Chapter IV

Phylogeographic patterns of *Scarabaeus (Pachysoma*) (Coleoptera: Scarabaeidae) as inferred from gene genealogies and coalescent theory.

Catherine L. Sole[1], Armanda D. S. Bastos[1, 2] and Clarke H. Scholtz[1]

[1] *Department of Zoology & Entomology, University of Pretoria, Pretoria, 0002, South Africa*

[2] *Mammal Research Institute (MRI), Department of Zoology & Entomology, University of Pretoria, Pretoria, 0002, South Africa*

Running title: Phylogeography of *Scarabaeus (Pachysoma)* (Scarabaeidae: Scarabaeinae).

Abstract

Mitochondrial cytochrome oxidase I (COI) sequence data were used to infer phylogeographic patterns of three species of *Scarabaeus (Pachysoma)*, a group of flightless dung beetles endemic to the arid west coast of southern Africa. Nested clade analysis in conjunction with historical demographic analysis allowed for the inference of historical bottlenecks followed by population expansion in response to climatic oscillations. All three species exhibit high overall haplotype diversity with no shared haplotypes between populations or collecting localities, refugia could, therefore, not be identified. Recent events imply human induced, environmental barriers and reduced vagility as having caused fragmentation, influencing the strong population structure seen in two of the three species. Coalescence for each species was calculated and it was estimated that all three species underwent population expansion within the Pleistocene era, in response to the formation of advective fog. The neighbor-joining trees showed *S. (P.) hippocrates* as having four distinct populations, *S. (P.) gariepinus* having three populations, two in South Africa and one in Namibia while *S. (P.) denticollis* comprised a single population along a dune field continuum in Namibia. AMOVA analysis confirmed the phylogenetic partitioning. Analysis of gene flow revealed a strong degree of south-north movement, consistent with the unidirectional wind regime, with some movement occurring in a southerly direction.

Keywords *Scarabaeus*, mitochondrial DNA, cytochrome oxidase I, Phylogeography, Namaqualand, Namib Desert

Introduction

Species consist of geographically structured populations, many of which have experienced little or no genetic contact for long periods of time due to the limited dispersal abilities of the individuals and/or as a result of habitat discontinuities (Carisio *et al*., 2004). In addition to selective forces, factors that contribute to these associations are past events such as colonisation history and current demography (Juan *et al*., 1998). By examining the variation among populations, their historical associations and the processes of genetic restructuring that may have lead to speciation, can often be revealed (Wright, 1931). Species complexes among geographically isolated populations of polytypic species have great potential for historic inferences (Kirchman *et al*., 2000). These geographically isolated populations represent the extreme of spatial patterning and are therefore of particular interest (Kirchman *et al*., 2000).

Phylogeography is the study of the principles and processes governing the geographical distributions of genealogical lineages, especially those within and amongst closely related species (Avise, 2000). Elucidating the phylogeographic patterns within the species of *Scarabaeus* (*Pachysoma)* will enable us to infer their evolutionary history, re-construct colonisation routes and identify possible refugia. Furthermore, it will be possible to identify and delineate genetically meaningful conservation units, evolutionary significant units (ESU's) and management units (MU's) (Moritz, 1994a; b), within the different species. This information will be useful for developing conservation management recommendations for preserving species of *Scarabaeus (Pachysoma).*

Scarabaeus (*Pachysoma)* represents an excellent group to study the effects of geographic isolation within species. The species are geographically isolated, occurring in pockets of discontinuous populations on coastal sands from Cape Town (33°56'S – 18°28'E) in South Africa to Walvis Bay (22°58'S – 14°30'E) in Namibia. Some of these areas are currently under threat of both anthropogenic and environmental factors. Threats to the habitat of *Scarabaeus* (*Pachysoma)* species come from the removal of the natural vegetation for large scale wheat farming in the south western Cape, commercial development on the west coast for holiday and recreational purposes e.g. Lambert's Bay and Strandfontein, mining for diamonds and other minerals as well as from exotic plant invaders modifying dune systems e.g. Port Jackson (*Acacia saligna*) and Rooikrans (*Acacia cyclops*). Furthermore, some of the species of *Scarabaeus* (*Pachysoma)* are potentially threatened as they are sought after by collectors (Harrison, 1999).

The species of *Scarabaeus (Pachysoma)* range from 2 – 5 cm in length, with *S. (P.) denticollis* and *S. (P.) rodriguesi*, representing the smallest and largest extremes in body size,

respectively. Three species, *S. (P.) hippocrates*, *S. (P.) gariepinus* and *S. (P.) denticollis* where selected for population based analyses as these species exhibited distinct south-north morphological clinal variation (Harrison, 1999).

Cytochrome oxidase I (COI) sequence data are used in the present study for comparisons within and between *S. (P.) hippocrates*, *S. (P.) gariepinus* and *S. (P.) denticollis*. We addressed the following questions: Firstly, to what degree has geographic isolation led to the genetic restructuring between populations of the same species? Secondly, what is the extent of gene flow between populations of the same species and does it correlate with patterns of geographic proximity? Thirdly, from which geographic location did the group originate and how are the populations of each species related to one another? Finally, what are the effective/actual population sizes of the species in question?

Materials and Methods

Population sampling, amplification and sequencing

For all three of the species, we sampled, where possible, a minimum of 10 individuals per designated population. Each of the species was divided into populations based on morphological and distributional data (Table 1). Total genomic DNA was extracted from 176 individuals from the thoracic muscle tissue and amplified using TL2–N-3014 and C1–J–1718 (Simon *et al*., 1994) targeting a 1345 base pair (bp) fragment. Thermal cycling parameters comprised an initial denaturation for 90 seconds at 94°C followed by 35 cycles at 94°C for 22 seconds, 48°C for 30 seconds and 72°C for 90 seconds with a final elongation step at 72°C for 1 min. Amplified COI products were purified from the tube using the High Pure PCR Product Purification Kit (Roche) according to manufacturer specifications. An analysis of the 1197 bp region generated for 46 individuals of the 13 morphological species of *Scarabaeus (Pachysoma)* (Sole *et al*., 2005) revealed that the 5' end of the COI gene was the most parsimoniously informative. Based on this we used only 960 bp from the 5' end of the partial COI gene for the present study. For this reason, each amplicon was sequenced with C1-J-1718 and C1-J-2183 (Simon *et al*., 1994). Sequencing reactions were performed at an annealing temperature of 48ºC with version 3.1 of the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer). Nucleotide sequences were determined through a capillary system on an ABI 3100 automated sequencer (Perkin-Elmer). Sequence chromatograms were visualised and edited in Sequence Navigator (Perkin-Elmer).

Analysis

Phylogenetic Analysis

Hierarchical likelihood ratio tests were performed using Modeltest 3.5 to determine the model of sequence evolution that best fit the data at hand (Posada & Crandall, 1998). Parameters such as the proportion of invariable sites and α parameter of the gamma distribution of rate variation among sites (Yang *et al*., 1994) were calculated in Modeltest. MEGA version 2.1 (Kumar *et al*., 2001) was used to calculate the transition/transversion (ti/tv) ratio. Maximum Parsimony (MP) (Kluge & Farris, 1969) and Maximum Likelihood (ML) (Felesenstein, 1973, 1981) trees were obtained using PAUP* v. 4.08b (Swofford, 1998). For MP trees, we used heuristic searches with tree-bisection-reconnection (TBR) as the branch-swapping algorithm and the nucleotides were treated as unordered characters. The starting tree was obtained via stepwise addition with random addition of sequences with 10 replicates. ML analysis was performed as above, using the values obtained from Model Test but given the computational time required no replicates were performed. The neighbor-joining (NJ) (Satou & Nei, 1987) option in the computer program MEGA was used to re-construct relationships between the populations within the three species with the model selected from Modeltest. Support for all relationships was estimated using 1000 bootstrap replicates (Felsenstein, 1985). Population assemblages within species were ascertained from mid-point rooted trees. As all tree topologies from the different tree drawing options were similar only the NJ trees are presented.

Molecular diversity

Mean nucleotide diversities within each population/assemblage were calculated with Arlequin 2.000 (Schneider *et al*., 2000). Hierarchical structuring of genetic variation was determined using AMOVA (Excoffier *et al*., 1992 new version: Schneider *et al*., 2000), which produces Φ - statistics similar to the F - statistics of Wright (1951; 1965). Φct describes the regional apportionment of genetic variation with respect to all haplotypes, Φsc describes the apportionment of variation within the populations of a given region and Φst characterises the variation between haplotypes in a single population relative to all haplotypes (Barber, 1999). Analyses were performed independently of groupings designated by the authors as well as on the assemblages identified from each neighbor-joining tree, to determine the hypothesis that best fit the data. Levels of significance of Φst - statistics were determined through 10,000 random permutation replicates (Schneider *et al*., 2000).

Historical Population Dynamics

Distance-based methods (mismatch distribution in Arlequin 2.000, Schneider *et al*., 2000), coalescence and maximum likelihood methods (LAMARC version 1.2.2 Kuhner *et al*., 2004; MIGRATE version 1.7.6.1 Beerli & Felsenstein, 1999; 2001) were used to estimate effective population size, exponential growth or shrinkage, time and rate of expansion and migration rates within and among populations/assemblages.

Population growth/decline based on two models

Stepwise Expansion Model

The program Arlequin was used to calculate the mismatch distribution (frequency of pairwise differences) between the haplotypes of a population. This evaluates the hypothesis of recent population growth (Rogers & Harpending, 1992) with the underlying assumption that population growth or decline leaves distinctive signatures on the DNA sequences compared with constant population size. Recent growth should generate a unimodal distribution of pairwise differences, but the exact mode of growth (exponential, stepwise or logistic) cannot be distinguished (Rogers & Harpending, 1992). This distribution is then compared with that expected under a model of population expansion (Rogers, 1995) calculating the estimator expansion time (τ) and the mutation parameter (θ) (Schneider & Excoffier, 1999). A non-linear least squares (Schneider & Excoffier, 1999) approach is used to estimate parameters for the stepwise growth model: $\theta_0 = 2\mu N_0$ (before expansion), $\theta_1 = 2\mu N_1$ (after expansion) and $t = \tau / 2v$ (time of expansion, note $v = m_T\mu$ which is the mutation rate for the entire DNA sequence under study where m_T is the number of nucleotides and μ is the mutation rate). N_0 and N_1 are the effective population sizes of females before and after population expansion respectively.

