ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2016) **14**(2); 81–90
doi:10.1017/S1479262115000106

Effects of domestication bottleneck and selection on fatty acid desaturases in Indian sesame germplasm

Nupur Mondal[1,2], K. V. Bhat[1]*, P. S. Srivastava[2] and S. K. Sen[3]

[1]*Division of Genomic Research, National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi 110 012, India,* [2]*Department of Biotechnology, Jamia Hamdard University, New Delhi 110 062, India and* [3]*Indian Institute of Technology, Kharagpur, West Bengal, India*

Received 6 December 2014; Revised 29 January 2015; Accepted 13 February 2015 – First published online 18 March 2015

Abstract

Sesame (*Sesamum indicum* L.) is one of the oldest and most nutritional oilseed crops, of which domestication history has been poorly understood. This study suggested that sesame has undergone domestication bottleneck during its use for a long time. In this investigation, the molecular analysis included 4.4 Mbp of the genomic DNA of sesame comprising stearoyl acyl desaturase (*sad*), fatty acid desaturase 2 (*fad2*) and omega 3 fatty acid desaturase (*o3fad*) genes in 99 accessions of four populations of sesame germplasm namely: wild species, landraces, improved cultivars and introgressed lines. Results indicated that the improved cultivars and landraces lost 46.6 and 36.7% of nucleotide diversity, respectively, which indicate that the genetic diversity of the crop had been eroded due to selection after domestication. However, there was no significant reduction in genetic diversity of improved cultivars compared with landraces, indicating that unique improved cultivars generated through crosses were of less frequency in this population. Moreover, introgressed lines retained only 17.77% (π) and 4.57% (θ) of landrace diversity. To evaluate the impact of selection across fatty acid biosynthetic pathway, individual nucleotide diversity at three major genes involved in the pathway was surveyed. The analysis between wild and improved cultivars supported positive selection in *fad2* and *o3fad* loci. Though locus-to-locus sequence variation was observed, positive results with two most important loci supported selection after domestication. Reduced diversity in these critical quality governing genes in improved cultivars suggested that future sesame cultivation would benefit from the incorporation of alleles from sesame's wild relatives.

Keywords: crop domestication; fatty acid desaturase; genetic diversity; sesame; SNP

Introduction

It is now an accepted fact that domestication has had far-reaching effects on crop genomes. A common effect has been reduction in genetic diversity in crops compared with their wild progenitors because of population bottleneck or intensive selection (Tanksley and McCouch, 1997). Reduced diversity in crop cultivars is a growing concern because such crops lose wider adaptability and consistent productivity. Hence, there have been increased efforts to widen genetic base of the crops. Understanding the domestication genetics will greatly

* Corresponding author. E-mail: kvbhat2014@gmail.com

facilitate these efforts through discovery and utilization of rare but potentially important alleles present in these genetic resources.

Sesame (*Sesamum indicum* L.) is one of the oldest, most important and nutritious oilseed crops in the world. Sesame seed with high oil content along with antioxidants, vitamins and minerals makes it a desirable oilseed for consumption. Owing to its high economic importance, it is expected that the crop has been genetically modified in several regions of world, to improve its agronomic traits. This is evident from the improved traits of domesticated sesame, such as loss of seed dormancy, seeds being smooth and of varying colours, higher in oil content and large-sized capsules compared with its wild progenitors (Bedigian, 2003). The history of domestication of sesame, particularly of Indian germplasm, has not been studied yet.

Domestication-associated genes offer an approach to reconstruct a crop's history of domestication using associated trait with phylogeny and phylogeographic resolution. No domestication-associated genes have yet been identified in sesame. Moreover, a variety of traits getting evolved during domestication make it more unclear as to what extent a single gene can be used to infer the cultivation history of whole crop genome.

Oil accounts for up to 50% of sesame seed (Nzikou *et al.*, 2009), making it one of the major determinants of both sesame seed yield and quality. Strong selection by humans on oil traits is thus expected during sesame domestication. It is likely to have occurred via the genes in the fatty acid biosynthetic pathway. Three key genes, found in this pathway, are known to play major roles in unsaturated fatty acid production: stearoyl acyl carrier protein desaturase 2 (*fad2*) and omega 3 fatty acid desaturase (*o3fad*). *Sad* is the plant enzyme that catalyses the conversion of stearic acid C18:0 into oleic acid C18:1. *Fad2* encodes 18:1 desaturase and they are responsible for conversion of oleic C18:1 into linoleic acid C18:2. *O3fad* is responsible for conversion of linoleic acid C18:2 into linolenic acid C18:3. No study has been undertaken in sesame using these or any other genes to study domestication history in the world.

The aim was, therefore, to explore the domestication history of Indian sesame germplasm using domestication-associated genes. For this study, *sad*, *fad2* and *o3fad* DNA sequence variations were evaluated within and among four populations of sesame germplasm: wild species, landraces, introgressed lines and improved cultivars. The objective was to assess how DNA sequence variation was altered through human activities of cultivation and intensive breeding approximately over the past 4000 years in India.

Materials and methods

Plant material

A total of 99 accessions of sesame were sampled from the core collection, based on morphological traits, maintained by the National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India. Genotypes from different groups were selected to have a wide array of genetic diversity for this study. These included 70 accessions of endemic landraces, 20 improved Indian cultivars, four accessions of wild progenitor (*S. malabaricum* syn *S. mulayanum* Nair.), one accession of a wild species (*S. radiatum* Schum. et Thonn.) and four accessions of introgressed lines (Table S1, available online).

