MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 4344–4358 doi: 10.1111/j.1365-294X.2012.05684.x

Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust

CHUAN MA,* PENGCHENG YANG,* FENG JIANG,* MARIE-PIERRE CHAPUIS,† YASEN SHALI,‡ GREGORY A. SWORD§ and LE KANG*

**State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China, †School of Biological Sciences, University of Sydney, Sydney, NSW 2006, Australia, ‡Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China, §Department of Entomology, Texas A&M University, College Station, TX 77843-2475, USA*

Abstract

The migratory locust, *Locusta migratoria*, is the most widely distributed grasshopper species in the world. However, its global genetic structure and phylogeographic relationships have not been investigated. In this study, we explored the worldwide genetic structure and phylogeography of the locust populations based on the sequence information of 65 complete mitochondrial genomes and three mitochondrial genes of 263 individuals from 53 sampling sites. Although this locust can migrate over long distances, our results revealed high genetic differentiation among the geographic populations. The populations can be divided into two different lineages: the Northern lineage, which includes individuals from the temperate regions of the Eurasian continent, and the Southern lineage, which includes individuals from Africa, southern Europe, the Arabian region, India, southern China, South-east Asia and Australia. An analysis of population genetic diversity indicated that the locust species originated from Africa. Ancestral populations likely separated into Northern and Southern lineages 895 000 years ago by vicariance events associated with Pleistocene glaciations. These two lineages evolved in allopatry and occupied their current distributions in the world via distinct southern and northern dispersal routes. Genetic differences, caused by the long-term independent diversification of the two lineages, along with other factors, such as geographic barriers and temperature limitations, may play important roles in maintaining the present phylogeographic patterns. Our phylogeographic evidence challenged the long-held view of multiple subspecies in the locust species and tentatively divided it into two subspecies, *L. m. migratoria* and *L. m. migratorioides*.

Keywords: adaptive evolution, genetic divergence, genetic structure, *Locusta migratoria*, migration

Received 30 March 2012; revision received 21 May 2012; accepted 23 May 2012

Introduction

Phylogeographic studies aim to elucidate the historical mechanisms and processes that are responsible for the geographic distribution of genetic lineages, especially within and among closely related species, thereby contributing to the understanding of evolutionary history (Avise *et al.* 1987; Avise 2009). Frequent gene flow will reduce differentiation levels among populations (Mayr 1963), whereas other factors, such as historical processes, geographic barriers and local adaptation, can lead to population subdivision (Avise 2009; Hickerson *et al.* 2010). The information of contemporary population structure and gene flow provided by phylogeographic studies has important implications for species management and conservation (Avise 2009).

Correspondence: Le Kang, Fax: +86 10 64807099;
E-mail: lkang@ioz.ac.cn

The migratory locust, *Locusta migratoria*, is a notorious agricultural pest worldwide. Like other locust species, the migratory locust expresses remarkable traits, such as density-dependent phase transition and long-distance migration most notably observed as flying swarms (Uvarov 1966, 1977; Pener & Simpson 2009). The migratory capacity of this locust species is reflected in its extensive distribution, ranging from temperate to tropical zones across the Old World (Uvarov 1966, 1977; Farrow & Colless 1980). Mass migration in swarms provides the opportunity for extensive gene flow among populations. However, the migratory locust exhibits considerable levels of differentiation among various geographic populations (Chapuis *et al.* 2008; Zhang *et al.* 2009; Tokuda *et al.* 2010). Therefore, it is an excellent model organism for studying the global-scale phylogeography of highly migratory species. More importantly, an understanding of historical and ongoing migration patterns of the locusts is essential for developing pertinent strategies to minimize economic losses because of locust plagues.

Recently, population genetics of the migratory locust have attracted increasing attention because of its high economic importance. A genetic analysis of 25 geographic populations of the locust using 14 microsatellite markers distinguished two major phylogeographic clusters. One cluster consists of individuals from Eurasian and African continents; the other is from the Pacific and Indian Ocean islands (Chapuis *et al.* 2008). However, the evolutionary direction of the unrooted phylogenetic tree in such a study could not be resolved due to the lack of out-group microsatellite genotypes. Zhang *et al.* (2009) selected an African population as the out-group taxon to compare 25 geographic populations from China using eight microsatellite markers. Nevertheless, the out-group selection in this study seems inappropriate because the southern China populations are genetically closer to African populations than to northern China populations (Tokuda *et al.* 2010). A recent study on the migratory locust populations mainly from Japan demonstrated that these populations could be separated into two clades corresponding to their climate regions, although only four mitochondrial DNA (mtDNA) fragments were used (Tokuda *et al.* 2010). However, the evolutionary origin, diversification and colonization routes of the migratory locust across its extensive range remain unknown.

MtDNA has obvious advantages for establishing the genetic diversity level and phylogeographic structure of a species because of its rapid mutation rate and short coalescence time. MtDNA sequence analysis enables application of the genealogical approach and phylogenetic tools for addressing population-level questions (Brito & Edwards 2009). Hence, mtDNA sequences are the most widely used genetic markers for phylogeographic investigations. Compared with partial mtDNA sequences, complete mitochondrial genome (mtgenome) sequences have even higher resolutions and have been applied to a variety of phylogenetic and phylogeographic studies (Ingman *et al.* 2000; Gilbert *et al.* 2008; Morin *et al.* 2010). All 13 mitochondrial protein-coding genes encode the essential components of the electron transport chain, which are involved in energy (ATP) production and heat generation. These genes have been shown to undergo adaptive evolution in humans and a great variety of other animals because of their different metabolic requirements (Balloux *et al.* 2009; Ning *et al.* 2010; Foote *et al.* 2011; Sun *et al.* 2011). The population subdivision of the migratory locust between temperate and tropical regions (Tokuda *et al.* 2010) may imply the adaptive differentiation of mitochondrial genes in response to different environments. However, little attention has been paid to the association between mtgenome nucleotide mutations and locust adaptive evolution.

Taxonomically, the migratory locust *L. migratoria* is the sole species in the genus *Locusta* of the subfamily Oedipodinae. Given their wide geographic distributions, many geographic populations of the migratory locust species display variations in morphology, life history, physiology and other biological characteristics. Nine to 11 subspecies have been described based on morphometric characters and geographic distribution (Uvarov 1966; COPR 1982; FAO website, http://www.fao.org/ag/locusts-CCA/common/ecg/1078/en/LMI-_Distribution_map3.pdf). Recently, the validity of these subspecies was challenged by morphometric (Farrow & Colless 1980; Kang *et al.* 1989; Kang & Chen 1991) and molecular evidence (Zhang & Kang 2005; Chapuis *et al.* 2008; Zhang *et al.* 2009; Tokuda *et al.* 2010). Thus, more effective molecular markers must be utilized to resolve the subspecies status of the migratory locust.

In this study, we sequenced 65 complete mtgenomes and three mitochondrial genes (*cox1*, *nad5* and *cob*) of 263 individuals from 53 sampling sites covering the main distribution regions and respective populations of the migratory locust in the world. The aim was to (i) obtain a global picture of the population genetic structure to infer the origin and colonization patterns of the migratory locust; (ii) explore historical events responsible for the present genetic distribution patterns; and (iii) test the validity of multiple subspecies with a phylogeographic and evolutionary perspective.