For the COI data we estimated the mutation rate by following the procedure in Rooney *et al*. (2001). Firstly, the number of nucleotide substitutions per site was estimated by comparing the in-group (classified as one of the three species within this study) with its sister taxa, (which was obtained from the phylogeny found in Sole *et al*., 2005; chapter 2) using the formula $d = (Tv + TvR)/m_T$, where Tv is the number of tranversions between the focal and sister species, R is the ratio of transitions to transversions within the focal species and m_T is the length of the investigated DNA sequence. Secondly, the rate of nucleotide substitution (γ) per site, per lineage, per year was estimated by $\gamma = d/2T$, where T stands for the divergence time of the two compared species (this was estimated in MEGA using Brower, 1994; 2.3 %

pairwise divergence per million years) - (Divergence estimates are under debate and surrounded by much contention. However, the decision to use the molecular clock to estimate divergence times in this study was conservative and used with caution (Graur & Martin, 2004)). Thirdly, it was possible to estimate mutation rate per nucleotide site, per generation (μ) by solving the equation $\mu = \gamma t_{g,}$ where t_g is generation time in years, which in this case was taken to be a single generation per year. The mutation rate per haplotype (v) was calculated by $v = m_T\mu$. Finally, the coalescence time (time to expansion) in generations was calculated by $t = \tau/2v$ (Rogers & Harpending, 1992) and the coalescence time in years was estimated by multiplying t with generation time.

Arlequin estimates approximate confidence intervals for θ_0, θ_1 and expansion time (τ) - which are substituted into the above equations to solve for $N_{0,}$ N_1 and t, respectively - by parametric bootstrapping of 10,000 replicates**.** If population growth applies, the validity of the stepwise expansion model is tested using the same bootstrap approach by a goodness of fit statistic (P), representing the probability that the variance in the simulated dataset is equal to or greater than that seen in the observed dataset. We also computed Harpending's Raggedness Index - R - and its significance in the same manner (Harpending, 1994).

Tajima's (1983) estimate of θ, was estimated in Arlequin, while Fu's (1994a, b) UPBLUE estimate of θ, was estimated by Fu's phylogenetic estimator of θ on line (http://hgc.sph.uth.tmc.edu/cgi-bin/upblue.pl). Tajima's estimate is based on the calculation of the mean number of pairwise differences of the sequences, while Fu's UPBLUE estimate is calculated by incorporating the genealogical information of the sequences (Su *et al*., 2001). Fu's UPBLUE estimate puts emphasis on recent mutations, revealing recent population processes, while Tajima's estimate puts more weight on ancient mutations, reflecting past population trends (Su *et al*., 2001). As $\theta = 2N\mu$ for the mitochondrial genome, the ratio of population size change is correlated with θ given a constant mutation rate (μ). Comparing Tajima's with Fu's UPBLUE estimate will give an idea of population size change in recent time. In addition, Fu's (1997) *F*s test of neutrality was carried out in Arlequin. Although the *F*s was originally designed as a test of neutrality, it has utility as an estimator of population growth (Smith & Farrell, 2005). The *F*s value tends to be negative if there is an excess of recent mutations (i.e. mutations that occur in a small number of individuals). A large negative value indicates an excess of recent mutations an outcome that can be caused by either population growth and/or selection (Su *et al*., 2001).

Exponential Expansion Model

The program LAMARC version 1.2.2 was used to estimate population parameters based on coalescence and maximum likelihood methods. LAMARC uses the Markov Chain Monte Carlo (MCMC) approach with the Metropolis-Hastings genealogy sampler to search through genealogies of which samples are taken at intervals from which to calculate a maximum likelihood of theta (θ). LAMARC calculates estimates of population exponential growth or decline, population size, recombination and migration rates. Ten short chains of 500 steps each, which were followed by 2 long chains of 10,000 steps, sampled every 20^{th} step and 4 heated chains were run at temperatures of 1, 1.3, 1.6 and 2. Each run was repeated 4 times to ensure consistency of results. We used LAMARC to estimate theta ($\theta = 2N_f(\mu)$) and growth, given as g ($\theta_{now} = \theta_t^{(-gt)}$) (or decline) rates for populations of the three species.

Migration

In order to evaluate the relationships among populations within species we used the program MIGRATE version 1.7.6.1 to estimate both effective population sizes ($\theta = 2\mu N_f$) and migration rates ($M = 2mN_f$) (Beerli & Felesenstein, 1999; 2001). MIGRATE searches through genealogy space using a likelihood ratio test and coalescent theory to estimate these parameters. It makes use of the MCMC approach with the Metropolis Hastings Green algorithm. It assumes constant effective population sizes for each population, but allows various effective population sizes for different populations (Zheng *et al*., 2003). Values of theta (θ) were estimated from the Fst-calculation. When using MIGRATE on our data we used the population subdivisions from Table 1 for each species running each analysis separately. Ten short chains with 1000 sampled genealogies and three long chains with 1,000,000 sampled genealogies each, were run. Heating was set to be active with four heated chains at temperatures of 1.00, 1.33, 1.66 and 2.00. Five runs were repeated for each species' dataset to check for consistency.

Network Estimation and Nested Clade Analysis

A haplotype network was estimated using statistical parsimony (Templeton *et al*., 1992) in TCS version 1.18 (Clement *et al.*, 2000). The method links haplotypes with smallest number of differences as defined by a 95 % confidence limit. Loops (= reticulations) in the network, which result from homoplasy in the data, were broken in accordance with the predictions derived from coalescent theory: i) common haplotypes are more likely to be found at interior

nodes of a cladogram, and rare haplotypes at the tips; ii) haplotypes represented by a single individual are more likely to connect to haplotypes from the same population than to haplotypes from different populations (Crandall & Templeton, 1993; Posada & Crandall, 2001).

Nested Clade Phylogeographical Analysis (NCPA; Templeton et al., 1995) was used to infer population history in each of the three species. NCPA first tries to reject the null hypothesis of no association between haplotype variation and geography and then attempts to interpret the significant associations (Crandall & Templeton, 1993; Templeton *et al*., 1995). The NCPA was constructed by hand, based on the parsimony network following the rules given in Templeton & Singh (1993). Such a nested design treats haplotypes as "0-step clades," groups of haplotypes separated by a single mutation as "1-step clades," groups of "1-step clades" separated by a single mutation as a "2-step clade" and so on.

GEODIS v2.0 (Posada *et al*., 2000) was used to calculate NCPA distance measures and their statistical significance. This method uses geographical distances between sampled populations to calculate two basic statistics: Dc (clade distance) and D*n* (nested clade distance). Dc measures the average distance of all clade members from the geographical centre of distribution. D*n* measures how widespread a clade is in relation to the distribution of its sister clades within the same nesting group. Random geographical distribution in coalescent theory allows one to distinguish between tip (with one connection to the remaining network) and interior (with two or more connections) clades by permutational tests of which we performed 10,000. We used Templeton's (2004) updated inference key to deduce the cause of significant associations between haplotypes. This would allow us to distinguish between historical (fragmentation, range expansion) and current (gene flow, genetic drift, system of mating) processes responsible for the observed patterns of genetic variation (Templeton *et al*., 1995).

The Mantel test (Mantel, 1967) was used to determine significant associations between genetic distances obtained in MEGA and geographic distances between the designated populations from Table 1. One thousand randomised permutations were performed using Mantel version 2.0 (Liedlof, 1999).

Table 1. Total number of individuals for each population and subpopulation. The populations were designated according to geographic and morphological differences.

Species	Population location	Population individuals	Subpopulation individuals
S. (P.) denticollis	**Population - Koichab Pan (KP)**	**13**	
	Koichab Pan		11
	Luderitz		2
	Population - Namib Rand (NR)	**8**	
	Tok Tokkie Trails		8
	Population - Gobabeb (GB)	**11**	
	Gobabeb - 5km SE Homeb		11
S. (P.) gariepinus	**Population - Langhoogte to Kommagas Road (LK)**	**12**	
	Langhoogte to Kommagas Road		12
	Population - Holgat River (HR)	**17**	
	40km N Port Nolloth - Holgat River		17
	Population (Namibia) - Hohenfells (HF)	**13**	
	Hohenfells Dunes		13
	Population (Namibia) - Daberas/Obib Dunes (DO)	**11**	
	Road from Daberas to Obib Dunes		11
	Population (Namibia) - Klinghardts Mountains (KM)	**14**	
	Klinghardts Mountains		14
S. (P.) hippocrates	**Population - Cape Town/Lamberts Bay (LA)**	**13**	
	10km West of Leipoldtville		13
	Population - Olifants/Green River (BV/KK)	**9**	
	Kommandokraal Farm (KK)		5
	Koekenaap (BV)		4
	Population - Green/Buffels River (SK)	**20**	
	Kleinsee - Sandkop		20
	Population - Buffels River/Port Nolloth (PN)	**11**	
	1km North of Port Nolloth		11

Chapter IV (a)

Genetic structure, phylogeography and demography of *S. (P.) hippocrates* based on inferences from Cytochrome Oxidase I

Introduction

The long-term persistence of many species is threatened by the loss of their natural habitat caused by human activities. Remaining habitats are often small and isolated from each other by less suitable habitat e.g. settlement areas, agriculture and roads. Isolation of local populations and reduction of suitable habitat are potential negative effects of a fragmented landscape. Causal factors of fragmentation may be human induced or environmental. Isolation is of particular significance when considering taxa with limited dispersal ability as they face an increased risk of extinction due to demographic and genetic factors.

Scarabaeus (Pachysoma) hippocrates represents a good species to evaluate the effects of geographic isolation caused in some instances by natural barriers and in others by human activities. *S. (P.) hippocrates* is one of the larger species of the flightless *Scarabaeus (Pachysoma)* occurring from Bloubergstrand (S33°48' – E18°27'), Cape Town, to Port Nolloth (S29°15' – E16°53') in Namaqualand. *S. (P.) hippocrates* prefer vegetated soft to firm sand of coastal hummocks and hillocks, the periphery of dune systems, riverbeds and banks (Harrison, 1999). The species is shown to exhibit a gradual morphological cline along this distribution with populations isolated by both natural and non-natural barriers. Habitat modification threatens certain populations of *S. (P.) hippocrates*, specifically those occurring at Port Nolloth, where the town is expanding into their habitat, and areas around Leipoldtville and Kommandokraal, where farming communities exist.

Materials and Methods

See body of Chapter 4. The localities where *S. (P.) hippocrates* were collected for this study are represented in Figure 1.

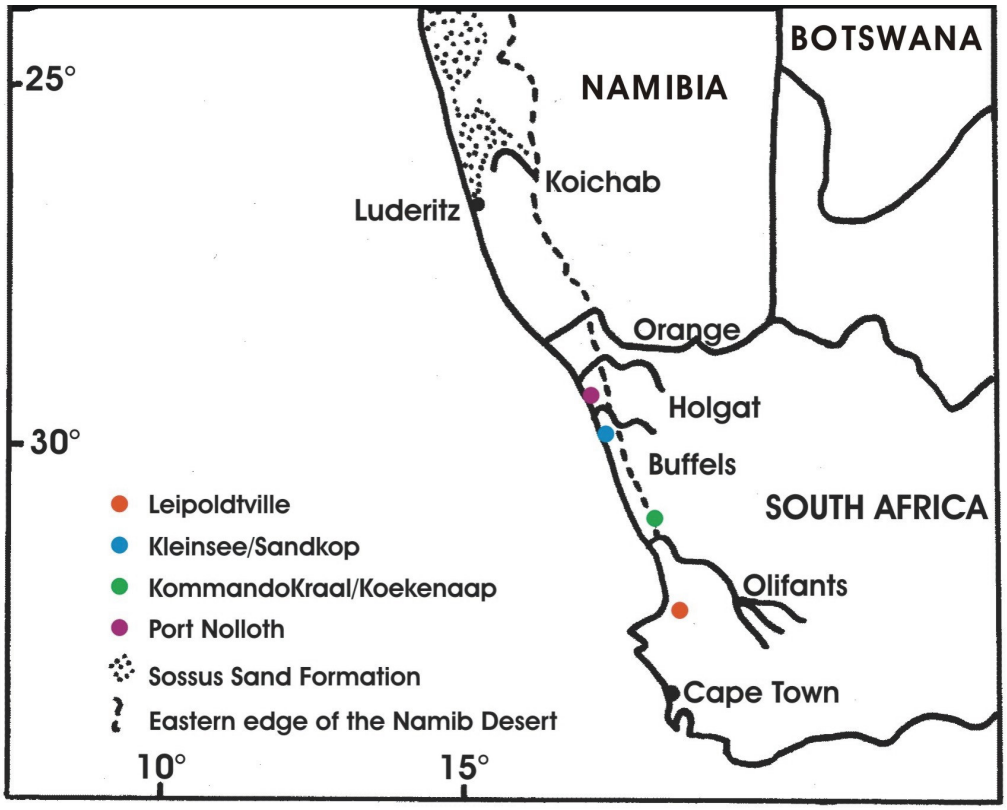


Figure 1. Collecting sites/localities, within South Africa, of *S. (P.) hippocrates* used in this study.

Results

Phylogenetic and molecular diversity

Population statistics

Overall we collected 53 individuals for molecular characterisation (GenBank Accession Numbers AY965154 – AY965206: Appendix 2). The sequences exhibited an overall A/T bias of 69.5 %. Un-corrected pairwise distances ranged from 1 to 12.3 % (data not shown). The model best fitting the data was the Transversional model with a gamma distribution of 0.0012 (TvM assuming unequal base frequencies and different transition and transversion rates, (Posada & Crandall, 1998; Nahum *et al.*, 2003). Individual LAPH13 was initially removed from the data analysis as it appeared as an outlier/out-group to *S. (P.) hippocrates* (tree not illustrated). However, as removal of the individual did not alter the relationships within the neighbor-joining tree it was included in all subsequent analyses. It may have appeared as an out-group as there is a known species complex between *S. (P.) hippocrates* and *S. (P.) glentoni* (Sole *et al.*, 2005; Chapter 2). The neighbor-joining tree can be seen in

Figure 2 (drawn in MEGA using the TvM model), indicates four distinct groups into which each population designated from Table 1 could be grouped. Parsimony and Maximum Likelihood trees exhibited similar topologies to the neighbor-joining tree with and without LAPH13 (results not shown).

Table 2 shows molecular diversity statistics obtained within each population as well as an overall estimate for the species as a whole. Mean nucleotide diversity was high for the Leipoldtville population and an order of magnitude lower in the other two populations where estimates were obtained. Port Nolloth had a single haplotype therefore estimates for this population were 0. Thirty-one haplotypes were identified among the four populations with each population having its own unique set of haplotypes specific to each geographic region. Accordingly, haplotype diversity expressed over the complete sample was relatively high (H = 0.948 +/- 0.004).

Genetic differentiation among populations

An analysis of molecular variance (AMOVA) was performed to estimate the fixation index using the optimal model of sequence evolution identified above. Analyses were performed with the species as a single group with each population defined as a region. The results of AMOVA revealed that 84.88 % of the variance resulted from among population differences while 15.12 % of the variation could be attributed to within population variation. The fixation index was high and significant at 0.849 ($p < 0.001$) indicating strong genetic differentiation between the four populations (Table 3).

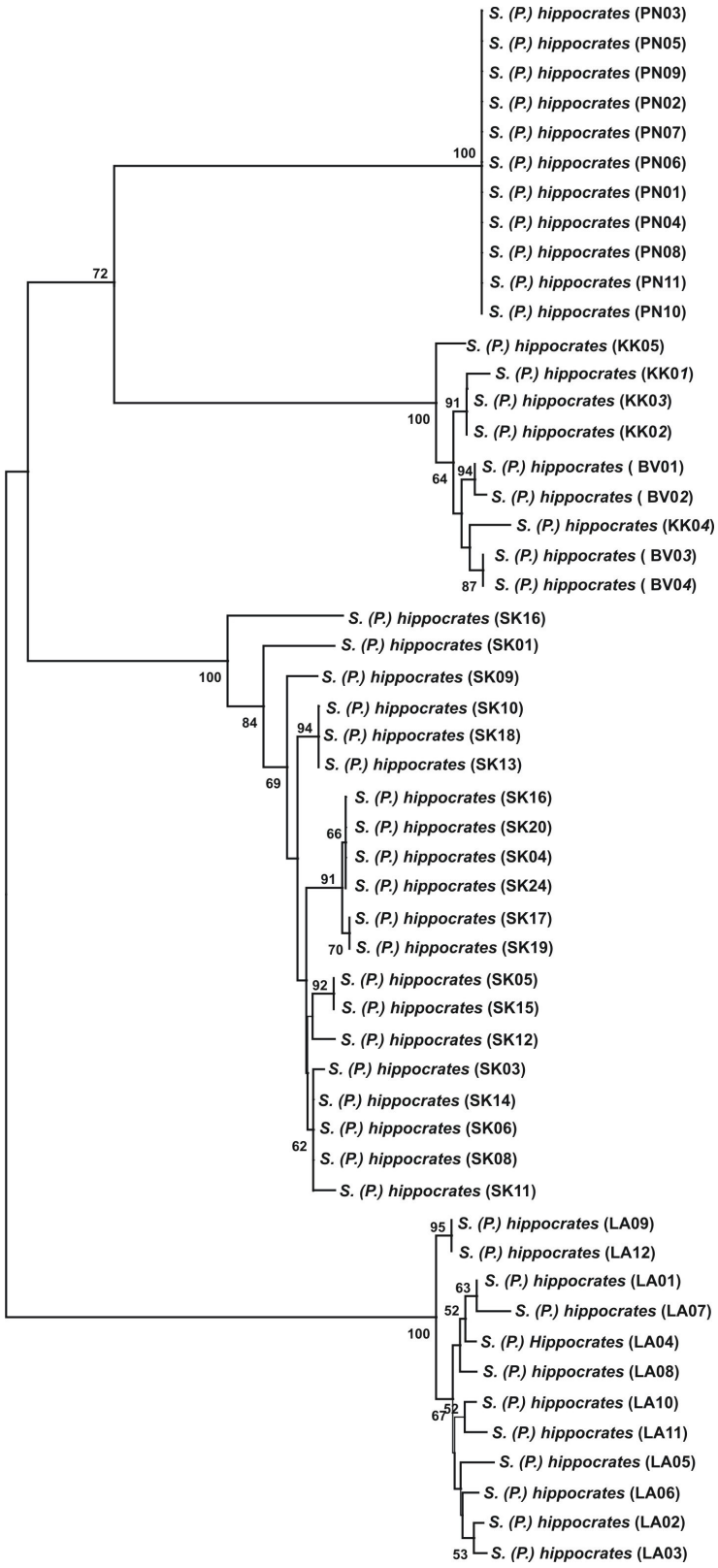


Figure 2. Mid-point rooted neighbor-joining tree for COI sequence data of *S. (P.) hippocrates*. Bootstrap values below 50 % were removed

Table 2. Summary of general nucleotide diversity statistics over the 960 bp region of *S. (P.) hippocrates*

Species	Assemblage	*N*	No. of haplotypes	Haplotype diversity	Nucleotide diversity	% Pairwise divergence	Variable Sites (V)	Parsimoniously Informative Sites (PI)	Singletons (S)
S. (P.) hippocrates	Leipoldtville	13	12	0.987 (0.035)	0.056 (0.029)	0.002 - 0.008			
	KommadoKraal/Koekenaap	9	7	0.944 (0.070)	0.007 (0.004)	0.001 - 0.008			
	Kleinsee - Sandkop	20	11	0.926 (0.034)	0.008 (0.004)	0.001 - 0.024			
	Port Nolloth	11	1	0	0	0			
Total		53	31	0.948 (0.021)	0.058 (0.003)	0.01 - 0.123	200 (20.83%)	133 (13.85%)	67 (6.98%)

[$] V, PI and S were only estimated for the overall dataset

Table 3. Summary of Fst statistics produced by AMOVA (Excoffier *et al*., 1992) for *S. (P.) hippocrates.*

Species		Φst	%	P
S. (P.) hippocrates	Among populations		84.88	<0.001
	Within populations		15.12	<0.001
	Fixation index	0.849		

[b] P values were determined from 10000 random permutations.

Demographic patterns based on the Stepwise and Exponential Expansion Models

Stepwise expansion model

Both tree topology (Fig. 2) and the mismatch distribution (Fig. 3) indicate recent sudden demographic expansion for each of the three populations. The branches of the tree within each population are small and similar in length suggesting a recent expansion in population size and geographic range. The haplotypic data (mismatch distributions) showed similar uni-model curves as expected in accordance with historically expanding populations. Both the variance (sum of the squared deviation (SSD)) and Harpendings Raggedness Index (HRI) suggested that the curves did not differ significantly under a model of population expansion.

Time of divergence

Using the 960 bp of the COI sequence we calculated the average number of nucleotide substitutions per site (*d*) and obtained a value of $d = 0.073$. The divergence time between *S. (P.) hippocrates* and *S. (P.) endroedyi,* which were shown to be sister taxa (see Sole *et al*., 2005; Chapter 2), was estimated to have occurred 2.3 million years ago. This gives the estimate of nucleotide substitutions per site per lineage per year (λ) to be $0.073/(2 \times 2{,}300{,}000) = 1.5 \times 10^{-8}$. The mutation rate per nucleotide site per generation (μ) would be 1.5×10^{-8}. The coalescence time was calculated from the (τ) values in Table 4a, in generations for each population (see below), using a mutation rate per haplotype (v) of 1.4×10^{-5}.

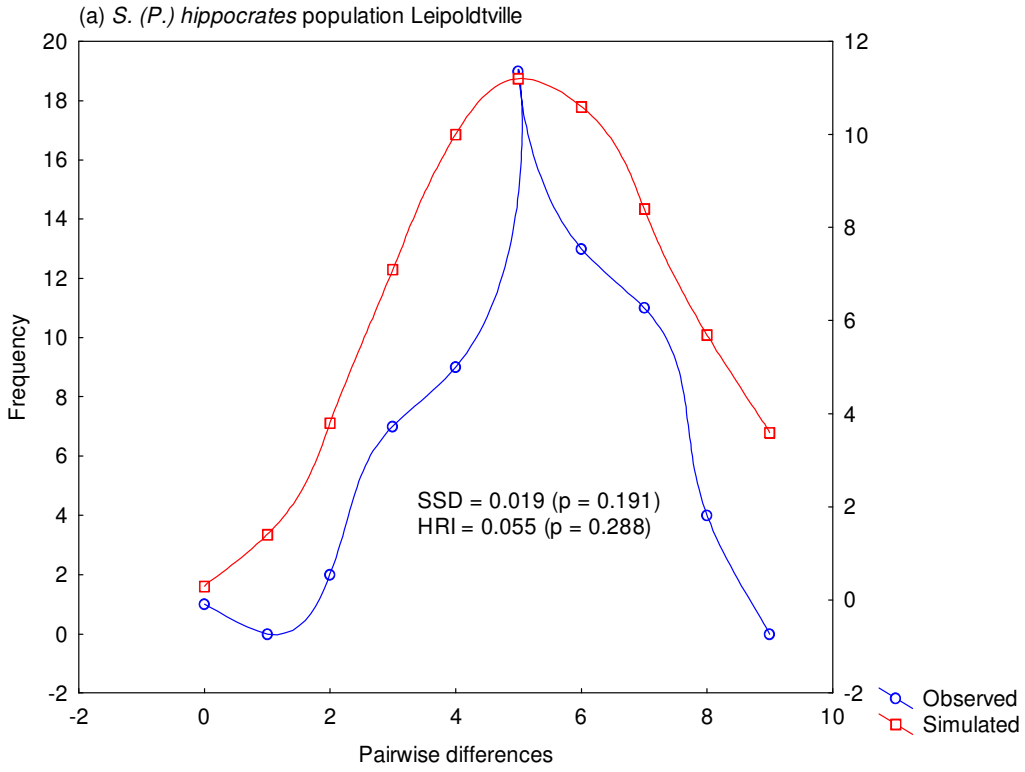
Based on the above calculated mutation rate and τ values of 5.655, 5.653 and 6.188 (Table 4a) the expansion of the Kommandokraal/Koekenaap and the Leipoldtville populations appeared to have been around 202,000 generations/years ago, while the Kleinsee - Sandkop population appeared to have undergone expansion at around 221,000 years/generations ago. Estimated effective female population size after expansion (N_1) was an order of magnitude higher than before expansion (N_0) for all three populations with the Kommandokraal/Koekenaap population having the lowest N_0 of approximately 100,000 individuals.

Exponential expansion model

The exponential expansion model indicates rapid growth for two of the populations namely, Kommandokraal/Koekenaap and Kleinsee - Sandkop having positive 'g' values (Table 4b). The Leipoldtville population, however, shows a strong negative 'g' value indicating overall population decline. Effective female population size estimated from theta (θ) did not show

large differences between populations but showed an overall tendency towards increasing population numbers.

The UPBLUE estimates for all the populations are an order of magnitude smaller than the Tajima's estimates (UPBLUE/Tajima; Table 5), except in the KommandoKraal/Koekenaap population, indicating that the Kleinsee - Sandkop and Leipoldtville populations are declining, whilst KommandoKraal/Koekenaap has a stable population. This is supported by Fu's *F*s statistics (Table 5) which are negative, but not significantly so for all the populations, indicating that the populations are experiencing minimal to no recent expansion.



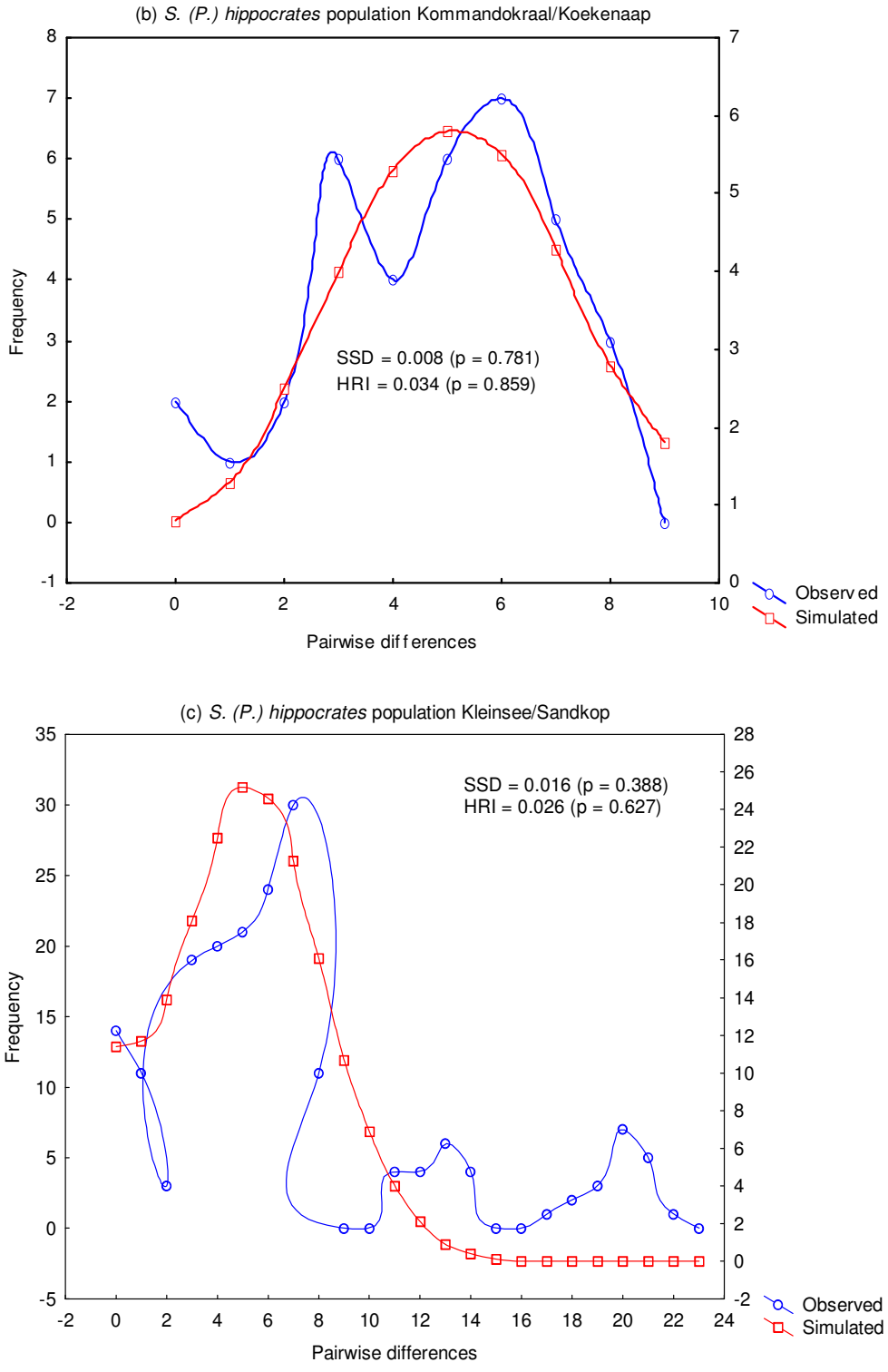


Figure 3(a–c). Mismatch frequency distributions of pairwise nucleotide differences for three of the four populations of *S. (P.) hippocrates*, with sum of the squared deviation (SSD) and Harpendings Raggedness Index (HRI) represented on the graphs.

Table 4. Estimated parameters for (a) Stepwise and (b) Exponential Expansion models of *S. (P.) hippocrates*

a) Stepwise Expansion Model

		Stepwise Expansion Model			
Species	**Populaion**	**τ**	**$\theta_0 = 2\mu N_0$**	**$\theta_1 = 2\mu N_1$**	**$t = \tau/2v$**
S. (P.) hippocrates	Leipoldtville	5.655	0	6656.250	202,000
	KommadoKraal/Koekenaap	5.653	0.003	52.598	201,900
	Kleinsee - Sandkop	6.188	0.084	16.449	221,000
	Port Nolloth	0	0	0	0

b) Exponential Expansion Model

		Exponential Expansion Model		
Species	**Population**	**$\theta = 2\mu Nf$**	**g**	**N_f**
S. (P.) hippocrates	Leipoldtville	0.0252	-19.439	840,000
	KommadoKraal/Koekenaap	0.0293	765.972	980,000
	Kleinsee - Sandkop	0.0169	74.623	564,000
	Port Nolloth	0	0	0

Table 5. Summary of estimations of Tajima's estimate, Fu's UPBLUE and Fu's *F*s statistic of *S. (P.) hippocrates*

Species		**Leipoldtville**	**KommadoKraal/Koekenaap**	**Kleinsee - Sandkop**	**Port Nolloth**
S. (P.) hippocrates	Tajima's estimate	40.078	5.476	9.839	0
	Fu's UPBLUE	0.291	5.267	0.020	0
	UPBLUE/Tajima	0.007	0.962	0.002	0
	Fu's *F*s	-1.236 (ns)	-1.366 (ns)	-0.604 (ns)	0

[$] **ns = non-significant**

Migration

The Port Nolloth population was removed from the migrate analysis after a number of runs as there appeared to be confusion as to where movement of the individuals was occurring. This was thought to be due to the fact that the Port Nolloth population consisted of a single haplotype. Once removed, MIGRATE showed movement from the southern populations (Leipoldtville and Kommandokraal/Koekenaap) into the northern Kleinsee - Sandkop population (Fig. 4).

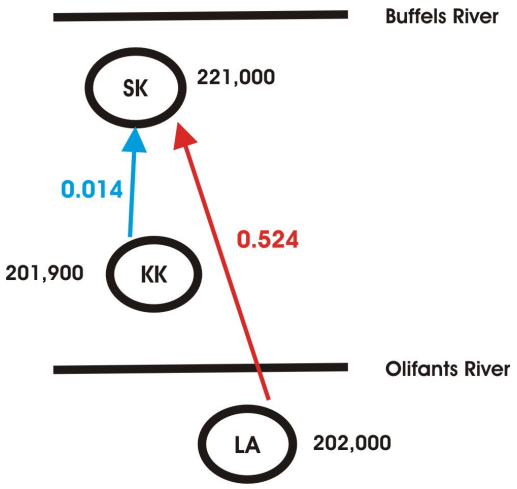


Figure 4. A schematic representation of the migration of individuals between populations of *S. (P.) hippocrates*. The coloured arrows indicate the direction of movement while the numbers in the same colour represent an approximation of the number of individuals moving/generation. The numbers in black represent coalescent times of the populations. (Abbreviations are as follows: LA = Leipoldtville, KK = KommandoKraal/Koekenaap, SK = Kleinsee - Sandkop)

Phylogenetic haplotype relationships

Using statistical parsimony the maximum number of mutational steps between two haplotypes, not including homoplasious changes and based on a probability of 95 %, was 13 mutational steps (as seen in Figure 4). The parsimony network was resolved except for the presence of two reticulations, which were broken. Haplotypes were collapsed for the drawing of the cladogram but not in the total nesting structure, which was used to calculate the contingency test values. It was not possible to determine a single root for the entire cladogram. There were three disjointed portions that could not be linked with 95 %

confidence, namely clades 6-1, 6-2 and 11-1. Furthermore, haplotypes LA13, PN01 and SK16 were separated from all other haplotypes by many mutational steps (14 - 109). Clades 6-1 and 11-1 were linked by 60 mutational steps while haplotype PN01 linked to these two clades by 64 mutational steps. Haplotype SK16 and clade 6-2 linked into the entire network with 65 mutational steps. Haplotype LA13 linked into the network with 109 steps. The network confirms the results seen in previous analyses in that there are relatively high divergence levels between the four populations, as the three disjointed portions represent the haplotypes from a single population (Table 1) with PN01 being representative of the single haplotype from 11 individuals from the Port Nolloth population.

Nested clade analysis did not reject the null hypothesis at the lower nesting levels, due to the fact that these nesting levels did not have haplotypes from different geographic regions in them. However, the contingency test showed significant geographical association of genotypes contained within haplotype groups 12-1, 13-1 and 14-1 (Table 6). All the inference events correspond to allopatric fragmentation of the populations, which would appear true as the populations are isolated by natural barriers (rivers), human habitation, farming areas and mining activities. There is no support for secondary contact between the populations as no shared haplotypes are contained between the populations.

The Mantel test revealed no significant association between genetic and geographic distances ($g = 1.282$, $r = 0.458$, $p > 0.05$). This indicated that an increase in geographic distance did not necessarily correlate with a greater degree of genetic distinction.

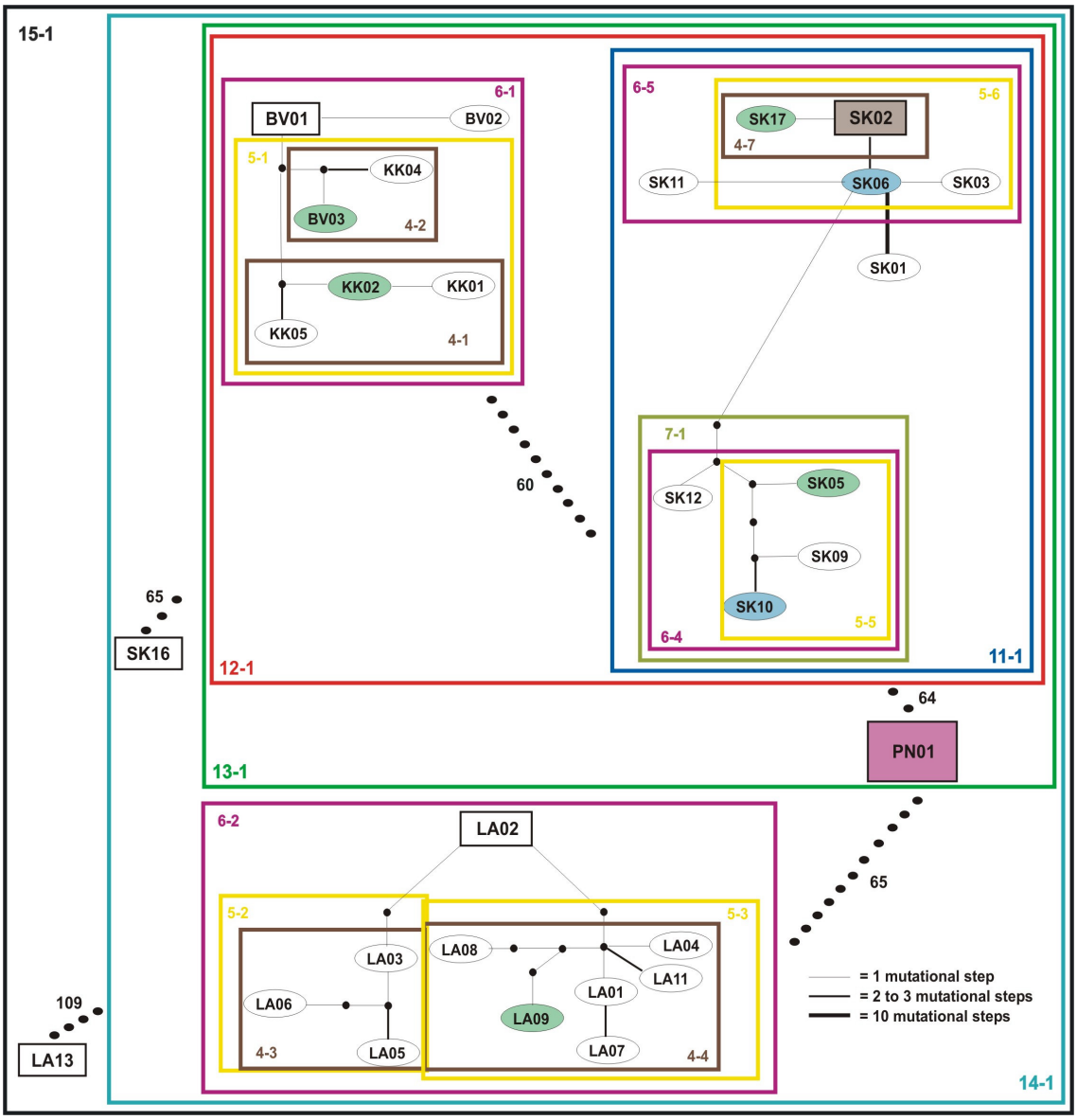


Figure 5. Statistical parsimony network and associated network design. Haplotypes are designated by letters, which represent the population from where the haplotype came (as seen in Table 1) and numbers representing the individual sequenced. The thickness of the connecting line corresponds to the number of mutational steps. Ancestral haplotypes are represented by squares. The relative frequency at which the different haplotypes occurred is indicated by different background colours, haplotypes with pink background occur 11 times, the grey background four times, a blue background occurred three times and a green background occur twice. The nested clades are represented from 4 - step clades upwards, as the mutations were only collapsed for drawing of the cladogram and not for calculation of the nested clade. Dotted lines represent alternative ambiguous connections.

Table 6. Geographical nested clade analysis for *S. (P.) hippocrates* as inferred from Templeton (2004).

Haplotype group	χ^2	P-value	Inference chain	
Clades within 12-1	28.000	<0.001	1-19-No	Allopatric fragmentation - also observed at high clade level
Clades within 13-1	39.000	<0.001	1-19-No	Allopatric fragmentation - also observed at high clade level
Clades within 14-1	53.333	<0.001	1-19-No	Allopatric fragmentation - also observed at high clade level

Discussion

Demographic patterns

S. (P.) hippocrates exhibits a high degree of genetic polymorphism as seen by the high overall haplotype diversity. Both the Stepwise and Exponential Expansion models show a strong inclination for population growth from an ancestral bottlenecked population. As expected, the four populations show appreciable mtDNA divergence in that strong support for all four of the designated populations was obtained. Strong genetic structure between the populations is clear in the highly significant fixation index value, as well as the high maximum pairwise divergence value of 12.3 % (Table 2), indicative of a relatively long historical separation.

The Mantel test shows no association between geographic and genetic distances hence distance could be ruled out as the defining factor for the distinct genetic isolation present. The most probable cause of genetic discontinuities displaying geographic disorientation could therefore be attributed to extrinsic barriers to gene flow. Extrinsic factors contributing to the genetic structure seen in *S. (P.) hippocrates* would be the Olifants and Holgat Rivers, Port Nolloth (the town), large tracts of areas on which farming occurs, as seen in the Leipoldtville and Kommandokraal areas, and mining at Kleinsee - Sandkop. This is clearly illustrated by the fact that each of the factors mentioned above relates to a specifically identified population. NCPA supports this by inferring allopatric fragmentation at all significant levels, indicating that this species may be on its way to speciation.

Previous studies dealing with other insect taxa have shown that speciation events have occurred in the sandy accumulations of river mouths in Namaqualand e.g. Thysanura (Irish, 1990). Harrison (1999) showed quite clearly that *S. (P.) endroedyi* and *S. (P.) glentoni* speciated around the Olifants River mouth with *S. (P.) endroedyi* occurring north of the Olifants River and *S. (P.) glentoni* south of the Olifants River. Both these species occur sympatrically with, and are sister to, *S. (P.) hippocrates* (Harrison & Philips, 2003; Sole *et al*., 2005). The Holgat River has also been shown to be the boundary of the northern distribution of *S. (P.) hippocrates* (Harrison *et al*., 2003). Anthropogenic influences can clearly be seen in the Port Nolloth population where the town is encroaching into the coastal dunes and destroying much of the available habitat. It would appear that anthropogenic factors, occurring over the last 100 years, affect Leipoldtville and Kommandokraal/Koekenaap populations in that individuals occur on disturbed farmland while the Kleinsee - Sandkop population is affected by mining in the area.

Rates of migration were estimated in an attempt to infer historical movements of species. Movement appeared to be in a south-north direction consistent with the unidirectional wind regime along the west coast. The Kleinsee - Sandkop population appears to have undergone expansion earlier than the other two populations, which is in contrast to the migration estimates as these show the Kleinsee - Sandkop population to be receiving individuals from the other two populations as opposed to movement from the Kleinsee - Sandkop population. However, as the expansion times between the populations appear small with the difference between the earliest and the latest coalescent events being as little as 18,000 years ago it would appear that fragmentation of these populations occurred over a similar time period after movement in a northerly direction. The coalescent events appear, therefore to, have occurred at approximately 200,000 years ago thus dating back to the late Pleistocene.

The modern semi-arid environment and winter rainfall within the south-western Cape seen today dates back to the Pliocene. Fossil pollen studies indicate that two invasions of temperate rain forest, and two wet intervals occurred between 33,000 and 45,000 years ago (van Zinderen Bakker, 1975). These climatic oscillations could have caused the historical population expansions seen in the genetic signal. The wet intervals were considered colder periods and the transitional Namib, the area between the Olifants and Orange Rivers, and the southern Namib, presently covered by gigantic dunes, would therefore have received more rain during this period. Increased rainfall would have resulted in stream rejuvenation causing an increase in the sediment source and giving rise to dune plumes occurring at the mouths of many ephemeral rivers along the west coast (van Zinderen Bakker, 1975). This highlights the importance of river mouths as barriers to gene flow and areas that could be designated as refugia for population expansion events of flightless species. These Namaqualand dunes, previously stabilised by vegetation, are extremely long and narrow and presently show signs of being overridden by shifting un-vegetated barchanoid dunes, fed by the erosion of the older coastal dunes (Tankard & Rogers, 1978).

Actual areas of refugia are difficult to identify, as there are no shared haplotypes between populations indicative of a possible ancestral population. Ancestral haplotypes are thought to be shared between, and widespread among populations (Avise *et al.*, 1987).

Current population trends

Scarabaeus (Pachysoma) hippocrates occupy the southernmost area of the total *Scarabaeus (Pachysoma)* distribution. Namaqualand is a winter-rainfall desert covering some 50,000km^2

- characterised by predictably low rainfall, and mild seasonal temperature changes (Cowling *et al*., 1999; Colville *et al*., 2002) - and exhibits a high degree of endemism in both fauna and flora. Namaqualand has not only recently been subject to extensive commercial and subsistence grazing, resulting in significant vegetation change, soil erosion and the loss of primary productivity (Colville *et al*., 2002), but also over past geological time major climatic oscillations have occurred. A combination of these factors could, therefore, have structured this species. Population fragmentation over geological time, leads to genetic divergence of populations, while human induced fragmentation acts over a short period of time eroding genetic diversity. The entire population has been fragmented by environmental factors over time which have not decreased the genetic divergence. However, within the same species, the Port Nolloth population shows a distinct loss of genetic diversity through human induced fragmentation. This indicates that both population- and human- induced fragmentation have affected this species as a whole.

Although both the stepwise and exponential expansion models indicate strong population growth these increases in population sizes may be misleading. There is no available census data for this species but considering current trends of species records and habitat destruction it is far easier to infer that population numbers are declining (Moya *et al*., 2004) and that *S. (P.) hippcrates* may be extinct in areas where it once occurred. This is supported to some extend by the UPBLUE/Tajima estimates that show that two out of four populations are in fact declining. However, it should be borne in mind that genetic signatures of population growth can be misleading (Lavery *et al*., 1996). Fu's UPBLUE estimator of θ and *F*s statistics may be an indicator that significant population growth is not occurring but recent demographic events may be masked by earlier events. The genetic signal of growth could, therefore, be an artefact of past demographic population increase as would have occurred during the Pleistocene (Lavery *et al*., 1996). It has been shown that after rapid population growth, subsequent periods of decline would have no great effect on the initial pattern of growth unless there is a major prolonged bottleneck or until equilibrium is approached (Lavery *et al*., 1996). As many species are not in equilibrium due to past demographic events and the more recent events are undetectable, the results we obtain may be misleading (Lavery *et al*., 1996).

Summary statistics show that *S. (P.) hippocrates* is a genetically and geographically well structured species. Migration estimates show gene flow to be unidirectional from the south to the north. Genetic diversity indicates that distinct genetic variability exists within

and between the populations. Genetic patterns can be related to both past geological events as well as recent fragmentation events.

In this case the anthropogenic and environmental forces discussed above are not mutually exclusive and the combination thereof provides a plausible explanation for the complex population demographic structure seen today.

Chapter IV (b)

Genetic structure, phylogeography and demography of *S. (P.) gariepinus* based on inferences from Cytochrome Oxidase I

Introduction

Scarabaeus (Pachysoma) gariepinus are distributed from the Buffels River (S29°33' – E17°24') in South Africa to the Agub Mountain (S26°59' – E15°58') in Namibia. Interestingly, they occur both south and north of the Orange River with their distribution covering two distinct biomes, Namaqualand, south of the Orange River, and the Namib Desert, north of the Orange River. Many ecological factors are found to influence distributional patterns of arthropods. These include temperature, rainfall, sand characteristics and availability of food among others. Limited vagility and narrow ecological and physiological tolerances may have promoted the present day distribution of *S. (P.) gariepinus*. The presence of this endemic species in the southwest arid regions, the fact that they exhibit south to north morphological clinal variation, are flightless with unique biology, and occur on either side of the Orange River warrants investigation into their biogeography.

Size, elytral sculpture, indument and size of the mesepisternal protuberance were found to vary within and between localities (Harrison, 1999). The populations south of the Orange River are characterized by smaller body size and red indument, while the Namibian populations are generally larger with their indument stained white to grey.

Materials and Methods

See general introduction to chapter. Details of specimen collection sites can be seen in Figure 6.

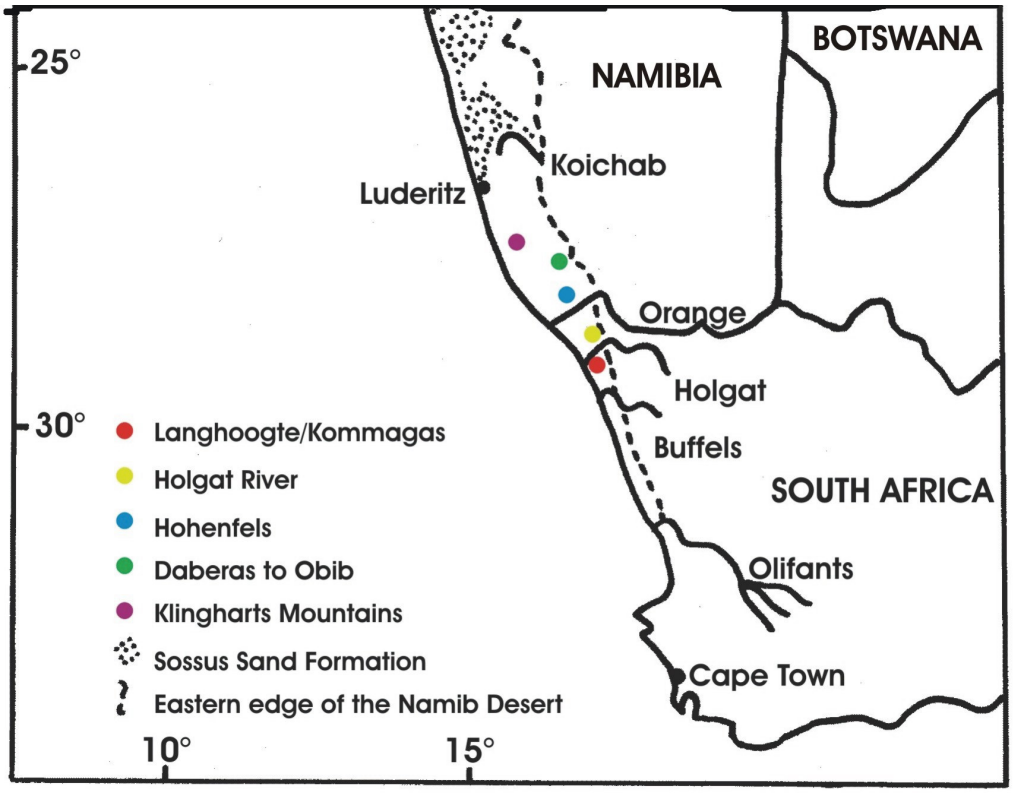


Figure 6. Localities in both Namibia and South Africa where *S. (P.) gariepinus* was collected for this study

Results

Phylogenetic and Molecular diversity

Population statistics

A total of 67 individuals were used for molecular characterisation (GenBank Accession numbers AY965087 – AY965153). The sequences exhibited an overall A/T bias of 69.10 %. Gamma distribution for the data was estimated at 1.0041 with the proportion of invariable sites being 0.7627, the transition/transversion ratio was 6.6 and the model best fitting the data selected by Modeltest was Tamura-Nei. The neighbor-joining tree of 67 individuals (Fig. 7) revealed three distinct assemblages. The first assemblage (labelled N; Fig. 7) had 100 % bootstrap support and consisted exclusively of individuals from the three populations in Namibia (Table 1). The second and third assemblages (labelled LK and HR; Fig. 7) had 100 % and 99 % bootstrap support, respectively and consisted of individuals from each of their respective populations, Langhoogte/Kommagas and Holgat River (Table 1). These three assemblages where each treated as distinct populations in further analyses. (Both the

Parsimony and Maximum Likelihood trees exhibited similar topologies to the neighbor-joining tree; data not presented).

Table 7 shows the molecular diversity statistics for each designated population from Table 1, the population designations from the neighbor-joining tree and the species as a whole. The sequenced COI fragment defined 62 unique haplotypes among the 67 individuals investigated. Accordingly, haplotype diversity expressed over the complete sample was very high (H = 0.997 +/- 0.004) (Table 7). All three assemblages contained numerous different haplotypes and no evidence was found for certain haplotypes being specific to a geographic region.

Genetic variation among populations

Mean nucleotide diversity was calculated across the three populations of the neighbor-joining tree (Table 8). The results of AMOVA revealed that differences among the three defined groups accounted for 49.6 % of the variance ($\Phi ct = 0.496$; $p = 0.001$). A high and significant Φst value of 0.704 ($p = 0{,}001$) indicated strong genetic structure between the three designated populations. Pairwise comparisons between the Φst therefore clearly support the distinctiveness of three populations. The remaining variation could be attributed to $\Phi sc = 0.415$ (among group within population variation) which was significant ($p = 0.001$), accounting for 20.9 % of the overall variation.

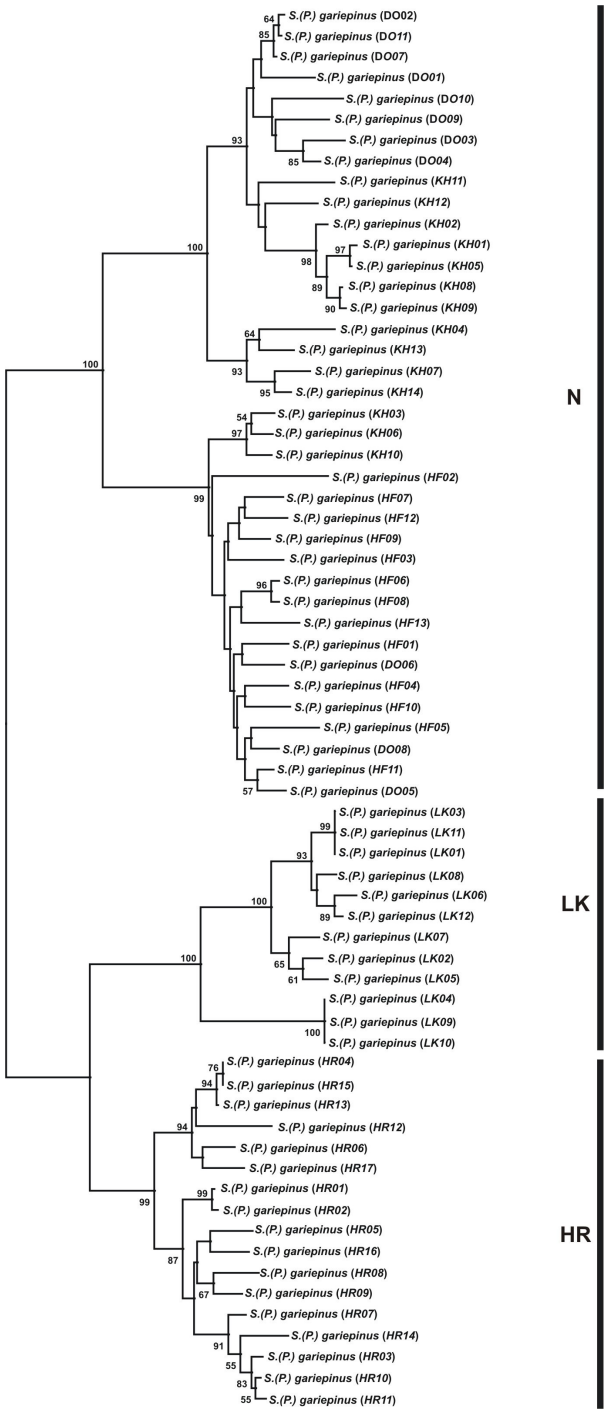


Figure 7. Mid-point rooted neighbor-joining tree for the COI sequence data of *S. (P.) gariepinus*. Bootstrap values below 50 % were removed.

Table 7. Summary of general diversity statistics of *S. (P.) gariepinus*

Species	Assemblage	*N*	No. of haplotypes	Haplotype diversity	Nucleotide diversity	% Pairwise divergence	Variable sites (V)	Parsimoniously Informative Sites (PI)	Singletons (S)
S. (P.) gariepinus	Langhoogte/Kommagas	12	8	0.909 (0.065)	0.026 (0.014)	0.004 - 0.039			
	Holgat River	17	16	0.993 (0.023)	0.023 (0.012)	0.001 - 0.035			
	Hohenfels	13	13	1.000 (0.030)	0.022 (0.012)	0.002 - 0.031			
	Daberas to Obib	11	11	1.000 (0.039)	0.038 (0.020)	0.001 - 0.063			
	Klingharts Mountains	14	14	1.000 (0.027)	0.042 (0.022)	0.001 - 0.065			
	Namibia assemblage	38	38	1.000 (0.000)	0.043 (0.021)	0.001 - 0.069			
Total		67	62	0.997 (0.004)	0.057 (0.001)	0.001 - 0.103	64 (6.67%)	30 (3.13%)	34 (3.54%)

[$] V, PI and S were only estimated for the overall dataset

Table 8. Summary of Fst statistics calculated by AMOVA (Excoffier *et al*., 1992) for *S. (P.) gariepinus*

Species		Φst	%	P
S. (P.) gariepinus	Among groups	Φct 0.496	49.6	<0.001
	Among groups within populations	Φsc 0.415	20.9	<0.001
	Within populations	Φst 0.705	29.5	<0.001

[b] P values were determined from 10000 random permutations.