Fatty acid profile

A profile of the fatty acid composition of all the accessions was generated by gas chromatography. The procedure followed including oil extraction, fatty acid methyl esterification and fatty acid profiling was the same as reported earlier by us (Mondal *et al.*, 2010)

DNA extraction

DNA was extracted from the leaf tissue of a single, 2-month-old plant of each accession by the cetyl trimethyl ammonium bromide (CTAB) method (Saghai-Maroof *et al.*, 1984). DNA was diluted to yield 20 ng/μl concentration for further use.

Polymerase chain reaction and sequencing

Polymerase chain reaction (PCR) primers were designed using the Oligos software (Institute of Biotechnology, University of Helsinki, Finland). Coding regions of the three candidate genes were targeted for designing the primers. For amplifying the three genes, 12 different primer pairs were used, because the primers were designed to amplify with approximately 100 bp overlap between the neighbouring amplified regions for each gene. PCRs were conducted in 25 μl of solution, which included 1× Taq buffer without $MgCl_2$, 2.5 mM $MgCl_2$, 1 U Taq DNA polymerase, 0.25 μM of each primer, 2 mM dNTP mix, 2 μl of 20 ng of genomic DNA and sterile deionized water. PCRs were conducted under the following conditions: denaturation at 94°C for 2 min, 30 cycles of denaturation at 94°C for 1 min, annealing at different temperatures depending upon T_m (optimal melting temperature) of particular primer for 1 min, extension at 72°C for 1 min and final extension at 72°C for 10 min. The primers and the annealing temperature for PCRs are listed in Table S2 (available online).

https://doi.org/10.1017/S1479262115000106 Published online by Cambridge University Press

PCR products were cleaned using Axygen commericial kits (Axygen Biosciences 33210 Central Avenue, Union City, CA 94587 USA), and then cycle sequenced using ABI Prism Big Dye Terminator cycle sequencer (Applied Biosystems, CA). In majority of the cases, PCR amplified products were sequenced from both directions to reduce the error rates. Sequences obtained were compared with the sesame genome using the BLASTN program and only the best hits were kept for further analyses. After these filtering steps, the consensus sequences were aligned with CLUSTAL W (Thompson *et al.*, 1994) and the alignment was manually corrected with a sequence editor. Before the analysis, alignments were trimmed to exclude regions or sequences with missing or low-quality data.

Population genetic analysis

Population genetic analyses of the aligned sequences were performed using the program DnaSP 5.10 (Librado and Rozas, 2009). Nucleotide diversity, the average pairwise nucleotide divergence, of a multiple alignment was calculated as π (Tajima, 1983; Nei, 1987), and the number of segregating sites as θ (Watterson, 1975). Several measures of sequence variation were obtained, and they were haplotype number, the signal of selection, i.e. deviation from neutrality (Tajima, 1989; Fu and Li, 1993), and the frequency of recombination, i.e. the minimum number of recombination events (Hudson and Kaplan, 1985). The positions of single-nucleotide polymorphisms (SNPs) and indels for each accession were generated. The strength of linkage disequilibrium (LD) was estimated using the Z_{nS} statistic (Kelly, 1997), which is the average of R^2 (squared correlation coefficient) (Hill and Robertson, 1968) over all pairwise comparisons. DnaSP was also used to calculate Z_a (Rozas *et al.*, 2001), the average of R^2 (Hill and Robertson, 1968) over all pairwise comparisons between adjacent polymorphic sites; ZZ (Rozas *et al.*, 2001) $= Z_a - Z_{nS}$; and Wall's B and Q tests (Wall, 1999).

Sequences were collapsed into haplotypes and analysed using Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010). Within- and between-population pairwise distances, population comparisons (F_{ST}, 100 permutations) and molecular diversity (π, as a measure of nucleotide diversity, i.e. the probability that two randomly chosen homologous nucleotides are different) were computed by the pairwise distance method at a 0.05 significance level.

Results

Nucleotide diversity in sesame

A total of 4.4 Mbp of genomic DNA was sequenced, which comprised the coding region of the three unsaturated fatty acid-producing genes. In all the 99 sesame genotypes, 1501 bp of *sad*, 1437 bp of *fad2* and 1470 bp of *o3fad* were targeted. A total of 716 variable sites were found (460 singleton and 240 parsimony informative sites). In addition, 578 indel sites were identified, which along with variable sites account for SNPs. The total number of mutations in these three genes was 514, of which 137 were synonymous changes and 277 were replacement changes.

To describe nucleotide diversity or sequence variation, the two common measures used were π, the expected heterozygosity per nucleotide site (Tajima, 1983), and θ, the number of segregating sites in a genotypic sample corrected for sample size (Watterson, 1975). The wild sesame had a substantial diversity with an average π value of 0.0116 and θ value of 0.0121, while the improved cultivars had comparatively reduced diversity ($\pi = 0.0062$ and $\theta = 0.0112$). The landraces showed high diversity, while that of introgressed lines was quite low (Table S4, available online). Overall, improved cultivars retained 84.39% π and 37.45% θ of landraces (Table 1).