Materials and methods

Collection of specimens

A total of 263 migratory locust specimens from 53 sampling sites were collected (Fig. 1, Table S1, Supporting

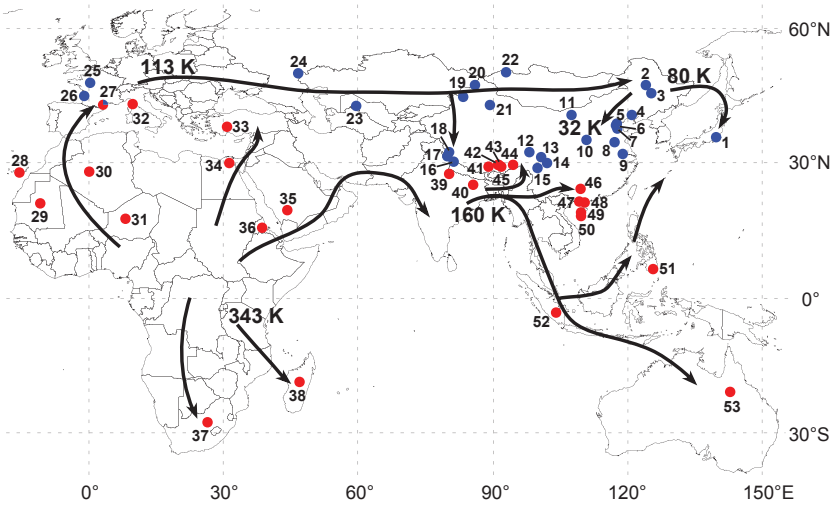


Fig. 1 Sampling localities of the migratory locust. Blue dots represent populations of the Northern lineage, and red dots represent populations of the Southern lineage. Detailed sampling information is presented in Table S1. A proposed scenario of early dispersal routes is illustrated with arrows.

information) and stored in high-concentration alcohol at 4 °C. The sampling covered the main distribution range of this species. Nearly, all traditionally defined subspecies were included according to their respective distribution ranges, as illustrated on the website of the Food and Agriculture Organization of the United Nations (FAO website, http://www.fao.org/ag/locusts-CCA/common/ecg/1078/en/LMI-_Distribution_map3.pdf).

MtDNA extraction, amplification and sequencing

Previous studies reported the presence of nuclear mitochondrial pseudogenes (numts) in the migratory locust (Gellissen *et al.* 1983; Zhang & Hewitt 1996; Song *et al.* 2008; Moulton *et al.* 2010; Berthier *et al.* 2011), which can potentially complicate subsequent phylogenetic analyses and even lead to false conclusions. To avoid numt co-amplification, mitochondria-rich tissues, an mtDNA-enrichment protocol (Ma *et al.* 2009) and long PCR amplifications were used. Nine pairs of primers were designed to amplify overlapping fragments covering the whole mtgenome, and three pairs of primers were designed to amplify *cox1*, *nad5* and *cob* genes from mtDNA-enriched genomic DNA (Table S2, Supporting information). The segment spanning the A + T-rich region was amplified using LA-Taq (Takara Co., Dalian, China) with the following thermal cycling conditions: 95 °C for 1 min; 30 cycles of 98 °C for 10 s, 60 °C for 10 s, and 65 °C for 140 s; 65 °C for 5 min. The other fragments were amplified with rTaq or ExTaq (95 °C for 3 min; 35 cycles of 95 °C for 30 s, 52 °C for 30 s, and 72 °C for 2–2.5 min; 72 °C for 5 min). All PCR products were purified and directly sequenced via primer walking. Sequencing data were assembled using the SeqMan software (DNAStar, Inc.). Transfer RNAs were identified by tRNAscan-SE 1.21 (Schattner *et al.* 2005), and the other genes were determined by comparing with those of the first sequenced migratory locust mtgenome (Flook *et al.* 1995). No numts were identified after comparisons with the published complete mtgenome (Flook *et al.* 1995) and examinations of sequence translation, nucleotide composition and codon composition biases for protein-coding genes.

Phylogenetic analysis

Gastrimargus marmoratus (GenBank accession numbers: NC_011114) and *Oedaleus asiaticus* (NC_011115) were selected as out-groups because of the close relationships of *Locusta* with *Gastrimargus* and *Oedaleus* (Fries *et al.* 2007; Ma *et al.* 2009). Each of the 37 genes and the A + T-rich region were aligned using BioEdit (Hall 1999) and concatenated for phylogenetic analysis. Poorly aligned nucleotide positions were omitted by the program Gblocks (Castresana 2000) with a more stringent selection criterion (i.e. do not allow many contiguous nonconserved positions). The phylogenetic relationships were inferred by Bayesian inference (BI) using MrBayes 3.1.2 (Huelsenbeck & Ronquist 2001) and maximum likelihood (ML) method using PhyML 3.0 (Guindon *et al.* 2010). For BI analysis, the data sets were divided into 17 partitions: each protein-coding gene, concatenated 22 tRNA genes, each rRNA gene, and the A + T-rich region. The best-fitting substitution model for each partition was selected by jModeltest (Posada 2008) under the Akaike information criterion. Two sets of four chains

were executed for 15 million generations using parameters unlinked among partitions. Each set was sampled every 1000 generations with a burnin of 20%. Bayesian posterior probabilities were estimated on a 50% majority rule consensus tree of the remaining trees. For ML analysis, as PhyML 3.0 does not support the partitioning strategy, the best-fitting model 'GTR + I + G' for the concatenated data set was implemented. The subtree pruning and regrafting (SPR) and nearest neighbour interchange (NNI) approach was used in tree improvement. Nodal support was evaluated by 100 bootstrap iterations.

Genetic diversity calculation, population differentiation estimation and haplotype network reconstruction

The mtgenome is an inheritance unit as a whole. Hence, the sequences of *cox1*, *nad5* and *cob* genes were concatenated into a data matrix and used for subsequent analyses. To assess how genetic diversity varied across geographic populations, we calculated the following summary statistics. Haplotype diversity (h), nucleotide diversity (π) and the mean number of pairwise differences were calculated to estimate DNA polymorphism using DnaSP5.10.01 (Librado & Rozas 2009). Analysis of molecular variance (AMOVA) and F_{ST} calculations were performed using Arlequin 3.5 (Excoffier & Lischer 2010) with 10 000 permutations based only on populations with more than three individuals. Genetic distance calculations among sequences were conducted using MEGA5 (Tamura *et al.* 2011) under the Kimura 2 model.

To compare genetic connections of geographic populations, a median-joining network was constructed with the software Network 4.6 (Bandelt *et al.* 1999). The initial analysis was based on the entire data set. Owing to the absence of shared variation (see Results and Discussion), the Southern and Northern lineages were then separately analysed. Given the relatively large number of haplotypes available, an initial star contraction with a maximum star radius of five was used to facilitate data representation and interpretation. To confirm our results further, partial *cox1* and *cob* sequences of 125 *L. migratoria* individuals as previously described (Tokuda *et al.* 2010) were also included to reconstruct the median-joining network with the above settings.

Estimation of divergence times

Divergence time estimation is necessary to identify historical events related to genetic clustering patterns of locust populations. Divergence times were estimated using BEAST (Drummond & Rambaut 2007) based on the *cox1* gene, with 'GTR+I+G' as the substitution model selected by jModeltest. Out-group taxa were excluded in the time dating analysis. Major clades within the Southern and Northern lineages were each constrained to be monophyletic. Two runs were executed for 50 million steps, sampling every 1000 steps and discarding the initial 20% as burnin. Coalescent tree priors were set to constant size model. No fossil or geological evidence was available for calibration; thus, the proposed insect molecular clock (*cox1* substitution rate = 1.77% per million years) (Papadopoulou *et al.* 2010) was adopted. To allow rate variation among branches, a relaxed clock with uncorrelated lognormal distribution was used (Drummond & Rambaut 2007). Samples from the two runs, which yielded similar results, were combined. Convergence of the chains was checked using the program TRACER v1.5 (http://tree.bio.ed.ac.uk/software/tracer) to ensure that effective sample sizes were above 200.

