Molecular Ecology (2008) **17**, 3640–3653 doi: 10.1111/j.1365-294X.2008.03869.x

Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles

M-P. CHAPUIS,*†‡§ M. LECOQ,‡ Y. MICHALAKIS,† A. LOISEAU,* G. A. SWORD,§ S. PIRY* and A. ESTOUP*

**Unité Mixte de Recherche Centre de Biologie et Gestion des Populations (INRA/IRD/Cirad/Montpellier SupAgro), Institut National de la Recherche Agronomique, Campus international de Baillarguet, CS 30016, F-34988 Montferrier-sur-Lez cedex, France, †Génétique et Evolution des Maladies Infectieuses Unité Mixte de Recherche 2724 CNRS-IRD, Institut de la Recherche et du Développement, 911 avenue Agropolis, 34394 Montpellier Cedex 5, France, ‡Unité Propre de Recherche Acridologie, Centre de coopération internationale en recherche agronomique pour le développement, 34398 Montpellier Cedex 5, France, §School of Biological Sciences, University of Sydney, The Macleay Building A12, Sydney, NSW 2006, Australia*

Abstract

An understanding of the role of factors intrinsic to a species' life history in structuring contemporary genetic variation is a fundamental, but understudied, aspect of evolutionary biology. Here, we assessed the influence of the propensity to outbreak in shaping worldwide genetic variation in *Locusta migratoria*, a cosmopolitan pest well known for its expression of density-dependent phase polyphenism. We scored 14 microsatellites in nine subspecies from 25 populations distributed over most of the species' range in regions that vary in the historical frequency and extent of their outbreaks. We rejected the hypothesis that *L. migratoria* consists of two genetically distinct clusters adapted to habitats either rarely (nonoutbreaking) or cyclically (outbreaking) favourable to increases in population density. We also invalidated the current subspecific taxonomic classification based on morphometrics. Bayesian inferences indicated evidence of a homogenizing effect of outbreaks on *L. migratoria* population structure. Geographical and ecological barriers to gene flow in conjunction with historical events can also explain the observed patterns. By systematically assessing the effects of null alleles using computer simulations, we also provide a template for the analysis of microsatellite data sets characterized by a high prevalence of null alleles.

Keywords: computer simulation, life-history evolution, locust, phase polyphenism, spatial dynamics, subspecies

Received 3 April 2008; revision received 8 June 2008; accepted 18 June 2008

Introduction

Consensus has emerged on the prevailing roles of geological processes and recent human perturbations in driving the range-wide genetic structure of most extant taxa (Avise 2000). However, the implications of other ecological factors that may influence range-wide patterns of genetic variation, such as those intrinsic to a species' life history, remain poorly understood (Costello *et al*. 2003). Assessing how ecological differences among contemporary populations affect the geographical structuring of intraspecific genetic variation has important implications for managing pests, conserving threatened populations, and predicting population responses to climate change (Mousseau 2000; Costello *et al*. 2003). Tackling this question requires phylogeographical analyses of species for which a detailed knowledge exists of the most important intrinsic features that might affect its evolution. To our knowledge, such range-wide genetic surveys are rare, particularly in terrestrial animal species (but see Heuertz *et al*. 2006 for a plant example). This is because (i) sampling sufficient portions of the range is difficult, and (ii) most terrestrial taxa with a wide range are either crop or forest pests (e.g. bruchids; black rat), domestic species (e.g. honeybees; pigs), or human commensals (e.g. houseflies;

Correspondence: Marie-Pierre Chapuis, Tel. +61 (02) 9351 7503; Fax: +61 (02) 9351 4119; E-mail: chapuimp@supagro.inra.fr

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

house mice) that have expanded their ranges primarily due to human activities (e.g. Larson *et al*. 2005).

In contrast, the migratory locust, *Locusta migratoria*, is naturally distributed over all grasslands of the Old World, from Northern Eurasia to New Zealand. This major agricultural pest displays irregular outbreaks across most of its range (Uvarov 1966; COPR 1982). Importantly, *L. migratoria* also expresses phase polyphenism, a form of phenotypic plasticity in which different rearing densities produce individuals that vary in a variety of traits including morphology, colour, physiology and migratory behaviour (Uvarov 1977; Pener & Yerushalmi 1998; Simpson & Sword 2008). Low densities result in relatively sedentary solitarious-phase individuals, whereas high population densities generate mass- migrating gregarious-phase locusts. This process of density-dependent phase change is referred to as gregarization. Based on morphometrics, taxonomists have described 11 geographical subspecies of *L. migratoria* across its range (reviewed in Remaudière 1940, 1947; COPR 1982). This taxonomy of *L. migratoria* is far from being considered definitive by some authors (e.g. Farrow & Colless 1980). The taxonomists used a morphometric technique originally devised to differentiate between the plastic gregarious and solitarious locust phases (Song 2004). The environmental effects of local population density on the phenotype may have confounded the use of morphometrics in *L. migratoria* taxonomy. To date, the validity of *L. migratoria* subspecies taxonomy has not been tested using genetic approaches.

Locusta migratoria may consist of genetically distinct clusters with different propensities to outbreak. This view has been suggested for another migratory pest insect, the flightless Mormon cricket, *Anabrus simplex* (Bailey *et al*. 2005). However, the presence of a strong geographical barrier to gene flow (i.e. the Rocky Mountains) between eastern nonoutbreaking and western outbreaking populations may well have a confounding effect on the factors shaping genetic variation in this species. The plausibility of genetic structuring based on the propensity to outbreak necessarily assumes that outbreak events are not strictly determined by environmental factors, but also at least in part by the expression of traits that are adaptive and under genetic control. In *L. migratoria*, Chapuis *et al*. (2008) recently provided potential support for such genetic variation by demonstrating that locusts from a historically outbreaking Malagasy population express density-dependent morphometrical and behavioural gregarization to a greater degree than do those from a historically nonoutbreaking French population. Such a population genetic divergence in phenotypic plasticity may be interpreted as a differential adaptation to habitats rarely (i.e. nonoutbreaking) vs. periodically (i.e. outbreaking) favourable to increases in population density (Dopman *et al*. 2002; Sword 2002). Such adaptive processes result from differentiation at selected loci, and may lead to differentiation of their surrounding regions due to linkage disequilibrium. In *L. migratoria,* the propensity to gregarize is a complex composite character with an elaborate underlying molecular basis involving the differential expression of more than 500 genes between the isolated solitarious and crowded gregarious phases (Kang *et al*. 2004). Therefore, many loci throughout the genome, including randomly chosen microsatellite loci, might exhibit differences in allele frequencies among populations due to contrasting selective regimes imposed by variation in outbreak frequency.

The association between neutral and adaptive divergences may be undermined, however, if adaptive traits of interest act on gene flow at the population level (e.g. Whiteley *et al*. 2004). Because *L. migratoria* outbreaking populations experience recurrent demographic flushes and mass movement, one may expect a lower neutral divergence throughout the genome in historically outbreaking areas relative to nonoutbreaking areas.

Here we assess how the propensity to outbreak has shaped genetic variation in the cosmopolitan pest, *L. migratoria*. To address this question, we genotyped 14 microsatellite loci in locusts from 25 populations with different propensities to outbreak, representing nine subspecies from across the species' range. Microsatellite loci are widely used to assess population genetic structure and to study evolutionary relationships below the species level (e.g. Estoup & Angers 1998). However, orthopteran species seem particularly affected by a high prevalence of microsatellite null alleles (Chapuis *et al*. 2005, in press). We therefore used computer simulations to evaluate the effect of null alleles on our inferences from microsatellite data.

Materials and methods

Population sampling and genotyping

From 2001 to 2004, we collected a total of 25 population samples (19–32 individuals per sample) covering much of the Old World distribution of *Locusta migratoria* (Fig. 1). The sampling included populations from nine subspecies referred to as *L. m. capito*, *L. m. cinerascens*, *L. m. manilensis*, *L. m. migratoria*, *L. m. migratorioides*, *L. m. gallica*, the Australian subspecies, the Arabian subspecies, and the Palavas form named after a small town in Southern France (Remaudière 1940, 1947; COPR 1982). The Tibetan and Indian subspecies were not sampled. We defined two contrasting historical patterns of outbreak events based on a literature survey (detailed in Fig. 1). Populations from some areas have experienced frequent and widespread outbreak events while others came from areas where outbreaks have been infrequent and local.