Domestication bottleneck

Domestication bottleneck represents the first occurrence of human-mediated selection (Hyten *et al.*, 2006). The nucleotide diversity retained by improved cultivars was 53.39% π and 92.56% θ of its wild population. The haplotype diversity (H_d) data showed a higher genetic diversity in wild species for two loci, *fad2* and *o3fad* (Table 2). Moreover, the matrix of pairwise F_{ST} showed that wild-improved cultivar populations had high values (0.1–0.2) in all three loci, reflecting appreciable

Table 1. Nucleotide diversity per base pair in coding regions within the four populations of sesame

Locus	Population	π	θ
Sad	Wild species	0.00 068	0.00 075
	Impr. cultivars	0.00 181	0.00 274
	Landraces	0.00 312	0.01 150
	Introgressed lines	0.00 183	0.00 182
Fad2	Wild species	0.01 168	0.01 248
	Impr. cultivars	0.00 720	0.01 050
	Landraces	0.01 357	0.05 515
	Introgressed lines	0.0	0.0
O3fad	Wild species	0.02 260	0.02 307
	Impr. cultivars	0.00 962	0.02 038
	Landraces	0.00 544	0.02 332
	Introgressed lines	0.00 210	0.00 230

Impr. cultivars, improved cultivars; π, the expected heterozygosity per nucleotide site; θ, the number of polymorphic sites in a genotypic sample corrected for sample size.

https://doi.org/10.1017/S1479262115000106 Published online by Cambridge University Press

Table 2. Linkage disequilibrium and haplotype diversity (H_d) within the four populations of sesame

Locus	Population	R_m	S	Z_{nS}	Z_a	ZZ	Wall's B	Wall's Q	H	H_d
Sad	Wild species	0	2	1	1	0.0	1	1	0.07 031	0.00 068
	Impr. cultivars	0	13	0.1832	0.4028	0.2196	0.3333	0.4615	0.00 597	0.00 181
	Landraces	8	75	0.0781	0.2906	0.2125	0.2192	0.3108	0.00 112	0.00 311
	Introgressed lines	0	4	1	1	0.0	1	1	0.09 877	0.00 182
Fad2	Wild species	32	0	0.5699	0.6272	0.0573	0.5484	0.5983	0.03 125	0.01 168
	Impr. cultivars	51	2	0.1427	0.4461	0.3034	0.4167	0.5306	0.00 076	0.00 720
	Landraces	21	342	0.2066	0.5314	0.3248	0.4655	0.5155	0.00 002	0.01 357
	Introgressed lines	–	–	0.0	0.0	0.0	0.0	0.0	0.0	0.0
O3fad	Wild species	0	66	0.3902	0.7146	0.3244	0.6406	0.6923	0.01 600	0.02 221
	Impr. cultivars	6	99	0.2715	0.4876	0.2161	0.4457	0.5269	0.00 428	0.00 947
	Landraces	6	151	0.1391	0.5528	0.4137	0.5035	0.5563	0.00 177	0.00 538
	Introgressed lines	0	9	0.6667	0.9167	0.2500	0.8750	1.0000	0.04 948	0.00 346

Impr. cultivars, improved cultivars; R_m, minimum number of recombination events in a sample; S, number of segregating sites; Z_{nS}, average of R^2 over all pairwise comparisons; Z_a, the average of R^2 over all pairwise comparisons between adjacent polymorphic sites; ZZ, $Z_a - Z_{nS}$; Wall's B, proportion of pairs of adjacent segregating sites that are congruent; Wall's Q, measure of linkage disequilibrium between adjacent pairs of segregating sites; H, Haplotype.

divergence between the two populations (Fig. 1, row 2). The neutrality tests (Tajima's D, Fu and Li's D, and Fu and Li's F) revealed a negative but not significant result in all the three loci for wild population (Table 3). Wild species showed a positive result of Fu's F test because of deficiency of alleles. This might be the result of population bottleneck or over dominant selection or small population size taken for the analysis. π non-synonymous