Demographic analysis

Signatures of population demographic changes were tested for (i) the East Asia clade, (ii) the Eurasian continent clade and (iii) the Southern lineage excluding an individual from Madagascar. First, Tajima's D and Fu's Fs statistics were calculated by Arlequin 3.5, with the deviation from neutrality determined from 10 000 coalescent simulations. Second, mismatch distributions of pairwise sequences were calculated using Arlequin with 1000 bootstrap replicates. To estimate the pattern of population size changes through time, Bayesian skyline plots (BSPs) implemented in BEAST 1.6.1 were constructed. For each BSP, the substitution model was selected using jModeltest. Samples were drawn every 1000 steps for 50 million steps under an uncorrelated lognormal relaxed clock model. The substitution rate was set to 1.77% per million years (Papadopoulou *et al.* 2010). The piecewise-linear skyline model and 10 groups were selected for Bayesian skyline coalescent tree priors. All other parameters used were default values. Demographic plots were visualized in Tracer v1.5 with a burnin of 20%.

Natural selection test

The ratio (ω) of the nonsynonymous (dN) to synonymous (dS) nucleotide substitution rate often serves as an indicator of natural selection on protein-coding genes. Positive, neutral and negative selection were represented by $\omega > 1$, $\omega = 1$ and $\omega < 1$, respectively (Hughes & Nei 1988). The codeml program of the PAML package (Yang 2007) was used to examine the potential adaptive evolution of mitochondrial protein-coding genes. The following analyses were based on the

tree topology derived from complete mtgenome sequences.

We first used the M0 (one-ratio) model, which assumes a constant ω value along all branches. To assess whether or not there were significant differences in the selective pressures on the Northern and Southern lineages of the migratory locust, we adopted two-ratio models assuming that the branches of interest (foreground branches) had different ratios from the background branches (Yang 2007). The Southern and Northern lineages were separately set as the foreground with all other taxa as the background. Significant difference between the one- and two-ratio models was evaluated by likelihood ratio tests. To determine the selective pressures on individual protein-coding genes between the Southern and Northern lineages, three ratios were defined: one for the Southern lineage, one for the Northern lineage and the third for the out-group taxa. The branch test often has little power to detect positive selection because of the stringent criterion that the ω ratio, which is averaged over all sites of the protein, should be more than one (Yang 2006). Therefore, we estimated the ω ratio for every site using the branch-site model A (Zhang *et al.* 2005) in combination with the Bayes empirical Bayes (BEB) method (Yang *et al.* 2005). Sites were considered under positive selection if the positive-selection model (Model A) fits the data significantly better than the corresponding null model (M1a).

Results

General features of mtDNA sequences

We sequenced 65 complete mtgenome sequences of the migratory locust from populations covering its main geographic area in Africa, Europe, Asia, Australia and the Pacific Ocean islands (Fig. 1, Table S1, Supporting information). These mtgenomes ranged in size from 15 901 to 16 057 bp, with length variations mainly in the A + T-rich regions that contained three or four copies of a 155-bp tandem repeat unit. The mtgenomes all harboured a typical set of 37 genes and an identical gene order to the first sequenced mtgenome of the migratory locust (Flook *et al.* 1995), except that the *trnW* gene in two individuals was inverted at its original position.

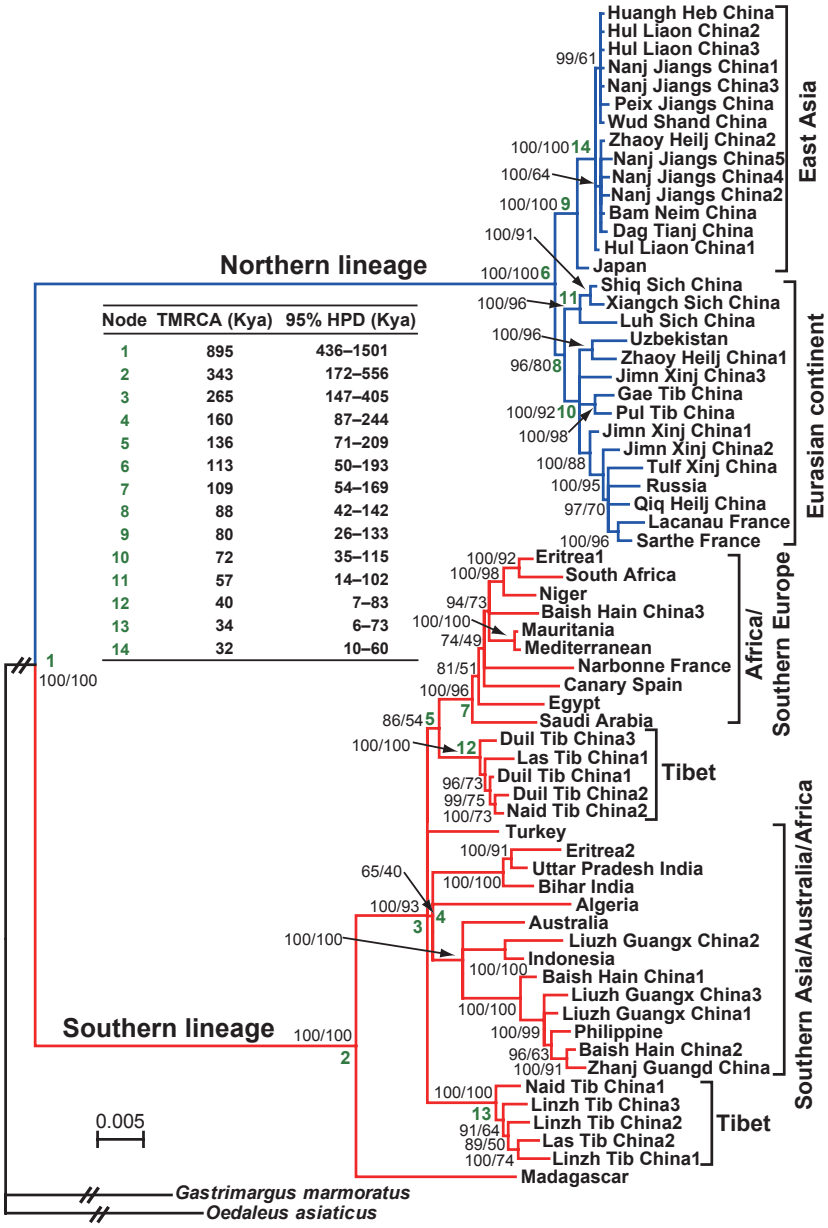
The three genes *cox1*, *nad5* and *cob* possessed the most parsimony-informative sites and partitioned Bremmer support values (Sorenson and Franzosa 2007) (Fig. S1, Supporting information), indicating that these genes provide the most phylogenetic information in complete mtgenome sequences. Thus, these genes were selected for use as efficient mtDNA markers in an expanded analysis of the locust phylogeography. The sequences of *cox1*, *nad5* and *cob* from 263 individuals were concatenated into a matrix 4374 bp in length, excluding stop codons. A total of 166 unique haplotypes were identified.

Phylogenetic reconstruction

Both BI and ML phylogenetic trees derived from 65 mtgenome sequences revealed two highly supported lineages: the Northern and Southern lineages (Fig. 2). The Northern lineage can be further divided into two clades: the East Asia and Eurasian continent. The East Asia clade included individuals from Japan, north-eastern China and the east coast of China, whereas the Eurasian continent clade included individuals from France, Russia, Uzbekistan, northern China and south-central China. Branch lengths in the East Asia clade were generally short, indicating low levels of genetic differentiation among populations. The Southern lineage, spanning a more extensive geographic range, comprised five major clades. The Madagascar clade was first separated from the Southern lineage and clearly differentiated from the other clades of this lineage. The Africa/southern Europe clade included individuals from the African mainland, Saudi Arabia, northern Mediterranean coast, southern France, Canary Island of Spain and Hainan Island of China. The Southern Asia/Australia/Africa clade was comprised of individuals from southern China, the Philippines, Indonesia, Australia, India, Turkey and Algeria. Interestingly, the individuals from south-eastern Tibet were separated into two clades. One clade was from the area upstream of the Brahmaputra River and close to the Africa/southern Europe clade. The other was mainly from subtropical areas of south-eastern Tibet. In contrast, individuals from north-western Tibet belonged to the Eurasian continent clade. Taken together, our phylogenetic analysis revealed multiple origins for the locust populations in Tibet.