Fourteen microsatellite loci (OZC9, OZC35, OZC76 in Zhang *et al*. 2003, and LM1-88, LM10-78, LM2-B, LM2-Λ, LM3-Ω, LMT-113, LMT-137, LMT-177, LM10-180, LM11-121,

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

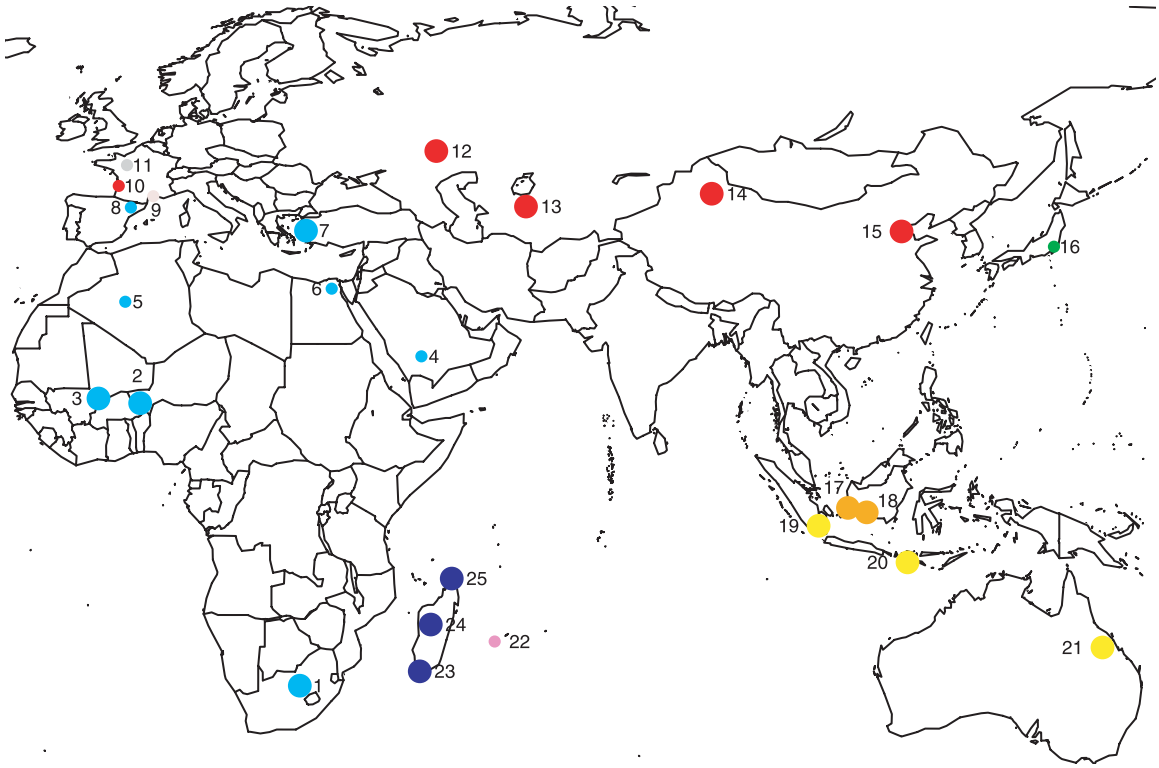

Fig. 1 Species range of *Locusta migratoria* with locations, taxonomy, outbreak patterns, and genetic clustering of sampled populations. Recorded outbreak events are depicted as either uncommon and local (small circle) or frequent and widespread (large circle). The outbreak patterns of sampled populations indicated by numbers were based on the following sources: [1] Brown (1986), Farrow (1987), and D. Brown (Plant protection Research Institute, South Africa, personal communication); [2,3] Betts (1961); [4] Farrow & Colles 1980; [5] Uvarov & Hamilton (1936) and Benfekih *et al*. 2002; [6] M. Lecoq, personal communication; [7] Karabag (1958); [8–13] Waloff (1940); [14] Chen (1991); [15] Chen (1991) and Zhang & Li (1999); [16] Hakomori & Tanaka (1992); [17–20] Lecoq & Surkino (1999); [21] Farrow (1979); [22] Wintrebert (1970); [23–25] Randriamanantsoa (1998). Sampling sites with similar colours belong to the same genetic cluster as assessed by the Bayesian method of Corander *et al*. (2003). Subspecies taxonomy as follows: *Au*, Australian subspecies; *Ar*, Arabian subspecies; *Ca*, *L. m. capito*; *Ci*, *L. m. cinerascens*; *Ga*, *L. m. gallica*; *Ma*, *L. m. manilensis*; *Md*, *L. m. migratorioides*; *Mi*, *L. m. migratoria*; *Pa*, Palavas form.

LMT-133 in Chapuis *et al*. 2005) were genotyped, using fluorescently labelled polymerase chain reaction (PCR) primers and a MegaBACE sequencing machine (Amersham Biosciences), as described in Chapuis *et al*. (2005). OZC9 was amplified within the multiplex set 1 of Chapuis *et al*. (2005) with a primer concentration of 0.5 μM, OZC35 within the multiplex set 2 with a primer concentration of 0.3 μM, and OZC76 within the multiplex set 2 with a primer concentration of 0.2 μM. To minimize genotyping errors, a minimum of two replicate PCRs per sample per locus were conducted and alleles were scored independently by two of the authors using GENETIC PROFILER version 1.0 software (Amersham Biosciences). Alleles included in the final consensus genotypes were observed at least twice; if observed only once, an additional PCR replicate was conducted. One negative and eight positive controls (samples with known genotypes) were included in each run of 96 PCRs to check for potential contamination and standardize genotyping across experiments.

Genetic variation within and among populations

For each *L. migratoria* population, we estimated over all loci the mean expected values of five statistics traditionally used for summarizing within-population genetic variation: the allelic diversity (A), allelic diversity corrected for subsamples of 20 genes (R_S; El Mousadik & Petit 1996), expected heterozygosity (H_E; Nei 1987), observed heterozygosity (H_O), and allele size variance in base pairs (V).

Genotypic differentiation between populations was tested using Fisher's exact tests (GENEPOP 3.3; Raymond & Rousset 1995). Corrections for multiple tests were performed using the false discovery rate approach (Benjamin & Hochberg 1995). The level of differentiation between populations was quantified by computing pairwise estimators of F_{ST} (Weir 1996). Following Chapuis & Estoup (2007), we used the *ENA* (excluding null alleles) correction method to efficiently correct for the positive bias induced by the presence of null alleles on F_{ST} estimation. $F_{ST}^{\{ENA\}}$ values were

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

computed using the program FREENA (http://www.montpellier.inra.fr/URLB/).

Identification of L. migratoria *genetic clusters*

We applied a Bayesian analysis implemented in the program BAPS 3.1 for estimating hidden substructure across the studied species' range of *L. migratoria* (Corander *et al*. 2003). The method determines clusters of population samples by minimizing Hardy–Weinberg and linkage disequilibrium within the clusters. The number of clusters is treated as an unknown parameter. This method has been shown to be very conservative in identifying population structure (Waples & Gaggiotti 2006). Hence, it determines the uppermost level of population structure, grouping together populations that frequently exchange migrants. As explained in the null alleles section below, we tested the effects of the occurrence of null alleles on this method.

A neighbour-joining (NJ) tree (Saitou & Nei 1987) relating the populations was constructed based on microsatellite data using the chord distance (Cavalli-Sforza & Edwards 1967; D_C) with 2000 bootstrap replications to assess the stability of the nodes (Hedges 1992). The trees were calculated with the POPULATION software package (Olivier Langella, CNRS UPR9034, France) and graphically displayed with TREEVIEW software (Page 1996). The genetic distance of Cavalli-Sforza & Edwards (1967) was chosen because (i) it seems to be the most efficient distance for obtaining a correct tree topology (Takezaki & Nei 1996); (ii) it makes no assumption regarding constant population size or mutation rates among loci; and (iii) the null allele bias for this genetic distance is low and similar for a large range of splitting times (Chapuis & Estoup 2007).

Testing outbreaks as a factor driving genetic structure

We used a binary explanatory factor, *Outbreaking*, taking values of 0 for populations having frequently experienced widespread outbreak events and 1 for populations coming from areas where outbreaks have been infrequent and local. Because *L. migratoria* genetic variation seemed to be influenced by major physical barriers to migration such as water masses (see Results section), we also considered an additional factor, *Insularity*. This factor takes values of 1 for insular populations (those in Japan, Indonesia, Australia, Madagascar, and the Reunion islands) and 0 for continental populations (those in mainland Europe, Africa, and Asia).