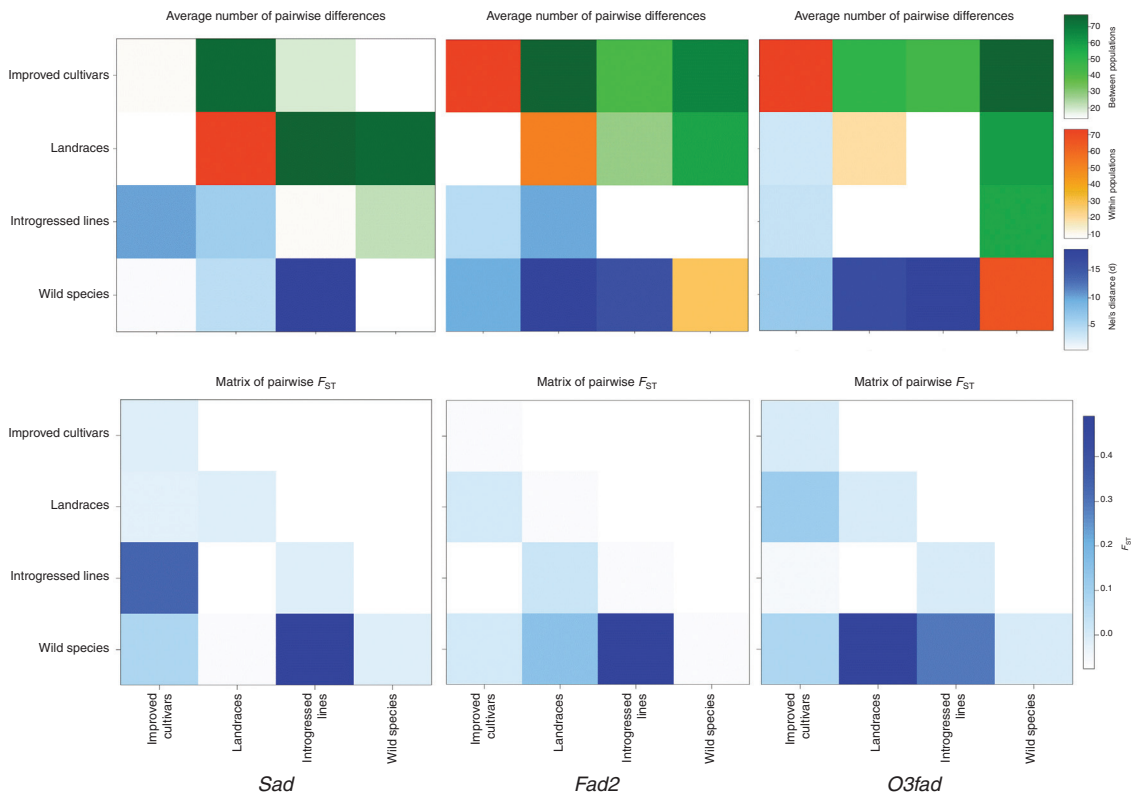

Fig. 1. Genetic diversity calculations for (a) *Sad*, (b) *Fad2* and (c) *O3fad* in four populations of sesame. First row diagrams plot the average number of population pairwise differences (i.e. the mean number of basepair differences between all pairs of sequences in a sample): above diagonal shows the average number of pairwise differences between populations, diagonal element indicates the average number of pairwise differences within population and below diagonal is the corrected average pairwise difference (Nei's distance). Second row diagrams plot matrices of population pairwise F_{ST}s. A high F_{ST} implies a considerable degree of differentiation among populations. Distance method: pairwise difference.

https://doi.org/10.1017/S1479262115000106 Published online by Cambridge University Press

Table 3. Neutrality tests for *sad, fad2* and *o3fad* loci within the four populations of sesame

Locus	Population	Tajima's *D*	Fu and Li's *D*	Fu and Li's *F*	Fu's *F* stats
Sad	Wild species	−0.70 990	−0.70 990	−0.60 427	1.099
	Impr. cultivars	−1.31 358	−1.70 351	−1.83 832	−3.113
	Landraces	−2.46 937**	−4.25 580*	−4.25 692*	−17.973
	Introgressed lines	n.d.	n.d.	n.d.	n.d.
Fad2	Wild species	−0.66 476	−0.66 476	−0.69 923	0.892
	Impr. cultivars	−1.38 525	−0.32 127	−0.74 508	−2.174
	Landraces	−2.76 192***	−5.92 034*	−5.54 868*	−23.901
	Introgressed lines	n.d.	n.d.	n.d.	
O3fad	Wild species	−0.28 204	−0.35 375	−0.36 827	1.002
	Impr. cultivars	−2.19 936**	−2.57 946*	−2.87 627*	0.735
	Landraces	−2.65 797***	−5.07 831*	−4.92 190*	−12.865
	Introgressed lines	−0.80 861	−0.80 861	−0.77 723	2.944

Impr. cultivars, improved cultivars; n.d., not detectable.
Values were significantly different (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

(π non-syn) was also more in wild population in all the three loci when compared with improved cultivars, supporting the reduction in genetic diversity as expected (Table 4). The number of alleles was high in the improved cultivars when compared with the wild relatives in all three loci (Fig. 2). The values of LD, Z_{nS} and Z_a decreased in improved cultivars when compared with their wild forms. The ZZ value depicted a decrease only in *o3fad* locus and not in other two loci of improved cultivars compared with their wild relatives, probably because of the less number of accessions of wild forms taken for study. Moreover, Wall's *B* and Wall's *Q* reduced remarkably in value in case of improved cultivars against wild species (Table 2).

The landraces had an average π value of 0.0074 and a θ value of 0.0299 (Table S4, available online) and were less diverse than the wild species, while it was more diverse than the improved cultivars. The landraces retained 63.26% (π) of diversity present in wild population. H_d showed a retention of 63.81% in landraces from wild species (Table 2). Landraces had a high average number of pairwise differences with wild species in all three loci (Fig. 1, row 1). Moreover, *o3fad* locus showed that there is a very high reduction within population of landrace compared with that of wild. The F_{ST} pairwise matrix of *fad2* and *o3fad* loci showed a very high value (0.2–0.4) resulting in high divergence between the two populations (Fig. 1, row 2). The landraces showed a very high number of alleles compared with the wild species (Fig. 2). All the neutrality tests (Table 3) in landraces gave significantly negative results indicating an excess of rare polymorphism in the population, which is consistent with either positive selection or an increase in population size. The values of Z_{nS}, ZZ and Z_a decreased in landraces when compared with wild species. Wall's *B* and Wall's *Q* of landraces also showed less value than those of wild species (Table 2).

