CHAPTER 4

Inter-continental phylogeography and regional gene flow in the field cricket, *Gryllus bimaculatus* De Geer (Orthoptera: Gryllidae)

Abstract

I sequenced part of the mitochondrial cytochrome *b* gene to determine intraspecific patterns of genetic variation in *Gryllus bimaculatus*, within and among wild crickets of seven South African and two European populations. For South Africa, there was a large degree of within-population genetic differentiation and 30 haplotypes were identified from 133 individuals. Two ancestral haplotypes occurred in all of the South African populations. A large amount of gene flow occurred between all the South African populations, and two possible migration routes were established: (a) between the northeastern regions of the country and (b) between the southeastern regions of the country. NCA predicted that isolation by distance compounded by substantial localised gene flow was the consequence of geographical differentiation in South Africa. For South Africa and Europe, there was a large degree of between-population differentiation, with some gene flow occurring between one South African population (Queenstown) and both the European populations. Only two haplotypes were identified from 28 individuals in the European populations, of which one was a unique haplotype occurring only in Spain, and the other occurred throughout South Africa. The within-population genetic differentiation in Europe was very low, possibly reflecting the influence of the severe seasonal temperature changes on population numbers. NCA could not be performed on the African-European data set, due to the large non-sampled areas. This study concludes that gene flow in *G. bimaculatus* is not restricted to one continent, or even regions, and that inter-continental gene flow as well as regional gene flow occurs in this cricket, either by means of air, land or sea transport.

Keywords: *Gryllus bimaculatus*, Inter-continental phylogeography, Gene flow, Mitochondrial DNA, Cytochrome *b*

Introduction

Gryllus bimaculatus has a global distribution, spanning Africa, Asia and southern Europe (Harrison & Bogdanowicz, 1995; Ragge, 1972). It is an irruptive insect with large numbers appearing locally over a short time span, with low local population densities between these peaks. Moreover, *G. bimaculatus* outbreaks are unpredictable in time as well as in space, probably because of the specific habitat requirements for egg and larval development. Oviposition sites have to be warm and moist (personal observation) and immatures cannot survive cold conditions (Ferreira, 2006a, Chapter 3). Southern Europe has stronger seasonal trends and relatively severe winters imposing strong limitations on cricket population numbers, compared to the mild winters and low seasonality of Southern Africa (Martyn, 1992; Schulze, 1994). Although some parts of Asia experience dramatic changes in climatic conditions between seasons, the temperate areas preferred by *G. bimaculatus*, have a relatively low seasonality (Martyn, 1992). The ecological conditions and climatic conditions on each continent modulate the population numbers of crickets. For instance, changes in weather systems probably lead to mass migrations of *G. bimaculatus* from the West African coast towards the sea in 1969 (Ragge, 1972). During March and April 2005, a heat wave probably caused population explosions of *G. bimaculatus* in the Western Cape Province of South Africa (Breytenbach, 2005).

Genetic analyses are very valuable in understanding the spatial dynamics of this species since it can shed light on the degree of isolation between local populations, as well as among populations on different continents. Phylogeographical analyses have the potential of yielding information about present as well as historical processes that may have given rise to the present geographical distribution and patterns of irruption. By measuring intraspecific patterns of genetic variation in mitochondrial (mt) DNA it is possible to make statements on the effective sizes of populations as well as the amount of gene flow between them. Phylogeographical analyses also allow one to separate historical events from recurrent forces to explain phylogeographical patterns (Templeton *et al*., 1995; Templeton, 1998).

Intraspecific phylogeographical patterns have been studied within the context of hybridization between three closely related North American *Gryllus* species (*G. firmus, G. pennsylvanicus* and *G. ovisopis*: Broughton & Harrison, 2003; Willet *et al*., 1997). Broughton & Harrison (2003) found little geographical

structure and no correlation between genetic distance and geographical proximity for the *Ef1α*, *Cam*, and *Cyt-c* introns of the nuclear genes in *G. firmus* and *G. pennsylvanicus*. However, they argue that ancestral polymorphism have persisted through the divergence of *G. firmus*, *G. pennsylvanicus* & *G. ovisopis* and that this event rather than gene flow explains the observed deep coalescence of the nuclear loci. On the other hand, the COI-COII mtDNA gene revealed substantial intraspecific genetic structure in *G. firmus* and *G. pennsylvanicus* with both species having a north-south split (Willet *et al*., 1997). These authors suggested that, although contemporary gene flow accounted for intraspecific patterns of allele frequency variation, the distinctness of mtDNA haplotypes between northern and southern clades argued against the gene flow hypothesis, and that the biogeograhical history of this north-south divergence event remains a puzzle.

Phylogeographical analyses can give fundamental insights to understand geographical variation in morphological as well as behavioural traits of crickets. For instance, clinal variation was observed in the advertisement call of the túngara frog, *Physalaemus pustulosus,* and Ryan *et al*. (1996) proposed that call similarities between populations located close to each other were probably due to gene flow or might have originated through a common ancestor. Ferreira (2006b, Chapter 5) and Ferreira & Ferguson (2002) found significant geographical variation in the calling song as well as morphological traits of *G. bimaculatus*. In order to fully understand the underlying mechanisms contributing to this geographical variation, it is therefore necessary to quantify the degree of contemporary intraspecific gene flow as well as patterns of historical geographical differentiation for several populations of *G. bimaculatus*.

The aims of this study are:

1. To determine geographical variation in cytochrome *b* mtDNA sequences in *G. bimaculatus* from South Africa and Europe.
2. To estimate gene flow and migration rates between different geographical areas in southern Africa and Europe.
3. To infer historical biogeographical processes from the observed genetic variation.

