

ORIGINAL ARTICLE

# Fine-scale spatial genetic dynamics over the life cycle of the tropical tree *Prunus africana*

DG Berens<sup>1,2</sup>, C Braun<sup>1</sup>, SC González-Martínez<sup>3</sup>, EM Griebeler<sup>1</sup>, R Nathan<sup>4</sup> and K Böhning-Gaese<sup>1,2,5,6</sup>

Studying fine-scale spatial genetic patterns across life stages is a powerful approach to identify ecological processes acting within tree populations. We investigated spatial genetic dynamics across five life stages in the insect-pollinated and vertebrate-dispersed tropical tree *Prunus africana* in Kakamega Forest, Kenya. Using six highly polymorphic microsatellite loci, we assessed genetic diversity and spatial genetic structure (SGS) from seed rain and seedlings, and different sapling stages to adult trees. We found significant SGS in all stages, potentially caused by limited seed dispersal and high recruitment rates in areas with high light availability. SGS decreased from seed and early seedling stages to older juvenile stages. Interestingly, SGS was stronger in adults than in late juveniles. The initial decrease in SGS was probably driven by both random and non-random thinning of offspring clusters during recruitment. Intergenerational variation in SGS could have been driven by variation in gene flow processes, overlapping generations in the adult stage or local selection. Our study shows that complex sequential processes during recruitment contribute to SGS of tree populations.

*Heredity* (2014) **113**, 401–407; doi:10.1038/hdy.2014.40; published online 21 May 2014

## INTRODUCTION

Gene flow via pollen and seeds is a key factor in determining the genetic make-up of plant populations (Loveless and Hamrick, 1984). Restricted gene flow is probably one of the most prevalent factors influencing fine-scale spatial genetic structure (SGS), that is, a decrease in relatedness with increasing spatial distance between individuals (Vekemans and Hardy, 2004). It results in spatial aggregations of siblings in close vicinity of the mother trees and, thus, in high offspring's SGS (for example, Chung *et al.*, 2003).

Within generations, that is, among different juvenile life stages stemming from the same adult population, the persistence of the initial genetic make-up is influenced by various processes. As only few individuals from the seed rain survive, SGS is likely to weaken in older stages (Chung *et al.*, 2003). This effect is expected even if mortality is random in space (random thinning). Yet, if mortality is non-random and driven by selection, the patterns of SGS can change more profoundly (Hamrick *et al.*, 1993; Jones and Hubbell, 2006). In addition, because of density- and distance-dependent predation and pathogen pressure, juveniles may encounter high mortality in the vicinity of the mother tree (Janzen, 1970; Connell, 1971), leading also to a strong thinning of the sib clusters that dilutes SGS.

Across generations, that is, between adults and juveniles, the level of SGS may vary because of different demographical processes: historic factors can lead to differences in underlying patterns of gene flow between adults and juveniles. For example, if the extent of seed dispersal has declined across generations, owing to, for example, anthropogenic disturbance, juveniles may show stronger SGS than

adults (for example, Farwig *et al.*, 2008). In contrast, if the adult population results from a generation with few founder trees it may show stronger levels of SGS than the present day juveniles (Jones *et al.*, 2006; Pardini and Hamrick, 2008). If the population is in equilibrium and adults of the previous generation are representative of the currently establishing generation, SGS should weaken from juvenile to adults owing to random and non-random mortality (Zhou and Chen, 2010).

Studying genetic diversity and SGS of different life stages and across generations of a species helps to understand the formation of genetic structure within adult plant populations (for example, Hamrick *et al.*, 1993; Latouche-Hallé *et al.*, 2003; Jacquemyn *et al.*, 2006; Zhou and Chen, 2010). Owing to overlapping generations, which are common in tree species, individuals from a younger life stage grow besides closely related individuals from older generations (Jones and Hubbell, 2006). Thus, SGS analyses on pooled life stages of trees can be misleading. Hence a multistage perspective can lead to a higher resolution of the processes underlying the formation of SGS in trees (Chung *et al.*, 2003; Jones *et al.*, 2006; Pardini and Hamrick, 2008). Only a few studies that assessed SGS from a multistage perspective incorporated the initial seed rain stage in their analyses (for example, Jones and Hubbell, 2006; Zhou and Chen, 2010). Including this stage may reveal how seed dispersal leads to the formation of the initial genetic template.

Here, we investigated the development of SGS in the insect-pollinated and vertebrate-dispersed tropical tree *Prunus africana* (Hook. f.) Kalkman (Rosaceae) in Kakamega Forest, western Kenya.

<sup>1</sup>Department of Ecology, Institute for Zoology, Johannes Gutenberg-University of Mainz, Mainz, Germany; <sup>2</sup>Department of Ornithology, National Museums of Kenya, Nairobi, Kenya; <sup>3</sup>Department of Forest Ecology and Genetics, CIFOR-INIA, Madrid, Spain; <sup>4</sup>Movement Ecology Lab, Department of Ecology, Evolution and Behavior, The Hebrew University of Jerusalem, Edmond J Safra Campus, Givat Ram, Jerusalem, Israel; <sup>5</sup>Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Frankfurt (Main), Germany and <sup>6</sup>Department of Biological Sciences, Goethe University, Frankfurt (Main), Germany

Correspondence: Dr DG Berens, Department of Ecology, Conservation Ecology, Faculty of Biology, Philipps-University Marburg, Karl-von-Frisch-Straße 8, 35032 Marburg, Germany.