Table 4. π non-synonymous (π non-syn) and π synonymous (π syn) of the three loci within four populations of sesame

Locus	Population	π syn	π non-syn	π non-syn:π syn
Sad	Wild species	0.0	0.00 090	–
	Impr. cultivars	0.00 529	0.00 035	0.0661
	Landraces	0.00 409	0.00 216	0.5281
	Introgressed lines	0.00 425	0.00 060	0.1411
Fad2	Wild species	0.01 797	0.00 872	0.487
	Impr. cultivars	0.00 377	0.00 661	1.753
	Landraces	0.01 248	0.01 296	1.083
	Introgressed lines	0.0	0.0	0.0
O3fad	Wild species	0.03 158	0.01 657	0.524
	Impr. cultivars	0.00 913	0.00 933	1.021
	Landraces	0.00 588	0.00 528	0.897
	Introgressed lines	0.00 739	0.00 206	0.278

Impr. cultivars, improved cultivars.

https://doi.org/10.1017/S1479262115000106 Published online by Cambridge University Press

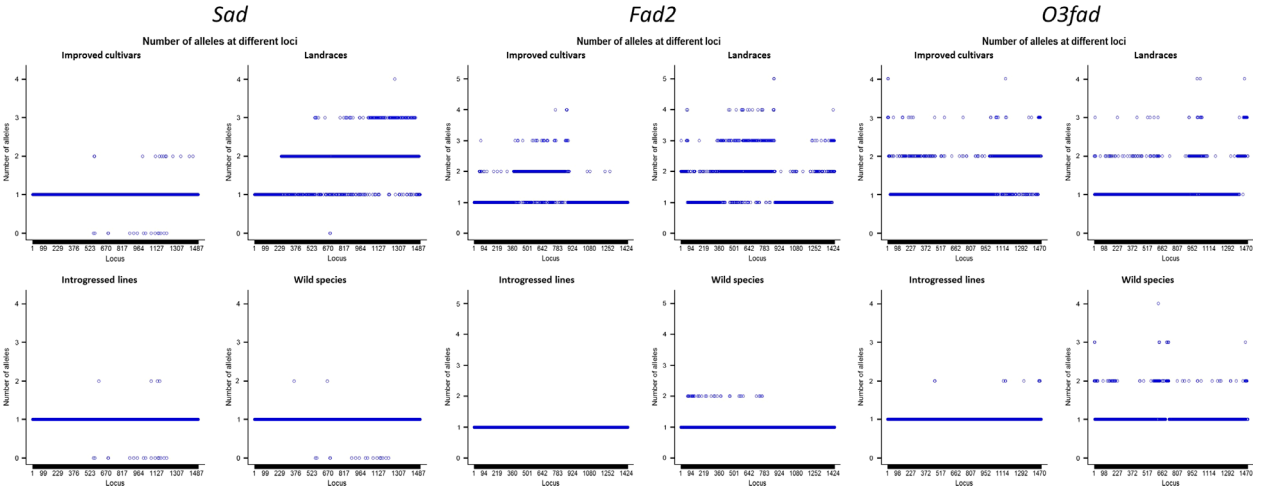

Fig. 2. A graphical presentation of variation in allele frequency at different loci in four populations of sesame using analysis of molecular variance (AMOVA).

Effect of selection on diversity

Improved cultivar versus landrace germplasm

It is generally expected that intensive artificial selection imposed in modern plant breeding programs over the last century has reduced the genetic diversity from that present in the landraces. In this study, a comparison of the improved cultivars with the landraces indicated, on an average, a reduction in nucleotide diversity in improved cultivars for both π and θ, though not significantly (Table S4, available online).

Haplotype is a term used to designate a specific combination of linked alleles within a contiguous segment of DNA, and thus H_d provides another measure for genetic diversity (Hyten *et al.*, 2006). The H_d data of improved cultivars depicted a decrease in *fad2* and *sad* loci when compared with those of landraces (Table 2). In improved cultivars' *o3fad* locus, there was an increase in the H_d. These data were supported by the coalescent simulation, which was run 10^3 times, according to which, the H_d in improved cultivars' *sad* and *fad2* showed a decrease while that of *o3fad* showed an increase, in comparison to those of landrace population. The average number of pairwise difference between the two populations is very high in *sad* and *fad2,* while in *o3fad* it is moderately high (Fig. 1, row 1). Moreover, the number of alleles in all the three loci showed a higher value in landraces compared with improved cultivars (Fig. 2).

The pairwise neutrality tests, Tajima's *D*, Fu and Li's *D*, and Fu and Li's *F* data revealed a significant negative result in case of landraces for all the three loci (Table 3). The values ranged between −2.4693 (*sad*) and −2.7619 (*fad2*) for Tajima's *D*; −4.2558 (*sad*) and −5.9203 (*fad2*) for Fu and Li's *D*, and −4.2569 (*sad*) and −5.5486 (*fad2*) in case of Fu and Li's *F*. The significant negative values indicated an excess of rare polymorphism in the population, which might be consistent with either positive selection or an increase in population size. Improved cultivars also showed a significant negative result for *o3fad* locus, while *fad2* and *sad* loci depicted a non-significant negative value. Fu's *F* analysis exhibited negative results for all the loci implying that there was an excess of alleles due to population expansion or genetic hitchhiking in both improved cultivars and landraces. The F_{ST} pairwise matrix of all the three loci showed a low value (0.0–0.1) causing mild divergence between the two populations (Fig. 1, row 2).

Landrace versus introgressed germplasm

The results indicated a significant reduction in nucleotide diversity in introgressed lines when compared with that of landraces. On an average, introgressed lines retained only 17.77% (π) and 4.57% (θ) of landrace diversity, respectively (Table S4, available online). This could be supported by the H_d of the three loci as all of them showed a decreased value in introgressed lines when compared with landraces (Table 2).

The neutrality tests, Tajima's *D*, Fu and Li's *D*, and Fu and Li's *F* values of landrace and introgressed lines population showed a significant difference in all three loci (Table 3). The data supported either a positive selection or an increase in population size in case of landrace. Fu's *F* value for all the three loci in landrace population provided negative results implying population expansion or genetic hitchhiking. The same could not be predicted for introgressed lines as *fad2* and *sad* were undetectable and *o3fad* value of Fu's *F* was positive.

https://doi.org/10.1017/S1479262115000106 Published online by Cambridge University Press

Selection in fatty acid desaturase loci

The genetic diversity in terms of π and θ was generally reduced in improved cultivars when compared with wild population. This reduction was recorded in two loci, *o3fad* and *fad2,* of improved cultivars, while *sad* locus depicted a considerable increase in these values (Table 1). There was also a decrease in H_d in these two loci of improved cultivars ($H_d = 0.00\,947$ and $0.00\,720$, respectively) when compared with that of wild species ($H_d = 0.02\,221$ and $0.01\,168$ respectively) (Table 2). In case of *sad* locus, there was an increase in improved cultivars ($H_d = 0.00\,181$) than that of wild species ($H_d = 0.00\,068$). Moreover, in both *o3fad* and *fad2* loci, the ratio of π non-synonymous:π synonymous (π non-syn:π syn) showed an increase in improved cultivars when compared with wild species. However, *sad* locus did not follow the trend as the ratio for wild was zero (Table 4). The neutrality tests of Tajima's *D*, Fu and Li's *F*, and Fu and Li's *D* presented negative values for all the three loci in wild as well as improved cultivars (Table 3). Comparison of these values in the two populations suggested that there was a remarkable decrease in improved cultivars compared with wild species. Only the *o3fad* gene of improved cultivars showed a significant decrease (away from neutral) in these values (Table 3). Even the Fu's Fs Stats showed a decrease in their values for the three loci in improved cultivars compared with wild species. Though not significant, the pairwise F_{ST} matrix of improved cultivar, wild in all three loci, suggested high divergence between the two populations (Fig. 1, row 2).

Discussion

Nucleotide diversity in sesame

This study represents a comprehensive analysis of DNA sequence in germplasm of four diverse populations of Indian sesame germplasm, namely wild species, landraces, improved cultivars and introgressed germplasm. Although there was locus-to-locus variation in terms of π and θ values, it was apparent that wild sesame contains substantial levels of nucleotide diversity. Indeed, wild sesame appears to harbour ten times more diversity ($\pi = 0.01\,165$; $\theta = 0.01\,210$) compared with that of many other crops including *Glycine soja* [$\pi = 0.00\,217$ and $\theta = 0.00\,235$; (Hyten *et al*., 2006)], wild *Arabidopsis thaliana* for *PgiC* locus [$\pi = 0.003$ and $\theta = 0.0065$; (Kawabe *et al*., 2000)] and loblolly pine [$\pi = 0.0040$ and $\theta = 0.0041$; (Brown *et al*., 2004)]. By contrast, improved cultivated sesame contains markedly less nucleotide variation, which showed an average π value of 0.0062 and a θ value of 0.0112 (Table S4, available online). When compared with other organisms, the average π value of improved sesame cultivars was similar to that of flanking region of the *sad2* gene of *Linum usitatissimum* [$\pi = 0.00\,650$; (Fu, 2011)]. Moreover, the nucleotide diversity was similar to the θ values of *GA20ox1* locus of *Populus tremula* [0.00 113 (Ingvarson, 2005)].

Domestication bottleneck

Crop cultivars are developed from wild progenitors through domestication and intensive selection for various features during breeding. It has been well documented that selection targeted at individual loci will reduce genetic diversity within and around the selected loci (Halliburton, 2004). In this study, the domestication was supported by a reduction in genetic diversity of improved cultivars when compared with landraces (*fad2* and *sad*) as well as wild population (*o3fad* and *fad2*). Overall, the effects of domestication combined with intensive selection in breeding have resulted in sequence diversity losses, in improved cultivars versus wild population, of 46.61% π, 7.44 % θ and 44.5% H_d for all three loci, on an average. Similar results have been reported while studying domestication history of soybean (Hyten *et al*., 2006), where the cultivars versus wild resulted in sequence diversity losses of 49% of π, 65% of θ and 44% of H_d. Though the value of θ in this study did not show any significant loss of diversity, the rest of the values (including 39.10% loss of π non-syn) indicated a bottleneck in sesame domestication.

The landraces of sesame developed by farmers in different agroecological niche environment are long-run adaptations and selections in wild sesame species. These local cultigens were expected to have less genetic diversity during the course of time due to domestication. In this study, landraces retained 63.26% (π) of diversity present in wild progenitor. Similar results have been reported in soybean, in which there was 66% retention of π in landraces compared with the wild population (Hyten *et al*., 2006). But sesame landraces showed an increase in θ when compared with that of wild relatives probably because the number of wild relatives taken for study is few compared with landraces. The H_d value also suggested a loss of genetic diversity in landraces (0.0073) when compared with wild species (0.0115), which was 63.47% retention. This is in agreement with that of soybean where 63% retention of H_d in landraces has been reported. Hence, the overall reduction in genetic diversity of landraces and improved cultivars suggested that genetic bottleneck had indeed occurred during the domestication of sesame due to the conscious

https://doi.org/10.1017/S1479262115000106 Published online by Cambridge University Press

and unconscious selection for oil quality in sesame by farmers over time.

It has been observed that the process of domestication leads to an increase in LD. But this study showed otherwise. Z_{nS}, Z_a and ZZ values decreased in improved cultivars and landraces when compared with the wild species. Same could be said about the values of Wall's *B* and Wall's *Q*. Main reason for decreased LD might be due to presence of frequent cross pollinations in landraces, which may result in recombinations between loci and thereby decreased LD. Further reason might be that sesame wild germplasm (seven species) in India is limited when compared with the 35 total species present in the Africas (Duhoon *et al.*, 2004). Thus, the genetic base for the domestication of sesame was very small leading to low LD in landraces and improved cultivars even after domestication.

The results of this study also suggested that there was no significant reduction in genetic diversity in improved cultivars compared with landraces. This was probably because involvement of hybridization was relatively less in frequency, during the development of improved cultivar from landraces. Majority of improved cultivars included in this study (12/20) were developed by direct selection from local collections/landraces, thereby resulting in direct representation of landrace diversity in released varieties (Table S1, available online).

There was very significant loss of diversity in the introgressed line population compared with all other populations included in this study. This might be explained by the fact that, through conventional method, introgressed line germplasm is developed during prebreeding by hybridization of selected germplasm lines followed by intensive selection for trait(s). The method reduced genetic diversity significantly due to imposed selection pressure. Moreover, three out of four introgressed lines, included in this study, were the same across combination and the size of the introgressed line population was too small resulting in reduced genetic diversity as expected.

The results revealed that not all the genes depicted a similar diversity loss in improved cultivars compared to wild relatives and landraces. Particularly, *o3fad* showed higher diversity in improved cultivars than in landraces; *sad* showed very high diversity in improved cultivars. These results indicated that, during development of improved cultivars from landraces, selection was not targeted for this gene or its related genes. Similarly, in the second case (*sad*), there was either no or limited domestication or selection effect on the improved cultivars. Overall, results did highlight that despite domestication and selection, the Indian improved cultivars retained diversity in genes or traits to an extent suitable for cultivation in diverse agroclimatic regions of the country. This trend was also evident in earlier reports, such as the Indian cultivated forms that were rich in diversity (Duhoon *et al.*, 2004), and the collections showed higher molecular (random amplified polymorphic DNA; RAPD) diversity compared with that of exotic (Bhat *et al.*, 1999) varieties. One of the possible reasons for this trend could be that the majority of the Indian cultivars (37/62) were developed through selection from local collections/landraces (Duhoon *et al.*, 2004) to suit diverse local agroclimatic regions of the country.

Selection in fatty acid desaturase loci

This study included three desaturases, out of which two key genes involved were *fad2* and *o3fad,* which catalyse the conversion of oleic acid into linoleic acid and linoleic into linolenic acid, respectively. The present results suggested that there was positive selection in them during domestication and/or improvement of cultivar in sesame. *O3fad* showed a reduction in genetic diversity as there was a loss of 57.44% of π, 11.67% of θ and 57.37% of H_d in improved cultivars when compared with wild population. Similarly, in *fad2* locus, there was a reduction in 38.36% π, 16.00% θ and 38.47% H_d in improved cultivars. Earlier, Maynard and Haigh (1974) stated that a strong positive selection leads to a loss of nucleotide diversity.

The ratio of π non-syn:π syn was 1.021 for *o3fad* and 1.753 for *fad2,* which further supported the theory of positive selection for the two loci during the course of domestication. Similar results have been recorded in *ods* locus of *Drosophila* (Ting *et al.*, 1998), *het-c* of Sordariaceae fungi (Wu *et al.*, 1998), S locus of various plants (Clark and Kao, 1991), 19 seminal protein genes of *Drosophila* (Swanson *et al.*, 2001) and *RPP1*complex, *RPP8* complex, *RPP13* of *Arabidopsis thaliana* (Botella *et al.*, 1998, McDowell *et al.*, 1998, Bittner-Eddy *et al.*, 2000, respectively) and in several others. These studies lead to direct postulation of positive selection among these genes because their ratio of π non-syn:π syn was more than 1, which further supports the present data. Moreover, the matrix of pairwise F_{ST} for all three loci in improved cultivar and wild species showed a significant elevation from zero. This suggested high divergence between two populations (Fig. 1, row 2).

Tajima's *D*, Fu and Li'*D*, and Fu and Li's *F*, along with Fu's *F* stats, were performed for studying neutrality in detail. Tajima's *D*, Fu and Li's *D*, and Fu and Li's *F* gave negative result for the improved cultivars of both *o3fad* and *fad2,* suggesting an excess of rare variant relative to neutral equilibrium model, and was thus consistent

https://doi.org/10.1017/S1479262115000106 Published online by Cambridge University Press

with a positive selection. But the values were significant only for improved cultivars of *o3fad*. This could be explained with the fact that demography was not considered in these loci. The population structure especially variation in the number of samples in different populations also affected the results.