Materials and methods

Sampling localities and genomic DNA extraction

Male crickets, *G. bimaculatus*, were sampled from 2000 to 2003 at seven localities in South Africa (Figure 4.1a; Dullstroom: 25°5'S, 30°00'E; Dundee: 28°15'S, 30°23'E; Hotazel: 27°14'S, 22°57'E; Makhado: 23°05'S, 29°59'E; Paarl: 33°14'S, 18°56'E; Pretoria 25°41'S, 28°13'E; Queenstown: 31°52'S, 26°52'E). One hundred and thirty three individuals from South Africa were sequenced. In addition, 28 crickets from two regions in Europe (Figure 4.1b; Italy and Spain) were sequenced. Crickets were stored at – 70°C after collection until DNA isolation. DNA was extracted from tarsi using 5% Chelex 100® (Walsh *et al*., 1991) and stored at – 20°C. The Italian sample comprised four males from Genova (44°24'N, 8°54'E), two from Modigliana (44°8'N, 11°46'E) and one each from Peccioli (43°32'N, 10°43'E) and Russi (44°22'N, 12°1'E). Due to the few animals from each locality in Italy, they were treated as a single Italian sample. Twenty crickets from Spain (Sevilla: 37°24'N, 6°59'W) were sequenced.

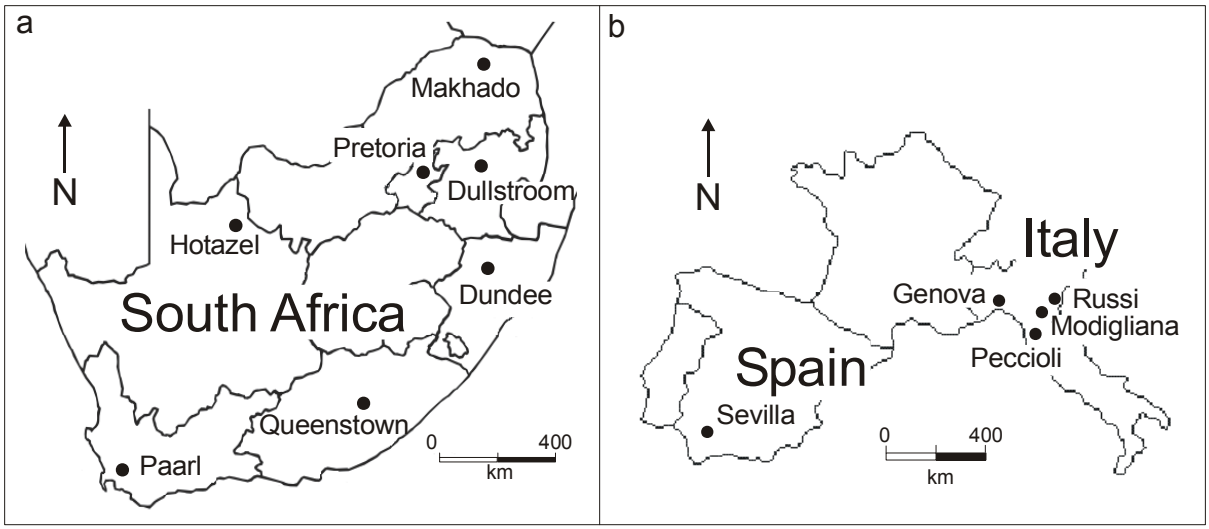

Figure 4.1. Map of sampling locations of *G. bimaculatus* in (a) South Africa and (b) Europe.

PCR amplification and sequencing of the cytochrome b region

A forward primer, labeled 197F (5'GTTGGACGTGGAATATATTA3') and complimentary reverse primer, labeled 925R (5'GCACCGATTCAAGTTAATAA3') were designed and used to amplify a 728bp fragment of the cytochrome *b* gene. Polymerase chain reaction (PCR) was performed using a GeneAmp system 2400 thermal cycler (Perkin Elmer Applied Biosystems, Norwalk, Connecticut) using a total reaction volume of 50μl. Concentrations and volumes of the reagents for each amplification reaction were as follows: 1μl of DNA template, 2.5μl of each primer (10pmol/μl), 4μl of dNTPs (2.5mM), 5μl of 10 X buffer, 1μl of *Taq* DNA polymerase (1U/μl; Biotools, Madrid, Spain) and 36.5μl of ddH_20. Thermal cycling parameters were as follows: 20 seconds preheat at 96°C, 1 cycle of initial denaturation at 95°C for 20 seconds, annealing at 50°C for 25 seconds, extension at 72°C for 60 seconds and 39 cycles of denaturation at 95°C for 15 seconds, annealing at 49.5°C for 15 seconds, extension at 72°C for 50 seconds and 1 cycle of final extension at 72°C for 60 seconds and held at 5°C. PCR negative controls containing no DNA template were included in each set of amplification reactions. After checking PCR products on a 1.5% agarose gel, stained with ethidium bromide, they were purified using the High Pure PCR Product Purification kit (Roche Diagnostics Corporation, Indianapolis, USA) following the manufacturer's specifications.

All PCR products were sequenced in both directions. Purified DNA was cycle-sequenced in 10μl reactions using 2μl of the Big Dye Terminator cycle Sequencing Ready Reaction kit (fourfold diluted; Perkin Elmer), 1μl of 5 X sequencing buffer, 1μl of primer (3.2 pmol/μl), 2 – 6μl of DNA template, 6 – 2μl of ddH_2O in a GeneAmp thermal cycler. The amount of DNA template and ddH_2O used were determined by the PCR product yield. Conditions for thermal cycling were as follows: preheat for 0.0 seconds at 96°C, followed by 25 cycles of denaturation at 96°C for 10 seconds, annealing at 54°C for 5 seconds and extension at 60°C for 4 min and held at 4°C. Sequencing was performed on ABI Prism 377 & 3100 DNA Sequencers (Perkin Elmer). Sequence chromatograms were visualized using Chromas V1.43 (Mc Carthy, 1996) and aligned and edited in Dapsa V4.9 (Harley, 2000).

