Phylogeographic pattern and regional evolutionary history of the maize stalk borer *Busseola fusca* (Fuller) (Lepidoptera: Noctuidae) in sub-Saharan Africa

Michel Sezonlin[1,3,*], Stéphane Dupas[1], Bruno Le Rü[2], Nathalie Faure[1], Philippe Le Gall[1] & Jean-François Silvain[1]

[1] IRD, UR 072 c/o CNRS, UPR 9034, Lab. Evolution, Génomes et Spéciation, avenue de la Terrasse, F-91198 Gif-sur-Yvette, France

[2] IRD, UR 072, c/o ICIPE, Noctuid Stem Borers Biodiversity Project, PO Box 30772, Nairobi, Kenya

[3] IRD, UR 072, c/o IITA, Biological Control Centre for Africa, 08 BP 0932 Cotonou, Bénin

* Corresponding author

Abstract. *Busseola fusca* (Fuller) (Lepidoptera: Noctuidae) is one of the major cereal pests in sub-Saharan Africa. Previous phylogeographic investigations on samples collected in Kenya, Cameroon and West-Africa showed the presence of three main clades (*W, KI, KII*) originated from populations isolated in West and East Africa around one million years ago. Demographic and phylogenetic analyses suggested that this event was followed by local demographic expansion and isolation by distance. These hypotheses were tested by a more comprehensive sampling across *B. fusca*'s geographic range in Africa. Comparisons of sequences of partial mitochondrial DNA gene (cytochrome *b*) from 489 individuals of 98 localities in southern, central, eastern and western African countries confirmed the presence of the three main clades. Phylogenetic, F-statistics, demographic parameters and nested clade phylogeographic analyses confirmed that the clades experienced geographic and demographic expansion with isolation by distance after their isolation in three refuge areas. The geographic range of clade *KII*, already known from East to Central sub-Saharan Africa was extended to Southern Africa. Mismatch distribution analysis and the negative values of Tajima's D index are consistent with a demographic expansion hypothesis for these three clades. Significant genetic differentiations were revealed at various hierarchical levels by analysis of molecular variance (AMOVA). Hypotheses about the geographic origin of the three main clades are detailed.

Résumé. Scénario phylogéographique et histoire évolutive régionale du foreur de graminées *Busseola fusca* (Fuller) (Lepidoptera : Noctuidae) en Afrique sub-saharienne. *Busseola fusca* (Fuller) (Lépidoptère : Noctuidae) est l'un des ravageurs majeurs des cultures céréalières en Afrique Subsaharienne. Une première étude phylogéographique portant sur des individus échantillonnés au Kenya, au Cameroun et en Afrique de l'Ouest a montré l'existence de trois clades principaux (*W, KI, KII*) issus de populations isolées à l'Ouest et à l'Est de l'Afrique il y a environ un million d'années. Les analyses démographiques et phylogénétiques indiquent que cet événement a été suivi d'une expansion démographique locale avec des phénomènes d'isolement par la distance. Ces hypothèses ont été testées à plus grande échelle grâce à un échantillonnage des populations de *B. fusca* couvrant désormais la majeure partie de son aire de distribution. Le séquençage d'un fragment du gène mitochondrial codant pour le cytochrome *b* chez 489 individus provenant de 98 localités des pays sud, centre, est et ouest africains confirme l'existence des trois clades observés précédemment. Les résultats des analyses phylogénétiques, les paramètres démographiques, les statistiques de Wright ainsi que les analyses des clades emboîtés confirment que ces trois populations, après avoir été isolées dans des aires refuges différentes, ont connu une expansion démographique et géographique avec un isolement par la distance. La distribution géographique du clade *KII*, connue de l'Afrique l'Est à l'Afrique centrale, s'étend jusqu'en Afrique Australe. L'analyse de 'mismatch distribution' et les valeurs négatives de l'indice D de Tajima sont bien en accord avec l'hypothèse d'une expansion démographique de ces trois clades. Des différenciations génétiques significatives ont été révélées aux différents niveaux hiérarchiques par l'analyse moléculaire de la variance (AMOVA). Les hypothèses sur l'origine géographique des trois clades sont précisées.

Keywords: Stem borer, population genetics, cytochrome *b*, Pleistocene, Africa.

E-mail: msezonlin@yahoo.fr, dupas@legs.cnrs-gif.fr, faure@legs.cnrs-gif.fr, legall@legs.cnrs-gif.fr, silvain@legs.cnrs-gif.fr, bleru@icipe.org

Accepté le 4 mai 2006

The most important cereal crops in sub-Saharan Africa are maize, millet, rice and sorghum. Most of these cereal crops serve as host plants for many stem borer insects among which *Busseola fusca* (Fuller 1901) (Lepidoptera: Noctuidae) is one of the most economically important pests (Polaszek & Khan 1998). *B. fusca* is an endemic species in sub-Saharan Africa with wide geographical distribution (Ajayi 1998; Moyal 1998; Ndemah *et al.* 2001; Haile & Hofsvang 2001; Kfir *et al.* 2002). It is a major pest of sorghum (*Sorghum bicolor* [L.] Moench) and maize (*Zea mays* L.) (Poaceae). The domestication of sorghum probably began some 5,000 years ago in North-East Africa (Dogett 1988; Murty & Renard 2001) whereas maize was introduced more recently, at the end of the 16th Century (Madeira Santos & Ferraz Torrao 1998; Chastanet 1998). *Busseola fusca* varies in its ecological preference across its geographical range. It is more adapted to lowland in West Africa than in East and Southern Africa (Kfir *et al.* 2002). Previous study on the genetic structure of *B. fusca* shows that this ecological preference is associated with major differences in partial DNA sequence of the cytochrome *b* mitochondrial gene (Sezonlin *et al.* 2006). According to that study, *B. fusca* populations are differentiated into three major clades of mitochondrial haplotypes, one located in West African region (*W*), one restricted to East Africa (*KI*) and one found in Central and East Africa (*KII*). The origin of these clades is likely related to Pleistocene climatic events. Partial geographic overlap was observed only between clades *KI* and *KII*. Biogeographic barriers likely corresponding to *B. fusca* ancient history on wild Poaceae have been shown to be the major factors of differentiation of this species. These barriers, namely the Cameroon Volcanic Line (CVL) region between Central and West Africa and the Rift Valley in East Africa, appear to be similar to those that shaped geographic differentiation of phytophagous mammals and rodents (Sezonlin *et al.* 2006).

However, no signature of sorghum domestication and maize introduction has been detected yet on the genetic structure of *B. fusca*, despite the expected important demographic consequences of this switch to cultivated plants.

The regional evolutionary history and the centres of origin of the mitochondrial clades remain unknown. To answer these questions and to more accurately estimate geographical distribution of each clade, the sampling of *B. fusca* was completed to include most of its geographic range from Western, Eastern to Southern Africa. Samples from Eritrea, Ethiopia, Malawi, Mozambique, Rwanda, Republic of South Africa, Uganda, Zambia and Zimbabwe were added. Partial sequences of the gene coding for cytochrome *b*, informative at the intrageneric level in Lepidoptera (Simmons & Weller 2001; Sezonlin *et al.* 2006) have been used. A sequential approach combining several phylogeographic and evolutionary methods (Bernatchez 2001) was used to analyse the molecular data and infer the demographic history of *B. fusca* populations in more details. Such analyses in sequential approach that start from phylogeny to evolutionary history via demography and genetic structure allow us to move from testing deeper phylogenetic splits to inferring recent patterns of population structure. This also may highlight the centres of origin of *B. fusca* populations and may elucidate the regional evolutionary history that has produced this genetic structure.