E-mail: dana.berens@staff.uni-marburg.de

Received 30 June 2013; revised 14 March 2014; accepted 17 March 2014; published online 21 May 2014

*P. africana* is a common species in Kakamega Forest, considered a late pioneer or an early climax tree species (for example, Swaine and Whitmore, 1988). Farwig *et al.* (2008) studied genetic structure of *P. africana* seedlings and adults at a larger spatial scale among eight populations distributed over Kakamega Forest and adjacent fragments. They found stronger genetic differentiation among seedling populations than among adults, showing that genetic diversity and structure of the species is sensitive to processes reducing gene flow, such as forest fragmentation and local disturbance (Farwig *et al.*, 2008). To better understand these patterns, we used parentage analyses to investigate the extent of gene flow across different life stages mediated by insect pollinators and vertebrate seed dispersers (Berens *et al.*, 2013). Our results suggested that seed dispersal was strongly distance restricted (mean distances of only 5 m), while pollen dispersal reached over 23-fold longer distances. Yet, strong non-random mortality led to a disproportionate increase of realized maternal gene flow distances at later life stages, as well as to strong variation in paternal gene flow distances (Berens *et al.*, 2013). In this study, we used a multi-stage approach to determine whether the observed patterns of SGS are consistent with expectations deriving from the observed biological processes. We hypothesize (i) strong SGS at the seed stage because of distance-restricted seed dispersal; (ii) a continuous decrease of SGS across older recruitment stages because of mortality and the thinning of clusters of related individuals; and (iii) an increase in SGS across generations (between adult trees and seedlings of the next generation) resulting not only from restricted dispersal but also from proposed changes in gene flow across generations (Farwig *et al.*, 2008).

## MATERIALS AND METHODS

### Study species

*P. africana* is a long-lived (>100 years) monoecious tree species whose range expands from East Africa to Madagascar and the Comores (Hall *et al.*, 2000).

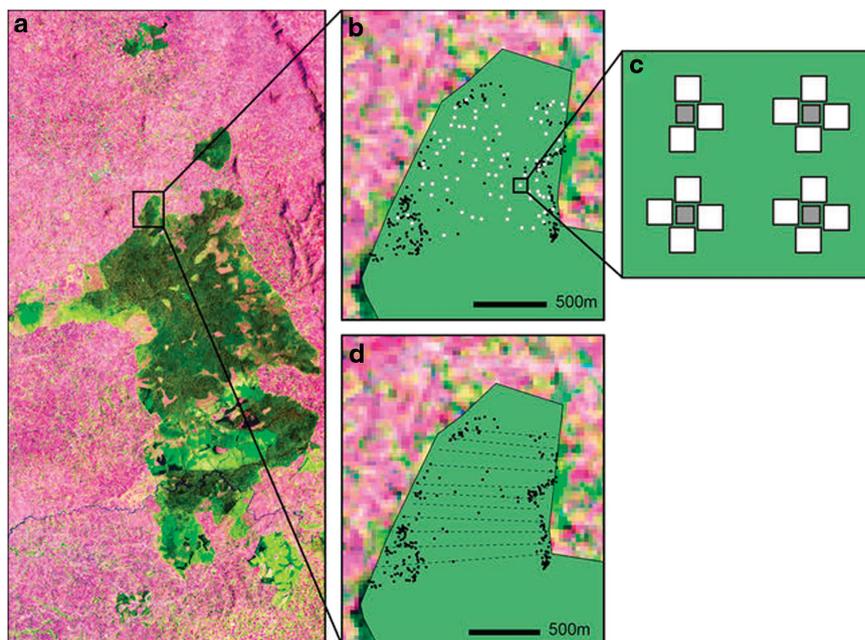
Owing to bark exploitation for medicinal use, the species is listed on CITES appendix II. *P. africana* has small white hermaphroditic protogynous flowers. Main pollinators are hymenoptera and diptera (Hall *et al.*, 2000). The species is mainly outcrossing (Hall *et al.*, 2000). Parentage analyses of the species revealed selfing rates of 5–13% (Berens, unpublished data). Main seed dispersers of the one-seeded fleshy fruits are birds and monkeys (Hall *et al.*, 2000). Secondary seed dispersal seems to be neglectable in the species, as experiments with thread-marked *P. africana* seeds did not record any caching, but only later predation of seeds (Svd Gönna and M Melcher, personal communication). Survival to the sapling stage is very low, potentially because of high light requirement as well as strong seed predation and seedling herbivory (Tsingalia, 1989; Tesfaye *et al.*, 2002).

### Study area

Kakamega Forest is located in western Kenya (Figure 1). The mid-altitudinal rainforest (1500–1700 m asl, KIFCON, 1994) comprises a main forest block and five fragments with a total area of 13 000 ha (Lung, 2004). Average annual temperature is 18.7 °C (Mitchell *et al.*, 2009) and average yearly precipitation is about 2000 mm per year (averaged from Forest Department records at Isecheno Forest Station from 1982 to 2005).

The study area within Kakamega Forest was a 120 ha forest peninsula in the northwest of the non-fragmented main forest block extending into the agricultural matrix (0°21'34.1" North, 34°51'31.0" East; Figure 1). It is part of the official Kakamega Forest Nature Reserve, managed by the Kenya Wildlife Service (KWS) and established in 1967. The study area mostly comprises primary forest, which has not been clear-cut for at least 100 years (Mitchell *et al.*, 2009). Smaller secondary forest patches within the study area are *ca.* 50 years old. Owing to the strict protection by the KWS, rates of human disturbance in terms of logging or bark harvesting of the species in the study area are low (Mitchell *et al.*, 2009).

Individuals of all life stages of *P. africana* in the study area are spatially clustered and the degree of clumping decreases with life stage (see also 'Drivers of SGS change across life stages' for details on clustering): the density of adult *P. africana* trees (with a diameter at breast height >10 cm; hereafter ADLTS) within our study area is 0.00022 individuals m<sup>-2</sup>. Adult trees are clustered in



**Figure 1** (a) Satellite image showing Kakamega Forest with forested area (green) and surrounding farmland (pink) and the location of the study area within the forest. (b) Spatial distribution of adult *P. africana* trees ( $n=260$ , black dots) and of sampling plots ( $n=86$ , white squares) for sampling of fruits and seeds (SDFRT) and current-year seedlings (SLYNG). (c) Spatial distribution of adult *P. africana* trees (black dots) and of 11 transects (dashed lines) for sampling of middle age seedlings (SLMID) and old seedlings (SLOLD). (d) Each sampling plot consisted of four seed traps (0.5 m<sup>2</sup>, hatched squares) and 15 seedling subplots (1 m<sup>2</sup>, white squares).

areas of high light availability, in particular near forest edges and in secondary regrowth in the west and east of the forest peninsula. Mean densities in our study plots decreased from 63.22 for propagules (seeds and fruits, hereafter SDFRT), 40.06 for current year seedlings (SLYNG), 0.01 for seedlings up to 3-year (SLMID) to 0.004 seedlings older than 3-year (SLOLD) individuals per  $m^2$  (see Plant material for life stage definition).