Eliminating the analysis of nuclear copies of the cytochrome b gene

Nuclear copies of mtDNA (mtDNA pseudogenes) have been found in several orthopteran species namely *Schistocerca gregaria* (Zhang & Hewitt, 1996), ten grasshopper species from the families Acrididae, Podisminae, Calliptaminae, Cyrtacanthacridinae and Gomphocerinae (Bensasson *et al*., 2000), therefore all cytochrome *b* sequences in this study were checked for characteristics consistent with mtDNA pseudogenes (Arctander, 1995; Bensasson *et al*., 2001). In addition, the chromatograms of these sequences were checked for double peaks at each base pair. Finally, pure mtDNA of five individuals (two from Pretoria, one from a laboratory-reared colony and two from Sevilla) was extracted using Cesium Chloride (CsCl; Lansman *et al*., 1981). The two hind femurs and front wings were removed from each of the five individuals and stored at – 20°C, after which the remainder of the cricket was homogenized. For each individual the cytochrome *b* gene of the purified mtDNA was amplified and sequenced as described above. Chromatograms were visualized using Chromas, aligned and edited using Dapsa. These sequences were then aligned and compared to those obtained from the genomic DNA extractions (Chelex extraction of the two Sevilla males or Roche kit extraction of the two laboratory-reared males and one Pretoria male, following manufacturer's specifications).

In order to verify that it was free of nuclear DNA, the extracted pure mtDNA of each male was amplified using nuclear ribosomal primers (18_sF and 5.8_sR), yielding a 962 region of the ITS gene (Ji *et al*., 2003). It was assumed that a CsCl extraction of mtDNA contained no nuclear DNA if there was no PCR product using the ribosomal primers. Application of these primers to amplify genomic extractions (Roche kit) from these crickets yielded clear PCR products. Polymerase chain reaction was performed using a GeneAmp thermal cycler with a total reaction volume of 50μl. Concentrations and volumes of the reagents for each amplification reaction were as follows: 1μl of DNA template, 2.5μl of each primer (10pmol/μl), 4μl of dNTPs (2.5mM), 5μl of 10 X buffer, 1μl of *Taq* DNA polymerase (1U/μl; Biotools) and 36.5μl of ddH_20. Thermal cycling parameters were as follows: 4 min preheat at 94°C, 35 cycles of denaturation at 95°C for 20 seconds, annealing at 50°C for 40 seconds, extension at 72°C for 90 seconds and 1 cycle of final extension at 72°C for 20 seconds and held at 4°C. One positive control (genomic DNA template using ITS primers) and one negative control (no

DNA template using ITS primers) were included with each set of amplification reactions. PCR products were checked on a 1.5% agarose gel, stained with ethidium bromide.

Gene flow and genetic differentiation within and between populations

All the analyses were performed on two separate data sets, the first comprising both the South African and European populations (African-European data set) and the second comprising only South African populations (South African data set).

Modeltest V3.06 (Posada & Crandall, 1998) suggested that the most appropriate approach for estimating within-population genetic diversity for both data sets was the Tamura-Nei (TrN) model (Tamura & Nei, 1993). The estimated transition/transversion ratio was 11.048 for the African-European data set and 10.501 for the South African data set. Haplotype diversity (*h*) and nucleotide diversity (π) were calculated for each population using Arlequin V2.000 (Schneider *et al*., 2000).

A X^2 contingency test was performed to determine significant geographical associations using 1000 randomized permutations (Roff & Bentzen, 1989). In addition to this, a Mantel test (Mantel, 1967) was performed to test the null hypothesis of no association between the genetic distances obtained in Arlequin and the geographical distances between populations. One thousand randomized permutations were performed using Mantel V2.0 (Liedloff, 1999).

An analysis of molecular variance (AMOVA; Excoffier *et al*., 1992), using Arlequin, was performed to evaluate the degree of within-population and among population genetic differentiation for both data sets using pairwise F_{ST} statistics. This allowed estimation of the number of migrants between populations per generation. In addition, Migrate V1.7.6.1 (Beerli, 1997 – 2002) was used to calculate maximum likelihood estimates for migration rates among populations. This approach uses an expansion of coalescent theory that includes migration. The analysis on the African-European data set was inconclusive, and it was therefore decided to perform this analysis only on the South African data set. Migration was assumed to be symmetrical. One hundred short chains with 500 sampled genealogies each and three long chains with 5000 sampled genealogies each were run. One of every 20 reconstructed genealogies was sampled, therefore the total number of reconstructed genealogies was 10^4 for each short

chain and 10^5 for each long chain. Heating was set to adaptive, with temperature settings of 1.0, 1.2, 1.5 and 3.0.

Phylogeography and nested clade analysis (NCA)

TCS V1.13 (Clement *et al*., 2000) was used to estimate a single mutation haplotype genealogy (probability > 0.95) from DNA sequences for both data sets as described by Templeton *et al*. (1992). Identical sequences were collapsed to single haplotypes and haplotype frequencies were calculated for estimating haplotype outgroup probabilities, which correlate with haplotype age (Donnelly & Tavaré, 1986; Castelloe & Templeton, 1994). The haplotype network was converted into a nested series of clades using the nesting rules in Templeton *et al*. (1987) and Templeton & Sing (1993). After construction of the cladogram a NCA was performed, using the inference key of Templeton (2004) and Geodis V2.0 (Posada *et al*., 2000). This analysis attempted to discriminate between phylogeographical associations due to recurrent forces (e.g. gene flow, genetic drift and mating systems) vs. historical events (e.g. fragmentation or range expansion) at the population level (Templeton, 1998). The analysis performed on the African-European data set was inconclusive, mainly due to the large non-sampled area between South Africa and Europe. The European data were therefore excluded from this analysis. One thousand random permutations were performed to obtain statistical inferences at the 5% level.

Results

MtDNA haplotype variation

A 575 bp region of the targeted 728 bp region of the cytochrome *b* gene was used in the analyses (Table 4.1). The 161 crickets from South Africa and Europe yielded 31 haplotypes (Tables 4.2 & 4.3) defined by 35 polymorphic sites (Tables 4.1 & 4.2). Six haplotypes were found at two or more localities (h1, h2, h4, h6, h7 and h28; Table 4.3). The most frequently observed haplotype (h2, *n* = 71) was found in all the South African and European populations. The second most frequently observed haplotype (h1, *n* = 49) was found only in South Africa. Only two haplotypes were found in Europe (h2 and h31; Table 4.3). The first haplotype, namely h2, was the only haplotype found in Italy and was one of two Spanish haplotypes. The second haplotype, namely h31, was found in one individual in Spain.

