations, without breaking genotypes to prevent any disequilibrium within loci (Hardy–Weinberg) to affect the significance of disequilibrium between loci.

For unphased genotypic data, we also estimated the composite disequilibrium coefficient defined by Weir (1996) as $\Delta_{AB}=p_{AB} + p_{A/B} - 2p_Ap_B$ (with $-0.5<\Delta_{AB}<0.5$), where p_{AB} and $p_{A/B}$ are the two-locus haplotypic (A and B in coupling) and non-haplotypic (A and B in repulsion) frequencies, respectively, and p_A and p_B are the allele frequencies at loci A and B, respectively. We performed this bi-allelic estimation for all pairs of linked alleles using GDA. Then we normalized this composite measure of LD by calculating the bi-allelic r^2_{AB} correlation coefficient defined by Weir (1996) for each pair of linked alleles, assuming higher order disequilibria could be neglected. This composite correlation is the correlation of the matrix G of allelic doses defined above. Finally, we obtained a multiallelic r^2 correlation coefficient for each pair of linked loci by summing the r^2_{AB} values weighted by p_Ap_B.

We inferred haplotypic data within each linkage group using a Bayesian method for reconstructing haplotypes from population genotype data (Stephens et al. 2001; Stephens and Donnelly 2003), implemented in the software PHASE version 2.1. We relaxed the assumption of stepwise mutation. Ten independent reconstructions were performed and the one with the best overall goodness-of-fit of the estimated haplotypes to the underlying approximate coalescent model was selected for the following analyses. An exact test of independence between all pairs of linked loci, based on haplotypic contingency tables, was performed with PowerMarker V3.23 (K. Liu and S. Muse, http://www.powermarker.net) using 50,000 permutations. The multiallelic r^2 correlation coefficient was also estimated with PowerMarker for all pairs of linked loci. This coefficient is the correlation of the intra-group

H matrices of allelic doses h_{ij}, where $h_{ij}=0$ or 1 if the reconstructed haplotype i contains 0 or 1 times allele j, respectively.

For all independence tests, we used an experiment-wise first type error rate of 5%. Since multiple comparisons were performed, we applied Bonferroni's correction.

Results

Core collection definition

The size of the core collection was set to 141 individuals (list of accessions given in S1). It contained 86% of the total collection diversity, as measured by the total number of classes for all variables, with a Nei index value of 0.57. It consisted of 42 T, 89 W and 10 WT individuals.

Sample diversity and structure

No locus had allele frequencies larger than 95%. The average number of alleles per locus was 9.8 in the total sample (Table 1). Observed heterozygosity varied broadly from 0.23 to 0.89 with a mean of 0.67. PCA axes 1, 2 and 3 explained 5.1%, 4.8% and 4.4% of total variability, respectively. Cultivars were not homogeneously distributed along the first three axes (Fig. 2). Some weak differentiation between T and W/WT grapevine cultivars was apparent on this figure. Fst was low but significant between T and W subsamples ($F\text{st}=0.027$, $P<0.01$), and between T and WT subsamples ($F\text{st}=0.017$, $P<0.05$), but not between W and WT subsamples ($F\text{st}=0.002$, $P>0.05$). The Fst estimates involving WT subsample should be interpreted with caution, since it contained only ten individuals. However, this differentiation remained undetected by the model-based analysis with non-admixed populations implemented in Structure. The estimated log probability of the data given the assumed number of ancestral populations K ($\ln \Pr(X|K)$) was highest for $K=3$, but the proportion of the sample assigned to each population was roughly symmetric (about one third in each population) and most individuals had fairly equivalent probabilities to belong to any of the three populations. This situation is typical of one with no population structure, as stated by J. K. Pritchard in the documentation of the software Structure. In view of these results, we decided to perform LD analysis not only on the total sample, but also on two subsamples (T and W + WT).