### Plant material

In 2005, we established 22 parallel transects crossing the study area from the eastern to the western forest edge and covering *ca.* 87 ha (Figure 1). Transects were between 694 and 1225 m long and were separated by 40 m. We recorded all *P. africana* ADLTS in the transect area, as well as all trees up to 100 m distance south of transect 1 and north of transect 22, as they may be potential parents of the offspring analyzed. In total, we detected 261 adult *P. africana* trees. Global positioning system (GPS) coordinates of each tree were taken. Leaf material of all adult trees was collected between January and May 2006.

In the same year, 86 geopositioned sampling plots were randomly established along the transects (Figure 1b). A number of 20 sampling plots were under *P. africana* trees. Each plot consisted of an array of four seed traps ( $0.7 \times 0.7$  m) made of a wire frame covered with mosquito mesh (Figure 1d). To avoid secondary dispersal from traps, traps were installed at a height of *ca.* 1 m above the ground. Propagules (SDFRT) were collected from seed traps during a fruiting peak of *P. africana* in our study site on a weekly basis between January and April, 2006. They were assigned to fruits (with pulp) or seeds (without pulp) and pooled for further analyses. Fruit pulp was manually removed for storage and all samples were dried with silica gel. In total, 8544 propagules were counted in the traps. We aimed at having similar sample sizes (*ca.* 300 individuals) for each life stage, evenly distributed across the study site. For genotyping, we subsampled 180 propagules from sampling plots, which were situated under *P. africana* trees (representing 2% of all propagules found there) and 131 propagules from sampling plots away from *P. africana* trees (representing 92% of all propagules in that category), resulting in a total sample of 311 genotyped propagules (SDFRT).

SLYNG were collected at the same 86 sampling plots between March and May, 2006. Fifteen  $1 m^2$  subplots were placed within each sampling plot (Figure 1c) and newly emerged seedlings were counted and collected on a biweekly basis. Newly emerged seedlings originated from the same fruiting period as SDFRT and could be identified because of the presence of cotyledons. Leaf samples were dried with silica gel. Again, 180 seedlings were subsampled from sampling plots under *Prunus* trees (representing <1% of all seedlings under *Prunus* trees), and the remaining 129 seedlings from plots away from *Prunus* trees (representing 27% of seedlings found at these stations), resulting in a total subsample of 309 out of the 48 991 seedlings for genetic analyses.

As a result of lower density of older seedlings (SLMID and SLOLD) compared with SDFRT and SLYNG, they were sampled, between January and February 2008, along every second of the 22 transects crossing the study area (Figure 1c). Each *P. africana* seedling encountered in these plots was recorded and the geographic position was taken. A leaf sample of each individual was taken and the seedling age was determined by an experienced field assistant considering the degree of lignification of the stem axis. Although this procedure implies a risk of misassignment to the wrong germination year, the approach is still valid for a differentiation between younger and older seedlings. Thus, a subsample of 298 out of the 861 SLMID and 301 out of the 368 SLOLD was taken for genetic analyses. SLMID were between 1 and 3 years old (average 3 years), and SLOLD from 4 years up to >6 years (average 5 years).

### Molecular analyses

All plant material (leaves of ADLTS, SLYNG, SLMID, SLOLD, and embryo tissue of SDFRT) was ground to fine powder with a Retsch mixer mill (Type MM 301, Retsch, Haan, Germany). DNA was extracted following the protocol described by Wang *et al.* (1993). The standard protocol was modified by using 40  $\mu$ l of NaOH and 95  $\mu$ l of Tris buffer for each sample. Samples were genotyped at six microsatellite loci that were originally developed for *Prunus avium* and *Prunus persica*: *UDP97-403* (Cipriani *et al.*, 1999), *P12A02* (Sosinski *et al.*, 2000), *BPPCT-002* (Dirlewanger *et al.*, 2002), *UDP96-005* and *UDP98-410* (Schueler *et al.*, 2003), and *EMPaS06* (Vaughan and Russell, 2004).

Primers were fluorescent 5'-end-labeled (6-FAM, NED, HEX; Applied Biosystems, Foster City, CA, USA). Polymerase chain reaction protocols followed Farwig *et al.* (2008) with annealing temperatures ranging from 54 to 62 °C and the number of cycles ranging from 44 to 46. For loci *UDP97-403* and *EMPaS06*, a final extension step of 60 °C for 30 min was added to the standard protocol. Samples were genotyped on ABI 3730 and 3130xl capillary sequencers. Allele scoring was performed using the software GeneMapper 4.0 (Applied Biosystems). For all loci, base shifts between scorings of the different sequencers were calibrated by running 48 samples per locus (3%) on both instruments. All samples except that of one adult tree were successfully amplified.

### Changes in clustering across life stages

In order to investigate the change in spatial clustering across life stages, we obtained the pairwise geographic distances between individuals in a fixed distance category of up to 100 m. We compared these individual pairwise distances across life stages using a generalized linear model with quasi-Poisson-distributed errors. Significance of the independent variable 'life stage' was determined by comparing the full model with a model only including the intercept via analysis of variance (ANOVA) and F-tests. Differences between life stages were analyzed with *post-hoc* pairwise Wilcoxon tests (with Bonferroni corrections) as implemented in R 3.0.1 (R Core Team, 2013).

### Genetic diversity and SGS

Rarefied allelic richness, which is a measure of the number of alleles per locus based on the minimum number of samples genotyped at all loci, as well as observed ( $H_o$ ) and expected heterozygosity ( $H_e$ ) were assessed for each locus and life stage using the software FStat 2.9.3.2 (Goudet, 1995; El Mousadik and Petit, 1996). Null allele frequencies were obtained from CERVUS 3.0.3 (Kalinowski *et al.*, 2007). Differences between  $H_o$  and  $H_e$  within life stages were tested through assessing the significance of  $F_{is}$  after Bonferroni correction in FStat. We conducted ANOVAs to test for differences in allelic richness,  $H_o$  and  $H_e$  among life stages. ANOVAs were done in R 3.0.1 (R Core Team 2013).