Haplotype network analysis

The median-joining network analysis on the *cox1*, *nad5* and *cob* genes of 263 individuals corroborated the split of the Southern and Northern lineages in phylogenetic trees (Fig. 3). The two lineages had no distribution range overlap except for the southern France population, among which three individuals belonged to the Southern and one to the Northern lineage. The individuals from Europe, Central Asia and north-eastern China shared some haplotypes, indicating the frequent gene flow across the Eurasian continent despite the long geographic distance. North-western Tibetan populations belonged to the Northern lineage, whereas all individuals from south-eastern Tibet belonged to the Southern



Node	TMRCA (Kya)	95% HPD (Kya)
1	895	436–1501
2	343	172–556
3	265	147–405
4	160	87–244
5	136	71–209
6	113	50–193
7	109	54–169
8	88	42–142
9	80	26–133
10	72	35–115
11	57	14–102
12	40	7–83
13	34	6–73
14	32	10–60

Fig. 2 Bayesian phylogram based on sequences of the 37 mitochondrial genes and the A + T-rich region. Branch lengths indicate the expected number of substitutions per site. The branches of the out-group taxa, *Gastrimargus marmoratus* and *Oedaleus asiaticus*, are truncated as indicated by slashes. Numbers at nodes indicate Bayesian posterior probabilities (in percentage) and maximum likelihood bootstrap support values. Divergence time estimates were based on the *cox1* gene. Median time to the most recent common ancestor and 95% highest posterior density (HPD) are shown for nodes marked with numbers in green.

lineage. The haplotypes of the individuals from southern France, Turkey, Saudi Arabia and India largely clustered with those from Africa.

To confirm our results further, we combined 125 *cox1* and *cob* sequences from Tokuda *et al.* (2010) with our corresponding sequences from 263 individuals to construct the haplotype network, which showed similar patterns as above (Fig. S2, Supporting information). Consistent with the results of Tokuda *et al.* (2010), both the Southern and Northern lineages coexisted in two localities of Japan.

Genetic diversity and population genetic structure

Values for haplotype (h) and nucleotide (π) diversity were higher for individuals from the Southern lineage (h = 0.990 ± 0.003, π = 0.471% ± 0.013%) than those from the Northern lineage (h = 0.874 ± 0.028, π = 0.182%

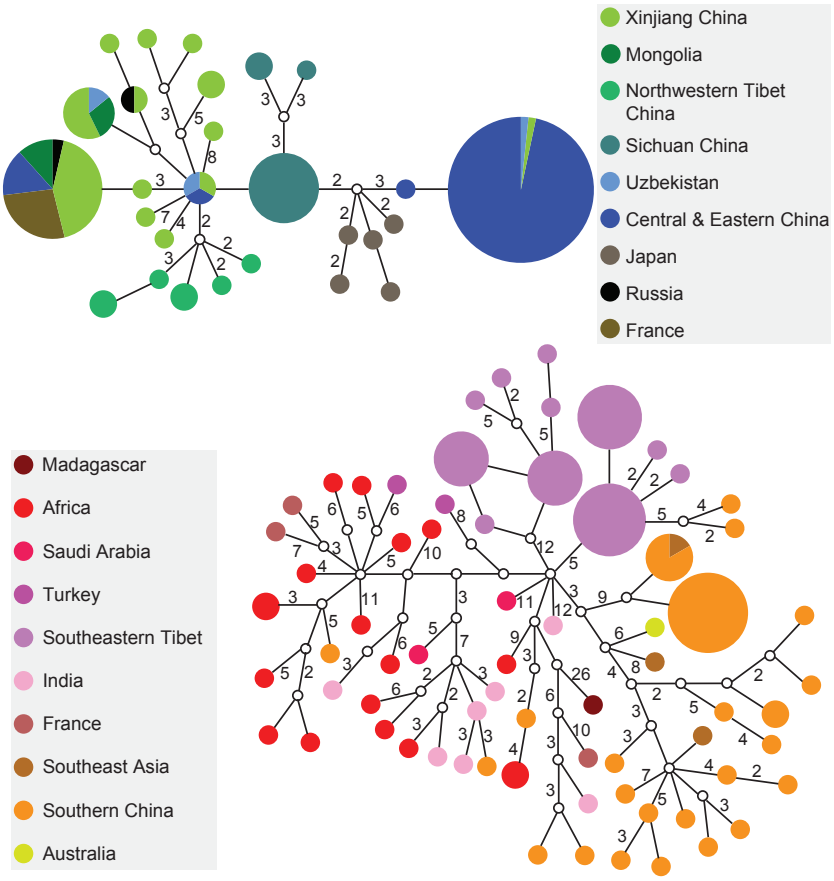


Fig. 3 Haplotype networks derived from concatenated *cox1*, *nad5* and *cob* genes. Networks of the Northern (above) and Southern (below) lineages were constructed separately. Sizes of circles are proportional to the haplotype frequencies. Small empty circles represent unsampled or extinct haplotypes. The number of mutations >1 is shown next to branches.

± 0.008%) (Table 1). Individuals from Africa had the highest genetic diversity (h = 0.993 ± 0.021, π = 0.462% ± 0.053%), whereas individuals of the East Asia clade displayed the lowest h and π values (0.537 ± 0.076 and 0.064% ± 0.015%, respectively).

The AMOVA analysis based on the combined data set (*cox1*, *nad5* and *cob* genes) revealed significant genetic differentiation among the worldwide populations of the migratory locust (F_{ST} = 0.924, P < 0.0001). The average F_{ST} values for the Northern and Southern lineages were 0.436 and 0.336, respectively. Most variation (87.45%) existed between the Southern and Northern lineages, followed by 7.58% within populations and 4.97% within each lineage. The average genetic distance between the two lineages (2.686%) was notably higher than that within each lineage (0.182% for the Northern lineage and 0.473% for the Southern lineage). In the Northern lineage, the mean genetic distance of the East

Table 1 Genetic diversity and neutrality test statistics

Group (n)	h	π (%)	k	Tajima's D	Fu's Fs
Northern lineage (133)	0.874 ± 0.028	0.182 ± 0.008	7.95	−2.18**	−24.65**
Southern lineage (130)	0.990 ± 0.003	0.471 ± 0.013	20.58	−2.08**	−23.87**
'East Asia' clade (66)	0.537 ± 0.076	0.064 ± 0.015	2.80	−1.97**	−7.59**
'Eurasian continent' clade (67)	0.990 ± 0.005	0.182 ± 0.010	7.97	−2.37**	−24.85**
Southern lineage excluding Madagascar (129)	0.990 ± 0.004	0.465 ± 0.012	20.32	−2.05**	−23.88**
Southern China (45)	0.969 ± 0.019	0.397 ± 0.032	17.37	−1.44	−9.96**
South-eastern Tibet (48)	0.956 ± 0.017	0.279 ± 0.013	12.19	−0.29	−5.76
Africa (18)	0.993 ± 0.021	0.462 ± 0.053	20.20	−1.70*	−2.89

n, individual numbers; h, haplotype diversity; π, nucleotide diversity; k, average number of nucleotide differences. $^{*}P$ < 0.05; $^{**}P$ < 0.01.

Asia clade was 0.064%, which was much lower than that of the Eurasian continent clade (0.183%). These results indicate conspicuous interlineage genetic differentiation and shallow intralineage differentiation.

Divergence time estimation

The mitochondrial molecular clock analysis estimated that the split of the Southern and Northern lineages was dated to 895 thousand years ago (Kya), with a 95% highest posterior density (HPD) of 436–1501 Kya (Fig. 2). The time to the most recent common ancestor was 113 Kya for the Northern lineage, 88 Kya for the Eurasian continent clade and 80 Kya for the East Asia clade. The divergence among populations from eastern China only occurred at 32 Kya. Major clades within the Southern lineage diverged earlier than those within the Northern lineage (Fig. 2). Within the Southern lineage, the Madagascar population diverged from others 343 Kya. The divergence time of the South Asia/Australia clade from the India and Arabian clades was estimated to be 160 Kya.