Table 1 Observed heterozygosity (H_{obs}), expected heterozygosity (Nei index) (H_{e}), total number of alleles, and number of alleles with frequency > 5%, in the total sample

Linkage group	Locus	H_{obs}	H_{e}	Nb alleles	Nb non-rare alleles
4	VrZAG83	0.76	0.72	5	4
4	VVIP77	0.82	0.83	14	5
4	VVIT30	0.66	0.61	5	3
4	VVIP37	0.81	0.79	10	4
4	VVMD32	0.86	0.83	12	7
4	VVIP25B	0.74	0.69	11	3
4	VMC2B5	0.23	0.73	11	5
4	VrZAG21	0.79	0.79	10	5
10	VrZAG25	0.80	0.74	10	5
10	VMC2A10	0.85	0.81	13	8
10	VVIN85	0.41	0.45	5	3
10	VMC4F9.2	0.63	0.63	6	3
10	VrZAG64	0.80	0.80	9	6
10	VVIH01	0.73	0.81	15	5
10	VMC3D7	0.68	0.68	9	4
11	VVS2	0.88	0.84	14	5
11	VVMD25	0.89	0.78	12	5
11	VMC2A12	0.50	0.48	9	2
11	VMC3E12	0.57	0.73	14	4
15	VVS16	0.50	0.51	7	3
15	VVIV67	0.79	0.84	17	8
15	VVIQ61	0.43	0.52	4	3
15	VVIP33	0.76	0.82	9	5
15	VMC5G8	0.82	0.77	9	5
15	VVIM42a	0.71	0.71	9	5
15	VVIV24	0.60	0.52	5	3
15	VVMD30	0.82	0.82	10	7
15	VMC4D9.2	0.74	0.82	12	6
15	VMC8G3.2	0.70	0.84	17	3
17	VVIB09	0.77	0.74	6	4
17	VMC9G4	0.67	0.79	11	5
17	VVIN73	0.31	0.34	5	3
17	VMC3A9	0.30	0.81	11	4
17	VVIQ22A	0.44	0.45	6	2
17	VVIQ22B	0.48	0.65	5	4
17	VMC2H3	0.43	0.72	15	4
17	SCU06vv	0.62	0.72	8	5
17	VMC3C11.1	0.81	0.76	13	5
Mean for all groups				9.8	4.5

Linkage disequilibrium

We performed 703 Fisher's exact tests to investigate the significance of genotypic disequilibrium between all loci. The comparison-wise significance threshold was therefore 7.1×10^{-5} (rounded up to 1×10^{-4}). A total of 44, 42 and 4 significant associations (Fig. 3) were recorded for the total sample, W + WT and T subsamples, respectively. All involved locus pairs were linked. Significant genotypic LD extended up to 16.8, 16.8 and 13.6 cM for the total sample (Fig. 4), W + WT and T subsamples (data not shown), respectively. The T subsample presented fewer loci in disequilibrium than the W + WT subsample, which could be due to a lack of power in this smaller subsample.

In order to determine whether the difference in genotypic LD patterns observed between W + WT and T subsamples was due to a difference in test power, we drew 50 subsamples of 42 individuals from the W + WT subsample (99 individuals) and performed the Fisher's exact test implemented in GDA for each of them. We found an average of 44.1 (standard deviation 27) locus pairs in genotypic disequilibrium over the 50 random subsamples, compared to 42 for W + WT and 4 for T subsamples. This result suggests that the difference in LD extent between T and W + WT subsamples was probably not due to a difference in power, which seems to be still large in subsamples as small as 42 individuals.