Material and methods

Moth sampling

The sampling of *Busseola fusca* individuals was carried out between 2001 and 2004. *B. fusca* individuals were sampled in West Africa (19 localities from Benin, Togo, Ghana, Mali and Burkina-Faso), in Central Africa (3 localities from Cameroon), in East Africa (55 localities from Kenya, Uganda, Rwanda, Tanzania, Ethiopia and Eritrea) and in Southern Africa (21 localities from Malawi, Mozambique, Zambia, Zimbabwe and Republic of South Africa) (figs.1a, b, c, d).

Moths rearing and conservation

Larvae and pupae collected were brought to the laboratories (IITA - Cotonou for Central and West Africa and ICIPE - Nairobi for Southern and Eastern Africa) to be reared to adulthood on semi-natural medium made of fresh stems of maize and cultivated sorghum (IITA) and artificial medium (ICIPE). The rearing of larvae allows the morphological identification of *B. fusca* moths among other stem borers species. The moths were killed just after emergence and preserved in absolute ethanol before DNA extraction.

Molecular analysis

Total DNA was extracted from insect thoraxes, using the DNeasy tissue kit (Qiagen GmbH, Germany). The number of individuals of *B. fusca* analyzed for each locality ranged from one to 13. The molecular marker used is a fragment of the gene coding for the cytochrome *b* for which approximately 1000 bp were amplified by PCR. The same primers and PCR protocol as described by Sezonlin *et al.* (2006) were used for all samples. Amplified PCR products were purified with the Quick protocol (Promega Wizard SV Gel and PCR Clean-Up System) and directly sequenced on an automated sequencer ABI prism 377 using the amplification primers in both directions. The consensus sequences obtained were aligned manually using MacClade 4.06 (Maddison & Maddison 2002).

Haplotype phylogeny

Phylogenetic relationships were estimated by means of maximum parsimony (MP) and Neighbour-Joining (NJ) using Maximum-Likelihood distances with PAUP* 4b10 (Swofford

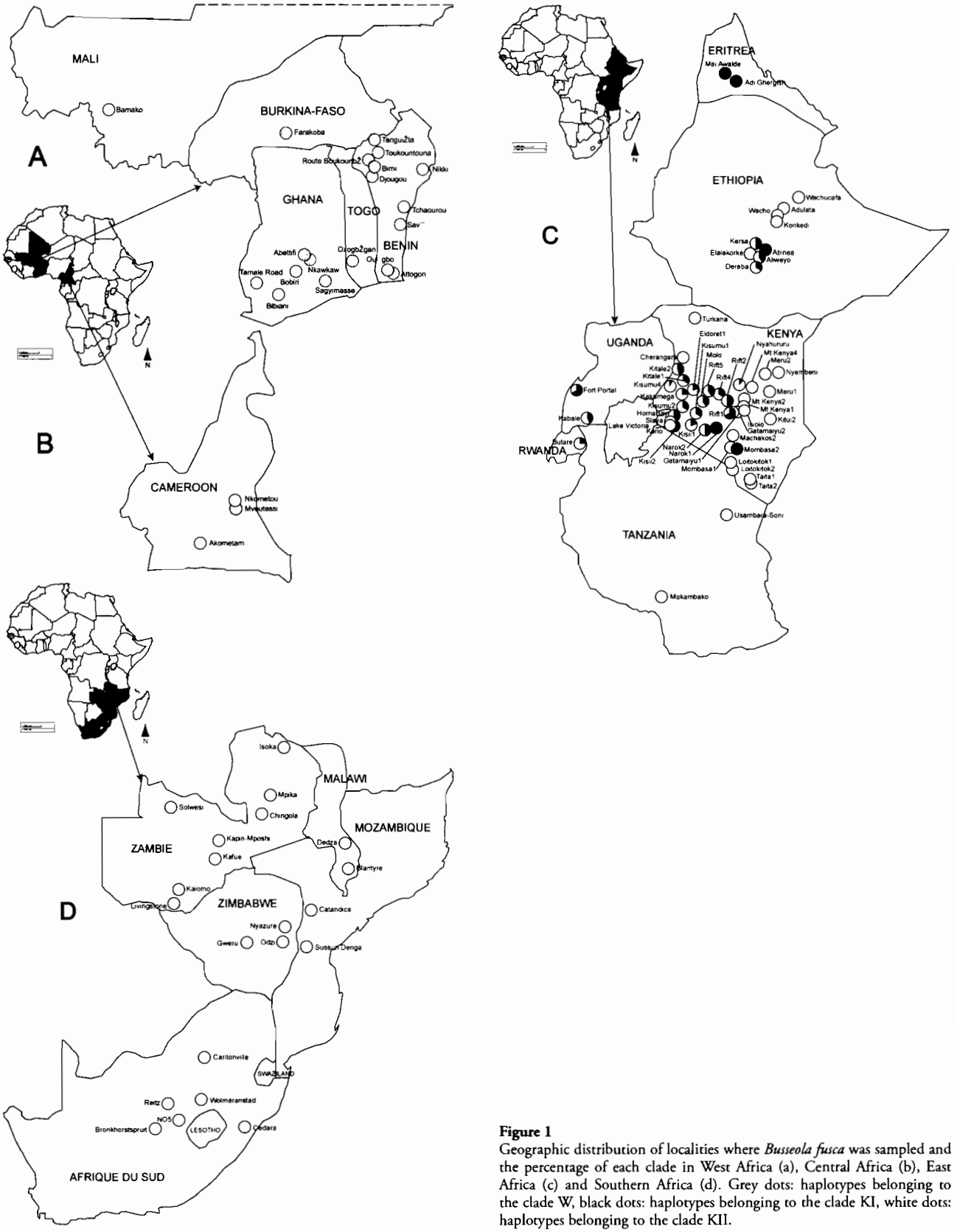

Figure 1
Geographic distribution of localities where *Busseola fusca* was sampled and the percentage of each clade in West Africa (a), Central Africa (b), East Africa (c) and Southern Africa (d). Grey dots: haplotypes belonging to the clade W, black dots: haplotypes belonging to the clade KI, white dots: haplotypes belonging to the clade KII.

2002). MP analyses were performed using a heuristic search strategy starting with stepwise addition trees replicated 10 times, using a random input order of sequences to get the initial tree for each replicate. Robustness of MP topologies was assessed by bootstrap with 100 replicates (full heuristic search) of 10 random stepwise addition replicates each, for all analyses.

Although the systematics of African lepidopteran stem borers is still rather confused (Holloway 1998), recent studies showed that *Busseola phaia* Bowden 1956 (Lepidoptera: Noctuidae), collected from various regions of East Africa, is the sister species of *B. fusca* (Moyal, *pers. com.*). Therefore, *B. phaia* was chosen as outgroup.

MODELTEST version 3.07 (Posada & Crandall 1998) was used to select the substitution model(s) that best describe the data. This software performs a hierarchical test of likelihood fits under 56 different models of character variation.

