# Chapter 4

## The effect of large annual population size fluctuations on spatial genetic pattern in the continuously distributed African Wild Silk Moth (*Gonometa postica*)

"Genomes can be regarded as genetic archives that contain information about the past history of changes in demography and of natural selection."

Veuille & Slatkin 2002

# The effect of large annual population size fluctuations on spatial genetic pattern in the continuously distributed African Wild Silk Moth (*Gonometa postica*)

Wayne Delport[1], Paulette Bloomer[1] & J. Willem H. Ferguson[2]

[1]Molecular Ecology and Evolution Programme, Department of Genetics, University of Pretoria, Pretoria, 0002, South Africa

[2]Centre for Environmental Studies, Department of Zoology and Entomology, University of Pretoria, Pretoria, 0002, South Africa

## Abstract

Several insect species exhibit large inter-annual fluctuations in population size, yet the effect of such population size fluctuations on the inference of demographic parameters from population genetic data is not well understood. Although some theoretical models suggest that population size fluctuations result in an increase in spatial genetic structure as a result of local genetic drift, this is not consistent with observed spatial genetic patterns in many cyclical species. The African Wild Silk Moth (*Gonometa postica*) is a species that exhibits large inter-annual population size fluctuations in the Kalahari region of southern Africa. We used morphological estimates of dispersal ability, mtDNA sequencing, microsatellite genotyping and simulation modeling to determine the connectivity of populations in this species. Our results indicate high levels of gene flow as found in other cyclical species. Our simulation results confirm the occurrence of increased spatial genetic pattern at low dispersal distances, yet we cannot demonstrate the conditions under which gross overestimates of dispersal are obtained. Rather we observe that low levels of dispersal, less than 1% of the species total range, are sufficient to prevent the genetic drift effects of population size fluctuations in continuously distributed species. The implications of these results in interpreting spatial genetic patterns from cyclical species are discussed.

**Keywords**: population cycles, continuously distributed, isolation by distance, simulations

**Introduction**

Since its advent, the principle aim in the field of phylogeography has been the inference of population processes from spatially distributed genetic data (Avise *et al.* 1987). Such inferences are typically made from (i) reconstructed gene genealogies (Bermingham & Avise 1985, Avise *et al.* 1987), (ii) statistical tests of population structure and partitioning of genetic variance (Weir & Cockerham 1984, Excoffier *et al.* 1992, Raymond & Rousset 1995) or (iii) statistical modeling of hypotheses and inference of parameters (Kuhner *et al.* 1998, Beerli & Felsenstein 1999, 2001, Nielsen & Wakeley 2001, Knowles & Maddison 2002, Knowles 2004). The latter generally incorporates the observation that gene sorting is stochastic (Kingman 1982, Hudson 1990), and thus accounts for the substantial variability possible in the observed genetic data under identical demographic processes. These backward-through-time coalescent models provide a computationally tractable method for the estimation of population parameters from genetic data, given their assumptions. Assumptions typically include the dispersal model (island model: Wright 1931, lattice model: Malecot 1951), substitution model, neutrality of gene regions considered, and the underlying demographic model. The suitability of these analytical techniques for a species of interest is dependent on these assumptions. Several demographic models have been incorporated into coalescent methods, where one can estimate (i) gene flow from a structured population (Beerli & Felsenstein 1999, 2001), (ii) continuous population growth or decline in an unstructured or structured population (Kuhner *et al.* 1998, Kuhner *et al.* 2004) (iii) the divergence time of two populations that exchange/d migrants (Nielsen & Wakeley 2001) and (iv) immigration rates given a spatial range expansion (Ray *et al.* 2003, Exoffier 2004). Violation of the assumptions of these models, however, typically leads to discrepancies in the indirect estimation of gene flow from genetic data (Hastings & Harrison 1994, Leblois *et al.* 2003). In addition to these maximum-likelihood computational approaches, simulation approaches have been used to gain an understanding of the effect of temporal changes in dispersal and density on the inference of demographic parameters in continuous populations (Leblois *et al.* 2004). These computational and simulation approaches have, however, thus far only considered demographic processes that occur over extended genealogical periods. Several insect species that have short generation times exhibit large inter-annual population size fluctuations (Bowers *et al.* 1993, Ginzburg & Taneyhill 1994, Turchin 2003), determined by the interaction between exogenous and endogenous factors (Turchin 2003). In particular, the continuously distributed African Wild Silk Moth (*Gonometa postica*) exhibits large inter-annual population size fluctuations (Veldtman 2004, Delport 2005 Chapter 2). Yet the effect of such complex demographics on the population genetic inference of demographic parameters such as migration is uncertain.

*G. postica* is currently of economic interest in southern Africa since both this species and its sister species, *G. rufobrunnea*, have been shown to possess a silk fibre of exceptional quality (Freddi *et al*. 1993). To this end the initiation of an African Wild Silk industry has been proposed as a potential means of poverty alleviation in rural southern African communities. Attempts have already been made at initiating such an industry, yet a consistent complaint is the lack of a continuous local supply of cocoons, a problem that has already resulted in the demise of Shase Silk Pty (Ltd) in northwest Botswana. Since silk is processed after emergence of the adult moth, this problem is one of cyclical population dynamics and not harvesting impact. The life cycle of *G. postica* is characterized by two generations per annum, the first starting in September-October with the emergence of adult moths. Moths have no feeding mouthparts and only survive for three-five days (Hartland-Rowe 1992), during which breeding occurs. Eggs are laid, larvae emerge and pass through six instar larval stages in approximately five weeks, after which larvae pupate and enter diapause (Hartland-Rowe 1992). Diapause is interrupted in either February or September of the following year. Approximately 12-50% of the first generation emerge as adults in February and proceed with an additional population cycle. The impact of this variable second generation on the incidence of eruptions is, however, uncertain. In addition to the temporal cyclical nature of the species, abundance varies spatially within a season (Delport 2005 Chapter 2). It is clear that thorough population dynamics research is required to address eruptions in this species. Thus, a principle aim of the current research was to determine the dispersal ability *G. postica* using indirect morphological and molecular methods.

Preliminary analysis of population genetic data from *G. postica*, however, indicated that the inference of spatial genetic pattern might be problematic (Delport 2005 Chapter 3). This was due in part to the low levels of microsatellite allelic variation in the species (Delport 2005 Chapter 3), and the potential for a complex demographic history to disrupt the spatial genetic pattern. Complex demographic effects and their influence on demographic inferences from population genetic data in continuous populations are seldom evaluated (Leblois *et al.* 2004). Leblois *et al.* (2004) have shown that temporal and spatial fluctuations of demographic parameters (density, dispersal and population size) have little effect on the inference of neighbourhood size using Rousset's (2000) spatial regression method. Yet a temporal increase in density, even a relatively ancient flush of low magnitude, results in a significant overestimate of neighbourhood size (Leblois *et al.* 2004). In addition, the effect of demographic changes on the inference of population parameters is largely dependent on the timing of demographic events, where an event more than twenty generations in the past is considered to have little effect (Leblois *et al.* 2004). *G. postica*, however, exhibits annual

population size fluctuations of large magnitude, potentially two to three orders (Veldtman 2004). Neighbourhood size has been proposed as a suitable statistic for inferring levels of gene flow in continuously distributed species (Rousset 2000), yet the effect of annual population size fluctuations on its inference is unknown. Theoretically, the degree of dispersal is likely to determine the effects of population size fluctuation on estimates of neighbourhood size. When dispersal is low, population size fluctuations could generate population genetic structure between demes, the result of local genetic drift (Wright 1940). However, when dispersal is high, size fluctuations could have a homogenizing effect on spatial genetic pattern as a result of global genetic drift and high connectivity between populations. Inferring the position of a species along this continuum using genetic data is a question that has not yet been explored, the result of which would be instrumental in our understanding of the effects of complex demographic processes. Therefore, in this manuscript we evaluate the dispersal ability of *G. postica* using indirect morphological and molecular methods and interpret these results in light of the effects of population size fluctuations on spatial genetic pattern, as derived with simulation modeling. Our results show a very weak signal of spatial genetic structure in *G. postica*. Furthermore, simulation results suggest that low levels of dispersal in continuously distributed species are sufficient to prevent an increase in population genetic structure as a result of population size fluctuations. These results are discussed in relation to observed spatial genetic patterns detected in other cyclical species.

## Materials and methods

### *Study site and morphological data analysis*

*G. postica* is an eruptive species that occurs in the Kalahari region of southern Africa (Figure 1). Although the species has been recorded in other areas, east and south of the distribution depicted, the Northern Cape province of South Africa and Kalahari region of southern Africa are considered the core region (Veldtman *et al.* 2002), where the species is regularly recorded in large numbers. We collected pupae (cocoons) from trees during the latent phase of their development in June 2002. Adult moths were then allowed to emerge in the laboratory and were frozen at -20°C for future genetic analyses. We estimated an indirect measure of dispersal-ability based on morphology. Wing load (i.e. the ratio of body mass to total wing area) has previously been used as a proxy for estimating dispersal or flying ability of Lepidoptera (Casey & Joos 1983). We pinned and photographed one fore- and one hindwing from each of 153 individuals (57 males, 96 females) at 2 mega pixel resolution with a FujiFilm FinePix S304 digital camera at constant height. Eight calibration photographs were taken at random intervals during the course of digitizing wings. Total wet body mass was measured, using a Scaltec Sac51 digital scale, before removal of wings. Dry