We first tested whether the propensity to outbreak affects *L. migratoria* population structure through reduced success of migrant genotypes between contrasted outbreaking vs. nonoutbreaking habitats. Under this hypothesis, the genetic divergence will be higher between the nonoutbreaking and outbreaking populations than within them, except in the presence of counteracting homogenizing effects. We used Mantel tests implemented in the software CADM (Legendre & Lapointe 2004) to test for the correlation between matrices of Euclidean distances for the *Outbreaking* factor and matrices of multilocus and single-locus D_C genetic distances. Corrections for multiple tests were performed using the false discovery rate approach (Benjamin & Hochberg 1995).

We then tested for a homogenizing effect of *L. migratoria* population structure due to outbreak dynamics. Under this hypothesis, the genetic differentiation is expected to be lower among outbreaking populations than among non-outbreaking populations. Such a lower genetic differentiation among outbreaking populations, however, can be also accounted for by larger effective population sizes (less genetic drift). Therefore, we analysed the effects of the propensity to outbreak on both the level of within-population variation and the level of genetic differentiation among populations, using general linear models (GLM). First, we performed a standard GLM analysis of the expected heterozygosity (H_E) with *Outbreaking* and *Insularity* as fixed factors, assuming a logit-link function and a binomial error using GENSTAT 8.1 (McConway *et al*. 1999). Second, we used a new hierarchical Bayesian method that estimates F_{ST} values as migration–drift factors for each local population and relates them to the *Outbreaking* and *Insularity* factors using a GLM-based approach, assuming nine models differing by the number of factors involved (Foll & Gaggiotti 2006). The posterior probabilities for each model relative to the others were estimated using a reversible jump Markov chain Monte Carlo (MCMC) method. For the latter treatment, we generated three independent estimation replicates using the program GESTE (http://www-leca.ujf-grenoble.fr/logiciels.htm). For each replicate, we used a sample size of 20 000 iterations within the MCMC run including a burn-in of 5000 iterations and a thinning interval of 20 iterations (total length of the chain = 500 000 iterations). Proposal distributions were adjusted by running 10 pilot runs of 200 iterations each. We reported the mode estimates of the regression coefficients and of σ^2, a measure of model fit, as well as their 95% highest probability density intervals (HPDI), obtained under the model with the highest probability and averaged over the three replicates.

Null alleles

Deviation from Hardy–Weinberg equilibrium (HWE) was tested for each locus separately and within each population using a Fisher's exact test implemented in the software GENEPOP 3.3 (Raymond & Rousset 1995). Corrections for multiple tests were performed using the false discovery rate approach (Benjamin & Hochberg 1995). When observed genotype frequencies deviated significantly from HWE, we used the program MICROCHECKER (van Oosterhout *et al*. 2004) to determine the most probable cause among various genotyping errors (e.g. short allele dominance, Wattier

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

et al. 1998) and the presence of null alleles. A maximum-likelihood estimate of the frequency of null alleles (Dempster *et al*. 1977) was then calculated for each locus and population using the program FREENA (Chapuis & Estoup 2007).

Because microsatellite null alleles were prevalent in *L. migratoria* (see Results section), we used computer simulations based on the coalescent (Hudson 1990; Leblois *et al*. 2003) to evaluate the effect of null alleles on our inferences and to assess the efficiency of a traditional correction method, the so-called INA (including null alleles) method. This method adjusts the genotype frequencies for each locus and population by (i) estimating null and visible allele frequencies following Dempster *et al*. (1977), and (ii) assuming a single new null allele size common to all genotyped populations (for details see the section 'Simulation method' in Chapuis & Estoup (2007)). This evaluation focused on (i) the statistics traditionally used to estimate within population genetic variation (A, R_S, H_E, H_O, and V), (ii) the Bayesian analysis implemented in the program BAPS, and (iii) the tree topologies obtained using the NJ algorithm and the D_C genetic distance. The effect of null alleles on F_{ST} estimation has already been studied by Chapuis & Estoup (2007). To our knowledge, the impact of the presence of null alleles in genotype data sets on the Bayesian GESTE approach has never been investigated. Unfortunately, this approach is computationally intensive and hence difficult to test using a reasonably large number of simulated data sets.

In our simulation survey of statistics summarizing within-population genetic variation (A, R_S, H_E, H_O, and V), we assumed a single isolated population at mutation–drift equilibrium of effective size N_e. Inference from BAPS was then tested assuming a population model of two populations of equal effective size N_e exchanging migrants at a rate m. Finally, the effect of null alleles on NJ tree topologies was evaluated by assuming a population split model without migration with five populations of equal effective size N_e (see Fig. S1, Supplementary material for details). For all the above population models, a triplet of genotype data sets (14 independent microsatellite loci and 60 genes per population, i.e. the sample values in our real *L. migratoria* data set) were simultaneously simulated. In the first, all alleles were assumed to be visible (i.e. absence of null alleles). The two other data sets in the triplet were based on the same individuals, but included null alleles, one being uncorrected and the other corrected using the INA method (Chapuis & Estoup 2007). For each parameter set, we simulated 100 or 1000 data set triplets depending on the studied method or statistics.

Mutations in the repeat region and the primer sites of microsatellite loci, leading to the presence of null alleles were simulated as described in Chapuis & Estoup (2007). The parameter $N_e\mu_R$ (with μ_R being the mutation rate in the repeat region) was set to 5, as this value generates mean heterozygosity values similar to those observed on average for *L. migratoria* microsatellite markers (i.e. 0.84 ± 0.01). This parameter was also set to 1 in order to test if our results hold for lower levels of within population genetic variation. The tested values of the parameter $N_e\mu_B$ (with μ_B as the mutation rate in the binding sites of the two primers) varied from 0.05 to 0.50 in order to explore a large range of null allele frequencies.

We assessed the performance of the NJ algorithm and the D_C genetic distance by measuring the degree of distortion of the topology of the reconstructed tree from the model tree (Fig. S1). The topology of a reconstructed tree was considered correct when the Robinson & Foulds's (1981) distortion index took a value of zero. Because most evolutionary biologists are interested in making a rooted tree, we measured topological errors by arbitrarily rooting trees on population 1 (see Fig. S1). The topological errors for unrooted trees are equal to or smaller than those for rooted trees, because in rooted trees, an additional error may be generated in the process of placing the root (Nei *et al*. 1983).

Results

Null allele prevalence in Locusta migratoria

Each sampled locust population showed significant heterozygote deficiencies in five to 13 single-locus exact tests. Samples that failed to amplify at some loci did not yield PCR products for these loci after two or three repeated PCRs. However, other loci were successfully amplified from these same DNA samples. This pattern strongly suggested the presence of null alleles for most loci and populations. In agreement with this, MICROCHECKER showed that the general excess of homozygotes was distributed across most allele size classes. Estimated frequencies of null alleles per locus per population ranged from 0 to 0.753, with frequencies averaged over loci varying from 0.077 to 0.250 depending on the population (see Table S1, Supplementary material for more details). The mean allele frequency over all populations and loci was 0.189.

Assessing the effects of microsatellite null alleles

Microsatellite data sets obtained using computer simulations showed that the presence of null alleles decreases estimates of all statistics summarizing genetic diversity within populations (Fig. 2). However, the effect of null alleles on the expected heterozygosity (H_E) was weak for a large range of null allele frequencies. For a mean null allele frequency of = 0.19 over all loci and populations (i.e. the value estimated in our real data set), H_E values were only 5% lower than that for = 0. Conversely, A, R_S, and H_O distributions for = 0.19 were almost separated from those for = 0. This was not the case for V due to the large variance

