ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

# Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

OPEN ACCESS Biotechnology

ISSN 1682-296X
DOI: 10.3923/biotech.2022.90.102

# Research Article
# Patterns of Interspecific Hybridization among Crop and Wild Sorghum in Open-Pollinated Fields

[1,4]Titus O. Magomere, [1]Eliud K. Ngugi, [2]Solomon I. Shibairo, [1]Eunice Mutitu and [3]Silas D. Obukosia

[1]Department of Plant Science and Crop Protection, University of Nairobi, P.O. Box 29053-00625, Nairobi, Kenya
[2]Masinde Muliro University of Science and Technology, P.O. Box 1699-50200, Bungoma, Kenya
[3]Africa Union Development Agency, New Partnership for Africa's Development, Regional Office for Eastern Africa, Nairobi, Kenya
[4]Department of Biochemistry Microbiology and Biotechnology, School of Pure and Applied Sciences, Kenyatta University, P.O. Box 43844, Nairobi, Kenya

## Abstract

**Background and Objective:** Spontaneous temporal and spatial interspecific hybridization between crops and their weedy progenitors may have potential environmental risks and implications in the adoption of genetically modified crops. This study determined the pattern of gene flow from *S. bicolor* to *S. halepense*, *S. sudanense* and *S. bicolor* ssp. *verticilliflorum* in open-pollinated field plots. **Materials and Methods:** Concentric field designs were planted to evaluate the pattern of gene flow from crops to weedy sorghums. *Sorghum halepense*, *Sorghum sudanense* and *Sorghum verticilliflorum* seed were sampled. Crop alleles in the weedy backgrounds were analysed using PCR analysis of SSR loci. A regression model with log transformation on geneflow was evaluated. **Results:** Results show the presence of spontaneous interspecific hybridization in unlimited flow design (UFD) plots. Up to 61% hybridization was observed at 5 m, 30% at 70 m and 45% at 90-100 m. A regression model $\log_{10}Y = \log_{10}A+Bx$ was determined to best fit and explain the short distance and long-distance gene flow. Species variations were seen in the flow of genes from crop to weedy sorghums in UFD plots. More interspecific hybridization events were observed above 100 m from pollen source in *S. halepense* (14.7%) and *S. sudanense* (18.8%) than in *S. verticilliflorum* (1.4%). **Conclusion:** There exist species differences in the flow of genes from crop to weedy sorghums. Therefore, isolation distances of more than 200 m and early detection of interspecific hybridization events using effective sampling and detection approaches are necessary to mitigate the potential risks.

Key words: Interspecific hybridization, gene flow, sorghum, weeds, PCR, cross-pollination, tillers

Citation: Magomere, T.O., E.K. Ngugi, S.I. Shibairo, E. Mutitu and S.D. Obukosia, 2022. Patterns of interspecific hybridization among crop and wild sorghum in open-pollinated fields. Biotechnology, 21: 90-102.

Corresponding Author: Titus O. Magomere, Department of Biochemistry Microbiology and Biotechnology, School of Pure and Applied Sciences, Kenyatta University, P.O. Box 43844, Nairobi, Kenya



Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

## INTRODUCTION

Flowering plants show varying degrees of outcrossing and spontaneous interspecific hybridization between domesticated and wild species[1]. Allogamous species tend to exhibit high intra-species outcrossing rates in agricultural fields. Spontaneous hybridization in crops and among crops and their wild relatives has potential risks and implications in the adoption of genetically modified crops. The interspecific hybrids would become more aggressive or invasive and thus cause direct hazards to humans or domestic animals[2].

Pollen flow has been shown to vary with distance, wind direction and year in which the experiments are performed[3]. In sorghums[4,5], previous studies have demonstrated the existence of natural pollen mediated geneflow within the conspecific crop and wild populations. In *S. sudanense,* natural cross-pollination was recorded at 0.1-13% and 0.5-9% in R-lines and B-lines respectively. Outcrossing ranged from 0 to near 100% on individual sudangrass plants and was highly variable[6]. In certain cases, outcrossing averages of 39 and 57% were demonstrated when most tillers were at anthesis[6].

Spontaneous outcrossing is reported in sorghum where intra-species hybridization or outcrossing rates from fertile to male-sterile *S. bicolor* plants has been shown to occur at 2.54% at 13 m, below 1% at a distance of 26 m or greater and eventually dropping to 0.06% at 158 m[7]. Hybridization of 50% occurred between male sterile and male fertile crop sorghums grown in concentric designs 1 m from the source, 14% at 10 m and reached 1% at 60 m from the pollen donor[8].

Temporal and spatial gene flow evaluation in the cross and self-pollinated crop species may provide information on the magnitude of transgene movement from crops to the wild species. Several factors influence the rate of gene flow, including reproductive biology, environmental conditions, pollen structure and stability, fluctuations in pollinator, plant abundance, density, distribution and cropping system[9]. Several approaches are normally used to estimate potential or realised gene flow via pollen dispersal, including, regression models based on empirical information[10], probabilistic models, grid-based models and mechanistic approaches[11].

Regression models have been extensively applied in explaining the relationship between gene flow and distance in experimental plots. Previous studies[8] have deduced a negative exponential regression model where Pollen Mediated Gene Flow (PMGF) and distance from the pollen source field were transformed through logarithmic and square-root transformation, respectively. The model explains the factors that influence the flow of pollen from a local Ochuti landrace to a pollen bait sorghum male-sterile crop. The maximum distance of gene flow was estimated at 203 m from the source field. Mathematical models (double log transformation and exponential) have been applied to estimate the maximum gene flow distance[7]. The maximum distance of gene flow from male fertile sorghum plants to male-sterile bait plants was shown to be approximately between 200 and 700 m. In other cereals like rice[12], wheat[13], oilseed rape[14] and maize[15] mathematical models have also been severally applied in explaining gene flow in open-pollinated fields in crops and in predicting hybridization patterns[3]. Species related and environmental factors have an impact on the introgression of crop genes into the wild plants[15].

An information gap exists on the patterns of interspecific hybridization and geneflow among weeds and crops in the sorghum genus. This study evaluated the influence of the distance from the edge of the pollen source, weedy type, location and wind direction on gene flow from *S. bicolor* to *S. sudanense*, *S. halepense* and *S. bicolor* ssp. *verticilliflorum* in agricultural fields. Experiments were therefore done to determine the pattern of gene flow from *S. bicolor* to *S. halepense*, *S. sudanense* and *S. bicolor* ssp. *verticilliflorum* in open-pollinated field plots. In addition regression model with log transformation on gene flow from *S. bicolor* to *S. halepense*, *S. sudanense*, *S. bicolor* ssp. *verticilliflorum* was evaluated.

## MATERIALS AND METHODS

Three locations were selected for the study in the year 2012. The first two were in Western Kenya at Alupe (+0° 28' 45.07", +34° 7' 11.39") and Kakamega (+0° 12' 39.50", +34° 57' 34.11") while a third was sited at the College of Agriculture and Veterinary Sciences (CAVS) (-1° 14' 59.72", +36° 44' 30.79") of the University of Nairobi. The soils and climatic conditions in the three sites favour sorghum production for both small scale and large scale producers. These sites represent sorghum production regions in high altitude areas of the country.