Diversity indices and demographic history of *Busseola fusca* clades

Haplotypic (h) and nucleotide (π) diversity values and all demographic parameters were performed with ARLEQUIN 2.000 software (Schneider *et al.* 2000) for groups of haplotypes in order to estimate their level of polymorphism and to localize the centre of origin of the different clades. These diversity indices are useful to examine the demographic history of a lineage (Grant & Bowen 1998) because their value does not depend on the length of the DNA fragment, nor on the sample size (Nei & Li 1979; Nei 1987). The centre of origin of each *B. fusca* clade was established by comparing the genetic diversity for different groups defined within this main population. In this case, populations were grouped according to geography in Central, South and East Africa and phytogeographic zones (White 1983) in West Africa. In West Africa three phytogeographic zones concerned our study, forest (drier types) region, forest and secondary grassland region and savannah region whereas in the rest of Africa, four geographic zones were retained, namely East – North, East – South, Austral and Cameroon. Centres of origin would have higher haplotype and nucleotide diversity than more recently founded populations (Althoff & Pellmyr 2002). This analysis could allow us to track the recent geographical expansions of these populations. The distribution of pairwise differences between individual sequences was analyzed by means of mismatch distribution analysis (Slatkin & Hudson 1991; Schneider & Excoffier 1999). A unimodal distribution would be expected for populations in expansion or for populations that have undergone a recent bottleneck, and a multimodal distribution for populations at demographic equilibrium (Slatkin & Hudson 1991). The raggedness index of the observed distribution (r) representing the modality of the distribution, and the sum of square deviation from the mismatch expected from a model of sudden population expansion (SSD) were calculated. Since the nucleotide substitution models selected by hierarchical likelihood ratio tests (hLRTs) (HKY + I + G) and Akaike information criterion (AIC) (K81uf + I + G) were not available in the ARLEQUIN 2.000 software, the r and SSD indices were calculated by using pairwise differences. The significance of these statistics was tested as implemented in ARLEQUIN. Finally, Tajima's D index was calculated with ARLEQUIN. This index can provide information about demographic history with demographic expansion leading to negative values, and subdivided populations leading to positive values (Tajima 1989a, b).

Genetic structure of *Busseola fusca* populations

In order to test for genetic differentiation, hierarchical levels of genetic divergence between various groups were calculated with the fixation index Φ_{ST} (Excoffier *et al.* 1992), an estimator that includes information on haplotype frequency and molecular distance. The significance of Φ_{ST} for population comparisons was assessed using 1000 permutations. The Φ_{ST} values and permutations were computed in ARLEQUIN 2.000 (Schneider *et al.* 2000).

Nested clade phylogeographic analysis (NCPA)

NCPA was performed as described by Templeton (1998, 2004). The genealogic relationships are represented through a haplotype parsimony network to define a series of nested clades. The probabilities of haplotype connections were calculated according to coalescent theory using TCS1.21 software (Clement *et al.* 2000) and the network with probabilities above the parsimony threshold (0.95) was selected. The hypothesis of random geographic distributions is tested through permutation tests for each clade and subclade components. These statistical analyses of geographical distances within and between clades were carried out with GeoDis 2.1 (Posada *et al.* 2000). GPS coordinates of all sampling localities were used. The geographical distances between centres of distributions of clades were tested for significance in permutation tests, within clade (*Dc*, the average distance of individuals from the clade's geographical centre), with nested clade centre (*Dn*, the average distance of individuals from the geographical center of all members of the nested clade) or between interior and tip at each level (*(I-T)Dc*, the average distance between interior and tip clades within a given clade and *(I-T)Dn*, the average distance between interior and tip clades in the nested clade). Significant geographic patterns were interpreted in terms of population history, using the latest inference key from Templeton (2004) from http://darwin.uvigo.es.

Results

Phylogenetic reconstruction

A fragment of 965 bp encoding cytochrome *b* was sequenced from 489 individuals of *Busseola fusca* across its geographic range from Western, Central, Southern and Eastern Africa. We observed 108 different haplotypes (GenBank accession numbers AY769536 to AY769605 and DQ284857 to DQ284895). The haplotypes with GenBank accession number AY769536 to AY769605 were re-used whereas those with accession number DQ284857 to DQ284895 were news. 123 nucleotide sites were variable (12.75%) and 58 were informative in parsimony analysis (6.01%). Parsimony analysis generated 438 equiparsimonious trees (length = 342, CI = 0.371, RI = 0.808). All trees were divided into the same three clades: a clade grouping sequences from the West African region only (*W*), a Kenya I clade (*KI*) and a Kenya II clade (*KII*), which also contained sequences from Cameroon and southern African countries (fig. 2). Discrepancies between these 438

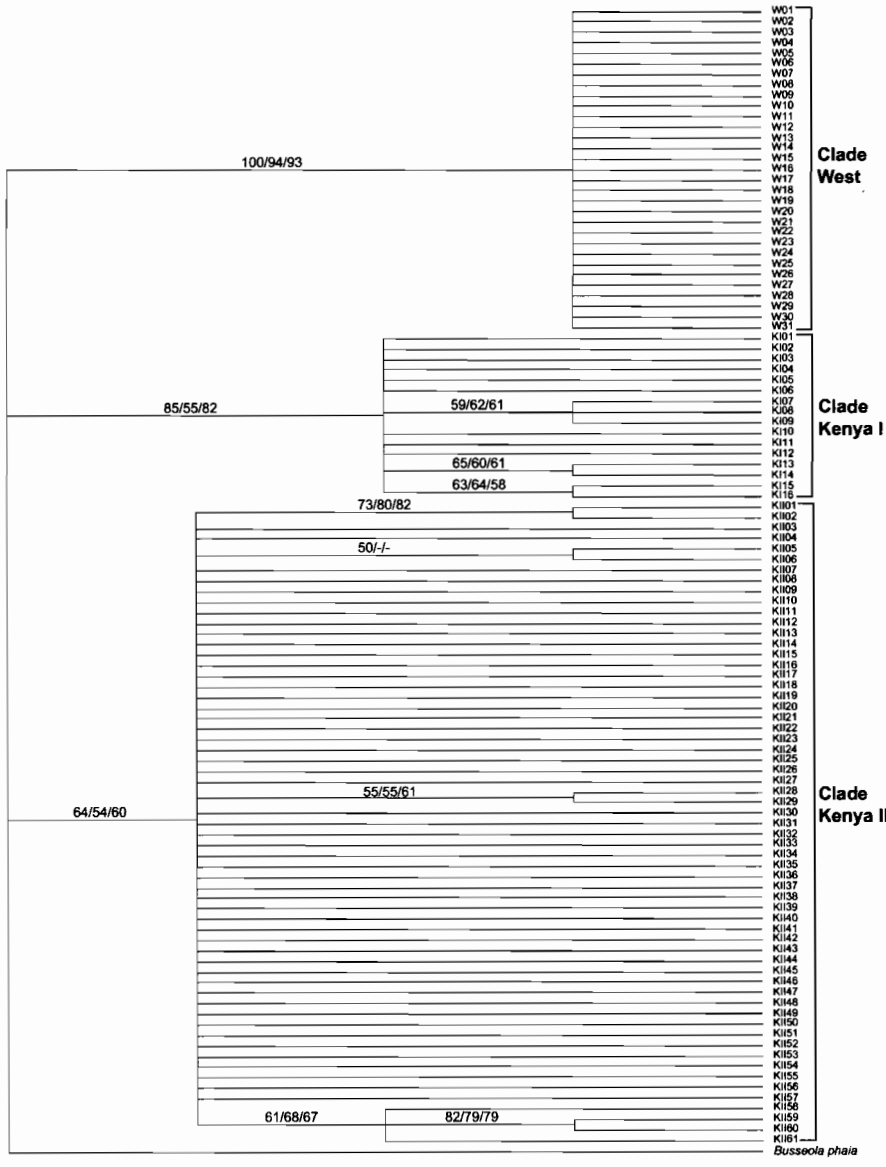

Figure 2
Majority consensus of the most parsimonious trees. Bootstrap support of >50% in both MP (first number), NJ with hLRT (second number), and NJ with AIC (third number) searches in 1000 replicates are given for the relevant nodes. *Busseola phaia* was used as the outgroup taxon.

equiparsimonious topologies concerned only the apical nodes. It was therefore possible to construct a majority rule consensus of the most parsimonious trees (fig. 2).