The results suggested that there was a positive selection in *o3fad* and *fad2*. The phenotypic data also supported this result because improved cultivars showed an increase in both linoleic and linolenic acid contents (%) compared with wild species (Table S3, available online).

The observed loss of diversity from wild to improved cultivars was likely to, at least in part, cause population bottlenecks during the domestication of sesame. Despite domestication, the improved cultivated sesame showed variable results with different genes in this study. This can be supported by the fact that sesame has been considered as a semi-domesticated crop having still few important characters of wild species (Bedigian, 2003).

Plant domestication, breeding and biotechnology have modified sesame genomes according to the needs of humanity with increasing efficiency and precision. Understanding such processes, crop domestication in particular, is crucial today because of the rising demand for improving yield and quality, as well as to obtain sesame with additional capability to grow in varied ecosystems. The background information obtained from this study indicated that, during crop improvement, selective genes were to be considered based on the domestication dynamics. Moreover, an increased number of wild species from around the world is needed to be included for further study to enhance the possibility of incorporation of desirable genes in cultivars from wild relatives.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262115000106

Acknowledgements

The authors thank Dr P. Dhureja, Agricultural Chemicals, IARI, for laboratory facilities and Mr R. Tanwar for technical assistance in performing gas chromatography for this study. Financial assistance to the senior author in the form of Junior and Senior Research Fellowship by the University Grant Commission, Government of India, is duly acknowledged. The authors also thank NBPGR and National Agricultural Innovation Project for supporting this work and paper.

References

Bedigian D (2003) Evolution of sesame revisited: domestication, diversity and prospects. *Genetic Resources and Crop Evolution* 50: 779–787.

Bhat KV, Babrekar PP and Lakhanpaul S (1999) Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *Euphytica* 110: 21–33.

Bittner-Eddy PD, Crute IR, Holub EB and Beynon JL (2000) RPP13 is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant Journal* 21: 177–188.

Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, Holub EB and Jones JDG (1998) Three genes of the Arabidopsis RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell* 10: 1847–1860.

Brown GR, Gill GP, Kuntz RJ, Langley CH and Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences* 101: 15255–15260.

Clark AG and Kao TH (1991) Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proceedings of the National Academy of Sciences* 88: 9823–9827.

Duhoon SS, Sharma SM, Lakhanpaul S and Bhat KV (2004) Sesame. In: Dhillon BS, Tyagi RK, Saxena S and Agrawal A (eds) *Plant Genetic Resources: Oilseed and Cash Crops*. Indian Society of Plant Genetic Resources, New Delhi, India: Narosa Publishing House, pp. 136–145.

Excoffier L and Lischer HEL (2010) Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Fu YX and Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.

Fu Y-B (2011) Genetic evidence for early flax domestication with capsular dehiscence. *Genetic Resources and Crop Evolution* 58: 1119–1128.

Halliburton R (2004) *Introduction to Population Genetics*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Hill WG and Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38: 226–231.

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

Hyten D, Song Q, Zhu U, Choi IY, Nelson RL, Costa JM and Specht JE (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences USA* 103: 16666–16671.

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

Kawabe A, Yamane K and Miyashita NT (2000) DNA polymorphism at the cytosolic phosphoglucose isomerase (PgiC) locus of the wild plant Arabidopsis thaliana. *Genetics* 156: 1339–1347.

Kelly JK (1997) A test of neutrality based on interlocus associations. *Genetics* 146: 1197–1206.

Librado P and Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Maynard SJ and Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetics Research* 23: 23–35.

McDowell JM, Dhandaydham M, Long TA, Aarts MG, Goff S, Holub EB and Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis. *Plant Cell* 10: 1861–1874.

Mondal N, Bhat KV and Srivastava PS (2010) Variation in fatty acid composition in Indian germplasm of sesame. *Journal of the American Oil Chemists' Society* 87: 1263–1269.

Nei M (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nzikou JM, Matos L, Bouanga-Kalou G, Ndangui CB, Pambou-Tobi NPG, Kimbonguila A and Desobry S (2009) Chemical composition on the seeds and oil of sesame (*Sesamum indicum* L.) grown in Congo-Brazzaville. *Advance Journal of Food Science and Technology* 1: 6–11.

Rozas J, Gullaud M, Blandin G and Aguadé M (2001) DNA variation at the *rp49* gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. *Genetics* 158: 1147–1155.

Saghai-Maroof MA, Soliman KM, Jorgensen RA and Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences* 81: 8014–8018.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF and Aquadro CF (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proceedings of the National Academy of Sciences* 98: 7375–7379.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437–460.

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

Tanksley SD and McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277: 1063–1066.

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

Ting CT, Tsaur SC, Wu ML and Wu CI (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282: 1501–1504.

Wall JD (1999) Recombination and the power of statistical tests of neutrality. *Genetics Research* 74: 65–69.

Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theoretical population biology* 7: 256–276.

Wu J, Saupe SJ and Glass NL (1998) Evidence for balancing selection operating at the *het-c* heterokaryon incompatibility locus in a group of filamentous fungi. *Proceedings of the National Academy of Sciences* 95: 12398–12403.