Theor Appl Genet (2005) 112: 58–65
DOI 10.1007/s00122-005-0103-3

ORIGINAL PAPER

Robin G. Allaby · Gregory W. Peterson · David Andrew Merriwether · Yong-Bi Fu

Evidence of the domestication history of flax (*Linum usitatissimum* L.) from genetic diversity of the *sad2* locus

Received: 26 May 2005 / Accepted: 1 September 2005 /

Abstract A phylogenetic analysis was conducted on 34 alleles of 2.5 kb sized stearoyl-ACP desaturase II (*sad2*), obtained from 30 accessions of cultivated and pale flax (*Linum* spp.), to elucidate the history of flax domestication. The analysis supports a single domestication origin for extant cultivated flax. The phylogenetic evidence indicates that flax was first domesticated for oil, rather than fibre. The genetic diversity of the *sad2* locus in cultivated flax is low when compared to that of the pale flax assayed. An absolute archaeological date could be applied to the synonymous substitution rate of *sad2* in cultivated flax, yielding a high estimate of 1.60–1.71×10^{-7} substitutions/site/year. The occurrence of nonsynonymous substitutions at conserved positions of the third exon in alleles from cultivated flax suggests that the locus may have been subjected to an artificial selection pressure. The elevated synonymous substitution rate is also compatible with a population expansion of flax since domestication, followed by a population decline in historic times. These findings provide new insight into flax domestication and are significant for the continuous exploration of the flax germplasm for utilization.

Keywords Flax · Crop domestication · Network analysis · Sequence variation · *Sad2* gene

Communicated by F. Salamini

R. G. Allaby · D. A. Merriwether
Department of Anthropology, Binghamton University SUNY,
Binghamton, NY 13902, USA

G. W. Peterson · Y.-B. Fu (✉)
Plant Gene Resources of Canada, Saskatoon Research Centre,
Agriculture and Agri-Food Canada, 107 Science Place,
S7 N 0X2 Saskatoon, SK, Canada
E-mail: fuy@agr.gc.ca
Tel.: +1-306-9567642
Fax: +1-306-9567246

Introduction

The earliest oil and fibre crop was flax (*Linum usitatissimum* L.), constituting part of the 'Neolithic package' of crops emanating from the Near East some 10,000 years ago (Zohary and Hopf 2000). Flax was a principal source of oil and fibre from prehistoric times until the early twentieth century, and still remains a crop of considerable economic importance. Thus, it is important to understand the history of flax domestication and for continuous exploration of the flax germplasm for utilization. However, the domestication process of flax is still shrouded in uncertainty (Zohary and Hopf 2000).

There are principally two morphotypes of cultivated flax, which are broadly designated as 'oil varieties' and 'fibre varieties'. The oil varieties tend to be shorter plants with large seeds that contain 40% oil, while the fibre varieties are taller, more sparsely branched plants with smaller and fewer seeds (Zohary and Hopf 2000). Archaeological records are unclear as to which of the two uses, oil or fibre, was the primary cause of flax domestication. In fact, both oil and fibre productions have been proposed as the first use of the crop (Diederichsen and Hammer 1995).

The cultivated form of flax (*L. usitatissimum*) is a self-pollinating diploid plant with a karyotype of $2n = 30$. Morphological (Diederichsen and Hammer 1995), genetic (Gill and Yermanos 1967), and molecular (Fu et al. 2002b) evidence suggest that the wild progenitor of cultivated flax is pale flax (*L. angustifolium* Huds.), with which it is interfertile. Cultivated flax has long upright stems in comparison to wild forms and capsules that generally do not dehisce. Unlike many of the wild progenitors of the Near East Neolithic package crop species, pale flax has a very wide biogeographical range spanning Western Europe and the Mediterranean, North Africa, Western and Southern Asia, and the Caucasus regions. Cultivated flax has many geographical centres of genetic diversity, including central Asian, Western Asian, Mediterranean, and Abyssinian regions (Vavilov 1951) as well as the

European–Siberian region (Zeven and de Wet 1982; Diederichsen and Hammer 1995). The centres of genetic diversity of cultivated crops were taken as evidence of the centres of origin of those crops (Vavilov 1926) which, in the case of flax, has led to the suggestion that the different diversity regions of flax may represent independent domestication events within the wide biogeographical range of pale flax (Harlan 1971).