The model selected by the Akaike information criterion of the maximum likelihood (ML) was the K81uf + I + G (-LnL = 2479.6731) (Kimura 1981). The parameters inferred from this substitution model were: A =0.337, C= 0.140, G = 0.106, T = 0.417; [AC substitution rate] = 1.000, [AG] = 22.5462, [AT] = 0.3998, [CG] = 0.3998, [AT] = 22.5462, [GT] = 1.000 with a certain proportion of invariable sites (I = 0.6918) and heterogeneous rate of substitution following a gamma distribution with alpha shape $\alpha = 0.7750$. According to the hierarchical likelihood ratio tests (hLRT), the HKY + I + G model of evolution (-LnL =2480.8875) (Hasegawa *et al.* 1985; Yang 1993;

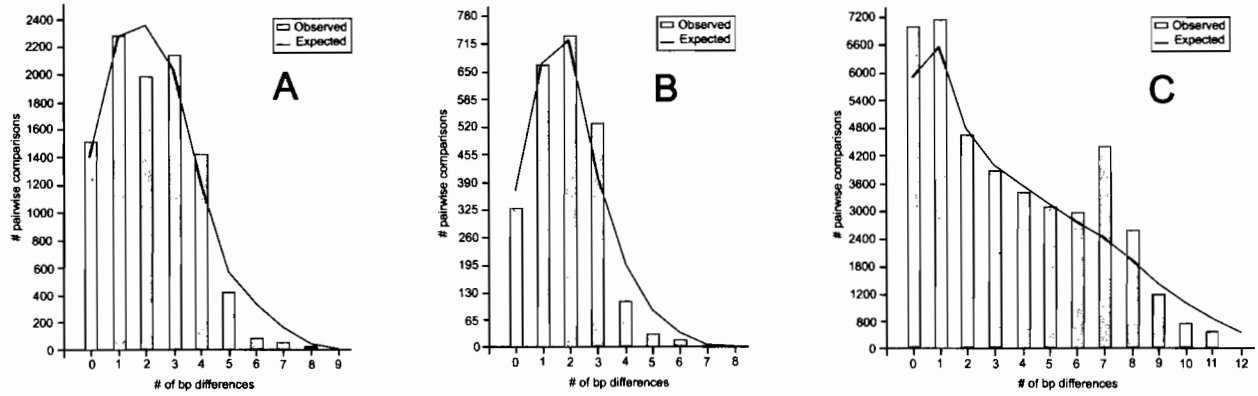

Figure 3
Mismatch distribution analysis showing histogram of observed and expected mismatch frequencies (A: West Africa population; B: Kenya I population; C: Kenya II population).

Gu *et al.* 1995) was selected. The parameters estimations were I = 0.6904 and α = 0.7714. Neighbour-Joining analyses of the ML distances obtained using the parameter estimates derived from each substitution model were performed. Bootstrap values for each NJ analysis were obtained from 1000 replications. Topology of NJ tree obtained using ML distances was similar to the one derived from MP analyses (fig. 2).

As already pointed out by Sezonlin *et al.* (2006), the three conspicuous clades of individuals and haplotypes are supported by bootstrap values exceeding 50% in both MP an NJ analyses. The smallest clade *KI* comprised 16 haplotypes and 70 individuals, all of which came from East Africa. The clade *KII* comprised 61 haplotypes and 280 individuals and had the largest distribution from East to Central Africa via southern Africa. Finally, the clade *W* comprised 31 haplotypes and 139 individuals and was found only in West Africa (fig. 1). No haplotype was shared between West African populations and East-Central-Southern African populations. Both in MP and NJ analyses, *W* and *KI* were supported by high bootstrap values whereas *KII* was supported by lower bootstrap values. The phylogenetic relationships between these three major clades remain unresolved. The sister group status of clades *W* and *KI* was observed only in NJ analyses using the substitution models selected by hLRT and AIC criterion. In both NJ analyses, the bootstrap values remain low (55% and 56% respectively for hLRT and AIC).

Table 1. Estimates of haplotype and nucleotide diversity for different population groupings of *B. fusca*. Forest region populations have the highest diversity for clade *W*, East-North region populations have the highest diversity for clade *KI*, East-South region populations have the highest diversity for clade *KII*.

Clade	Region	Haplotype diversity	Nucleotide diversity (%)
W	Forest (drier types)	0.879 +/- 0.024	0.254 +/- 0.155
	Forest and secondary grassland	0.772 +/- 0.094	0.209 +/- 0.137
	Savannah	0.659 +/- 0.072	0.154 +/- 0.105
KI	East – North	0.897 +/- 0.041	0.209 +/- 0.138
	East – South	0.789 +/- 0.032	0.166 +/- 0.110
KII	East – North	0.495 +/- 0.151	0.256 +/- 0.165
	East – South	0.862 +/- 0.021	0.379 +/- 0.213
	Austral	0.720 +/- 0.069	0.108 +/- 0.080
	Cameroon	0.199 +/- 0.112	0.029 +/- 0.020

Genetic structure of the *Busseola fusca* populations

Most of the molecular variation was accounted for by the differentiation between the three clades highlighted by phylogenetic analyses with $\Phi_{ST} = 0.868$ ($P < 10^{-5}$). At fine scale, the genetic structure was observed within local populations in each major clade. The different values were 0.226 ($P < 10^{-5}$), 0.238 ($P < 10^{-5}$), 0.344 ($P < 10^{-5}$) respectively for *W*, *KI* and *KII* clades.

Diversity and demographic history of *Busseola fusca*

Haplotype and nucleotide diversity were calculated for different groups considered within each clade identified (tab. 1). The values of these indices vary greatly within each clade. For clade *W*, the forest region has the highest haplotype diversity. The East-North region is more diverse for clade *KI* than for other regions. Finally for *KII*, the East-South has the highest haplotype diversity. Both the variance (SSD)