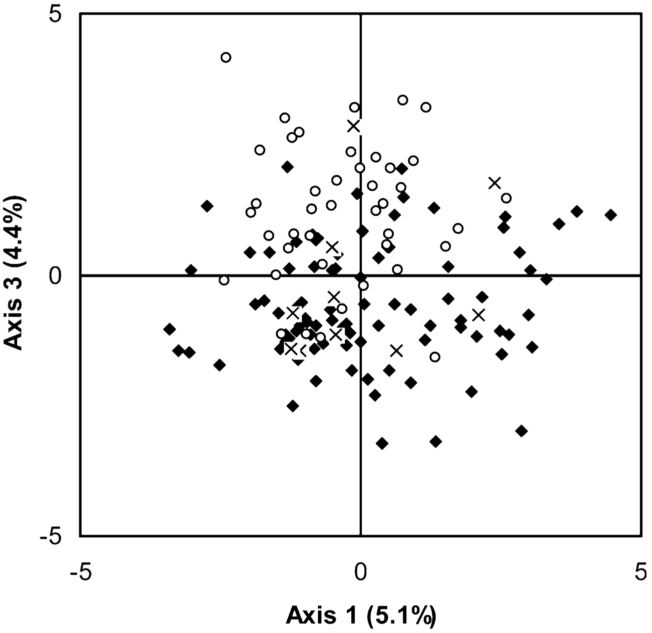
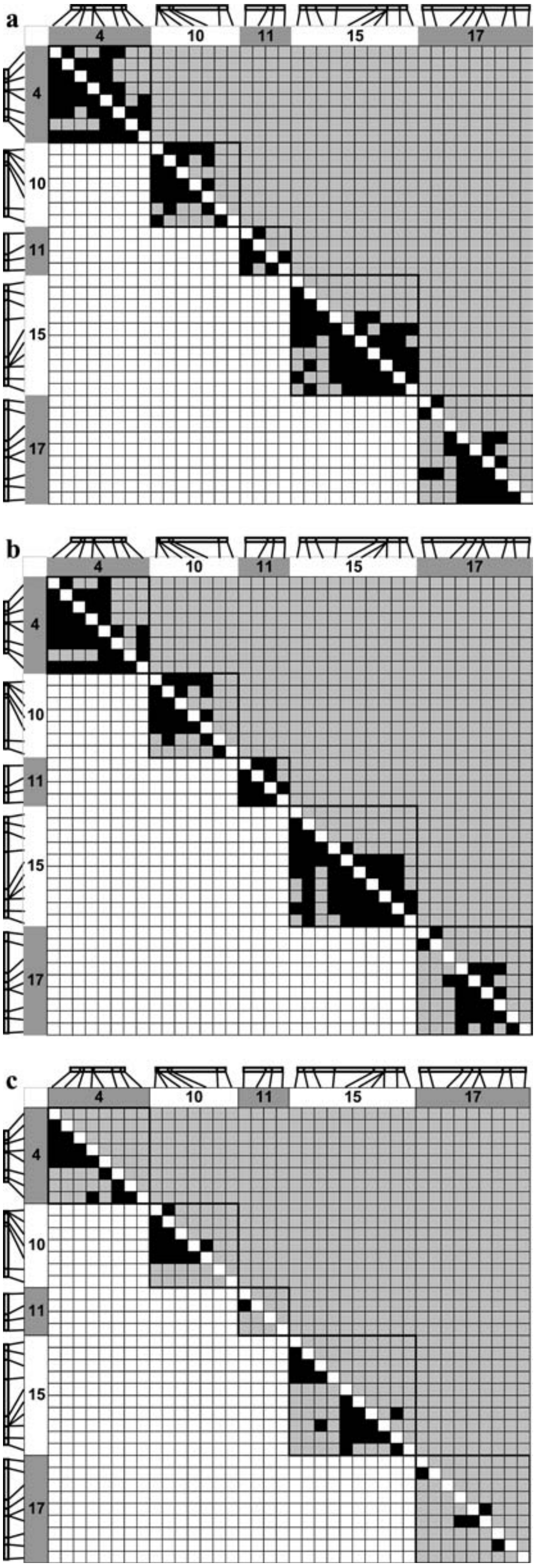


Fig. 2 Principal component analysis (axis 1 and 3) based on 38 SSR markers for a core collection of 141 *V. vinifera* table (*open circle*), wine (*filled diamond*), and double use (*cross*) cultivars

We estimated r^2 (correlation of the G matrix) from composite Δ for all pairs of loci within linkage groups. Figure 4 shows the relationship between r^2 and distance (cM) for the total sample. As expected, r^2 declined with distance. There were a few differences among linkage groups, with group 4 showing higher r^2 values than the other ones.