In order to test for a significant decrease in relatedness across distance, a regression between the kinship coefficient  $f_{ij}$  and the logarithm of pairwise geographic distances of individuals was computed using SpaGeDi (Hardy and Vekemans, 2002). Standard errors of the regression slope,  $b$ -log, were computed using a jackknife procedure over loci. Values were tested for significant difference from a random pattern using 10 000 permutations. Analyses were conducted for each life stage separately across the full distance range.

Spatial autocorrelation of genetic relatedness across geographic distance was assessed using GenAIEx 6.501 (Peakall and Smouse, 2006; Peakall and Smouse 2012). We performed random permutations ( $n=999$ ), to test the null hypothesis that genotypes are spatially interchangeable (Smouse *et al.*, 2008). First, analyses were conducted for each life stage separately. Distance intervals were set at 100-m intervals (0–100, 101–200 and so on). Maximum distance intervals were 1100 for SDFRT and SLYNG, 1600 for SLMID and SLOLD, and 1700 for ADLTS (for the mean number of pairs per distance interval see Table 1). We performed a heterogeneity test in order to evaluate the significance of correlograms at a  $P$ -level of 0.01 (Banks and Peakall, 2012). Second, we conducted a heterogeneity test to see whether relatedness differed among life stages for each distance interval and across distance intervals (Smouse *et al.*, 2008). The maximum distance interval was set to 1000 m to assure comparability among life stages.

Assuming that SGS has reached equilibrium, we used an iterative procedure implemented in SpaGeDi to indirectly estimate the axial variance of gene dispersal distance  $\sigma$  for adult trees from its effective density  $D_e$  (Hardy and Vekemans, 2002).  $D_e$  is normally unknown but different values of  $N_e/N$ , whereby  $N_e$  is the effective population size and  $N$  the census one, can be used to approximate this quantity (Vekemans and Hardy, 2004; Born *et al.*, 2008). As the ratio  $N_e/N$  typically ranges between 0.1 and 0.5 (Frankham 1995; Born *et al.*, 2008), we used the observed adult density, 0.00022 individuals  $m^{-2}$  as an upper limit estimate, as well as 0.00022/10 as a lower limit estimate of the effective population density  $D_e$ . The distance range for the regression was set between  $\sigma$  and 20  $\sigma$ , assuming a mutation rate  $\mu$  of  $10^{-3}$  for nuclear microsatellites (see Heuertz *et al.*, 2003).

## RESULTS

### Changes in clustering across life stages

The clustering rate of individuals in terms of the mean  $\pm$  s.e. distances of individual pairs in the smallest distance category (up to 100 m distance) was  $56.2 \pm 0.5$  m for ADLTS ( $n=2322$ ) and increased from  $21.4 \pm 0.4$  m for SDFRT ( $n=5603$ ),  $23.4 \pm 0.3$  m for SLYNG ( $n=6961$ ),  $39.2 \pm 0.4$  m for SLMID ( $n=5543$ ) to  $41.2 \pm 0.3$  m for SLOLD ( $n=8972$ ). The differences in spatial distance between individuals in the smallest distance category were significant between all life stages except for SLMID and SLOLD (pairwise Wilcoxon rank sum tests, all  $P$ -values  $<0.001$  after Bonferroni correction).

### Genetic diversity

The microsatellite loci were highly polymorphic with a mean  $\pm$  s.e. allelic richness of  $18.7 \pm 1.5$  (from 16.7 in SDFRT to 29.2 in ADLTS, based on a minimum number of 252 diploid individuals; Table 1). Mean null allele frequency was 0.07 across loci and life stages and ranged from  $-0.06$  to 0.26 (locus R1 in SLMID). Observed heterozygosity ( $H_o$ ) ranged between 0.66 (SDFRT) and 0.77 (ADLTS),

while gene diversity (expected heterozygosity,  $H_e$ ) was very similar (0.82–0.83) across all life stages (Table 1). In addition,  $H_o$  was significantly lower than  $H_e$  in all cases ( $P$ -values for all  $F_{is}$ -values  $<0.0017$ ; Table 1). Neither allelic richness, nor  $H_o$  or  $H_e$  differed between life stages as shown by ANOVA (all  $P$ -values  $>0.05$ ).

