Theor Appl Genet (2010) 121:659–672
DOI 10.1007/s00122-010-1338-1

ORIGINAL PAPER

Carotenoid biosynthesis genes provide evidence of geographical subdivision and extensive linkage disequilibrium in the carrot

Jérémy Clotault · Emmanuel Geoffriau · Eric Lionneton · Mathilde Briard · Didier Peltier

Received: 21 October 2009 / Accepted: 3 April 2010 / Published online: 22 April 2010

Abstract According to the history of the cultivated carrot, root colour can be considered as a structural factor of carrot germplasm. Therefore, molecular variations of carotenoid biosynthesis genes, these being involved in colour traits, represent a good putative source of polymorphism related to diversity structure. Seven candidate genes involved in the carotenoid biosynthesis pathway have been analysed from a sample of 48 individual plants, each one from a different cultivar of carrot (*Daucus carota* L. ssp. s*ativus*). The cultivars were chosen to represent a large diversity and a wide range of root colour. A high single nucleotide polymorphism (SNP) frequency of 1 SNP per 22 bp (mean $\pi_{sil} = 0.020$) was found on average within these genes. The analysis of genetic structure from carotenoid biosynthesis gene sequences and 17 putatively neutral microsatellites showed moderate genetic differentiation between cultivars originating from the West and the East ($F_{ST} = 0.072$), this being consistent with breeding history, but not previously evidenced by molecular tools. Surprisingly, carotenoid biosynthesis genes did not exhibit decay of LD (mean $r^2 = 0.635$) within the 700–1,000 bp analysed, even though a fast decay level of LD is expected in outcrossing species. The high level of intralocus LD found for carotenoid biosynthesis genes implies that candidate-gene association mapping for carrot root colour should be useful to validate gene function, but may be unable to identify precisely the causative variations involved in trait determinism. Finally this study affords the first molecular evidence of a genetic structure in cultivated carrot germplasm related to phylogeography.

Communicated by M. Havey.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1338-1) contains supplementary material, which is available to authorized users.

J. Clotault · E. Geoffriau (✉) · M. Briard
Agrocampus Ouest, INHP, IFR 149 Quasav,
UMR 1259 GenHort, 2 Rue Le Nôtre,
49045 Angers, France
e-mail: emmanuel.geoffriau@agrocampus-ouest.fr

E. Lionneton
Clause Vegetable Seeds, 49070 Beaucouzé, France

D. Peltier
Université d'Angers, IFR 149 Quasav, UMR 1259 GenHort,
49045 Angers, France

Introduction

Morphological characteristics lead to a division of the cultivated carrot (*Daucus carota* subsp. *sativus*) into two botanical varieties: var. *atrorubens* and var. *sativus* (Small 1978). Var. *atrorubens* refers to carrots originating from the East, exhibiting yellow or purple storage roots and poorly indented, grey-green, pubescent foliage. Var. *sativus* refers to carrots originating from the West and exhibiting orange, yellow or sometimes white roots, and highly indented, non-pubescent, yellow-green foliage (Small 1978). Many intermediate variants exist between these two types.

Despite this taxonomic differentiation between geographical groups, no population structure has been found in carrot germplasm by examining random molecular markers such as isozymes (St. Pierre and Bayer 1991; St. Pierre et al. 1990); random, amplified, polymorphic DNA (RAPD; Grzebelus et al. 2002; Nakajima et al. 1998, 1997); amplified-fragment length polymorphism (AFLP; Bradeen et al. 2002; Nakajima et al. 1998; Shim and Jørgensen 2000); and inter-simple sequence repeat (ISSR; Bradeen et al. 2002). This lack of structure, despite the

morphological evidence, has been explained by the outcrossing mating system and frequent crossings within carrot germplasm (Bradeen et al. 2002).

The carrot is believed to originate from Afghanistan before the 900s, as this area is described as the primary centre of greatest carrot diversity (Mackevic 1929), Turkey being proposed as a secondary centre of origin (Banga 1963). The first cultivated carrots exhibited purple or yellow roots. Carrot cultivation spread to Spain in the 1100s via the Middle East and North Africa. In Europe, genetic improvement led to a wide variety of cultivars. White and orange-coloured carrots were first described in Western Europe in the early 1600s (Banga 1963). Concomitantly, the Asiatic carrot was developed from the Afghan type and a red type appeared in China and India around the 1700s (Laufer 1919; Shinohara 1984). According to this history, it makes sense to envisage that colour should be considered as a structural factor in carrot germplasm.

Carotenoid composition determines the white, yellow, orange or red root colour in the carrot (Nicolle et al. 2004; Surles et al. 2004). The variability of carotenoid accumulation in crops is generally determined by allelic variation in genes from the carotenoid biosynthesis pathway. Single nucleotide polymorphisms (SNPs) in the coding region of *LCYB*, encoding lycopene β-cyclase, co-segregates with flesh colour phenotypes in a population of segregating derivatives from an intercross between canary-yellow- and red-coloured watermelon varieties (Bang et al. 2007). Many polymorphic sites show strong association with the maize endosperm colour phenotype at *Y1*, encoding phytoene synthase (Palaisa et al. 2003). Candidate-gene association mapping shows an association between *LCYE* (lycopene ε-cyclase) polymorphisms and maize kernel carotenoid composition (Harjes et al. 2008). Association mapping in durum wheat identified new quantitative trait loci (QTLs) and found a significant association between *Psy1-B1*A, a phytoene synthase gene, and the yellow pigmentation of kernel (Reimer et al. 2008). In the carrot root, the genetic determinism of carotenoid composition remains unclear. However, some major genes and QTLs involved in carrot root colour control were recently shown to co-localise with genes involved in the carotenoid biosynthesis pathway (Just et al. 2009). Carotenoid biosynthesis genes therefore represent valuable candidate genes to investigate differential root pigmentation in the carrot. Polymorphism within such genes, having been involved in a trait that has been highly bred for in this species' history, should reflect demographic events such as population subdivision, gene flow, or genetic drift, as well as selective events (van Tienderen et al. 2002).

Population structure affects molecular polymorphism by modifying allele frequencies amongst each population. This modification of the probability of allele exchange between populations therefore impacts on linkage disequilibrium (LD), i.e., the non-random association between two markers. Population subdivision can generate background linkage disequilibrium, even amongst unlinked markers such as markers from different linkage groups. Apart from population structure, LD is largely affected by the balance between the mutation rate and the recombination rate. New mutations arising in a genome region increase LD levels, which subsequently decrease by recombination events (Rafalski and Morgante 2004). The mating system is expected to influence LD: the effective recombination rate is higher in outcrossing species such as maize (Remington et al. 2001) than in selfing species such as *Arabidopsis thaliana* (Nordborg et al. 2002), because they have more loci with heterozygous alleles in outcrossing than in selfing species. Linkage disequilibrium has never been characterised in the carrot, an outcrossing, biennial, diploid species.

Improving our knowledge of population structure and linkage disequilibrium would be informative for designing association-mapping strategies. The distance over which LD persists determines the density of the markers required, the mapping resolution expected, and therefore the appropriate experimental design for association studies (Flint-Garcia et al. 2003; Gupta et al. 2005). Association studies in a subdivided population can result in spurious associations. A thorough knowledge of LD and population structure is therefore prerequisite to choosing between starting a candidate-gene or genome-wide association analysis, to help constitute the sample investigated, and to select appropriate analysis methods.