Two field sites at Alupe and Kakamega were utilized to isolate the role of various factors responsible for the movement of crop genes to conspecific wild and weedy sorghums. The position of the *S. bicolor* pollinators relative to *S. sudanense*, *S. halepense* and *S. bicolor* ssp. *verticilliflorum* was used to evaluate gene flow. A concentric design on a plot of 210 m in diameter was used to estimate the maximum distance of pollen flow from crop sorghum to the weedy sorghums was designed as follows: *S. bicolor* was planted in the centre and *S. sudanense*, *S. halepense* and *S. bicolor* ssp. *verticilliflorum* were each planted to radiate around the

centre plot at 0.5, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 and 200 m. Additional weedy plants were grown at 200 m from the pollen source. The experiment was replicated in two sites.

A 10×10 m central plot was sowed with crop sorghum (*S. bicolor*- BTX623) at a spacing of 0.3 m between plants and 0.5 m within rows. The weedy sorghums were planted along with compass directions (North, North-East, North-West, West, South, South-East, South-West and East) in radial circles. The weedy species were sown at a spacing of 0.3 m between plants and 0.5 m between rows in 1×3 m plots. Three rows of weedy species with 11 plants per row were planted at each of the subplots positioned at 0.5, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 m from the central plot. *S. bicolor* was planted in 2 weeks stagger to synchronize the flowering period with that of the weedy species which were planted later.

The plots were sown with 20-40 kg $ha^{-1}$ of N (DAP) and top-dressed with 50-80 kg $ha^{-1}$ of N (CAN). Mites were controlled by Ortus®, SC 5% (fenpyroximate) foliar spray (0.05 kg $ha^{-1}$). The pesticide rates of 0.28 kg $ha^{-1}$ of Dursban® 50W (chlorpyrifos) were used to control Lepidopteran and Homopteran. Panicles of each of the three weedy species and the crop were harvested and transported to the laboratory for analysis.

**DNA analysis and PCR:** Seeds were collected from both the main stems and tillers of mature plants from open-pollinated *S. bicolor and S. bicolor* ssp. *verticilliflorum*, *S. halepense* and *S. sudanense* plots under UFD designs in Western Kenya. Main stems and tillers were sampled to allow for differences in pollinations among species that have extended flowering due to temporal extensive tillering. The seeds were packed in labelled packets and dried under greenhouse conditions for 3-4 weeks. Harvested seeds from each of the three weedy species, from each of the plots in UFD designs, were bulked into two samples. This was done to increase the chance of detecting a hybridization event on the loci evaluated during PCR analysis. Mature seeds (approximately, 30) were harvested dried and threshed. Endosperms and the zygotic embryos were ground into fine flour using a kitchen grinder. A modified CTAB-based seed extraction protocol[16-18] was utilized. A 2 mL Eppendorf tube containing 1.5 mL of extraction buffer was placed in a water bath at 65 . About 0.3 g of ground flour was transferred into the tube. The tube was inverted five times to mix the contents before incubating at 65°C for 10 min. During the incubation period, the contents were mixed gently by inversion at least twice. After incubation, the 2 mL tubes were centrifuged at approximately 7200×g for 10 min at room temperature. The supernatant (~1.2 mL) was transferred to a clean 2 mL Eppendorf tube and nucleic acids were extracted with an equal volume of chloroform: Isoamyl alcohol (24:1). The upper aqueous phase (~1 mL) was collected after centrifugation (7200×g for 10 min). Pre-warmed (55) 10% CTAB Solution was added at L/l0th of the volume of the collected aqueous phase (~100 µL). This was mixed well by gentle inversion. A second extraction was done with an equal volume of chloroform: Isoamyl alcohol (24:1), by inverting the tube 20 times. The mix was centrifuged at approximately 7200×g for 10 min at room temperature. The upper aqueous layer was pipetted into a fresh 2 mL tube and 3 volumes of precipitation buffer were added. This was mixed gently and incubated at room temperature for 10 min. The DNA pellet was collected by centrifugation at approximately 7200×g for 15 min at room temperature. The buffer solution was discarded and the pellet was washed twice with 1 mL of 70% ethanol. The pellet was stirred from the Eppendorf wall and gently swirled in the ethanol for 10 sec to wash off the salts.

The contents were centrifuged at approximately 7200×g for 5 min at room temperature and ethanol was removed carefully by decanting. Washing the pellet was repeated once. Residual alcohol was allowed to evaporate (pellet drying) then the DNA pellet was dissolved in 100 µL TE. To obtain good elution the mix was incubated at 4 for 2 hrs. The sample was further incubated at 65 for 15 min and mixed gently to ensure complete pellet elution. In the case of un-dissolved cell debris, the sample was centrifuged at maximum speed for 5 min to pellet the debris. The supernatant was pipetted into a clean tube and stored at -20. Primers were designed based on the polymorphic SSR sequences from crop sorghum physical map genomic clones[19-21] on five linkage groups (chromosomes) as shown in Table 1. The primers were selected based on their polymorphism and ability to distinguish the crop and wild sorghum accessions.

The polymerase chain reaction was done in 0.5 mL reaction tubes with a hybaid® thermocycler. Reaction volumes of 11 µL were used in all experiments. The reaction conditions were set in conventional PCR format using Invitrogen™ PCR reagent system, Taq DNA polymerase with (W-1) Invitrogen™ and 10 mM dNTP Mix, PCR grade Invitrogen™ (As per manufacturer recommendations). The PCR products were separated and analysed using a 4% UltraPure™ Agarose, Invitrogen™ (As per manufacturer recommendations). Agarose gel electrophoresis for genomic DNA extracted from the young leaf, old leaf and seed tissues was done before PCR to evaluate its quantity through densitometric analysis and quality.

**Modelling of gene flow in UFD designs:** The percent gene flow data from UFD design plots in site 1 and 2 were plotted

Table 1: Loci and sequences of primers used to detect crop alleles in weedy genotypes

| Marker names | Forward primer (5'->3') | Reverse primer (5'->3') |
|---|---|---|
| SB1764 | CTTGTGCTTGCTTGCACCATATTC | GTCGATGAGGAGCTTCATGCTCAG |
| SB3420 | GAGCCAGCATGCATGATAATTGTT | CACAAAGGCATGACAGTCAATCAA |
| SB5058 | GAGAATTGGAAGAAAGCCTCGGTT | CAGAGCTCCTAAACGGTCCTCAAA |
| SB5458 | AATGTGGTGGTGTTGTGTCTCCAT | TCTACTGCTATCATCGCCTCCACC |
| SB3258 | ATTGTTGTCCTCCCCTCCACC | AGTACCTGAACCAGGCGTCGCT |

as a function of distance. Linear regressions (y = A+Bx) were fitted on the relationship between gene flow and distance in UFD plots. Where A is the percent gene flow immediately adjacent to the source and x is the distance from the edge of the source (in meters). B is an empirical factor, possibly related to several physical and biological factors such as wind speed, wind direction, pollen size, species, site, flower structure and biological compatibility. Non linear regressions and curves fitted on the relationship between gene flow and distance in UFD plots included, non linear regressions:

$$y = \frac{A + B}{1 + D \times X}$$

exponential relationship:

$$Y = A+B*(R**X)$$

and double exponential curve:

$$Y = A+B*(R**X)+C*(S**X)$$

A log-transformed regression model $\log_{10}Y = \log_{10}A+BX$ was tested. The models were tested using GEN STAT 14 (VSN 2012). Log transformation of the y variate normalized the residuals on rug plots and improved the power of the F test in evaluating the model in the three datasets in site 1, site 2, site 1 and site 2. Meteorological data was obtained from the Kenya meteorology department. Meteorological parameters are useful in estimating gene flow in wind-pollinated allogamous species in open-pollinated field designs. Data were obtained on wind speed (m/s) and wind direction for a 2 month period during which natural pollination took place in the experimental fields. Three daily readings on the two parameters were obtained.