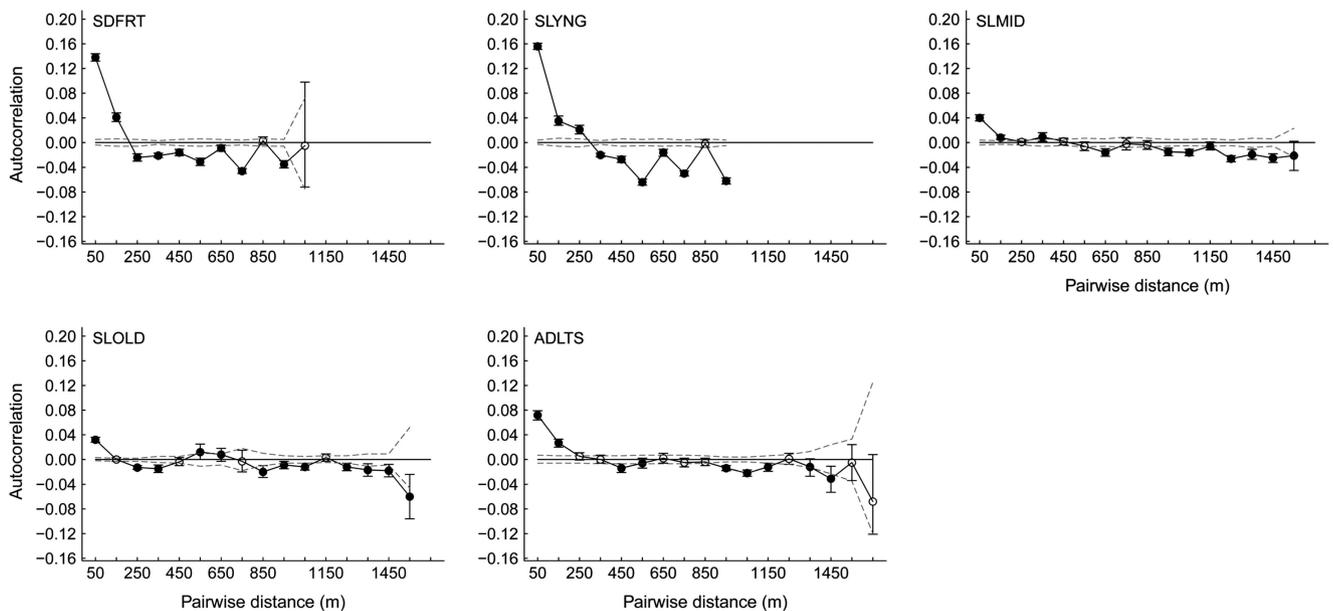
### Fine-scale SGS

We found significant SGS in all life stages (Figure 2, Table 1). Relatedness was positive and significant up to 200 m in ADLTS, SDFRT and SLMID, up to 300 m for SLYNG and up to 100 m for SLOLD. The heterogeneity test revealed that across autocorrelograms, relatedness differed among all life stages, particularly in the first distance interval (Table 2, Supplementary Table S1). Relatedness among individual pairs in the first distance interval up to 100 m was three times stronger for SDFRT and SLYNG than for SLMID and SLOLD, indicating stronger SGS in the earlier life stages (Table 1). Relatedness was intermediate in ADLTS (Table 1). Historical neighborhood sizes and gene dispersal distances (based on the adult life

**Table 1** Sample size, mean allelic richness, observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ) and  $F_{is}$ -values and their significance ( $*P < 0.0017$ ) across six microsatellite loci, jackknifed slopes of the regression of kinship ( $f_{ij}$ ) across the logarithm of geographic distance and their significance ( $b$ -log;  $*P < 0.001$ , after 10 000 permutations), autocorrelation of individuals in the first distance interval (up to 100 m), the multiclass test criterion omega and significance ( $*P = 0.001$ ) of autocorrelograms after 999 permutations in five different life stages of *Prunus africana*

Life stage	Sample size	Allelic richness	$H_o$	$H_e$	$F_{is}$	$b$ -log	$r_{(100)}$	Omega	Minimum number of pairs
SDFRT	311	16.67	0.66	0.82	0.19*	$-0.014^*$	0.14	129.33*	2692
SLYNG	309	18.06	0.75	0.82	0.08*	$-0.015^*$	0.16	128.78*	2197
SLMID	298	19.64	0.69	0.82	0.16*	$-0.001^*$	0.04	154.83*	1070
SLOLD	301	19.06	0.67	0.83	0.19*	$-0.001^*$	0.03	160.34*	290
ADLTS	260	20.30	0.77	0.83	0.08*	$-0.013^*$	0.07	132.72*	1703

Abbreviations: ADLTS, adult trees; SDFRT, propagules; SLMID, medium age seedlings; SLOLD, old seedlings and saplings; SLYNG, young seedlings. The minimum number of pairs per distance interval is given for each life stage up to a distance of 1000 m (distance used for comparison among life stages).



**Figure 2** Fine-scale correlograms for five life stages of *P. africana*. Shown are  $r$ -values  $\pm$  s.e. and 95% confidence interval per distance class (see Materials and methods section for details on distance classes) and geographic distance. Filled circles are significantly different from zero (after 999 permutations), open circles are not. ADLTS, adult trees; SDFRT, propagules; SLYNG, young seedlings; SLMID, medium age seedlings; SLOLD, old seedlings and saplings.

**Table 2** Results of the heterogeneity test testing for differences in SGS between life stages

	SDFRT	SLYNG	SLMID	SLOLD	ADLTS
SDFRT	—	0.001	0.001	0.001	0.001
SLYNG	83.69	—	0.001	0.001	0.001
SLMID	118.49	111.37	—	0.001	0.001
SLOLD	123.05	123.37	111.37	—	0.001
ADLTS	107.62	112.88	59.46	80.42	—

Abbreviations: ADLTS, adult trees; SDFRT, propagules; SLMID, medium age seedlings; SLOLD, old seedlings and saplings; SLYNG, young seedlings.  
Given are values of the multiclass test criterion omega (below diagonal) and P-values (above diagonal).

stage) were estimated to be 152 and 236 m for  $D_e = 0.00022$  and 85 and 558 m for  $D_e/10$ , respectively.

## DISCUSSION

All life stages of *P. africana* from Kakamega Forest showed SGS. Although genetic diversity did not differ across the various life stages, the strength of SGS varied considerably.

### Inbreeding and genetic diversity

For all life stages, we found lower observed ( $H_o$ ) than expected ( $H_e$ ) heterozygosity. The frequency of null alleles was problematically high only in one locus (R1) in two life stages (SLMID = 0.26, SLOLD = 0.21) and are thus not considered to be the main cause for this pattern. The significant autocorrelation in the first distance interval for adults indicates a close genetic relationship of reproductive nearby individuals. As a large proportion of pollination events normally occur within spatial clusters of conspecific trees (Berens *et al.*, 2013), biparental inbreeding (that is, mating between relatives) may explain the observed heterozygosity deficit.

Genetic diversity, in terms of allelic richness and heterozygosity, was similar throughout the life stages in *P. africana*. It shows that effective population sizes were large enough at each life stage to prevent the effects of genetic drift. It further indicates that the extent of gene flow is large for all life stages, and that no historic events have led to a fundamental alteration of processes across life stages. This is in line with other studies that have also shown that genetic diversity is similar across life stages within populations (for example, Chung *et al.*, 2003; Jacquemyn *et al.*, 2006). The levels of genetic diversity in our study were similar to those of another study on *P. africana* in Kakamega Forest (Farwig *et al.*, 2008) and to those of European *Prunus* species (Stoekel *et al.*, 2006, *P. avium*:  $H_e$ : 0.448–0.918) but higher than the diversity in an Asian *Prunus* species (Pakkad *et al.*, 2003, *P. cerasoides*:  $H_e$ : 0.292–0.689).