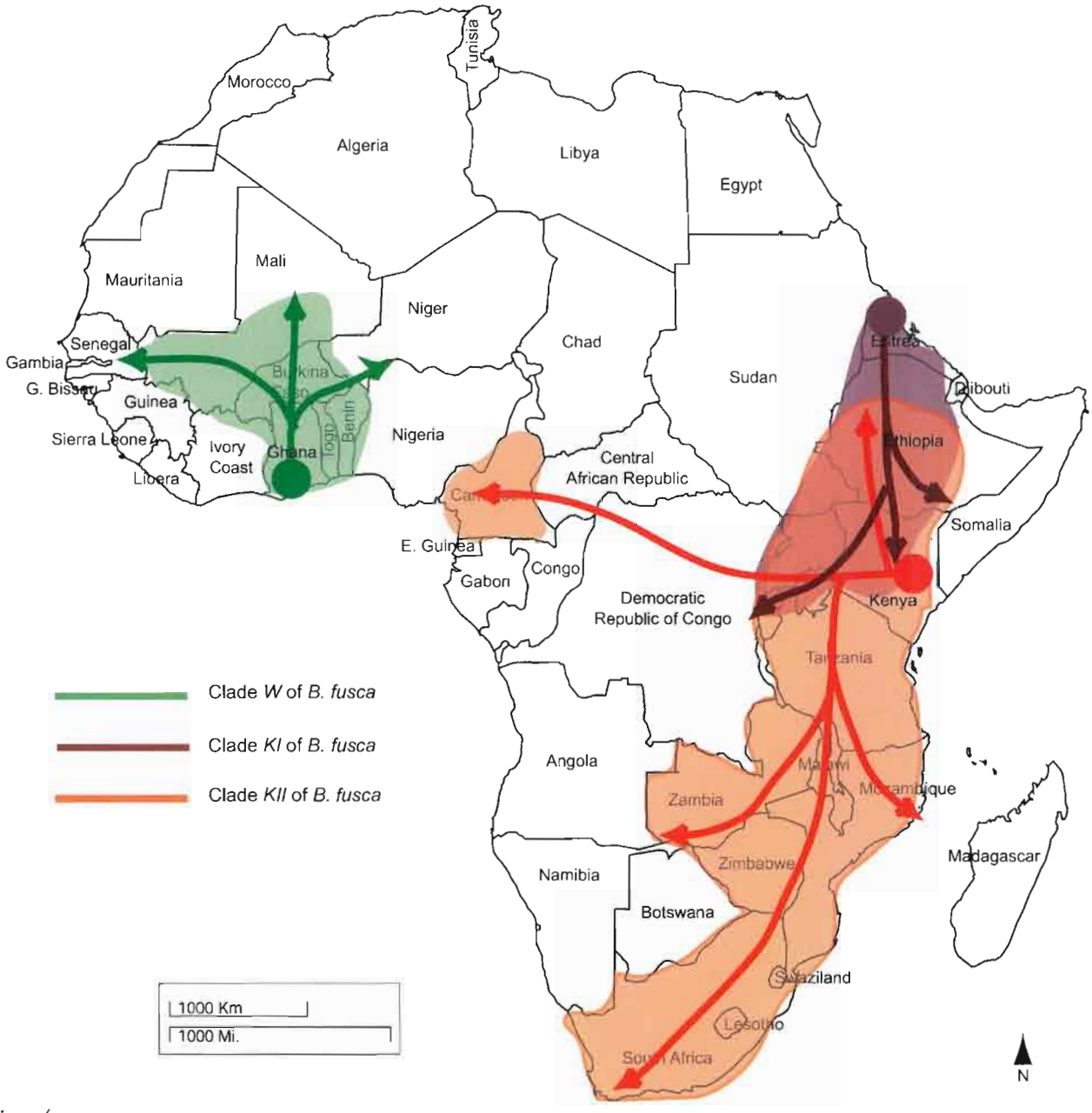

Figure 4
Current geographic distribution of *Busseola fusca* and the different putative centres of origin of its three main clades highlighted by phylogenetic analyses.

and raggedness index (r) tests suggested that the curves (figs 3a, 3b, 3c) do not significantly differ from the distribution under a model of population expansion (P_{SSD} = 0.41 and P_r = 0.69 for *W*; P_{SSD} = 0.10 and P_r = 0.27 for *KI*; P_{SSD} = 0.87 and P_r = 0.94 for *KII*). Similarly, the negative values obtained for Tajima's D index for each clade (-1.62105; -1.63863; -1.49025 for clades *W*, *KI* and *KII*, respectively) are all consistent with the hypothesis of population expansion since the origin of the clades. The current geographic distribution of *B. fusca* with centre of origin of each clade is illustrated fig. 4.

Nested clade phylogeographic analysis

The NCPA network calculation identified the same three clades revealed by MP and NJ analyses. The networks of these three clades were represented by figs 5a, 5b, 5c. The West African clade contained the 31

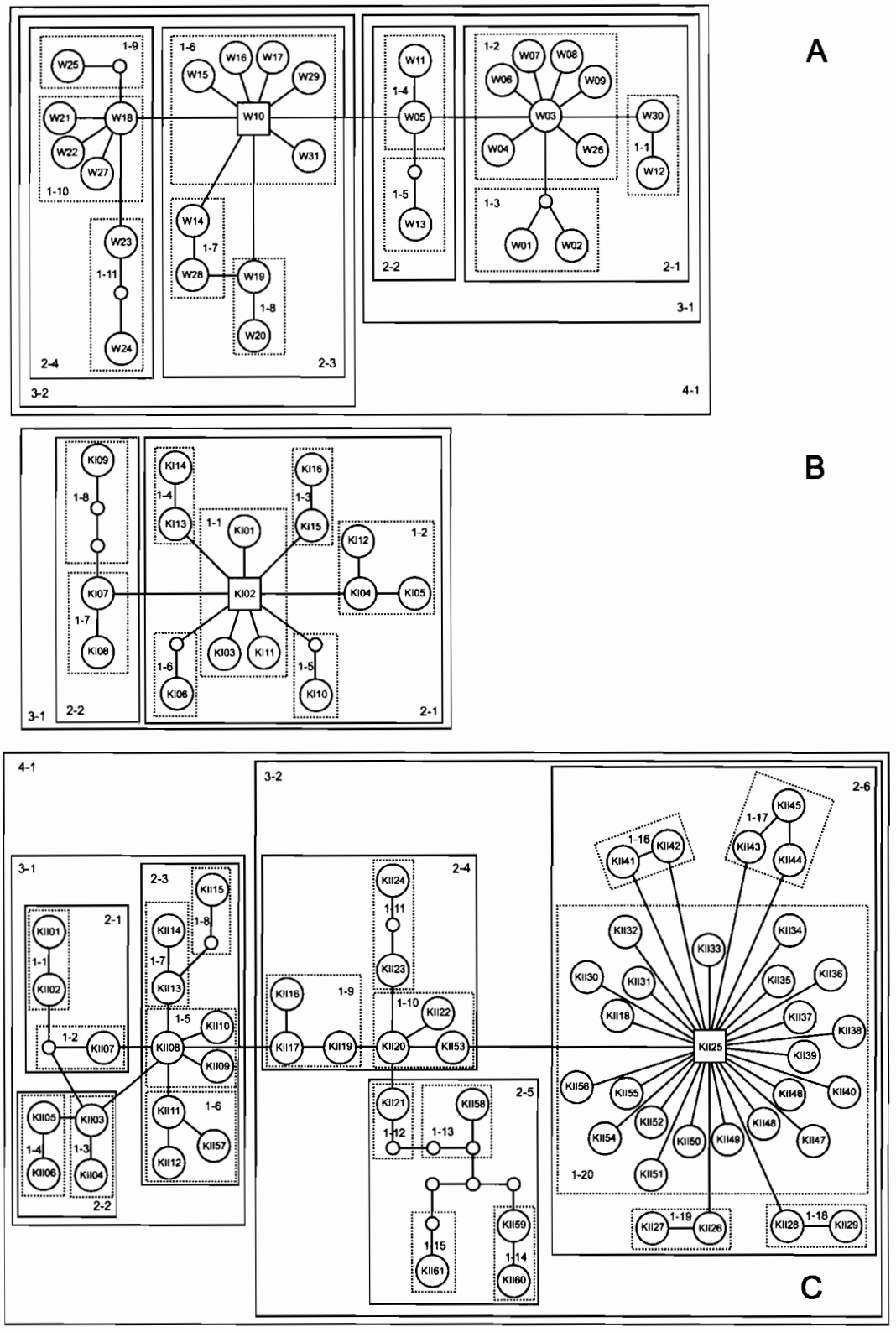

Figure 5
Haplotype network of all haplotypes detected for *Busseola fusca*. Each haplotype is labelled by its number. Hypothetical haplotypes are designated by small circles (A: West Africa population; B: Kenya I population; C: Kenya II population).

haplotypes observed exclusively in this region and four hypothetical intermediate haplotypes, hierarchically grouped into 11 'one-step clades', four 'two-step clades', and two 'three-step clades'. The *KI* clade contained the 16 observed haplotypes and four hypothetical haplotypes: 8 'one-step clades' and 2 'two-step clades'. Finally, the *KII* clade, with 61 observed haplotypes and 10 hypothetical is organized into 20 'one-step' clades, 6 'two-step clades' and 2 'three-step clades'.

Nested contingency analysis on the haplotype network revealed significant geographic associations in the three major networks at all clade levels. These significant values were interpreted using Templeton's (2004) inference key (tab. 2). Most of the clades displaying geographic associations were interpreted by restricted gene flow with isolation by distance although some of them were interpreted as restricted gene flow or dispersal but with some long distance dispersal and one by contiguous range expansion.