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

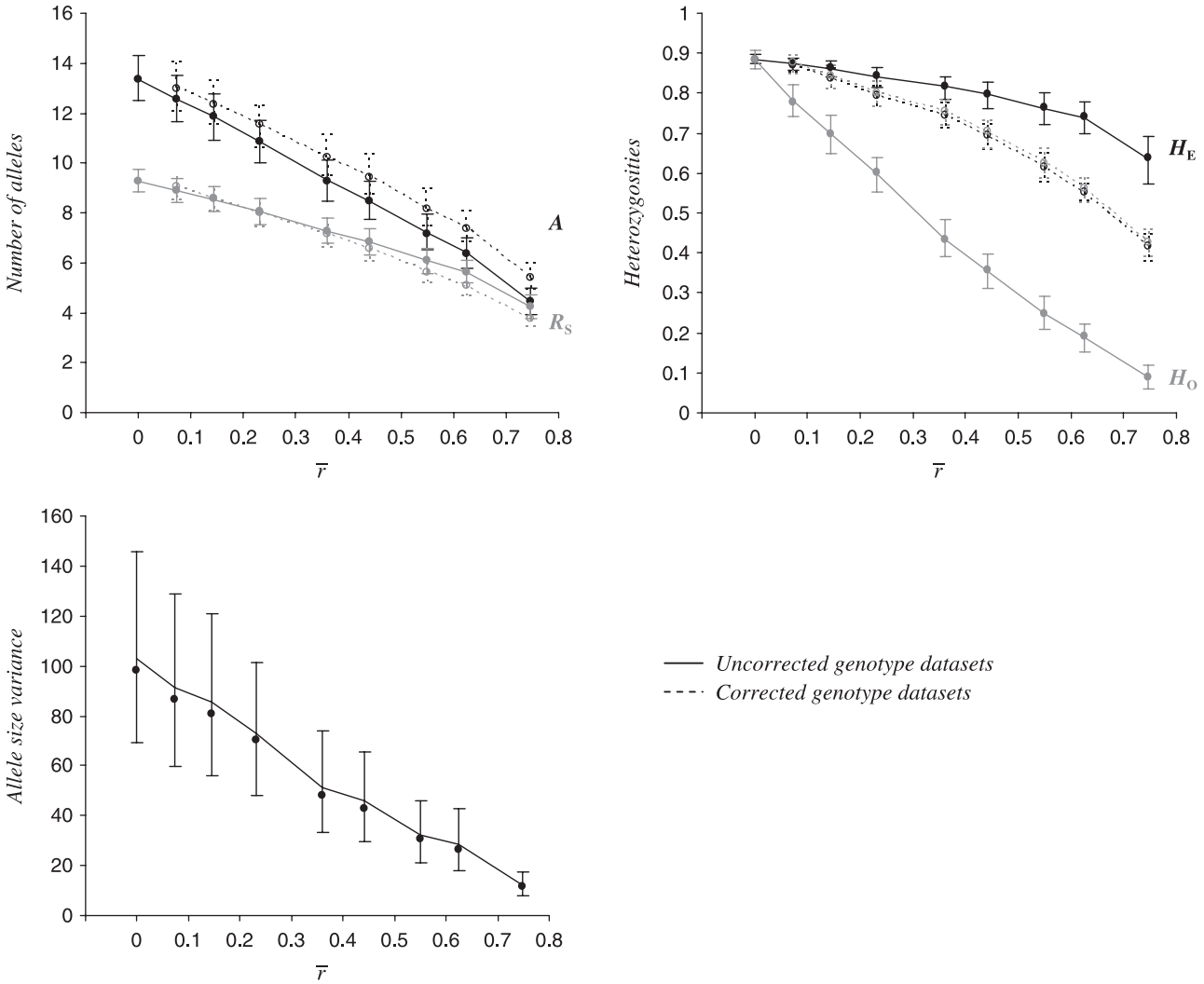

Fig. 2 Effects of null alleles and efficiency of the INA correction method on statistics describing within-population genetic variation. Over 1000 replicates, the simulated mean null allele frequencies were grouped into classes of 0.1 units. For each class, allelic diversity (A), allelic diversity corrected for subsamples of 20 genes (R_S), expected heterozygosity (H_E; Nei 1987), observed heterozygosity (H_O), and allele size variance in base pairs (V) were computed and averaged for both uncorrected and INA-corrected data sets (dotted lines). V was not computed for corrected data sets because the size of the null allele after correction is unknown and arbitrarily fixed (see Chapuis & Estoup 2007). Only genotype data sets with at least 20 genes with visible states were analysed (a number close to the smallest number of genes successfully genotyped in our samples). $N_e\mu_R$ was set to 5 (see text for details).

of this statistic, but mean values decreased by 29% between = 0 and = 0.19. The traditionally used INA correction procedure for null alleles partially rectified the bias induced by null alleles in A and H_O, but continued to generate underestimated values. This correction procedure increased the negative bias induced by null alleles in H_E and R_S with estimates reaching values lower than those obtained with uncorrected data sets.

With regards to the BAPS clustering method of Corander *et al*. (2003), our simulation results show that intermediate frequencies of null alleles (i.e. = 0.19) increased the power of BAPS to discriminate gene pools especially for high levels of gene flow (Fig. 3). The INA correction method partly decreased the former upper power of BAPS in data sets harbouring null alleles. For instance, at $N_e\mu_R = 5$ and $N_em = 2$ in a data set that was composed of two populations and originally contained 19% null alleles, BAPS found two different populations in 18%, 55%, and 83% of 100 simulated files in which null alleles were removed, corrected, or uncorrected, respectively. However, the power of BAPS was low at high levels of gene flow even for data sets with null alleles (e.g. two populations found in only 10% of simulated files without null allele correction at $N_em = 4$). This is in agreement with the results of Waples & Gaggiotti (2006),

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

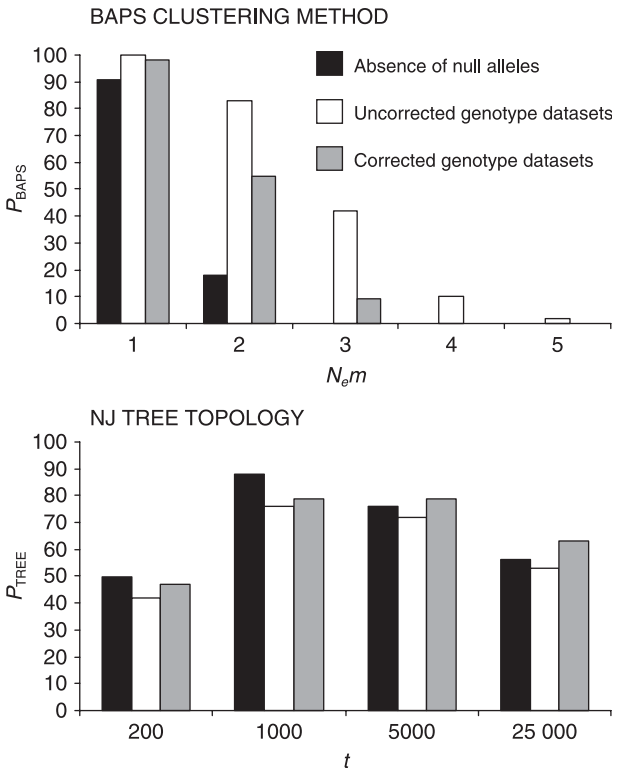

Fig. 3 Effects of null alleles and efficiency of the INA correction method on the BAPS test and NJ clustering method. One hundred simulated triplet genotype data sets (each triplet included one file without null alleles, and both uncorrected and corrected files with null alleles) were generated for each value of the effective number of migrants per generation (N_em) for the BAPS test and of the divergence time since the ancestral split event (t; Fig. S1) for the NJ clustering method. For simulated data sets with null alleles, the mean null allele frequency = 0.19 ± 0.01 (i.e. a value close to the mean value over all loci and populations of our real data set). P_{BAPS}, percentage of replicates in which BAPS inferred the correct number of two populations; P_{TREE}, percentage of replicates in which the correct topology was obtained. $N_e\mu_R$ was set to 5.

who showed that BAPS could reliably discriminate different gene pools only under restricted migration (i.e. $N_em = 1$). Additional simulations considering genotype data sets of two independent samples drawn from a single panmictic population showed that, in the presence of null alleles, BAPS correctly inferred a single population for all simulated data sets with sample size equal to 20 or 60 genes.

Simulation results regarding the performance of the NJ algorithm and the Cavalli-Sforza & Edwards' (1967) distance to detect the correct topology in the presence and absence of null alleles are summarized in Fig. 3. Topological errors did not increase linearly with splitting times either in the absence or presence of null alleles. The power to detect the correct topology was higher for intermediate splitting times (t = 1000 and 5000 generations) than for both low (t = 200) or large (t = 25 000) splitting times. These results parallel those of Takezaki & Nei (1996), who showed that the sampling error of genetic distances is large for low divergence levels and that the average values of the D_C genetic distance reach a plateau at high divergence levels. The presence of null alleles only slightly decreased the percentage of correct topologies. For small splitting times, the correction procedure partially rectified the errors in recovering correct topologies induced by null alleles. For large splitting times, the percentages of correct topologies were greater than those obtained in the absence of null alleles.