### Prevalence of SGS in *P. africana*

The high relatedness we found in the smallest distance categories and the significant decrease in relatedness across geographic distance in all stages can potentially be ascribed to limited seed dispersal in *P. africana*. Usually, seed dispersal by frugivorous birds and large mammals should sustain gene flow across large distances, thereby leading to low levels of SGS. In fact, indirect estimations of gene dispersal distances (both pollen and seeds combined) from microsatellites revealed distances of 236–558 m. Further, also a direct assessment of contemporary pollen and seed dispersal distances using parentage analyses revealed seed dispersal distances of up to 700 m within our study population (Berens *et al.*, 2013), with related individuals found even at far distances. However, direct estimates of

pollen and seed dispersal revealed that the majority of seeds were dispersed across short distances, leading to a mean seed dispersal distance of only 5 m (Berens *et al.*, 2013). This low average distance may result from the behavior of seed dispersers. Although numerous dispersers of *P. africana* swallow the fruits, some drop them directly under the maternal tree (Farwig *et al.*, 2006). Especially monkeys spend a lot of time in the tree feeding on fruits and a large proportion of seeds is spit out and dropped under the crown (DG Berens, personal observation). This probably leads to aggregates of closely related offspring, which was in turn mirrored in marked SGS. Our assumption is supported by the fact that SGS was not significantly different from zero in distance categories exceeding 300 m. Thus, individuals in larger spatial distances were not more related to one another than expected by chance. Thus, the patterns of SGS we found in this study agree with our direct observations and are in line with several other studies that also found a correlation between restricted seed dispersal and SGS in tropical trees (for example, Latouche-Hallé *et al.*, 2003).

In addition to distance-restricted dispersal, high light requirements of the species may also have contributed to the overall high levels of SGS across stages. Adult trees are strongly clustered near forest edges and in secondary forest patches, leading to high densities of offspring in these areas. Enhanced recruitment potential in areas of high light availability may again lead to a spatial clustering of related individuals, expressed in pronounced SGS. Previous studies have also shown that microhabitat selection can lead to SGS in plant species (Kalisz *et al.*, 2001; Ueno *et al.*, 2002). Thus, a combination of distance-restricted seed dispersal and microhabitat requirements are potentially the main drivers of SGS found in our study.

### SGS changes across life stages

In contrast to genetic diversity, the strength of SGS differed across life stages in *P. africana*. SGS was pronounced in early stages (SDFRT and SLYNG), but drastically decreased from young to older juvenile stages (SLMID, SLOLD). SGS in adults (ADLTS) was slightly stronger than in old juvenile stages. Other studies looking at SGS from a multistage perspective have either found an increase (for example, Jacquemyn *et al.*, 2006), decrease (for example, Zhou and Chen, 2010) or no change (for example, Berg and Hamrick, 1995) in SGS with age. This absence of a common pattern highlights the importance of species- and site-specific postdispersal and early selection, as well as differences across generations (for example, owing to different levels of gene flow).

### Decrease of SGS from early to late juvenile stages

We consider several valid explanations for the observed decrease in SGS from early to late recruitment stages. One likely cause could be an interannual variation in allele frequencies within the older juveniles and adults. Indeed, SLMID, SLOLD and ADLTS life stages comprise individuals stemming from different recruitment years. Individual pollen and seed dispersal patterns within the *P. africana* population may have varied among these years, producing distinct SGS patterns. In addition, pooling individuals across different recruitment years may cause an overall veiling of the SGS in comparison with SGS formed in progeny of a single year. A second likely factor causing the decrease in SGS could be offspring mortality (that is, stand thinning with age), which would reduce relatedness within offspring clumps.

In their study on *Jacaranda copaia*, Jones and Hubbell (2006) also detected a decrease in SGS from young seedlings to saplings. Comparable to many other tree species (Petit and Hampe, 2006), recruitment rates of young *P. africana* seedlings to older saplings are

low, potentially because of high light requirements (Tsingalia, 1989; Tesfaye *et al.*, 2002). Thus, multiple mortality factors, both random and non-random, may lead to the observed thinning and drive the initial decrease in SGS. In fact, our analyses of spatial clustering revealed thinning of individuals from offspring clusters across life stages and support these results. Specifically, differing pollen and seed dispersal distances directly estimated at different recruitment stages by means of parentage analyses indicated a non-random density and distant-dependent mortality in our study population, which could be driving the observed patterns (Berens *et al.*, 2013).

### Stronger SGS in adults than in late juveniles

Contrary to our expectations, SGS was stronger in adult trees than in late juvenile stages. We consider several valid explanations for this pattern. First, one has to consider that differences in SGS between juvenile stages and the adult stage may actually mirror differences between generations, for example, in gene flow and demographic processes. Jones and Hubbell (2006) attributed the decrease in SGS from adults to subadults in *J. copaia* to non-equilibrium demographic processes. *J. copaia* shows strong temporal variation in reproduction, suggesting that adult trees are descendants of one or few past recruitment events (Jones and Hubbell, 2006). Although phenology data do not suggest such irregular massive recruitment pulses in *P. africana* (DG Berens, personal observation), the extent and spatial patterns of seed dispersal events and selection processes may have been different at the time the adult generation established (Farwig *et al.*, 2008). Although the study area nowadays is well protected, the extent of historic anthropogenic influences, such as selective logging and habitat alteration are unknown. Past logging may have led to altered light regimes and may have affected seed disperser communities at the time the adult generation recruited, before the establishment of the nature reserve in 1967. Furthermore, adjacent forest areas had potentially been deforested at the beginning of the last century (Mitchell *et al.*, 2009), which could have led to reduced gene flow at the time of the parental generation (the ADLTS class) establishment.