Table 2. Inference chain results of geographical distance analysis from Fig. 5a, 5b, 5c.

Clade	Chain of inference	Inference
1 – 1 (*W*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
1 – 2 (*W*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
1 – 4 (W)	1-2-3-4: NO	Restricted gene flow with isolation by distance
1 – 6 (*W*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
1 – 10 (*W*)	1-2-11-17: NO	Inconclusive outcome
2 – 1 (*W*)	1-2-3-5-6-7: YES	Restricted gene flow / Dispersal but with some long distance dispersal
2 – 2 (*W*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
2 – 3 (*W*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
2 – 4 (*W*)	1-2-11-12: NO	Contiguous range expansion
3 – 2 (*W*)	1-2-3-5-6-7: YES	Restricted gene flow / Dispersal but with some long distance dispersal
1 – 2 (*KI*)	1-2-3-5-6-7: YES	Restricted gene flow / Dispersal but with some long distance dispersal
2 – 1 (*KI*)	1-2-3-5-6-7: YES	Restricted gene flow / Dispersal but with some long distance dispersal
1 – 3 (*KII*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
1 – 5 (*KII*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
1 – 18 (*KII*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
1 – 20 (*KII*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
2 – 2 (*KII*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
2 – 3 (*KII*)	1-2-11-17: NO	Inconclusive outcome
2 – 6 (*KII*)	1-2-3-5-6-7: YES	Restricted gene flow / Dispersal but with some long distance dispersal
3 – 1 (*KII*)	1-2-3-4: NO	Restricted gene flow with isolation by distance
3 – 2 (*KII*)	1-2-3-5-6-7: YES	Restricted gene flow / Dispersal but with some long distance dispersal

Discussion

The goal of this study was to reconstruct the phylogeographic pattern of *Busseola fusca* across the whole geographic range of the species, to evaluate the current geographic distribution of each clade and to determine their centres of origin.

Genetic structure and phylogeographic pattern of *Busseola fusca*

All phylogenetic analyses confirm the separation of *B. fusca* into three major clades corresponding to three geographical units: one localized in West African region (*W*), one restricted to East Africa (*KI*) and one found from Central to East Africa via Southern Africa (*KII*). Partial geographic overlap was observed only between the clades *KI* and *KII*. The genetic distances between the clades suggest that the differentiation occurred during the Pleistocene (Sezonlin *et al.* 2006). Major climatic changes occurred in sub-Saharan Africa during the Pleistocene. The period of climatic instability started 3.3 to 2.45 Ma, oscillating between hot/humid and cooler, drier periods (Wagner 2002). A shift to arid, open conditions occurred in near 2.8 Ma, 1.7 Ma, and 1.0 Ma (de Menocal 1995). De Menocal (1995) concluded that this alternation of cold, dry periods and warmer, wetter periods led to oscillations in savannah biotope expansion. Pleistocene events also played an important role in differentiation of African vertebrates (Quérouil *et al.* 2003). The present study confirmed that *B. fusca* populations are differentiated in three clades that were isolated during Pleistocene in three different refuges (Sezonlin *et al.* 2006). This hypothesis was also supported by genetic structure analyses that show high values of the fixation index between the three clades. The fragmentation hypothesis can be inferred between West African and other clades that do not overlap at all, but is questionable between the *KI* and *KII* clades that overlap in wide areas. According to Templeton *et al.* (1995), only mainly non-overlapping populations can clearly be inferred as a product of past fragmentation.

All the mitochondrial variation revealed significant and historical separation between clades *W*, *KI* and *KII*. The monophyly of the western populations allows the designation of at least one Evolutionary Significant Unit (ESU) (Alpers *et al.* 2004) within the *B. fusca* species. It is the same for *KI* and *KII* that can be considered as ESUs. If these clades are confirmed by nuclear polymorphism analyses (microsatellites

data), pheromones and behavioural ecology, therefore different biological strategies must be used to control *B. fusca* populations.

Origin and evolutionary history of each clade

In the present *B. fusca* study, the population of forest region has higher haplotype and nucleotide diversity for clade *W* than other phytogeographic regions of West Africa. It would likely be the centre of origin for this clade. Alpers *et al.* (2004) found also a center of origin in Ghana for the West African populations of the roan antelope (*Hippotragus equinus*) (Desmarest 1804), a herbivorous species that has a sub-Saharan distribution similar to that of *B. fusca*. Concerning the *KI* clade, the eritrean region, where only *KI* individuals were found was more diverse than other regions. Accordingly, we will consider it as the possible/putative centre of origin for clade *KI*. Finally for clade *KII*, the highest haplotype and nucleotide diversity was found among population of South-East region. The centre of origin of this clade is likely in this geographic area. Consequently, we can suggest that the centres of origin were likely localized in forest, East-North and East-South regions respectively for clades *W*, *KI* and *KII*. We observed that the strongest haplotype diversity was associated with the lowest levels of nucleotide diversity. This accumulation of haplotypes suggests that the clades experienced bottlenecks at their origins, followed by major population demographic expansion (Grant & Bowen 1998; Avise 2000). However, Petit *et al.* (2003) have shown that for some European tree species the highest diversity is not observed in the centre of origin but rather in secondary contact zones. Therefore in *B. fusca* case, further studies will be necessary to confirm the reality of these possible centres of origin. The fact that clades *KI* and *KII* originate from East Africa gives support to some previous studies (Livingstone 1982, Arctander *et al.* 1999). Indeed, these authors have described the East African region as a mosaic of secondary refuge zones for herbivorous mammals, with periodic exchanges between refuge zones through temporary contact bridges in the East African Rift Valley. Although the East African populations of *B. fusca* are now overlapping, the Rift Valley was pointed out as one of the main factors that explain most of the molecular variation in East Africa (Sezonlin *et al.* 2006). The Rift Valley seems to act as an important natural barrier to maintain population structure for other African species in East Africa (Arctander *et al.* 1999, Pitra *et al.* 2002).

The current distribution of *B. fusca* populations can be explained by contiguous range expansion or by dispersal with some long distance dispersal as it is highlighted by NCPA inferences. This pattern is probably linked to the expansion of its wild host plant (*Sorghum arundinaceum* (Desv. Stapf) (Poaceae) (Haile & Hofsvang 2002) during the Pleistocene. The presence of individuals of the large population unit, clade *KII* in Central Africa (Cameroon) is consistent with the hypothesis of a faunistic link between these two regions (Bruhl 1997), which are separated by a distance of 3000 km. An eastern origin of central populations is a possibility as it was suggested for a butterfly species in Cameroon (De Jong & Congdon 1993). De Jong & Congdon (1993) argued that the low animal species diversity in highland forests of Cameroon suggests that these species originated from long distance migration from East Africa. The faunistic link between Eastern and Central Africa also exists for some vertebrate species (Pitra *et al.* 2002). However, the nature of this faunistic link was not elucidated by our study and remains unknown for *B. fusca*. The present study shows that the clade *KII* is also present in Southern Africa. This geographic expansion toward Southern Africa is consistent with the patterns highlighted by the study of some African vertebrates (Faulkes *et al.* 2004). Climatic and topographic differences between major biogeographic African regions (de Menocal 1995) might explain the different processes that govern the current geographic distribution of each *B. fusca* clade.