We also performed an haplotypic test of association based on reconstructed haplotypes within each linkage group (Fig. 3). For this test, we used the same comparison-wise significance threshold as for the genotypic test, i.e. 7.1×10^{-5} (rounded up to 1×10^{-4}), for comparison purposes. In all three samples, we found extensive haplotypic LD for linked loci, up to 33.6, 33.6 and 23.7 cM for the total sample (Fig. 4), W + WT and T subsamples (data not shown). There were more locus pairs in haplotypic than in genotypic disequilibrium (most probably due to a larger power of the association test for haplotypes than for genotypes) and all locus pairs in genotypic disequilibrium were also in haplotypic LD.

LD estimates (r^2) for reconstructed haplotypes (correlation of the intra-group H matrices) were slightly larger than for raw genotypic data, but they exhibited the same pattern of decline with distance and the same minor differences between linkage groups (Fig. 4).



▶

Fig. 3 Significance of independence tests between all loci. Within groups, loci are sorted according to the order of the map in Fig 1. The results of the genotypic tests are given in the upper right half of matrices, and the haplotypic tests (only within linkage groups) in the lower left half. *Black squares* represent significant associations (5% experiment-wise), *grey squares* non-significant associations, and white squares untested associations. *Thick lines* delimit within groups locus pairs. **a** total sample, **b** W + WT subsample, **c** T subsample