Moreover, a founder effect with few, related individuals in the founding generation could have led to stronger SGS in the adult population (Jones and Hubbell, 2006; Pardini and Hamrick, 2008). Few patches in our study area had been clear-cut >50 years ago. Recolonization of formerly clear-cut patches by single individuals may have led to patterns similar to those deriving from a founder effect within the adult population. However, an overall founder effect causing high levels of SGS in adults is unlikely because the *P. africana* population we studied is situated in mostly primary forest (Mitchell *et al.*, 2009). Furthermore, if the adult population had established from a founder generation, adults should show a lower allelic richness than the younger stages, as allelic richness would have increased with time because of gene flow (Jones and Hubbell, 2006).

Finally, assuming that the population is in equilibrium and ADLTS of the previous generation are representative of the establishing generation, the differences in SGS between adults and late juveniles in *P. africana* could also be driven by microhabitat heterogeneity and local selection (Linhart and Grant, 1996; Jones and Hubbell, 2006). Related individuals may have similar microhabitat requirements. Consequently, spatial heterogeneity in abiotic conditions, for example, soil nutrients or especially light environment, would allow for the survival of different groups of related individuals in later life stages leading to significant SGS in adults despite lower SGS in late juveniles (Jones and Hubbell, 2006). If related individuals require similar microhabitat conditions, and generations of related individuals

overlap, neighboring individuals can be a mixture of related parents and progeny, leading to pronounced SGS (Jones and Hubbell, 2006).

### Limitations

SGS is a complex function of the density of individuals, behavior of pollinators and seed dispersers, and site history (Jones *et al.*, 2006). Moreover, the ability to detect SGS is highly influenced by sampling schemes (Vekemans and Hardy, 2004). Although we have a rather detailed understanding on how the ecological processes pollination and seed dispersal may have driven the observed patterns of SGS, our knowledge on site history is rather scarce. Thus, conclusions based on potential historic effects remain speculative, although site history is an important determinant of current SGS in tree populations. Moreover, one potential caveat of our study lies in different sampling designs for different life stages (transects vs sampling stations). Naturally, older recruitment stages occur in low densities, which makes sampling along transects an appropriate method to achieve a representative sample of the population. In contrast, young recruitment stages occur in high density. Therefore, we had to base our sampling scheme on sampling stations to keep field work tractable. Consequently, sampling of older life stages mirrored the clustered distribution of *P. africana* trees, while for younger stages sampling in areas of *P. africana* tree clusters may have been underrepresented. Moreover, for SDFRT and SLYNG, we used only a small proportion of the offspring found under *Prunus* trees. However, this sampling scheme should give conservative SGS estimates for these life stages. Thus, we still consider the conclusions drawn from our results to be valid.

### CONCLUSIONS

Changes in genetic structure have to be conceived as a multi-step process. Caused mostly by distance-restricted seed dispersal, the seed rain and early seedling stages of our study species showed significant SGS, resulting in spatial clusters of closely related individuals. During the next transition to older seedling and sapling stages, random and non-random mortality have led to the thinning of family clusters and, eventually, to the reduction of SGS. In addition, across generations, non-equilibrium processes, spatial heterogeneity in microhabitat conditions and overlapping generations are potential drivers of changes in SGS between adults and juveniles of the next generation. Corroborating the findings of previous studies, our results highlight the importance of a multistage perspective to understand genetic patterns within populations of trees. Further, the comparison of data on SGS and on contemporary gene flow across different life stages can help to assess the development of SGS of trees.

### DATA ARCHIVING

The genotype data of all individuals of the five life stages are available from the Dryad Digital Repository: doi:10.5061/dryad.39fs7.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

We thank the KWS for granting access to Kakamega Forest. Fieldwork greatly benefited from the assistance of BB Chityui, NK Sajita, S Tillmann, J Kitschke, R Berens, R Dohm and numerous other people. We also thank J Albrecht, N Farwig, A Hampe, S Liepelt and three anonymous reviewers for useful input and E and N Fouts for English proofreading. Financial support was provided by the BMBF (01LC0625E1). RN gratefully acknowledges support from the Israel Science Foundation (ISF-FIRST 1316/05 and ISF 150/07), the US National Science Foundation (DEB-0453665), the Adelina and Massimo Della

Pergola Chair of Life Sciences, and the Friedrich Wilhelm Bessel Research Award of the Humboldt Foundation. DGB is funded by the Robert Bosch Foundation. This study is part of the PhD thesis of DGB at the University of Mainz.