Demographic history

The East Asia clade, the Eurasian continent clade and the Southern lineage excluding the individual from Madagascar all had significantly negative Tajima's *D* and Fu's *Fs* values (Table 1), indicating population size expansion, genetic hitchhiking and/or selection in evolutionary history (Tajima 1989; Fu 1997). The unimodal mismatch distributions with small, nonsignificant Harpending's raggedness index (Fig. 4) demonstrated recent demographic expansion (Rogers & Harpending 1992) or a range expansion with high levels of migration between neighbouring demes (Ray *et al.* 2003; Excoffier 2004).The mismatch distribution for the East Asia clade was strongly biased toward low divergence values as distinguished by 0- and 1-nucleotide change (Fig. 4A), indicating a relatively recent expansion from a small number of ancestors.

The BSP of the effective population size through time revealed different profiles of historical demography. The population size of the East Asia clade remained stable for a long period, followed by a sharp rise at 5 Kya (Fig. 4A). The population size of the Eurasian continent clade rapidly expanded between 35 Kya and 15 Kya (Fig. 4B). Obviously different from the Northern lineage, the population size of the Southern lineage excluding the Madagascar clade increased in a slow manner over a long term (Fig. 4C).

Selection test

To test for signatures of selection on the mitochondrial protein-coding genes, we first examined the ω (*dN/dS*)

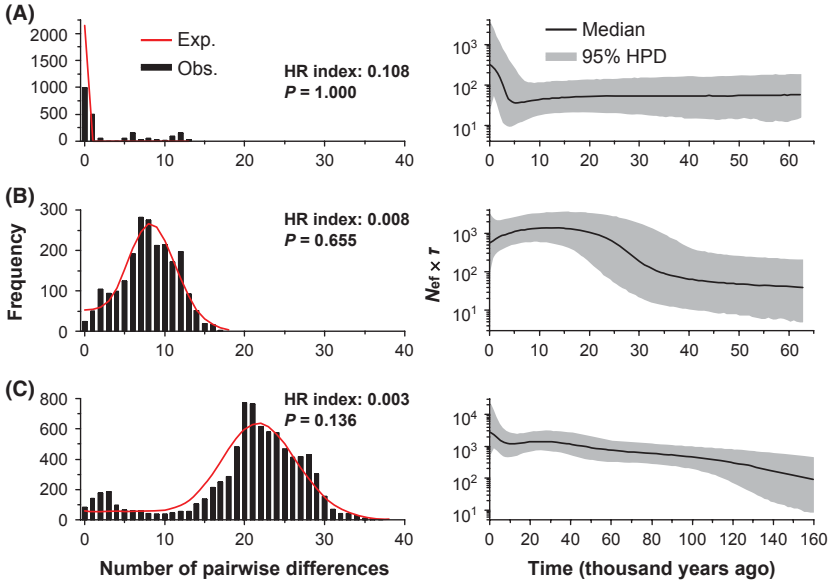


Fig. 4 Mismatch distributions (left) and Bayesian skyline plots (right). (A) The East Asia clade, (B) the Eurasian continent clade and (C) the Southern lineage excluding the individual from Madagascar were each calculated based on the concatenated sequences of the *cox1*, *nad5* and *cob* genes. The observed mismatch distribution is denoted by vertical bars, and the expected distribution under the population expansion model is represented by red lines. Harpending's raggedness (HR) indices are shown. For Bayesian skyline plots, middle lines represent median estimates of the effective population size (N_{ef}) multiplied by the generation time (τ) in years, and shaded areas represent 95% highest posterior densities (95% HPD). $N_{ef} \times \tau$ is presented on a logarithmic scale.

ratios for the combined sequences of all 13 genes using branch models (Yang 2007). The ω ratio in the one-ratio model was low ($\omega = 0.042$), suggesting that the locust mitochondrial protein-coding genes have evolved under strong selective pressure, which is consistent with the essential function of mitochondrial proteins. In the two-ratio model, the branch of the Southern lineage had a significantly lower ω ratio ($\omega = 0.028$; $P < 0.001$) than all other branches ($\omega = 0.046$), whereas the ω ratio for the Northern lineage ($\omega = 0.055$) was not significantly higher than that for other branches ($\omega = 0.040$; $P = 0.07$). This result indicates that the Southern lineage may have been shaped by stronger purifying selection to remove deleterious mutations.

We then tested whether all mitochondrial protein-coding genes evolved under the same selective pressures. As shown in Fig. 5, the ω ratios varied among these genes, indicating that they accumulated different amounts of nonsynonymous mutations because of various selective forces. All genes but *cox1* had higher ω ratios in the Northern lineage than the Southern one. Particularly, the *nad4L* gene had the largest difference in ω ratio, which was the highest in the Northern lineage and lowest in the Southern lineage. These results demonstrated that the different selective constraints existed not only among various mitochondrial genes, but also between the Southern and Northern lineages.

Finally, the branch-site model (Zhang *et al.* 2005) was used to identify positively selected sites for the Southern and Northern lineages, respectively. When the Southern lineage was selected as the foreground lineage, the 56th amino acid residue of ATP6 was positively selected with a posterior probability of 99.7% using the BEB procedure. Specifically, 23 of the 35 Southern lineage individuals had an asparagine substituted for serine at this site. When the Northern lineage was selected as the foreground lineage, the 83rd residue of NAD2 was potentially under positive selection with a posterior probability of 94.9% using the BEB approach and 98.2% using the Naive Empirical Bayes approach (Nielsen & Yang 1998). At this site, 29 of the 30 Northern lineage individuals encoded glycine, whereas all Southern lineage individuals encoded methionine.

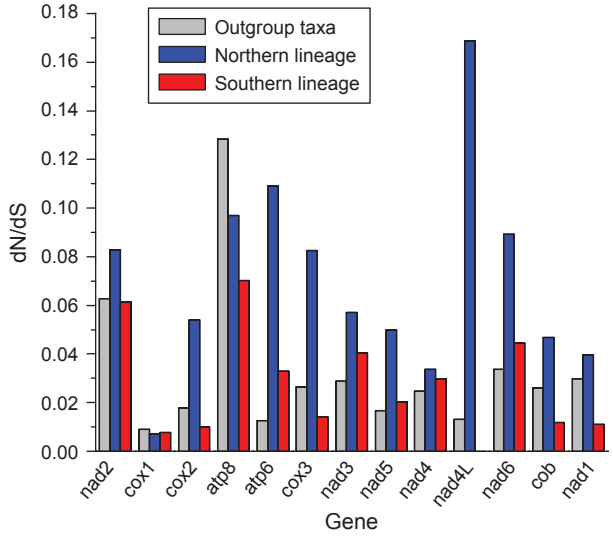


Fig. 5 Ratios of nonsynonymous to synonymous substitution rates (*dN/dS*).

Discussion

The maternal mode of mitochondrial inheritance means that mtDNA-based phylogeography reflects the historical processes only in females. If males and females of a species have different evolutionary histories, mtDNA alone cannot reveal the history of the species as a whole. Indeed, sex-biased dispersals have been reported in some insect species (Gade 2002; Lagisz *et al.* 2010). However, evidence suggests that such gender-specific dispersal may not exist in the migratory locust. Similar flight abilities have been demonstrated in males and females of the migratory locust (Liu *et al.* 2006). Consistent with this, the phylogeographic patterns we obtained using mtDNA data are overall congruent with microsatellite-based results (Zhang *et al.* 2009). Consequently, mtDNA has a strong resolving power in inferring the population phylogeographic relationships of the migratory locust.