Table 4.1. The 575 bp cytochrome *b* region for Pretoria male, G4. This sequence also represents haplotype 1. The 35 variable base pair positions reported on in Table 4.2 are indicated with asterisks and printed in bold

```
  *
TCATAGCTGCAGCTTTTATAGGATATGTATTACCATGAGGACAAATATCA
                              *   *    *
TTTTGAGGAGCTACTGTAATTACTAATCTTCTATCAGCAATTCCTTATTT
    *                                        *
AGGGACTGATTTAGTTCAATGAGTATGAGGAGGATTTGCAGTTGATAATG
      *                 *                    *
CCACACTAACTCGATTTTTTACATTCCATTTCATAATCCCATTTATCGTT
    *                    *          *
GCAGCATTCGTAATAATTCACTTACTTTTCCTTCACCAAACAGGATCTAA
    *                             *     *      *
CAACCCAATAGGAATTAATAGAAATCTAGATAAAATCCCATTCCATCCAT
     *               *   *                  *    *
ATTTTACTTTTAAGGACATCATAGGATTTTTAATCATACTAATATCATTA
  *
ACCATTTTATCACTTACAAATCCTTATTTATTAGGGGATCCAGATAATTT
           *                        *  *                    *
CACTCCTGCTAATCCTTTAGTTACCCCAGTTTCATATTCAACCAGAATGAT
                *       *    *    *          *
ATTTTCTTTTTGCTTTACGCCATTCTACGATCAATTCCTAACAAATTAGGA
                *                          *
GGAGTAATAGCTCTAATTGCATCTATTGCCATTCTATTTTTATTACCCTT
       *      *
TATATCCAATAATAAATTCCGAAGA
```

Table 4.2. Variable positions in the 575 bp of cytochrome *b* region defining the 31 haplotypes for *G. bimaculatus* sequenced from South Africa, Italy and Spain. h1 is the corresponding haplotype of Pretoria male G4 in Table 4.1

Haplotype	Nucleotide position
h1	c t t c g g c t c g c c c c t c t c c c c c t g t a t c c g t c c c t
h2	. . . . . t c
h3	. . . t . . . . . t c . . . . .
h4	. . . . . c . . . . .
h5	t . . . . . c . . t . . . . .
h6	 c
h7	. . . . . c . . . t . . . . . c . . . . .
h8	. . . . . t . . . . . a . . . . . t c
h9	. . . . . t . . . . .
h10	. . . . . c . . . t . . . . . c . . . . . t .
h11	. . . . c . . . . .
h12	. . . . . t . .
h13	. . . . . t . . . . .
h14	. . . . . c . . . t . t . . . . . c . . . . .
h15	. . . . . t . . . . .
h16	. . . . . t . . . . . t . . . . . t c
h17	. . c . . . . .
h18	. . . . . a . . . . . t . . . . . t c
h19	. . . . . t . . . . . t c
h20	. . . . . t . . . . .
h21	. . . . . t . . . . .
h22	. . . . . t . . . . . t . . . t c
h23	. . . . . c g . . . . c . . . .
h24	. . . . . t . . . . . t . . . . . t c
h25	. c . . . . . t . . . . . t c
h26	. . . . . c . t . . . . .
h27	a . . . . .
h28	. . . . . t c
h29	. . . . . t . t . . . . . c . . . . .
h30	. . . . . t . . . . . t . . c
h31	. . . . . a . . . . . t c

Table 4.3. The number of individuals sequenced (*n*) and the number of individuals representing each haplotype (h) are given below for each locality. The haplotype diversity (*h*) and nucleotide diversity (π) of the cytochrome *b* region of *G. bimaculatus* are also given for each locality

		Cytochrome *b* haplotype (h)																															
Locality	*n*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	*h*	π (SD)
Dullstroom	20	8	6	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0.720	0.058 (0.04)
Dundee	14	6	4	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0.714	0.059 (0.039)
Hotazel	20	9	4	0	0	1	0	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0.748	0.077 (0.048)
Makhado	20	5	7	0	0	0	1	2	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0.790	0.07 (0.044)
Paarl	20	8	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	2	0	0	0	0.755	0.056 (0.037)
Pretoria	20	10	4	0	1	0	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0.690	0.059 (0.038)
Queenstown	19	3	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0.430	0.041 (0.029)
Italy	8	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000	0.000 (0.000)
Sevilla	20	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.095	0.004 (0.007)

Testing for mtDNA pseudogenes

Firstly, all the cytochrome *b* sequences translated successfully into proteins and no stop codons were encountered. Secondly, although isolated incidences of double peaks were encountered in some of the sequences, they did not occur consistently at the same positions of both forward and reverse sequences and it is suspected that these peaks were artifacts of the quality of the sequencing reaction itself. Thirdly, when the sequences obtained through CsCl extractions were compared with those from genomic DNA extractions, not a single inconsistency was found between the two sources of DNA. Fourthly, a large degree of sequence variation was not observed, comparable to those found in the nuclear copies of mitochondrial genes in grasshoppers (Zhang & Hewitt, 1996). The ITS primers yielded a good PCR product when using genomic DNA extractions. On the other hand, the CsCl extractions did not allow amplification of the ITS gene in any of the five individuals tested, suggesting that the purified mtDNA extraction was free of nuclear DNA.

Within-population genetic differentiation

Haplotype diversity (h) and nucleotide diversity (π) differed largely among populations (Table 4.3), with the lowest haplotype and nucleotide diversity values observed in the European populations. Nucleotide diversity in the total data set was 6.09%. For South Africa, the Hotazel population had the highest nucleotide diversity and Queenstown the lowest nucleotide diversity, indicating a large degree of within-population differentiation in the Hotazel population and a small degree of within-population differentiation in the Queenstown population. The low nucleotide diversity in the Queenstown population is also reflected in its haplotype diversity. Six of the seven South African populations had haplotype diversities larger than 0.69, indicative of a high degree of overall within-population genetic diversity. The low nucleotide diversity of the European populations was also reflected in the haplotype diversities and indicated a small degree of within-population genetic differentiation. The current effective maternal population sizes (estimated using Migrate; Table 4.4) reflected the degree of within-population differentiation (based on nucleotide diversity; Table 4.3) in some of the South African populations. While the Queenstown population had the smallest effective population size and the lowest nucleotide diversity, the Dullstroom population had

the largest effective population size and its nucleotide diversity was almost similar to three other populations.