Evolution of host plant specialization in *Busseola fusca*

Many ecological studies have shown that the introduction of exotic plants can lead to host plant shifts in oligo- or monophagous insect species. This is the case for *Rhagoletis pomonella* (Walsh 1867) (Diptera: Tephritidae) which in North America switched from hawthorn to introduced apple tree (Bush 1994); the bug *Jadera haematoloma* (Herrich-Schaeffer 1847) (Hemiptera: Rhopalidae), which added to its host plant spectrum an introduced ornamental plant of the family Sapindaceae (Carroll & Boyd 1992). Other examples of such host switches are the nymphalid *Euphydrias editha* (Boisduval 1852) (Lepidoptera: Nymphalidae) which switched to *Plantago lanceolata* L. (Plantaginaceae), introduced by North American breeders (Singer *et al.* 1993), the groundnut beetle *Caryedon serratus* (Olivier 1790) (Coleoptera: Bruchidae) which added groundnut, *Arachis hypogea* L. (Fabaceae), an introduced Papilionoideae to its native host plant range (Delobel 1995) and finally African cereal stem borers such as *B. fusca* and *Sesamia calamistis* Hampson 1910 (Lepidoptera, Noctuidae) which shifted to maize after its introduction into Africa.

Host plant shifts appear as a major factor promoting ecological specialization in phytophagous insects.

Within an insect population on a given host plant, some genotypes procuring better fitness to their bearers can be selected. This phenomenon may occur within the various populations of an insect species associated with different host plants and leads to genetic differentiation between these populations. A study by Via *et al.* (2000) on the aphid *Acyrthosiphon pisum* (Harris 1776) (Hemiptera: Aphididae) that feed on two different legume species shows that direct selection against migrants and hybrids in each parental environment could favour the evolution of more precise or efficient habitat choices. Moreover, these authors argued that the differential phenology of host plants strengthens ecological specialization by increasing the reproductive isolation between populations. This phenomenon promotes mating within the same habitat. As is well established for some phytophagous insects, such host plant shifts can lead to ecological segregation of populations with appearance of host races and genetic differentiation. This is well known with *R. pomonella* (Feder *et al.* 1988; McPheron *et al.* 1988), *R. cerasi* (L. 1758) (Diptera: Tephritidae) (Schwartz *et al.* 2003), the European corn borer *Ostrinia nubilalis* (Hübner 1796) (Lepidoptera: Pyralidae) (Thomas *et al.* 2003), *Spodoptera frugiperda* (Smith & Abbott 1797) (Lepidoptera: Noctuidae) (Pashley Prowell *et al.* 2004), *C. serratus* (Sembène 2000) and the pea aphid *A. pisum* (Via *et al.* 2000; Caillaud & Via 2000). Genetic differentiation of such host races can occur within a limited period of time. For example, it has taken a few centuries for *R. pomonella* and *C. serratus* populations to differentiate, several tens of years for *J. haematoloma* and less than 20 years for *E. editha*. Considering the age of sorghum domestication and of maize introduction in Africa, it is thus likely that these major agricultural events could have been at the origin of such a phenomenon of genetic divergence among *B. fusca* populations and possibly at the origin of host race appearance in this species. The fact that *B. fusca*, considered as oligophagous (Le Rü *et al.* 2006) now preferentially uses maize and cultivated sorghum in most part of its distribution areas suggests that these host plants offer this species very suitable resources. The preference for these cultivated plants (Kfir *et al.* 2002, Le Rü *et al.* 2006) would represent one of the factors that might have led *B. fusca* to start ecological specialization between populations. This ecological segregation between *B. fusca* populations exploiting wild and cultivated host plants or cultivated sorghum and maize could be highlighted by further molecular studies. However, a preliminary population genetic study using mitochondrial marker (Sezonlin *et al.* 2006) has not demonstrated any genetic structure between cultivated sorghum and maize *B. fusca* populations. An extensive survey carried out in East and West Africa indicates that *B. fusca* is now mostly associated with maize and sorghum crops and is generally uncommon in the wild habitat (Le Rü *et al.* 2006; Sezonlin *et al.* unpublished) Therefore, we have not been able until now to test ecological segregation between wild host plants and cereal crops, but the recent discovery of *B. fusca* populations associated to *Phragmites mauritianus* Kunth. (Poacaea) in Ethiopia and Eritrea and to *Setaria megaphylla* (Steud.) (Poaceae) T. Duran & Schinz in Kenya (Le Rü *et al.* 2006) will soon help us to test this hypothesis.

Acknowledgements. This study was supported by IRD UR072. We are grateful to Dr. G. Goergen, Collection Manager for IITA Biodiversity Centre - Station Cotonou in Benin, for his support. We would like to thank the members of the ICIPE Noctuid Stem Borer Biodiversity Project (NSBBP) mainly L. Ngala, B. Musyoka, J. Mungai and H. Njambi for technical assistance during insect collection and insect rearing in the laboratory in Kenya. Anonymous referee also helped greatly in improving this manuscript.

References

Ajayi O. 1998. Sorghum: West Africa, p. 39-45 *in*: **Polaszek A. (ed.)**, *African cereal stem borers: economic importance, taxonomy, natural enemies and control.* CABI, Wallingford, UK.

Alpers D.L., Van Vuuren B.J., Arctander P., Robinson T.J. 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology* **13**: 1771-1784.

Althoff M.D., Pellmyr O. 2002. Examining genetic structure in bogus yucca moth: A sequential approach to phylogeography. *Evolution* **56**: 1632-1643.

Arctander P., Johansen C., Coutellec-Vreto M. 1999. Phylogeography of three closely related African bovids (Tribe *Alcelaphini*). *Molecular Biology and Evolution* **16**: 1724-1739.

Avise J.C. 2000. *Phylogeography. The History and Formation of Species.* Harvard University Press, Cambridge, Massachusetts, 447 p.

Bernatchez L. 2001. The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution* **55**: 351-379.

Bruhl C.A. 1997. Flightless insects: a test case for historical relationships of African mountains. *Journal of Biogeography* **24**: 233-250.

Bush G.L. 1994. Sympatric speciation in animals: new wine in old bottles. *Trends in Ecology and Evolution* **9**: 285-288.

Caillaud C.M., Via S. 2000. Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *The American Naturalist* **156**: 607-621.

Carroll S.P., Boyd C. 1992. Host race radiation in the soapberry bug: Natural history with the history. *Evolution* **46**: 1052-1069.

Chastanet M. 1998. Introduction et place du maïs au Sahel Occidental (Sénégal-Mauritanie), p. 251-275 *in*: **Chastanet M. (ed.)**, *Plantes et paysages d'Afrique. Une histoire à explorer.* Karthala-CRA, Paris.

Clement M., Posada D., Crandall K.A. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657-1660.

De Jong R., Congdon T.C.E. 1993. The montane butterflies of the eastern Afrotropics, p. 133-164 *in*: **Lovett J. C., Wasser S. K. (eds.)**, *Biogeography and ecology of the rain forests of eastern Africa*. Cambridge University Press, Cambridge.

Delobel A. 1995. The shift of *Caryedon serratus* (Ol.) from wild Caesalpiniaceae to groundnuts took place in West Africa (Coleoptera: Bruchidae). *Journal of Stored Products Research* **31**: 101-102.

Doggett H. 1988. *Sorghum*. 2nd ed, Longman Scientific and Technical Publishers, London, 512 p.

Excoffier L., Smouse P.E., Quattro J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Faulkes C.G., Verheyen E., Verheyen W., Jarvis J.U.M., Bennett N.C. 2004. Phylogeograpical patterns of genetic divergence and speciation in African mole-rats (Family: Bathyergidae). *Molecular Ecology* **13**: 613-629.

Feder J.L., Chilcote C.A., Bush G.L. 1988. Genetic differentiation between sympatric host races of the apple maggot fly, *Rhagoletis pomonella*. *Nature* **336**: 61-64.

Grant S.W., Bowen B.W. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity* **89**: 415-426.

Gu X., Fu Y.X., Li W.H. 1995. Maximum likelihood estimation of heterogeneity of substitution rate among nucleotide sites. *Molecular Biology and Evolution* **12**: 546-557.