The evolutionary history of the migratory locust

The higher genetic diversity observed among African locust populations in this study indicates a possible African origin of the migratory locust because ancestral populations are expected to possess higher genetic diversity than derived populations (Savolainen *et al.* 2002). This finding is supported by the origins of other related genera in the same subfamily Oedipodinae. Two genera, *Oedaleus* Fieber and *Gastrimargus* Saussure, are considered to be most closely related to *Locusta* based on mtgenome analysis (Ma *et al.* 2009). The genus *Oedaleus* has been suggested to originate from the Ethiopian region, and 15 of the 23 extant species in the genus *Gastrimargus* are distributed in Africa (Ritchie 1981, 1982). The genera *Humbe* Bolívar, *Oreacris* Bolívar and *Locustana* Uvarov, all closely related to *Locusta*, are endemic to the Ethiopian region (Ritchie 1981, 1982). Such a distribution pattern of these closely related genera further supports the African origin of the migratory locust.

The divergence time estimation in our study was based on a revised insect mitochondrial divergence rate of 3.54% per million years, which was obtained from

an interspecific analysis (Papadopoulou *et al.* 2010). Mutation rates at the population level have been found to be much higher than interspecific mutation rates possibly due to the delayed effect of purifying selection (Ho *et al.* 2005, 2007). This time dependency of the clock may lead to overestimation of divergence times within species, highlighting the rough nature of the divergence time estimates in our study. We estimated that the divergence time between the two lineages is 895 Kya (95% HPD: 436–1501 Kya), which is later than that (860–1890 Kya) obtained by Tokuda *et al.* (2010) using a divergence rate of 2.0–2.3% per million years (Brower 1994). Regardless of the divergence rates used, the divergence time just falls within Pleistocene (2.588 million years BP to 11 550 years BP), which is characterized by repeated glaciation events. Pleistocene glaciations are often interpreted as a major factor in shaping the biodiversity of most extant species (Hewitt 2004). Therefore, for the migratory locust, the vicariance events that initiated separation of the two lineages are likely related to Pleistocene glaciation cycles. During glacial periods, most areas in the Northern Hemisphere were not suitable for the survival of many organisms. Only a minority likely survived in favourable refuge areas and then recolonized formerly glaciated regions, resulting in population bottlenecks and a concomitant loss of genetic diversity. The relatively low genetic diversity within the Northern lineage of the migratory locust is probably a consequence of Pleistocene glaciations. The southern peninsulas of Iberia, Italy, the Balkans and the Caspian/Caucasus region are considered refuge areas for a variety of organisms (Hewitt 2004) and may have provided favourable habitats for the migratory locust during glaciation periods. The adjacent areas, such as the Black Sea, Caspian Sea, and, possibly, Aral Sea basin regions, have been speculated as refuge areas for the migratory locust despite the lack of direct genetic evidence (Zhang *et al.* 2009). To pinpoint the refuge areas of the Northern lineage, further investigation with increased sampling is required. Our results demonstrate that the locust populations from south-central China (Pop. 12–15 in Fig. 1 and Table S1, Supporting information) form a clade and share no common haplotype with other populations. These populations are probably descendants of refugees in the Hengduan Mountains region, which has provided refuge areas for many organisms (Qiu *et al.* 2011). Nevertheless, further dispersal of these populations to other regions is hampered by mountain barriers.

The recent coalescence time (32 Kya, 95% HPD: 10–60 Kya) for the eastern China populations coincides with the last glacial maximum (LGM; ~20 Kya), supporting that the eastern China region was recolonized by the locusts from refuge areas after the LGM (Zhang *et al.* 2009). The sudden demographic expansion of these populations (~5 Kya; Fig. 4A) is likely associated with warm climates and early agricultural development, which started >10 Kya (Lu *et al.* 2009). Deforestation and crop cultivation likely provided favourable habitats and food sources for the locust populations to expand rapidly. The association of demographic expansion, human population growth and agricultural development has been similarly suggested in other insects such as the stable fly (Dsouli-Aymes *et al.* 2011). The flood land expansions of the Yellow and Huaihe Rivers provided a large number of suitable habitats for the locusts after the glacial period such that they are now the most important breeding areas of the locusts in China (Ma 1965).

Based on the network analyses, we can hypothesize the following scenario for the Southern lineage. The migratory locust first dispersed from Africa to South Asia along the coasts of the Arabian Peninsula. In India, some individuals proceeded to the Tibetan Plateau, whereas others expanded to southern China and further spread to South-east Asia and Australia. Surprisingly, this dispersal route well matches the 'southern route' of modern humans (Macaulay *et al.* 2005; Mellars 2006). The divergence time estimates indicate that the colonization history of the locusts predates that of modern humans. However, if the intraspecific divergence rate of the locusts exceeds that of the interspecific rate we used, our inferred divergence times would have been overestimated, and thus, there is a possibility that locust range expansion followed human dispersal. Therefore, intraspecific calibration points are required in future studies to test whether the locust expansion patterns are shaped by human dispersals. The Madagascar clade, which is sister to all other clades in the Southern lineage, represents an early descendant of common ancestors of the locusts. We hypothesize that there was one dispersal event from the African mainland to Madagascar, but subsequent gene flow between them was restricted. Our results reveal that there are two distinct clades in the south-eastern Tibet populations with their coalescence time (265 Kya, 95% HPD: 147–405 Kya) remarkably later than the uplift of the Tibetan Plateau to its present altitude, which occurred ~8 Ma (Harrison *et al.* 1992). These two clades likely represent two separate intrusions of the locusts into the Brahmaputra River Basin of south-eastern Tibet. Another possible explanation is that there was only one intrusion into the Tibetan Plateau and that these two clades were derived from two separate refuge areas in the Tibetan Plateau (e.g. the Hengduan Mountain region) after glacial periods. If so, the two clades should be respectively monophyletic and form sister-group relationships. However, this assumption was not supported by our results, thus ruling out such a possibility.

Underlying factors maintaining the current population structure

Our analyses reveal that the locust populations are highly structured and clearly divided into two lineages corresponding to distinct climate regions. Considering the high mobility trait of the migratory locust, the strong cleavage between the two lineages is surprising because persistent gene flow in highly migratory species is expected to prevent genetic differentiation (Mayr 1963). Indeed, there is a strong west–east gene flow, as reflected by haplotype sharing between the France and northern China populations, as well as the close relationships between the Africa and Southern Asia populations. Climatic factors have been proposed to affect genetic variations among the locust populations (Chapuis *et al.* 2008). Consequently, we hypothesize that different latitudes, and especially, temperatures, may have exerted different selective pressures on the two lineages. In this study, this hypothesis is supported by different ω ratios for mitochondrial protein-coding genes between the two lineages. Furthermore, the identification of positively selected amino acid sites in the Southern (the 56th residue of ATP6) and Northern (the 83rd residue of NAD2) lineages provides direct evidence for the adaptive differentiation of the two lineages. Adaptive differences were previously assumed to maintain the current separation of the migratory locust populations in China (Zhang *et al.* 2009). Such adaptive evolution is further supported by the significantly different cold tolerance traits between the northern and southern China populations (Jing & Kang 2003; Wang & Kang 2005; Tanaka & Zhu 2008). We predict that such geographic separation of the two lineages will persist, although more secondary contact zones other than southern France and Japan will be detected when the sampling coverage is expanded in future research.