The earliest archaeological finds of pale flax come from Tell Abu Hureyra in Northern Syria (11,200–10,500 years ago) (Hillman 1975). Pale flax occurred throughout the Near East by the eighth millennium BC in archaeological records (Zohary and Hopf 2000). The first occurrence of cultivated forms of flax is evidenced in archaeological records by an increase in the seed size, at Tell Ramad in Syria 9000 years ago (Zeist and Bakker-Heeres 1975). Flax then spread from the Near East to Europe and the Nile Valley. The flax varieties that spread into the Danube valley were winter oil varieties. However, in Eastern Europe, summer fibre varieties were developed which spread into central Europe and replaced the original varieties (Helbaek 1959; Diederichsen and Hammer 1995). While it is uncertain whether this spread of fibre flax from Eastern Europe resulted from a domestication event independent to that of fibre flax in the Near East, it is suggested though that all modern fibre varieties in use today originated from Eastern Europe (Helbaek 1959).

In recent years RAPD and AFLP data have been collected to assess the genetic diversity of cultivated flax and to discern the genetic structure of extant flax germplasm (Everaert et al. 2001; Fu et al. 2002a; Fu 2005). Generally, RAPD variation in cultivated flax is low and fibre varieties have a narrower genetic base than oil varieties (Fu et al. 2002a). Also, three major global divisions of flax variation, roughly corresponding to the Indian subcontinent, Western Asia, and Europe, were observed (Fu 2005). It is unclear whether these three major divisions of genetic variation could represent independent domestication events, as supported by Harlan (1971), or are simply the result of the reproductive isolation of the three groups.

The objective of this study was to examine the genetic diversity of the stearoyl-ACP desaturase II (*sad2*) locus in 30 unique flax accessions, with the hope of retrieving phylogenetic information for understanding the early events of flax domestication. Informative phylogenetic inferences require the use of linked characters such as DNA or amino acid sequences, rather than anonymous genomic data, as the latter data can lead to misinterpretations of monophyly or polyphyly (Allaby and Brown 2003). The *sad* gene is responsible for converting stearoyl-ACP to oleoyl-ACP, by introducing a double bond at C_9, and, thus, can increase the unsaturated fatty acid content of the plant (Ohlrogge and Jaworski 1997). This gene is of commercial interest for the manipulation of unsaturated fatty acids in major crop plants (Knutzon et al. 1992) and it has been well characterized (Shanklin and Sommerville 1991; Singh et al. 1994), thus making it an attractive target for phylogenetic study. There are two paralogous *sad* loci in flax, *sad1* and *sad2*, which are differentially expressed in the plant (Jain et al. 1999). The *sad2* locus is more strongly expressed than *sad1* and, consequently, the more agronomically important locus of the two.

Materials and methods

All flax accessions were obtained from the Plant Genetic Resources of Canada germplasm bank (Table 1). The accessions included six *L. angustifolium* (La) accessions and 24 cultivated *L. usitatissimum* (Lu) accessions (nine landrace, seven fibre, and eight oilseed varieties) of diverse geographic origin.

DNA extraction

Plants were grown from the seed for 2 weeks in a greenhouse. Leaf tissue was collected from individuals from each accession and freeze–dried for 1 week. Dried leaf tissues, from 5 individuals of the accession LS 29 and from 1 individual of each remaining accession, were ground in a mixer mill and 20 mg of the resulting fine powder was subjected to nucleic acid extraction using a Plant DNeasy Mini kit (Qiagen, Mississauga, ON, Canada). The DNA was quantified using a fluorometer and Hoechst 33258 dye (Sigma, Oakville, ON, Canada), with excitation and emission wavelengths of 360 and 460 nm, respectively.

PCR

Oligonucleotide primers for PCR, to specifically amplify the *sad2* locus, were designed based on the AJ006957 and AJ006958 Genbank accession entries for *sad1* and *sad2* in *L. usitatissimum*, respectively, using Primer3 (Rozen and Skaletsky 1998) and manufactured by Invitrogen Canada Inc. (Burlington, ON, Canada). The *sad2* gene was amplified in two halves to facilitate sequencing. The upstream primer pair was forward [5′-CCATTCAATTCAATATCTCACATTC-3′] and reverse [5′-GAAAAGGATAAAGGGGCAGAT-3′], which were expected to produce a 1403 bp product. The downstream primer pair was forward [5′-TACTCACATCCTATCTGCCCCT-3′] and reverse [5′-CTCAACTCTCGGGCAAACTC-3′], which was expected to give a product of 1302 bp. The primer combination of the upstream forward and downstream reverse was expected to give a product of 2560 bp. The thermal cycling conditions used on an M.J. Research PTC-200 thermocycler were as follows: 94°C 3 min (94°C 1 min, 63°C 1 min, 72°C 1 min 30 s) × 4, (94°C 1 min, 60°C 1 min, 72°C 1 min 30 s) × 4, (94°C 1 min, 58°C 1 min, 72°C 1 min 30 s) × 25, 72°C 20 min, 12°C hold. The PCR cocktail contained one time PCR Assay Buffer (Promega, Madison, WI, USA), 2.5 mmol/l $MgCl_2$ (Promega), 0.2 mmol/l each dNTP

