Heredity (2007) 98, 375–384
© 2007 Nature Publishing Group All rights reserved 0018-067X/07 $30.00
www.nature.com/hdy

ORIGINAL ARTICLE

Recent degeneration of an old duplicated flowering time gene in *Brassica nigra*

P Sjödin[1], H Hedman[2,3], O Shavorskaya[2,4], C Finet[1,5], M Lascoux[1] and U Lagercrantz[1]

[1]*Evolutionary Functional Genomics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden and* [2]*Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden*

Gene and genome duplications play a major role in the evolution of plant species. The *Brassica nigra* genome is highly replicated as a result of ancient polyploidization events. Two copies of the flowering time gene *CONSTANS* (*COa* and *COb*) have been identified in *B. nigra*, and previous studies showed that *COa* is functional. In the present study, the polymorphism of 92 *COb* alleles sampled in seven populations was analyzed. Both polymorphism and recombination levels were elevated and varied strongly among populations and 8% of *COb* alleles exhibit apparently disabling mutations. Sequence data, however, do not provide unambiguous support for the presence of relaxed selective constraint on *COb* as compared to known functional *CO* genes. On the one hand, some of the disabling mutations reached high-frequency arguing for a loss of function but, on the other hand, the ratio of nonsynonymous to synonymous nucleotide polymorphism and diversity is low and similar to that observed in other *B. nigra CO* and *CO*-like genes, supporting the conservation of some function. We also showed that *COb* is still transcribed. Finally, the flowering time of *Arabidopsis thaliana co* mutant plants transformed with *COb* alleles with and without apparent disabling mutations was similar. We propose that *COb* was retained for a long period after duplication, but a recent fixation of a detrimental mutation, possibly as an effect of a bottleneck, resulted in its nonfunctionalization. We also speculate as to the presence of subsequent selection for rapid degeneration of the gene.

Heredity (2007) **98,** 375–384; doi:10.1038/sj.hdy.6800951; published online 7 March 2007

Keywords: Brassica; pseudogene; duplication; DNA polymorphism

Introduction

Gene duplication has long been predicted to play a major role in evolution and the adaptation of organisms to the environment as duplicated genes may be free to evolve and acquire new functions (Ohno, 1970; Hughes, 1994; Lynch and Conery, 2000). Recent genome sequencing efforts have indeed confirmed this prediction. The evolution of plants, in particular, seems to be characterized by recurrent polyploidization (Wendel, 2000). Even the small *Arabidopsis thaliana* genome contains numerous large duplicated chromosomal segments (Blanc *et al.*, 2000; Vision *et al.*, 2000), suggesting at least two large-scale duplication events (Blanc *et al.*, 2003), and the genomes of several Brassicaceae species have experienced additional rounds of polyploidization after their divergence from *A. thaliana* (UN, 1935; Lagercrantz and Lydiate, 1996; Lagercrantz, 1998). The high rate of polyploidization observed in plants, together with frequent local gene duplication events, such as tandem duplications, results in a large number of duplicated genes (Arabidopsis Genome Initiative (AGI), 2000). Thus, a high degree of redundancy is expected and understanding the fate of duplicated genes and the significance of gene duplication in general remains one of the main issues of plant evolution and speciation.

Theoretical models predict three alternative fates of duplicated genes. First, as most mutations are disabling (Lynch and Walsh, 1998) and duplicate genes are generally assumed to be redundant, most models predict that one copy of a duplicate pair will eventually be silenced and become a pseudogene. If population sizes are not very large, the expected time to silencing will be relatively short (a few million years or less) (Watterson, 1983; Lynch and Force, 2000). Second, one copy may acquire a novel function whereas the other copy retains the original gene function (Ohno, 1970). Third, the functions of both copies are compromised by mutation so that each gene fulfils separate parts of the original gene function (subfunctionalization) (Force *et al.*, 1999). The relative importance of these three alternatives is still unclear and may be difficult to assess. Notably, many genes which were classified as pseudogenes due to the presence of disabling mutations (e.g., stop codons, frameshifts indels), have actually been shown to have acquired/retained a function, for instance in the regulation of other genes (Balakirev and Ayala, 2003; Hirotsune *et al.*, 2003).

Correspondence: Professor U Lagercrantz, Evolutionary Functional Genomics, Evolutionary Biology Centre, Uppsala University, Norbyv. 18D, Konsumv. 18a, Uppsala 75236, Sweden.
E-mail: Ulf.Lagercrantz@ebc.uu.se
[3]*Current address: Evolutionary Functional Genomics, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden.*
[4]*Current address: Genetics and Biology Laboratory, BioSciences Institute, University College Cork, Western Road, Cork, Ireland.*
[5]*Current address: Laboratoire Reproduction et Developpement des Plantes, UMR 5667, Ecole Normal Supérieure de Lyon, Lyon Cedex, France.*

Received 4 September 2006; revised 3 November 2006; accepted 2 February 2007; published online 7 March 2007

The *A. thaliana CONSTANS* (*CO*) gene is a putative transcription factor that promotes flowering in response to long days (Putterill *et al*., 1995). *CO* is a member of the family of *CO*-like (*COL*) genes, which is characterized by two conserved domains, two B-box type zinc fingers, and a C-terminal domain named CCT which is also present in CO, COL and TOC proteins (Strayer *et al*., 2000; Robson *et al*., 2001). The closest paralog to *CO* in *A. thaliana* is *COL1*, which is located about 3 kb upstream of CO and is probably the result of a tandem duplication. A *COL1* homologue is also present in *B. nigra,* showing that this duplication took place before the divergence of the lineages leading to *Arabidopsis* and *Brassica*.

The diploid *B. nigra* has been inferred to descend from a hexaploid ancestor (6*n*). The progenitor of the hexaploid ancestor, a diploid plant with a genome size similar to that of *A. thaliana*, went through a triplication after the lineages leading to *Brassica* and *Arabidopsis* separated (Lagercrantz, 1998). Consequently, single-copy genes in *A. thaliana* are expected to be represented by at least three copies in *B. nigra*. Two copies of *CONSTANS* (*COa* and *COb*) have so far been identified in *B. nigra* (Lagercrantz and Axelsson, 2000). We have previously shown that *B. nigra COa* is functional as it can complement an *Arabidopsis co* mutant (Kruskopf-Österberg *et al*., 2002). Previous quantitative trait locus (QTL) mapping identified a major QTL for flowering time mapping close to *COa* and also a smaller QTL for the same trait mapping close to *COb* (Lagercrantz *et al*., 1996). Although a copy of *COL1* is, as in *A. thaliana*, located 3 kb upstream of *COa*, we have not been able to find a *COL1* homolog near *COb*.