Genetic variation within and between L. migratoria *populations*

In our real *Locusta migratoria* data set, microsatellites showed a high level of polymorphism corresponding to $N_e\mu_R$ around 5. Based on our simulation results, we subsequently focused our comparative study of within-population diversity on H_E. Mean H_E values over loci were globally high in all populations, ranging from 0.68 to 0.92 according to the population sample (see Fig. S2, Supplementary material for estimated values and 95% confidence intervals of H_E). GLM analysis showed that the factors *Outbreaking* and *Insularity* did not have significant effects on H_E (likelihood-ratio chi-squared tests; deviance ratio = 0.160 and 0.036, respectively; P = 0.690 and 0.848, respectively), nor did their interaction (likelihood-ratio chi-squared test; deviance ratio = 0.135; P = 0.713).

Significant genotypic differentiation ($P \leq 0.05$) was observed in 282 of the 300 pairwise comparisons between *L. migratoria* populations (see Table S2, Supplementary material for Fisher's exact test P values). Nonsignificant differentiation was found among many African and all Malagasy pairwise comparisons. The global level of population differentiation is moderate with an $F_{ST}^{\{ENA\}} = 0.074$ (95% confidence intervals computed by bootstrapping 10 000 replicates over all loci = [0.054–0.095]). Figure 4 shows that the pairwise $F_{ST}^{\{ENA\}}$ distribution was skewed, with an excess of small values (e.g. 75% of values ≤ 0.1) and a long tail with large values (e.g. Japan vs. Reunion islands: $F_{ST}^{\{ENA\}} = 0.265$).

Identification of L. migratoria *genetic clusters*

The BAPS and NJ-tree analyses both detected large-scale patterns of genetic population structure across the range of *L. migratoria*. These patterns, in turn, clearly revealed the poor congruence between genetic clustering and traditional *L. migratoria* subspecies taxonomy.

Because our simulations indicated that BAPS more effectively discriminates gene pools in the presence of null alleles, while not detecting more gene pools than actually exist, we used the raw uncorrected microsatellite data set

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

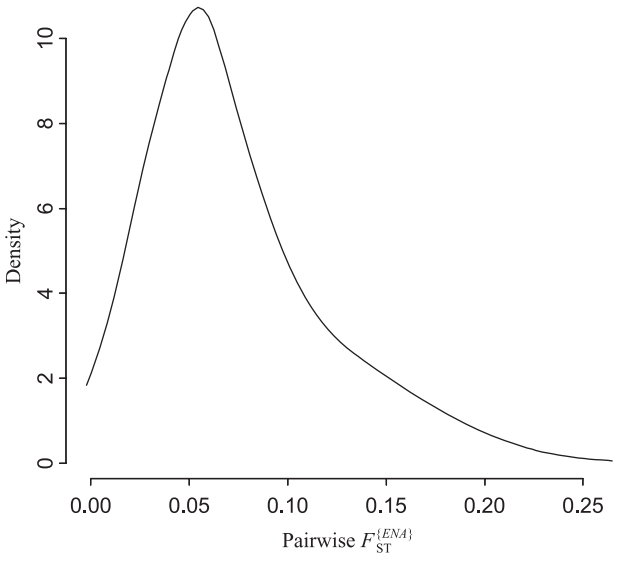

Fig. 4 Density estimation of pairwise $F_{ST}^{\{ENA\}}$ values computed between 25 *Locusta migratoria* populations genotyped at 14 microsatellite loci. We used the locfit function (Loader 1996) implemented in version 2.1.1 of the R package (Ihaka & Gentleman 1996; http://cran.r-project.org).

to infer the number of population clusters in *L. migratoria*. We found the highest posterior probability for structuring into nine population clusters (probability of the model = 1) depicted with colour codes in Fig. 1. Some of the genetic clusters of populations were widely distributed across continental land masses. One encompassed the whole African continent, the Arabian Peninsula and Southern Europe, as well as three of the subspecies determined by the classical taxonomic approach, *L. m. migratorioides*, *L. m. cinaerescens*, and the Arabian subspecies. Another included Chinese, Uzbekistan, Russian and French populations, and was comprised of the *L. m. manilensis*, *L. m. migratorioides*, and *L. m. gallica* subspecies. At a finer scale, two French populations located in the northwestern limit of the species' range sorted out separately in a single genetic cluster (populations 9 and 11). The other genetic clusters included island populations that, despite belonging to the same subspecies, were geographically restricted to their respective islands, except for the populations of Sumatra, Sumba, and Australia that formed a single genetic cluster. For instance, the *L. m. manilensis* subspecies fell out in three different genetic clusters from China, Borneo, and the Sumatra–Sumba–Australia complex.

As expected from our simulation study, the NJ trees obtained from our real *L. migratoria* data set had identical topologies when using uncorrected or corrected data sets (see Fig. 5 for NJ tree based on corrected data set). The identified clusters of populations are largely congruent with those obtained using BAPS. The NJ tree discriminated two main biogeographical clusters; one including the Eurasian and African continents, and a second including populations of the Pacific and Indian Ocean islands. Although the Eurasia–African group included populations from seven of the nine sampled subspecies determined by the classical taxonomic approach, the relationships among populations could not be resolved further within this continental cluster (cf. low bootstrap values). In contrast, the group of island populations appeared highly structured genetically (cf. high bootstrap values), although the classical taxonomy defined only four subspecies within this group.

Testing outbreaks as a factor driving L. migratoria *genetic structure*

BAPS genetic clustering failed to support the idea of outbreaks as a pertinent factor underlying either genetic structure or taxonomy in *L. migratoria*. For instance, the geographically close populations from Borneo and Sumba belonged to two distinct BAPS genetic clusters, although they shared the same high frequency and intensity of outbreak events. In agreement with this pattern, the NJ tree indicates that the cluster composed of European, African, and Asian populations is comprised of locusts from populations with a mixture of outbreak patterns. Mean D_C genetic distances between *L. migratoria* populations were not significantly correlated to differences in propensity to outbreak (Mantel test; $P = 0.164$). Similarly, none of the single-locus D_C genetic distances accounted for differences in propensity to outbreak (Mantel tests with false discovery rate correction; $P > 0.198$).

Statistical analysis based on the program GESTE gave the highest probability to explain the observed pattern of local genetic differentiation for the model including a constant, the factors *Outbreaking* and *Insularity*, and their interaction (model 9; posterior probability averaged over replicate analyses $P = 0.40$; Table 1). The second most probable model also included the factors *Insularity* and *Outbreaking*, but not the interaction term (model 7; $P = 0.38$; Table 1). The only other considered model included a constant and *Insularity* (model 3; $P = 0.22$; Table 1). Models that did not include *Insularity* were readily excluded, suggesting that the effect of this factor is strong. In contrast, the factor *Outbreaking* has a much weaker influence on local genetic differentiation, since the model with a constant and *Insularity*, which does not include this factor, has a non-negligible posterior probability (model 3; $P = 0.22$; Table 1).

Under the most probable model, the estimates of the α_i regression coefficients were similar over the three simulation replicates. We obtained a positive value for $\alpha_1 = 0.77$ (0.45; 1.10), indicating that local genetic differentiation increased with *Insularity*, and a negative value for $\alpha_2 = -0.42$ (–0.75; –0.11), indicating that genetic differentiation tends to decrease for outbreaking populations. The negative value