**Data analysis:** The results were scored based on the absence or presence of specific bands representing each of the species. Polymorphisms were scored on the differential size of the bands amplified per reaction. Maximum distance of gene flow from the *S. bicolor* centre plot to *S. bicolor* ssp. *verticilliflorum*, *S. halepense* and *S. sudanense* peripheral plots were measured using hybridization frequencies. These hybridization frequencies in the field were estimated using the crop SSR loci amplified in the weedy populations. Effective meteorological data considered the time of the year when pollination occurred on main stems/tillers and the time of the day when flowers were open for natural pollination. Effective meteorological data was analyzed to include wind speed and direction parameters that were important during natural pollination. To obtain the maximum distance of gene flow distances to both weeds in the 'UFD' analysis of variance was done and mean differences were separated at $p \leq 0.05$. Regression analysis and parameter correlations for geneflow, distance from the pollinator, plant direction, wild species types, wind direction and wind speed were calculated using GEN STAT 14 (VSN 2012).

## RESULTS

Plant genomic DNA extraction from seed consisting of the endosperm and the zygote embryo portion yielded between 180-300 ng of genomic DNA from *S. halepense*, *S. bicolor* and *S. sudanense* (Fig. 1a). This procedure had a higher DNA yield than what was obtained from routine CTAB based leaf genomic DNA extraction products shown in (Fig. 1b). The quality of the DNA in Fig. 1a was better with low shredding of DNA. Both procedures, however, gave DNA of good PCR grade, resulting in the expected fragments from target loci.

Seeds obtained from open-pollinated UFD plots demonstrated the presence of spontaneous hybridization between *S. bicolor* and the weedy accessions. Alleles from weedy sorghums were found in other weedy species suggesting the presence of cross-pollination amongst the wild populations (Fig. 2, 3 and Table 2). Analysis of *S. sudanense* seed showed the presence of *S. bicolor*, *S. halepense* and *S. verticilliflorum* alleles on locus SB1764, SB3420, SB5058, SB5458 and SB3258 (Fig. 2 and Table 2). For instance on locus SB3420, *S. bicolor* produced 300 bp, *S. sudanense* produced 290 bp, *S. halepense* produced 0 bp and *S. verticilliflorum* produced 580 bp. An interspecific hybrid between *S. bicolor* and *S. sudanense* produced both 300 and 290 bp fragments, parental species from *S. sudanense* and *S. halepense* produced 290 bp while *S. sudanense* and *S. verticilliflorum*

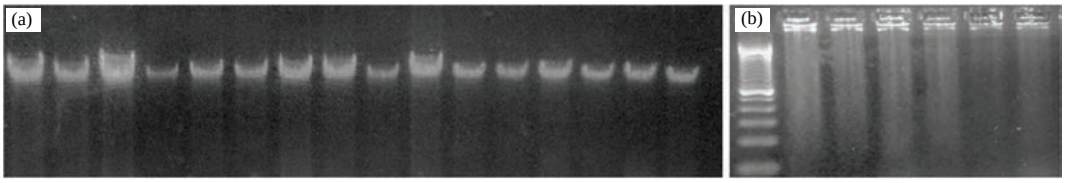


Fig. 1(a-b): Genomic DNA was extracted from (a) Seed and (b) Young leaf

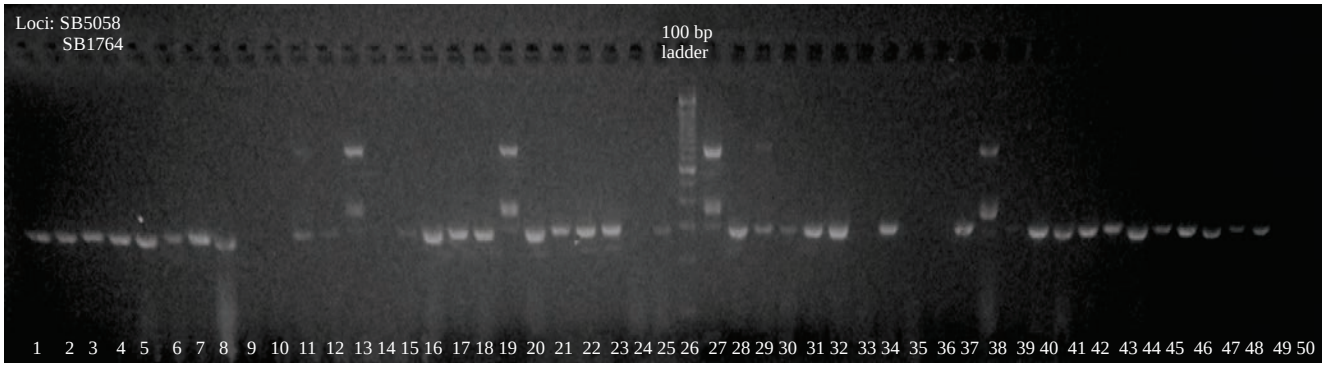


Fig. 2: Agarose gel of 49 *S. sudanense* PCR products on loci SB1764 and SB5058

Showing spontaneous hybridization with *S. halepense* (320 and 850 bp-lane 15, 20, 27, 38), *S. verticilliflorum* (890 bp-lane 13, 29), *S. bicolor* (300 bp-lane 21, 22, 23, 31, 32) in open-pollinated UFD plots on loci SB1764 and SB5058. Self-pollination in *S. sudanense* (300 bp-lanes 1, 2, 3, 4, 14, 25) and an *S. bicolor* control, lane 50

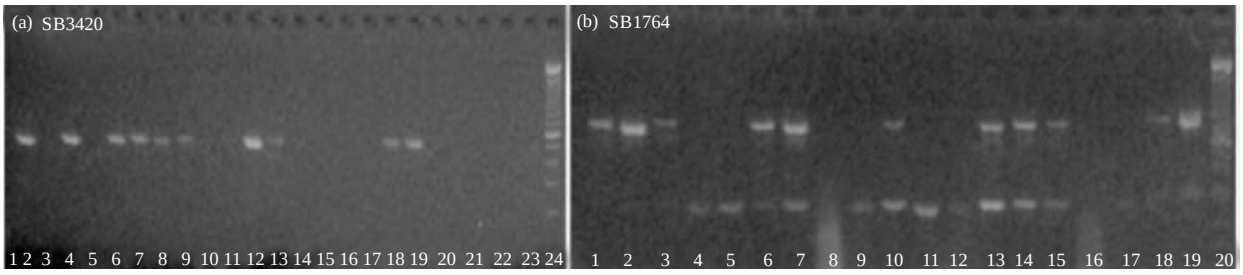


Fig. 3(a-b): Agarose gel of *S. halepense* and *S. sudanense* PCR products on loci SB3420 and SB1764, (a) Agarose gel showing open pollinated *S. halepense* samples with *S. verticilliflorum* alleles (580 bp-lane 1, 2, 5, 6, 7, 8, 11, 12) and without (0 bp-lane 14, 15, 16, 19-23) on locus SB3420 and (b) Open pollinated *S. sudanense* plants in UFD plots showing *S. verticilliflorum* alleles (890 bp-lane 1, 3, 6, 14, 15, 19), *S. halepense* alleles (850 bp-lane 2, 7, 10, 19 b) and *S. bicolor* plus *S. sudanense* alleles-(300 bp-lane 4,5,9,11,12,17 b) on locus SB1764

Table 2: Alleles in *S. bicolor* and in the wild species on SSR loci assayed

| Locus | *S. bicolor* | *S. sudanense* | *S. halepense* | *S. verticilliflorum* |
|---|---|---|---|---|
| SB1764 | 300 | 300 | 850 | 890 |
| SB3420 | 300 | 290 | 0 | 580 |
| SB5058 | 250 | 0 | 320 | 390 |
| SB5458 | 300 | 0 | 0 | 0 |
| SB3258 | 300 | 400/290 | 280 | 290 |

had both 290 and 580 bp. Similar results from interspecific hybridizations between *S. halepense* and *S. verticilliflorum* were seen in seeds from UFD fields (Fig. 3a-b and Table 2).