This study brings together information about population structure in cultivated carrot germplasm and linkage disequilibrium in carotenoid biosynthesis genes. We sequenced seven carotenoid biosynthesis genes and genotyped 17 microsatellite loci for 48 individuals. This combinatorial approach between carotenoid biosynthesis genes and simple sequence repeats allowed us to test genetic structure according to geographic origin or root colour within cultivated carrot germplasm. The extent of intralocus and interlocus linkage disequilibrium and nucleotide sequence diversity was assessed in this outcrossing crop. These results will be discussed regarding the geographical and historical elements which could have imprinted cultivated carrot germplasm diversity structure, and according to the development of an association mapping strategy for identifying alleles involved in carotenoid content.

Materials and methods

Plant material

Forty-eight cultivars of carrot (*Daucus carota* L. ssp. *sativus*) were sampled (Table 1). These cultivars, obtained

Table 1 Set of carrot cultivar samples

Code	Cultivar name	Geographical origin	Source	Root color	Varietal type
106	Snow White	USA	Territorial seeds	White	Open-pollinated
111	Amber White	USA	Kitchen garden seeds	White	Open-pollinated
301	Sugarsnax	USA	Nunhems	Orange	Hybrid
104	Blanche Collet Vert très Hors Terre	France	INH	White	Open-pollinated
105	Blanche Demi Longue des Vosges	France	INH	White	Open-pollinated
210	Jaune du Doubs	France	INH	Yellow	Open-pollinated
311	Nantaise améliorée 4	France	INH	Orange	Open-pollinated
312	Bellot	France	INH	Orange	Open-pollinated
316	Parisienne Market 2	France	INH	Orange	Open-pollinated
337	De Colmar à Cœur Rouge 2	France	INH	Orange	Open-pollinated
109	White Belgian	United Kingdom	HRI	White	Open-pollinated
100	Kuttiger	Switzerland	INH	White	Open-pollinated
107	White Satin	Holland	Bejo	White	Hybrid
201	Yellowstone	Holland	Bejo	Yellow	Open-pollinated
313	Amsterdam 2 Sweetheart	Holland	HRI	Orange	Open-pollinated
108	Long White Green Top	Denmark	HRI	White	Open-pollinated
208	Gelbe Lobbereicher	Germany	HRI	Yellow	Open-pollinated
303	Boléro	Western Europe	Vilmorin	Orange	Hybrid
336	Nandor	Western Europe	Clause vegetable seeds	Orange	Hybrid
339	Premia	Western Europe	Syngenta seeds	Orange	Hybrid
345	Ceres	Western Europe	Clause vegetable seeds	Orange	Hybrid
348	Adour	Western Europe	Syngenta seeds	Orange	Hybrid
349	Siroco	Western Europe	Vilmorin	Orange	Hybrid
928	Nairobi	Western Europe	Bejo	Orange	Hybrid
901	Tavola	Western Europe	Nunhems	Orange	Hybrid
526	PD526	Middle East	INH	Purple	Breeding material
520	PM520	Middle East	INH	Purple	Open-pollinated
515	PT2 515	Middle East	INH	Purple	Open-pollinated
500	Anthocyanee	Middle East	INH	Purple	Open-pollinated
521	PD521	Middle East	INH	Purple	Open-pollinated
522	PA522	Middle East	INH	Purple	Open-pollinated
366	Mestnaya Zheltaya	Uzbekistan	VIR	Orange	Open-pollinated
227	LR Mestnaya	Afghanistan	VIR	Yellow	Open-pollinated
514	Afghan Purple	Afghanistan	HRI	Purple	Open-pollinated
403	Annual Red Rawalpindi	Pakistan	HRI	Red	Open-pollinated
420	Pusa Kesar	India	HRI	Red	Open-pollinated
516	PD516	Asia	INH	Purple	Open-pollinated
411	Red Queen	India	Sungro	Red	Hybrid
224	Ch-Wy	Asia	Mikado-Kyowa seeds	Yellow	Breeding material
226	YC226	Asia	INH	Yellow	Open-pollinated
407	Pink Selection	China	HRI	Red	Open-pollinated
421	JF421	Asia	INH	Red	Open-pollinated
307	Kuroda	Japan	INH	Orange	Open-pollinated
426	Honbeni Kintoki	Japan	Takii	Red	Open-pollinated
225	Yellow BM	Asia	Vilmorin	Yellow	Breeding material
423	Red Pink BM	Asia	Vilmorin	Red	Breeding material
400	Nutrired	Asia/USA	Seminis	Red	Open-pollinated
200	Kinby	Asia/USA	Johnny seeds	Yellow	Hybrid

from gene banks and seed companies, were chosen to maximise diversity, based on geographic origin, root colour and shape. The allelic diversity was sampled by analysing one individual for each cultivar. DNA was extracted from root powder following a standard CTAB protocol (Briard et al. 2000).

Microsatellite genotyping

Individuals were genotyped at 17 microsatellite loci. The PCR mix contained 8 pmol of reverse primers, 8 pmol of fluorescent dye-labelled M13 primer, and 2 pmol of the forward primer in a final 10 µl reaction volume containing Invitrogen PCR buffer, 2.5 mM $MgCl_2$, 0.25 mM dNTPs, 0.5 U *Taq* DNA polymerase (Invitrogen) and 50–100 ng DNA template. PCR was performed with a GeneAmp PCR system 2700 (Applied Biosystems, Foster City, CA) with a thermal cycle under the following conditions: 94°C for 5 min, then 10 cycles at 94°C for 15 s, 55°C for 15 s with a −0.5°C per cycle and 72°C for 30 s, followed by 30 cycles at 94°C for 15 s, 50°C for 15 s, 72°C for 30 s, and a final extension at 72°C for 10 min. Subsequently, PCR products were diluted in 1/40 ultrapure water; 5 µl of diluted PCR products was added to 5 µl of formamide and 500 LIZ size standard (Applied Biosystems). Amplified fragments were analysed by capillary electrophoresis (ABI 3730 DNA Analyzer, Applied Biosystems, Foster City, CA), and the labelled PCR products were automatically sized with Genemapper Software (Applied Biosystems, Foster City, CA). Microsatellite primers are available on request to the corresponding author.

Choice of sequenced loci

Seven carotenoid biosynthesis genes were chosen according to their position in the pathway: *IPI*, *PDS*, *CRTISO*, *LCYB1*, *LCYE*, *CHXE* and *ZEP* (Just et al. 2007). *IPI* and *PDS* were chosen because they act upstream in the pathway. *CRTISO* is required to complete lycopene synthesis in dark-grown tissues (Isaacson et al. 2002; Park et al. 2002) and is therefore worth studying in the carrot root. Sequence polymorphisms of an orthologue of *LCYB1* are involved in flesh colour variations in watermelon (Bang et al. 2007). LCYE is involved in channelling metabolites towards the pathway branch generating α-carotene and lutein contained in some carrot types (Surles et al. 2004). Transcript levels of *LCYE* were higher in yellow carrots as compared to other types (Clotault et al. 2008). *CHXE* encodes the enzyme catalysing lutein synthesis, which is the major pigment in yellow carrots (Surles et al. 2004). *CHXE*, *PDS* and *ZEP* were mapped in regions including the most significant QTLs for carotenoid levels in the carrot (Just et al. 2007, 2009; Santos and Simon 2002). These loci are located on six of the nine linkage groups of the carrot genome.

PCR, cloning and sequencing

A list of primers used for PCR and sequencing is available in Table 2. PCR was performed with an MJ Research PTC-100 thermal cycler under the following conditions: 35 cycles of 94°C for 30 s, annealing temperature for 45 s and 72°C for 2 min. The 25 µl-PCR mix contained 50 ng DNA, 0.2 µM of each primer, 0.2 mM of each deoxynucleotide triphosphate, 2 mM $MgCl_2$, 1× buffer (Interchim), and 1 unit of Taq Uptitherm DNA polymerase (Interchim) in sterile water. After directly sequencing PCR products, 30–50% of the amplified fragments did not give readable sequence because of their heterozygous status. These fragments were then cloned into the pGEM-T Easy vector (Promega; Madison, WI). Eight individual clones were chosen for each cloned PCR product and were sequenced.