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

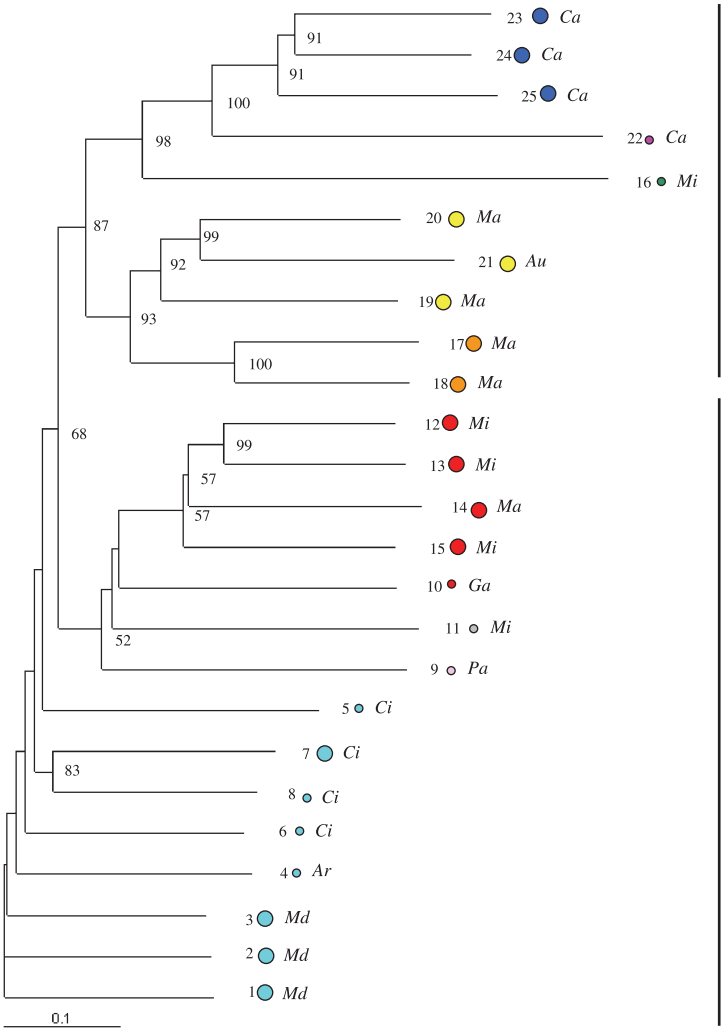

Fig. 5 Neighbour-joining tree of 25 *Locusta migratoria* populations based on Cavalli-Sforza & Edwards' (1967) distances computed from 14 microsatellite loci. Bootstrap support values (1000 replicates) are shown only where the support exceeded 50%. The tree is unrooted because all microsatellite loci developed for this species did not show cross-species applicability (Zhang *et al.* 2003; Chapuis *et al.* 2005), making it impossible to obtain outgroup genotypes. Sampling numbers with similar colours belong to the same genetic cluster as assessed by the Bayesian method of Corander *et al.* (2003). Outbreak events are depicted as uncommon and local (small circle) or frequent and widespread (large circle) (see Fig. 1). Subspecies taxonomy as follows: *Au*, Australian subspecies; *Ar*, Arabian subspecies; *Ca*, *L. m. capito*; *Ci*, *L. m. cinaerescens*; *Ga*, *L. m. gallica*; *Ma*, *L. m. manilensis*; *Md*, *L. m. migratorioides*; *Mi*, *L. m. migratoria*; *Pa*, Palavas form.

Table 1 Posterior probabilities of the nine possible models to explain local genetic differentiation in *L. migratoria* as a function of the factors *Insularity* and *Outbreaking*

Model	Factors included	Replicate 1	Replicate 2	Replicate 3	Mean over replicates
1	Constant	0	0	0	0
2	*Insularity*	0	0	0	0
3	Constant and *Insularity*	0.25	0.17	0.23	0.22
4	*Outbreaking*	0	0	0	0
5	Constant and *Outbreaking*	0	0	0	0
6	*Insularity* and *Outbreaking*	0	0	0	0
7	Constant, *Insularity* and *Outbreaking*	0.36	0.41	0.37	0.38
8	*Insularity*, *Outbreaking* and Interaction	0	0	0	0
9	**All**	**0.39**	**0.42**	**0.40**	**0.40**

The model with the highest probability is shown in bold.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

of $\alpha_3 = -0.36$ (−0.69; −0.04), indicates that the decrease of local genetic differentiation with propensity to outbreak was greater in continental populations than in island populations (see Fig. S3, Supplementary material for mode estimates and HPDIs of local F_{ST} values).

The estimated value of σ^2, a measure of model fit, was 0.53 [0.29; 1.00], for the most probable model 9, which indicated that the factors *Insularity* and *Outbreaking* are far from explaining the entire pattern of local genetic differentiation, and hence, that some other, unaccounted for, environmental or life-history factors are likely to play a role.

Discussion

Because we found *Locusta migratoria* microsatellites highly prone to null alleles, we first discuss the results of our computer simulations on the effects of frequent null alleles on analyses of population genetic variation. We then address the worldwide genetic structure of *L. migratoria*, with a particular reference to the influence of the propensity of some of its populations to outbreak.

Analysis of data sets with high prevalence of null alleles

Because microsatellite null alleles have been found in a wide range of taxa (Dakin & Avise 2004) and our results are consistent across a large range of levels of genetic diversity (see Figs S4 and S5, Supplementary material for simulation results with $N_e\mu_R = 1$), our conclusions about the effects of null alleles should be generally applicable to a variety of evolutionary and population genetics studies. The presence of null alleles led to an underestimation of all statistics traditionally used to summarize genetic variation within populations, but the bias was lower for Nei's (1987) expected heterozygosity and particularly small for large levels of genetic diversity. Moreover, the INA method for correcting genotype data with null alleles performed poorly for all studied statistics. This is because this correction makes the unrealistic assumption of a single null allele state regardless of the null allele frequency. Therefore, although the bias induced on HWE is rectified, the bias induced on the level of genetic diversity remains large if more than one allele is null.

Our simulations also showed that different gene pools were more frequently discriminated by the Bayesian analysis of Corander *et al*. (2003) when data sets harboured null alleles, at least for intermediate null allele frequencies (i.e. = 0.19) and high levels of gene flow ($N_em \geq 1$). This result is consistent with the presence of null alleles causing an upward bias of genetic differentiation, as recently shown for F_{ST} and genetic distances by Chapuis & Estoup (2007). The discrimination bias of BAPS is expected to become negligible for low null allele frequencies because of the minimal effects on genetic differentiation in this case (Chapuis & Estoup 2007). The dramatic loss of information inherent to markers with high null allele frequencies is also expected to limit the increase of BAPS power in the presence of null alleles. The INA correction procedure partly decreased the upper power of BAPS to discriminate populations in the presence of null alleles. This result is consistent with the correction method partly decreasing the bias induced by null alleles in population differentiation estimates (Chapuis & Estoup 2007).

Finally, we confirmed the expectation of Chapuis & Estoup (2007) that null alleles had a limited effect on NJ-tree topologies based on D_C genetic distances. Additional simulations showed that the performance of other genetic distances traditionally used for recovering tree topologies (i.e. those of Nei *et al*. 1983, Nei 1978, and Goldstein *et al*. 1995) was poorer than that of D_C in both the absence and presence of null alleles (see Fig. S6, Supplementary material for more details). Furthermore, data sets corrected for null alleles were only slightly better in recovering the correct tree topology than data sets without null alleles, especially for large splitting times. Chapuis & Estoup (2007) showed that genetic distances calculated from corrected data sets were underestimated, at least when null allele frequencies were high. It is therefore plausible that the correction method sufficiently underestimated large genetic distances to move them far from the plateau where the power to detect the correct topology of the NJ algorithm and the D_C genetic distance is limited.

Influence of the propensity to outbreak on genetic structure

To date, *L. migratoria* subspecific taxonomy has been based on morphometrical criteria traditionally used to define the plastic phenotypic phase states of this species in the field. The variable effect of the environment on *L. migratoria* morphology likely explains why its worldwide genetic structure inferred from microsatellite data is largely incongruent with the established subspecific taxonomy. Furthermore, our results are not consistent with the view that *L. migratoria* populations have differentiated with respect to their propensity to outbreak. Genetic population clusters contained a combination of populations from frequent and widespread historical outbreak areas as well as uncommon and local historical outbreak areas. This result held for the entire multilocus set and for each single locus analysed independently.

The poor genetic clustering in relation to the propensity to outbreak observed in our study might suggest that none of the 14 neutral markers are in linkage disequilibrium with the genes involved in the differential expression of the phase polyphenism among populations of *L. migratoria*. However, one cannot exclude the possibility of linkage disequilibrium of at least some of our microsatellite markers

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

with some of the genes under differential selection. In this case, our results might be explained by large levels of gene flow counteracting adaptive gene frequency changes at these genes and on linked (microsatellite) loci.