Preliminary studies of sequence diversity of *COb* showed that some *COb* alleles carry disabling mutations (e.g., premature stop codons) whereas others are free of them. Could then the previously identified flowering time QTL be due to the segregation of functional and nonfunctional alleles at *COb*? To test this hypothesis, we analyzed the sequence polymorphism of a large number of *COb* alleles from different populations of *B. nigra*. We also transformed an *A. thaliana co* mutant with apparently functional and nonfunctional *COb* alleles, and assayed the transformants for flowering time.

Materials and methods

Study organism and plant material

The black mustard, *Brassica nigra* ($2n=16$), is an outcrossing annual, closely related to *A. thaliana*. In temperate regions, *B. nigra* was a major mustard crop before the 1950s but has been replaced by *B. juncea* and is today mainly a weed. However, it is still used as a condiment in India and Ethiopia. Even if *B. nigra* is not currently the object of extensive breeding, its use as a condiment certainly implies a proportionally strong human influence. Its current natural distribution area ranges from the Atlantic Ocean eastward to India and from Southern Scandinavia southward to Ethiopia. It has also been introduced in North America. Its phylogeography and recent history are still poorly documented, but a previous study using nuclear microsatellites delineated three main groups, namely Europe, India and Ethiopia (Westman and Kresovich, 1999). At a finer scale, preliminary results based on chloroplast DNA variation (Alström-Rapaport *et al*., unpublished) also suggest ancient admixture in populations from Northern Europe, with parental populations located in the Iberian Peninsula, Italy and the Balkans. All population samples used in this study were *ex situ* germplasm accessions. Seed samples originating from France (Plouvenez-Lochrist, IPK CR 2140), Germany (IPK BRA1269), Greece (Smila, IPK CR2104 and IPK BRA187), Italy (IPK BRA26 and IPK BRA27), India (Uttar Pradesh, USDA BN-PI-175073), Ethiopia (IPK BRA1163 and USDA BN-PI-273641) and Turkey (Balikesri, USDA BN-PI-592737) were obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany and from the USDA, ARS, North Central Plant Introduction Station, Iowa State University, IA, USA.

DNA sequencing

Genomic DNA was prepared from leaf tissue as described by Liscum and Oeller (1997). Genomic fragments of *B. nigra COb* were amplified using primers CO89 (CAA GAT GAT GGA TAC ACG AAT) and CO122 (GTA TTT ATG TTT ATG CGG GTG AAG C). Amplified products were treated with ExoSAP-IT (USB Corporation) and both strands were sequenced directly on an ABI 377 sequencer. PCR products from heterozygous individuals were also cloned, and both alleles were sequenced. At least two clones of each allele were sequenced. The GenBank accession numbers of the sequences are xxx.

The *CO* from *A. thaliana* was accession Y10556.

RT-PCR

RNA was extracted from leaves using the RNeasy Mini Kit (QIAGEN Inc., Valencia, CA, USA). First-strand cDNA was synthesized using 1 μg DNAse-treated total RNA, random hexamer primers and Invitrogene's Superscript II (Invitrogen Corp., Carlsbad, CA, USA). As a control for DNA contamination, parallel cDNA syntheses reactions were also performed without reverse transcriptase. *COb* was amplified from cDNA with 35 PCR cycles using primers CO18 (CTT CCC GCC ATA AAC GTG TCC) and CO 60 (GAA CAA TGC TGT ATC CTG TGT CAA). As a control for RNA quantity and cDNA synthesis efficiency, PCR amplifications (25 cycles) were also performed using QuantumRNA 18S Standards (Ambion Inc., Austin, TX, USA, Competimer to primer ratio 7:3).

DNA sequence analysis

Sequence data were aligned using AutoAssembler™ 2.0 software (Applied Biosystems, Foster City, CA, USA). Most sequence data analyses were carried out using DNAsp 4.0 (Rozas *et al*., 2003).

To test for differentiation among populations, we used a permutation test based on either the K_{ST} (Hudson *et al*., 1992) or the nearest-neighbor (S_{nn}) statistics (Hudson, 2001).

The population recombination rate, $\rho=4N_{\mathrm{e}}r$, where N_{e} is the effective population size and r is the per-locus recombination rate per generation, was estimated with Hudson's method for all data (Hudson, 1987), for each country and sample taken separately, as well as for the largest haplotype groups defined based on disabling mutations.

The age of some of the disabling mutations was estimated using two different approaches. First, assuming neutrality and an infinite allele model, mean allele age and an MLE of allele age and its approximate confidence interval were estimated from the observed allele frequencies in the sample following Griffiths (2003) and Slatkin and Rannala (2000). Allele age estimates were also obtained using the approach outlined in Aldrich *et al*. (2002) and Toomajian *et al*. (2003). The basic idea is to estimate an allele's age by the decay of ancestral haplotype sharing (DHS, McPeek and Strahs, 1999). In the present case, the haplotype free of disabling mutations was assumed to be the ancestral one. The DHS was estimated using the DHSMAP v.1.04 software (McPeek and Strahs, 1999). The program takes as input a set of marker haplotypes from affected individuals, haplotypes from the ancestral haplotype class, a genetic map of the markers, and a mutation rate for each marker. We do not have pedigree data and therefore the genetic distance between markers was inferred from the observed recombination fraction in the sampled sequences. The program then infers the ancestral haplotype surrounding the specified mutation and produces as an output a maximum-likelihood estimate of the time to the most recent common ancestor (TMRCA) in generations for the allele class. Strictly speaking, this TMRCA is not the time when a new allele arose by mutation but represents the lower boundary for that time; that is, the age of the variant will generally be somewhat higher than that of the most recent common ancestor in a population (Slatkin and Rannala, 2000). This method assumes that there is no selection at marker loci but can allow for mutations and for multiple origins of a variant.

We also analyzed *COb* sequence polymorphism data using the composite likelihood ratio (CLR) for the detection of selective sweeps developed by Kim and Stephan (2002) and further extended by Kim and Nielsen (2004). Jensen *et al*. (2005) showed that bottlenecks or population structure can lead to a high percentage of false positives. To accommodate this, they developed a 'goodness of fit' (GOF) statistic. Briefly, by simulating a selective sweep using the estimated location of the site under selection and the estimated value of the strength of selection, the distribution of GOF values is obtained and the estimated GOF value can be compared to this distribution. The programs are available at http://128.151.242.156/~orrlab/YuseobPrograms.html. We analyzed two alignments. The first alignment included only sequences without long deletions or missing data. In the second alignment sites with indel variation were removed. The first alignment thus consisted of longer sequences but fewer individuals than the second. As the results were similar for the two alignments but the second alignment had fewer sites, we will only present results for the first alignment.