Apart from differences in the size of fragments obtained on amplification, loci SB3420, SB5058 and SB5458 had null alleles that did not have amplification in *S. halepense*,

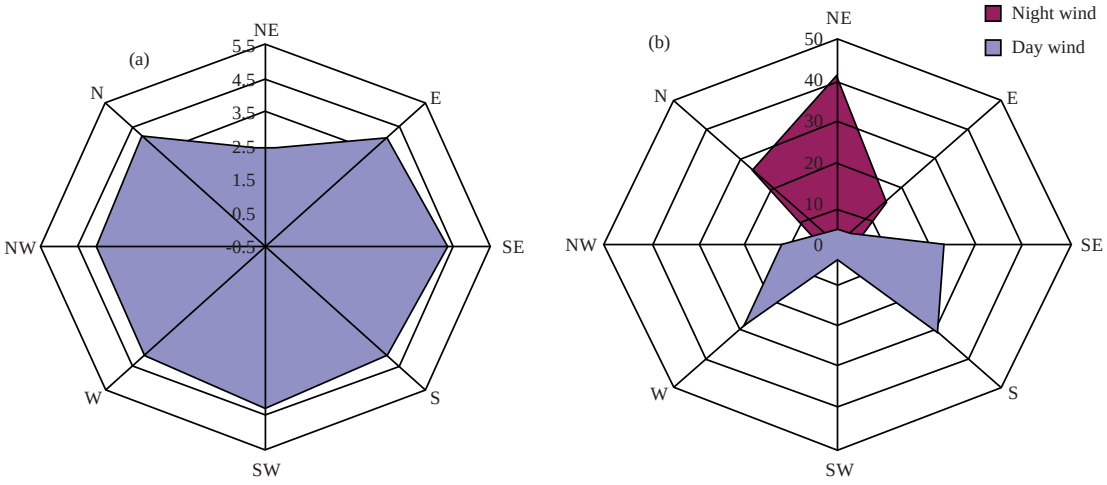


Fig. 4(a-b): Day wind speed between (a) 6 am and 6 pm and (b) Day and night wind direction, at sites 1 and 2

*S. sudanense* and *S. verticilliflorum*, respectively from open-pollinated fields (Table 2). These alleles were recovered in situations where hybridization with other weedy species or crop accessions occurred (Fig. 2 and 3).

**Meteorological data in the study sites:** The analysis of wind direction and wind speed showed that on average wind direction during the day blew towards the northwest, west and southeast and east. At night wind blew toward the north and northeast directions during pollination. Wind speed averaged 3.271 m $sec^{-1}$ at both sites. Mean wind speeds were recorded at between 3 and 3.7 m $sec^{-1}$ (Fig. 4a). Wind towards the southwest had large means of 3.5 m $sec^{-1}$ while wind towards the south-east had means of 3.7 m $sec^{-1}$, northeast east and north had average wind speeds of between 3.22 and 3.24 m $sec^{-1}$. Low speeds were observed in the south at 3.1 m $sec^{-1}$and northwest at 2.5 m $sec^{-1}$ (Fig. 4a). During the 2 month pollination period, day wind direction records showed that southeast and south directions had a higher number of times when the wind blew in the two directions, 23 and 30, respectively. The west and northwest had high values of 28 and 12 (Fig. 4b). Northeast, east, southwest and north, each had a count of 4. The night wind had high counts towards the northeast (41) followed by north (25) and east (15). Low counts of 4 were observed towards southeast, south, southwest, west and northwest (Fig. 4b).

**Maximum geneflow distance of crop genes into weedy sorghums in UFD field designs in two sites:** Significant differences in the flow of *S. bicolor* genes on a distance gradient in UFD plots were observed ($p \leq 0.05$). From the analysis up to 61% of the samples from *S. halepense*, *S. sudanense* and *S. verticilliflorum* had *S. bicolor* alleles at 5 m in both UFD plots from site 1 and site 2. The presence of *S. bicolor* alleles diminished with distance from the pollinator to between 40-30% at 60-70 m. There was an increase of up to 45% of weedy populations with *S. bicolor* alleles at 80-90 m from pollinators, demonstrating long-distance gene flow. The *S. bicolor* alleles in the weedy sorghums continued to decrease and these alleles were absent at 180-200 m from the source of pollen (Fig. 5a-b).

In the second site, *S. bicolor* gene flow above 100 m from pollinator was observed on the southern and western arms (Fig. 5a). *S. bicolor* alleles were detected at 100 m from the centre plot on the southeast arm, South-West arm and Northwest arm. The least distance of gene flow was recorded on the eastern arm (<50 m from source) (Fig. 5a).

In site 1 gene flow due to natural wind processes was more pronounced towards the south. On the southern arm in UFD plots *S. bicolor* alleles were detected 200 m from the pollen source. All other arms in this site did not have the presence of the *S. bicolor* allele after 100 m from the source. The east and northeast wing in this site showed a maximum gene flow of 60 m from the pollen source (Fig. 5b).

Significant differences were observed in the pattern of gene flow from *S. bicolor* to *S. sudanense*, *S. halepense* and *S. verticilliflorum* in the eight directions in UFD plots ($p \leq 0.05$). At 5 m, all 8 directions had high values of hybridization between *S. halepense* and *S. bicolor*. The southern and western arms showed elevated short distance and long-distance gene flow than all the other arms. The western arm showed the presence of *S. bicolor* alleles at 100 and 200 m from pollinator plots in both sites (Fig. 6a).

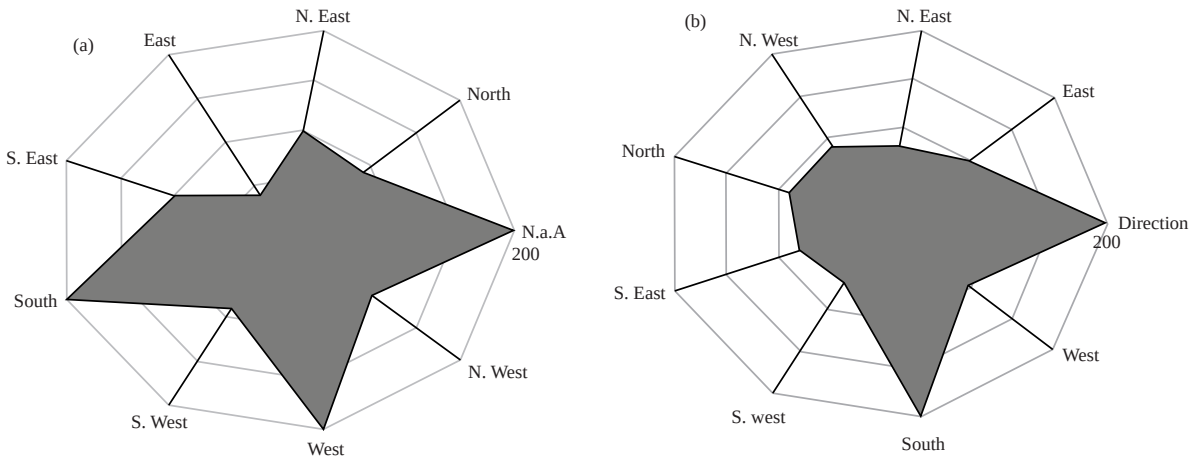


Fig. 5(a-b): Pollen mediated gene flow from *S. bicolor* to *S. verticilliflorum*, *S. halepense* and *S. sudanense* in UFD plot at (a) Site 2 and (b) Site 1