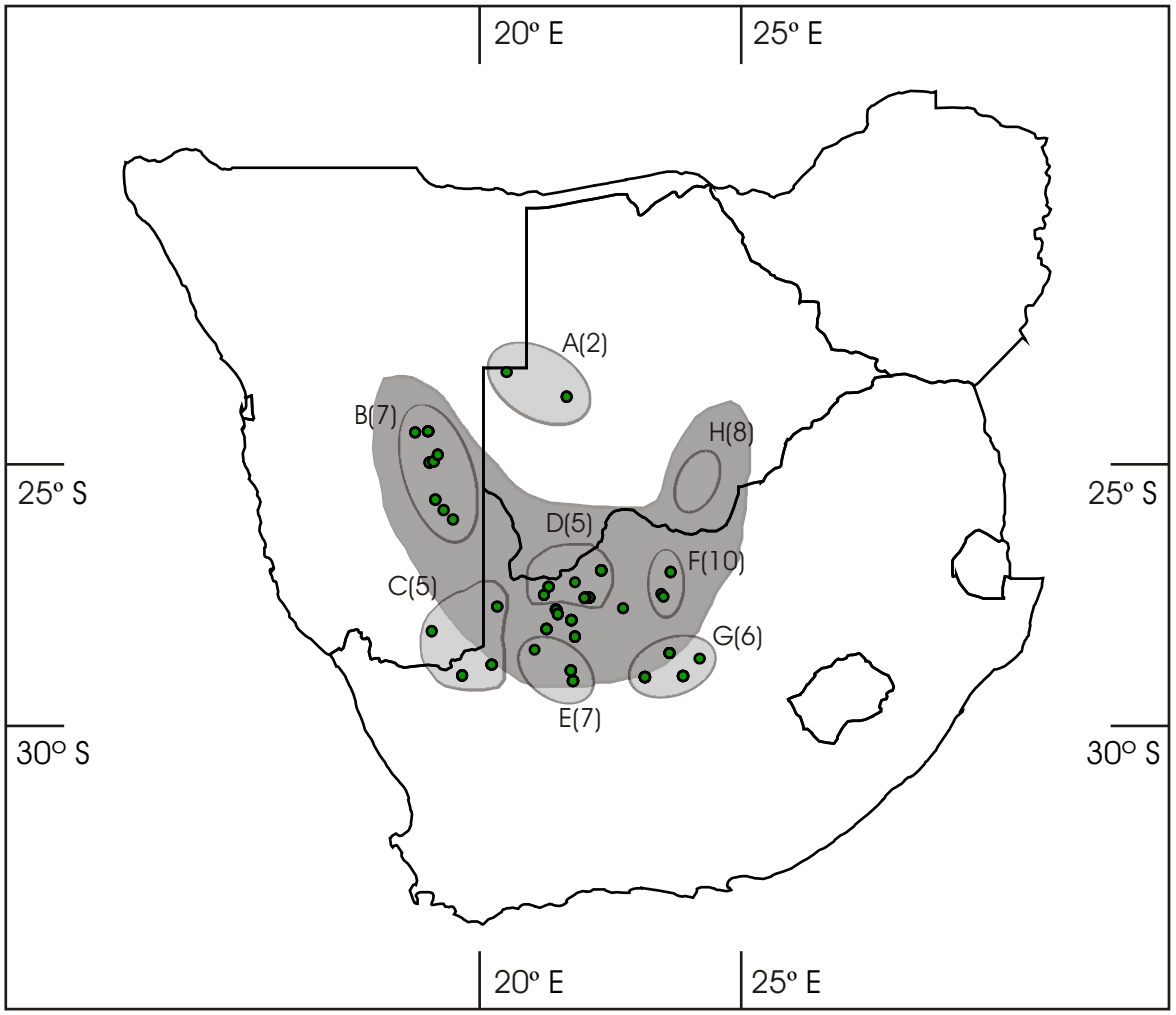



**Figure 1**: Distribution of *Gonometa postica* in southern Africa and sampling localities for mtDNA and microsatellite samples. The distribution of eruptive populations is depicted as a large gray polygon, whereas small light gray polygons depict locations of sampling sites for mtDNA analysis (with sample sizes in parentheses). Microsatellite sampling localities are represented by small filled circles.

mass was not determined since the moths were frozen at -20ºC for subsequent DNA extraction and analyses. The sexes of the moths were easily scored due to the large sexual size dimorphism evident in the species (Veldtman *et al.* 2002). Total wing area was measured from digitized images. Images (JPEG format) were first converted to grayscale images in ImageJ (Rasband 1997) and wing areas were measured with the particle analyze tool (where all particles greater than 10000 pixels were analyzed). Scale (in mm) was set from calibration images.

*Genetic sampling and laboratory methods*

Total genomic DNA was extracted from 137 individuals collected from 23 localities (Figure 1, mean 4.9 individuals per locality) using a Qiagen Dneasy extraction kit. Partial fragments of the mitochondrial DNA control region were PCR-amplified, for 50 individuals collected across the distribution range (Figure 1), in a total volume of 50µl in thin-walled microcentrifuge tubes. The reaction mix contained approximately 50ng genomic DNA, 2mM $MgCl_2$, 1x reaction buffer, 0.2mM of each of four nucleotides, 1.5U Super-therm DNA polymerase (Southern Cross Biotechnology) and 12.5 picamol of each primer (12S-332+ Taylor *et al.* 1993: 5' TAGGGTATCTAATCCTAGTT 3'; *Bombyx mori* position 9582, tRNA-ile-r: 5' ATCAGAATAATCCTTTATTCAGGC 3'; *B. mori* position 10475). A Geneamp PCR system 9700 (Applied Biosystems) was used to cycle the reaction mix through the conditions: 94°C for 2min; 35 cycles of 94°C for 30 seconds, 59°C for 30 seconds and 72°C for 45 seconds; followed by an extended elongation at 72°C for 10 minutes. PCR products were ethanol precipitated and cycle sequenced using BigDye (Applied Biosystems) and electrophoresed on an ABI3100 capillary sequencer (Applied Biosystems), yielding a 250bp fragment of the control region. Difficulties with sequencing through the AT-rich repeat region, characteristic of the insect mtDNA control region, prevented the sequencing of the full 487bp PCR product. Mitochondrial DNA sequences are available in Genbank (Accession numbers: DQ020593-DQ020611). In addition to DNA sequencing we genotyped 137 individuals for six polymorphic microsatellite loci (Gon6, Gon60, Gon65, Gon55, Gon107, Gon120). Details of protocols for microsatellite development and amplification conditions are detailed in Delport (2005 Chapter 3).

*Statistical genetics methods*

Before conducting advanced statistical inference we first tested for neutrality of mutations, in the mtDNA data, using Tajima's (1989) D and Fu & Li's (1993) $F_s$ test statistics. These statistics are based on the premise that (i) in a gene under selection, $\theta$ estimated from segregating sites will be substantially greater than $\theta$ estimated from nucleotide diversity, since rare mutations selected

against are down-weighted in the calculation of the latter (Tajima's *D*), or (ii) purifying selection is evident as an excess of mutations at the tips of a genealogy, versus those occurring on internal branches of the genealogy (Fu & Li's $F_s$). Failure to reject the null hypothesis of neutrality is evidence for the absence of selection in the gene considered; yet the rejection of neutrality could be the result of either selection or the demographic process of rapid population growth following a bottleneck (Tajima 1989). Mitochondrial DNA sequence data were used to draw an allele network, a hypothesis of the genealogical relationships among alleles, using TCS (Clement *et al.* 2000). Before making inferences from the allele network, we used a process of eliminating homoplasy, where older alleles of higher frequency have greater probability of being interior nodes (Posada & Crandall 2001), and where similar alleles are more likely to be derived from adjacent or closely situated localities. A structured population is not consistent with the isolation by distance model most appropriate for this species, yet due to insufficient mtDNA sample sizes we grouped localities based on geographic proximity and the temporal persistence of eruptions (Delport 2005 Chapter 2, Figure 1). *F*-statistics, based on pairwise differences, were calculated using Arlequin (Schneider *et al.* 2000). Finally, we used the mtDNA data to test for the existence of population structure with 1000 permutations of the chi-square test (Roff & Bentzen 1989). Microsatellite data were evaluated for linkage disequilibrium, the test of non-random association of alleles at different loci, and for Hardy-Weinberg equilibrium using the Arlequin software package (Schneider *et al.* 2000).

The statistical methods we used to further analyze the microsatellite data were based on a continuous population and isolation by distance model, where the probability of identity in state between two neutral genes decreases with the distance between them (Wright 1943, 1969). This model is most suitable when analyzing data from species that are continuously distributed, or where demes or subpopulations cannot be easily demarcated. The analyses were subdivided into three approaches: (i) spatial autocorrelation statistics, (ii) spatial regression statistics, and (iii) hierarchical *F*-statistics. Although, the calculation of *F*-statistics is not based on the isolation by distance model *per se*, the method of the sequential pooling of individuals into populations allows the inference of the scale at which populations may be structured.

*(i) Spatial autocorrelation statistics: S*patial autocorrelation methods were used to assess the spatial distribution of allelic frequencies among localities. Spatial autocorrelation methods typically involve the calculation of some spatial statistic, such as Moran's *I* (Sokal & Oden 1978, Epperson & Li 1996) at subsequent distance intervals. Theoretically, a decrease in the statistic with distance is indicative of the degree of isolation by distance (Hardy & Vekemans 1999, Diniz-Filho & Telles 2002,Vekemans & Hardy 2004). Since the choice of distance intervals can bias the interpretation of

results in the spatial analyses (Fenster *et al.* 2003), Hardy & Vekemans (2003) have suggested the use of two statistics that prevent this potential bias. The % partic is the proportion of all individuals represented at least once in the particular distance interval; whereas the CV partic is the coefficient of variation of the number of times each individual is represented in the distance interval (Hardy & Vekemans 2003). As a rule of thumb the % partic should be greater than 50% and CV partic less than or equal to 1 for each distance interval. We tried alternate distance intervals and chose the set which satisfied these conditions (Table 1). Spatial autocorrelation statistics were calculated for each of the above distance classes using SPAGeDi (Hardy & Vekemans 2003) and the relationship between spatial autocorrelation coefficients and distance (correlograms) was plotted to quantify isolation by distance.

*(ii) Spatial regression statistics:* The spatial regression of genetic distance with geographic distance was used to estimate neighborhood size (Nb). Rousset's (1997) multilocus estimator of an $F_{st}/(1-F_{st})$ between populations, was plotted against the log of geographical distance, and regression statistics calculated. We used a population-based statistic over an individual-based statistic (Rousset 2000) since we had sampled multiple individuals from single trees. An individual-based approach (Rousset 2000) would require the exclusion of data where more than one individual was sampled at a particular tree, and thus we treated individual trees as populations, and included populations located 1 to 500 km apart with more than 3 individuals. We calculated multi-locus estimates of pairwise differentiation, regressed against the logarithm of geographic distance and tested the significance of the observed correlation, between genetic and geographic distances, with a randomized permutation procedure (10000 permutations). *P*-values were estimated as the proportion of the permuted regression values that were greater, and thus indicative of greater isolation by distance, than that observed. Neighborhood size was calculated as the inverse of the slope of the regression between genetic and the logarithm of geographic distance. All spatial regression analyses were performed with GENEPOP v3.4 (Raymond & Rousset 1995).

*(iii) Hierarchical F-statistics*: Another potential method for detecting the scale at which a population may be structured is that proposed by Goudet *et al.* (1994), known as hierarchical F-statistics. The method comprises the sequential pooling of individuals into populations by increasing the extent of each population in each subsequent level, until all individuals are included as a single population. At each level *F*-statistics are calculated and plotted against the level of pooling. Although the method used to pool individuals is subjective, pooling on the basis of geographic distance is acceptable (Goudet *et al.* 1994, Arnaud *et al.* 2001). We used six levels for the calculation of *F*-statistics: within 1km, 5km, 10km, 50km, 100km and finally northern versus

**Table 1**: Statistics for the optimal distance classes chosen for spatial correlogram analysis. "Maximum distance" is the upper bound of each distance class, "Number of pairs" is the number of pairwise comparisons within the distance class. "% partic" and "CV partic" are indications of the proportion of all individuals utilised to calculate the statistics of interest within each distance class.

| **Distance Class** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| Maximum Distance (km) | 10 | 50 | 100 | 150 | 200 | 250 | 300 | 400 | 500 | 600 |
| Number of pairs | 384 | 1104 | 1159 | 1608 | 1473 | 556 | 753 | 1453 | 664 | 26 |
| % partic | 95.6 | 94.1 | 91.9 | 93.4 | 100 | 90.4 | 96.3 | 100 | 64.7 | 11.0 |
| CV partic | 0.72 | 0.71 | 0.82 | 0.65 | 0.60 | 1.14 | 0.90 | 0.86 | 1.24 | 3.64 |
| Mean distance (km) | 1.5 | 38.8 | 76.1 | 128 | 171 | 225 | 278 | 344 | 448 | 511 |
| Mean ln(distance) | -0.35 | 3.62 | 4.31 | 4.85 | 5.14 | 5.41 | 5.62 | 5.84 | 6.10 | 6.24 |

southern populations (localities A-B and C-G in Figure 1 respectively). For each pooling level, adjacent populations within the specified distance of one another were combined as a single population. Weir & Cockerham's (1984) *F*-statistics, $F_{ST}$ and $F_{IS}$, were calculated, using custom software for each sequential grouping, and 95% confidence intervals were obtained by bootstrapping over loci.

*Simulation model*

To investigate the relationship of dispersal distance and the inference of spatial genetic pattern under fluctuating population size, we constructed a custom forward-through-time simulation lattice model of the process of population turnover in *G. postica*. The lattice model (Malécot 1951), in which each node represents an individual, has been used extensively for both analytical (Rousset 2000) and simulation (Sokal & Wartenberg 1983, Epperson 1995, Leblois *et al.* 2003, Leblois *et al.* 2004) approaches to modeling continuous populations. We used a 500 x 500 lattice model to simulate the effect of annual population size fluctuations on the inference of Rousset's (2000) pairwise-individual estimate of neighbourhood size. The model comprised four steps; (i) population size was determined stochastically, (ii) breeding pairs were formed between individuals and progeny produced, (iii) progeny genotypes were chosen at random from their parents, and mutated, (iv) lattice positions for progeny were chosen from empty lattice nodes in the subsequent generation.