We found that geographic isolation and other environmental factors also play important roles in shaping the phylogeographic patterns of the locusts. The Madagascar population is highly divergent from the African mainland populations because of the ocean barrier. This finding is also the case for northern Japan populations, which are genetically differentiated from the north-eastern China populations. Even in Tibet, the south-eastern populations are genetically different from the northwestern populations that belong to the Northern lineage because of the complicated topography of the land. Chapuis *et al.* (2008) hypothesized that the mountain belt ranging from the Alps to Tibet represents a barrier for the Eurafrican and Eurasian locusts. Despite the long geographic distance and oceanic barrier, the locust populations from the southern Japan islands, southern China, South-east Asia and Australia have close affinity with one another. This closeness is likely attributable to the wind currents above sea level and the strong dispersal ability of the migratory locust (Tokuda *et al.* 2010). Although long-distance migration across the Atlantic Ocean was observed for the desert locust *Schistocerca gregaria* (Rosenberg & Burt 1999; Lovejoy *et al.* 2006), there is no wild population of the migratory locust in the American continent. Therefore, the isolation of the Atlantic Ocean and Pacific Ocean, as well as the low temperature of the northern Eurasian continent, limits the locust dispersal into the New World. An alternative possibility is that trans-ocean dispersal indeed succeeded, but the locusts failed to colonize the American continent because of specific ecological niches there. Altogether, a complex combination of intrinsic (the adaptive differentiation and strong dispersal ability that is further strengthened by wind) and extrinsic (physical barriers and temperature) factors contributes to the current population genetic structure of the locusts.

Taxonomic implications

For the migratory locust, 9–11 subspecies were previously described on the basis of morphometrics and geographic distribution (Uvarov 1966; COPR 1982; FAO website, http://www.fao.org/ag/locusts-CCA/common/ecg/1078/en/LMI-_Distribution_map3.pdf), but the biological relevance of these subspecific designations is still questionable. Univoltine and multivoltine populations of the locusts could be easily separated into two groups, namely the diapause and nondiapause egg populations (Farrow & Colless 1980). The desert in the Northern Hemisphere acts as a natural barrier for gene flow between geographic populations of the migratory locust. However, in eastern China, such a barrier is not apparent and the locust populations are continuously distributed from the south to the north, where the annual generation numbers of the locusts vary from four to one (Kang *et al.* 1989; Kang & Chen 1991). The proportion of diapause eggs increases from the south to the north of China and Japan (Tanaka 1994; Tanaka & Zhu 2008). Therefore, the criteria for subspecific classification based on morphometrics, distribution, generation number and diapause eggs are not clear enough to identify the subspecific status for a given locust population.

In contrast, this study uses a molecular approach with a phylogeographic and evolutionary perspective. Based on the criterion that individuals belonging to the same subspecies have lower sequence variations and should be monophyletic in the phylogenetic tree (Zink 2004), we provide robust evidence against the traditional subspecific classification. For example, previously defined subspecies, such as populations from Africa,

Arabia, southern China, India and Australia, are closely related despite the large-scale geographic ranges. In particular, two genetically distinct clades are present in south-eastern Tibetan populations, thus opposing the current taxonomic status of the subspecies *L. m. tibetansis* and *L. m. burmana*. An inconsistency between morphologically defined locust subspecies and genetic clustering was also found in previous studies (Chapuis *et al.* 2008; Zhang *et al.* 2009).

Our findings reveal two highly divergent lineages as a result of long-term allopatric evolution, and AMOVA indicates that 87.45% of total molecular variation is explained by variation between the two lineages. Logically, the two lineages represent two subspecies of the migratory locust. The Eurasian populations should be considered as the Asiatic migratory locust (*L. m. migratoria* L.), whereas populations from tropical areas of Africa, Asia and Australia should be considered as the African migratory locust (*L. m. migratorioides* Reiche & Fairmaire). Among all other previously defined subspecies (Uvarov 1966; COPR 1982; FAO website, http://www.fao.org/ag/locusts-CCA/common/ecg/1078/en/LMI-_Distribution_map3.pdf), *L. m. capito* (Saussure), *L. m. manilensis* (Meyen), *L. m. tibetensis* Chen, *L. m. burmana* Ramme, Indian subspecies, Australian subspecies and Arabian subspecies are geographic populations of the African migratory locust (*L. m. migratorioides* Reiche & Fairmaire), whereas *L. m. gallica* (Remaudiere), *L. m. remaudierei* Harz, *L. m. cinerascens* (Fabricius) and *L. m. rossica* Uvarov & Zlotarevsky are geographic populations of the Asiatic migratory locust (*L. m. migratoria* L.). Therefore, these subspecies designations require further revision and justification if they are to be accepted.

Conclusions

This study represents the most comprehensive investigation to date of the global-scale phylogeography of the migratory locust. We show that high levels of genetic differentiation exist among worldwide populations of this locust species, which features strong dispersal abilities. Most molecular variations are present between the Northern and Southern lineages, which cover distinct climate regions on the planet (temperate and tropical areas, respectively). We infer that the locust species originated from Africa and dispersed via different routes across the world. Pleistocene glaciations appear to have influenced the evolutionary history of the migratory locust, especially the Northern lineage. The adaptive differentiation of the two lineages, which is driven by different selective pressures in their respective distribution areas, counteracts the considerable dispersal ability of the locusts and likely accounts for the maintenance of the south–north cleavage pattern. Other factors, such as geographic barriers and temperature, also play important roles in maintaining the present phylogeographic patterns of the migratory locust. Our study does not support the previously defined subspecies of the migratory locust and provides a genetic basis for a phylogeography-based subspecific taxonomy. Clearly, additional molecular markers and a more extensive sampling coverage of the migratory locust will help refine the picture of its current genetic structure and evolutionary history.

Acknowledgements

We thank N. Jiang, G. Köhler, F. Li, Q. Li, G. Liang, C. Liu, S. Tanaka, M. Usmani, R. Wang, B. Zhang, D. Zhang and other collectors for offering locust specimens. We are grateful to X. Jiang, J. Li, T. Ning, M. Peng, W. Wang, J. Yu and F. Zhao for their suggestions in data analyses. This work was supported by the National Basic Research Program of China (Grant No. 2012CB114102) and the National Natural Science Foundation of China (No. 30770307).

References

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Balloux F, Handley LJ, Jombart T, Liu H, Manica A (2009) Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 3447–3455.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Berthier K, Chapuis MP, Moosavi SM, Tohidi-Esfahani D, Sword GA (2011) Nuclear insertions and heteroplasmy of mitochondrial DNA as two sources of intra-individual genomic variation in grasshoppers. *Systematic Entomology*, **36**, 285–299.

Brito P, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, **135**, 439–455.

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 6491–6495.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540–552.

Chapuis MP, Lecoq M, Michalakis Y *et al.* (2008) Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles *Molecular Ecology*, **17**, 3640–3653.

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

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Dsouli-Aymes N, Michaux J, De Stordeur E, Couloux A, Veuille M, Duvallet G (2011) Global population structure of the stable fly (*Stomoxys calcitrans*) inferred by mitochondrial and nuclear sequence data. *Infection, Genetics and Evolution*, **11**, 334–342.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 853–864.

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

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 subspeciation in the tropics and affinities of the Australian race. *Acrida*, **9**, 77–99.

Flook PK, Rowell CH, Gellissen G (1995) The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *Journal of Molecular Evolution*, **41**, 928–941.

Foote AD, Morin PA, Durban JW *et al.* (2011) Positive selection on the killer whale mitogenome. *Biology Letters*, **7**, 116–118.

Fries M, Chapco W, Contreras D (2007) A molecular phylogenetic analysis of the Oedipodinae and their intercontinental relationships. *Journal of Orthoptera Research*, **16**, 115–125.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Gade G (2002) Sexual dimorphism in the pyrgomorphid grasshopper *Phymateus morbillosus*: from wing morphometry and flight behaviour to flight physiology and endocrinology. *Physiological Entomology*, **27**, 51–57.

Gellissen G, Bradfield JY, White BN, Wyatt GR (1983) Mitochondrial DNA sequences in the nuclear genome of a locust. *Nature*, **301**, 631–634.

Gilbert MTP, Drautz DI, Lesk AM *et al.* (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8327–8332.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Harrison TM, Copeland P, Kidd WS, Yin A (1992) Raising Tibet. *Science*, **255**, 1663–1670.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183–195.