Haile A., Hofsvang T. 2001. Survey of lepidopterous stem borers of sorghum, maize and pearl millet in Eritrea. *Crop Protection* **20**: 151-157.

Haile A., Hofsvang T. 2002. Host plant preference of the stem borer *Busseola fusca* (Fuller) (Lepidoptera: Noctuidae). *Crop Protection* **21**: 227-233.

Hasegawa M., Kishino H., Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **21**: 160-174.

Holloway J. D. 1998. Noctuidae, p. 79-86 *in*: **Polaszek A. (ed)**, *African cereal stem borers: economic importance, taxonomy, natural enemies and control*. CABI, Wallingford, UK.

Kfir R., Overholt A.W., Khan R.Z., Polaszek A. 2002. Biology and management of economically important lepidopteran cereal stem borers in Africa. *Annual Review of Entomology* **47**: 701-731.

Kimura M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences of the USA* **78**: 454-458.

Le Rü B.P., Ong'amo G.O., Moyal P., Muchugu E., Ngala L., Musyoka B., Abdullah Z., Kauma-Matama T., Lada V.Y., Pallangyo B., Omwega C.O., Schulthess F., Calatayud P.-A., Silvain J.-F. 2006. Geographic distribution and host plant ranges of East African noctuid stem borers. *Annales de la Société Entomologique de France* **42**(3-4): 353-361.

Livingstone D.A. 1982. Quaternary geography of Africa and refuge theory, p. 523-536 *in*: **Prance G. T. (ed.)**, *Biological diversification in the tropics: Proceedings of the 5th international symposium of the association for tropical biology*. Columbia University Press, New York.

Maddison D.R., Maddison W.P. 2002. *MacClade: Analysis of phylogeny and character evolution*. Version 4.02. Sinauer Associates, Sunderland, Massachusetts.

Madeiras Santos M.E., Ferraz Torrao M.M. 1998. Entre l'Amérique et l'Afrique, les îles du Cap-vert et Sao Tomé: les cheminements des milhos (mil, sorgho et maïs), p. 69-83 *in*: **Chastanet M. (ed.)**, *Plantes et paysages d'Afrique. Une histoire à explorer*. Karthala-CRA. Paris.

McPheron B.A., Smith C.D., Berlocher S.H. 1988. Genetic differences between host races of *Rhagoletis pomonella*. *Nature* **336**: 64-66.

Menocal P.B. de 1995. Plio-Pleistocene African climate. *Science* **270**: 53-59.

Moyal P. 1998. Infestation patterns and parasitism of the maize stalk borer, *Busseola fusca* (Fuller) (Lepidoptera: Noctuidae) in the Ivory Coast. *African Entomology* **6**: 289-296.

Murty D.S., Renard C. 2001. Sorgho (*Sorghum bicolor* (L.) Moench), p. 105-125 *in*: **Raemaekers R. H. (ed.)**, *Agriculture en Afrique Tropicale*. Direction Générale de la Coopération Internationale, Bruxelles, Belgique.

Ndemah R., Schulthess F., Korie S., Borgemeister C., Cardwell K.F. 2001. Distribution, relative importance and effect of lepidopterous borers on maize yields in forest zone and mid-altitude of Cameroon. *Journal of Economic Entomology* **94**: 1434-1444.

Nei M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York, 512 p.

Nei M., Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the USA* **76**: 5269-5273.

Pashley Prowell D., McMichael M., Silvain J.-F. 2004. Multilocus genetic analysis of host use, introgression, and speciation in host strains fall armyworm (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America* **97**: 1034-1044.

Petit R.J, Aguinagalde I., de Beaulieu J-L., Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., Mohanty A., Müller-Starck G., Demesure-Musch B., Palmé A., Martín J.P., Rendell S., Vendramin G.G. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563-1565.

Pitra C., Hansen J.A., Lieckfeldt D., Arctander P. 2002. An exceptional case of historical outbreeding in African sable antelope populations. *Molecular Ecology* **11**: 1197-1208.

Polaszek A., Khan Z.R. 1998. Host plants, p. 3-10 *in*: **Polaszek A. (ed.)**, *African cereal stem borers: economic importance, taxonomy, natural enemies and control*. CABI, Wallingford, UK.

Posada D., Crandall K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Posada D., Crandall K.A., Templeton A.R. 2000. GeoDis: A program for the cladistic nested analysis of geographical distribution of genetic haplotypes. *Molecular Ecology* **9**: 487-488.

Quérouil S., Verheyen E., Dillen M., Colyn M. 2003. Patterns of diversification in two African forest shrews: *Sylvisorex johnstoni* and *Sylvisorex ollula* (Soricidae, Insectivora) in relation to paleo-environmental changes. *Molecular Phylogenetics and Evolution* **28**: 24-37.

Schneider S.D., Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise distances when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**: 1079-1089.

Schneider S.D., Roessli D., Excoffier L. 2000. *Arlequin V. 2.000: a software for genetic data analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Schwartz D., McPheron A.B., Hartl B.G., Foller F.E., Hoffmeister S.T. 2003. A second case of genetic host races in *Rhagoletis*? A population genetic comparison of sympatric host populations in the European cherry fruit fly, *Rhagoletis cerasi*. *Entomologia Experimentalis et Applicata* **108**: 11-17.

Sembène M. 2000. *Variabilité de l'Espaceur Interne transcrit (ITS1) de l'ADN ribosomique et polymorphisme des locus microsatellites chez la bruche d'arachide Caryedon serratus (Olivier) : différenciation en races d'hôtes et infestation de l'arachide au Sénégal*. Thèse de doctorat d'Etat ès sciences, Université Cheik Anta Diop de Dakar, Dakar, 214p.

Sezonlin M., Dupas S., Le Rü B., Le Gall P., Moyal P., Calatayud P.-A., Giffard I., Faure N., Silvain J.-F. 2006. Phylogeography and population genetics of the maize stalk borer *Busseola fusca* (Lepidoptera, Noctuidae) in sub-Saharan Africa. *Molecular Ecology* **15**: 407-420.

Simmons R.B., Weller S.J. 2001. Utility and evolution of cytochrome *b* in insects. *Molecular Phylogenetics and Evolution* **20**: 196-210.

Singer M.C., Thomas C.D., Parmesan C. 1993. Rapid human-induced evolution of insect-host associations. *Nature* **366**: 681-683.

Slatkin M., Hudson R.R. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555-562.

Swofford D.L. 2002. PAUP*. *Phylogenetic Analysis Using Parsimony (*and other methods)*. Sinauer Associates, Sunderland, Massachusetts.

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

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

Templeton A.R., Routman E., Phillips C.A. 1995. Separating population structure from population history: a cladistic analysis of the geographical dispersion of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140**: 767-782.

Templeton A.R. 1998. Nested clade analysis of phylogeographic data: testing hypothesis about gene flow and population history. *Molecular Ecology* **7**: 381-397.

Templeton A.R. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**: 789-809.

Thomas Y., Bethenod M.T., Pelozuelo L., Frérot B., Bourguet D. 2003. Genetic isolation between two sympatric host-plant races of the European corn borer, *Ostrinia nubilalis* Hübner. I. Sex pheromone, moth emergence timing, and parasitism. *Evolution* **57**: 261-273.

Via S., Bouk C.A., Skillman S. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* **54**: 1626-1637.

Wagner T. 2002. Late Cretaceous to early Quaternary organic sedimentation in the eastern Equatorial Atlantic. *Palaeogeography, Palaeoclimatology, Palaeoecology* **179**: 113-147.

White F. 1983. The vegetation of Africa, a descriptive memoir to accompany the UNESCO / AETFAT / UNSO vegetation map of Africa. *UNESCO, Natural Resources Research* **20**: 1-356.

Yang Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* **10**: 1396-1402.