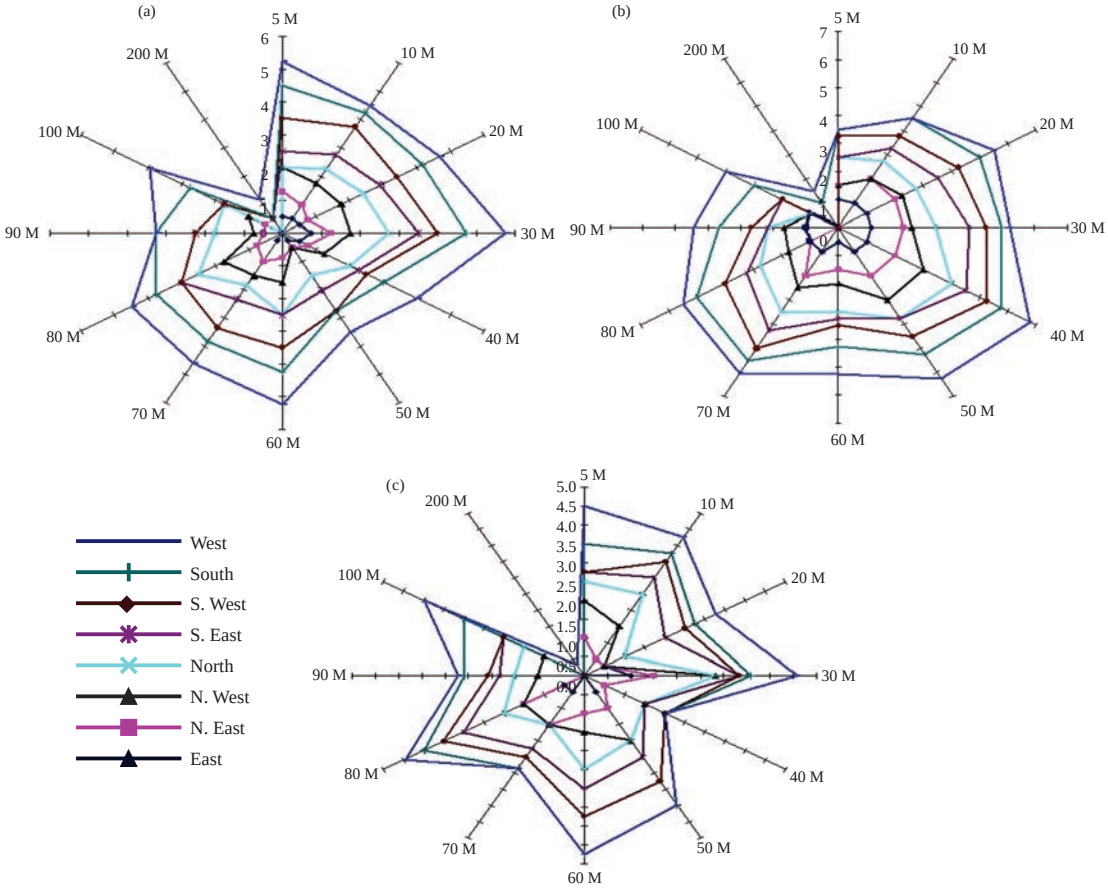


Fig. 6(a-c): Pattern of crop gene flow into (a) *S. halepense*, (b) *S. sudanense* and (c) *S. verticilliflorum*

*S. bicolor* alleles were detected in *S. sudanense* plants grown in UFD designs in the two sites (Fig. 6b). However, plants grown close to the pollinators had more hybridizations than those grown distally in all directions except the south, west and northwestern arms. Here, low inter-specific hybridizations were recovered earlier at 40-50 m from the

Table 3: Fitting regression models to explain the relationship between gene flow and distance from pollen source while considering factors (species, direction, wind direction and speed) that influence the pattern

| | | Parameter estimates | | | | | |
|---|---|---|---|---|---|---|---|
| Models | Data source | a | se | B | se | p-value | $r^2$ |
| Linear | Site 1+Site 2 | 0.6393 | 0.0265 | -0.001774 | 0.000374 | <0.001 | -0.778 |
| Y = A+BX | Site 1 | 0.6671 | 0.0328 | -0.002262 | 0.000424 | <0.001 | -0.773 |
| | Site 2 | 0.6942 | 0.0427 | -0.005416 | 0.000723 | <0.001 | -0.808 |
| log transformed | Site 1+Site 2 | -0.4640 | 0.0487 | -0.002811 | 0.000826 | <0.001 | -0.779 |
| $\log_{10}Y = \log_{10}A+BX$ | Site 1 | -0.4067 | 0.0604 | -0.00399 | 0.00103 | <0.001 | -0.770 |
| | Site 2 | -0.5756 | 0.0840 | -0.00021 | 0.00144 | <0.001 | -- |
| Linear divided by linear | Site 1+Site 2 | 1.131 | 0.436 | -0.433 | 0.460 | <0.001 | -0.996 |
| Y = A+B/(1+D*X) | Site 1 | 0.728 | 0.189 | -0.127 | 0.157 | <0.001 | -0.993 |
| | Site 2 | -- | -- | -- | -- | <0.001 | -- |
| Exponential | Site 1+Site 2 | 0.827 | 0.300 | -0.143 | 0.319 | <0.001 | 0.948 |
| Y = A+B*(R**X) | Site 1 | 0.6169 | 0.0749 | -0.0205 | 0.0415 | <0.001 | -0.955 |
| | Site 2 | 0.5671 | -- | -8.023E-17 | -- | <0.001 | -- |
| Exponential | Site 1+Site 2 | 0.8882*1 | 0.0397 | 1.00538*2 | 0.00506 | <0.001 | 0.495*3 |
| Y = A+B*(R**X)+C*(S**X) | Site 1 | 1.0016*1 | 0.0311 | 1.0016*2 | 0.0311 | <0.001 | -0.786*3 |
| | Site 2 | 1.000*1 | -- | 1.000*2 | -- | <0.001 | -- |

*1,*2, *3R and S parameters were estimated

pollinator. Hybridizations after 90 m were observed in southeastern, southern and western arms in both sites. Detection of hybrids between *S. sudanense* and *S. bicolor* was seen at 200 m from pollinator plots at the southern and western arms (Fig. 6b).

Distance of high dispersal of *S. bicolor* alleles to *S. verticilliflorum* plots was observed between 5-10 m. This was followed by a decrease in hybridizations between 30-60 m where there was a surge in hybridization. *S. bicolor* alleles diminished gradually from 70-100 m in all directions. Hybridization beyond 100 m was recovered in the western arm only (Fig. 6c).