(i) Population size in the subsequent generation was determined with a discrete-time Ricker model (Ricker 1954) of population growth. This discrete analogue of the logistic growth model was used since the model is capable of both stability, and of limit cycles and chaos (Turchin 2003). The Ricker model is defined by:

$$N_{t+1} = N_t e^{r_t(1-\frac{N_t}{k})}$$

where, $N_t$ is the population size at time $t$, $r_t$ the population growth rate, and $k$ is the carrying capacity. Stochasticity, and thus population size fluctuations, was introduced into the model by drawing a parameter $\varepsilon_t$, from a normal distribution with mean = 0, and variance $\sigma^2$, at each generation and calculating a new growth rate ($r_{t+1}$) by adding $\varepsilon_t$ to $r_t$ (Turchin 2003). Through changing the variance, $\sigma^2$, the degree of population size fluctuations could be altered. We simulated population turnover under three levels of variance (constant size, $\sigma^2 = 0.1$, $\sigma^2 = 0.001$) to observe the effects of size fluctuations on the inference of population statistics.

(ii) Once population size had been determined, breeding pairs were chosen by first choosing a

female, and then choosing a male located within *d* lattice units of the female. A single progeny was produced from the union of male and female.

(iii) The genotypes for each progeny were constructed by randomly choosing an allele each from the female and male of the breeding pair. Mutations were implemented as a stepwise mutation model (SMM: Ohta and Kimura 1973) with locus specific mutation rates drawn randomly from a gamma (2, 2.5 x $10^{-4}$) distribution. This procedure accounts for the observation that mutation rates at individual loci are variable, yet with a mean mutation rate considered to be the average in several taxa (Leblois *et al.* 2003). More advanced mutation models have been proposed (Generalised Stepwise Mutation Model, e.g. Pritchard *et al.* 1999, Infinite Allele Model: Kimura & Crow 1964, K-Allele Model: Crow & Kimura 1970), yet Leblois *et al.* (2003) have shown that the impact of more advanced mutation models in the calculation of neighbourhood size is negligible. Thus we have only implemented a SMM in this model.

(iv) Finally, the lattice position of the progeny was chosen randomly from within *d* lattice units of the breeding female. If the progeny was unable to find a lattice position within a specified search time, it was abandoned and not carried through to the subsequent generation. The above procedure of finding a breeding pair, creating genotypes and populating the subsequent generation was repeated until the calculated population size had been reached. By sampling breeding pairs with replacement, and limiting the search time (10000 permutations) for progeny to fill a lattice node, variance in reproductive success of individuals was introduced.

Population turnover was simulated over 1000 generations for three levels of dispersal ($d$ = 2, 4, 6), where *d* is the dispersal parameter and represents the maximum number of lattice units for breeding pair formation and natal dispersal. Although the increase in *d* is small, it translates into large increases in the breeding pair and progeny lattice search area (4 x 4, 8 x 8, 12 x 12). Preliminary runs of the simulation model under constant population size indicated that these three dispersal parameters generated neighbourhood sizes of approximately 10, 25 and 50, respectively. Thus we started the simulation from a random distribution of alleles among nodes, and simulated restricted dispersal ($d$ = 2, 4, 6) for 1000 generations. At each generation pairs were formed, genotypes constructed and progeny positions were chosen, as described above. Six microsatellite loci were simulated, where the starting number of alleles in each locus was drawn from a gamma (6, 5) distribution. This method accounted for the observed patterns of allelic diversity in *G. postica*, where a few loci were highly diverse, yet most had moderate to low diversity (Delport 2005 Chapter 3). The three starting populations were subsequently used in all simulations of fluctuating population size. When testing for the effect of population size changes on a calculated statistic such as neighborhood size, two potential sources of variation are encountered: variation in the

calculation of the statistic from a sub-sample of the population, and variation in the statistic as a result of demographic instability. Since we are interested in inferring the latter, we used a permutation procedure (n = 1000) where neighbourhood size was calculated from the generated starting population, with 200 samples drawn randomly at each permutation. In this way variance in the calculation of neighborhood size under a constant population size could be determined (Table 2), and could thus be accounted for when observing trends under fluctuating population size. We simulated population turnover for 100 generations under three levels of dispersal, three neighbourhood size starting conditions and three levels of population size fluctuation.

## Results

### *Morphological results*

Consistent with the size dimorphism results of Veldtman *et al.* (2002) we detected significant differences between male and female body mass, hind wing area, forewing area and total wing area (Table 3). Females are approximately six times heavier than males, yet have approximately only four times more wing surface area than males (Table 3). This result translates into a statistically significant higher wing load for females, than for males.

### *General statistics*

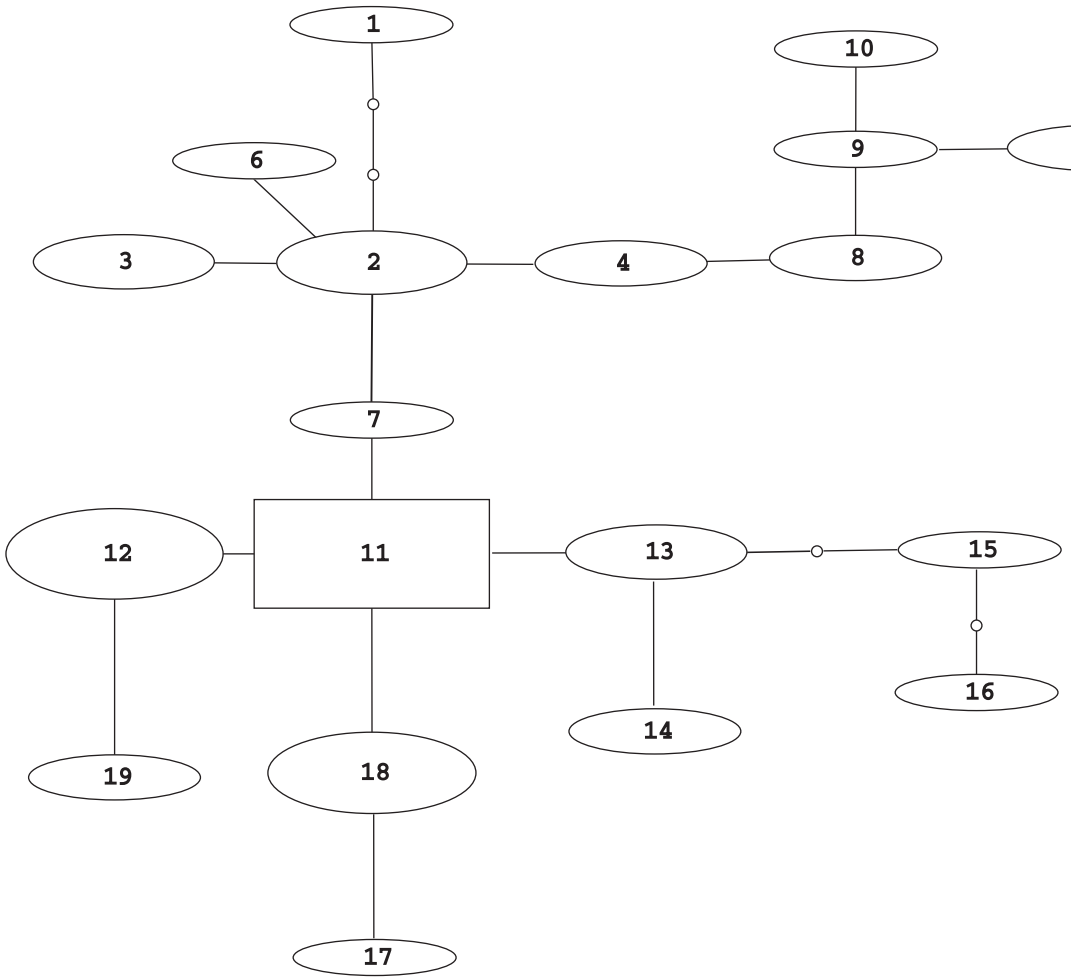
Statistical tests for deviations from neutrality indicated that the AT rich mtDNA region sequenced did not show significant deviations from neutrality (Tajima's D = -0.875, P = 0.19; Fu's $F_s$ = -2.821, P = 0.067) where significance was determined with 1000 permutations. Both neutrality statistics, however, were negative, which is indicative of an excess of homozygotes. Given that *G. postica* populations experience large annual population size fluctuations this result is expected. However, whether the observed deviations from neutrality, although non-significant, are the result of demographic instability or a selective sweep would require analysis of another unlinked marker. The allele network (Figure 2) indicates little genetic structure among localities, with a few high frequency widely distributed alleles. Allele 11 in particular is found in six of the eight locality groups (Figure 2). Again this pattern is typical of a population that has experienced population growth (Rogers & Harpending 1992, Avise 2000). Although the overall pattern appears to be panmixia, some populations have unique alleles. *F*-statistics (Table 4) indicate that most locality groups have high levels of maternal gene flow, yet groups G and H exhibit significant structure when compared with the remaining locality groups. Finally, the probability of obtaining the observed distribution of alleles among localities at random is extremely low ($X^2$ = 161.63, *P* <

**Table 2:** Sampling variance in the calculation of neighbourhood size (Nb) from starting populations used in the simulation model. Mean and standard deviations of 1000 permutations of the calculation of neighbourhood size (Nb) using 200 samples drawn at random from each of the three starting populations (Nb = 10, Nb = 25, Nb = 50).

| | **Nb = 10** | **Nb = 25** | **Nb = 50** |
|---|---|---|---|
| **Mean** | 13.08 | 23.11 | 39.14 |
| **Sd** | 2.39 | 6.27 | 17.62 |

**Table 3**: Indirect morphological estimates of dispersal ability in *Gonometa postica*. Mass, hindwing, forewing and total wing areas are shown, as is the wing load; the ratio of body mass to total wing area. The results of a t-test, for unequal sample sizes, of the difference in means between male and female morphological characteristics is shown, $F_{(df1,df2)}$, where *df* = degrees of freedom, as is the associated *P* value.

| | **male** | | **female** | | $F_{(1,149)}$ | ***P*** |
|---|---|---|---|---|---|---|
| | **mean** | **sd** | **mean** | **sd** | | |
| **Mass (g)** | 0.533 | 0.111 | 2.973 | 0.652 | 769.587 | <<0.001 |
| **Hindwing area ($mm^2$)** | 63.013 | 12.778 | 268.366 | 52.585 | 823.228 | <<0.001 |
| **Forewing area ($mm^2$)** | 124.936 | 20.804 | 446.481 | 85.430 | 764.604 | <<0.001 |
| **Total wingarea ($mm^2$)** | 187.949 | 32.758 | 714.847 | 136.210 | 808.305 | <<0.001 |
| **Wing load ($g/mm^2$)** | 0.00291 | 0.001 | 0.00420 | 0.001 | 85.595 | <<0.001 |
| **Wing load ($N/m^2$)** | 28.64 | 7.63 | 41.24 | 8.34 | | |




| Allele | Locality (frequency) | Total Frequency |
|---|---|---|
| 1 | A | 1 |
| 2 | A (1), E (2), F (1) | 4 |
| 3 | B (1), F (2) | 3 |
| 4 | H (2) | 2 |
| 5 | C (1), D (1) | 2 |
| 6 | B (1) | 1 |
| 7 | B (1) | 1 |
| 8 | B (1), F (1) | 2 |
| 9 | C (1) | 1 |
| 10 | B (1) | 1 |
| 11 | B (2), C (1), D (1), E (2), F (2), G (1) | 9 |
| 12 | F (1), G (5), H (1) | 7 |
| 13 | F (1), H (2) | 3 |
| 14 | C (2) | 2 |
| 15 | H (1) | 1 |
| 16 | H (1) | 1 |
| 17 | D (1) | 1 |
| 18 | E (3), F (2), H (1) | 6 |
| 19 | D (2) | 2 |

**Figure 2**: Mitochondrial DNA allele network constructed in TCS (Clement *et al.* 2000). Frequencies of each allele are depicted by size and in the accompanying table. Frequencies of each allele at each mtDNA sampling locality group are presented.

**Table 4**: Calculation of mtDNA pairwise-population $F_{ST}$ values using Arlequin (Schneider *et al.* 2000) for subpopulations/demes defined in Figure 1. Subpopulations A and B were combined for the purpose of this analysis. $F_{ST}$ values are presented below the diagonal, and significance of deviations of $F_{ST}$ from zero are shown above the diagonal (10000 permutations), where significance at the 0.05 level is indicated in bold. Sample sizes for each subpopulation/deme are indicated in parentheses.

| | **A&B (9)** | **C (5)** | **D (5)** | **E (7)** | **F (10)** | **G (6)** | **H (8)** |
|---|---|---|---|---|---|---|---|
| **A&B** | | 0.46273 | 0.46203 | 0.24126 | 0.93110 | **0.00129** | **0.00040** |
| **C** | 0.01656 | | 0.61439 | 0.10583 | 0.13108 | **0.01495** | **0.02158** |
| **D** | 0.01656 | 0.02174 | | 0.10751 | 0.12821 | **0.01346** | **0.02099** |
| **E** | 0.03724 | 0.12454 | 0.12454 | | 0.54965 | **0.00505** | **<0.0001** |
| **F** | -0.04553 | 0.04306 | 0.04306 | -0.02728 | | **0.01257*** | 0.14979 |
| **G** | **0.28605** | **0.38120** | **0.38120** | **0.41312** | **0.24121** | | **0.00416** |
| **H** | **0.06014** | **0.09744** | **0.09744** | **0.16667** | 0.03903 | **0.28292** | |

0.001), as determined with permutations of the chi-square test (Roff & Bentzen 1989). Distributions of microsatellite allele frequencies (Figure 3) indicate that most loci have high frequencies of one to two common alleles, yet low frequencies of the remaining alleles. Of the six microsatellite loci, four (Gon6, Gon60, Gon65, Gon120) showed significant deviations from Hardy-Weinberg equilibrium (Table 5), of which two (Gon60, Gon65) were highly significant ($P < 0.001$). No evidence for linkage among loci was detected.

*Statistical genetics methods*

*(i) Spatial autocorrelation*

Spatial autocorrelation statistics plotted against distance showed no clear pattern of isolation by distance (Figure 4), with most correlograms showing random fluctuations in spatial autocorrelation statistics with distance. Furthermore, results from permutation tests of the regression of spatial statistics against ln (distance) indicate that combined-locus correlograms do not indicate a significant association of spatial statistics with distance (Table 6). However, some locus-specific correlograms, specifically Gon55, do show significant regressions with distance for the two kinship statistics (Loiselle *et al.* 1995, Ritland 1996) and Moran's I (Table 6). Gon55, a perfect tetranucleotide repeat, exhibits a spatial pattern for some statistics that is significantly greater than that of a permuted sample. Yet, examination of the spatial regression pattern on correlograms (Figure 4) indicates that these statistics show a substantial increase or decrease at the 500 or 600km level. The % partic and CV partic (Table 1) at these distance levels indicate that fewer individuals are represented at these distance classes (low % partic), and some individuals may be overrepresented (high CV partic). This is most likely the cause of the observed deviations from no pattern of spatial autocorrelation that is predominant in the data.

*(ii) Spatial regression*

The regression analysis of pairwise $F_{ST}/(1-F_{ST})$ against geographic distance yielded a positive relationship (Figure 5). However, permutation tests determined that the probability of a random correlation between geographic and genetic distances being greater than that observed, was intermediate ($P = 0.432$) and thus the observed correlation is non-significant. Calculation of neighbourhood size from the slope of the regression yielded an estimate of 167 individuals. Given a demographic density estimate, and the calculated neighbourhood size, it is possible to infer dispersal (Rousset 2000). McGeoch *et al.* (2004) provide preliminary density estimates of 150-1075 cocoons per hectare (0.015 – 0.1075 cocoons/$m^2$). However, since *G. postica* cocoons are parasitized extensively (Veldtman *et al.* 2004) we adjusted the density for no loss, 25% loss and

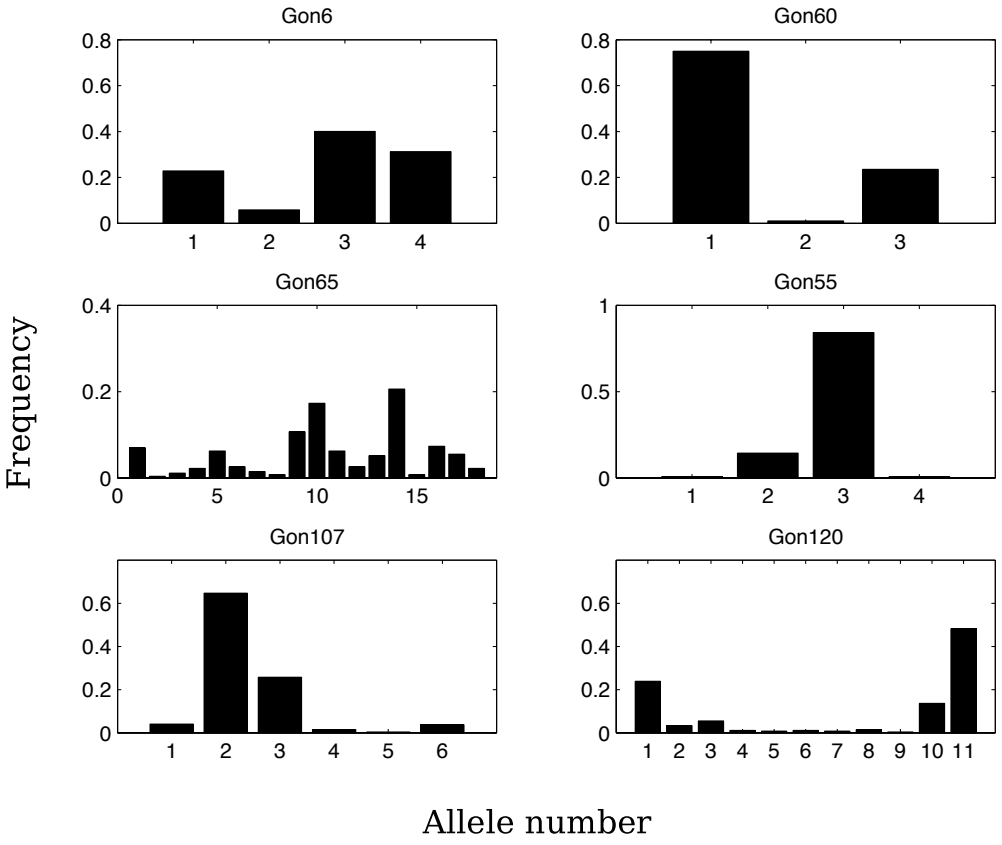



**Figure 3**: Distribution of microsatellite allele frequencies for each of the six loci analysed for 137 *G. postica* samples collected across the distribution of the species.

**Table 5**: Tests of the deviation from Hardy-Weinberg, the random-association of alleles within diploid individuals as calculated using Arlequin v2.0 (Schneider *et al.* 2000). Significant deviations from Hardy-Weinberg equilibrium for each microsatellite locus was calculated with 100000 steps in the Markov chain, with 1000 dememorisation steps. $H_0$ = observed heterozygosity, $H_e$ = expected heterozygosity, $n$ = number of alleles per locus.

| | ***Ho*** | ***He*** | ***P*** | ***n*** |
|---|---|---|---|---|
| Gon6 | 0.780 | 0.690 | 0.029* | 4 |
| Gon60 | 0.162 | 0.391 | < 0.001** | 3 |
| Gon65 | 0.493 | 0.893 | < 0.001** | 17 |
| Gon55 | 0.272 | 0.277 | 1.000 | 5 |
| Gon107 | 0.500 | 0.514 | 0.606 | 11 |
| Gon120 | 0.662 | 0.694 | 0.002* | 4 |

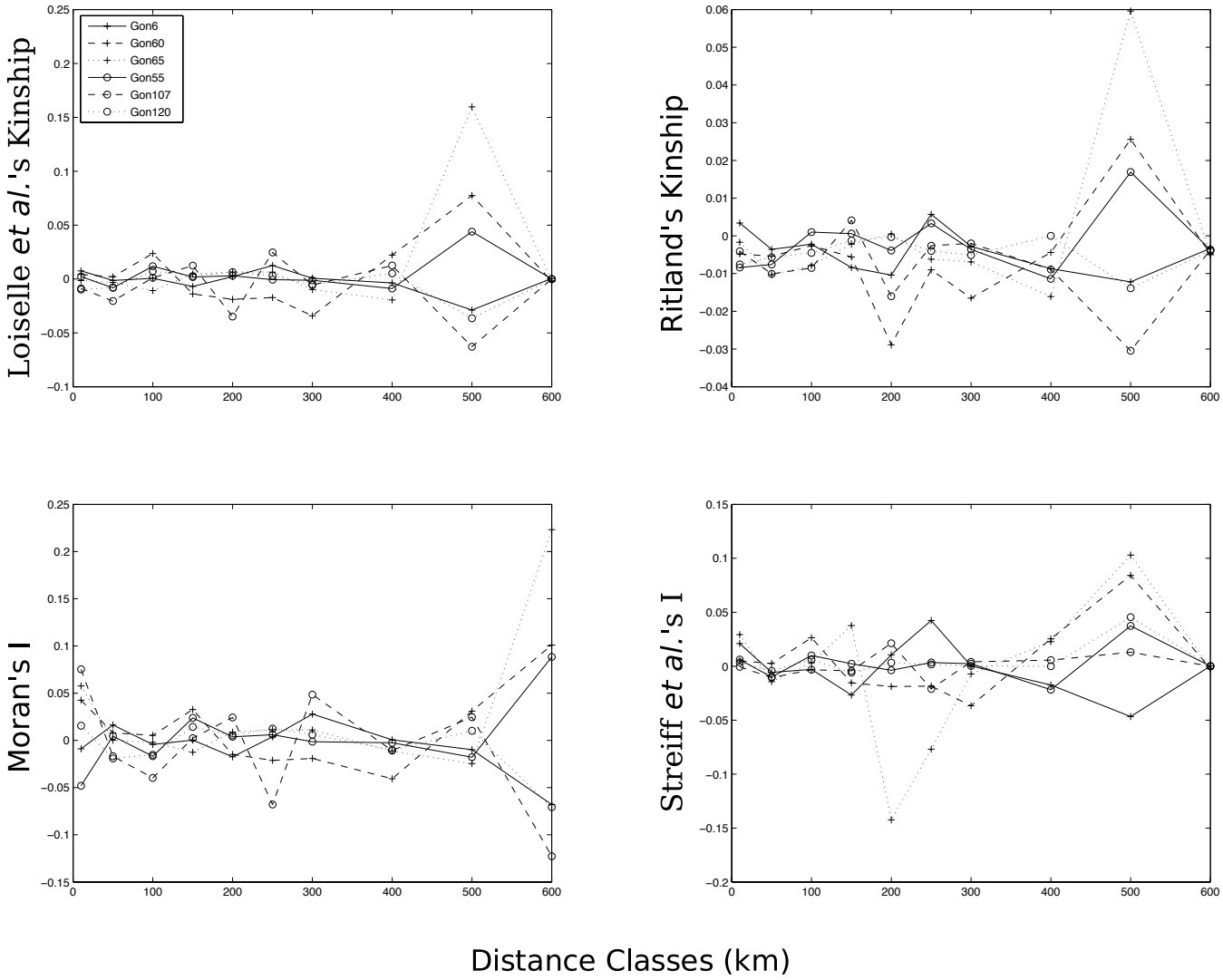



**Figure 4**: Correlograms of spatial autocorrelation statistics plotted against geographic distance classes. In each case the spatial autocorrelation statistic calculated is plotted against increasing distance for each of the six loci.

**Table 6**: Results of permutation tests for the regression of spatial autocorrelation statistics against ln (distance). $K_L$= Loiselle *et al.*'s (1995) kinship; $K_R$ = Ritland's (1996) kinship; $I_k$ = Moran's I (Sokal & Oden 1978); I = Streiff *et al.*'s (1998) I. For each case the probability, *P*, that a permuted regression statistic is greater than that observed (obs < per), less than that observed (obs > per) and not equal to that observed (obs <> per) is represented. * = significance at 0.05; ** = significance at 0.01. obs = observed regression statistics, per = permuted regression statistic. A Bonferroni corrected alpha level of 0.0125 was used for significance at the 0.05 level.

| | **$K_L$** | **$K_R$** | **$I_k$** | **I** |
|---|---|---|---|---|
| **All Loci** | | | | |
| *P* (H: obs < per) | 0.57 | 0.3 | 0.58 | 0.35 |
| *P* (H: obs > per) | 0.43 | 0.7 | 0.42 | 0.65 |
| *P* (H: obs <> per) | 0.85 | 0.59 | 0.84 | 0.71 |
| **Gon6** | | | | |
| *P* (H: obs < per) | 0.92 | 0.84 | 0.93 | 0.47 |
| *P* (H: obs > per) | 0.08 | 0.16 | 0.07 | 0.53 |
| *P* (H: obs <> per) | 0.16 | 0.33 | 0.15 | 0.94 |
| **Gon60** | | | | |
| *P* (H: obs < per) | 0.2 | 0.25 | 0.21 | 0.21 |
| *P* (H: obs > per) | 0.8 | 0.75 | 0.79 | 0.79 |
| *P* (H: obs <> per) | 0.41 | 0.51 | 0.42 | 0.41 |
| **Gon65** | | | | |
| *P* (H: obs < per) | 0.04 | 0.04 | 0.04 | 0.18 |
| *P* (H: obs > per) | 0.96 | 0.96 | 0.96 | 0.82 |
| *P* (H: obs <> per) | 0.09 | 0.08 | 0.09 | 0.37 |
| **Gon55** | | | | |
| *P* (H: obs < per) | 0.99 | 0.99 | 0.99 | 0.87 |
| *P* (H: obs > per) | **<0.01**** | **<0.01**** | **0.01*** | 0.13 |
| *P* (H: obs <> per) | 0.03 | **<0.01**** | 0.03 | 0.25 |
| **Gon107** | | | | |
| *P* (H: obs < per) | 0.91 | 0.7 | 0.92 | 0.64 |
| *P* (H: obs > per) | 0.09 | 0.3 | 0.08 | 0.36 |
| *P* (H: obs <> per) | 0.18 | 0.59 | 0.17 | 0.71 |
| **Gon120** | | | | |
| *P* (H: obs < per) | 0.66 | 0.74 | 0.66 | 0.62 |
| *P* (H: obs > per) | 0.34 | 0.26 | 0.34 | 0.38 |
| *P* (H: obs <> per) | 0.69 | 0.51 | 0.68 | 0.75 |

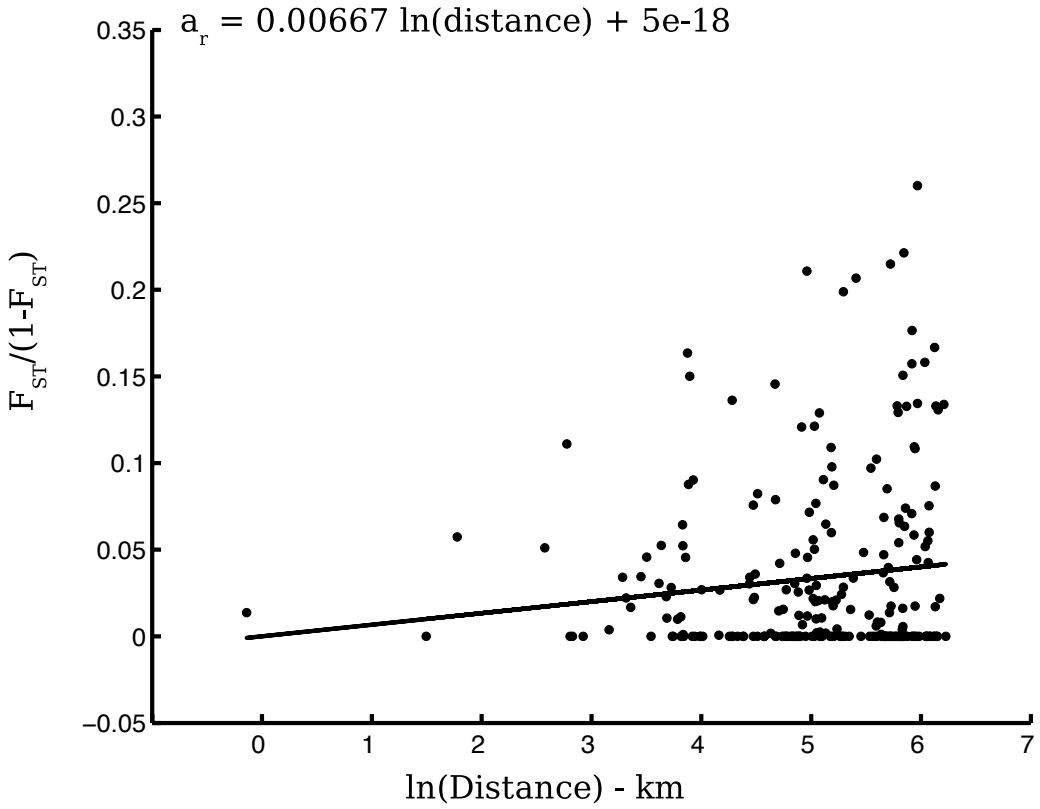



**Figure 5**: Genetic differentiation in the African Wild Silk Moth, *Gonometa postica*, represented as the regression of pairwise genetic distance against the natural logarithm of distance (km). $a_r$ = 0.00667 ln(distance) + 5e-18 . Permutations (10000) indicated that the probablity of obtaining a randomised correlation greater than that observed was moderate (P = 0.433).

50% loss to parasitism and predation. These corrected densities of adults/$m^2$ have not been sufficiently replicated for certainty; yet do provide a potential dispersal range of *G. postica*, given the observed microsatellite data. The adjusted densities, combined with a neighbourhood size of 167 individuals, suggest a mean parent-offspring dispersal distance of between 11.1 and 42.1m.

*(iii) Hierarchical F-statistics*

Hierachical *F*-statistics indicated a slight increase in $F_{IS}$ with the subsequent pooling levels (Figure 6a). Since $F_{IS}$ is a measure of deviation from Hardy-Weinberg within subpopulations, an increase in $F_{IS}$ is expected over subsequent pooling levels. This result is the consequence of combining non-interbreeding subpopulations, and thus violating the Hardy-Weinberg equilibrium assumption of random mating within populations (Wahlund effect, Wahlund 1928). In contrast, subsequent pooling levels show a decrease in $F_{ST}$ (Figure 6b), especially evident after the $4^{th}$ level, where localities occurring within 50km of one another were pooled. Since $F_{ST}$ is a measure of genetic differentiation over subpopulations, the rapid decrease in $F_{ST}$ at this point is indicative of where the pooling level represents an essentially panmictic population, i.e. between subpopulation genetic differentiation is pooled into a single panmictic unit. Furthermore, we tested whether the mean permuted $F_{ST}$ values at pooling levels 4 and 5 differed significantly using a one-way-ANOVA. The mean of permuted $F_{ST}$ values at pooling levels 4 and 5 were significantly different, and thus reinforce the slight decrease in overall population genetic structure observed at this pooling level. Thus, although the overall levels of $F_{ST}$ are low, the pattern observed using hierarchical $F_{ST}$ analysis does allow the inference of the geographical extent of a subpopulation.

*Simulation model*

It is evident that quite different patterns of population size fluctuation can be achieved through adjusting the parameter $\sigma^2$ in the calculation of per generation population size (Figure 7). These population size fluctuations translate into different patterns of change in neighbourhood size over time dependent on the dispersal parameter. First, when population size is kept constant, low dispersal parameters ($d$ = 2, d = 4) appear to reach equilibrium at neighbourhood sizes of approximately 10 and 25-30 respectively (Figure 8, Table 7), irrespective of the neighbourhood size of the starting population. A larger dispersal parameter ($d$ = 6) shows a steady increase in neighbourhood size over time, and when run over increased number of generations (1000) achieves equilibrium at a neighbourhood size of 40-50 individuals (Table 7). When subjected to population size fluctuations neighbourhood sizes show (i) a steady decrease, (ii) a rapid decrease or (iii) no apparent difference in comparison to constant population size. Steady decreases in neighbourhood size are observed at a small starting neighbourhood size (Nb = 10) and low dispersal ($d$ = 2), for

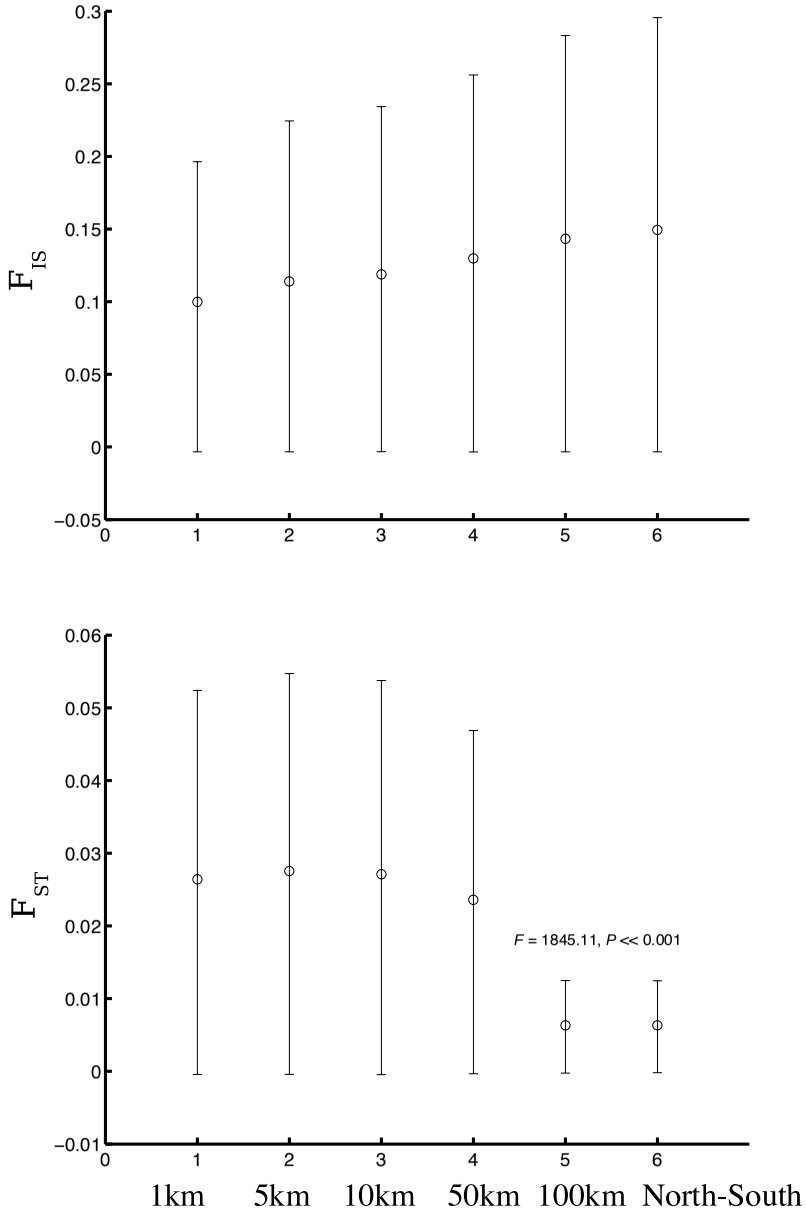



**Figure 6**: Changes in $F_{ST}$ and $F_{IS}$ calculated over six polymorphic microsatellite loci for each subsequent pooling level in the hierarchical $F_{ST}$ analysis. Pooling levels are described in the methods. Mean and standard errors of the mean, as determined from bootstrapping over loci, are shown. Results of a one-way ANOVA of bootstrapped $F_{ST}$'s for pooling levels 4 and 5 are indicated by the test-statistic, *F*, and associated *P* value.

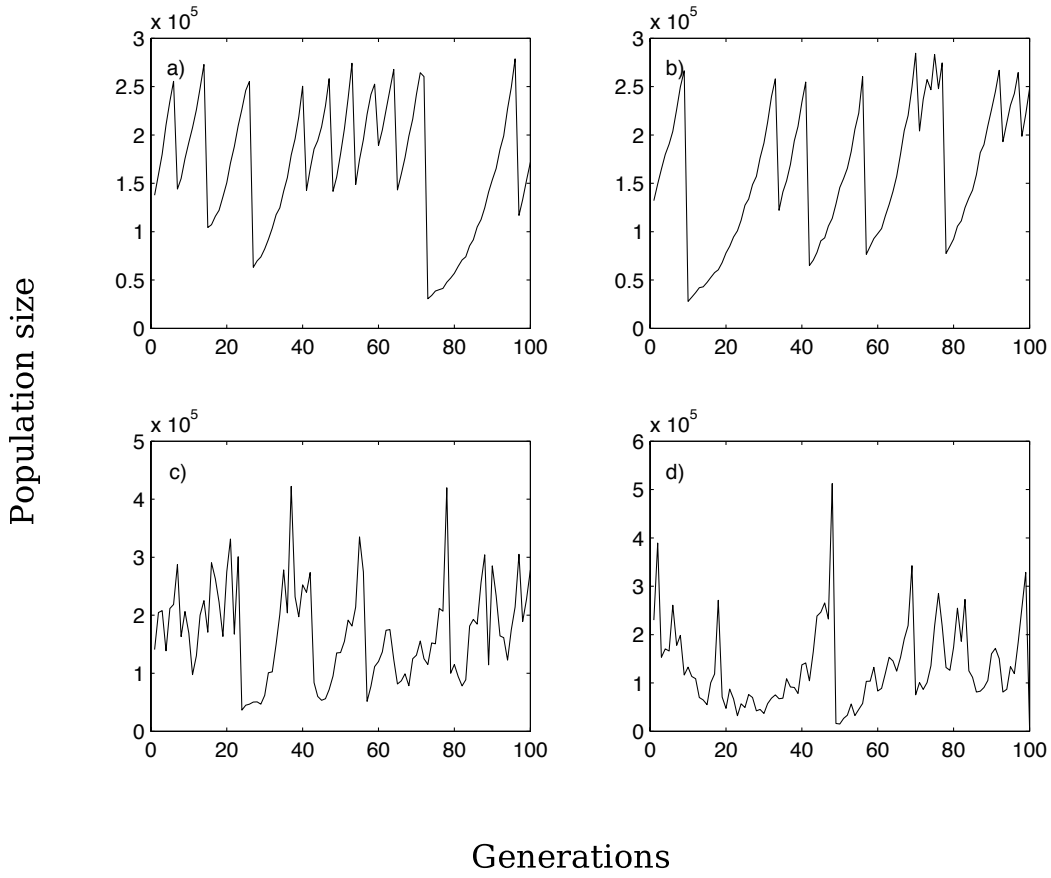



**Figure 7**: Examples of population size fluctuations achieved with the Ricker logistic growth model (Ricker 1954). a & b) low variance $\sigma^2 = 0.001$; c & d) high variance $\sigma^2 = 0.1$.

both high and low population size variance (Figure 8), although the latter is characterized by a sudden drop in neighbourhood size after seventy generations. In addition, for low to intermediate dispersal (d = 2, d = 4) the final neighbourhood sizes after 100 generations are lower than that of constant population size, given the same starting conditions and dispersal parameters. At a small starting neighbourhood size, subpopulation genetic differentiation is expected to be high, and when dispersal is low, this subpopulation genetic differentiation should be maintained by random genetic drift within local subpopulations. Neighbourhood sizes are most likely smaller since the probability for alternate alleles to become fixed in different subpopulations is greater, than in a constant population size scenario. This notion is supported by the observed reduction in heterozygosity under population size fluctuations (Figure 9).

Rapid decreases in neighbourhood size are observed for intermediate and large starting neighbourhood sizes under high population size variance and to an extent under low population size variance (Figure 8). Again, given low levels of dispersal the probability of alternate alleles becoming fixed, due to random genetic drift when population size is substantially reduced, is greater than when dispersal is greater. In general, population size fluctuations, and lower dispersal distances cause one to underestimate neighbourhood size, the degree of which is dependent on dispersal distance. In addition, the speed at which neighbourhood size is reduced under low levels of dispersal is dependent on the variance in population size fluctuations, with larger fluctuations causing faster reductions in inferred neighbourhood size. Finally, no apparent differences between neighbourhood sizes calculated from a fluctuating versus constant population are observed at large dispersal distances (d = 6). Deviations from this generalization are evident at an intermediate starting neighbourhood size, where neighbourhood size after 100 generations is substantially lower than in the constant population size scenario. Since each simulation is an independent stochastic event, deviations that are the result of the independent population size changes over time are expected. In particular, the population size change profile for the anomalous simulation was characterized by a large population crash after sixty generations, and a failure to recover from low numbers before the end of the simulation. This large population crash, would first decrease allelic diversity overall (Figure 9), but through random choosing of alleles for the generation subsequent to the crash allow for alternate alleles to become fixed in different positions on the lattice. This would have the overall result of reducing neighbourhood size; and preventing a subsequent recovery to normal levels, with a population size increase, before the end of the simulation. However, the general result for high dispersal, of limited effect, is intuitive since large dispersal distances will tend to homogenize, through gene flow, any local genetic differentiation that has occurred as a result of a decrease in population size. It is fascinating, however, that a mean parent-offspring distance of only 6 lattice units (less than 1.5% of the dimensions of the lattice) is sufficient for

**Table 7**: Relationship between dispersal parameter, *d*, and the equilibrium neighbourhood size for a continuous population of constant size. Mean neighbourhood size (and standard deviations) are presented for the last 200 generations, of a total of 1000 generations simulated, for three levels of dispersal, *d*, and three levels of starting population neighbourhood size, *Nb*.

| | ***d* = 2** | ***d* = 4** | ***d* = 6** |
|---|---|---|---|
| ***Nb* = 10** | 9.66 (1.0) | 30.80 (4.10) | 51.28 (20.58) |
| ***Nb* = 25** | 12.03 (1.69) | 31.72 (5.02) | 47.67 (22.55) |
| ***Nb* = 50** | 10.56 (1.13) | 25.31 (6.82) | 40.95 (10.13) |

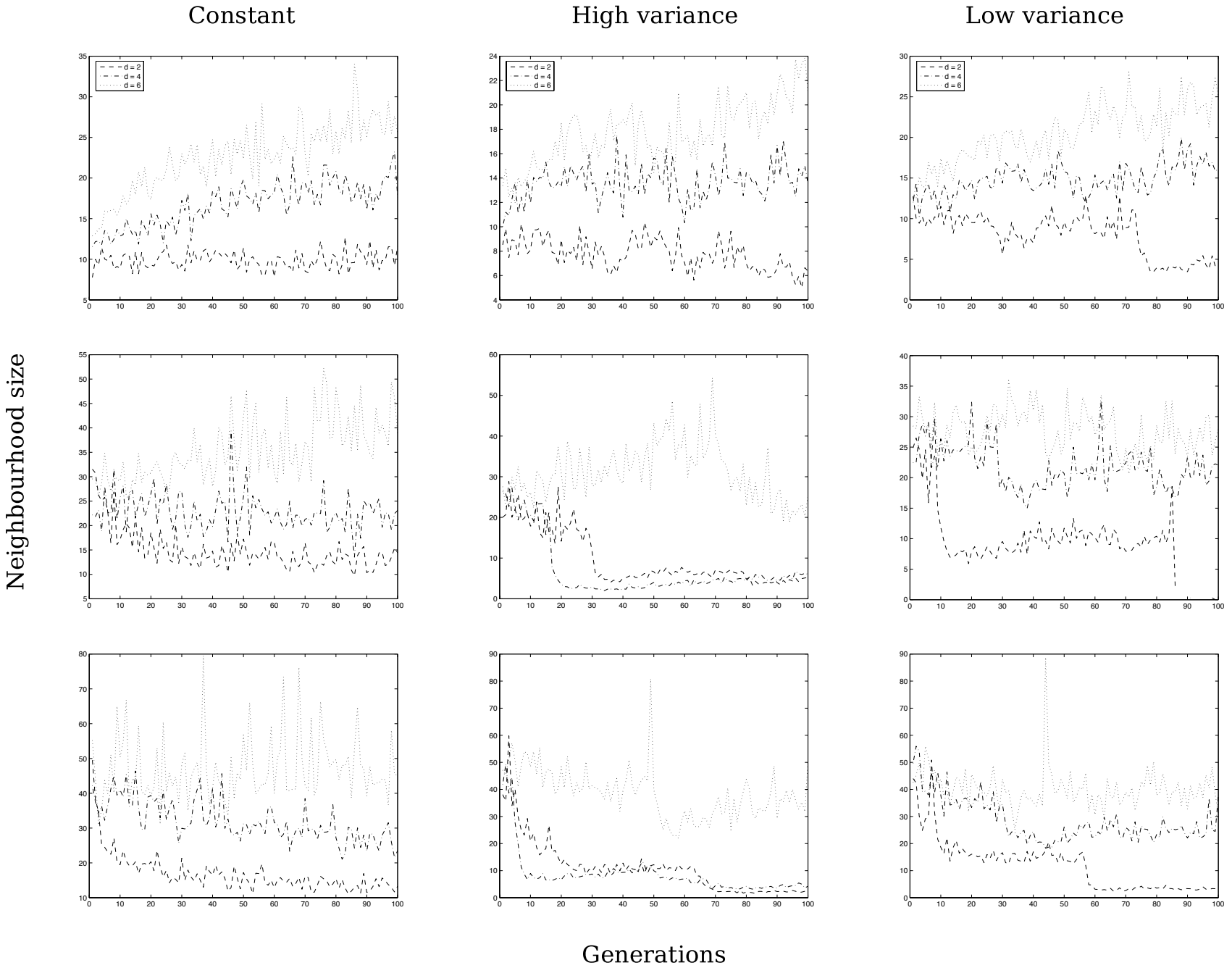



**Figure 8**: Changes in neighbourhood size over time under constant population size and fluctuating population size for three levels of dispersal, and three starting population neighbourhood sizes.

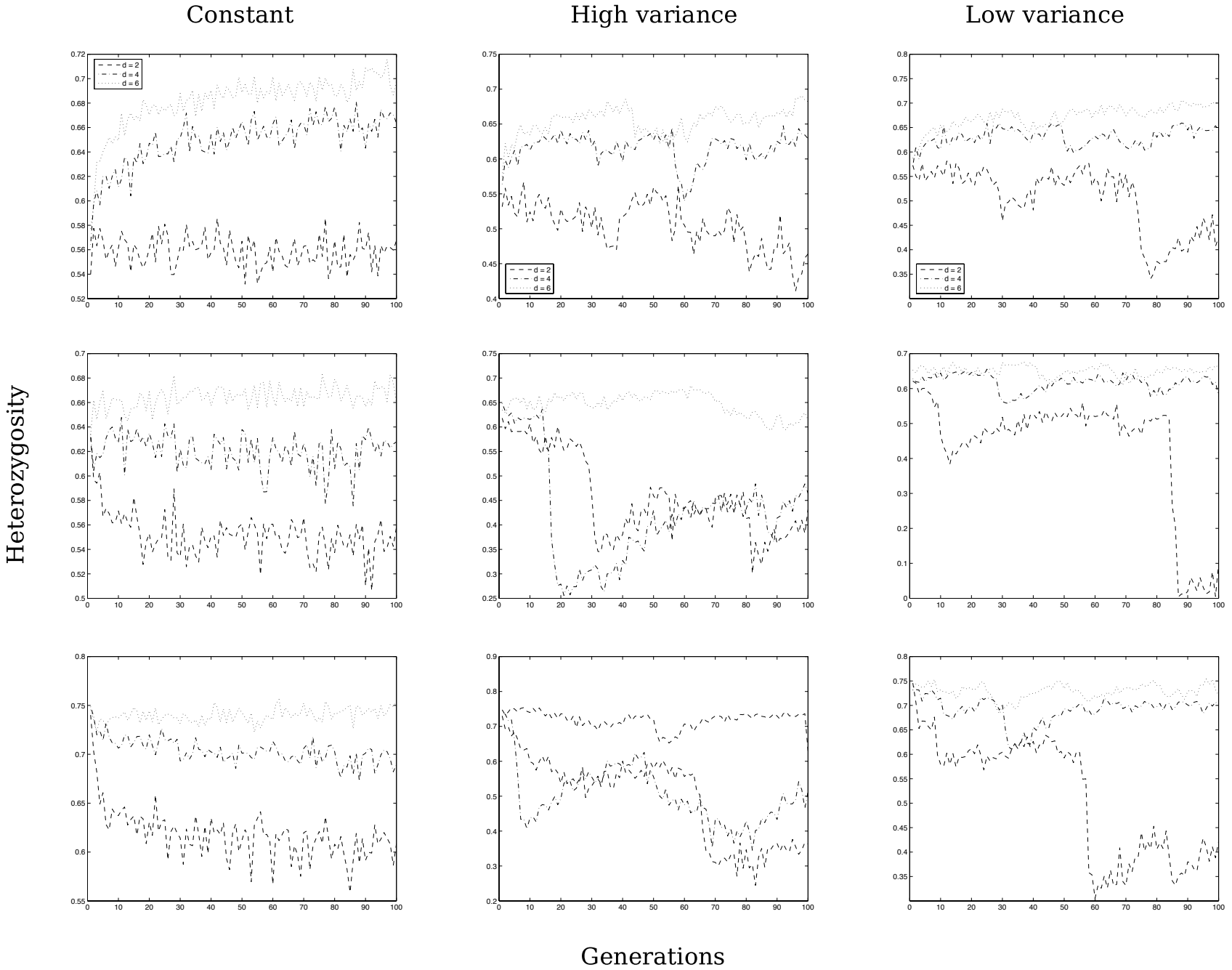



**Figure 9**: Changes in heterozygosity over time under constant population size and fluctuating population size for three levels of dispersal, and three starting population neighbourhood sizes.

population size fluctuations not to have any visible effect on the estimation of neighbourhood size.

## Discussion

The major conclusion in this study is that the inference of demographic parameters from genetic data is highly susceptible to the assumptions of the demographic model used in analyzing the data. Thus far no analytical models are suitable for analyzing population genetic data from species with large annual population size fluctuations. We have shown using simulation modeling that the inference of neighbourhood size from microsatellite data obtained from such a species is dependent on both the degree of population size fluctuations, and on the level of dispersal. Theoretically, given a species with low dispersal ability, one would tend to underestimate neighbourhood size, as a result of local subpopulation differentiation. However, the challenge remains to interpret the data derived from *G. postica*, our focal species in this study.

### *Estimates of dispersal ability*

The morphological estimates of dispersal as determined by wing loading give an indication of dispersal ability of *G. postica*. In Lepidoptera, there is a positive correlation between body mass and wing loading (Casey & Joos 1983), and generally a lower wing load translates into more energy-efficient flying. This relationship exists since the wing stroke frequency decreases with body size (Casey & Joos 1983). At the extreme of wing load in moths are the sphingid moths, which are considered the fastest flying group of insects (Matthews 1992), and are renowned for their hovering ability when drinking nectar from flowers on the wing. Sphingid moths are known to disperse and migrate long distances (Janzen 1984, Haber & Frankie 1989). The tobacco hawk moth (*Manduca sexta*), one of the largest sphingid species, has a mass which varies between 1 and 3g, wing length of 40-60mm and wing loadings of 7-15 N $m^{-2}$ (Stevenson *et al.* 1995). *G. postica* is a moth of similar mass (males: 0.26 – 0.91g, females: 1.35 – 4.15g), yet has substantially higher wing load, the result of smaller wing surface area. Given such a high wing load, *G. postica* would need an extremely high wing stroke frequency for sustained flight, which would be energetically expensive. Efficient flying is necessary for *G. postica*, since the adult moths have no feeding mouth parts and rely on reserves from the pupal stage to sustain flight. Species with high wing stroke frequencies have flights that are irregular and characterized by erratic movements (Rydell & Lancaster 2000). Field observations of *G. postica* females, at light traps, support this notion (pers obs). Males on the other hand appear to be more efficient fliers, due to lower wing load. In addition, males begin warming up their wings with high frequency wing strokes soon after emergence from the cocoon

(pers obs). Rydell & Lander (2000) attribute the need for warming-up flight muscles to species with higher wing loads that need to achieve a higher thoracic temperature than species with lower wing loads. Given this observed difference between male and female *G. postica*, females may be more sedentary and males dispersive. The indirect evidence presented here, however, is problematic since wing load may not be sufficient to determine dispersal ability. For example, some small-bodied geometrid moths are known to have very low wing loads, yet are poor fliers (Rydell *et al.* 1997). Furthermore, non-feeding saturniid moths, a closely related family of lasiocampids, have been shown to be good dispersers (Waldbauer & Sternberg 1982), despite large body sizes and high wing loads. Furthermore, wing muscle mass may be a more important determinant of dispersal ability than wing load *per se*, and would therefore be important to consider in future studies. The evidence in support of low dispersal ability in *G. postica* is limited. Yet the absence in *G. postica* of a stability-enhancing fulcrum and spine between the hind- and forewings, characteristic of good fliers such as the Hawk moths, provides additional support for low dispersal ability in this species. In conclusion, indirect evidence of dispersal does suggest low vagility, yet these results would need to be confirmed with capture-mark recapture studies.

The potential for differences in male and female dispersive ability would certainly lead to differences in the inferred spatial genetic patterns from mitochondrial and nuclear loci. However, in this study little population genetic structuring is observed in both data sets. $F_{ST}$ values for mtDNA (Table 4), are generally greater than that of microsatellites (Figure 6). This result would be expected if females were less vagile than males, yet population size fluctuations could reduce mtDNA diversity. Since organellar genes (mtDNA) have one quarter the population size of nuclear genes the loss of organellar genetic diversity under severe bottlenecks is greater (Wilson *et al.* 1985, Grant & Leslie 1993). In particular, Grant and Leslie (1993) note that contrary to Northern Hemisphere species, greater levels of diversity are often observed in nuclear genes than in mitochondrial genes from southern African taxa. The authors attribute this loss of mtDNA diversity to the frequent extinctions and recolonisations occurring in southern African populations as a result of drought and rainfall cycles. Thus the observed population genetic pattern in mtDNA may be the result of reduced female versus male dispersal coupled with the loss of diversity as a result of population size fluctuations.

Through simulations we have shown that population size fluctuations combined with low dispersal can lead to gross underestimates of neighbourhood size from microsatellite data, and thus substantial genetic structure. However, intermediate dispersal will tend to have no apparent effect on the inference of neighbourhood size. The microsatellite data generated in this study is

characterized by a few high frequency alleles, yet there is no obvious pattern of fixing of alternate alleles in different subpopulations. Rather, the data are representative of an essentially panmictic population with a very weak signal of isolation by distance. Thus we believe the data observed are the result of population fluctuations with intermediate levels of dispersal, whereby genetic diversity is homogenized across the distribution of the species. A potential problem with comparing simulation results to patterns observed in real data is interpreting the spatial scale at which the effects occur. For example, a level of dispersal less than 1.5% of the size of the lattice area is sufficient for genetic diversity to withstand the effects of population size fluctuations in a simulation, yet this is not easily translated to a species dispersal distance. In particular, making inferences of the geographic extent of subpopulations or eruptions in such a case would be futile. Rather, one should consider the spatial analysis results.

The test for spatial autocorrelation of alleles with geographic distance indicated a pattern not significantly different from a random association. These results are partly due to the inability for spatial autocorrelation statistics to incorporate potential sources of error in population genetic data (Slatkin & Arter 1991). The authors stressed that population genetic data are characterized by three sources of variance: sampling, stochastic and parametric. These sources of variance amount to basing inference on only a subset of samples, the stochasticity of the lineage sorting process, and averaging across loci with potentially divergent mutation rates. Although these criticisms of spatial autocorrelation methods are robust, we believe the patterns observed in our data are rather the result of the failure for equilibrium to be reached. Sokal & Wartenberg (1983) have shown that correlograms are established within 50 generations, and furthermore within 150 generations correlograms reach a quasi-stationary state, which persists for hundreds of generations (Epperson 1995). Given the little we know of population size fluctuations in *G. postica* it is unlikely that correlograms, indicative of isolation by distance, would have reached an equilibrium state. Hierarchical F-statistic analysis showed a substantial reduction in $F_{ST}$ values beyond a geographical extent of 50km. Spatial regression of $F_{ST}$ against distance however, combined with rudimentary estimates of density, indicated a natal dispersal distance of between 11 and 42m. However, since the regression of $F_{ST}$ against distance is not significantly different from zero, this estimate is an absolute minimum dispersal distance. A less steep regression would imply larger neighbourhood sizes and thus larger dispersal distances. Whether, occasional long-distance dispersal or stepping-stone-effects with low dispersal distances can generate the observed patterns is questionable. Most likely, the observed patterns are the result of the combined effects of population size fluctuations, loci with low to intermediate mutation rates and low to intermediate levels of dispersal. Similar studies on continuously distributed cyclical species indicate little population genetic structuring. For example, the observed high levels of gene flow in the cyclic snowshoe hare (*Lepus americanus*) has

been attributed to a stepping-stone model of gene flow influenced by density cycles, where local bottleneck populations expand to previously unsuitable habitat and thus homogenize genetic diversity across the distribution (Burton *et al.* 2002). Another cyclic species that showed low genetic differentiation in a continuous distribution, the collared lemming, has its population genetic structure attributed to potential long-distance dispersal events (Ehrich *et al.* 2001). Indeed Burton *et al.* (2002) have noted the proliferation of high gene flow genetic signals derived from population genetic studies of cyclical species, and suggested this association may be worthy of further investigation. In this manuscript we report a similar occurrence in cyclical populations of African Wild Silk Moth. Through the use of simulations we show that neighbourhood size can be reduced, at low dispersal, yet cannot be increased as a result of population size fluctuations. Thus large neighbourhood sizes, and reduced population genetic structure in *G. postica*, does not appear to be the result of population size fluctuations. Though this observation would be dependent on the mutation rate of the loci considered. Loci with high mutation rates might be resistant to population size fluctuations, whereas loci with lower mutation rates may be more susceptible to increased homoplasy under population size fluctuations. Given lower levels of allelic diversity, the potential for the same allele to become fixed in different subpopulations, during population crashes, is greater than with higher levels of allelic diversity. This could result in the observed inference of high levels of gene flow in population genetic studies of cyclical species. We are currently investigating this process with simulation modeling.

One of the motivating factors for a population genetic study on *Gonometa postica* was the potential to guide the initiatives of the African Wild Silk industry in southern Africa. This study forms part of a larger genetic and ecological-based research programme that has been developed to gain an understanding of population cycles in this species. The purpose of the initial genetic survey was to determine the dispersal ability of *G. postica* and thus the appropriate scale at which future population dynamics studies should be conducted. Although, the genetic signal presented in this manuscript is weak, the reduction in $F_{ST}$ at a pooling level greater than 50km, in the hierarchical $F_{ST}$ analysis, provides a good indication of the extent of eruptions in this species. The addition of more microsatellite loci may provide greater resolution. In addition, temporal genetic sampling and analysis may assist in further evaluating this result. Since eruptions may result in an increased frequency of a locally rare allele, the presence of these rare alleles in adjacent regions in subsequent years may allow the inference of dispersal from the preceding eruption. We are currently investigating the potential for temporal rare allele frequencies to be informative in estimating dispersal. Knowledge of the dispersal ability of *G. postica* is crucial for the understanding of population cycles in this species, since it determines whether the species will persist in all regions given recent population crashes. The ability for relict populations to reseed regions where the

species has declined determines the long-term sustainability of the species and thus the long-term sustainability of a harvesting programme. Ultimately, dispersal estimates and the degree of spatial and temporal connectivity will be used to complement population dynamic modeling that is planned for this species. We look forward to a novel opportunity to incorporate genetic, ecological and modeling approaches to understanding the population biology of this species, and thereby providing harvesting recommendations to the Wild Silk Industry.

**Acknowledgements**

This work was funded by the Mellon Foundation Grant to J. W. H. Ferguson, P. Bloomer and W. Delport, and by a National Research Foundation grant to P. Bloomer (Gun: 2053653). The opinions and views presented in this article are however, not necessarily those of the National Research Foundation. Furthermore, we would like to thank Louis Hauman, Duncan McFadyen and E. O. Oppenheimer & Son for accommodation assistance during sampling. Thanks also to Fourie Joubert and deepthought (http://deepthought.bi.up.ac.za), a 64-node cluster, for computational support.

**References**

Arnaud JF, Madec L, Guiller A, Bellido A (2001) Spatial analysis of allozyme and microsatellite DNA polymorphisms in the land snail *Helix aspera* (Gastropoda: Helicidae). *Molecular Ecology*, **10**, 1563-1576.

Avise JC (2000) Phylogeography: The history and formation of species. Harvard University Press, Harvard.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA & Saunders NC. (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology & Systematics*, **18**, 489-522.

Beerli P & Felsenstein J. (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763-773.

Beerli P & Felsenstein J. (2001) Maximum-likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563-4568.

Bowers RG, Begon M, Hodgkinson (1993) Host-pathogen population cycles in forest insects: Lessons from simple models reconsidered. *Oikos*, **67**, 529-538.

Burton C, Krebs CJ, Taylor EB (2002) Population genetic structure of the cyclic snowshoe hare (*Lepus americanus*) in southwestern Yukon, Canada. *Molecular Ecology*, **11**, 1689-1701.

Casey TM, Joos BA (1983) Morphometrics, conductance, thoracic temperature, and flight energetics of Noctuid and Geometrid moths. Physiological Zoology, **56**, 160-173.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1659.

Delport W (2005) Characterisation of six microsatellite loci in the African Wild Silk Moth (*Gonometa postica*, Lasiocampidae). Chapter 3, University of Pretoria thesis, Pretoria, South Africa.

Delport W (2005) Temporal and spatial distribution of African Wild Silk Moth, *Gonometa postica*, eruptions in southern Africa. Chapter 2, University of Pretoria thesis, Pretoria, South Africa.

Diniz-Filho JAF, Telles MPC (2002) Spatial autocorrelation analysis and the identification of operational units for conservation in continuous populations. *Conservation Biology*, **16**, 924-935.

Ehrich D, Jorde PE, Krebs CJ, Kenney AJ, Stacy JE, Stenseth NC (2001) Spatial structure of lemming populations (*Dicrostonyx groenlandicus*) fluctuating in density. *Molecular Ecology*, **10**, 481-495.

Epperson BK & Li T (1996) Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. *Proceedings of the National Academy of Sciences, USA*, **93**, 10528-10532.

Epperson BK (1995) Spatial distributions of genotypes under isolation by distance. *Genetics*, **140**, 1431-1440.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **136**, 343-359.

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.

Fenster CB, Vekemans X, Hardy OJ (2003) Quantifying gene flow from spatial genetic structure data in a metapopulation of Chamaecrista fasciculata (Leguminosae). Evolution, **57**, 995-1007.

Freddi G, Bianchi Svilokos A, Ishikawa H, Tsukada M (1993) Chemical composition and physical properties of *Gonometa rufobrunnea* silk. *Journal of Applied Polymer Science*, **48**, 99-106.

Ginzburg LR, Taneyhill DE (1994) Population cycles of Forest Lepidoptera: A maternal effects hypothesis. *Journal of Animal Ecology*, **63**, 79-92.

Goudet J (1995) Fstat version 1.2: a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.

Goudet J, De Meeüs T, Day AJ, Gliddon CJ (1994) The different levels of population structuring of the dogwhelk, *Nucella lapillus*, along the South Devon coast. In: *Genetics and Evolution of Aquatic Organisms* (ed. Beaumont AR), pp. 81-95. Chapman & Hall, London.

Grant WS, Leslie RW (1993) Effect of metapopulation structure on nuclear and organellar DNA variability in semi-arid environments of southern Africa. *South African Journal of Science*, **89**, 287-293.

Haber WA & Frankie GW (1989) A tropical hawkmoth community: Costa Rican dry forest Sphingidae. Biotropica, **21**, 155-172.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145-154.

Hardy OJ & Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Hardy OJ & Vekemans X (2003) SPAGeDi 1.1: a program for *S*patial *P*attern *A*nalysis of *Ge*netic *Di*versity, User's manual.

Hastings A, Harrison S (1994) Metapopulation dynamics and genetics. *Annual Review of Ecology and Systematics*, **25**, 167-188.

Hudson RR (1990) Gene genealogies and the coalescent process. In: Oxford Surveys in Evolutionary Biology Volume 7 (eds. Futuyama D & Antonovics J), pp 1-43. Oxford University Press, Oxford.

Janzen DH (1984) Two ways to be a tropical big moth: Santa Rosa saturniids and sphingids. In: Oxford Surveys in Evolutionary Biology Volume 1 (eds ), pp 85-140. Oxford University Press, Oxford.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725-738.

Kingman JFC (1982) The coalescent. Stochastic Processes and their Applications, **13**, 235-248.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623-2635.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1-10.

Kuhner MJ, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on coalescent. *Genetics*, **149**, 429-434.

Kuhner MK, Yamato J, Beerli P, Smith LP, Rynes E, Walkup E, Li C, Sloan J, Colacurcio P, Felsenstein J (2004) LAMARC v 1.2.1. University of Washington, http://evolution.gs.washington.edu/lamarc.html.

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

Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics*, **166**, 1081-1092.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of tropical understory shrub, Psychotria officinalis (Rubiaceae). *American Journal of Botany*, **82**, 1420-1425.

Malécot (1951) Un traitement stochastique des problémes linéaires (mutation, linkage, migration)

en génétique des populations. *Annales de l'Université de Lyon A*, **14**, 79-117.

Matthews P (1992) (ed.) The Guinness Book of Records 1993. New York.

McGeoch MA, Veldtman R, Scholtz CH (2004) Feasibility of annual cocoon yield requirements. Report to Wild Silk Africa. University of Stellenbosch, South Africa.

Nielsen R, Wakeley JW (2001) Distinguishing Migration from Isolation: an MCMC Approach. *Genetics*, **158**, 885-896.

Ohta T & Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, **22**, 201-204.

Posada D & Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37-45.

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *American Journal of Human Genetics*, **65**, 220-228.

Rasband WS (1997) *ImageJ*, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/.

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

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

Ricker WE (1954) Stock and recruitment. *J. Fish. Res. Board Can.*, **11**, 559–663.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research Cambridge*, **67**, 175-185.

Roff DA & Bentzen P (1989) The statistical analysis of mitochondrial DNA polymorphisms: chi-square and the problem of small samples. *Molecular Biology and Evolution*, **6**, 539-545.

Rogers AR & Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetics differences. *Molecular Biology and Evolution*, **9**, 552-569.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*,**145**, 1219-1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58-62.

Rydell J, Skals N, Surlykke A & Svensson M (1997) Hearing and bat defence in geometrid winter

moths. *Proceedings of the Royal Society of London B*, **264**, 83-88.

Rydel J, Lancaster WC (2000) Flight and thermoregulation in moths were shaped by predation from bats. *Oikos*, **88**, 13-18.

Schneider S, Roessli D & Excoffier L (2000) ARLEQUIN version 2: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Slatkin M & Arter HE (1991) Spatial autocorrelation methods in population genetics. *The American Naturalist*, **138**, 498-517.

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society*, **10**, 199-228.

Sokal RR & Wartenberg DE (1983) A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics*, **105**, 219-237.

Stevenson RD, Corbo K, Baca LB & Le QD (1995) Cage size and flight speed of the tobacco hawkmoth *Manduca sexta*. Journal of Experimental Biology, **198**, 1665-1672.

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

Turchin (2003) Complex Population Dynamics: a theoretical/empirical synthesis. *Monographs in Population Biology*, **35**. Princeton University Press, Princeton.

Vekemans X & Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921-936.

Veldtman R (2004) The ecology of southern African wild silk moths (*Gonometa* species, Lepidoptera: Lasiocampidae): consequences for their sustainable use. University of Pretoria PhD thesis.

Veldtman R, McGeoch MA & Scholtz CH (2002) Variability in cocoon size in southern African wild silk moths: implications for sustainable harvesting. *African Entomology*, **10**, 127-136.

Veldtman R, McGeoch MA & Scholtz CH (2004) The parasitoids of southern African wild silkmoths (Lepidoptera). *African Entomology*, **12**, 117-122.

Wahlund S. 1928. Zusammensetzung von Populationen und Korrelation-serscheinungen von Standpunkt der Verebungslehre aus betrachtet. *Hereditas*, **11**, 65-106.

Waldbauer GP & Sternburg JG. (1982) Long mating flights by Hyalophora cecropia. *J. Lepid. Soc.*, **36**, 154-155.

Weir BS & Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD & Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.*, **26**, 375-400.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97-159.

Wright S (1940) Breeding structure of populations in relation to speciation. *American Naturalist*, **74**, 232-248.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114-138.

Wright S (1969) *Evolution and the Genetics of Populations.* Vol 2. *The Theory of Gene Frequencies*. University of Chicago Press, Chicago.