Our results indeed suggest that substantial gene flow occurs, at the very least, across continental areas. Indeed, the level of genetic structure was very low among continental populations, except for a few populations in the northwestern limit of the species' range. Importantly, the GESTE Bayesian analysis suggests that the propensity to outbreak itself leads to increased gene flow. This analysis gave support for outbreaking populations being less genetically structured than nonoutbreaking populations (posterior probability $P = 0.78$). Moreover, we found estimates of genetic diversity to be relatively similar among outbreaking and nonoutbreaking populations. Altogether, these results suggest similar harmonic means of effective population sizes and higher gene flow among historically outbreaking populations. Outbreak events may also impact *L. migratoria* neutral genetic structure due to long-distance colonization events from outbreaking to nonoutbreaking habitats. In agreement with this, the West European *L. m. gallica* population was included in the Asian and East European BAPS genetic cluster and was clearly related to Asian and East European populations in the NJ gene tree. However, the level of differentiation of *L. m. gallica* with Asian and East-European populations [$F_{ST}^{\{ENA\}}$ = (0.028–0.032)] was higher than that among Asian and East-European populations [$F_{ST}^{\{ENA\}}$ = (0.002–0.012)], suggesting that this population has been isolated for a substantial time since its foundation. This population may have been founded during the last Western European outbreak event that occurred at the 14th century and originated from the Black Sea area, a scenario suggested by Waloff (1940).

The advent of genomics resources in locusts has fostered further efforts to elucidate the molecular genetic mechanisms underlying locust phase change (Kang *et al*. 2004; De Loof *et al*. 2006; Ma *et al*. 2006; G.A. Sword *et al*. unpublished data). Understanding the sequence variation and transcript expression patterns of such candidate genes in a phylogeographical context will be of great value in further examining the importance of selected genetic factors in promoting locust phase change and outbreaks.

Biogeographical factors associated with worldwide genetic structure

Several lines of evidence suggest that the detection of moderate reductions in gene flow between historically outbreaking and nonoutbreaking areas may have been confounded by biogeographical barriers to gene flow. First, our microsatellite data indicate that oceans and high mountain ranges represent significant barriers to dispersal for *L. migratoria*. Populations from Indian and Pacific Ocean islands showed significantly higher levels of genetic differentiation than populations from continental Europe, Asia, and Africa (GESTE posterior probability $P = 1.00$; see Fig. S3 for mode estimates and HPDIs of local F_{ST} values). This pattern suggests lower levels of gene flow due to geographical isolation on islands. Furthermore, the continental Eurafrican and Eurasian groups of populations separated by the BAPS genetic clustering had intermediate pairwise $F_{ST}^{\{ENA\}}$ values, varying from 0.026 to 0.066. These two major biogeographical groupings appear to be physically separated by the Northern Hemisphere mountain belt stretching from the Alps to Tibet. Unfortunately, we could not obtain samples from populations of the Indian and Tibetan subspecies that would have been necessary to better address this hypothesis. Interestingly however, the western part of the European continent comprised populations belonging to both the Eurafrican and the Eurasian BAPS genetic clusters, suggesting that this area is a contact zone between these two relatively isolated groups of populations.

Second, climatic factors appeared to impact genetic variation among populations. Australian and Indonesian population samples were more closely related to African [$F_{ST}^{\{ENA\}}$ = (0.035–0.070)] than to Chinese [$F_{ST}^{\{ENA\}}$ = (0.085–0.127)] populations. Farrow & Colless (1980) already suggested that Australian *L. migratoria* were more closely related to African than to Asian locusts based on morphometric criteria. Because Africa, Australia, and Indonesia share tropical environments, as opposed to the Palaeoarctic conditions of China, our microsatellite survey suggests that locusts may disperse and colonize more effectively within than between different biomes (tropical vs. temperate).

Because at a worldwide scale the effects of outbreak events might have been confounded, at least partly, with those of the above biogeographical factors, additional studies based on more regional sampling scales are needed to gain further support and better assess the homogenizing effect of outbreak events in *L. migratoria*.

Acknowledgements

We are grateful to S. Arsol, Mr Badaruddin, L. Baoping, D. Brown, N. Clinquennois, O. Deveci, A. El-Heneidy, P. Fontana, A. Foucart, A. Guendouz-Benrima, O. Hasbullah, Y. Hidayat, D. Hunter, Mr Kadiran, I. Kamal, Mr Karsono, A. Khan, P. Meunier, O.N. Naumovich, D. Llucià Pomares, F. Ramli, T. Rachadi, G. Reynaud, X. Shengquan, S. Soumavro, Mr Sumarli, F. Tami, S. Tanaka, S. Yskak, T. Yulianti, S. Wangpeng, Mr Widji, Mr Winarto, D.-X. Zhang, G. Zhiyong, and other collectors for providing us with *L. migratoria* population samples. M.P.C. was supported by a Bourse Docteur Ingénieur grant from the Centre National de la Recherche Scientifique from November 2002 to November 2005 and has been supported by a Marie-Curie Outgoing Fellowship within the Sixth European Community Framework Programme since February 2007. This work was funded by the Institut National de la Recherche Agronomique (Santé des Plantes et Environnement) and the Centre de coopération internationale en recherche

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

agronomique pour le développement. YM acknowledges support from the Centre National de la Recherche Scientifique and the Institut de la Recherche et du Développement.

References

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Bailey NW, Gwynne DT, Ritchie MG (2005) Are solitary and gregarious Mormon crickets (*Anabrus simplex*, Orthoptera, Tettigoniidae) genetically distinct? *Heredity*, **95**, 166–173.

Benfekih L, Chara B, Doumandji-Mitiche B (2002) Influence of anthropogenic impact on the habitats and swarming risks of *Dociostaurus maroccanus* and *Locusta migratoria* (Orthoptera, Acrididae) in the Algerian Sahara and the semiarid zone. *Journal of Orthoptera Research*, **11**, 243–250.

Benjamin Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B: Series B*, **57**, 289–300.

Betts E (1961) Outbreaks of the African migratory locust (*Locusta migratoria migratorioides* R. & F.) since 1871. *Anti-Locust Memoir*, **6**, 1–25.

Brown HD (1986) New locust problem. *Antenna*, **10**, 11–13.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular and Biology Evolution*, **24**, 621–631.

Chapuis M-P, Loiseau A, Michalakis Y, Lecoq M, Estoup A (2005) Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta migratoria*. *Molecular Ecology Notes*, **5**, 554–557.

Chapuis M-P, Estoup A, Augé-Sabatier A *et al*. (2008) Genetic variation for parental effects on the propensity to gregarise in *Locusta migratoria*. *BMC Evolutionary Biology*, **8**, 37.

Chapuis M-P, Popple J-A, Simpson SJ *et al*. (in press) Eight polymorphic microsatellite loci for the Australian plague locust, *Chortoicetes terminifera*. *Molecular Ecology Resources*. doi: 10.1111/j.1755-0998.2008.02204.x.

Chen Y-L (1991) *The Migratory Locust, Locusta Migratoria, and its Asiatic Subspecies*. Ortopterists' Society and Lyman Entomological Museum, Québec.

COPR (1982) *The Locust and Grasshopper Agricultural Manual*. Centre for Overseas Pest Research, London.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163**, 367–374.

Costello AB, Down TE, Pollard SM, Pacas CJ, Taylor EB (2003) The influence of history and contemporary stream hydrology on the eolution of genetic diversity within species: an examination of microsatellite DNA variation in bull trout, *Salvelinus confluentus* (Pisces: salmonidae). *Evolution*, **57**, 328–244.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504–509.

De Loof A, Claeys I, Simonet G *et al*. (2006) Molecular markers of phase transition in locusts. *Insect Science*, **13**, 3–12.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B: Statistical Methodology*, **39**, 1–38.

Dopman EB, Sword GA, Hillis DM (2002) The importance of the ontogenetic niche in resource-associated divergence: evidence from a generalist grasshopper. *Evolution*, **56**, 731–740.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among population of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco]. *Theoretical Applied Genetics*, **92**, 832–839.

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: *Advances in Molecular Ecology* (ed. Carvalho G), pp. 55–86. Amsterdam NATO Press, Amsterdam.

Farrow RA (1979) Causes of recent changes in the distribution and abundance of the migratory locust (*Locusta migratoria* L.) in Australia in relation to plagues. *CSIRO Division of Entomology Commonwealth Scientific and Industrial Organization, Canberra*, **9**, 32.

Farrow RA (1987) Effects of changing land use on outbreaks of tropical migratory locust, *Locusta migratoria migratorioides* (R. & F.). *Insect Scientific Application*, **8**, 969–975.

Farrow RA, Colless DH (1980) Analysis of the interrelationships of geographical races of *Locusta migratoria* (Linnaeus) (Orthoptera: Acrididae) by numerical taxonomy, with special reference to sub-speciation in the tropics and affinities of the Australian race. *Acrida*, **9**, 77–99.