Introduction of *B. nigra COb* alleles into *A. thaliana*

Four different constructs were produced. Two of them contained *COb* alleles with apparently functional coding regions and two others contained alleles with a premature stop codon (St3; Table 1) eliminating a significant part of the conserved C-terminal CCT domain. St3 is always associated with at least one other disabling mutation (Sp1, D1, d2) located in the other exon and even, in the case of D1, in another functional domain, the zinc finger. Hence, alleles carrying St3 are generally also impaired in other parts of the genes. Alleles were amplified by PCR and cloned in pPZP211 (Hajdukiewicz *et al*., 1994). The constructs contained genomic clones comprising the coding region with a 700 bp upstream sequence and a 200 bp downstream sequence. The resulting clones were resequenenced to exclude mutations from the PCR amplification. The plasmids were transformed into *Agrobacterium* strain GV3101 that was used to vacuum infiltrate an *A. thaliana co* mutant (*co-2*; Nottingham stock centre no. N175). Transformed lines were selfed to obtain homozygous lines. Five homozygous lines originating from independent transformation events were chosen from each construct for the flowering time experiments.

Twenty seeds from each of the five homozygous lines from each of the four constructs were randomly sown in 96-well flats. In addition, seeds from wild type and co-2 mutant were included. The plants were grown in a growth chamber under a 16-h photoperiod. Flowering time was scored as the number of days from germination to the opening of the first flower.

Table 1 Frequency of apparently disabling mutations in *B. nigra COb*

Pop	n	*St1*	*D1*	*St2*	*d2*	*d3*	*Sp1*	*Sp2*	*D4*	*St3*	*I1*	*I2*	*None*
		4[a]	*33*	*264*	*630*	*643*	*723*	*724*	*863*	*977*	*987*	*1056*	
			197						*1032*		*988*	*1057*	
Eth	22	0.05	0.05	0	0.05	0.09	0.41	0	0.09	0.41	0	0.18	0.27
Fra	8	0	0.38	0.38	0	0	0	0	0	0	0.38	0.38	0
Ger	10	0	0	0	0	0	1	0	0	1	0	0	0
Gre	20	0	0.15	0	0.10	0.30	0.65	0	0.20	0.40	0	0	0.30
Ind	10	0	0.50	0	0	0	0.70	0	0	0.80	0	0.20	0
Ita	26	0	0	0.04	0.35	0.15	0.27	0.04	0.15	0.42	0.04	0.08	0.19
Tur	10	0	0.60	0	0	0	0	0	0	0.10	0	0.50	0.30
Tot	106	0.01	0.17	0.04	0.11	0.11	0.43	0.01	0.09	0.44	0.04	0.15	0.19

Abbreviations: *n*, number of alleles sequenced.
Mutations denoted St result in premature stop codons, D denotes large deletions, d denotes one bp deletions resulting in frame shifts, Sp a mutation in a splice signal, and I denotes 2 bp insertions resulting in frame shifts.
[a] Numbers below the names of mutations represent its position in nucleotides from the start codon. The positions of different mutations are also shown in Figure 1.

Table 2 Haplotypes defined by disabling mutations

	S D S d d S S D S I I *t 1 t 2 3 p p 4 t 1 2* *1 2 1 2 3*	N	*Disabling mutations*
Hap 1	T G G A A G T T C – –	18	None
Hap 2	. . . . A – –	1	St3
Hap 3	. –	1	D1
Hap 4	. . . . A	3	I2
Hap 5	. . . . . C . . A . .	29	Sp1, St3
Hap 6	. . . . – C . – – . .	7	d3, Sp1, D4
Hap 7	. – . . – C . – – . .	1	D1, d3, Sp1, D4
Hap 8	. – . . . . . . A . .	1	D1, St3
Hap 9	. – . . . C . . A . .	2	D1, Sp1, St3
Hap 10	. . . – . . . . A . .	9	d2, St3
Hap 11	. . T . . . . . . G .	4	St2, I1
Hap 12	. . . . G .	2	I1
Hap 13	. – . . . . A	12	D1, I2
Hap 14	. . . . . . C	1	Sp2
Hap 15	A	1	St1
		92	

The disabling mutations are defined in Table 1.

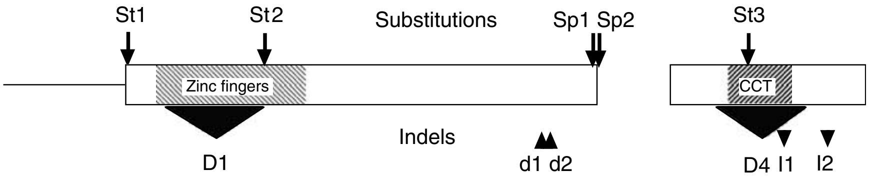

Figure 1 Schematic drawing of *COb* in *B. nigra* indicating identified putative disabling mutations. The different mutations are described in Table 1. Boxes indicate exons. Zinc-fingers and CCT denote the two conserved domains in *CO*-like genes.

Results

Nucleotide variation and disabling mutations at the *COb* locus

Ninety-two sequences of *COb*, corresponding to 76 haplotypes, and originating from seven countries, were obtained (Table 1). *COb* contains two putative exons separated by a short intron. The region sequenced included about 200 bp of the 5′ UTR and the whole gene except the last 56 bp of exon 2, resulting in a total of about 1300 bp, although some sequences were considerably shorter due to large deletions. In fact, the most prominent feature of the variation along the *COb* locus is the prevalence of (apparently) disabling mutations. In total, 11 different disabling mutations were identified, defining a total of 15 haplotype classes (Table 2). These mutations include base substitutions resulting in premature stop codons and alterations of the 5′-GU splice signal, as well as deletions eliminating large parts of the two conserved domains, and small deletions and insertions resulting in frame-shifts (Figure 1). A GATA-1 box was reported in the promoters of *CO* in both *A. thaliana* (Putterill *et al.*, 1995) and *B. napus* (Robert *et al.*, 1998). A corresponding GATA-1 box was also present in *B. nigra COa* (unpublished data). In all of the sequenced *COb* haplotypes however, a 12 bp deletion destroys this putative GATA-1 box.

The frequency of many of the disabling mutations varied widely across populations (Table 1), and the frequency of alleles without obvious disabling mutations was low in all samples (average 21%). These data suggest that *COb* might be a pseudogene, although the segregation of functional and nonfunctional alleles at *COb* could also explain the presence of a QTL for flowering time identified close to *COb* (Lagercrantz *et al.*, 1996). The geographical distribution of the disabling mutations suggests that they are rather old. However, this is not reflected by the age estimates obtained by the two methods used in the present study. The DHS method consistently led to very small estimates of the mutation age, but relies on unknown parameters such as genetic distance between markers. Assuming an effective population size of 5000 ($2N_e = 10\,000$), mutation ages estimated through allele frequencies were larger, but still surprisingly low (Table 3).