Table 4.4. The estimated migration rates per generation between South African populations of *G. bimaculatus* derived from a maximum likelihood estimation (Beerli, 1997-2002). Migration was assumed to be symmetrical. '*n*' indicates the number of individuals sequenced at each locality

		Effective population size	Migration rate (4Nm)					
Population	*n*	(θ = 4Ne mu)	Dullstroom	Dundee	Hotazel	Makhado	Paarl	Pretoria
Dullstroom	20	0.01141						
Dundee	14	0.00587	37.79690					
Hotazel	20	0.00141	38.43680	0.02190				
Makhado	20	0.00037	61.97450	16.54800	0.01250			
Paarl	20	0.00088	86.76480	0.00630	22.29770	0.21390		
Pretoria	20	0.00170	58.82090	0.00017	0.02180	0.17910	58.21410	
Queenstown	19	0.00010	10.35310	0.01490	2.10070	0.31780	0.77200	0.00002

Between-population genetic differentiation

Although there was a larger degree of between-population genetic differentiation in the African-European data set than in the South African data set, the overall patterns of gene flow between the South African populations were the same for both analyses. For the African-European data set, AMOVA indicated that 80.51% of the genetic variation occurred within populations and 19.49% among populations, revealing significant between-population genetic differentiation (F_{ST} = 0.195, P < 0.001). For the South African data set, AMOVA also revealed significant between-population genetic differentiation (F_{ST} = 0.049, P < 0.05) and indicated that 5% of the genetic variation occurred among populations and 95% of the variation occurred within populations. As expected, the estimated number of migrants per generation was higher between the South African populations than between the South African and European populations (Table 4.5). The results of the Roff-Bentzen test supported that of the AMOVA. There was significant geographical differentiation in the African-European data set (observed X^2 = 271.5, P < 0.05, n = 161) as well as within the South African data set (observed X^2 = 190.19, P < 0.05, n = 133). This was probably due to the small number of haplotypes in Europe compared with South Africa in the African-European data set, and the large number of haplotypes found in single South African populations in the South African data set (Table 4.3). Three possible migration routes were suggested within South Africa, based on the migration estimates of Migrate (Maximum log likelihood = 3.585; Table 4.4). It was assumed that a migration route existed among populations if the migration rates among these populations were higher compared with other populations. Firstly, large degrees of gene flow occurred among the populations in the northeastern region of South Africa (Dullstroom, Dundee, Makhado and Pretoria). Secondly, migration took place among the Hotazel, Paarl and Queenstown populations that are situated in the southern and western regions of the country. Lastly, a smaller degree of migration occurred between the southern and northeastern regions of the country.

Table 4.5. Population pairwise F_{ST} statistics (Tamura-Nei distance parameter) for the South African and European populations of *G. bimaculatus*, derived from AMOVA (lower diagonal area of the table). Asterisks indicate significant F_{ST} statistics ($P < 0.05$ *, $P < 0.01$ ** and $P < 0.001$ ***), while negative values indicate a larger degree of within-population variation than between-population genetic differences. The upper diagonal area of the table shows the estimated number of migrants per generation (2Nm). 'Inf' indicates an infinite number of migrants per generation among populations

	Population pairwise F_{ST}'s and number of migrants per generation								
Population	Dundee	Dullstroom	Hotazel	Italy	Makhado	Paarl	Pretoria	Queenstown	Sevilla
Dundee		749.090	Inf	0.515	22.134	Inf	48.022	1.039	0.332
Dullstroom	0.001		Inf	0.629	Inf	Inf	89.959	1.585	0.444
Hotazel	-0.015	-0.016		0.703	51.417	Inf	Inf	1.278	0.494
Italy	0.493***	0.443***	0.416***		1.218	0.444	0.559	3.421	Inf
Makhado	0.022	-0.038	0.010	0.291***		Inf	52.020	3.157	0.842
Paarl	-0.092	-0.054	-0.181	0.53***	-0.156		Inf	1.406	0.315
Pretoria	0.010	0.006	-0.025	0.472***	0.010	-0.054		0.998	0.395
Queenstown	0.325***	0.24***	0.281**	0.128	0.137	0.262*	0.333***		2.171
Sevilla	0.601***	0.53***	0.503***	0.000	0.373***	0.613***	0.559***	0.187*	

Isolation by distance

No significant isolation-by-distance effect was observed among the South African populations (Mantel test: g = 0.306, r = 0.096, $P > 0.05$, $n = 133$), suggesting that, although there was significant between-population genetic differentiation, simple isolation by geographic distance was not the major determinant of these genetic differences. A significant isolation-by-distance effect was observed for the African-European data set (Mantel test: g = 3.607, r = 0.776, $P < 0.05$, $n = 161$), suggesting that, on a broader scale, isolation by geographic distance was a determinant of between-continent genetic differentiation.

Phylogeography and nested clade analysis

Haplotype networks constructed for both spatial data sets showed large similarities for the South African populations. Since inclusion of the European data did not result in robust conclusions for the African-European data set, only the results of the South African data set will be discussed.

The haplotype network included a loop, and two alternative cladograms were constructed that comprised two 2-step clades (two-step cladogram; Figure 4.2a) and two 3-step clades (three-step cladogram; Figure 4.2b) respectively, depending on where the loop was broken.

Two-step cladogram: The cladogram consisted of 30 haplotypes, nine one-step clades and two 2-step clades. Clades 2-1 & 2-2 included haplotypes from all the South African populations. Clade 2-1 mostly comprised haplotypes from the southern region of South Africa (Hotazel, Paarl and Queenstown), while clade 2-2 mostly comprised haplotypes from the northeastern region (Dullstroom, Dundee, Makhado and Pretoria).

Three-step cladogram: This cladogram differed from the 2-step cladogram by having three 2-step clades and two 3-step clades. Although clade 1-3 was now included in clade 3-1, this had a minor influence on the spatial representation of the populations. Most of the populations from the southern region of South Africa were nested as clade 3-1 and the populations from the northeastern region nested as clade 3-2.