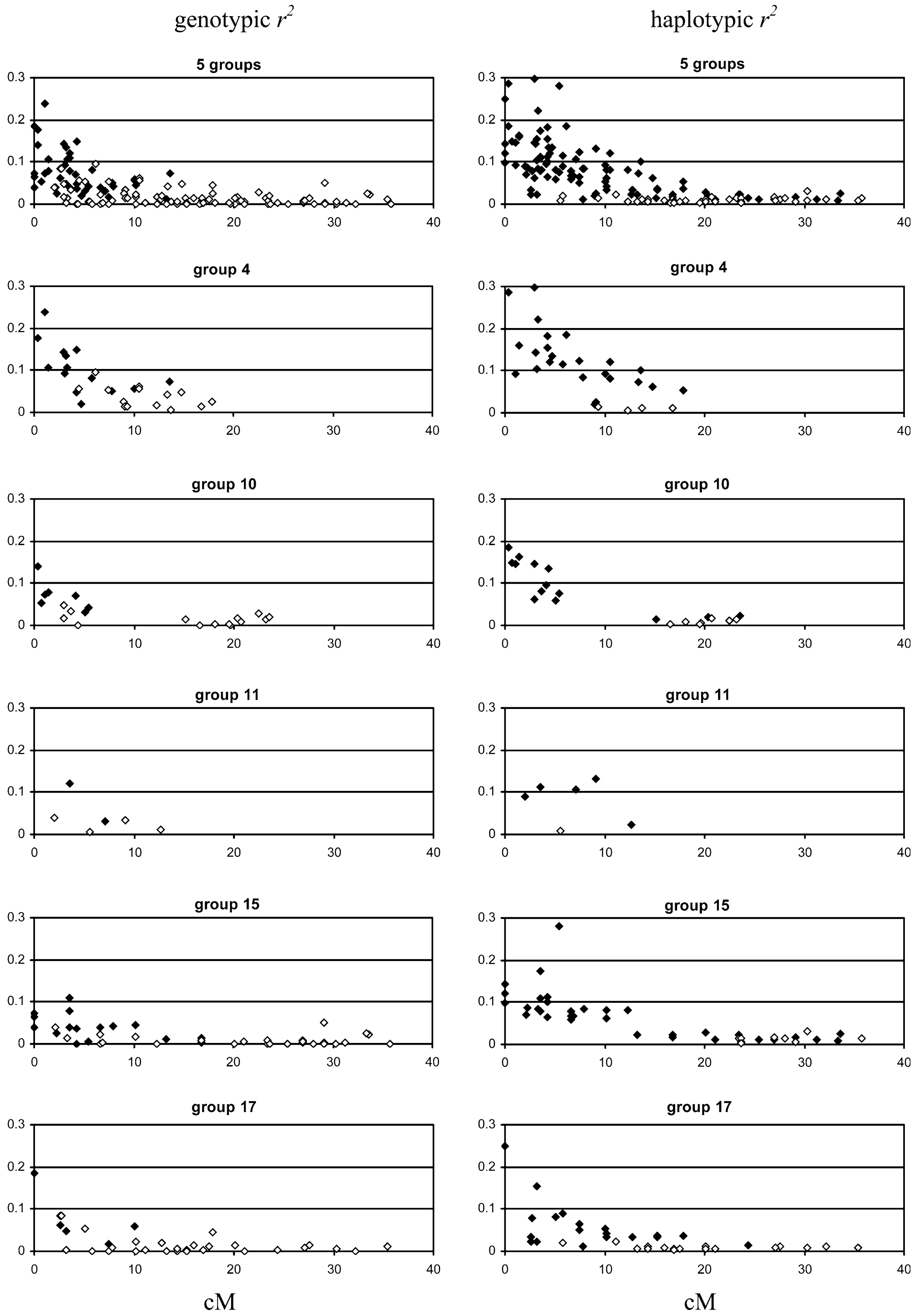


Fig. 4 r^2 values estimated on raw genotypic data or reconstructed haplotypic data as a function of distance (cM) for the total sample. Significant disequilibria are represented by (*filled diamond*) and non-significant ones by (*open diamond*)

Discussion

The first result obtained here, as a preliminary step for LD study, was the definition of the first core collection ever proposed in grapevine. Future collections will certainly be based on large-scale molecular information and should take both allelic and genotypic diversity into account. However, the overall SSR diversity (mean Nei index of 0.71) found within this core collection was largely consistent with that in previous studies of grape collection diversity (Sefc et al. 2000; Aradhya et al. 2003).

We provide the first assessment of the extent of LD in cultivated grapevine, in view of future applications such as LD mapping or marker assisted selection. Genotypic LD extended up to 16.8 cM, according to distances of the map used here (Fig. 1). Since distances in the map used in this study were smaller than distances in other published maps (total map length 23% smaller than in Riaz et al. 2004 and 6% smaller than in Adam-Blondon et al. 2004), LD in grapevine may even extend further. Moreover, we used the Bonferroni correction. Less stringent thresholds could yield more locus pairs in significant LD.

According to several authors (Nei and Li 1973; Pritchard and Prezworski 2001), structure within studied samples tends to increase LD all over the genome. The absence of genotypic LD between unlinked loci, both in total sample and in subsamples, suggests that no spurious LD was created by structure within this core collection. However, seven genotypic associations between linked loci were significant in the total sample whereas not in the W + WT or T subsamples, which could result either from structure or the larger sample size. We showed the existence of some structure in our sample by PCA and *F*st analyses, but it was weak, confirming the results of previous studies (Aradhya et al. 2003). Although the analysis we performed with Structure did not reveal any differentiation pattern, further ones with the admixture model, with more than five unlinked loci and/or assuming allele frequency correlations between ancestral populations, could reveal such weak differentiation. Anyway, it did not seem to have a large impact on LD.

The T sample showed fewer loci in genotypic disequilibrium than the total and W + WT samples. This difference did not seem to be due to a difference in sample sizes, since our test of the effect of sample size on power showed that genotypic disequilibrium analysis can efficiently detect LD even in samples as small as the T subsample. The lower amount of LD within T cultivars might rather be due to either a larger genetic base, more recombinations, or less intensive selection. For W + WT cultivars, the cultivation area is wider, which has probably lead to more differential selection for adaptation to varied environments.