Prior to sequencing, PCR products were purified with ExoSAP-IT (USB Corporation, Cleveland, USA) according to the supplier's instructions. Amplicons were sequenced using an Applied Biosystems ABI Prism 3730 sequencer (Weiterstadt, Germany) and BigDye-terminator v3.1 chemistry. Sequences were aligned with ClustalW (Thompson et al. 1994) and edited by eye with BioEdit 7.0.5.3 (Hall 1999). If cloning results revealed two different alleles, a single allele was randomly chosen for the following analyses to avoid representation bias among genes. For each gene, a reference nucleotide sequence was deposited in GenBank under the accession numbers FN662487–FN662493.

Sequence polymorphism

DNA sequences were computed using DnaSP 4.9 (Rozas et al. 2003). Sites with alignment gaps were excluded from analysis. Nucleotide polymorphism θ_w (Watterson 1975), nucleotide diversity π (Nei 1987), the number of haplotypes h and haplotype diversity Hd (Nei 1987) were calculated for each locus. Thus π depends on both the number of segregating sites and their frequency and θ_w depends on the number of segregating sites only; π was calculated for all sites (π_T), silent sites (i.e., intronic regions plus synonymous sites; π_{sil}), synonymous sites (π_{syn}) and non-synonymous sites (π_{nonsyn}).

Genetic structure

The genetic structure was investigated using microsatellites owing to the putative neutral status of these loci. We used the model-based programme STRUCTURE 2.2 (Pritchard et al. 2000a) to infer population structure and to assign

Table 2 Sequence and position of primers used for amplification of carotenoid biosynthesis gene parts

Genes	Gene function	GenBank cDNA accession number[a]	GenBank partial genomic DNA accession number[b]	Forward primer		Reverse primer		Annealing temperature (°C)
				Sequence	Position[c]	Sequence	Position[c]	
IPI	Isopentenyl pyrophosphate isomerase	DQ192183	FN662489	caacgttctgccaccaaggtaa	Exon 4	ctgaactattcctatgcggtgg	3′-UTR	55
PDS	Phytoene desaturase	DQ222429	FN662490	gctggtcacaagcccatatt	Exon 3	atgctccttccactgcaatc	Exon 5	57
CRTISO	Carotenoid isomerase	DQ192188	FN662491	gcccttgaattttgggtttt	Exon 1	cgccaatgcttgagttatca	Exon 2	55
LCYB1	Lycopene *β*-cyclase 1	DQ192190	FN662492	tgatggtgtgaccattcaagctgc	cds[d]	ccgtgccttttgccatgatctcta	cds[d]	57
LCYE	Lycopene *ε*-cyclase	DQ192192	FN662493	ctgcagtatgaggttgggggtccc	Exon 5	tgcctccatgcctggaacttttg	Exon 6	55
CHXE	*ε*-ring carotene hydroxylase	DQ192196	FN662487	agcattctacggttcctgct	Exon 3	ggaggccttccttctaaaactc	Exon 5	55
ZEP	Zeaxanthin epoxidase	DQ192197	FN662488	caccgctgttggggatgaaattat	Exon 2	atctttgccaccagcaggttcatt	Exon 5	55

[a] GenBank cDNA sequence from which primers were designed

[b] GenBank partial genomic DNA sequence from the present study

[c] Putative exon position within cDNA of carrot genes was determined by comparison with gDNA of orthologues

[d] Coding sequence. No putative intron was found in *LCYB1*

individuals to populations using a burn-in period of 200,000 iterations, a run length of 100,000 iterations, and a model allowing for admixture and correlated allele frequencies. STRUCTURE was used without any prior information on the origin of each individual. The number of populations (K) was tested from 1 to 10. Ten independent runs were carried out for each K. The most probable K was determined by comparing the rate of change in the log probability of data between successive Ks (Evanno et al. 2005). We also used a second software, BAPS 5.2 (Corander et al. 2003), to implement a different Bayesian model-based clustering method in order to check STRUCTURE assignation. Pairwise population differentiation was calculated based on F_{ST} (Weir and Cockerham 1984) using GENETIX 4.05 (Belkhir et al. 2004). Signification was tested by 10,000 permutations of individuals among groups. The level of unbiased, expected heterozygosity *He* was calculated as a measure of the degree of genetic variability (Nei 1978).

The study of the genetic structure evidenced by sequence polymorphism was performed from haplotype data rather than segregating sites because differences in nucleotide diversity between genes could bias analysis, and the independence of loci is an *a priori* requirement for the use of Bayesian, model-based clustering methods. All parsimony-informative (non-singleton) mutations were used to define haplotypes with SNAP (Price and Carbone 2005). Within each locus, haplotype fragments with independent genealogical history were defined on the basis of evidence of recombinations inferred from a four-gamete test (Hudson and Kaplan 1985). Independence among haplotype segments was tested using a Fisher's Exact Test for linkage as implemented in GDA 1.1 (Lewis and Zaykin 2001). Seven independent haplotype segments, each obtained from one gene, were considered for analysis. The partitioning of individuals between two clusters was tested by 10,000-run repetitions using BAPS.

The differentiation of carotenoid biosynthesis gene sequences among groups defined previously by STRUCTURE analysis from microsatellite data was measured by G_{ST} and $F_{ST,}$ and its significance was tested with $S_{nn,}$ using DnaSP. The F_{ST} was calculated as the ratio between the estimated nucleotide diversity within groups over the estimate of total nucleotide diversity (Hudson et al. 1992). In contrast, G_{ST} is the ratio between the estimated haplotype diversity within groups over the estimate of the total haplotype diversity (Nei 1973). S_{nn} (i.e., the nearest-neighbour statistic) is a measure of how often the nearest neighbours of sequences are found in the same group (Hudson 2000). We assessed S_{nn} significance using a 10,000-replicate permutation test.

Linkage disequilibrium

For the analysis of LD, only biallelic single nucleotide polymorphisms (SNPs) of at least 10% frequency were considered, as rare alleles can have large variances in LD estimates. The correlation coefficient r^2 between each SNP pair (Hill 1974) was calculated using TASSEL 2.0.1 (Bradbury et al. 2007).

Intragenic LD was investigated by plotting r^2 for all pairwise comparisons among SNPs within the same gene as a function of physical distance. The decay of LD with distance was evaluated by non-linear regression analysis (PROC NLIN in SAS software). We compared two models (Remington et al. 2001) evaluating the expected value of r^2 with observed data: $E(r^2) = 1/(1 + 4Nc)$, where N is the effective population size, c is the recombination fraction between sites, and Eq. 1 in Remington et al. (2001). Intergenic LD was investigated by plotting r^2 for the pairwise comparisons of polymorphic sites between pairs of genes.

Results

Nucleotide diversity

The sequencing of seven carotenoid biosynthesis genes in 48 carrot individuals resulted in 6,675 bp of aligned sequences (Table 3). Sequence length ranged from 718 bp in *ZEP* to 1,393 bp in *IPI*, with an average of 954 bp. The insertions/deletions (indels) cumulated length ranged from 0 (*LCYB1*) to 511 bp (*IPI*). The longest indel consisted in a 468 bp-miniature inverted-repeat transposable element (MITE) in intron 4 of *IPI* (data not shown). Excluding indels, a total of 5,829 bp of aligned sequences was analysed. Coding regions and non-coding regions contributed respectively to 36 and 64% of the sequences analysed. Within the seven gene fragments, 265 single nucleotide polymorphisms (SNPs) were found. SNP frequency ranged from 1 SNP per 11 bp for *CRTISO* to 1 SNP per 48 bp for *PDS*, with an average of 1 SNP per 22 bp (Table 3).

CRTISO ($\theta_w = 0.0198$, $\pi_T = 0.0345$), *LCYE* ($\theta_w = 0.0148$, $\pi_T = 0.0246$) and *CHXE* ($\theta_w = 0.0120$, $\pi_T = 0.0116$) showed the highest nucleotide polymorphism and diversity values, whereas *PDS* ($\theta_w = 0.0047$, $\pi_T = 0.0038$) showed the lowest indexes (Table 4). The average silent-site diversity ($\pi_{sil} = 0.0204$) and synonymous-site diversity ($\pi_{syn} = 0.0294$) levels were higher than non-synonymous diversity ($\pi_{nonsyn} = 0.0007$), in accordance with the assumption that coding regions are subject to background selection. The number of haplotypes per locus ranged from 9 for *CHXE* to 16 for *PDS* among the 48 individuals with an average of 11.9. Haplotype diversity (*Hd*) ranged from 0.523 for *CHXE* to 0.851 for *IPI* with an average of 0.751.

Population structure

The population genetic structure can be affected either by natural or artificial selection or by demographic events. Differentiation between these factors requires the use of markers unaffected by selection. Therefore, we chose to study the population structure by using 17 microsatellite markers first and then carotenoid biosynthesis genes. Microsatellite loci confirmed the high diversity found in the carrot: 17 loci exhibited a total of 155 alleles (size ranging from 113 to 448 bp with 3 to 18 alleles per locus and an average of 9.1 alleles per locus as given in Supplementary Table 1) for 47 individuals (data missing for Bellot). The mean expected heterozygosity (*He*) was 0.729 in the sample. The Bayesian model-based clustering approach implemented by STRUCTURE was used to infer population structure and to assign individuals to populations sharing similar genotypes (Pritchard et al. 2000a). This approach showed that the sample was most probably divided between two clusters (Evanno et al. 2005; Fig. 1a). Despite the choice of a model allowing for admixture between several clusters, only individuals 515 and 500 showed less than 80% of probability of assignment to one of the two clusters. These two individuals were therefore excluded from subsequent analyses of geographical differentiation based on carotenoid biosynthesis gene sequences. The three most admixed individuals (515, 500 and 520) correspond to open-pollinated cultivars originating from Middle East. The first cluster contained 26 individuals, only sampled into cultivars originating from Europe or North America, except the Japanese cultivar Kuroda (307) and Asian breeding material Yellow BM (225). We considered this cluster as the 'Western group' in further analyses. The second cluster contained 19 individuals (excluding the two most admixed cultivars), only sampled into cultivars originating from Asia, including the Middle East, or of undetermined origin, i.e., cultivars bred in the USA but with supposed Asian origins. We considered this cluster as the 'Eastern group' in further analysis. When using BAPS, a second Bayesian clustering approach built on a different model, the individuals were assigned very similarly to the two clusters. The F_{ST} among the Western and the Eastern groups defined by STRUCTURE is 0.072 ($P < 0.001$). Even if this F_{ST} value should be considered as moderate, it is highly significant, reinforcing the assignation obtained by Bayesian methods.

The structure inferred from putative neutral microsatellite loci was compared with the structure evidenced by carotenoid biosynthesis gene polymorphism (Fig. 1b). The

Table 3 Nucleotide polymorphism of seven carotenoid biosynthesis genes in a sample of 48 cultivated carrots

Genes	No. of sequences	Sequences with indels		Sequences without indels				
		Total length (bp)	Indel length (bp)	Total length (bp)	Coding region length (bp)	Non-coding region length (bp)	No. of SNPs	SNP frequency (SNPs/bp)
IPI	47	1,393	511	882	417	465	33	1/27
PDS	48	1,055	1	1,054	153	901	22	1/48
CRTISO	48	996	130	866	252	614	76	1/11
LCYB1	48	770	0	770	770	0	20	1/38
LCYE	48	936	172	764	99	665	50	1/15
CHXE	48	807	21	786	120	666	42	1/19
ZEP	48	718	11	707	273	434	22	1/32
Total	335	6,675	846	5,829	2,084	3,745	265	–
Average	48	954	121	833	291	535	38	1/22

Table 4 Sequence diversity estimates of carotenoid biosynthesis genes in a sample of 48 cultivated carrots

Genes	θ_w	π_T	π_{sil}	π_{syn}	π_{nonsyn}	π_{nonsyn}/π_{syn}	h	Hd
IPI	0.0085	0.0082	0.0123	0.0292	0.0012	0.0407	11	0.851
PDS	0.0047	0.0038	0.0043	0	0	–	16	0.830
CRTISO	0.0198	0.0345	0.0440	0.0770	0.0014	0.0179	12	0.799
LCYB1	0.0059	0.0088	0.0297	0.0297	0.0023	0.0778	15	0.842
LCYE	0.0148	0.0246	0.0273	0.0430	0	0	10	0.711
CHXE	0.0120	0.0116	0.0131	0.0104	0	0	9	0.523
ZEP	0.0070	0.0087	0.0122	0.0165	0	0	10	0.699
Average	0.0104	0.0143	0.0204	0.0294	0.0007	0.0227	11.9	0.751

Total nucleotide polymorphism θ_w and nucleotide diversity π are expressed per site. π was calculated for all sites π_T, silent sites π_{sil}, synonymous sites π_{syn} and non-synonymous sites π_{nonsyn}. h is the number of haplotypes. *Hd* is haplotype diversity

clustering of individuals between two clusters was determined using BAPS from independent haplotype segments corresponding to seven carotenoid biosynthesis genes. Clustering obtained by this method was compared to assignations previously obtained from SSR genotypes. From the 26 individuals assigned to the Western group by SSR analysis, 18 (69%) were attributed to the same group by haplotype analysis. The second group obtained by haplotype analysis included 17 out of 19 individuals (89%) inferred to the Eastern group by SSR analysis. The geographical subdivision shown in SSR data is also revealed with gene sequences. The group of white carrots showed the highest difference of assignation depending upon the markers used (5 differentially assigned individuals). As the genes investigated are involved in carotenoid biosynthesis, it should be possible that the assignment differences according to marker types (i.e., microsatellites vs. gene haplotypes) result from selection in some colour types.

In order to determine if the geographical subdivision evidenced by putatively neutral microsatellite loci influenced the nucleotide polymorphism of each carotenoid biosynthesis pathway gene, three differentiation statistics testing for genetic structure were calculated at individual loci (Table 5). F_{ST} and G_{ST} were calculated respectively from nucleotide and haplotype diversity. The disadvantage of the first index is its independence to mutation rate, whereas the latter does not take into account the level of divergence between haplotypes. F_{ST} and G_{ST} indexes usually converged for the same gene, except for *IPI* ($F_{ST} = -0.0074$; $G_{ST} = 0.0242$) and *CHXE* ($F_{ST} = 0.1028$; $G_{ST} = 0.0118$). These results suggest that nucleotide diversity reflects geographic subdivision for *CHXE* but not for *IPI*. The effect of genetic differentiation by geographical origin was not significant for *LCYE*, *CHXE* and *PDS*, ($G_{ST} = -0.009$–0.022; $P > 0.05$). Therefore, we can hypothesise that *LCYE* and *PDS*, and maybe *CHXE* and *IPI,* may have contributed to differences of assignment observed between SSR and carotenoid biosynthesis gene analyses. Conversely, *CRTISO* and *ZEP* exhibited highly significant haplotype diversity ($G_{ST} = 0.067$–0.068; $P < 0.01$) and nucleotide diversity ($F_{ST} > 0.25$, while SSR loci gave $F_{ST} = 0.07$), showing strong differentiation between Western and Eastern carrot cultivars.

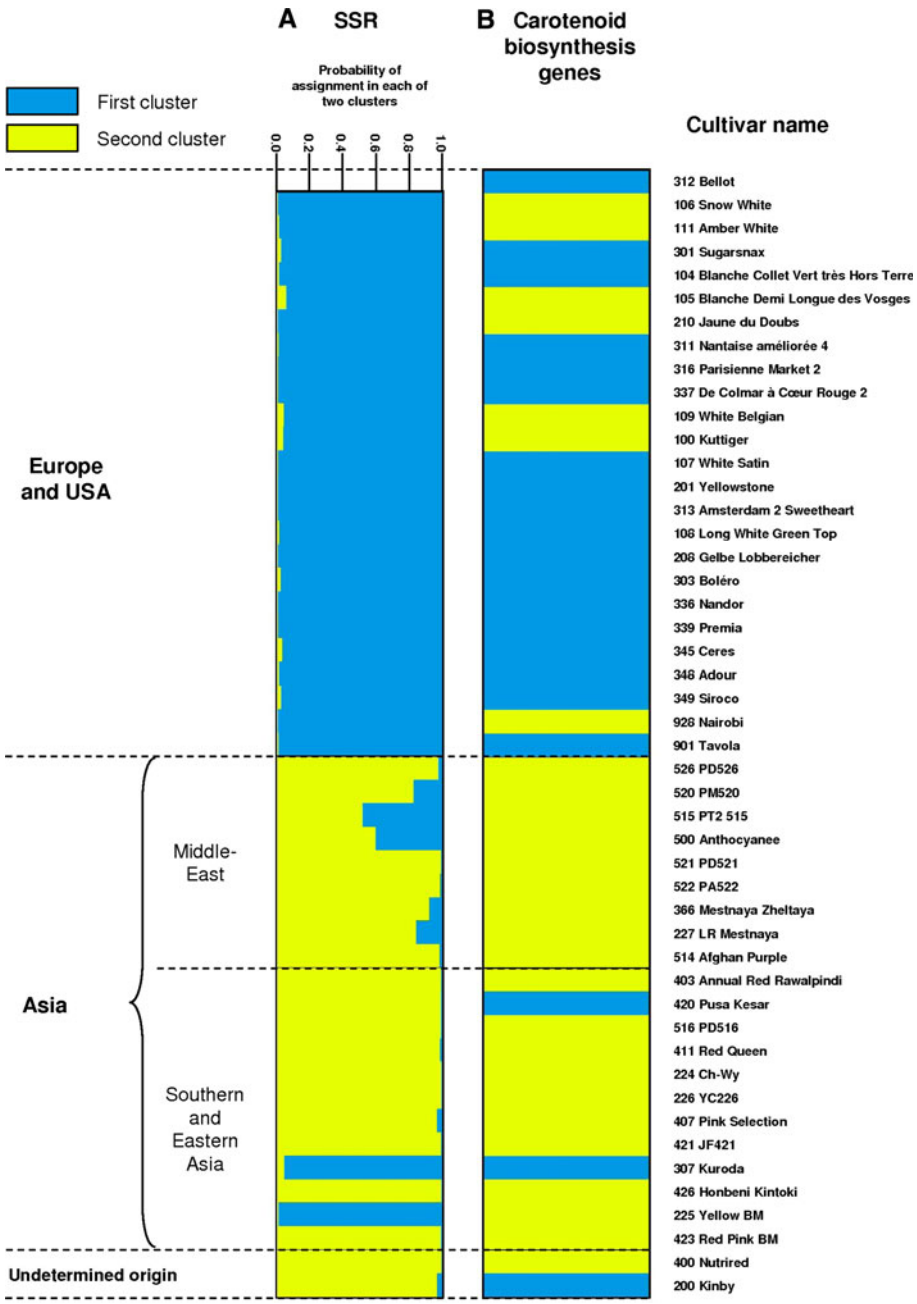

Fig. 1 Genetic structure in cultivated carrot germplasm. Assignment of carrot cultivars into two clusters were inferred by STRUCTURE from 17 SSR data (**a**) and by BAPS from haplotypes of seven carotenoid biosynthesis genes (**b**). STRUCTURE provided a measure of probability of assignment in each of two clusters, whereas BAPS did not. Assignation results were visualised with DISTRUCT (Rosenberg 2004)

Linkage disequilibrium

Linkage disequilibrium was investigated between pairwise segregating sites in order to predict the expected resolution and marker density needed for candidate-gene association mapping and to discuss global LD with reproductive biology characteristics of the carrot. The expected resolution and marker densities needed for a candidate-gene association mapping study depend of the decay of LD, i.e., the average distance until two base pairs are no longer in linkage disequilibrium. The decay of LD is shown by plots of r^2 as a function of physical distance between the SNPs (Fig. 2). Two models of non-linear regression were applied to data (Remington et al. 2001). These models did not fit the data better than a model without LD decay within the analysed portion ($P > 0.1$). We did not detect any LD decay within 700–1,000 bp analysed for each gene.

We evaluated LD levels among genes by averaging r^2 within each gene. The average r^2 ranged from 0.34 for *IPI* to 0.93 for *LCYE* with an average of 0.63 for the seven carotenoid biosynthesis genes. Three genes exhibited particularly high LD levels: *LCYE* (average $r^2 = 0.93$), *CHXE* (average $r^2 = 0.88$) and *CRTISO* (average $r^2 = 0.86$; Fig. 2). These three genes also exhibited the highest level

Table 5 Nucleotide (F_{ST}) and haplotype (G_{ST}) differentiation between the Western and Eastern groups

	F_{ST}	G_{ST}	S_{nn}
IPI	−0.0074	0.0242	0.6296*
PDS	0.0188	0.0216	0.5718 NS
CRTISO	0.3363	0.0668	0.6895***
LCYB1	0.0722	0.0245	0.6291**
LCYE	−0.0437	−0.0095	0.4974 NS
CHXE	0.1028	0.0118	0.4746 NS
ZEP	0.2661	0.0679	0.6727***

Significance of the nearest neighbour statistic (S_{nn}) was assessed using a 10,000-replicate permutation test

* $P < 0.05$

** $P < 0.01$

*** $P < 0.001$

of nucleotide polymorphism (θ_w; Table 4), respectively 0.0148, 0.0120 and 0.0198. Two genes showed medium LD levels: *LCYB1* (average $r^2 = 0.52$) and *ZEP* (average $r^2 = 0.51$). *IPI* (average $r^2 = 0.34$) and *PDS* (average $r^2 = 0.34$) exhibited the lowest LD levels (Fig. 2). LD matrix for each gene is given in Supplementary Fig. 1.

In addition to physical linkage, other factors such as demographic events can influence the extent of LD between a pair of SNPs. These demographic factors have a genome-wide influence and so their impact can be estimated by measuring the background LD between unlinked SNPs within the genome. We measured background LD levels by calculating r^2 between each pair of interlocus SNPs among genes (Fig. 3). Except for the *LCYE-ZEP* gene pair, the interquartile range for comparisons between genes from different or undetermined linkage groups (Just et al. 2007) was <0.1, although $r^2 = 0.1$ is often considered as being the threshold of LD persistence. Interlocus LD levels were highest for the *CHXE-PDS* comparison (median 0.1125; interquartile range 0.1125, 0.1600). This weak LD level could be related to map position, since *CHXE* and *PDS* are mapped to LG2 and separated by 45–50 cM (Just et al. 2007).

Discussion

A high level of diversity expected in an outcrossing species, but an unexpectedly high LD

This study brings the first characterisation of carrot germplasm genetic diversity using gene polymorphism and/or simple sequence repeat (SSR) markers. The sample investigated exhibits a high level of average silent nucleotide diversity ($\pi_{sil} = 0.020$) within the studied carotenoid biosynthesis genes. One explanation for this high degree of polymorphism is the predominately high outcrossing rate in carrot species, estimated at 95% (Le Clerc 2001). The level of sequence diversity is expected to be higher in outcrossing species than in selfing species, mainly because of larger effective population size (Pollak 1987; Glémin et al. 2006). Silent nucleotide diversity observed in the carrot is in the range of other outcrossing species such as maize ($\pi_{sil} = 0.013$; Tenaillon et al. 2004), *Populus tremula* ($\pi_{sil} = 0.016$; Ingvarsson 2005) or *Arabidopsis lyrata* ($\pi_{sil} = 0.023$) (Wright et al. 2003). It is higher than for most selfing species such as *Oryza sativa* ($\pi_{sil} = 0.003$; Caicedo et al. 2007) or *Arabidopsis thaliana* ($\pi_{sil} = 0.010$; Aguade 2001).

The mating system is also expected to strongly influence linkage disequilibrium (LD). The extent of LD is inversely proportional to the population recombination rate. Selfing tends to increase the level of loci with homozygous alleles, therefore involving a lower effective recombination rate in selfing than in outcrossing species. In outcrossing species, significant LD levels ($r^2 > 0.2$) are detected within 1,500 bp in maize (Remington et al. 2001), 500 bp in ryegrass (Xing et al. 2007), but only 200 bp in *Populus tremula* (Ingvarsson 2008). In selfing species, LD decays within 10–250 kb for *Arabidopsis thaliana* (Kim et al. 2007; Nordborg et al. 2002, 2005), 5–10 cM for barley (Kraakman et al. 2004; Malysheva-Otto et al. 2006) and 10 cM for durum wheat (Maccaferri et al. 2005). Therefore, we would expect a fast decay rate of LD in an outcrossing species like the carrot. However, we were unable to detect a decay of linkage disequilibrium within 700–1,000 bp for carotenoid biosynthesis genes in cultivated carrots. Besides the mating system, several factors are known to influence LD decay. Natural or artificial selection as well as demographic factors like population bottlenecks or population subdivisions could lead to increased LD levels. We could propose that selection directing toward these genes putatively involved in root colour could have increased LD levels for some genes investigated. No clear association between carotenoid biosynthesis gene polymorphisms and colour was shown until now, even if *CHXE*, exhibiting a large LD, or *PDS* and *CHXE*, exhibiting an intermediate LD, co-localise with QTLs for carotenoid content (Just et al. 2009). Linkage disequilibrium between SNPs within genes mapped to different linkage groups remained negligible, indicating that the genetic structure shown in the sample did not affect the background LD.

A moderate subdivision between European and Asian carrots: isolation by distance during species history

Two clusters were revealed both by carotenoid biosynthesis gene polymorphism and microsatellite marker analyses,

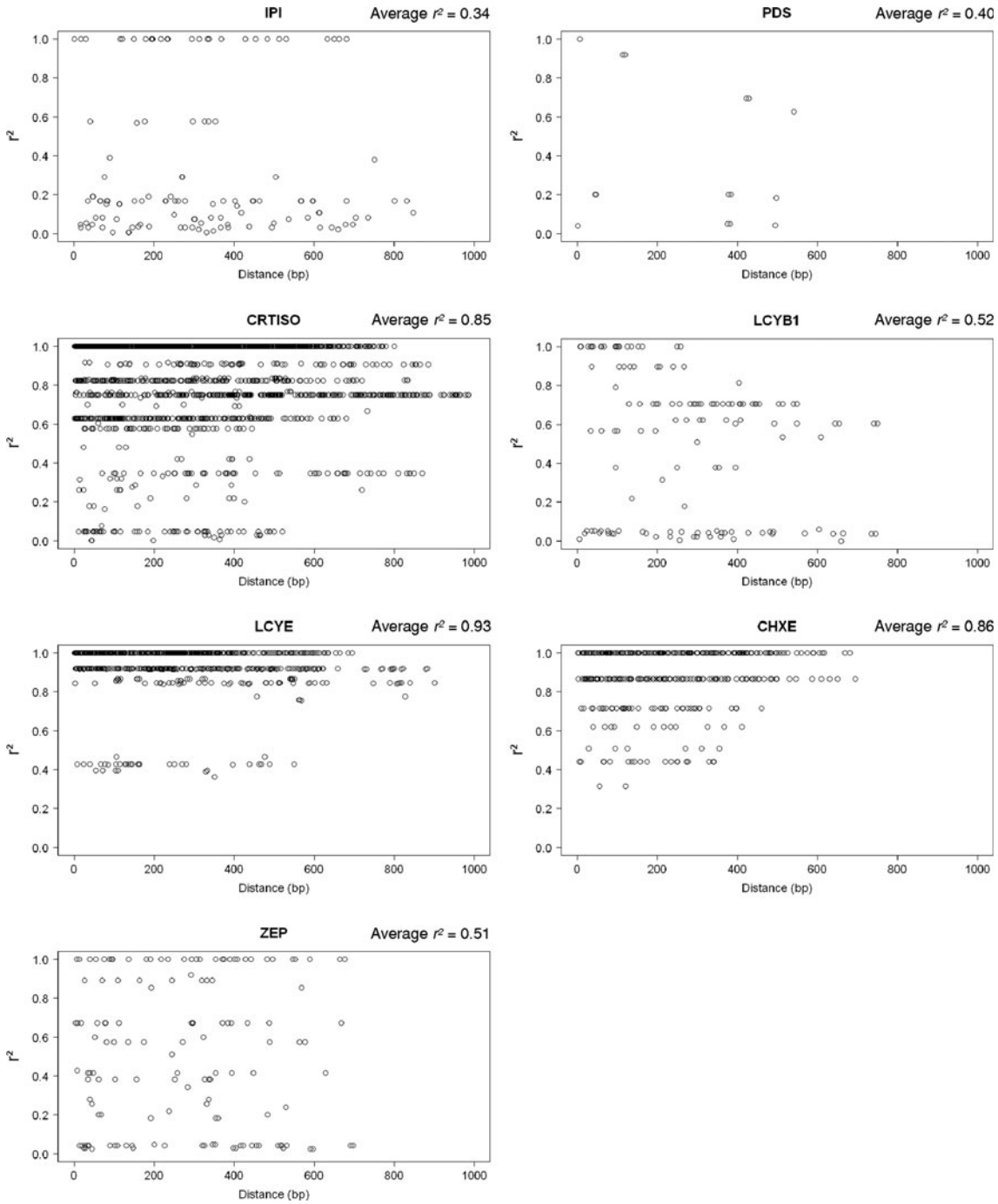

Fig. 2 Linkage disequilibrium (LD) of carotenoid biosynthesis genes. LD level was measured by squared correlations of allele frequencies (r^2) against distance between pairs of polymorphic sites within seven carotenoid biosynthesis genes

fitting almost perfectly their geographic origin: European-North American origin versus Asian origin. This result is coherent with current carrot history. It is supposed that the carrot was domesticated in Afghanistan before the 900s and was then introduced in Europe around the 1100s where it was highly bred (Banga 1963). North American carrot types derived from European cultivars (Simon 2000). Carrot breeding history in Asia is less documented. The Afghan carrot would have been introduced in China in the 1300s (Laufer 1919) and then in Japan in the 1600s (Shinohara 1984). In Japan, European types were adapted for cultivation. As an example, the most cultivated type

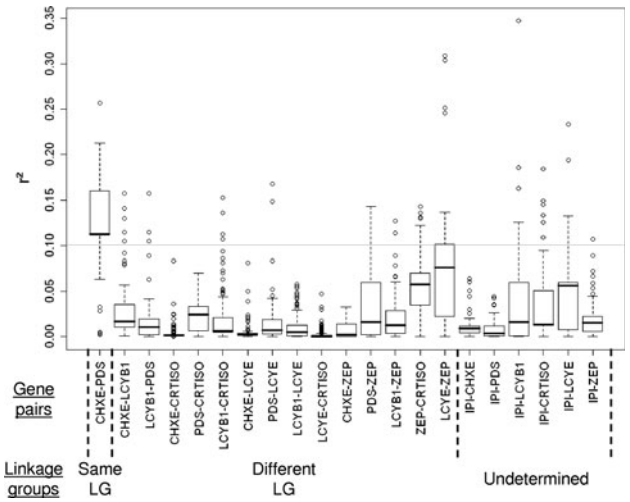

Fig. 3 Pairwise interloci linkage disequilibrium level between carotenoid biosynthesis genes. *Box* plot of squared correlations of allele frequencies (r^2) for interlocus pairwise comparisons of polymorphic sites was displayed. The box means the interquartile range. Mean is symbolised by the *thick line* in each box. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. *Open circles* represent outliers. Linkage groups are shown according to Just et al. (2007). The location of *IPI* in carrot linkage map is unknown

'Kuroda' was developed by breeding from the original European orange type 'Early Long Horn' (T. Sakai, personnel communication). This would explain the assignation of the individual sampled from the Kuroda cultivar, into the Western cluster in microsatellite analysis. Clustering methods used in this study rely on assignment of alleles to clusters. Alleles were sampled from cultivars according to their geographic origin. Therefore allelic distribution reflects population structure, generated during the carrot geographical spreading. Nevertheless these results must be considered with caution when extending the assignment obtained for single individuals sampled from cultivars to cultivars themselves. Thorough study of population structure in carrot will have to deal with the high intra-varietal diversity in this species (Shim and Jørgensen 2000). Although we chose to analyse a single haplotype for heterozygous sequences in order to avoid the possible effects of sequence overrepresentation depending on the level of heterozygosity at each locus, it is noticeable that we identified a similar population subdivision by gene polymorphism analyses compared to microsatellite analyses, despite the resulting loss of information and the decrease of power analysis.

Besides fitting geographical isolation during carrot history, population structure shown here is also coherent with a numerical taxonomy study which divided cultivated carrots into two botanical varieties based upon morphological differences between carrots from the West and the East (Small 1978). Three individuals originating from the Middle East evidenced high admixture between Western and Eastern clusters, showing the existence of putative intermediates. Except for these individuals, the assignment of investigated individuals to either of the two groups is clear-cut. It is noticeable that although intensive breeding exchanges since the nineteenth century, the background structure coming from demographic and early cultivation history still persists in cultivated carrot germplasm.

The use of Bayesian clustering methods reinforces the assignation of individuals to the two groups, since this method performs well for very low levels of population differentiation, and without a priori characterisation (Latch et al. 2006). None of the previous studies based on anonymous markers have revealed population structure among carrot cultivars, relating this fact to the outcrossing mating system, and to frequent crossings within carrot germplasm (Bradeen et al. 2002). This also could be due to small samples or insufficient representativity of worldwide material, especially from continental Asia. Our results show that population structure can be detected both by analysis of microsatellite loci and candidate gene polymorphism. Compared to other anonymous molecular markers, microsatellite loci are probably the powerful tools to characterise population structure, as they show a high polymorphism information content (PIC) due to their multiallelic and highly variable status (Powell et al. 1996). Biallelic and less variable SNPs could be less efficient to characterise population structure. Nevertheless, it was recently shown by coalescent simulations that SNP haplotypes provide a similar power to detect population structure than microsatellite markers (Haasl and Payseur 2010). Our study confirms that candidate gene haplotypes provide information about population structure, besides their interest for identifying the genetic basis of agronomic traits. The use of candidate genes haplotypes to characterise population structure and manage genetic resources should be reconsidered in the coming years.

In our study, as in most species (Malysheva-Otto et al. 2006; Schmid et al. 2006; Kwak and Gepts 2009), geographical origin of accessions represents a determining factor to explain population structure, to be included in sampling strategies when evaluating crop genetic resources. The genetic subdivision of carrot germplasm between Western and Eastern groups, found for the first time in this study, would require for carrot genetic resources managers and breeders to collect and preserve material originating from both differentiated genetic pools, especially when broadening the genetic base of breeding material or for the constitution of core collections, i.e., a subset of accessions from the entire collection that captures most of the genetic diversity of the species (Brown 1989).

Consequences for association studies

Besides providing useful guidance to genetic resource managers and breeders, the subdivision of carrot germplasm between two geographical groups should be considered carefully when designing association study experiments. Association studies will have to be wary of spurious associations between phenotype and markers that are not linked to any causative loci. Such associations occur when phenotypes for traits investigated are not distributed similarly among populations (Pritchard et al. 2000b). Searching for associations between carotenoid content and markers in carrots should be carried out with caution, especially for lycopene, present almost exclusively in Eastern red cultivars and α-carotene, present mainly in Western orange cultivars (Sun et al. 2009; Surles et al. 2004). Future experiments should take advantage of statistical methods (Devlin and Roeder 1999; Price et al. 2006; Pritchard et al. 2000b) which have recently been developed to account for population structure by using information from random molecular markers across the genome.

The extent of LD detected within genes is highly informative for the possible resolution of LD mapping. In maize, the low level of LD allowed the use of association mapping to efficiently localise SNPs that are significantly correlated to trait phenotypes within candidate genes (Harjes et al. 2008; Palaisa et al. 2003). On the other hand, more extensive LD levels make genome-wide association mapping suitable, as in *Arabidopsis thaliana* (Aranzana et al. 2005). Our study did not provide a decay of linkage disequilibrium within 700–1,000 bp for carotenoid biosynthesis genes. The extent to which LD decays remains to be determined. Genes with very extensive LD (*CHXE*, *LCYE*, *CRTISO*) could be considered as very valuable candidate genes for association mapping as selection is expected to improve LD by causing a selective sweep around advantageous sites (Palaisa et al. 2004, 2003). Such a hypothesis allows testing for association for a trait within a specific candidate-gene without investigating the full-gene sequence. However it may be difficult to identify the causative variations involved in trait determinism. *PDS*, *CHXE* and *LCYE*, did not show haplotype differentiation for geographic origin, contrary to microsatellite loci and other carotenoid biosynthesis genes. This could reflect diversifying selection (Beaumont 2005). We showed that *CRTISO* and *ZEP* exhibited the strongest differentiation for sequence diversity between Western and Eastern carrots. Therefore it would be valuable to explore the driving force leading to this pattern, especially carrot breeding.

The evidenced polymorphism within carotenoid genes and the genetic diversity structure, along with the extensive linkage disequilibrium along these genes, will help with the development of an adapted association mapping strategy regarding marker density and material sampling. From a species diversity point of view, our results, based on two types of markers and a wide range of material, provide the first evidence on a molecular basis of clear genetic structure within carrot germplasm, which is supported by both the demographic and cultivation history of the carrot. Further studies will assess intra-cultivar diversity. Finally our results show the need to consider these two geographical groups, to improve the overall management of genetic resources, and to maximise the efficiency of trait screening for plant breeding.

Acknowledgments We would like to thank Françoise Gros and Marie-France Le Cunff, of the *'Plate-forme de séquençage & de génotypage'* (IFR26, Nantes) for sequencing PCR products. We are grateful to Domenica Manicacci for providing script for LD decay analysis. We wish to thank Gérard Simon for his critical reading of this paper. This study was supported by grants from the *region Pays de la Loire*. This project is part of collaboration with Vilmorin SA, Clause Vegetable Seeds and Diana Naturals. Jérémy Clotault is a PhD student funded by the French Ministry of Research. *Ethical standards*: Authors declare that the experiments carried out comply with the current laws of the country in which they were performed, i.e., France.

Conflict of interest statement The authors declare that they have no conflict of interest.

References

Aguade M (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. Mol Biol Evol 18:1–9

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:e60

Bang H, Kim S, Leskovar D, King S (2007) Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene β-cyclase (*LCYB*) gene. Mol Breed 20:63–72

Banga O (1963) Main types of the western carotene carrot and their origin. W.E.J. Tjeenk Willink, Zwolle

Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? Trends Ecol Evol 20:435–440

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171. Université de Montpellier II, Montpellier (France)

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

Bradeen JM, Bach IC, Briard M, Le Clerc V, Grzebelus D, Senalik DA, Simon PW (2002) Molecular diversity analysis of cultivated carrot (*Daucus carota* L.) and wild *Daucus* populations reveals a genetically nonstructured composition. J Am Soc Hortic Sci 127:383–391

Briard M, Le Clerc V, Grzebelus D, Senalik D, Simon P (2000) Modified protocols for rapid carrot genomic DNA extraction and AFLP™ analysis using silver stain or radioisotopes. Plant Mol Biol Rep 18:235–241

Brown AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, Bustamante CD, Purugganan MD (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet 3:e163

Clotault J, Peltier D, Berruyer R, Thomas M, Briard M, Geoffriau E (2008) Expression of carotenoid biosynthesis genes during carrot root development. J Exp Bot 59:3563–3573

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 163:367–374

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Glémin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc R Soc B 273:3011–3019

Grzebelus D, Baranski R, Kotlinska T, Michalik B (2002) Assessment of genetic diversity in a carrot (*Daucus carota* L.) germplasm collection. Plant Genet Resour Newslett 130:51–53

Gupta P, Rustgi S, Kulwal P (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Haasl RJ, Payseur BA (2010) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity. doi:10.1038/hdy.2010.21

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. Science 319:330–333

Hill WG (1974) Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229–239

Hudson RR (2000) A new statistic for detecting genetic differentiation. Genetics 155:2011–2014

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589

Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). Genetics 169:945–953

Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. Genetics 180:329

Isaacson T, Ronen G, Zamir D, Hirschberg J (2002) Cloning of *tangerine* from tomato reveals a carotenoid isomerase essential for the production of *β*-carotene and xanthophylls in plants. Plant Cell 14:333–342

Just BJ, Santos CAF, Fonseca MEN, Boiteux LS, Oloizia BB, Simon PW (2007) Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping. Theor Appl Genet 114:693–704

Just BJ, Santos CAF, Yandell BS, Simon PW (2009) Major QTL for carrot color are positionally associated with carotenoid biosynthetic genes and interact epistatically in a domesticated × wild carrot cross. Theor Appl Genet 119:1155–1169

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 39:1151–1155

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

Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor Appl Genet 118:979–992

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conserv Genet 7:295–302

Laufer B (1919) The carrot. In: Sino-Iranica: Chinese contributions to the history of civilization in ancient Iran with special reference to the history of cultivated plants and products. Field Museum of Natural History, Chicago, pp 451–454

Le Clerc V (2001) Etude de la diversité génétique chez la carotte (*Daucus carota* L.): mise au point de stratégies d'analyse et de régénération des ressources génétiques. Ph.D. Thesis. Université d'Angers

Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data. Version 1.1 edn. University of Connecticut, Hartford

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–290

Mackevic VI (1929) The carrot of Afghanistan. Bull Appl Bot Genet Plant Br 20:517–562

Malysheva-Otto L, Ganal M, Roder M (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genet 7:6

Nakajima Y, Yamamoto T, Oeda K (1997) Genetic variation of mitochondrial and nuclear genomes in carrots revealed by random amplified polymorphic DNA (RAPD). Euphytica 95:259–267

Nakajima Y, Oeda K, Yamamoto T (1998) Characterization of genetic diversity of nuclear and mitochondrial genomes in *Daucus* varieties by RAPD and AFLP. Plant Cell Rep 17:848–853

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

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

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nicolle C, Simon G, Rock E, Amouroux P, Remesy C (2004) Genetic variability influences carotenoid, vitamin, phenolic, and mineral content in white, yellow, purple, orange, and dark-orange carrot cultivars. J Am Soc Hortic Sci 129:523–529

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim

S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J (2005) The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol 3:e196

Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15:1795–1806

Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. Proc Natl Acad Sci USA 101:9885–9890

Park H, Kreunen SS, Cuttriss AJ, DellaPenna D, Pogson BJ (2002) Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. Plant Cell 14:321–332

Pollak E (1987) On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics 117:353–360

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Price EW, Carbone I (2005) SNAP: workbench management tool for evolutionary population genetic analysis. Bioinformatics 21:402–404

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

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

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. Am J Hum Genet 67:170–181

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 20:103–111

Reimer S, Pozniak CJ, Clarke FR, Clarke JM, Somers DJ, Knox RE, Singh AK (2008) Association mapping of yellow pigment in an elite collection of durum wheat cultivars and breeding lines. Genome 51:1016–1025

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

Rosenberg NA (2004) *DISTRUCT*: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Santos CAF, Simon PW (2002) QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots. Mol Genet Genomics 268:122–129

Schmid K, Törjék O, Meyer R, Schmuths H, Hoffmann M, Altmann T (2006) Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. Theor Appl Genet 112:1104–1114

Shim SI, Jørgensen RB (2000) Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. Theor Appl Genet 101:227–233

Shinohara S (1984) Introduction and variety development in Japan. In: Vegetable seed production technology of Japan elucidated with respective variety development histories, particulars. Shinohara's Authorized Agricultural Consulting Engineer Office 4-7-7, Tokyo, pp 273–282

Simon PW (2000) Domestication, historical development, and modern breeding of carrot. Plant Breed Rev 19:157–190

Small E (1978) A numerical taxonomic analysis of the *Daucus carota* complex. Can J Bot 56:248–276

St. Pierre MD, Bayer RJ (1991) The impact of domestication on the genetic variability in the orange carrot, cultivated *Daucus carota* ssp. sativus and the genetic homogeneity of various cultivars. Theor Appl Genet 82:249–253

St. Pierre MD, Bayer RJ, Weis IM (1990) An isozyme-based assessment of the genetic variability within the *Daucus carota* complex (Apiaceae: Caucalideae). Can J Bot 68:2449–2457

Sun T, Simon PW, Tanumihardjo SA (2009) Antioxidant phytochemicals and antioxidant capacity of biofortified carrots (*Daucus carota* L.) of various colors. J Agric Food Chem 57:4142–4147

Surles RL, Weng N, Simon PW, Tanumihardjo SA (2004) Carotenoid profiles and consumer sensory evaluation of specialty carrots (*Daucus carota.* L.) of various colors. J Agric Food Chem 52:3417–3421

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol 21:1214–1225

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002) Biodiversity assessment using markers for ecologically important traits. Trends Ecol Evol 17:577–582

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

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

Wright SI, Lauga B, Charlesworth D (2003) Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. Mol Ecol 12:1247–1263

Xing Y, Frei U, Schejbel B, Asp T, Lübberstedt T (2007) Nucleotide diversity and linkage disequilibrium in 11 expressed resistance candidate genes in *Lolium perenne*. BMC Plant Biol 7:43