Hickerson MJ, Carstens BC, Cavender-Bares J *et al.* (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291–301.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for time dependency of molecular rate estimates. *Systematic Biology*, **56**, 515–522.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. *Nature*, **335**, 167–170.

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**, 708–713.

Jing XH, Kang L (2003) Geographical variation in egg cold hardiness: a study on the adaptation strategies of the migratory locust *Locusta migratoria* L. *Ecological Entomology*, **28**, 151–158.

Kang L, Chen YL (1991) The analysis of numerical taxonomy to the interrelationships among different geographical populations of *Locusta migrtoria* L. phase solitaria (Orthoptera, Acrididae). *Sinozoologia*, **8**, 71–82.

Kang L, Li HC, Chen YL (1989) Analysis of numerical character variations of geopgraphic populations of *Locusta migratoria* phase solitaria in China. *Acta Entomologica Sinica*, **32**, 418–425.

Lagisz M, Wolff K, Sanderson RA, Laskowski R (2010) Genetic population structure of the ground beetle, *Pterostichus oblongopunctatus*, inhabiting a fragmented and polluted landscape: evidence for sex-biased dispersal. *Journal of Insect Science*, **10**, 105.

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

Liu H, Li K, Yin J, Du G, Zhao X, Cao Y (2006) Comparative studies on the flight ability of the social type and scattered type of *Locusta migratoria manilensis*. *Plant Protection*, **33**, 34–37.

Lovejoy NR, Mullen SP, Sword GA, Chapman RF, Harrison RG (2006) Ancient trans-Atlantic flight explains locust biogeography: molecular phylogenetics of *Schistocerca*. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 767–774.

Lu H, Zhang J, Liu KB *et al.* (2009) Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 7367–7372.

Ma SJ (1965) *On the Breeding Area of the Oriental Migratory Locusts in China*. Science Press, Beijing.

Ma C, Liu C, Yang P, Kang L (2009) The complete mitochondrial genomes of two band-winged grasshoppers, *Gastrimargus marmoratus* and *Oedaleus asiaticus*. *BMC Genomics*, **10**, 156.

Macaulay V, Hill C, Achilli A *et al.* (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*, **308**, 1034–1036.

Mayr E (1963) *Animal Species and Evolution*. Belknap Press, Cambridge.

Mellars P (2006) Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science*, **313**, 796–800.

Morin PA, Archer FI, Foote AD *et al.* (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research*, **20**, 908–916.

Moulton MJ, Song HJ, Whiting MF (2010) Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources*, **10**, 615–627.

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**, 929–936.

Ning T, Xiao H, Li J, Hua S, Zhang YP (2010) Adaptive evolution of the mitochondrial ND6 gene in the domestic horse. *Genetics and Molecular Research*, **9**, 144–150.

Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, **27**, 1659–1672.

Pener MP, Simpson SJ (2009) Locust phase polyphenism: an update. *Advances in Insect Physiology*, **36**, 1–272.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Qiu YX, Fu CX, Comes HP (2011) Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution*, **59**, 225–244.

Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**, 76–86.

Ritchie J (1981) A taxonomic revision of the genus *Oedaleus* Fieber (Orthoptera: Acrididae). *Bulletin of the British Museum (Natural History) Entomology*, **42**, 83–183.

Ritchie J (1982) A taxonomic revision of the genus *Gastrimargus* Saussure (Orthoptera: Acrididae). *Bulletin of the British Museum (Natural History) Entomology*, **44**, 239–329.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Rosenberg J, Burt PJA (1999) Windborne displacements of Desert Locusts from Africa to the Caribbean and South America. *Aerobiologia*, **15**, 167–175.

Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science*, **298**, 1610–1613.

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, **33**, W686–W689.

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13486–13491.

Sorenson M, Franzosa E (2007) *TreeRot, Version 3*. Boston University, Boston, MA, USA.

Sun YB, Shen YY, Irwin DM, Zhang YP (2011) Evaluating the roles of energetic functional constraints on teleost mitochondrial-encoded protein evolution. *Molecular Biology and Evolution*, **28**, 39–44.

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

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tanaka H (1994) Geographic variation of embryonic diapause in the migratory locust, *Locusta migratoria* (Orthoptera: Acrididae), in Japan. *Japanese Journal of Entomology*, **62**, 629–639.

Tanaka S, Zhu DH (2008) Geographic variation in embryonic diapause, cold-hardiness and life cycles in the migratory locust *Locusta migratoria* (Orthoptera : Acrididae) in China. *Entomological Science*, **11**, 327–339.

Tokuda M, Tanaka S, Zhu DH (2010) Multiple origins of *Locusta migratoria* (Orthoptera: Acrididae) in the Japanese Archipelago and the presence of two major clades in the world: evidence from a molecular approach. *Biological Journal of the Linnean Society*, **99**, 570–581.

Uvarov B (1966) *Grasshoppers and Locusts*, **Vol. 1**. Cambrideg University Press, Oxford.

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

Wang XH, Kang L (2005) Differences in egg thermotolerance between tropical and temperate populations of the migratory locust *Locusta migratoria* (Orthoptera: Acridiidae). *Journal of Insect Physiology*, **51**, 1277–1285.

Yang Z (2006) *Computational Molecular Evolution*. Oxford University Press, New York.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107–1118.

Zhang DX, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, **11**, 247–251.

Zhang MZ, Kang L (2005) Genetic divergence among geographical populations of the migratory locust in China. *Science China Series C*, **48**, 551–564.

Zhang JZ, Nielsen R, Yang ZH (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, **22**, 2472–2479.

Zhang DX, Yan LN, Ji YJ, Hewitt GM, Huang ZS (2009) Unexpected relationships of substructured populations in Chinese *Locusta migratoria*. *BMC Evolutionary Biology*, **9**, 144.

Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **271**, 561–564.

C.M. is currently using molecular and evolutionary methods to study the population genetics of the migratory locust. P.Y. is interested in comparative genomics of insects. F.J. has an interest in understanding mechanisms of locust genome size expansion. M.-P.C. is conducting research on population evolution of locust species. Y.S. is conducting numerical analyses on morphometrics of locusts. G.A.S. incorporates multiple approaches to study insect ecology, evolution, genetics, and behavior. L.K.'s current research focuses mainly on ecogenomics of the

migratory locust and underlying mechanisms of locust phase transition.

Data accessibility

DNA sequences: GenBank accessions JN858148–JN858806. Sampling locations and GenBank numbers for each individual can be found in Table S1 (Supporting information). The sequence alignment of 37 mitochondrial genes and the A + T-rich region for phylogenetic reconstruction: Appendix S1 (Supporting information). Combined sequences of *cox1*, *nad5* and *cob* genes from 263 individuals for haplotype network: Appendix S2 (Supporting information). Data matrix of concatenated partial *cox1* and *cob* sequences from 388 individuals: Appendix S3 (Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 (A) Partitioned Bremer supports and (B) number of parsimonious sites for protein-coding genes and the A + T-rich region. The Partitioned Bremer support values were calculated using TreeRot.v3 (Sorenson and Franzosa 2007).

Fig. S2 Haplotype networks based on combined partial *cox1* and *cob* genes. Our gene sequences were truncated to the same length of partial *cox1* and *cob* in Tokuda *et al.* (2010), and all 388 sequences were integrated into a large data matrix for the network construction. Networks of the Northern (above) and Southern (below) lineages were separately constructed. Circle sizes are proportional to the haplotype frequencies. Small empty circles represent unsampled or extinct haplotypes. Numbers on branches indicate more than one mutation.

Table S1 Sampling information.

Table S2 Sequences of PCR primers.

Appendix S1 The sequence alignment of 37 mitochondrial genes and the A + T-rich region for phylogenetic reconstruction.

Appendix S2 Combined sequences of *cox1*, *nad5*, and *cob* genes from 263 individuals for haplotype network.

Appendix S3 Data matrix of concatenated partial *cox1* and *cob* sequences from 388 individuals.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.