The nested analysis showed significant differences for within-clade (*Dc*) and nested clade (*Dn*) distances (Appendix 1) for both cladograms.

Two-step cladogram: Significant geographical associations were only observed at the highest level (X^2 = 16.23, P = 0.01). Historic events rather than recurrent forces were the most likely explanation for the patterns in the clades. Restricted gene flow with isolation by distance was the most likely explanation for the pattern in clade 1-1. Clade 1-1 included individuals from all the South African populations (the furthest distance between two populations being 1770 km and closest distance being 220 km; Figure 4.1a), suggesting that, from a spatial distribution point of view, an isolation-by-distance effect be the best explanation for the pattern of gene flow. Within clade 1-1, the interior haplotype (assumed to be the oldest haplotype) represented all of the South African populations. The majority of the Queenstown individuals (n = 14) were represented by this interior haplotype. The tip haplotypes represented all of the South African populations, except the Dundee population, suggesting that although there were old haplotypes present in the populations, derived haplotypes evolved throughout most of the geographical distribution (Figure 4.2a). Restricted gene flow with isolation by distance was the most likely explanation for the pattern in clade 2-2. Clade 2-2 also comprised individuals from all of the South African populations and the same reasoning as for clade 1-1 would probably account for the pattern of gene flow in this clade, with the exception that the tip clades in clade 2-2 represented all of the South African populations. Clade 1-9 also had significant genetic structure but the sampling design was inadequate to discriminate between isolation by distance versus long distance dispersal. The geographical sampling was inadequate to explain the pattern in clade 1-2.

Three-step cladogram: Significant geographical associations were observed at the total cladogram level (X^2 = 12.21, P < 0.05). Clades 1-1, 1-2 & 1-9 comprised the same haplotypes as in the two-step cladogram, therefore yielding the same inferences. Inferences for clades 2-2 & 3-1 were inconclusive. Restricted gene flow with isolation by distance was the most likely explanation for the pattern in clade 2-3. The oldest interior clade (1-9) and the younger tip clades comprised haplotypes from all the South African populations. However, with the exception of clade 1-8, the geographical representation of the tip clades were restricted to certain regions of the country (clades 1-4 & 1-5 represented the northeastern region) indicating restricted gene flow in certain areas of the country.

Both cladograms indicated that, due to restricted gene flow between the South African populations, gene flow mainly occurred within two regions of the

country namely the northeastern (represented by clade 1-9) and southeastern (represented by clade 1-1) regions.

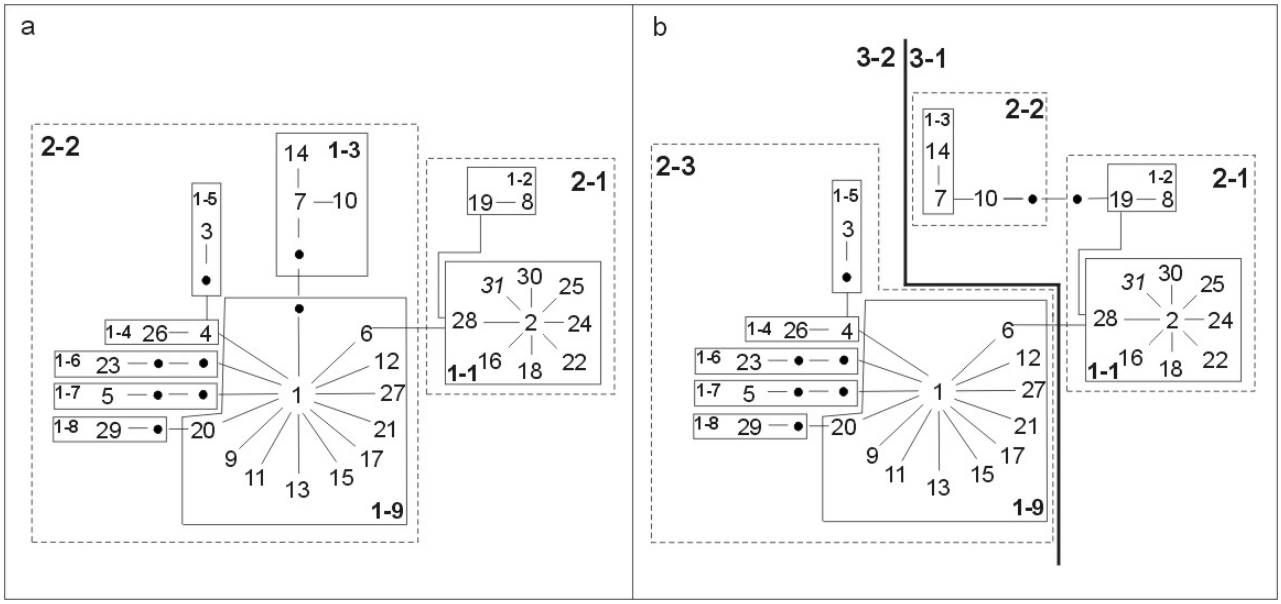

Figure 4.2. The cladogram and associated nested design for cytochrome *b* haplotypes from seven South African populations of *G*. *bimaculatus* based on the statistical parsimony of Templeton *et al*. (1987) and Templeton & Sing (1993). Haplotypes from Italy and Spain were excluded form the NCA due to large non-sampled areas between South Africa and Europe. The unique haplotype (h31) from the Spanish population is indicated in italics on the network (clade 1-1) and is only intended for explanatory purposes. Separate analyses performed to accommodate an internal loop in the haplotype network resulted in (a) a 2-step cladogram and (b) a 3-step cladogram. Haplotype numbers do not contain the prefix ‘h’ and are separated by one mutational step. Hypothetical haplotypes are indicated by a black dot. Numbers in bold indicate clade numbers and thin line boxes enclose one-step clades, dotted line boxes enclose two-step clades and the thick line indicates three-step clades for the nested clade analysis.

Discussion

Nuclear copies of the mtDNA have been identified in several orthopteran taxa, resulting in highly variable DNA sequences of particular genes (Bensasson *et al*., 2000; Zhang & Hewitt, 1996). Purified mtDNA sequences of five individuals matched their genomic DNA sequences, indicating that the results in this study reflected variation at the mitochondrial cytochrome *b* locus of *G. bimaculatus*, and that no nuclear products were amplified with the cytochrome *b* primers used in this study.