Although a considerable number of LD coefficients have been developed (Hedrick 1987; Lewontin 1988), the majority is suitable only for haplotypic data. The normalized coefficients *D′* and r^2 (Hedrick 1987) are the most widely used indices. Previous studies have shown that they are more independent of allele frequencies than non-normalized ones (Hedrick 1987; Zapata 2000), although they remain sensitive (Lewontin 1988; Nordborg and Tavaré 2002). Here we used the extension of r^2 defined by Weir (1996) for unphased genotypic data, which allowed us to compare between LD estimates on the raw data and on reconstructed haplotypes. One of the main interests of a LD measure is also to allow comparisons between genome regions, populations, or species. Based on a preliminary estimate of 130–216 kb/cM for the correspondence between genetic and physical distances in grapevine (Adam-Blondon et al. 2005), a cautious comparison of LD extent is possible with r^2 results published for other plant species. In the present study, r^2 values decreased to 0.1 within ca. 5 cM/650–1,080 kb (for genotypic data) or 10 cM/1,300–2,160 kb (for haplotypic data). Thus LD extended farther than in maize (Remington et al. 2001) and rice (Garris et al. 2003), where r^2 values reached 0.1 within ca. 2 and 100 kb, respectively. Its extent was similar to that in barley (Kraakman et al. 2004), in soybean (Zhu et al. 2003) and in another study of maize (Tenaillon et al. 2001), where 0.1 values of r^2 were found around 10 cM, 10 cM and 1,000 kb, respectively. It was smaller than in durum wheat (Maccaferri et al. 2005), where some r^2 values larger than 0.1 were maintained up to 30–40 cM. Therefore, LD extent as measured by r^2 decline seems to be moderate in grapevine as compared to other species, despite significant genotypic and haplotypic associations extending up to more than 15 and 30 cM, respectively.

Despite the high polymorphism of SSR markers and small sample sizes, which we expected to seriously hinder the detection of genotypic associations, extensive significant genotypic LD could be found in the total and W + WT samples. Moreover, r^2 values were comparable for genotypic and haplotypic data. Our results therefore demonstrate that studying LD at the genotypic level can yield valuable information and is potentially useful for all species for which there is neither information on genealogy nor straightforward access to haplotypes.

One important question arising from our results is whether the genotypic LD found is mainly of haplotypic origin (alleles associated in coupling). Two elements of answer can be proposed. First, when inspecting genotype contingency tables (data not shown), we noticed that the genotype associations responsible for significant LD largely involved double homozygotes, suggesting mainly haplotypic disequilibrium. Unfortunately, the level of heterozygosity was too high to perform a genotypic association test on homozygous loci only to quantify this observation. Second, we tested for haplotypic equilibrium after reconstructing haplotypes. Most loci in genotypic disequilibrium were also in haplotypic disequilibrium, strongly suggesting that the LD found in our study was mainly of haplotypic origin.

The rather extensive LD revealed here provides some hints on the factors that have lead to present diversity. Primary domestication of grapevine probably took place in the Near East or in the Transcaucasus, 6,000–4,000 years BC, and it then rapidly spread over Europe and Northern Africa (reviewed by Grassi et al. 2003). Domestication involved hermaphroditism selection, probably inducing severe bottlenecks and the creation of LD. It was also associated to the adoption of vegetative propagation, that may have maintained LD through limited recombination. For example, some cultivars still widely cultivated today were created in the Middle Ages (Bowers et al. 1999). Nevertheless, Arroyo-Garcia et al. (2002) and Grassi et al. (2003) suggested the occurrence of secondary domestication events in Europe, which could have increased local diversity and decreased LD. In our study, we found rather extensive LD in five linkage groups. This LD pattern is suggestive of a quite narrow genetic base with limited recombination and/or selection.

The rather large extent of LD also suggests that genome-wide QTL mapping strategies exploiting LD could be effective in grapevine, particularly for W + WT germplasm. Contrary to human, where a very high density of markers is necessary (Kruglyak 1999) except for LD blocks (Stumpf and Goldstein 2003), for grapevine as for barley (Kraakman et al. 2004), whole genome scan is likely to be a viable approach with less markers. Conversely, fine mapping resolution may be limited.

Our results also indicate that it could be possible to use genotypes from the germplasm collection as progenitors for crosses in marker assisted selection programs, without having to systematically reexamine the association between QTL and linked marker alleles.

Acknowledgements This research was funded by INRA. We thank D. Vares for Vassal collection plant and data management, V. Laucou for providing the DNA, T. Bataillon for his help with Mstrat, and two anonymous reviewers, B. Courtois, M. Deu, V. Laucou, J. P. Péros, and H. A. Tsitrone for useful comments on a previous version of the manuscript.