**Linear and non-linear-exponential regressions models for geneflow in UFD plots (Fig. 7a-f):** Linear regressions of geneflow with distance while grouping with the direction of gene flow or the weedy species (Fig. 7) gave negative coefficients. Similar results were also obtained when correlations were done on geneflow and distance in all datasets.

This situation was observed when the correlations were grouped with species or direction of gene flow, except towards the south and the west (Fig. 7a). All the eight directions showed negative correlations when geneflow was correlated with wind direction (Fig. 7b). Positive correlations were observed between geneflow and wind speed in all the directions for *S. verticilliflorum*, *S. sudanense* and *S. halepense* (Fig. 7c). The three weedy species also demonstrated negative correlations when geneflow was correlated with distance (Fig. 7d) Two species *S. verticilliflorum* and *S. halepense* showed a similar pattern of gene flow in UFD plots. *S. sudanense* had significantly higher long-distance geneflow (Fig. 7e). However, there was no correlation between gene flow and wind speed in each of the three species (Fig. 7f).

Non-linear regressions showed a significant relationship between gene flow and distance while statistically grouped with the direction of gene flow (Table 2). The relationship was not significant while statistically grouped with species (Fig. 8a-b), (Table 2). An exponential relationship described the presence of high levels of gene flow at 5 m followed by a decrease till 200 m when grouped by species and direction of gene flow (Fig. 8c-d), (Table 2). A double exponential curve described the relation between geneflow and distance in the UFD plots showing both short-distance gene flow (5 m from pollen source) and long-distance gene flow (80-100 m from pollen source). This was then followed by a decrease to 200 m when grouped with species or with the directions of gene flow (Fig. 8e-f), (Table 3).

The regression model $\log_{10}Y = \log_{10}A+BX$ best fitted and explained the short distance and long-distance flow of crop genes. F test done for the regression gave a significant result ($p \leq 0.05$). The probability of the log-transformed model was <0.001 on the three data sets from site 1, site 2 and both site 1 plus site 2. Negative parameter correlations ($r^2$) were seen in both linear and log-transformed models while tested on the three datasets (Table 3).

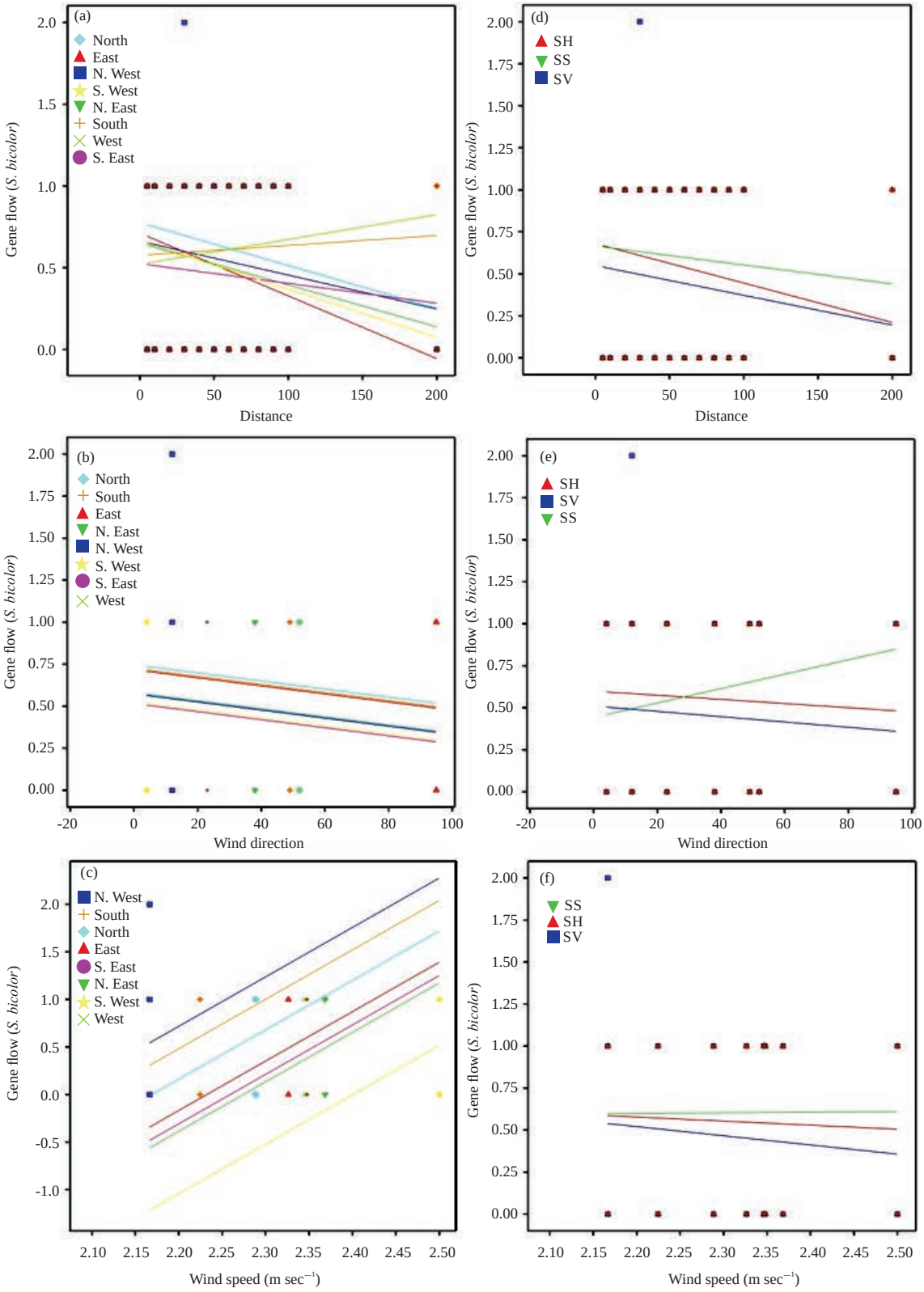


Fig. 7(a-f): Linear regression analysis for gene flow with direction and species (a-c) Gene flow with direction (d-f) Gene flow with species from *S. bicolor* to *S. halepense* (SH), *S. sudanense* (SS) and *S. verticilliflorum* (SV)