Genetic differentiation within South Africa and Europe

The degree of within-population genetic variation was much higher in the South African populations than in the European populations (Table 4.5). In Europe, severe regional seasonal changes in climate probably caused irruptive cricket populations in the warmer summer months with only a small portion of the population surviving through the colder winter months, consequently, reducing genetic variation (Table 4.5). For Italy and Sevilla, the coldest winter month had a 24-hour mean temperature of 1.3°C at Bologna (period 1808 – 1990), 8.7°C at Genova (period 1981 – 1990) and 10.7°C at Sevilla (period 1951 – 1990), while the hottest summer month had a 24-hour mean temperature of 25°C at Bologna (period 1808 – 1990), 24.6°C at Genova (period 1981 – 1990) and 26.7°C at Sevilla (period 1951 – 1990) (www.worldclimate.com). In addition, one would expect crickets in Europe to have fewer generations per year than crickets in South Africa, due to the less favourable climatic conditions in Europe (www.worldclimate.com). This could in turn contribute to the reduced genetic variation in the European populations compared to the larger degree of genetic variation in the South African populations (Table 4.5). On the other hand, South African populations experience less severe regional seasonal and annual climate changes (Schulze, 2001) that probably ensure a high survival rate through the winter months, resulting in higher haplotype and nucleotide diversities compared to the European populations. The coldest winter month had a 24-hour mean temperature of 12.4°C at Paarl (period 1994 – 2002), 11.2°C at Pretoria (period 1960 – 1991) and 10.5°C at Queenstown (period 1940 – 1991), while the hottest summer month had a 24-hour mean temperature of 25.4°C at Paarl (period 1994 – 2002), 22°C at both Pretoria (period 1960 – 1991) and Queenstown (period

1940 – 1991) (www.worldclimate.com). Although regional climate conditions potentially influenced the population dynamics, other factors such as food availability and rainfall probably also influenced population numbers, and therefore genetic diversity in this cricket.

Gene flow and migration rates among the South African and European populations

For the South-African data set, a large amount of gene flow occurred among populations, with the exception of Queenstown where only a small amount of gene flow occurred between Queenstown and the other South African populations (Table 4.5). Both the maximum likelihood estimation and the NCA revealed two similar general migration routes within South Africa (Table 4.4; Appendix 1). Firstly, migration occurred among the populations situated in the northeastern regions (Appendix 1) and among the populations situated in the southeastern regions (Appendix 1), corresponding largely to different climatic regions (Preston-Whyte & Tyson, 1988; Schulze, 1994). While the northeastern region of South Africa is categorised as a summer rainfall area with a high annual rainfall, the southern and western regions are categorised as a winter rainfall area with drier conditions (Preston-Whyte & Tyson, 1988; Schulze, 1994). Secondly, migration occurred along a North-South gradient among the northeastern and southeastern regions. One of the major inland road transport routes in South Africa follows a southwest to northeast direction (Brett & Mountain, 1997) and could incidentally be a major cause of the intraspecific gene flow within South Africa. Indeed, it is known that *G. bimaculatus* were accidentally transported from the Western Cape province to the Gauteng province in South Africa by road, covering a distance of more than 1000km (Spinman, 2005). For the African-European data set, there was no between-population genetic differentiation among the Italian and Spanish samples due to the large estimated number of migrants that migrate between these countries (Table 4.5). Although this number of migrants is based on the AMOVA's F_{ST} estimates that have been criticized by Whitlock & McCauley (1999), migration rates could not be obtained with Migrate since this study only included two populations from Europe and there are large unsampled areas between South Africa and Europe. In addition, more extensive sampling is required for a better understanding of cricket movements in Europe. As expected, the South African populations were significantly different from the European populations, except for

the Queenstown population (Table 4.5). Transport routes connecting South Africa and Europe through air and sea could increase the chances of inter-continental gene flow in *G. bimaculatus*, thereby maintaining the genetic diversity in this species. Although it is rare for these insects to fly long distances, swarms of *G. bimaculatus* were reported to land on ships passing the West coast of Africa on route from the Canary Islands to South Africa, the furthest ship being 933 km from land (Ragge, 1972).

Demographic processes explaining the genetic differentiation within southern Africa

The significant between-population genetic differentiation in South Africa was reflected by the large number of haplotypes from single localities (26 of 31 haplotypes). This is similar to the results of Willett *et al*. (1997) who found that, for the COI-COII mtDNA gene in *G. firmus* and *G. pennsylvanicus*, all except one of the mtDNA haplotypes were restricted to single localities. The Mantel test (correlating genetic distances with linear geographic distances) indicated that a simple model of isolation by distance was not suitable for explaining the geographical differentiation within South Africa. A more complex model is therefore required. Following the NCA, restricted gene flow with isolation by distance was the most likely explanation for the gene flow patterns for genetically differentiated populations from South Africa (Appendix 1). The inference of large-scale gene flow between some of the populations is supported by the migration estimates from AMOVA and maximum likelihood estimates (Tables 4.4 & 4.5). Geographical genetic differentiation could therefore be brought about by a combination of isolation by distance overlaid by unequal gene flow between geographical areas.

Maintenance of genetic variation in G. bimaculatus

Europe experiences severe seasonal changes that probably lead to large fluctuations in population densities, and could in turn, give rise to the small degrees of within-population genetic variation in the Italian and Spanish populations. One South African population, namely Queenstown, also has a small degree of within-population genetic variation. Crickets were not always abundant at Queenstown, and it is therefore possible that large fluctuations in this population's density might cause the small degree of within-population genetic

variation in Queenstown. On the contrary, there are large degrees of within-population genetic variation within the other South African populations. This suggests that a large gene pool of *G. bimaculatus* is potentially maintained in this region, and that the large amounts of gene flow that occur between populations probably maintain the genetic diversity across the geographical range of this species. It is also suggested that between-population variation in the phenotypic traits of *G. bimaculatus* reflect the between-population genetic variation among South African and European populations.