References

Adam-Blondon AF, Roux C, Claux D, Butterlin G, Merdinoglu D, This P (2004) Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics. Theor Appl Genet 109:1017–1027

Adam-Blondon AF, Bernole A, Faes G, Lamoureux D, Pateyron S, Grando MS, Caboche M, Velasco R, Chalhoub B (2005) Construction and characterization of BAC libraries from major grapevine cultivars. Theor Appl Genet 110:1363–1371

Aradhya MK, Dangl GS, Prins BH, Boursiquot JM, Walker MA, Meredith CP, Simon CJ (2003) Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. Genet Res 81:179–192

Arroyo-Garcia R, Lefort F, de Andrés MT, Ibañez J, Borrego J, Jouve N, Cabello F, Martinez-Zapater JM (2002) Chloroplast microsatellite polymorphisms in *Vitis* species. Genome 45:1142–1149

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2002) GENETIX 4.04, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France

Bowers J, Boursiquot JM, This P, Chu K, Johansson H, Meredith C (1999) Historical genetics: the parentage of Chardonnay, Gamay, and other wine grapes of Northeastern France. Science 285:1562–1565

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Nat Acad Sci USA 101:15255–15260

Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith CP, Edwards KJ, This P (2002) Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. Theor Appl Genet 105:780–795

Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Res 10:220–227

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

Grassi F, Labra M, Imazio S, Spada A, Sgorbati S, Scienza A, Sala F (2003) Evidence of a secondary grapevine domestication centre detected by SSR analysis. Theor Appl Genet 107:1315–1320

Gouesnard B, Bataillon TM, Decoux G, Rozale C, Schoen DJ, David JL (2001) Mstrat: an algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. J Hered 92:93–94

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

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331–341

Jannoo N, Grivet L, Dookun A, D'hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. Theor Appl Genet 99:1053–1060

Jorde LB (2000) Linkage disequilibrium and the search for complex disease genes. Genome Res 10:1435–1444

Jung M, Ching A, Bhattramakki 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

Kraakman ATW, Niks RE, Van den Berg PMMM, 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

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

Lewis PO, Zaykin D (2002) GDA 1.1. The University of Connecticut

Lewontin RC (1988) On measures of gametic disequilibrium. Genetics 120:849–852

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–289

MacRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J (2002) Linkage disequilibrium in domestic sheep. Genetics 160:1113–1122

Mohlke KL, Lange EM, Valle TT, Ghosh S, Magnuson V L, Silander K, Watanabe RM, Chines PS, Bergman RN, Tuomilehto A, Collins FS, Boehnke M (2001) Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in Finns. Genome Res 11:1221–1226

Nei M (1978) Eestimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Nei M, Li WH (1973) Linkage disequilibrium in subdivided populations. Genetics 75:213–219

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

Nsengimana J, Baret P, Haley CS, Visscher PM (2004) Linkage disequilibrium in the domesticated pig. Genetics 166:1395–1404

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Nat Acad Sci USA 98:11479–11484

Riaz S, Dangl GS, Edwards KJ, Meredith CP (2004) A microsatellite based framework linkage map of *Vitis vinifera* L.. Theor Appl Genet 108:864–872

Schoen D, Brown A (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc Nat Acad Sci USA 22:10623–10627

Sefc KM, Lopes MS, Lefort F, Botta R, Roubelakis-Angelakis KA, Ibanez J, Pejic I, Wagner HW, Glossl J, Steinkellner H (2000) Microsatellite variability in grapevine cultivars from different European regions and evaluation of assignment testing to assess the geographic origin of cultivars. Theor Appl Genet 100:498–505

Simko I (2004) One potato, two potato: haplotype association mapping in autotetraploids. Trends Plant Sci 9:441–448

StatSoft France (2005). STATISTICA (logiciel d'analyse de données), version 7.1. http://www.statsoft.fr

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

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction. Am J Hum Genet 73:1162–1169

Stumpf MPH, Goldstein DB (2003) Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. Curr Biol 13:1–8

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 Nat Acad Sci USA 98:9161–9166

Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM (2003) Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. J Anim Sci 81:617–623

Weir SB, Cockerham C (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Weir SB (1996) Genetic data analysis II. Sinauer Associates, Inc., Sunderland, pp 91–138

Zapata C (2000) The *D*' measure of overall gametic disequilibrium between pairs of multiallelic loci. Evolution 54:1809–1812

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