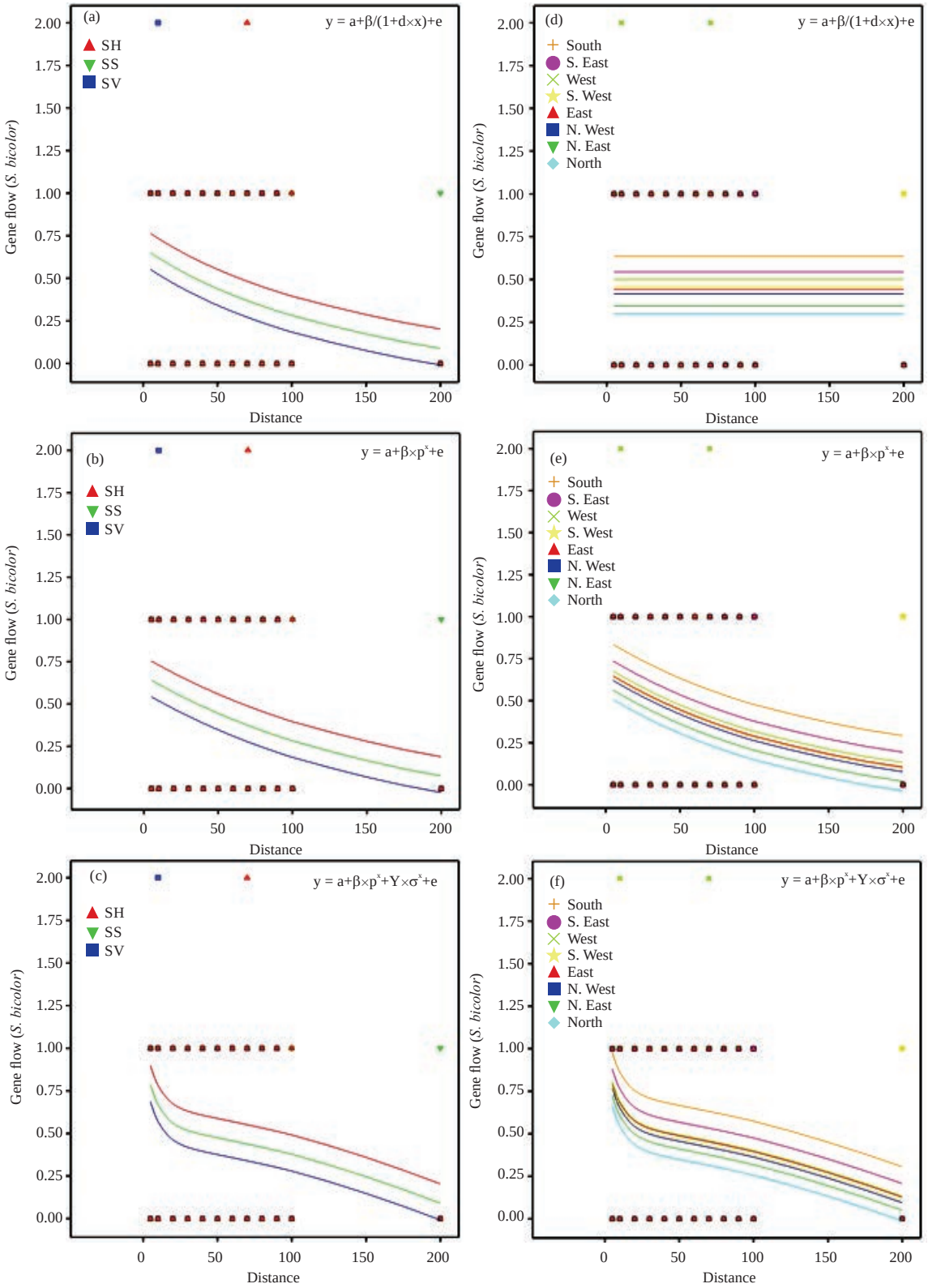


Fig. 8(a-f): Non-linear regression analysis for gene flow with direction and species (a-c) Gene flow with species (d-f) Gene flow with direction from *S. bicolor* to *S. halepense* (SH), *S. sudanense* (SS) and *S. verticilliflorum* (SV)

## DISCUSSION

A seed sampling and DNA extraction procedure described in this study gave high yields of PCR grade DNA (Fig. 1). The DNA extract from this technique was amenable to the evaluation of hybridization using PCR assays since the endosperm and the zygotic embryo contain genetic material from both male and female patients. There was a reduction in the time required and cost of the PCR tests for hybridization due to bulking of the sampled seed from each plot. This procedure showed an improved chance of capturing a hybridization event as opposed to the leaf-based assay where one plant was sampled. A study[22] applying a CTAB based seed endosperm DNA extraction method gave similar PCR grade DNA from maize Kernel. This seed bulking procedure approach has been severally utilized in reducing time and cost requirements in the assay of plants with RFLP and SSR markers[23,24].

Wind speed and wind direction influenced the maximum distance of gene flow in concentric UFD plots. Analysis of UFD field designs showed that wind direction and wind speed had a significant influence on the movement of *S. bicolor* genes *S. halepense*, *S. sudanense* and *S. verticilliflorum* populations (Fig. 5a-b). These two parameters were important when the main stems and tillers had flowered. It was also observed that gene flow was relatively higher in the south to west direction and *S. bicolor* alleles were detected at longer distances from the donor in the weedy species (Fig. 5a-b). These higher frequencies of hybridization correlated well with the wind direction in the two sites, since wind direction during the day was observed more towards the south to west direction. Wind speed was also highest in the southwest direction.

Wind speed and wind direction influenced pollen mediated gene flow from *S. bicolor* to the weedy species *S. halepense*, *S. sudanense* and *S. verticilliflorum*. As a consequence of the wind speed and wind direction high rates of hybridization were recovered between 5-30 m from pollen donors in all species except *S. sudanense* which had relatively lower values that could be attributed to its bushy growth habit (Fig. 5a-b). Weedy sorghums including *S. sudanense* are generally taller with several tillers, branches above the first internode and leaves making them bushy[25-27]. The species thus responds to wind causing it to lean and thus shielding the flowered panicle away from the pollen donor. However, short distance (5-30 m) and long-distance (80-90 m) gene flow was seen in all species. The long-distance surge could also have been a result of the wind effect and the pseudo block provided by the taller weedy species growing 5-10 m from the shorter pollinator. Changes in wind speed during the day were less than those observed in wind direction during the day thus the effect of wind speed on gene flow in each of the three species was less pronounced and therefore there was a lack of correlation between wind speed and geneflow when data were grouped with the species. There was a positive correlation in *S. sudanense* between wind direction and geneflow when data were grouped with the species probably due to the differences in growth habits.

Several gene flow studies in crops have reported the significant effect of wind parameters on pollen mediated gene movement. Studies in rice showed that the monsoon climate and area topography influenced outcrossing[12]. The previous studies[15] showed the wind effect on the movement of Bt genes from GM maize to non-GM along a distance gradient with the presence of border rows. Wind parameters were also important in gene flow studies in wheat[3]. The current study concurs with previous studies in sorghum where the importance of wind has been shown in crop to crop gene flow[7,8]. The wind was also important in the flow of *S. bicolor* alleles into *S. halepense* in field experiments[8].

The log-transformed regression model:

$$\log_{10}Y = \log_{10}A+Bx$$

best fitted and described the flow of crop genes considering the short distance and long-distance situations. Short dispersal of pollen mediated gene flow was recorded at 5 m from the central donor plot while the long "jump" dispersal was recorded at between 80-100 m. The maximum distance of gene flow in the UFD plots was seen at 200 m (Fig. 8c,f). Exponential curves were fitted on the data yielding a Leptokurtic left-skewed distribution showing a higher frequency of hybridization closer to the pollinator plot in all sites. The frequency of hybridization gradually diminished to 200 m from the pollinator. This gradual decrease could be attributed to gene flow factors, such as wind speed, flower morphology, plant stature and the distance from the pollinator[9]. Previous studies have described pollen dispersal models using inverse power and negative exponential transformation[9,28]. In addition, studies[7] observed hybridization at 158 m from the donor plot. Using mathematical models to estimate gene flow in crop sorghum plots they estimated a maximum gene flow distance of between 200-700 m. A negative exponential regression model with logarithmic transformation of PMGF and square root transformation of distance was applied to estimate the maximum gene flow distance at 200 m in crop sorghum[8].

This study will help the researchers to uncover the critical patterns of gene flow in traditional cropping systems where crops and transgenic crop genes could be found in weedy progenitors. Thus there is a need to consider isolation distances of more than 200 meters to ensure containment of robust transgenes in crop sorghums. Constant surveillance and early detection of such interspecific hybridization and crop to crop intraspecific hybridization events are recommended.