Acknowledgements

I would like to thank the farmers from Dullstroom, Dundee, Makhado, Paarl and Queenstown for allowing me to catch crickets on their properties. Luke Verburgt is thanked for collecting crickets in Europe. Wayne Delport, Glen Malherbe and Jenny Edrich are thanked for their laboratory and statistical assistance. The National Research Foundation (NRF) and University of Pretoria funded this project.

References

Arctander, P. 1995. Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proceedings of the Royal Society of London Series B* **262**: 13-19.

Beerli, P. 1997 – 2002. MIGRATE: documentation and program, part of LAMARC. V1.5. Revised august 7, 2002. Distributed over the Internet, http://evolution.genetics.washington.edu/lamarc.html. [Downloaded: 10 May 2004].

Bensasson, D., Zhang, D-X & Hewitt, G. M. 2000. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Molecular Biology and Evolution* **17**: 406-415.

Bensasson, D., Zhang, D-X, Hartl, D. L. & Hewitt, G. M. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16**: 314-321.

Brett, M. & Mountain, A. 1997. *Touring atlas of southern Africa*. Newlands, A & LeMaitre, A (Eds.). Struik Publishers (Pty) Ltd, Cape Town.

Breytenbach, K. 2005. Population of crickets explodes in heat, drought. In: *Cape Times*, Cape Town.

Broughton, R. E. & Harrison, R. G. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics* **163**: 1389-1401.

Castelloe, J. & Templeton, A. R. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution* **3**: 102-113.

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

Donnelly, P. & Tavarè, S. 1986. The ages of alleles and a coalescent. *Advances in Applied Probability* **18**: 1-19.

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

Ferreira, M. 2006a. High levels of environmentally-induced variation in the calling song of the field cricket, *Gryllus bimaculatus* De Geer (Orthoptera: Gryllidae). Chapter 3, MSc thesis, University of Pretoria, Pretoria, South Africa.

Ferreira, M. 2006b. Environmentally-induced geographical variation in the calling song of the field cricket, *Gryllus bimaculatus* De Geer (Orthoptera: Gryllidae). Chapter 5, MSc thesis, University of Pretoria, Pretoria, South Africa.

Ferreira, M. & Ferguson, J. W. H. 2002. Geographic variation in the calling song of the field cricket *Gryllus bimaculatus* (Orthoptera: Gryllidae) and its relevance to mate recognition and mate choice. *Journal of Zoology London* **257**: 163-170.

Harley, E. H. 2000. Dapsa: A program for DNA and protein sequence analysis, Version 4.9. Department of chemical pathology, University of Cape Town.

Harrison, R. G. & Bogdanowicz, S. M. 1995. Mitochondrial DNA phylogeny of North American field crickets: perspectives on the evolution of life cycles, songs, and habitat associations. *Journal of evolutionary biology* **8**: 209-232.

Ji, Y-J., Zhang, D-X. & He, L-J. 2003. Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer regions in insects and other invertebrates. *Molecular Ecology Notes* **3**: 581-585.

Lansman, R. A., Shade, R. O., Shapira, J. F. & Avise, J. C. 1981. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. III. Techniques and potential applications. *Journal of Molecular Evolution* **17**: 214-226.

Liedloff, A. 1999. Mantel nonparametric test calculator for Windows, Version 2.0. School of Natural Resource Sciences, Queensland University of Technology.

Mantel, N. 1967. The detection of disease clustering and a generalised regression approach. *Cancer research* **27**: 209-220.

Martyn, D. 1992. *Developments in atmospheric science, 18: Climates of the world*. Polish Scientific Publishers PWN Ltd., Warszawa.

Mc Carthy, C. 1996. Chromas, Version 1.43. School of Biomolecular and Biomedical Science, Griffith University.

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

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

Preston-Whyte, R. A. & Tyson, P. D. 1988. Atmospheric circulation and weather over southern Africa. In: *The atmosphere and weather of Southern Africa*, pp207-249. Oxford University Press, Cape Town.

Ragge, D. R. 1972. An unusual case of mass migration by flight in *Gryllus bimaculatus* DeGeer (Orthoptera Gryllidae). *Bulletin de l'Institut fondamental d'Afrique noire. Series A.* **34**:869-878.

Roff, D. A. & Bentzen, P. 1989. The statistical analysis of mitochondrial DNA polymorphisms: X^2 and the problem of small samples. *Molecular Biology and Evolution* **6**: 539-545.

Ryan, M. J., Rand, A. S. & Weigt, L. A. 1996. Allozyme and advertisement call variation in the túngara frog, *Physalaemus pustulosus*. *Evolution* **50**: 2435-2447.

Schneider, S., Roessli, D. & Excoffier, L. 2000. Arlequin, Version 2.000. A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schulze, B. R. 1994. *Climate of South Africa*. Part 8: General survey. Weather bureau, Department of Environment affairs, South Africa.

Spinman, T. H. K. 2005. http://www.spinman.co.za/default.asp?id=11581&des=conte nt&scat=supercycling/mountainbiking&cl=yes.

Tamura, K. & Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution* **10**: 512-526.

Templeton, A. R. 1998. Species and speciation: geography, population structure, ecology, and gene trees. In: *Endless forms: species and speciation*, pp32-43. Howard, D. J. & Berlocher, S. H. (Eds.). Oxford University Press, Oxford.

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

Templeton, A. R., Boerwinkle, E. & Sing, C. F. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117**: 343-351.

Templeton, A. R., Crandall, K. A. & Sing, C. F. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619-633.

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

Templeton, A. R. & Sing, C. F. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134**: 659-669.

Walsh, P. S., Metzger, D. A. & Higuchi, R. 1991. Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* **10**: 506-513.

Whitlock, M. C. & McCauley, D. E. 1999. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity* **82**: 117-125.

Willett, C. S., Ford, M. J. & Harrison, R. G. 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. *Heredity* **79**: 484-494.

Zhang, D-X, & Hewitt, G. M. 1996. Highly conserved nuclear copies of the mitochondrial control region in the desert locust *Schistocerca gregaria*: some implications for population studies. *Molecular Ecology* **5**: 295-300.