A total of 145 segregating sites (182 including segregating sites in indels) were observed in the sample of 92 *COb* sequences. In the coding region, 37 synonymous and 63 replacement changes were found. In addition to the apparently disabling indels described above, there were five short in frame indels (3 or 6 bp). The nucleotide diversity estimates for *COb* are relatively high, both for replacement and synonymous sites but the corresponding ratio, π_A/π_S, is comparatively low (0.31) (Table 4).

Hudson's estimate of the population recombination rate, $\rho = 4N_e r$, was also high (25.3) when all data were used. The minimum number of recombination events for *COb* was estimated to 23 by Hudson and Kaplan (1985) and to 36 by Myers and Griffiths (2003). LDhat (McVean *et al.*, 2001) gave almost the same estimate as Hudson

Table 3 Age estimates for disabling mutations observed in *COb* gene

Method for age estimation	*Disabling mutations*										
	St1	*D1*	*St2*	*D2*	*d3*	*Sp1*	*Sp2*	*I1*	*D4*	*I2*	*St3*
Decay of haplotype sharing											
Age (generations)	~0		951	626	720	727	720	928	720	626	1052
95% CI (coalescent history)	0–1049		659–1041	1–1051	383–1052	715–741	0–1052	360–621	197–1052	432–708	872–1052
95% CI (star-shaped history)	0–1049		269–1051	30–1051	197–1052	705–764	0–1052	638–979	382–1052	253–1052	936–1052
Frequency of allele											
2Ne = 10 000											
Mean age (generations)	808	7160	2480	5469	5469	12706	808	2480	4825	6929	1263
MLE age (generations)	94[a]	1672	195	1012	1012	5502	94[a]	195	802	1447	5670
95% CI (MLE)	0–5259	771–17416	17–11872	442–16330	442–16330	2484–18959	0–5259	17–11872	334–15741	660–17131	2552–18982

The age obtained by the method based on allele frequency assumes an effective population size of $2N_e = 10\,000$.
[a] Assuming sample size is infinite, otherwise negative values are obtained (we used $-\ln(1-p)$ instead of $-\ln(1-p)-2/n$ with $n = 106$ as in the other cases).

Table 4 DNA polymorphism in the *COb* gene of *B. nigra*

	All	*Non-coding*	*Syn.*	*Nonsyn.*	*Syn. and noncoding*	π_A/π_S
S	145	44	37	63	81	
π	0.028	0.032	0.059	0.018	0.043	0.31
θ_w	0.038	0.046	0.060	0.028	0.052	0.46

S is the number of segregating sites, π is the nucleotide diversity, and θ_w is the Watterson (1975) estimate. The segregating sites associated with indels and ambiguous nucleotides were excluded from the calculations.

Table 5 Differentiation among populations

	Fra	*Gre*	*Ita*	*Ger*	*Tur*	*Eth*	*Ind*
Fra		0.177	0.144	0.534	0.169	0.104**	0.299
Gre	1		0.027 NS	0.165**	0.149	−0.011 NS	0.102**
Ita	0.916	0.516 NS		0.172	0.140	0.023 NS	0.123
Ger	1	0.709*	0.880		0.514	0.158**	0.095*
Tur	0.858	1	0.935	1		0.106**	0.115**
Eth	0.885**	0.475 NS	0.584 NS	0.748**	0.888**		0.101**
Ind	0.812	0.681*	0.750*	0.558 NS	0.883	0.654 NS	

Above diagonal: K_{st}. Below diagonal: S_{nn}. All values are highly significant unless indicated (NS: nonsignificant; *$P<0.05$, **$P<0.01$).

(1987) method (25.5 instead of 25.3). The population recombination rate was also estimated for each population separately as well as for some of the haplotype classes based on the presence or absence of disabling mutations. The population recombination for the most frequent haplotype carrying disabling mutation (H5), $\rho = 0.001$, was much lower than for H1, the haplotype free of disabling mutations, $\rho = 2.2$.

Although both measures of population differentiation, K_{ST} and S_{nn}, were highly significant, their values were relatively low – 0.254, $P<0.01$ and 0.38345, $P<0.001$, respectively – and the resulting grouping of population was difficult to interpret (Table 5). Interestingly, we did not observe at *COb*, the strong differentiation between Ethiopia and European populations and the low genetic differentiation among European populations found for *COL1* (Lagercrantz *et al*., 2002).

Strong differentiation between haplotypes defined by disabling mutations and weak separation between localities is also evident in the neighbor-joining tree (Figure 2). To a large extent, the major groups in the tree represent disabling mutation haplotypes. Haplotype 1 that lacks obvious disabling mutations was present in samples from Ethiopia, Greece, Italy and Turkey, whereas Haplotype 5 dominated in Germany and India and was also found in other populations (Ethiopia, Italy and Greece).

Using the method of Nei and Gojobori (1986), the replacement (K_A) and synonymous (K_S) divergences between *B. nigra COb* and *B. nigra COa* were 0.19 and 0.48, respectively. This yields an estimate of the ratio of replacement to synonymous substitutions of 0.40. Using the codon-based model of Goldman and Yang (1994), the estimate of this ratio was 0.41 for the whole gene (Lagercrantz and Axelsson, 2000). This estimate is actually lower than those between other presumably functional genes in the gene family (Lagercrantz and Axelsson, 2000).

Compared to *A. thaliana*, we found nine fixed indels with an average length of deletions of 6.75 bp and 11 polymorphic indels with an average deletion length of 36.63 bp in the coding region. Moreover, although all the fixed indels are within the reading frame (have lengths that are multiples of 3), only three of the 10 polymorphic indels are.

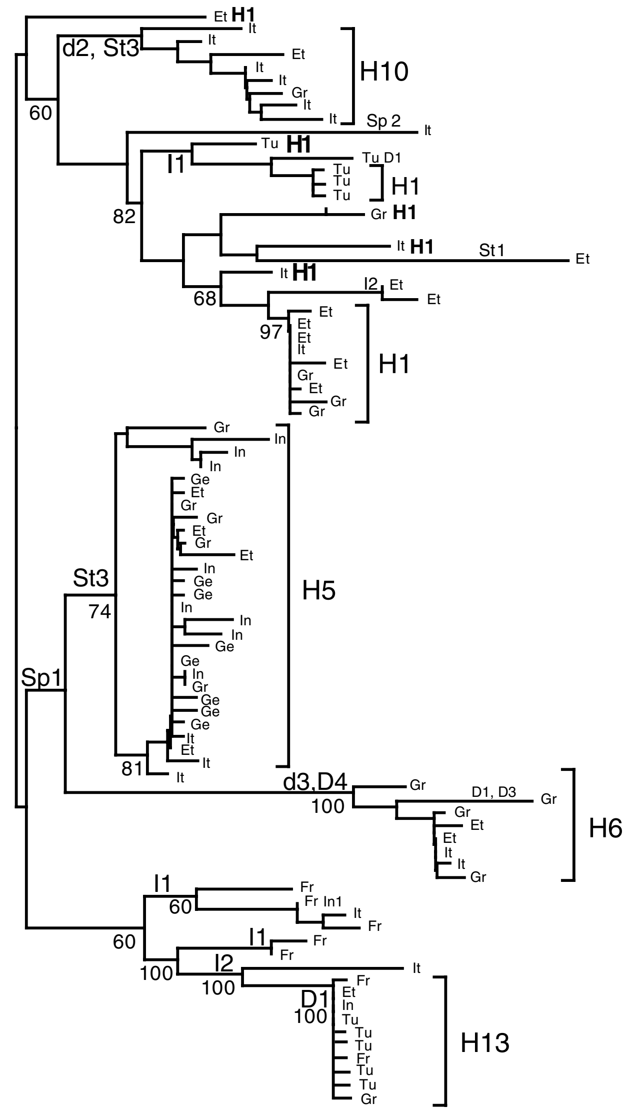

Figure 2 Neighbor-joining tree based on the Kimura two-parameter model. Codes (e.g., d2) at nodes or above branches indicate mutations described in Table 1; numbers below nodes indicate bootstrap values. Main haplotypes (Hap) defined in Table 2 are also indicated.

Evolution of *COb* does not conform to the standard neutral model

Tajima's D value was -0.843 and not significantly different from zero when all sequences were considered. Fu and Li D^* and F^* statistics, on the other hand, were both significant ($P<0.05$) with values -3.37 and -2.79, respectively, as well as Ramos-Onsins and Rozas, (2002) R_2 statistic ($R_2=0.0697$, $P<0.05$). The latter compares the expected number of singletons on a genealogy branch after severe recent population growth with the observed one. A lower value is expected under population growth than under constant population size. The distribution of polymorphism at *COb*, hence, departs from that expected under the standard neutral model.

If *COb* is a pseudogene as suggested by the high frequency of disabling mutations, the surprisingly low replacement to synonymous mutation ratio could be accounted for if the majority of the polymorphic mutations are older than the mutation that disabled *COb*. Recombination can have such an effect as it allows different parts of a gene to have different genealogies. If the variance of the time to the most recent common ancestor (TMRCA) along the gene is exceptionally high, this would strengthen our case even further. There are well-known factors that would have precisely this effect, in particular a bottleneck or a selective sweep. We estimated the CLR of a selective sweep model to a neutral constant size model. The resulting CLR was 10.75. To estimate the significance of this value, we simulated the neutral model a thousand times with the same values of mutation and recombination and estimated CLRs of the simulated data. 2.5% of the simulated data had higher CLR than our data and we could therefore reject the neutral model at the nominal level. We then used the GOF test of Jensen *et al*. (2005) to assess whether the departure was due to a selective sweep or to demographics. The observed GOF values were significantly higher than the simulated values, and we therefore rejected the (simple) model of selective sweep implemented in the program. We also computed a sliding window of the likelihood of a specific region being the focus of the selected sweep. We found three potential regions (Figure 3) that correspond to (1) the very beginning of our sequences (the 5′ UTR region), (2) the end of the first B-box in the zinc-finger region and (3) a region close to the 'semi' -conserved mr1 region (see Griffiths *et al*., 2003) after the zinc-finger region. We interpret this as these regions having an exceptionally short time to most recent common ancestor (TMRCA).

COb alleles are transcribed and their proteins have similar effects on flowering time

As the coding sequences of some haplotypes of *COb* showed characteristics indicating that they might be functional, we used reverse transcription-polymerase chain reaction (RT-PCR) to test whether they were transcribed. PCR primers specifically amplifying *COb* were used in PCR reactions from total cDNA. *COb* cDNA from leaf tissue was amplified from a number of *B. nigra* plants from different populations, as well as from *A. thaliana* plants transformed with different *COb* alleles. Low levels of amplification products could be seen from all samples with cDNA, but no products were obtained from control samples where the RT enzyme was omitted from the cDNA synthesis (Figure 4). Thus, we conclude that transcribed RNA of *COb* was present in all tested plants, both those with and without disabling mutations. RT-PCR from some plants produced multiple bands, possibly an effect of alternative splicing that could be due to the observed splice site mutations. The level of transcription was low (35 PCR cycles were required to obtain a product visible on ethidium bromide stained gels), but this is also true for the functional *COa* paralog in *B. nigra* (Lagercrantz and Axelsson, 2000) and for *CO* in *A. thaliana* (Putterill *et al*., 1995).

To test if apparently functional and nonfunctional *COb* alleles showed different effects on flowering time, they were transformed into *A. thaliana*. We cloned two apparently functional alleles, and two alleles with a premature stop codon (St3; Figure 1; Table 1), and introduced these alleles into an *A. thaliana co* mutant. Plants from five homozygous lines of each of the four

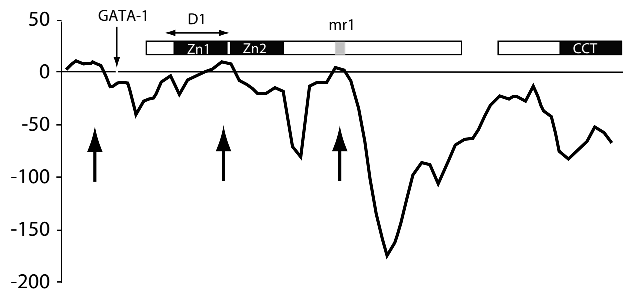

Figure 3 A sliding window of the likelihood ratio that the selective sweep has acted on a particular region of the sequence. Window length was set to 20 bp. The position of the missing GATA-1 box, the conserved regions including middle region1, and deletion D1 (not present in this alignment, see text) is indicated. Arrows indicate the possible locations of the mutations that rendered *COb* a pseudogene.

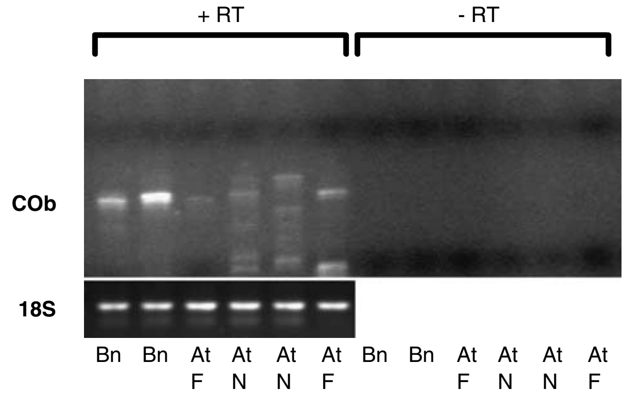

Figure 4 Detection of mRNA of *B.nigra COb* alleles by RT-PCR. RNA was isolated from leaves of *B. nigra* (Bn) and transformed *A. thaliana* plants (At) and used to synthesize cDNA, which was amplified by PCR using *Cob*-specific primers. Amplifications of 18S RNA were used as a control of cDNA amount and quality. To check for contamination of genomic DNA, amplifications were conducted on cDNA synthesis reactions both with (+ RT) and without reverse transcriptase (−RT). F and N denote apparently functional and nonfunctional alleles (see text).

constructs were grown in long day conditions and assessed for flowering time. In this experiment, wild-type *A. thaliana* flowered after an average of 24.3 days, and the *co* mutant after 34.5 days (Table 6). Transformed lines of all four alleles flowered slightly earlier than the mutant (31.8 and 31.4 days for potentially functional and nonfunctional, respectively). The difference in flowering time between mutant and transformed lines was statistically significant ($F_{2,8} = 12.7$; $P = 0.003$). However, the transformed lines with potentially functional and nonfunctional alleles showed almost identical flowering times. These data therefore do not suggest that the two types of alleles have a different effect on flowering time, and it is also possible that the slightly earlier flowering in transformed lines as compared to the untransformed mutant could be due to the transformation process itself.

Discussion

COb shows high levels of variability

COb showed surprisingly high polymorphism at both synonymous and non-synonymous sites. Indeed, estimates of nucleotide variation at *COb* were about three times higher than for the closely related *CONSTANS LIKE1* gene (*COL1*; Lagercrantz *et al*., 2002) estimated in a similar but more limited *B. nigra* sample ($\pi = 0.028$ and 0.008, respectively) and also relatively high compared to estimates obtained for genes in other plant species (Liu *et al*., 1998; Filatov and Charlesworth, 1999). This high level of polymorphism in *COb* could well be related to the high population recombination rate that was observed in some of the populations. A positive correlation between recombination and variability has been reported in both animals (Kaplan *et al*., 1989; Nachman, 1997; Przeworski *et al*., 2000; Hellmann *et al*., 2003) and plants (Dvorak *et al*., 1998; Kraft *et al*., 1998; Stephan and Langley, 1998). A positive correlation between recombination and sequence variability is generally explained by the presence of background selection or hitchhiking. Both background selection and hitchhiking will reduce genetic variation at linked neutral sites, resulting in higher loss in regions characterized by low recombination. Background selection and/or hitchhiking seem to fit well with the data in *Drosophila melanogaster* but, in humans, both intraspecies and interspecies differences increased with recombination rates, suggesting an intrinsic and neutral relationship between the mutation and recombination processes (Hellmann *et al*., 2003). One possible explanation for the high polymorphism in *COb* compared to *COL1* could therefore be that *COb* is located in a recombination hot spot. A high level of silent polymorphism could potentially also be the result of a balanced polymorphism. However, a balanced polymorphism is not consistent with the negative values of Tajima's *D* and Fu and Li's *D** and *F**. Negative values of Tajima's *D* were also obtained for *COL1* (Lagercrantz *et al*., 2002), indicating that these deviations from neutrality are more likely due to demography than selection.

Is *COb* a pseudogene?

Part of our data strongly suggests that since the duplication event, *COb* has lost its function and become a pseudogene, whereas other aspects seem difficult to reconcile with the loss of function. The large number of apparently disabling mutations and the deletion in the *COb* promoter of part of a putative GATA-1 box identified in *A. thaliana CO* are the main arguments for the loss of function. GATA-1 boxes are present in the promoters of the light-regulated chlorophyll a/b binding protein (*CAB*) genes of different species (Gidoni *et al*., 1989), and have been shown to be involved in the activation of an Arabidopsis *CAB* gene by light and by the circadian clock (Anderson and Kay, 1995). A combination of light and circadian clock regulation is important for the induction of flowering by *CO* in *A. thaliana* (Mouradov *et al*., 2002). This suggests that the identified GATA-1 box could be essential for proper *CO* function.

There are also a number of observations that suggest that *COb* might not be a pseudogene. First, *COb* alleles

Table 6 Flowering time of *A. thaliana co* lines transformed with different *COb* alleles

Construct/line	*Potentially functional*	n	*Flowering time (s.e.)*
WT		46	24.3 (0.26)
Co		34	34.5 (0.26)
179	Yes	103	31.7 (0.17)
183	Yes	107	32.0 (0.14)
69	No	107	31.2 (0.12)
273	No	103	31.7 (0.18)

are transcribed at low level. Second, at least some *COb* alleles show a circadian rhythm (unpublished data) as do *B. nigra COa* and *A. thaliana CO*. This, of course, does not exclude the possibility that the pattern of expression of *COb* has changed in a way that renders *COb* nonfunctional. Finally, the pattern of nucleotide variation at *COb* is not easy to explain under the assumption that *COb* is a pseudogene. A pseudogene should, by definition, be free of selective constraints, resulting in a ratio of replacement to synonymous substitutions close to one. However, *COb* does not show a higher replacement to silent divergence ratio than *COa*, which is functional. In particular, estimates of the ratio of replacement to synonymous substitutions for the two conserved domains known to be functionally important in *A. thaliana CO* were lower in comparisons including *COb* than in those including the functional *COa* gene.

To explain these apparently contradictory results, it might be useful to distinguish three alternatives: first, all alleles are functional, that is, *COb* is not a pseudogene; second, some alleles are functional but others are nonfunctional; and finally all alleles are nonfunctional or, in other words, *COb* is indeed a pseudogene. None of these three alternatives is a priori inconceivable, but we will argue that the former two are unlikely.

Obviously, the extreme variation in predicted amino-acid sequences is strong evidence against the first alternative. A majority of the alleles lack at least one of the conserved domains (zinc-finger or CCT domain) that are known to be vital for *CO* function. If all *COb* alleles were to be functional, we have to hypothesize a completely novel function for *COb*.

Our data also discredit the possibility that there are two classes of alleles, a functional class represented by the H1 haplotype and a nonfunctional one, that are maintained by balancing selection, as, for instance, Fu and Li's D^* were significantly negative. This, of course, does not exclude balancing selection, as the negative values may be a consequence of strong demographical factors masking the effect of balancing selection. Moreover, and even if selection is very strong, a high rate of null mutations such that there is a one-way flow of alleles from the ('functional') H1 class to the nonfunctional class would not create the deep split in the genealogy characteristic of balancing selection. However, although there is no definite argument to rule out balancing selection (or frequency-dependent selection), the lack of flowering time difference between alleles from the H1 class and another class in the transformation experiments strongly undermines the initial reason for this hypothesis: the QTL for flowering time found close to *COb*. In this case, our original classification of functional alleles as the H1 class must be wrong.

The major problem in assuming that *COb* is a pseudogene is the low value of the replacement to synonymous mutation ratio. This, however, can be accounted for if we assume that *COb* only recently became a pseudogene. If the TMRCA of the sequence region where the putative mutation that rendered *COb* nonfunctional is much shorter than the average TMRCA of the coding region, then most of the segregating sites in our sample will be due to mutations in a still functional gene while the disabling mutations most likely occurred recently in a nonfunctional or possibly even detrimental gene. The high estimated recombination rate assures that variation in TMRCA is present within *COb*. Furthermore, the age estimates of the disabling mutations that are all rather recent even if we assume a very large effective population size lends support to this hypothesis.

It is well known that a bottleneck or a selective sweep has the effect of increasing the variance of the TMRCA. An area within the gene with exceptionally low TMRCA is detectable by a local reduction in genetic variation and a negative value of Fay and Wu's *H* in areas close by.

Although we were not able to confirm the occurrence of a selective sweep, we could reject the standard neutral model and identify three candidate regions with the trademark signs of a local star shaped genealogy: (1) the very beginning of our sequences (the 5′ UTR region), (2) the end of the first B-box in the zinc-finger region and (3) a region close to the conserved middle region 1 (mr1) (see Griffiths *et al*., 2003) after the zinc-finger region. Of these three regions we found strong candidates to the mutation that impaired the function of *COb* in the first and second regions. The first region lacks the putative GATA-1 box, and in the second region the second B-box starts with CESCEC in *COb* instead of the CESCER amino-acid motif as in *CO* in *A. thaliana* and *COa* and *COL1* in *B. nigra*. The latter also seems to be a good candidate as it is very close to, and in between, the null mutations found in *Arabidopsis* mutants *co-2* and *co-4* and disrupts the spacing of C residues that is highly conserved in B-boxes (Griffiths *et al*., 2003).

Additionally, it is interesting to note that most of the disabling mutations remove the zinc finger or the CCT domain, so there appears to be a deficit of mutations impairing the middle region, which is less conserved in all *CO*-like genes. More precisely, given that 12 (including *St1.5)* disabling mutations occurred in 342 codons, the presence of a stretch of 121 contiguous codons without any disabling mutations seems unlikely. To qualify this we estimated the probability (through simulations) of having more than 121 contiguous codons without disabling mutations given that 12 disabling mutations were present in 342 codons. Although the estimated value was not exceptionally low ($P<0.07$), and as *COb* is still expressed, this may indicate that directional selection is acting to remove parts that could interfere detrimentally with the normal *COa* product.

Obviously, the fate of a duplication depends on the function of the duplicated gene. The silencing or functional modification of *COb* contrasts with the situation for another important integrator of flowering time control, *FLC*. *FLC* is a key component of the vernalization response pathway in *Arabidopsis* (Michaels and Amasino, 1999). In *B. rapa*, which most likely shares the whole-genome duplications present in the *B. nigra* genome (Lagercrantz and Lydiate, 1996), four *FLC* homolog have been cloned and shown to retain an original function (Schranz *et al*., 2002). Although both *CO* and *FLC* act in a dose-dependent manner, *FLC* is expressed at high levels (Henderson *et al*., 2003) but *CO* is expressed at very low levels (Putterill *et al*., 1995). An interesting question for future studies is to what extent duplicated *CO* and *FLC* genes are retained or silenced in other related species that share the same genome duplication events.

Conclusions

In conclusion, our data suggest that the minor QTL for flowering time in the genomic region harboring *COb* is not due to variation at this locus. Instead we suggest that *COb* is a recent pseudogene. A bottleneck may have led to the fixation of the GATA-1 box deletion in the 5′ UTR or the cysteine residue at the beginning of the second B-box but did not affect the majority of the coding region as strongly. Bottlenecks have been invoked before, for instance in *Buchnera*, to explain the transition of genes from a functional to a nonfunctional status (Wernegreen and Moran, 1999). Although the result from the GOF simulations only suggests that the most simple selective sweep scenario is unlikely and does not give direct evidence for the presence of a bottleneck, we favor the bottleneck hypothesis. It has the potential to account for the transition from a functional and selectively constrained status to a nonfunctional or detrimental status as the efficiency of selection is decreased during a bottleneck. This scenario could explain why *COb*, although kept under selection for a long time, was eventually lost.

Our data also suggest that *COb* was functional and conserved for a long period of time, but a recent fixation of a detrimental mutation resulted in selection for removal of still functional domains *COb*. The two long polymorphic deletions observed in the conserved regions could be interpreted as a first step towards such a removal. If the evolutionary scenario we have suggested for *COb* is common, duplicate copies may both be maintained by purifying selection, but once a copy loses its original function, it is quickly removed from the genome by deletions. Hence, if a pseudogene is found, it is likely to be still marked by selective constraints. This may, at least partly, explain the contradictory signals found in several pseudogenes (Balakirev and Ayala, 2003).

Acknowledgements

We thank Tanja Slotte and two anonymous reviewers for valuable comments on the manuscript. This research was supported by the Swedish Research Council (VR), the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (FORMAS), the Erik Philip Sörensen Foundation, and the Nilsson-Ehle Foundation.

References

Arabidopsis Genome Initiative (AGI) (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.

Aldrich C, Wambebe C, Odama L, Di Rienzo A, Ober C (2002). Linkage disequilibrium and age estimates of a deletion polymorphism (1597DeltaC) in HLA-G suggest non-neutral evolution. *Hum Immunol* **63**: 405–412.

Anderson SL, Kay SA (1995). Functional dissection of circadian clock- and phytochrome-regulated transcription of the *Arabidopsis* CAB2 gene. *Proc Natl Acad Sci USA* **92**: 1500–1504.

Balakirev ES, Ayala FJ (2003). Pseudogenes: are they ‘junk’ or functional DNA? *Annu Rev Genet* **37**: 123–151.

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000). Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**: 1093–1101.

Blanc G, Hokamp K, Wolfe KH (2003). A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res* **13**: 137–144.

Dvorak J, Luo MC, Yang ZL (1998). Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing aegilops species. *Genetics* **148**: 423–434.

Filatov DA, Charlesworth D (1999). DNA polymorphism, haplotype structure and balancing selection in the Leavenworthia PgiC locus. *Genetics* **153**: 1423–1434.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.

Gidoni D, Brosio P, Bond-Nutter D, Bedbrook J, Dunsmuir P (1989). Novel cis-acting elements in Petunia Cab gene promoters. *Mol Gen Genet* **215**: 337–344.

Goldman N, Yang Z (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11**: 725–736.

Griffiths RC (2003). The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theor Pop Biol* **64**: 241–251.

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003). The evolution of CONSTANS-like gene families in barley, rice and *Arabidopsis*. *Plant Physiol* **131**: 1855–1867.

Hajdukiewicz P, Svab Z, Maliga P (1994). The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol Biol* **25**: 989–994.

Hellmann I, Ebersberger I, Ptak SE, Paabo S, Przeworski M (2003). A neutral explanation for the correlation of diversity with recombination rates in humans. *Am J Hum Genet* **72**: 1527–1535.

Henderson IR, Shindo C, Dean C (2003). The need for winter in the switch to flowering. *Annu Rev Genet* **37**: 371–392.

Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S *et al*. (2003). An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**: 91–96.

Hudson RR (1987). Estimating the recombination parameter of a finite population model without selection. *Genet Res* **50**: 245–250.

Hudson RR (2001). Two-locus sampling distributions and their application. *Genetics* **159**: 1805–1817.

Hudson RR, Boos DD, Kaplan NL (1992). A statistical test for detecting geographic subdivision. *Mol Biol Evol* **9**: 138–151.

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.

Hughes AL (1994). The evolution of functionally novel proteins after gene duplication. *Proc Roy Soc London B* **256**: 119–124.

Jensen JD, Kim Y, Bauer DuMont V, Aquadro CF, Bustamante CD (2005). Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* **170**: 1401–1410.

Kaplan NL, Hudson RR, Langley CH (1989). The ‘hitchhiking effect’ revisited. *Genetics* **123**: 887–899.

Kim Y, Stephan W (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765–777.

Kim Y, Nielsen R (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics* **167**: 1513–1524.

Kraft T, Sall T, Magnusson-Rading I, Nilsson NO, Hallden C (1998). Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. maritima). *Genetics* **150**: 1239–1244.

Kruskopf-Österberg M, Shavorskaya O, Lascoux M, Lagercrantz U (2002). Naturally occurring indel variation in the *B. nigra COL1* gene is associated with variation in flowering time. *Genetics* **161**: 299–306.

Lagercrantz U (1998). Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150**: 1217–1228.

Lagercrantz U, Axelsson T (2000). Rapid evolution of the family of CONSTANS LIKE genes in plants. *Mol Biol Evol* **17**: 1499–1507.

Lagercrantz U, Kruskopf-Österberg M, Lascoux M (2002). Sequence variation and haplotype structure at the putative flowering-time locus COL1 of *Brassica nigra*. *Mol Biol Evol* **19**: 1474–1482.

Lagercrantz U, Lydiate DJ (1996). Comparative genome mapping in *Brassica*. *Genetics* **144**: 1903–1910.

Lagercrantz U, Putterill J, Coupland G, Lydiate D (1996). Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *Plant J* **9**: 13–20.

Liscum M, Oeller P (1997). AFLP: not only for fingerprinting, but for positional cloning. http://carnegiedpbstanford.edu/methods/aflp.html.

Liu F, Zhang L, Charlesworth D (1998). Genetic diversity in *Leavenworthia* populations with different inbreeding levels. *Proc Roy Soc London B Biol Sci* **265**: 293–301.

Lynch M, Conery JS (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.

Lynch M, Force A (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473.

Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer: Sunderland, MA.

McPeek MS, Strahs A (1999). Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. *Am J Hum Genet* **65**: 858–875.

McVean G, Awadalla P, Fearnhead P (2001). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.

Michaels SD, Amasino RM (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**: 949–956.

Mouradov A, Cremer F, Coupland G (2002). Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* **14** (Suppl): S111–S130.

Myers SR, Griffiths RC (2003). Bounds on the minimum number of recombination events in a sample history. *Genetics* **163**: 375–394.

Nachman MW (1997). Patterns of DNA variability at X-linked loci in Mus domesticus. *Genetics* **147**: 1303–1316.

Nei M, Gojobori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.

Ohno S (1970). *Evolution by Gene Duplication*. Springer Verlag: Berlin.

Przeworski M, Hudson RR, Di Rienzo A (2000). Adjusting the focus on human variation. *Trends Genet* **16**: 296–302.

Putterill J, Robson F, Lee K, Simon R, Coupland G (1995). The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc-finger transcription factors. *Cell* **80**: 847–857.

Robert LS, Robson F, Sharpe A, Lydiate D, Coupland G (1998). Conserved structure and function of the *Arabidopsis* flowering time gene CONSTANS in *Brassica napus*. *Plant Mol Biol* **37**: 763–772.

Robson F, Costa MM, Hepworth SR, Vizir I, Pineiro M, Reeves PH *et al.* (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. *Plant J* **28**: 619–631.

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

Schranz ME, Quijada P, Sung SB, Lukens L, Amasino R, Osborn TC (2002). Characterization and effects of the replicated flowering time gene FLC in *Brassica rapa*. *Genetics* **162**: 1457–1468.

Slatkin M, Rannala B (2000). Estimating allele age. *Annu Rev Genomics Hum Genet* **1**: 225–249.

Stephan W, Langley CH (1998). DNA polymorphism in lycopersicon and crossing-over per physical length. *Genetics* **150**: 1585–1593.

Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Mas P *et al.* (2000). Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science* **289**: 768–771.

Toomajian C, Ajioka RS, Jorde LB, Kushner JP, Kreitman M (2003). A method for detecting recent selection in the human genome from allele age estimates. *Genetics* **165**: 287–297.

UN (1935). Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn J Bot* **7**: 389–452.

Vision TJ, Brown DG, Tanksley SD (2000). The origins of genomic duplications in *Arabidopsis*. *Science* **290**: 2114–2117.

Watterson GA (1983). On the time for gene silencing at duplicated loci. *Genetics* **105**: 745–766.

Wendel JF (2000). Genome evolution in polyploids. *Plant Mol Biol* **42**: 225–249.

Wernegreen JJ, Moran NA (1999). Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. *Mol Bio Evol* **16**: 83–97.

Westman AL, Kresovich S (1999). Simple sequence repeat (SSR)-based marker variation in *Brassica nigra* genebank accessions and weed populations. *Euphytica* **109**: 85–92.