## CONCLUSION

This study deduced that wind parameters including wind speed and direction increased pollen mediated gene flow from *S. bicolor* to the weedy species *S. halepense*, *S. sudanense* and *S. verticilliflorum*. Results from the study further demonstrated that *S. verticilliflorum* had lower gene flow frequencies in most unlimited gene flow directions relative to *S. halepense* and *S. sudanense*. This provided significant evidence of species variations and differences in gene flow due to differential biological orientation of the species in the UFD plots. More interspecific hybridization events and the presence of crop genes (from the donor plot) were observed above 100 m from pollen sources in *S. halepense* and *S. sudanense.* In addition, short dispersal gene flow and long "jump" dispersal were observed at 5 and 100 m, respectively. The maximum distance of gene flow in the UFD plots was seen at 200 m. These results provide valuable insights into the nature of pollen mediated dispersal of crop genes/transgenes to their feral progenitors in situations where substantial diversity exists within interbreeding species. This will provide useful input in decision making on the judicious use of transgenic crops in traditional cropping systems.

## REFERENCES


1. Ellstrand, N.C., 2003. Current knowledge of gene flow in plants: Implications for transgene flow. Philos. Trans. R. Soc. B: Biol. Sci., 358: 1163-1170.
2. Campbell, L.G., A.A. Snow and C.E. Ridley, 2006. Weed evolution after crop gene introgression: Greater survival and fecundity of hybrids in a new environment. Ecol. Lett., 9: 1198-1209.
3. Gustafson, D.I., M.J. Horak, C.B. Rempel, S.G. Metz, D.R. Gigax and P. Hucl, 2005. An empirical model for pollen mediated gene flow in wheat. Crop Sci., 45: 1286-1294.
4. Barnaud, A., M. Deu, E. Garine, J. Chantereau and J. Bolteu *et al*., 2009. A weed-crop complex in sorghum: The dynamics of genetic diversity in a traditional farming system. Am. J. Bot., 96: 1869-1879.
5. Mutegi, E., F. Sagnard, M. Labuschagne, L. Herselman and K. Semagn *et al*., 2012. Local scale patterns of gene flow and genetic diversity in a crop-wild-weedy complex of sorghum (*Sorghum bicolor* (L.) Moench) under traditional agricultural field conditions in Kenya. Conserv. Genet., 13: 1059-1071.
6. Pedersen, J.F., D.B. Marx and D.L. Funnell, 2003. Use of $A_3$ cytoplasm to reduce risk of gene flow through sorghum pollen. Crop Sci., 43: 1506-1509.
7. Schmidt, M. and G. Bothma, 2006. Risk assessment for transgenic sorghum in Africa: Crop to crop gene flow in *Sorghum bicolor* (L.) Moench. Crop Sci., 46: 790-798.
8. Rabbi, I.Y., H.K. Parzies, D. Kiambi, B.I.G. Haussmann, R. Folkertsma and H.H. Geiger, 2011. Experimental studies on pollen-mediated gene flow in *Sorghum bicolor* (L.) Moench using male-sterile bait plants. Plant Breed., 130: 217-224.
9. Gunog, S. and F. Lutscher, 2011. Spread rates under temporal variability: Calculation and application to biological invasions. Math. Models Methods Appl. Sci., 21: 2469-2489.
10. Klinger, T., 2001. Variability and Uncertainty in Crop-to-Wild Hybridization. In: Genetically Engineered Organisms: Assessing Environmental and Human Health Effects, Letourneau, D.K. and B.E. Burrows (Eds.), CRC Press, New York, ISBN-13: 9781420042030, pp: 1-15.
11. Beckie, H.J. and L.M. Hall, 2008. Simple to complex: Modelling crop pollen-mediated gene flow. Plant Sci., 175: 615-628.
12. Yao, K., N. Hu, W. Chen, R. Li and Q. Yuan *et al*., 2008. Establishment of a rice transgene flow model for predicting maximum distances of gene flow in Southern China. New Phytol., 180: 217-228.
13. Bruˆl' Babel, A.L., C.J. Willenborg, L.F. Friesen and R.C. van Acker, 2006. Modeling the influence of gene flow and selection pressure on the frequency of a GE herbicide tolerant trait in non GE wheat and wheat volunteers. Crop Sci., 46: 1704-1710.
14. Devaux, C., C. Lavigne, F. Austerlitz and E.K. Klein, 2007. Modelling and estimating pollen movement in oilseed rape (*Brassica napus*) at the landscape scale using genetic markers. Mol. Ecol., 16: 487-499.
15. Gustafson, D.I., I.O. Brants, M.J. Horak, K.M. Remund, E.W. Rosenbaum and J.K. Soteres, 2006. Empirical modeling of genetically modified maize grain production practices to achieve European Union Labeling Thresholds. Crop Sci., 46: 2133-2140.
16. Demeke, T. and G.R. Jenkins, 2010. Influence of DNA extraction methods, PCR inhibitors and quantification methods on real-time PCR assay of biotechnology-derived traits. Anal. Bioanal. Chem., 396: 1977-1990.

17. Datukishvili, N., I. Gabriadze, T. Kutateladze, M. Karseladze and B. Vishnepolsky, 2010. Comparative evaluation of DNA extraction methods for food crops. Int. J. Food Sci. Technol., 45: 1316-1320.
18. Abdel-Latif, A. and G. Osman, 2017. Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. Plant Methods, Vol. 13. 10.1186/s13007-016-0152-4.
19. Bowers, J.E., C. Abbey and S. Anderson, C. Chang and X. Draye *et al*., 2003. A high-density genetic recombination map of sequence-tagged sites for *Sorghum* as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics, 165: 367-386.
20. Yonemaru, J.I., T. Ando, T. Mizubayashi, S. Kasuga, T. Matsumoto and M. Yano, 2009. Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of Sorghum (*Sorghum bicolor* (L.) Moench). DNA Res., 16: 187-193.
21. Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak and J. Grimwood *et al*., 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature, 457: 551-556.
22. Gao, S., C. Martinez, D.J. Skinner, A.F. Krivanek, J.H. Crouch and Y. Xu, 2008. Development of a seed DNA-based genotyping system for marker-assisted selection in maize. Mol. Breed., Vol. 22. 10.1007/s11032-008-9192-4.
23. Xu, Y. and J.H. Crouch, 2008. Marker-assisted selection in plant breeding: From publications to practice. Crop Sci., 48: 391-407.
24. van Treuren, R., L.J.M. van Soest and T.J.L. van Hintum, 2001. Marker-assisted rationalisation of genetic resources collections: A case study in flax using AFLPs. Theor. Appl. Genet., 103: 144-152.
25. Dillon, S.L., F.M. Shapter, R.J. Henry, G. Cordeiro, L. Izquierdo and L.S. Lee, 2007. Domestication to crop improvement: Genetic resources for *Sorghum* and *Saccharum* (Andropogoneae). Ann. Bot., 100: 975-989.
26. Ejeta, J. and C. Grenier, 2005. Sorghum and its Weedy Hybrids. In: Crop Ferality and Volunteerisim, Gressel, J. (Eds.), CRC Press, Boca Raton, FL., Pages: 448.
27. Follak, S. and F. Essl, 2013. Spread dynamics and agricultural impact of *Sorghum halepense*, an emerging invasive species in Central Europe. Weed Res., 53: 53-60.
28. Marceau, A., D.I. Gustafson, I.O. Brants, F. Leprince and X. Foueillassar *et al*., 2013. Updated empirical model of genetically modified maize grain production practices to achieve European Union Labeling Thresholds. Crop Sci., 53: 1712-1721.