Foll M, Gaggiotti OE (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–891.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences, USA*, **92**, 6723–6727.

Hakomori T, Tanaka S (1992) Genetic control of diapause and other developmental traits in Japanese strains of the migratory locust, *Locusta migratoria* L.: univoltine vs. bivoltine. *Japanese Journal of Entomology*, **60**, 319–328.

Hedges SB (1992) The number of replications needed for accurate estimation of the bootstrap *P*-value in phylogenetic studies. *Molecular and Biology Evolution*, **9**, 366–369.

Heuertz M, Carnevale S, Fineschi S *et al*. (2006) Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (Oleaceae): roles of hybridization and life history traits. *Molecular Ecology*, **15**, 2131–2140.

Hudson RR (1990) Gene genealogies and the coalescent process in. In: *Oxford surveys in Evolutionary Biology* (eds Futuyama DJ, Antonovics J), pp. 1–44. Oxford University Press, New York.

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computation and Graphical Statistics*, **5**, 299–314.

Kang L, Chen X, Zhou Y *et al*. (2004) The analysis of large-scale gene expression correlated to the phase changes of the migratory locust. *Proceedings of the National Academy of Sciences, USA*, **101**, 17611–17615.

Karabag T (1958) *The Orthoptera Fauna of Turkey. A Synonymic and Distributional Catalogue of Turkish Orthoptera*. Ankara University Sciences Faculty, Ankara, Turkey.

Larson G, Dobney K, Albarella U *et al*. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **307**, 1618–1621.

Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a 'continuous' population under isolation by distance. *Molecular Biology and Evolution*, **20**, 491–502.

Lecoq M, Sukirno (1999) Drought and exceptional outbreak of the oriental migratory locust, *Locusta migratoria manilensis* (meyen

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

1835) in Indonesia (Orthoptera: Acrididae). *Journal of Orthoptera Research*, **8**, 153–161.

Legendre P, Lapointe F-J (2004) Assessing congruence among distance matrices: single-malt scotch whiskies revisited. *Australian & New Zealand Journal of Statistics*, **46**, 615–629.

Loader CR (1996) Local likelihood density estimation. *Annals of Statistics*, **24**, 1602–1618.

Ma Z, Yu J, Kang L (2006) LocustDB: a relational database for the transcriptome and biology of the migratory locust (*Locusta migratoria*). *BMC Genomics*, **7**, 11.

McConway KJ, Jones MC, Taylor PC (1999) *Statistical Modelling using* GENSTAT. Edward Arnold, London.

Mousseau TA (2000) Intra- and interpopulation genetic variation: explaining the past and predicting the future. In: *Adaptive Genetic Variation in the Wild* (eds Mousseau TA, Sinervo B, Endler JA), pp. 219–250. Oxford University Press, New York.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

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

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*, **19**, 153–170.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357–358.

Pener MP, Yerushalmi Y (1998) The physiology of locust phase polymorphism: an update. *Journal of Insect Physiology*, **44**, 365–377.

Randriamanantsoa M (1998) *Manuel sur la Lutte Antiacridienne*. Direction de la protection des végétaux et Deutsche Gesellschaft für Technische Zusammenarbeit GmbH, Antananarivo, Madagascar.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Heredity*, **86**, 248–249.

Remaudière G (1940) Contribution à l'étude des *Locusta migratoria migratoria* L. ph. solitaria de la région de Palavas (Hérault) 1ère partie. *Revue de Pathologie Végétale et d'Entomologie Agricole de France*, **17**, 147–163.

Remaudière G (1947) Sur l'existence en France d'une nouvelle sous-espèce de *Locusta migratoria* L. *Comptes Rendus Des Séances de l'Académie Des Sciences*, **225**, 1025–1026.

Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Mathematical Biosciences*, **53**, 131–147.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstruction of phylogenetic trees. *Molecular and Biology Evolution*, **4**, 406–425.

Simpson SJ, Sword GA (2008) Phase polyphenism in locusts: Mechanisms, population consequences, adaptive significance and evolution. In: *Phenotypic Plasticity of Insects: Mechanisms and Consequences* (eds Whitman D, Ananthakrishnan T), pp. 93–135. Science Publishers Inc., Plymouth.

Song H (2004) On the origin of the desert locust *Schistocerca gregaria* (Forskål) (Orthoptera: Acrididae: Cyrtacanthacridinae). *Proceedings of Royal Society B: Biological Sciences*, **271**, 1641–1648.

Sword GA (2002) A role for phenotypic plasticity in the evolution of aposematism. *Proceedings of the Royal Society B: Biological Sciences*, **269**, 1639–1644.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.

Uvarov BP (1966) *Grasshoppers and Locusts*, Vol. 1. Cambridge University Press, Oxford, UK.

Uvarov BP (1977) *Grasshoppers and Locusts*, Vol. 2. Centre for Overseas Pest Research, London, UK.

Uvarov BP, Hamilton AG (1936) Phase variation and rate of development in the Algerian race of the migratory locust (*Locusta migratoria*, L.). *Bulletin of Entomological Research*, **27**, 87–90.

Waloff ZV (1940) The distribution and migrations of *Locusta* in europe. *Bulletin of Entomological Research*, **31**, 211–246.

Waples RS, Gaggiotti OE (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*, **7**, 1569–1573.

Weir BS (1996). *Genetic Data Analysis II*. Sinauer Associates, Sunderland, Massachusetts.

Whiteley AR, Spruell P, Allendorf FW (2004) Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. *Molecular Ecology*, **13**, 3675–3688.

Wintrebert D (1970) Identité, écologie et comportement du criquet migrateur dans le sud-ouest Malgache. *Annales de la Société Entomologique de France*, **6**, 35–152.

Zhang Z, Li D (1999) A possible relationship between outbreaks of the oriental migratory locust (*Locusta migratoria manilensis* Meyen) in China and the El Niño episodes. *Ecological Research*, **14**, 267–270.

Zhang DX, Yan LN, Ji YJ *et al*. (2003) Isolation, characterization and cross-species amplification of eight microsatellite DNA loci in the migratory locust (*Locusta migratoria*). *Molecular Ecology Notes*, **3**, 483–486.

M-P. Chapuis is currently using molecular and evolutionary ecology approaches to understand the evolution of population outbreaks and extreme density-dependent phenotypic plasticity in locust species. M. Lecoq is currently interested in understanding the population dynamics and underlying mechanisms of outbreak formation of various Orthopteran pest species, including the migratory locust and the red locust in Madagascar, as well as the Senegalese grasshopper in Niger. The focus of his work is also on improving the preventive control strategies of the desert locust in Western Africa. Y. Michalakis main interests are host-parasite interactions and the evolution of life-history traits, reproductive system and dispersal. A. Loiseau is a molecular biology technician working on various insect and rodent populations. G. A. Sword is conducting research on a variety of major topics in ecology, evolution and behaviour of insects including migration, collective movement, phenotypic plasticity, the evolution of warning coloration and plant-herbivore interactions. S. Piry is a computer scientist mainly specialized on GIS cartography and treatment, and development of program for population genetics. A. Estoup's current research focuses mainly on the evolutionary biology of non equilibrium populations with a particular interest in invading species.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

Supplementary material

The following supplementary material is available for this article:

Fig. S1. Population split model for the simulation-based study of the effects of microsatellite null alleles on neighbour-joining tree reconstruction.

Fig. S2. Estimated values and 95% confidence intervals of expected heterozygosity (Nei 1987).

Fig. S3. Mode estimates and 95% HPDIs of local F_{ST} values obtained with the program GESTE under the model with the highest probability (i.e. model 9). The model 9 includes a constant and, the factors *Outbreaking* and *Insularity*, and their interaction. Results are those obtained from the first replicate run.

Fig. S4. Effects of null alleles and efficiency of the INA correction method on statistics describing within-population genetic variation for $N_e\mu_R = 1$.

Fig. S5. Effects of null alleles and efficiency of the INA correction method on BAPS analysis and NJ clustering method for $N_e\mu_R = 1$.

Fig. S6. Effects of null alleles and efficiency of the INA correction method on genetic distances traditionally used for recovering correct NJ tree topologies.

Table S1. Heterozygosities, HWE and null allele frequency estimation

Table S2. Pairwise $F_{ST}^{\{ENA\}}$ and P values of Fisher exact tests for genotypic differentiation between populations for 25 *Locusta migratoria* populations using 14 microsatellite loci

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03869.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd