**PLOS ONE**  
Functional gene polymorphism to reveal species history: the case of the CRTISO gene in cultivated carrots  
--Manuscript Draft--

| Manuscript Number: |  |
|--------------------|---|
| **Article Type:**  | Research Article |
| **Full Title:**    | Functional gene polymorphism to reveal species history: the case of the CRTISO gene in cultivated carrots |
| **Short Title:**   | CRTISO gene reveals cultivated carrot history |
| **Corresponding Author:** | Emmanuel Geoffriau  
INRA AGROCOMPUS OUEST  
ANGERS, FRANCE |
| **Keywords:**      | Daucus carota L.;  root colour;  CRTISO;  breeding history;  balancing selection;  linkage disequilibrium |
| **Abstract:**      | Background: Carrot is a vegetable cultivated worldwide for the consumption of its root. Historical data indicate that root colour have been differentially selected over time and according to geographical areas. Root pigmentation depends on the relative proportion of different carotenoids for the white, yellow, orange and red types but only internally for the purple one. The genetic control for root carotenoid content might be partially associated with carotenoid biosynthetic genes. Carotenoid isomerase (CRTISO) has emerged as a regulatory step in the carotenoid biosynthesis pathway and could be a good candidate to show how a metabolic pathway gene reflects a species genetic history.  

Methodology/Principal Findings: In this study, the nucleotide polymorphism and the linkage disequilibrium among the complete CRTISO sequence, and the deviation from neutral expectation were analysed by considering population subdivision revealed with 17 microsatellite markers. A sample of 39 accessions, which represented different geographical origins and root colours, was used. This species was divided into two genetic groups: one from Middle East and Asia (Eastern group), and another one mainly from Europe (Western group). The Western and Eastern genetic groups were suggested to be differentially affected by selection: a signature of balancing selection was detected within the first group whereas the second one showed no selection. A focus on orange-rooted carrots revealed that cultivars cultivated in Asia were mainly assigned to the Western group but showed CRTISO haplotypes common to Eastern carrots.  

Conclusions: The carotenoid pathway CRTISO gene data proved to be complementary to neutral markers in order to bring critical insight in the cultivated carrot history. |
| **Order of Authors:** | Vanessa Soufflet-Freslon  
Matthieu Jourdan  
Jeremy Clotault  
Sebastien Huet  
Mathilde Briard  
Didier Peltier  
Emmanuel Geoffriau |
| **Suggested Reviewers:** | Elizabeth H Hemond  
Marine Science Center, Northeastern University, Nahant, Massachusetts, United States of America  
hemond.e@husky.neu.edu  
She has developed a similar approach on a different study model  
Norma B. Paniego |
Dr Paniego has studied SNP polymorphism and linkage disequilibrium in sunflower.
Functional gene polymorphism to reveal species history: the case of the CRTISO gene in cultivated carrots

Vanessa SOUFFLET-FRESLON*123, Matthieu JOURDAN*123, Jérémy CLOTAULT123, Sébastien HUET123, Mathilde BRIARD123, Didier PELTIER123, Emmanuel GEOFFRIAUV123

* These authors contributed equally to this work.

1 Agrocampus-Ouest, UMR1345 Institut de Recherche en Horticulture et Semences, SFR 4207 QUASAV, F-49045 Angers, France
2 Université d’Angers, UMR1345 Institut de Recherche en Horticulture et Semences, SFR 4207 QUASAV F-49045 Angers, France
3 INRA, UMR1345 Institut de Recherche en Horticulture et Semences, SFR 4207 QUASAV F-49071 Beaucouzé, France

Corresponding author: Emmanuel Geoffriau (emmanuel.geoffriau@agrocampus-ouest.fr, TEL 33.2.41.22.54.31, FAX 33.2.41.22.55.15)
Abstract

**Background:** Carrot is a vegetable cultivated worldwide for the consumption of its root. Historical data indicate that root colour have been differentially selected over time and according to geographical areas. Root pigmentation depends on the relative proportion of different carotenoids for the white, yellow, orange and red types but only internally for the purple one. The genetic control for root carotenoid content might be partially associated with carotenoid biosynthetic genes. Carotenoid isomerase (*CRTISO*) has emerged as a regulatory step in the carotenoid biosynthesis pathway and could be a good candidate to show how a metabolic pathway gene reflects a species genetic history.

**Methodology/Principal Findings:** In this study, the nucleotide polymorphism and the linkage disequilibrium among the complete *CRTISO* sequence, and the deviation from neutral expectation were analysed by considering population subdivision revealed with 17 microsatellite markers. A sample of 39 accessions, which represented different geographical origins and root colours, was used. This species was divided into two genetic groups: one from Middle East and Asia (Eastern group), and another one mainly from Europe (Western group). The Western and Eastern genetic groups were suggested to be differentially affected by selection: a signature of balancing selection was detected within the first group whereas the second one showed no selection. A focus on orange-rooted carrots revealed that cultivars cultivated in Asia were mainly assigned to the Western group but showed *CRTISO* haplotypes common to Eastern carrots.

**Conclusions:** The carotenoid pathway *CRTISO* gene data proved to be complementary to neutral markers in order to bring critical insight in the cultivated carrot history.
Introduction

Selection events along the species or breeding history generally lead to signatures of selection at the molecular level, i.e. local variations of diversity or allele frequencies at genes underlying phenotypic variations or in their surrounding region. The detection of signatures of selection encounters various issues. Firstly, demographic events like population subdivision can modify genetic polymorphism and mimic selective events [1]. Secondly, selection pattern can vary along a gene [2, 3]. Signatures of selection are therefore detected or not, according to the targeted genomic region and the linkage disequilibrium (LD; i.e., the non-random association of alleles at different loci) throughout the gene.

The history of carrot (*Daucus carota* subsp. *sativus*), cultivated worldwide, might have lead to such selection signatures. The domestication of this species is thought to have occurred in the Afghanistan region before the 900s [4]. The first cultivated carrots were purple or yellow rooted. Their cultivation spread along trade routes, reaching the Middle East and North Africa, and then to Europe in the Middle Ages. They were gradually replaced by white- and orange-rooted forms, which appeared in the 1600s [5]. Concomitantly, a red type appeared in Asia, and particularly in Japan in the 1700s [6,7]. Finally, since the nineteenth century, orange-rooted carrots have spread from Europe to other continents and have become predominant commercially. Diversity analyses were recently assessed in cultivated carrot and revealed a genetic subdivision between Western (European and American) and Eastern (Asian) accessions [8,9].

Root pigmentation depends on the relative proportion of different carotenoids, in combination with anthocyanins for the purple type [10]. Orange- and red-rooted carrots accumulate large amounts of carotenoids, mainly β- and α-carotene for the orange type or lycopene and β-carotene for the red type. Yellow-rooted carrots present low amounts of carotenoids, especially lutein and β-carotene. White-rooted carrots contain negligible amounts of carotenoids. The genetic control for root carotenoid
content can be associated with carotenoid biosynthetic genes in different species including carrot [11,12,13]. These data suggest that the carotenoid biosynthetic genes may have been targeted by human selection.

The carotenoid biosynthetic pathway has been established in higher plants [14,15]. This pathway involves a series of desaturations, cyclisations, hydroxylations and epoxidations. Most of the carotenoid genes were identified in various species [14, 15, 16, 17, 18] including carrot [19]. Among genes, the carotenoid isomerase (CRTISO) has emerged as a regulatory step in the carotenoid biosynthesis pathway. CRTISO enzyme catalyses the cis-to-trans isomerisation reactions leading to all trans-lycopene, the substrate for the subsequent lycopene cyclisation to form all trans-α/β-carotene [20]. The identification of CRTISO as an isomerase was also confirmed by its functional expression in Escherichia coli in which it was able to convert cis-carotenoids to all trans-carotenoids [17,18]. CRTISO mutants, such as ccr2 (Arabidopsis), phs3 (rice) and tangerine (tomato), result in accumulation of cis-carotenoids in the dark-grown plants [17, 21] and fruits [18].

During plant evolution, paralogous genes for several carotenoid biosynthesis enzymes have been conserved and subjected to subfunctionalization: the paralogs are differentially expressed in photosynthetic or non-photosynthetic tissues [22]. A mutation affecting the paralog expressed in non-photosynthetic tissues would therefore not affect the carotenogenesis needed for photosynthesis. In most species in which carotenoid isomerase was characterized (except maize, [23]), only one gene was found (tomato [17], carrot [18], rice [24]). One hypothesis to this tendency of CRTISO to be single-copy is that CRTISO activity could be partially redundant in the light, because of photoisomerisation. Indeed, light could substitute for the lack of isomerase activity in ccr2 mutants which then synthesize efficiently most carotenoids [18]. A similar light-induced isomerisation of prolycopene to all trans-lycopene was observed in the outer green tissues in immature fruit of tangerine mutants that
were exposed to light but not in non-photosynthetic tissues of flowers (petals), ripe fruit, and the innermost parts of the green fruit [17]. Consequently, the function of CRTISO in plants presumably is to enable carotenoid biosynthesis to occur in the dark and in non-photosynthetic tissues, such as root like carrot. By comparison to other crucial steps in the carotenoid pathway, CRTISO may have evolved with fewer constraints due to photoisomerisation, and may have been more prone to be subjected to artificial selection in non-photosynthetic organs, like root. Clotault et al. [25] showed that CRTISO gene has undergone through selection events in cultivated carrot but the polymorphism pattern was observed among partial CRTISO sequence (only 700-1000 pb). The particular status of this gene and preliminary results suggest that CRTISO gene could be a good candidate for selection signature research. The analysis of the nucleotide polymorphism and the LD among the complete CRTISO sequence will enable to clarify the selection pattern, depending on the gene structure and in relation with colour types. Moreover, selection effect can spread over genomic regions around the gene by a selective sweep. The study of the intrachromosomal LD will allow testing this hypothesis.

Our work aimed to know whether polymorphism within CRTISO, a key gene involved in the carotenoid biosynthesis pathway especially in root organs, should reflect the cultivated carrot history, including selective events during breeding. Actually, the present results showed a signature of selection pattern at the CRTISO gene and the critical contribution of CRTISO polymorphism for explaining the cultivated carrot history, especially regarding material from Asia and Central Asia.

**Materials and methods**

**Plant material**

Thirty nine cultivars of carrot (*Daucus carota* subsp. *sativus*) were sampled (Table 1); each one was represented by a single individual.
These cultivars were chosen to maximize the diversity according to geographical origin, root colour and shape. Seeds were obtained from several seed banks and breeding companies. Genomic DNA was extracted from 50 mg of young freeze-dried leaves using a modified CTAB protocol [26].

**Microsatellite genotyping**

In order to study genetic structure of the sample, 17 microsatellite primer pairs were chosen regarding their reproducibility and coverage of the carrot genome [27]. Five microsatellite markers were located around *CRTISO* gene to investigate the linkage disequilibrium on the linkage group 4 (Table S1).

PCR reactions were performed in 20 µL volume with 10-50 ng genomic DNA, PCR buffer 1X, 0.2 mM dNTPs, 2 mM MgCl₂, 0.25 µM of fluorescent dye-labelled forward primer, 0.25 µM of reverse primer, and 1U Taq DNA polymerase (Interchim, Montluçon, France). Amplifications were carried out using a MyCycler thermalcycler (Biorad, Hercules, CA). The thermalcycler was programmed as follows: initial denaturation at 94°C for 2 min, 35 cycles at 94°C for 30 s, annealing temperature for 30 s, 72°C for 30 s, and a final extension at 72°C for 10 min.

The amplified fragments were analyzed by capillary electrophoresis (ABI 3730 DNA Analyzer, Applied Biosystems, Foster City, CA), and labelled PCR products were automatically sized with Genemapper 3.7 software (Applied Biosystems, Foster City, CA).

**Genetic structure**

Microsatellite data were used to investigate the genetic structure of the sample with STRUCTURE 2.1 software [28]. The genetic model was used without information on the origin of each individual and allowing for admixture and allele correlated frequencies. Ten independent simulations, with a burn-in period length of 10⁵ and a run length of 10⁶, were used for each number of clusters (K) from one to ten. The most likely number of genetic clusters was selected using Evanno’s method [29]. For each
individual, the proportion $q$ of its genome assigned to each cluster and its 95% confidence interval were calculated. An individual was assigned to a cluster if $q > 0.5$. Fixation indexes were estimated among the obtained genetic clusters ($F_{ST}$) and colour groups ($F_{CT}$), by using FSTAT 2.9 software [30] (Goudet 1995) according to Weir and Cockerham formula [31].

**Allele sequencing**

Gene-specific primers were designed and synthesized (Figure 1), based on the mRNA sequence of *Daucus carota* subsp. *sativus* (GenBank accession number DQ192188) in order to obtain the complete *CRTISO* sequence (exons and introns). PCR reactions were performed in 25 µL volume with 10-50 ng genomic DNA, 1X PCR buffer, 2 mM MgCl$_2$, 0.2 mM dNTPs, 0.2 µM of both forward and reverse primers, and 1U Taq DNA polymerase (Interchim, Montluçon, France). Amplifications were carried out using a MyCycler thermalcycler (Biorad, Hercules, CA). The cycling conditions were: initial denaturation at 94°C for 3 min, 35 cycles at 94°C for 30 s, annealing at 55°C for 45 s, 72°C for 2 min, and a final extension at 72°C for 5 min.

PCR products were purified with ExoSAP-IT (GE Healthcare, France) and sequenced with BigDye® Terminator 3.1 chemistry (Applied Biosystems, Weiterstadt, Germany) in an ABI Prism 3730 sequencer (Applied Biosystems, Weiterstadt, Germany) according the supplier’s instructions. The primers used to amplify PCR fragments were also used for the sequencing. Eight partial genomic regions were sequenced with forward and reverse primers to discard PCR errors (Table 2). Nucleotide sequences were read with Sequence Scanner 1.0 software (Applied Biosystems, Weiterstadt, Germany) and the credibility between the two reads was checked with Geneious 5.0.2 software [32]. For each individual, the eight partially overlapping sequences were assembled in a contiguous sequence with a program compiled in C++ [33].
The contiguous sequences of the 39 individuals were aligned with ClustalW [34]. For heterozygous individuals, allelic phases were reconstructed by using the algorithm provided by PHASE implemented in DnaSP 5.10 software [35].

**Sequence polymorphism**

The population genetic parameters were defined by using DnaSP 5.10 software [35]. Sites with alignment gaps were excluded from analysis. The population-level genetic variation was estimated as nucleotide polymorphism ($\theta_w$; [36]) and nucleotide diversity ($\pi$;[37]). $\theta_w$ is based on the number of segregating sites while $\pi$ is based on the pairwise differences between sequences in the population. The analyses of nucleotide diversity ($\pi_T$) and nucleotide polymorphism ($\theta_w$) were based both upon the overall CRTISO sequence and upon a 100 bp window with a step size of 25 bp.

Haplotypes were inferred using DnaSP 5.10 software by including infinite-site violations. The number of haplotypes $h$ and haplotype diversity $Hd$ were then calculated [37]. A median-joining network establishing relationships among haplotypes based on the number of polymorphisms (each insertion/deletion -indel- was considered as a single nucleotide polymorphism -SNP-) was performed with Network 4.61 software [38] (Bandelt et al. 1999).

**Linkage disequilibrium**

Linkage disequilibrium (LD) is the non-random association of alleles at different loci. The level of LD is influenced by many factors such as population subdivision, selection, genetic linkage, recombination rate [39]. We chose to measure LD using $r^2$ because it summarizes both recombination and mutation histories [40].

LD was investigated by using TASSEL 3.0 software [41] not only along complete CRTISO sequence but also along the linkage group 4 [27] where CRTISO gene and five microsatellite markers were mapped. Since rare alleles can get a large variance in LD estimates, only biallelic loci with at
least 5% frequency were used to calculate $r^2$, which was plotted for all pairwise comparisons among SNPs. Given that population subdivision influences LD, this one was calculated in each genetic group revealed by population structure analysis.

**Neutrality tests**

Tajima’s $D$ test [42], implemented in DnaSP 5.10 software [35], was used to estimate the allele frequency departure from neutral expectation, by considering the dataset as a whole, the genetic and colour groups separately. Positive and negative $D$ values reveal an excess of intermediate and low frequency variants, respectively. Tajima’s $D$ test was calculated for the overall CRTISO sequence and upon a 100 bp window with a step size of 25 bp. The statistical significance for the above tests was inferred provided that the observed value was included within the 95% confidence interval of neutral distribution, which was calculated using 10,000 coalescent simulations in DnaSP by assuming no recombination.

**Results**

**Population structure**

All 17 microsatellite loci were polymorphic. A total of 174 alleles were identified with a mean of 10.2 alleles per locus, and a range from two to 19 alleles per locus. The Bayesian model-based clustering approach implemented by STRUCTURE showed that the sample studied was most probably divided into two clusters (Figure 2) according to Evanno’s method [29]. Among the 39 individuals studied, 34 were assigned to a cluster with a probability higher than 0.90 and five with a lower probability from 0.68 to 0.82. The first cluster contained 19 individuals, including 12 from Europe, three from Central Asia and four from Japan. These individuals showed mainly orange, white or yellow roots. In accordance with the distinction from [8] and [9], this cluster was considered as the Western genetic group. However, the assignment probability in this cluster
was the lowest for three orange rooted cultivars from Japan, along with high confidence interval, which shows an intermediate status of these cultivars. In contrary, two out of three central Asia orange rooted cultivars were significantly assigned to this first cluster. The second cluster contained 20 individuals, only sampled in Middle East or Asia, except the Swiss cultivar ‘Kuttinger’ (code 100). Most of these individuals showed red, purple or yellow roots. This second cluster was considered as the Eastern genetic group.

The divergence between the Western and Eastern genetic groups was moderate but significant ($F_{ST} = 0.115$, $P < 0.05$), reinforcing the assignation obtained by Bayesian clustering approach. Fixation indexes among colour groups ($F_{CT}$) ranged from 0.003 to 0.142 (Table 3).

**Genomic structure of the carotenoid isomerase in carrot**

The sequencing of CRTISO gene resulted in an alignment of 4,234 bp for 39 carrot diploid individuals.

The CRTISO gene in carrot exhibits the same genomic structure than other species (Arabidopsis, tomato, rice): these genes are split into 13 exons and 12 introns [17, 18, 24] (Figure 1). The coding and non-coding regions contributed respectively to 45 and 55% of the analysed sequences (excluding indels).

**Linkage disequilibrium**

The complete CRTISO sequence exhibited a high LD level with an average $r^2$ of 0.61 for the Western genetic group and 0.64 for the Eastern genetic group. LD showed no decay within 4,234 bp (Figure 3), when tested with the method described in [43].

LD was also evaluated along LG4 [12] by using five microsatellite loci flanking CRTISO gene. Low LD level was observed between markers and CRTISO gene (Figure 4). Indeed, $r^2$ ranged from 0 to 0.24 for both Western and Eastern genetic groups. Only GSSR6 and GSSR96, separated
by 11.5 cM, were in linkage disequilibrium ($r^2 = 0.47$) in the Western genetic group.

**Nucleotide diversity and neutrality test**

The observed nucleotide variation of the *CRTISO* genomic sequence, among the whole dataset, the Western versus Eastern genetic group, and the colour groups, is summarized in table 4.

*Within the whole dataset*

Among the 78 allelic sequences studied, 15 haplotypes were identified, of which five were found only once (singleton haplotypes). Haplotypes differed by 212 polymorphisms, consisting of 196 nucleotide substitutions and 16 indels. Among the 55 substitutions found in coding regions, 18 were non-synonymous. The overall nucleotide diversity was moderate ($\pi = 0.01753$). The synonymous ($\pi_{syn} = 0.02788$) diversity value was much higher than the non-synonymous one ($\pi_{nonsyn} = 0.00354$), yielding a $\pi_{nonsyn}/\pi_{syn}$ ratio of 0.12697. The nucleotide diversity was higher within non-coding regions ($\pi_{non-coding} = 0.02433$) than within coding regions ($\pi_{coding} = 0.00928$) (Table 4). High diversity values were observed within most of the introns, and particularly within intron 1 of *CRTISO* sequence (Figure 5a). Tajima’s $D$ statistic detected an excess of intermediate-frequency variants in the whole dataset, which yielded a significant positive $D$ value ($D = 2.76, P < 0.01$). High and significant $D$ values were also obtained within non-coding regions (Table 5, Figure 5a).

*Within and between groups*

Only the Western genetic group showed a similar pattern to the whole dataset for nucleotide diversity indexes and neutrality test (Tables 4 and 5; Figures 5a and 5b). The high nucleotide diversity observed within this genetic group was consistent with a balancing selection revealed by a positive and significant $D$ value.
Otherwise, the Eastern genetic group showed a low nucleotide diversity (Table 4, Figure 5c) and a non significant negative Tajima’s $D$ (Table 5).

Among the colour groups, the overall nucleotide diversity ($n_T$) varied from 0.01105 to 0.02071, with an average value of 0.01614. The white-, orange- and yellow-rooted carrots showed globally higher nucleotide diversity than the red and purple ones (Table 4). These results were consistent with the detected selection pattern. Indeed, the yellow and orange presented a positive $D$ value for both coding and non-coding regions revealing a balancing selection, whereas the white type presented a significant one only for the non-coding region. Notice that the red and purple types showed a trend of negative Tajima’s $D$ values even if they did not reach statistical significance (Table 5).

**Relationships among haplotypes**

The haplotype network revealed four main clusters (Figure 6a). The cluster 1 was the most distant one with 137 substitutions from the closest cluster. Based on haplotype 2 only, it contained mainly individuals belonging to the Western genetic group with white, yellow or orange roots (Figure 6b). The cluster 2 corresponded to eleven haplotypes and cultivars originating all from Middle-East and Asia: yellow-, red- and purple-rooted carrots (and one white) belonging to the Eastern genetic group, and orange ones belonging predominantly to the Western genetic group but cultivated in Japan or Central Asia. In this cluster, all the orange Japanese material was separated from other Asiatic cultivars (haplotypes 9 to 12 only).

Finally, the clusters 3 (haplotype 1) and 4 (haplotypes 3 and 15) corresponded only to one or two individuals, with white or orange roots. These clusters are closer to the cluster 2, even if it contains 3 out of 4 cultivars from the western genetic group. It could constitute an intermediate cluster.
Discussion

Variation of polymorphism along the complete gene sequence

In the present study, both results from microsatellite and CRTISO gene showed that cultivated carrot was divided into two genetic groups: one from Asia, and another one mainly from Europe. This subdivision is consistent with previous studies about microsatellite markers or genes [8, 9] and with the historical mention of two independent migration events from the domestication centre in Afghanistan: the first one to West in the 1200s, and the second one to East in the 1400s [5,6]. The haplotype network of CRTISO is especially marked by the genetic subdivision between the Western and the Eastern genetic groups: the three clusters 2, 3, 4 are composed mainly of Western group individuals, while the fourth cluster (1) is mainly made by Eastern individuals.

No decay of linkage disequilibrium was detected within 4,234 bp for CRTISO gene. This was consistent with results obtained by [8] which did not detect any decay of LD within 700-1,000 bp for some carotenoid biosynthesis genes in cultivated carrot. This result is unexpected for an allogamous species like carrot [44]. Indeed LD is known to decay quicker in outcrossing species than in selfing species [39]. For example, LD decays within 1,500 bp in maize [43], 200 bp in Populus tremula [45] or 500 bp in ryegrass [46]. Nevertheless, different factors could explain the observed high LD level such as low recombination rate, selection, and demographic events like population bottlenecks or population subdivisions.

By considering population differentiation obtained here, LD remained high in the Western and Eastern genetic groups, suggesting the population subdivision could not explain the absence of LD decay. However, CRTISO gene was mapped on the linkage group 4 near the centromeric region [27], and some studies revealed a high LD level in centromeric regions due to low recombination rate for several species (maize [43], human [47], [48]). Finally, carrot is a recently domesticated species: this has
resulted in a few recombination events in this biennial species, following a potential domestication bottleneck. Despite the absence of LD decay in \textit{CRTISO} gene, the Tajima’s \( D \) value varied a lot along the sequence, varying from significant to non significant values, especially in the whole dataset and the Western genetic group. The region analyzed in [8] and [25] was among the significant results. Even if in this case, the partial sequence lead to a similar pattern than the complete sequence, caution is needed when conducting candidate-genes selection analysis on partial gene sequence.

\textbf{The effect of selection and demography at CRTISO}

The possibility of photoisomerisation in leaves made us hypothesize that \textit{CRTISO} gene may have evolved with less constraints and may have been more prone to artificial selection, for example for root colour. Indeed, the analysis of the complete \textit{CRTISO} sequence demonstrated a significant positive Tajima’s \( D \) value, which indicates a departure from the neutral expectation. This significant positive value was observed for the whole dataset and the Western genetic group whereas the Eastern genetic group exhibited a neutral evolution pattern. Therefore it appears that \textit{CRTISO} evolved differentially in the Western and Eastern genetic groups. Indeed, these results suggest balancing selection as the force governing \textit{CRTISO} evolution in the Western genetic group, and are congruent with the high linkage disequilibrium [49] and the high silent-site nucleotide diversity detected in this genetic group. Furthermore, the results obtained in our study are in agreement with the previous results [8,25] based on the analysis of a partial \textit{CRTISO} sequence corresponding mainly to the intron 1. Otherwise, the Eastern genetic group showed no selection with a low nucleotide diversity.

One explanation of the potential balancing selection observed in the Western genetic group would be selection for some colour types. A signature of balancing selection was found at \textit{CRTISO} gene within yellow and orange carrots \((D = 3.08636, P < 0.001; D = 2.18514, P < 0.05,\)
respectively). Several models of selection maintaining diversity have been suggested: heterozygote advantage at a locus, frequency-dependent selection, temporally or spatially heterogeneous selection [49]. The latter hypothesis is the most plausible for carrot colour selection. Carrot has been bred for new colours between the domestication of this species and the eighteenth century and these selective events occurred in the centre of origin (Afghanistan region), in Western Europe and in Eastern Asia. Therefore, the breeding objectives varied temporally and spatially for carrot root colour.

However, caution is needed while interpreting the CRTISO polymorphism in yellow and orange carrots as evidence for signature of selection. Indeed, the differentiation between colour groups is considerably influenced by the geographical subdivision between Western and Eastern genetic groups. In microsatellite data, most of the variation between the colour groups ($F_{CT}$) is explained by the differences among genetic backgrounds, Western versus Eastern ($F_{ST}$). Orange and white carrots are mostly in the Western genetic group, 10 among 12 and four among six, respectively. Otherwise, all red and purple carrots (except the cultivar ‘Honbeni Kintoki’, code 426 unexpectedly) were assigned to the Eastern genetic group. The yellow group is half constituted by Western and Eastern individuals concordantly for microsatellite and CRTISO data, while the orange group is composed of individuals mainly assigned to the Western genetic group, with some individuals cultivated in Asia showing CRTISO haplotypes common to Eastern carrots. Therefore, these two colour groups are probably the most made of the two genetic groups. The respective effect of population subdivision, which could mimic balancing selection, and real balancing selection in these colour types should be investigated in the future.

Selection footprint and high LD extent detected among CRTISO gene might lead to a selective sweep around this gene [50]. The low intrachromosomic LD observed in our study could not confirm this hypothesis. New markers in CRTISO genomic region could be developed
thanks to carrot genome sequencing, and therefore could allow improving the study of selection events in this species.

**Insights into the history of cultivated carrots**
Neutral loci and *CRTISO* polymorphism analyses allowed testing the hypotheses about the cultivated carrot history. Microsatellite markers gave information about the entire genome and are therefore very suitable to understand gene pool origin. The incongruity between the *CRTISO* haplotype network and structure given by microsatellite analysis informed about genetic admixture. Literature considers the purple and yellow carrots as the two original colour types [4,5]. The purple type only found in Middle-East is clearly assigned to the Eastern group, both from microsatellite and *CRTISO* sequence data. By comparison, the clustering of yellow cultivars half into Western and half into Eastern groups, whether through microsatellite or *CRTISO* sequence analyses, suggest that the yellow type has experienced two separate migration events, following domestication. According to their assignment probability and large confidence interval estimated by STRUCTURE analysis, orange carrots sampled from Japan and Central Asia were assigned to the Western genetic group but with some admixture between Western and Eastern carrots whereas they belong clearly to the Eastern group depending on *CRTISO* data. This is congruent with the breeding of this colour type. Indeed some European cultivars were exported and adapted for cultivation to Japan in the 1900s. They were used in breeding programs and crossed with Asian carrots [51]. In the same way, carrots from Central Asia might come from European cultivars that were either exported and cultivated straight in Central Asia, or went by Japan before being used in Central Asia. Two out of three were significantly assigned to the Western genetic group based on microsatellite, but to the Asiatic cluster and an intermediate one (haplotype 15) based on *CRTISO* data. It remains therefore unclear if
Central Asia cultivars come from European material crossed locally with eastern material, or they originate from an eastern migration from Japan.

**Conclusion**

Neutral loci and CRTISO polymorphism analyses were shown to be complementary in order to recount the breeding history of cultivated carrot. Indeed human activities (seed exchange and transport, and breeding) could affect allele diversity, which resulted in the distinction of two genetic groups and large diversity at the CRTISO gene. These results may suggest that human preferences for carrot root colour have varied depending on periods and geographical areas. It brings new insights about the history of orange carrots. The orange-rooted carrots spread from Europe to other continents including Asia. Therefore the breeding of this carrot form was adapted to different markets. In spite of nucleotide polymorphism and the distinction of several haplotype clusters within the orange carrots, corresponding individuals belong mainly to the Western genetic group. This would confirm that the introduction of European carrots in Asian breeding programs is relatively recent, as genetic background remains common to both plant materials. The knowledge of the breeding history of this species, based both on neutral and functional loci, allows a better characterisation of plant material, and offers substantial insight to secure, manage and exploit carrot genetic resources while maximising genetic diversity.

**Acknowledgments**

We would like to thank Charles Poncet, of the ‘Plateforme GENTYANE’ (UMR1095 GDEC, Clermont-Ferrand, France), for microsatellite genotyping. We wish to thank Françoise Gros and Marie-France Le Cunff, of the ‘Plateforme de Séquençage et de génotypage’ (IFR26, Nantes,
France), for sequencing PCR products. We thank Sylvain Gaillard for the sequence assembly program.

References

1. Hudson R (1990) Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology 7:1–44

2. Camus-Kulandaivelu L, Chevin L, Tollon-Cordet C et al (2008) Patterns of molecular evolution associated with two selective sweeps in the Tb1–Dwarf8 region in maize. Genetics 180(2):1107-1121

3. Fu Z, Yan J, Zheng Y et al (2010) Nucleotide diversity and molecular evolution of the PSY1 gene in Zea mays compared to some other grass species. Theor Appl Genet 120:709–720

4. Mackevic V (1929) The carrot of Afghanistan. Bull Appl Bot Genet Plant Breed 20:517–562

5. Banga O (1963) Origin and distribution of the Western cultivated carrot. Genetica Agraria 17:357-370

6. Laufer B (1919) 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 451–454

7. Shinohara S (1984) Introduction and variety development in Japan. In: 4-7-7SAACEO (ed) Vegetable seed production technology of Japan elucidated with respective variety development histories, particulars. pp 273–282

8. Clotault J, Geoffriau E, Lionneton E et al (2010) Carotenoid biosynthesis genes provide evidence of geographical subdivision and extensive linkage disequilibrium in the carrot. Theor Appl Genet 121:659–72

9. Baranski R, Makselewicz-Kaul A, Nothnagel T et al (2011) Genetic diversity of carrot (Daucus carota L.) cultivars revealed by analysis of SSR loci. Genet Resour Crop Evol 59:163–170

10. 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
11. Maass D, Arango J, Wüst F, Beyer P, Welsch R (2009) Carotenoid Crystal Formation in Arabidopsis and Carrot Roots Caused by Increased Phytoene Synthase Protein Levels. PLoS ONE 4(7): e6373

12. 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 x wild carrot cross. Theor Appl Genet 119:1155–69

13. Hannoufa A, Hossain Z (2012) Regulation of carotenoid accumulation in plants. Biocatalysis and Agricultural Biotechnology 1:198–202

14. Cunningham FX, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol 49:557–583

15. Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. Current Opinion in Plant Biology 4:210–218

16. Bartley GE, Scolnik PA (1995) Plant carotenoids: pigments for photoprotection, visual attraction, and human health. The Plant cell 7:1027–1038

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

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

19. Just BJ, Santos CAF, Fonseca MEN et al (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

20. Isaacson T, Ohad I, Beyer P, Hirschberg J (2004) Analysis in vitro of the enzyme CRTISO establishes a poly-cis-carotenoid biosynthesis pathway in plants. Plant Physiol 136:4246–4255

21. Fang J, Chai C, Qian Q et al (2008) Mutations of genes in synthesis of the carotenoid precursors of ABA lead to pre-harvest sprouting and photo-oxidation in rice. The Plant Journal 54:177–189

22. Galpaz N, Ronen G, Khalfa Z et al (2006) A chromoplast-specific carotenoid biosynthesis pathway is revealed by cloning of the tomato white-flower locus. The Plant Cell 18:1947–1960
23. Li Q, Farre G, Naqvi S et al (2010) Cloning and functional characterization of the maize carotenoid isomerase and β-carotene hydroxylase genes and their regulation during endosperm maturation. Transgenic Res 19:1053–1068

24. Chai C, Fang J, Liu Y et al (2011) ZEBRA2, encoding a carotenoid isomerase, is involved in photoprotection in rice. Plant Mol Biol 75:211–221

25. Clotault J, Peltier D, Soufflet-Freslon V et al (2012) Differential selection on carotenoid biosynthesis genes as a function of gene position in the metabolic pathway: a study on the carrot and dicots. PLoS ONE 7(6): e38724

26. Briard M, Le Clerc V, Grzebelus D et al (2000) Modified protocols for rapid carrot genomic DNA extraction and AFLP™ analysis using silver stain or radioisotopes. Plant Molecular Biology Reporter 18:235–241

27. Cavagnaro PF, Chung S-M, Manin S et al (2011) Microsatellite isolation and marker development in carrot - genomic distribution, linkage mapping, genetic diversity analysis and marker transferability across Apiaceae. BMC Genomics 12:386

28. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

29. 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

30. Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86:485–486

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

32. Drummond AJ, Ashton B, Buxton S et al (2011). Geneious v5.1 created by Biomatters. Available from: http://www.geneious.com

33. Dutheil J, Gaillard S, Bazin E et al (2006) Bio++: a set of C++ libraries for sequence analysis, phylogenetics, molecular evolution and population genetics. BMC Bioinformatics 7:188

34. 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 Research 22:4673–4680
35. Librado P, Rozas J (2009) DnaSP v5: software for comprehensive analysis of DNA polymorphism data. Bioinformatics (Oxford, England) 25:1451–1452

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

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

38. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

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

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

41. Bradbury PJ, Zhang Z, Kroon DE et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

42. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

43. Remington DL, Thornsberry JM, Matsuoka Y et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

44. Stein, M. and Nothnagel, Th. (1995), Some remarks on carrot breeding (Daucus carota sativus Hoffm.). Plant Breeding, 114: 1–11

45. Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of Populus tremula. Genetics 180:329–340

46. Xing Y, Frei U, Schejbel B et al (2007) Nucleotide diversity and linkage disequilibrium in 11 expressed resistance candidate genes in Lolium perenne. BMC Plant Biology 7:43

47. Stenzel A, Lu T, Koch W et al (2004) Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. Hum Genet 114:377–385

48. Smith AV, Thomas DJ, Munro HM, Abecasis GR (2005) Sequence features in regions of weak and strong linkage disequilibrium. Genome Res 15:1519–1534
49. Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2(4):e64

50. Caicedo AL, Williamson SH, Hernandez RD et al (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet 3:1745–1756

51. Simon PW, Freeman RE, Vieira JV et al (2008) Carrot. In: Prohens J, Nuez F (eds) Springer New York, pp 327–357
Figure Legends

Figure 1. Position of primers used for CRTISO amplification. Boxes and lines represent exons and introns, respectively, of the CRTISO sequence. Arrows indicate partial sequences amplified with primers. The hatched box above the CRTISO sequence corresponds to the partial sequence analysed by [25].

Figure 2. Genetic structure of 39 carrot cultivars based on a Bayesian approach on microsatellite data and assuming two clusters. Each individual is represented by a single vertical box broken in two coloured segments, with length proportional to each cluster assignment probability (y-axis). Bars represent the 95% confidence interval. The individuals are placed on the x-axis according to the approximate longitude of their country of origin, from France (left) to Japan (right).

Figure 3. Linkage disequilibrium along CRTISO gene for the (A) Western and (B) Eastern genetic groups. LD level was measured by squared correlations of allele frequencies ($r^2$) against physical distance between pairs of SNPs.

Figure 4. Linkage disequilibrium along linkage group 4 [27] for the (A) Western and (B) Eastern genetic groups. LD level and its significance are represented at the top and at the bottom of the matrix, respectively. LD level was measured by squared correlations of allele frequencies ($r^2$) against genetic distance. Significance of LD for microsatellite pairs are given as p-values determined by permutations. Each matrix compares the LD between pairs of markers displayed at the left and the bottom.

Figure 5. Overall nucleotide diversity ($\pi_T$) (grey line), nucleotide polymorphism ($\theta_w$) (dashed line), and Tajima’s D value (black line), estimated along CRTISO gene for (A) the whole dataset, the (B) Western and (C) Eastern genetic groups. The analysis is based upon a 100 bp window with a step size of 25 bp. The exons and introns are shown as grey and white boxes, respectively. The significance for Tajima’s D is displayed at the top of each plot.

Figure 6. Median joining network derived from reconstructed DNA sequence haplotype of CRTISO gene. Haplotypes are displayed at network nodes and are symbolised by circles whose diameter is proportional to the frequency in the sample. The code number of the haplotypes is written in the circle centre. Perpendicular hashes on the lines joining haplotypes represent substitutions. When too numerous, they are displayed as slanting hashes with the number of substitutions
differentiating two network nodes. The main clusters are identified by dotted boxes. (A) The black and grey proportions of each circle represent respectively the proportion of individuals from the Western and Eastern genetic groups (estimated by STRUCTURE analysis) sharing the same haplotype. (B) Colours represent the proportion of different colour groups: white, yellow, orange, red and purple.
Table 1. List of carrot cultivar samples.
ACO-IRHS : Agrocampus Ouest - Institut de Recherche en Horticulture et Semences (FR) ; WGRU : Warwick Genetic Resources Unit (UK) ; VIR : Vavilov Research Institute (RUS)

| Code | Cultivar name          | Root colour | Geographical origin | Source                       |
|------|------------------------|-------------|---------------------|------------------------------|
| 100  | Kuttinger              | White       | Switzerland         | ACO-IRHS                     |
| 104  | Blanche Collet Vert Hors Terre | White | France             | ACO-IRHS                     |
| 105  | Blanche des Vosges     | White       | France              | ACO-IRHS                     |
| 108  | Long White Green Top   | White       | Denmark             | WGRU                         |
| 109  | White Belgian          | White       | United Kingdom      | WGRU                         |
| 112  | SAR112                 | White       | Middle East         | ACO-IRHS                     |
| 208  | Gelbe Lobbereicher     | Yellow      | Germany             | WGRU                         |
| 210  | Jaune du Doubs         | Yellow      | France              | ACO-IRHS                     |
| 222  | Golden Promise         | Yellow      | Asia                | Mikado-Kyowa Seeds           |
| 223  | Lobberich              | Yellow      | Italy               | ACO-IRHS                     |
| 224  | Ch-Wy                  | Yellow      | Asia                | Mikado-Kyowa Seeds           |
| 226  | YC226                  | Yellow      | Asia                | ACO-IRHS                     |
| 230  | Taborska               | Yellow      | Czech Republic      | Nohel Garden                 |
| 231  | Shima Ninjin           | Yellow      | Japan               | Mikado-Kyowa Seeds           |
| 307  | Kuroda                 | Orange      | Japan               | ACO-IRHS                     |
| 313  | Amsterdam 2 Sweetheart  | Orange      | Netherlands         | WGRU                         |
| 337  | De Colmar à Cœur Rouge 2 | Orange | France             | ACO-IRHS                     |
| 360  | Flakee                 | Orange      | Netherlands         | ACO-IRHS                     |
| 365  | Mestnaya LR            | Orange      | Kyrgyzstan          | VIR                          |
| 366  | Mestnaya Zheltaya      | Orange      | Uzbekistan          | VIR                          |
| 369  | Mestnaya LR            | Orange      | Kazakhstan          | VIR                          |
| 370  | Sapporo futo           | Orange      | Japan               | Mikado-Kyowa Seeds           |
| 372  | Koizumi Riso           | Orange      | Japan               | Mikado-Kyowa Seeds           |
| 373  | Kokubu Senso Oonaga    | Orange      | Japan               | Mikado-Kyowa Seeds           |
| 374  | Heian - 3 Sun          | Orange      | Japan               | Mikado-Kyowa Seeds           |
| 3015 | Nantaise améliorée     | Orange      | France              | ACO-IRHS                     |
| 403  | Annual Red Rawalpindi  | Red         | Pakistan            | WGRU                         |
| 407  | Pink Selection         | Red         | China               | WGRU                         |
| 411  | Red Queen              | Red         | India               | Sungro                       |
| 420  | Pusa Kesar             | Red         | India               | WGRU                         |
| 421  | JF421                  | Red         | Asia                | ACO-IRHS                     |
| 426  | Honbeni Kintoki        | Red         | Japan               | Takii                        |
| 500  | AnthocyaneG            | Purple      | Middle East         | ACO-IRHS                     |
| 514  | Afghan Purple          | Purple      | Afghanistan         | WGRU                         |
| 515  | PT2 515                | Purple      | Middle East         | ACO-IRHS                     |
| 516  | PD516                  | Purple      | India               | ACO-IRHS                     |
| 520  | PM520                  | Purple      | Middle East         | ACO-IRHS                     |
| 521  | PD521                  | Purple      | Middle East         | ACO-IRHS                     |
| 522  | PA522                  | Purple      | Middle East         | ACO-IRHS                     |
Table 2. Sequences of primers used for CRTISO amplification.
F, forward primer; R, reverse primer

| Amplified fragment | Primer sequence (5’-3’) | Length (bp) |
|--------------------|-------------------------|-------------|
| ① 5'UTR - Exon1    | F aatcacctcctcccccaag R tcactgaagccaaacatcaca | 20          |
|                    | F tgggtggagctctggatatt R aaatggacagtactggtggtca | 21          |
| ② Exon1 - Exon2    | F cttaaccttgataactcaagcattgg | 25          |
|                    | R cagtgttaggcattcccataggc | 22          |
| ③ Exon2 - Exon3    | F tgccttgcaactcattggaac | 20          |
|                    | R tgcgttgatcattggtcttct | 20          |
| ④ Exon3 - Exon5    | F ctcaaaatgctggagactagc | 22          |
|                    | R tccctactgtagcatttga | 20          |
| ⑤ Exon4 - Exon7    | F ctggcgaatggaaatgagat | 20          |
|                    | R tccctcttgagcaatgtag | 21          |
| ⑥ Exon7 - Exon9    | F tgggctaatgataggatgag | 22          |
|                    | R ggccttcttgagcataagc   | 22          |
| ⑦ Exon9 - Exon11   | F agacacacaggccgtacctt | 20          |
|                    | R caacctctgtgccctctatgt | 20          |
| ⑧ Exon11 - 3'UTR   | F caacccctctgcccctcag   | 20          |
Table 3. Fixation indexes ($F_{CT}$) among colour groups based on microsatellite data.

|       | Yellow   | Orange   | Red      | Purple   |
|-------|----------|----------|----------|----------|
| White | 0.019(NS)| 0.049*   | 0.094*   | 0.114*   |
| Yellow| 0.003(NS)| 0.015(NS)| 0.057*   | 0.142*   |
| Orange|          | 0.109*   |          |          |
| Red   |          |          | 0.053*   |          |

NS, non significant; *, $P < 0.05$
Table 4. Nucleotide polymorphism of *CRTISO* in a sample of 39 carrot cultivars.
(Sequences without indels, total length 4234 bp, coding region length 1845 bp)

|          | No. of sequences | $\theta_w$ | $n_T$  | $n_{coding}$ | $n_{non\text{-}coding}$ | $n_{sil}$ | $n_{syn}$ | $n_{nonsyn}$ | $n_{nonsyn}/n_{syn}$ | $h$  | $Hd$  |
|----------|------------------|------------|--------|--------------|--------------------------|-----------|-----------|--------------|------------------------|------|-------|
| Total    | 78               | 0.00973    | 0.01753| 0.00928      | 0.02433                  | 0.02496   | 0.02788   | 0.00354      | 0.12697                | 15   | 0.837 |
| West     | 38               | 0.01018    | 0.01808| 0.00912      | 0.02547                  | 0.02591   | 0.02845   | 0.00335      | 0.11775                | 8    | 0.644 |
| East     | 40               | 0.00971    | 0.00880| 0.00479      | 0.01211                  | 0.01242   | 0.01308   | 0.00201      | 0.15367                | 10   | 0.826 |
| White    | 12               | 0.01462    | 0.02071| 0.01099      | 0.02869                  | 0.02952   | 0.03411   | 0.00409      | 0.11991                | 5    | 0.742 |
| Yellow   | 16               | 0.01119    | 0.01915| 0.00984      | 0.02681                  | 0.02735   | 0.02947   | 0.0037       | 0.12555                | 4    | 0.692 |
| Orange   | 24               | 0.01139    | 0.0176 | 0.00931      | 0.02444                  | 0.02503   | 0.0283    | 0.00364      | 0.12862                | 7    | 0.786 |
| Red      | 12               | 0.01222    | 0.0122 | 0.00603      | 0.01728                  | 0.01757   | 0.01797   | 0.00209      | 0.11630                | 3    | 0.667 |
| Purple   | 14               | 0.01191    | 0.01105| 0.00541      | 0.01569                  | 0.01598   | 0.01698   | 0.00177      | 0.10424                | 5    | 0.78  |
Table 5. Tajima’s $D$ value calculated for the whole dataset, the genetic and colour groups.

| No. of sequences | Tajima’s $D$ value |
|------------------|--------------------|
|                  | Total | Coding regions | Non-coding regions |
| **Total**        | 78    | 2.75530**      | 1.7608            |
| **West**         | 38    | 2.89351**      | 2.3006*           |
| **East**         | 40    | -0.34631       | -0.7079           |
| **White**        | 12    | 1.95597        | 1.2712            |
| **Yellow**       | 16    | 3.08636***     | 2.8237**          |
| **Orange**       | 24    | 2.18514*       | 2.0361*           |
| **Red**          | 12    | -0.00587       | -0.0431           |
| **Purple**       | 14    | -0.32423       | -0.5050           |

*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$

Supporting Information File Legends

Table S1. Characteristics of 17 carrot microsatellite markers (Cavagnaro et al. 2011)
Figure 3

Click here to download high resolution image

![Graph A](image)

**A**

Average $r^2 = 0.61$

![Graph B](image)

**B**

Average $r^2 = 0.64$