Table 1 List of 30 accessions sequenced, with their species/type and origin country

LS[a]	Accession number[b]	Species/type[c]	Description	Origin
1	CN 107293	La		Russia[d]
2	CN 107257	La		France[e]
3	CN 19021	La		France[e]
4	CN 107258	La		France[e]
5	CN 19022	La		Czech Republic
6	CN 19023	La		India
8	CN 100911	Lu–LR	Cremone	Italy
9	CN 100895	Lu–LR	Karbin	Ethiopia
10	CN 101070	Lu–LR	Landrace	Russia
11	CN 19009	Lu–LR	Landrace	China
12	CN 101021	Lu–LR	Landrace (oil)	Afghanistan
13	CN 100890	Lu–LR	Svapo	France
14	CN 100896	Lu–LR	Giza	Egypt
15	CN 19010	Lu–LR	Landrace	Iran
16	CN 100909	Lu–LR	Palestina	Israel
17	CN 101120	Lu–Fibre	Liana	Poland
18	CN 101160	Lu–Fibre	Wiko	Azerbaijan
19	CN 98935	Lu–Fibre	Motley fibre	Belarus
20	CN 18991	Lu–Fibre	Nike	Russia
21	CN 101086	Lu–Fibre	Ariadna	Hungary
22	CN 98946	Lu–Fibre	Talmune fibre	Netherlands
23	CN 101017	Lu–Fibre	Baladi	China
24	CN 19003	Lu–Oil	AC McDuff	Canada
25	CN 98965	Lu–Oil	New River	USA
26	CN 97888	Lu–Oil	Tomagoan	Iran
27	CN 98256	Lu–Oil	Arreveti	India
28	CN 98263	Lu–Oil	Chaurra Olajlen	Hungary
29	CN 100832	Lu–Oil	Barbarigo	Czech Republic
30	CN 100917	Lu–Oil	Raluga	Romania
31	CN 97436	Lu–Oil	Giza	Egypt

[a] *LS* Linum sequencing number

[b] *CN* Canadian National accession number at the Plant Gene Resources of Canada, Saskatoon SK Canada

[c] La Linum angustifolium, Lu L. usitatissimum, LR landrace, Fibre fibre flax, Oil oil flax

[d] Source seed was from All Russian Flax Research Institute, VNIIL, Torzhok, Russia. Country of origin is unknown

[e] Source seed was from Jardin Botanique de la Ville et de l'Universite de Caen, France. Country of origin is unknown

(Invitrogen), 0.08 U/µl *Taq* polymerase(New England Biolabs, Pickering, ON, Canada), and 4 ng/µl flax DNA template in a final volume of 25 µl. The PCR products were separated on 2% agarose (Sigma, Oakville, ON, Canada) with one time TAE (2 mol/l Tris, 1 mol/l glacial acetic acid, 50 mmol/l EDTA pH 8) for 3 h at 100 V. The agarose gels were stained for 20 min in 1 mg/l ethidium bromide and recorded using a digital gel documentation system.

Sequencing

Amplicons were excised from agarose gel and purified using a QiaQuick Gel purification kit (Qiagen) and resuspended in 16 µl Qiagen Elution Buffer. Two microlitres of the gel-purified amplicon was used as a template to reamplify the amplicon to produce sufficient template for sequencing. The PCR reamplification product was gel purified as above and sequenced using an Applied Biosystems capillary DNA sequencer (DNA Technologies Unit, Plant Biotechnology Institute, National Research Council of Canada, SK, Canada). The sequence plots were proofread by eye and edited with ChromasLite (McCarthy 1998).

Sequence analysis

The sequences were aligned using ClustalX (Thompson et al. 1997). Networks were constructed as described previously (Allaby and Brown 2001). Briefly, networks provide a graphical approach to describing character conflict, instances where characters support different trees, as reticulations. The resulting graphs may then be interpreted as either containing all most parsimonious trees or as a visualization of recombination events. The population parameter of ρ was estimated by identifying the founding node, within the network for a group of alleles, and calculating the mean number of synonymous substitutions between the alleles and the ancestral node (Forster et al. 1996). The distribution of ρ values between various groups was statistically tested using the Mann–Whitney test (Saillard et al. 2000). The test of Tajima's *D* (Tajima 1989a) was used to determine deviations from neutrality.