- Banks SC, Peakall R (2012). Genetic spatial autocorrelation can readily detect sex-biased dispersal. *Mol Ecol* **21**: 2092–2105.
- Berens DG, Griebeler EM, Braun C, Chity BB, Nathan R, Böhning-Gaese K (2013). Changes of effective gene dispersal by pollen and seeds across successive life stages in a tropical tree. *Oikos* **122**: 1616–1625.
- Berg EE, Hamrick JL (1995). Fine-scale genetic-structure of a Turkey oak forest. *Evolution* **49**: 110–120.
- Born C, Hardy OJ, Chevallier M-H, Ossari S, Attéké C, Wickings EJ *et al.* (2008). Small-scale spatial genetic structure in the Central African rainforest tree species *Aucoumea klaineana*: a stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. *Mol Ecol* **17**: 2041–2050.
- Chung MY, Epperson BK, Chung MG (2003). Genetic structure of age classes in *Camellia japonica* (Theaceae). *Evolution* **57**: 62–73.
- Cipriani G, Lot G, Huang WG, Marrazzo MT, Peterlunger E, Testolin R (1999). AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L.) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. *Theor Appl Genet* **99**: 65–72.
- Connell JH (1971). On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: den Boer PJ, Gradwell GR (eds) *Dynamics of Populations*. Centre for Agricultural Publishing and Documentations: Wageningen, pp 298–310.
- Dirlwanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A *et al.* (2002). Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor Appl Genet* **105**: 127–138.
- El Mousadik A, Petit RJ (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theor Appl Genet* **92**: 832–839.
- Farwig N, Böhning-Gaese K, Bleher B (2006). Enhanced seed dispersal of *Prunus africana* in fragmented and disturbed forests? *Oecologia* **147**: 238–252.
- Farwig N, Braun C, Böhning-Gaese K (2008). Human disturbance reduces genetic diversity of an endangered tropical tree, *Prunus africana* (Rosaceae). *Conserv Genet* **9**: 317–326.
- Frankham R (1995). Effective-population size adult-population size ratios in wildlife—a review. *Genet Res* **66**: 95–107.
- Goudet J (1995). FSTAT (Version 1.2): a computer program to calculate F-statistics. *J Hered* **86**: 485–486.
- Hall JB, Ó'Brien EM, Sinclair FL (2000). *Prunus africana*—a monograph. School of Agricultural and Forest Sciences. Publication Number 18, University of Wales, Bangor.
- Hamrick JL, Murawski DA, Nason JD (1993). The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio* **107–108**: 281–297.
- Hardy OJ, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.
- Heuertz M, Vekemans X, Hausman JF, Palada M, Hardy OJ (2003). Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Mol Ecol* **12**: 2483–2495.
- Jacquemyn H, Brys R, Vandepitte K, Honnay O, Roldan-Ruiz I (2006). Fine-scale genetic structure of life history stages in the food-deceptive orchid *Orchis purpurea*. *Mol Ecol* **15**: 2801–2808.
- Janzen DH (1970). Herbivores and number of tree species in tropical forests. *Am Nat* **104**: 501–528.
- Jones FA, Hubbell SP (2006). Demographic spatial genetic structure of the neotropical tree *Jacaranda copaia*. *Mol Ecol* **15**: 3205–3217.
- Jones FA, Hamrick JL, Peterson CJ, Squires ER (2006). Inferring colonization history from analyses of spatial genetic structure within populations of *Pinus strobus* and *Quercus rubra*. *Mol Ecol* **15**: 851–861.
- Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* **16**: 1099–1106.
- Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001). Spatial population genetic structure in *Trillium grandiflorum*: the roles of dispersal, mating, history, and selection. *Evolution* **55**: 1560–1568.
- KIFCON (1994). Kakamega guide. The official guide. Kenya Indigenous Forest Conservation Programme, Nairobi.
- Latouche-Hallé C, Ramboer A, Bandou E, Caron H, Kremer A (2003). Nuclear and chloroplast genetic structure indicate fine-scale spatial dynamics in a neotropical tree population. *Heredity* **91**: 181–190.
- Linhart YB, Grant MC (1996). Evolutionary significance of local genetic differentiation in plants. *Annu Rev Ecol Syst* **27**: 237–277.
- Loveless MD, Hamrick JL (1984). Ecological determinants of genetic-structure in plant populations. *Annu Rev Ecol Syst* **15**: 65–95.
- Lung T (2004). Landbedeckungsänderungen im Gebiet 'Kakamega Forest und assoziierte Waldgebiete' (Westkenia)—Multispektrale Klassifikation von Landsat-Satellitenbilddaten und Auswertung mittels Methoden im Raster-GIS. Karlsruhe Geowissenschaftliche Schriften—Reihe A 15, Karlsruhe University of Applied Sciences, Karlsruhe, Germany.
- Mitchell N, Schaab G, Wägele JW (2009). Kakamega Forest ecosystem: an introduction to the natural history and the human context. Karlsruhe Geowissenschaftliche Schriften, Reihe A, Band 17, Karlsruhe University of Applied Sciences, Karlsruhe.
- Pakkad G, James C, Torre F, Elliott S, Blakesley D (2003). Genetic variation of *Prunus cerasoides* D. Don, a framework tree species in northern Thailand. *New Forests* **27**: 189–200.
- Pardini EA, Hamrick JL (2008). Inferring recruitment history from spatial genetic structure within populations of the colonizing tree *Albizia julibrissin* (Fabaceae). *Mol Ecol* **17**: 2865–2879.
- Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.
- Petit RJ, Hampe A (2006). Some evolutionary consequences of being a tree. *Ann Rev Ecol Evol Syst* **37**: 187–214.
- Peakall R, Smouse PE (2012). GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research— an update. *Bioinformatics* **28**: 2537–2539.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, URL: <http://www.R-project.org/>.
- Schueler S, Tusch A, Schuster M, Ziegenhagen B (2003). Characterization of microsatellites in wild and sweet cherry (*Prunus avium* L.)—markers for individual identification and reproductive processes. *Genome* **46**: 95–102.
- Smouse PE, Peakall R, Gonzales E (2008). A heterogeneity test for fine-scale spatial genetic structure. *Mol Ecol* **17**: 3389–3400.
- Sosinski B, Gannavarapu M, Hager LD, Beck LE, King GJ, Ryder CD *et al.* (2000). Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. *Theor Appl Genet* **101**: 421–428.
- Stoekel S, Grange J, Fernandez-Manjarres JF, Bilger I, Frascaria-Lacoste N, Mariette S (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species—*Prunus avium* L. *Mol Ecol* **15**: 2109–2118.
- Swaine M, Whitmore TC (1988). On the definition of ecological species groups in tropical rain forests. *Vegetatio* **75**: 81–86.
- Tesfaye G, Teketay D, Fetene M (2002). Regeneration of fourteen tree species in Harenna forest, southeastern Ethiopia. *Flora* **197**: 461–474.
- Tsingalia MH (1989). Variation in seedling predation and herbivory in *Prunus africana* in the Kakamega Forest, Kenya. *Afr J Ecol* **27**: 207–217.
- Ueno S, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2002). Size-class differences in genetic structure and individual distribution of *Camellia japonica* L. in a Japanese old-growth evergreen forest. *Heredity* **89**: 120–126.
- Vaughan SP, Russell K (2004). Characterization of novel microsatellites and development of multiplex PCR for large-scale population studies in wild cherry *Prunus avium*. *Mol Ecol Notes* **4**: 429–431.
- Vekemans X, Hardy OJ (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* **13**: 921–935.
- Wang H, Qi MQ, Cutler AJ (1993). A simple method of preparing plant samples for PCR. *Nucleic Acids Res* **21**: 4153–4154.
- Zhou H-P, Chen J (2010). Spatial genetic structure in an understory dioecious fig species: the roles of seed rain, seed and pollen-mediated gene flow, and local selection. *J Ecol* **98**: 1168–1177.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)