Results and discussion

All primer pairs produced amplicons of the expected size in the PCR; with the exception of the first half of the gene in the six *L. angustifolium* accessions studied, which were approximately 50 bp shorter than expected. The nucleotide sequences of *sad2* were obtained for 34 individuals of 30 unique flax accessions, including 5 individuals from the accession LS 29. Sequences have been submitted to GenBank (DQ157225-DQ157258). Flax is a self-fertilizing crop and, thus, the assayed accessions are usually homozygous. All accessions in this study appeared to be homozygous, with the exception of one (sample Lin 23, CN101017). The obtained sequences were aligned using the paralogue *sad1* from

L. usitatissimum (McGregor) as an outgroup (Genbank accession AJ006957).

Most polymorphic sites occurred in introns 1 and 2 (63%), intron 2 being the more variable of the two (Table 2). Exons 1 and 2 accounted for only 5% of the polymorphic sites, none of which occurred within the cultivated flax alleles, while 23% of polymorphic sites occurred in exon 3. Thus, exons 1 and 2 appeared to have been strongly conserved while less so for exon 3. Interestingly, two nonsynonymous changes were observed in the *sad2* alleles in this study. Both occurred in exon 3, in alleles assayed from cultivated flax, and both occurred at positions that were highly conserved in the *sad* genes of other plant species. The first resulted in an amino acid change from serine to proline, at position 225 in the SAD2 protein, and the second generated an amino acid change from glycine to serine at position 282. The former substitution was observed in the Genbank accession AJ006958, while the second substitution occurred in 8 accessions in this study, defining group I in the phylogenetic analysis below. Additionally, a 46 bp deletion in intron 1 occurred in all the pale flax accessions studied, but did not occur in any of the cultivated accession alleles or *sad1*. The segment of the *sad1* sequence, where the deletion occurred, appeared to have some symmetry and may be capable of forming a loop structure that may have facilitated its excision in the pale flax accessions.

Evidence for a single origin of cultivated flax

A network was constructed from 33 of the 34 alleles, from accessions in this study, and the two sequences of *sad1* and *sad2* from Genbank (Fig. 1). The sequences obtained from cultivated accessions occupy 7 nodes (I–VII in Fig. 1). The network has two areas of reticulation; the simpler associated with the alleles of cultivated flax and the more complex associated with the pale flax alleles. Considering the very low numbers of substitutions relative to the length of sequence, it is more likely that these reticulations represent recombination events rather than homoplasies. If this is the case only a single recombination event is evident within the alleles from cultivated flax close to the base of the clade. The sequence obtained from sample LS 23 was judged to be a chimeric product of two alleles because the front half of the gene had characters synapomorphic with group III, while the second half had characters synapomorphic with group I. Thus, this sequence was not included in any analysis. The network was rooted using the *sad1* gene sequence of flax and, consequently, the oldest node of the *sad2* network was identified. Surprisingly, only two *sad2* allele sequences were obtained from the 6 pale flax accessions from botanic gardens around the world, indicating the low diversity of these stocks. The alleles were highly divergent, indicating their antiquity. The branches of the two *sad2* alleles from pale flax did not join the network at positions that were directly ancestral to the alleles from cultivated flax. This observation suggests that pale flax populations containing wild alleles more closely related to the alleles of cultivated flax may exist but have yet to be sampled. In particular the occurrence of the 46 bp deletion in the pale flax alleles supports the assertion that they were not directly related to the common ancestor of the cultivated flax accessions. Clearly, the network shown in Fig. 1 resolves a single branch leading from the wild alleles and *sad1* outgroup, indicating a single common ancestor for this group of alleles for all geographic regions. Thus, a monophyletic origin of cultivated flax from pale flax is supported.

Studies of anonymous RAPD markers have previously shown that flax accessions fall into a few major geographical groupings, each supported by a single branch in the dendrogram of diverse flax germplasm (Fu 2005). However, the flax grouping based on *sad2* alleles revealed different geographic patterns (Fig. 1b). The accessions of Western and Eastern Europe were combined into a group that also included accessions from the New World, because of their common origin. The remaining geographical groups were the Near East, Africa, and Southern Asia. The alleles obtained from the European group occurred throughout the network, whereas those of Africa and the Near East were restricted to one and two nodes, respectively. While this data appears to support the notion of the centre of genetic diversity of flax being in Europe (including Russia and the New World) rather than the Near East, the

Table 2 Distribution of polymorphic sites in *sad2*

Type	Flanking region		Exon 1			Intron 1		Exon 2			Intron 2		Exon 3			Total		
	ti	tv	ti	tv	ns	ti	tv	ti	tv	ns	ti	tv	ti	tv	ns	ti	tv	ns
Wild	0	1	1	0	0	3	0	1	0	0	8	5	4	2	0	17	8	0
Cultivated	1	1	0	0	0	1	3	0	0	0	4	0	2	1	2	8	5	2
Total	1	2	1	0	0	4	3	1	0	0	12	5	6	3	2	25	13	2

The substitutions were separated into two categories, pale flax and cultivated flax, depending on where they occur on the network in Fig. 1. The substitutions were attributed to the cultivated group, if they occurred after the common ancestor of the cultivated alleles, and to the pale flax group, if they occurred before. Substitutions on the branch leading to the paralogue *sad1* were not included

ti Transitions, *tv* transversions, *ns* nonsynonymous substitution

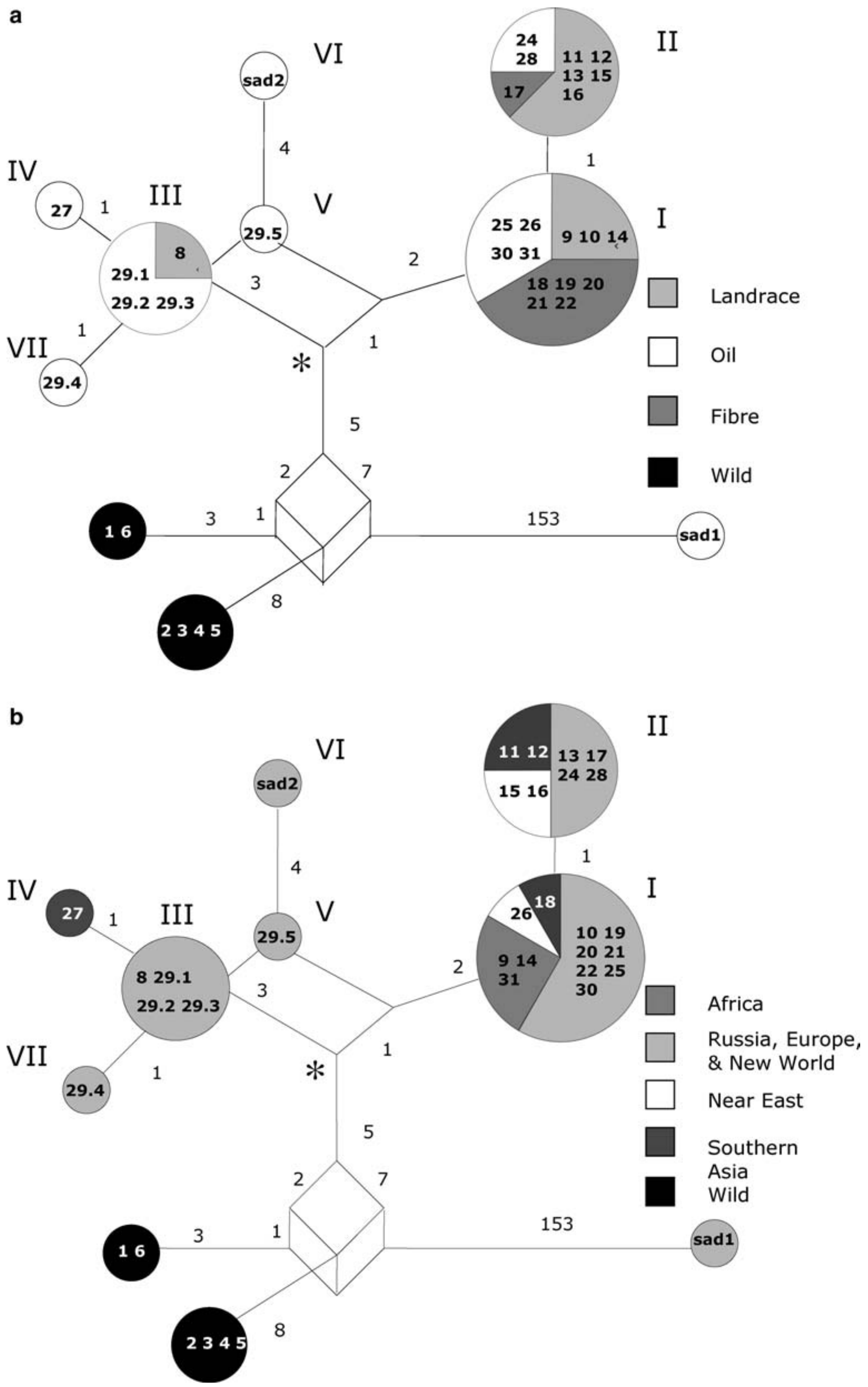

Fig 1 The network diagram of phylogenetic relationships of *sad1* and *sad2* nucleotide sequences. The numbers at node positions relate to accessions (see Table 1). The Genbank entries AJ006957 and AJ006958 are labelled as *sad1* and *sad2*, respectively. The size of node circles relates to sample frequency. The nodes, including alleles from cultivated flax accessions, are designated as groups I–VII. The numbers, by branches, indicate the number of substitutions for that branch. Character conflicts are described as reticulations within the network. The position of the most likely ancestor to all alleles in cultivated flax is indicated by an *asterisk*. **a** Phenotypic traits of accessions. **b** Geographical origin of accessions

sample size is too small to show a statistically significant difference in their ρ distributions. No alleles were obtained that occupy the common ancestor node position for cultivated alleles in the network and, thus, there is no direct phylogenetic evidence for the geographic origin of cultivated flax. It may be the case that such ancestral alleles have become extinct in modern flax populations. Groups I and III represent the most plesiomorphic allele types found in this study and such a pattern is compatible with an origin of flax in the Near East, which has alleles in Group I in the network, followed by a spread and expansion in Europe as envisioned by Helbaek (1959). Thus, the network obtained offers little resolution on the originating place of flax domestication.

Oil and fibre use

It has been the subject of some debate whether flax was domesticated primarily as a fibre or oil crop (Diederichsen and Hammer 1995). Figure 1a shows that the *sad2* alleles associated with fibre flax accessions appear to be restricted within the network, while the *sad2* alleles associated with oil varieties seem to be widespread throughout the network. The common ancestral node of the oil flax associated alleles is at the same point as the common ancestor for all alleles (Fig. 1). However, the common ancestor for the fibre flax associated alleles appears to be at the more derived node where group I occurred, given the present data. As a result, the ρ distributions of oil and fibre associated alleles differed with significance at the 0.001 level. This data strongly supports the antiquity of oil flax over fibre flax, implying that it was an oil-based crop for which flax was originally domesticated around 10,000 years ago. The relatively recent appearance of fibre-associated alleles is apparent in the fact that only a single segregating site occurred within that group. Interestingly, this site is where the two nonsynonymous changes were observed and it may reflect that the artificial selection process acted on the *sad2* locus for oil quality. It is noteworthy that no nonsynonymous substitutions occurred within the group alleles obtained from pale flax and, even more significantly, the *sad1* locus carried only a single nonsynonymous substitution out of 153 substitutions, which resulted in an amino acid change of glutamic acid to aspartic acid at position 357 of the SAD1 protein. This particular amino acid position is not highly conserved in other plant SAD proteins for either one of these two amino acids that are physiochemically very similar. The two nonsynonymous changes in alleles from cultivated flax at conserved amino acid positions appear to be remarkably unusual events within *sad* genes, and may be the result of artificial selective processes. Unfortunately, there is no data at the present time regarding the effect on the functionality of the SAD2 protein caused by these amino acid changes.

Sequence diversity of *sad2*

The diversity of the *sad2* alleles in cultivated flax was measured using π (Table 3). Globally, the diversity was low when compared to the divergent alleles present in the pale flax accessions. The oil producing flax accessions had alleles with the highest π values while the corresponding value for alleles from fibre producing varieties was tenfold lower. In the geographical groupings, both Europe (including Russia and the New World) and South Asia gave relatively high π values whereas the Near East and Africa gave tenfold lower π values. The domestication process is associated with a dramatic population expansion of the domesticated species. However, no grouping gave a statistically significant value for Tajima's D value, comparing nucleotide diversity to an estimate of the statistic θ based on the number of segregating sites divided by the harmonic mean of the sample size. This test works on the basis that the number of segregating sites increases more rapidly than the nucleotide diversity in a recently expanded population. If the domestication process is associated with a population bottleneck, then the signal for this test statistic should become stronger because the nucleotide diversity was more severely affected than the number of segregating sites (Tajima 1989b). However, it is similarly the case that the number of segregating sites declines more rapidly than the nucleotide diversity during a population contraction and, thus, the expansion signal in the data may theoretically be reduced by a subsequent population contraction (Tajima 1989b).

Elevated substitution rate

The statistic ρ, unlike π, can be applied to calculate the intraspecific substitution rate of particular allelic lineages, if the founding ancestral node is identified within the network (Forster et al. 1996). In this case, the ancestral node has been identified and archaeological records provide an absolute date for the earliest occurrence of flax in human contexts of 11,200–10,500 years ago (Hillman 1975). The synonymous substitution value for ρ for the *sad2* alleles from cultivated flaxes is 1.79×10^{-3} substitutions/site which yields an estimate of $1.60–1.71\times10^{-7}$ substitutions/site/year. This estimate is in considerable elevation of several general interspecific values calculated for the synonymous substitution rate in plants. Wolfe et al. (1989) calculated a value of $5.1–7.1\times10^{-9}$ substitutions/site/year for angiosperms based on *Chs*, *Adh*, *waxy*, *shrunken-1*, and *gapC* genes in grass lineages. This value range was later re-affirmed for grasses, but that of palms was found to be about twofold lower, based on alcohol dehydrogenase (*Adh*) genes (Gaut et al. 1996). However, higher rates have been found for members of the Brassicaceae at 1.5×10^{-8} substitutions/site/year based on *Chs* and *Adh* genes (Koch et al. 2000). It should be noted that all the above rate estimates depend on paleontological dates, knowingly with a limited degree of precision. Koch et al. (2000) generated a more precise and recent date based on fossil pollen evidence than the other two studies. Zhang et al. (2002) found that the synonymous substitution rate in *Arabidopsis* varied by as much as 13.8-fold within a group of 242 duplicated genes. The rate found in this study represents a tenfold increase on previous estimates of synonymous rates in plants, based on a relatively precise absolute date. A similar finding of intraspecific synonymous substitution rate elevation,

Table 3 Estimates of nucleotide parameters

Group	π/bp	D	ρ/bp
Region 1	0.001698	0.913	–
Region 2	0	0	–
Region 3	0.000195	−0.046	–
Region 4	0.001628	0.328	–
Oil	0.001728	0.336	0.001952
Fibre	0.000130	−0.075	0[a]
Landrace	0.000765	−0.524	0.001655
All cultivated	0.001311	0.046	0.001793
Pale flax	0.003598	1.456	–

Nucleotide diversity (π), Tajima's D (D) (Tajima 1989a) and Rho (ρ) (Forster et al. 1996) were calculated for various groupings of flax accessions. All nucleotide sites were used to calculate π (2560 bp), but only synonymous sites (upstream, intronic and exonic third position; 1815 bp) were used to estimate ρ. The four geographical regions were Region 1 (Europe, Russia and the New World), Region 2 (Africa), Region 3 (Near East), and Region 4 (South Asia). The accessions were grouped, by their phenotypic traits, into oil, fibre, landrace, and pale flax

[a] For the ρ calculations, the node position of the common ancestor of the cultivated alleles was used for oil varieties and landrace accessions, but the ancestor of fibre variety associated alleles was defined as group I (see Fig. 1)

based on precise radiocarbon dates, was also found for maize (Freitas et al. 2003); although, in this case, the rate was elevated by 100-fold.

The substitution rate may vary for several reasons. Kimura (1983) asserted that under ideal population conditions, neutrally evolving alleles substitute nucleotides at a rate equal to the mutation rate, regardless of the population. However, deviation from neutrality or ideal population conditions can cause this elegant equality to cease. A dramatic increase in population size will cause a drop in the lineage extinction rate for the duration of the expansion, thereby increasing the chance of each allele lineage acquiring new mutants. This may have the consequence of increasing the number of substitutions acquired during the expansion period. The domestication process is associated with a dramatic population increase, but in this study the classic signature for population expansion was not found using Tajima's D. However, a population contraction can have the effect of cancelling the expansion signal because of the rapid loss of segregating sites. In historic times fibre flax use has decreased globally, particularly since the 17th century as cotton has replaced it as the primary global fibre crop (Zohary and Hopf 2000). At the same time, the cultivation of oil flax may have increased considerably, due to the use of linseed oil for linoleum production and in paint industry. Alternatively, artificial selection of agronomic traits may also cause an increase in the substitution rate and may be also expected to be associated with nonsynonymous change. The *sad2* locus could also fit this scenario with the anomalous occurrence of two nonsynonymous changes within cultivated flax accessions.

Conclusive remarks

The network analysis presented here shows that cultivated flax is probably descended from the single domestication of a pale flax plant, apparently for its oil qualities. The genetic diversity of the *sad2* locus observed is compatible with the scenario of early artificial selection acting on that locus and/or a population contraction in historic times, resulting in diversity loss as evidenced by an elevated substitution rate without a population expansion signal. However, the generality of these findings may need to be determined with other linked characters as the use of the *sad2* locus might bias the inference toward oil varieties. Also, a higher resolution of flax domestication for the originating place and centre of flax diversity could be achieved with an assay of many more samples, particularly of pale flax with diverse geographic origins.

The findings of this study are also significant for the continuous exploration of flax germplasm for utilization. The diversity of the *sad2* locus offers a confirmation on the patterns of RAPD variation reported (Fu et al. 2002a), underscoring the need for broadening the genetic base of flax breeding materials, particularly for fibre breeding. While the geographic origin of flax cultivation remains uncertain, a focus of collection for diverse flax germplasm should still be placed on the Near East and Europe regions (Fu 2005). The germplasm of pale flax is currently limited worldwide (Diederichsen and Hammer 1995) and effort is needed to collect pale flax germplasm from various geographic regions; not only for studies of flax domestication but also for the exploration of new sources of genetic diversity (Fu et al. 2002b).

Acknowledgements The authors would like to thank Drs. Axel Diederichsen and Bruce Coulman for their helpful comments on an early version of the manuscript.

References

Allaby RG, Brown TA (2001) Network analysis provides insights into the evolution of 5S rDNA arrays in *Triticum* and *Aegilops*. Genetics 157:1331–1341

Allaby RG, Brown TA (2003) AFLP and the origins of agriculture. Genome 46:448–453

Diederichsen A, Hammer K (1995) Variation of cultivated flax (*Linum usitatissimum* L. subsp. *usitatissimum*) and its wild progenitor pale flax (subsp. *angustifolium* (Huds.) Thell.). Genet Res Crop Evol 42:262–272

Everaert I, de Riek J, de Loose M, van Waes J, van Bockstaele E (2001) Most similar variety grouping for distinctness evaluation of flax and linseed (*Linum usitatissimum* L.) varieties by means of AFLP and morphological data. Plant Variet Seeds 14:69–87

Forster P, Harding R, Torroni A, Bandelt H (1996) Origin and evolution of native American mtDNA variation: A reappraisal. Am J Hum Genet 59:935–945

Freitas FO, Bendel G, Allaby RG, Brown TA (2003) DNA from primitive maize landraces and archaeological remains: implications for the domestication of maize and its expansion into South America. J Archaeol Sci 30:901–908

Fu YB (2005) Geographic patterns of RAPD variation in cultivated flax. Crop Sci 45:1084–1091

Fu YB, Diederichsen A, Richards KW, Peterson G (2002a) Genetic diversity within a range of cultivars and landraces of flax (*Linum usitatissimum* L.) as revealed by RAPDs. Genet Res Crop Evol 49:167–174

Fu YB, Peterson G, Diederichsen A, Richards KW (2002b) RAPD analysis of genetic relationships of seven flax species in the genus *Linum* L. Genet Res Crop Evol 49:253–259

Gaut BS, Morton B, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at plastid gene *rbcL*. Proc Natl Acad Sci USA 93:10274–10279

Gill KS, Yermanos DM (1967) Cytogenetic studies on the genus *Linum*. Crop Sci 7:623–631

Harlan JR (1971) Agricultural origins: centers and noncenters. Science 174:468–474

Helbaek H (1959) Domestication of food plants in the Old World. Science 130:365–372

Hillman G (1975) The plant remains from Tell Abu Hureyra: A preliminary report. Proc Prehist Soc 41:70–73

Jain RK, Thompson G, Taylor DC, MacKenzie SL, McHughen A, Rowland GG, Tenaschuk D, Coffey M (1999) Isolation and characterisation of two promoters from linseed for genetic engineering. Crop Sci 39:1696–1701

Kimura M (1983) The neutral theory of evolution. Cambridge University Press, Cambridge, MA

Knutzon DS, Thompson GA, Radke SE, Johnson WB, Knauf VC, Kridl AC (1992) Modification of *Brassica* oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene. Proc Natl Acad Sci USA 89:2624–2628

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase in *Arabadopsis, Arabis* and related genera (Brassicaceae). Mol Biol Evol 17:1483–1498

McCarthy C (1998) ChormasLite v1.0., Technelysium Pty. Ltd., http://www.technelysium.com.au/chromaslite.html

Ohlrogge JB, Jaworski JG (1997) Regulation of fatty acid synthesis. Ann Rev Plant Physiol Plant Mol Biol 48:109–136

Rozen S, Skaletsky HJ (1998) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) MtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67:718–726

Shanklin J, Sommerville C (1991) Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proc Natl Acad Sci USA 88:2510–2514

Singh S, McKinney S, Green A (1994) Sequence of a cDNA from *Linum usitatissimum* encoding the stearoyl-acyl carrier protein desaturase. Plant Physiol 104:1075

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

Tajima F (1989b) The effect of change in population size on DNA polymorphism. Genetics 123:597–601

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876—4882

van Zeist W, Bakker-Heeres JAH (1975) Evidence for linseed cultivation before 6000 BC. J Archaeol Sci 2:215–219

Vavilov NI (1926) Studies on the origin of cultivated plants. Inst Bot Appl et d' Amelior. des Plantes. Leningrad, State Press (In Russian and English)

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Chronica Bot 13:1–366

Wolfe KH, Sharp PM, Li WH (1989) Rates of synonymous substitution in plant nuclear genes. J Mol Evol 29:208–211

Zhang L, Vision TJ, Gaut BS (2002) Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabadopsis thaliana*. Mol Biol Evol 19:1464–1473

Zeven AC, de Wet JMJ (1982) Dictionary of cultivated plants and their regions of diversity. Centre for Agricultural Publishing, Wageningen

Zohary D, Hopf M (2000) Domestication of plants in the Old World, 3rd edn. Oxford University Press, Oxford, pp 125–132