Historical population dynamics based on the Stepwise and Exponential Expansion Models

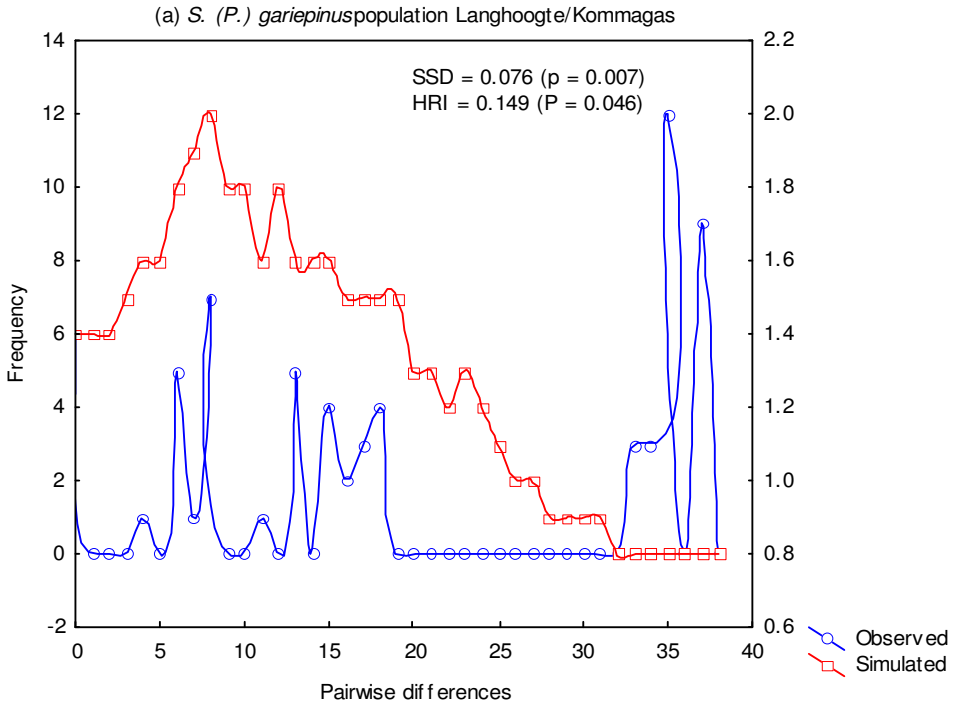
Stepwise Expansion Model

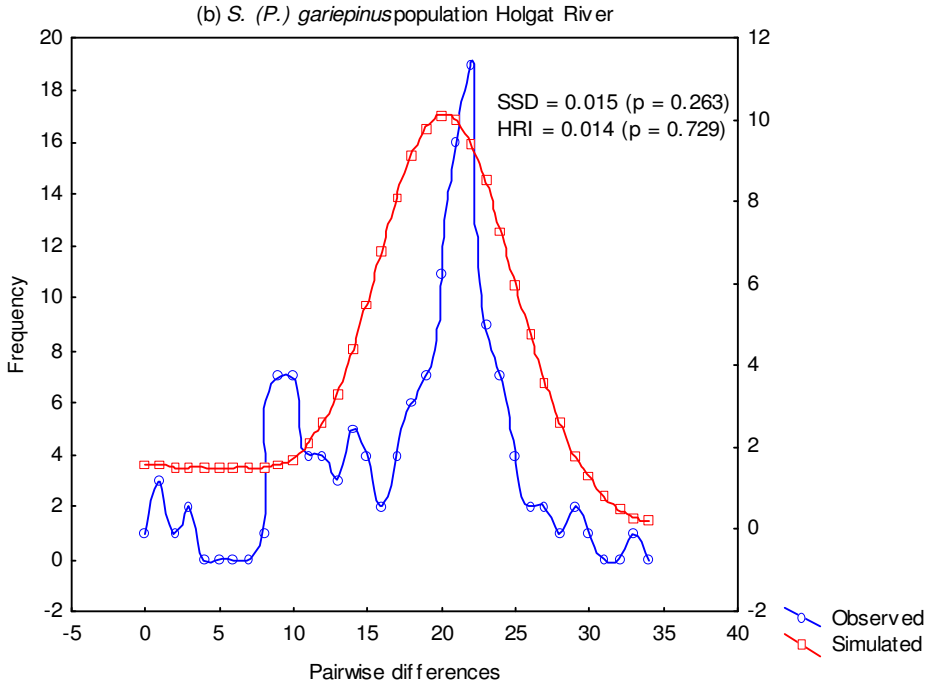
The frequency distribution for the pairwise nucleotide differences was investigated for the five populations separately (Table 1) as well as for the three populations indicated by the neighbor-joining tree. As the three populations within Namibia showed population expansion only the mismatch distribution for the Namibia population is presented. The tree topology (Fig. 7) with branches that are small and of similar length indicates recent sudden expansion. The mismatch distributions (Fig. 8) show similar uni-model curves as expected with a historically expanding population. Both the variance (Sum of the Squared Deviation - SSD) and Harpendings Raggedness Index (HRI) suggest that the simulated and expected curves do not differ significantly under a model of expansion.

Time of divergence

Using the 960 bp of the COI sequence we calculated the average number of nucleotide substitutions per site (d) and obtained a value of 0.08. The divergence time between *S. (P.) gariepinus* and *S. (P.) bennigseni* (Sole *et al.*, 2005; Chapter 2) was estimated to have occurred 2.8 million years ago. This gives the estimate of nucleotide substitutions per site, per lineage, per year (γ) to be $0.08/(2 \times 2{,}800{,}000) = 1.4 \times 10^{-8}$. The mutation rate per nucleotide site, per generation (μ) was therefore 1.4×10^{-8}. The coalescence time in generations for each population was calculated based on the τ values in Table 9a and a haplotype mutation rate (v) of 1.34×10^{-5}

The expansion of the Langhoogte/Kommagas population was estimated at around 231,000 generations/years ago. This appears to be the most recent expansion event while the Holgat River and Klingharts populations appeared to have undergone expansion much earlier at around 800,000 and 961,000 years ago, respectively. Estimated effective population size after expansion (N_1) was an order of magnitude higher than before expansion (N_0) in all populations.


(a) S. (P.) gariepinus population Langhoogte/Kommagas
SSD = 0.076 (p = 0.007)
HRI = 0.149 (P = 0.046)
Frequency
Pairwise differences
Observed
Simulated


(b) S. (P.) gariepinus population Holgat River
SSD = 0.015 (p = 0.263)
HRI = 0.014 (p = 0.729)
Frequency
Pairwise differences
Observed
Simulated

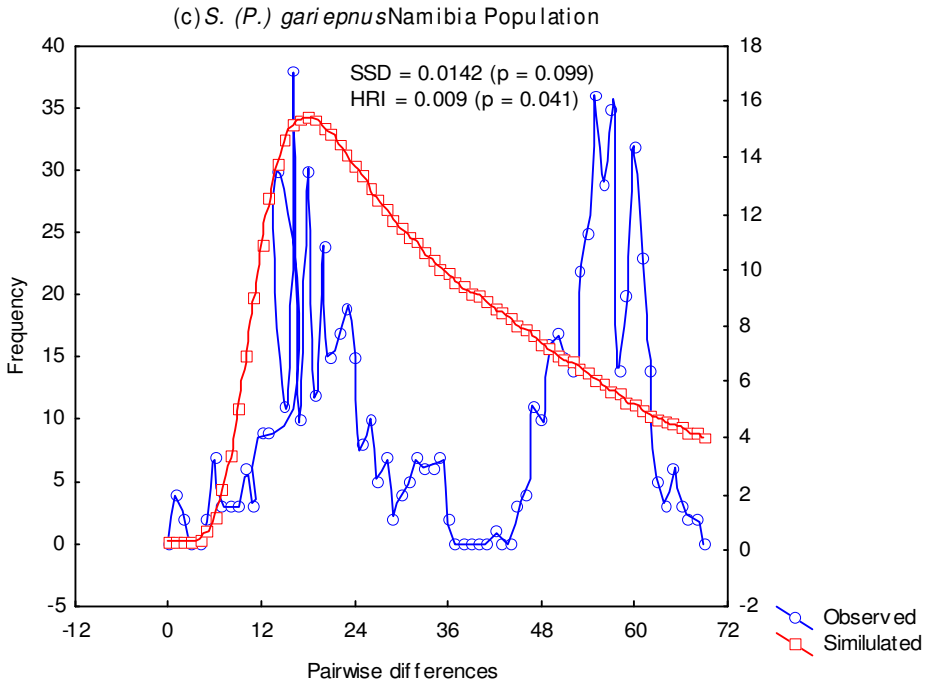


Figure 8 (a – c). Mismatch frequency distributions of pairwise nucleotide differences for the three population assemblages of *S. (P.) gariepinus*, with sum of the squared deviation (SSD) and Harpendings Raggedness Index (HRI) represented on the graphs.

Exponential Population Expansion

The exponential expansion model also indicated a rapid increase in effective population size (positive 'g' value; Table 9b) for all populations. Effective female population size estimated from θ differed markedly between populations with the Langhoogte/Kommagas population having the smallest effective female population size. These values, however, showed a consistent increase in population sizes.

The UPBLUE/Tajima estimate does not show a major recent increase in population size for any of the populations (Table 10). The Holgat River has a slightly significant negative *F*s estimate, indicating a small possible increase in population size while the Langhoogte/Kommagas population has a positive value for the *F*s estimate, indicating no recent increase in population size (Table 10). The Namibian population has a highly significant negative value indicating recent mutations leading to population growth, which is in contrast to the UPBLUE estimate.

Table 9. Estimated parameters for (a) Stepwise and (b) Exponential Expansion Models for *S. (P.) gariepinus*.

(a) Stepwise Expansion Model

		Stepwise Expansion Model			
Species	**Population**	**τ**	**$\theta_0 = 2\mu N_0$**	**$\theta_1 = 2\mu N_1$**	**$t = \tau/2\nu$**
S. (P.) gariepinus	Langhoogte/Kommagas	6.202	25.039	45.274	231,000
	Holgat River	21.371	0	85.156	800,000
	Hohenfels	16.294	0	5156.25	608,000
	Daberas to Obib	8.438	23.414	157.812	315,000
	Klingharts Mountains	26.75	28.812	79.023	961,000
	Namibia Assemblage	11.789	37.853	2710	440,000

(b) Exponential Expansion Model

		Exponential Expansion Model		
Species	**Population**	**$\theta = 2\mu Nf$**	**g**	**N_f**
S. (P.) gariepinus	Langhoogte/Kommagas	0.0279	51.045	996,000
	Holgat River	0.1371	239.111	4,900,000
	Hohenfels	1.7982	504.436	6,400,000
	Daberas to Obib	0.0722	60.816	2,600,000
	Klingharts Mountains	0.0956	76.029	3,400,000
	Namibia assemblage	0.4300	158.565	15,000,000

Table 10. Summary of estimations of Tajima's estimate, Fu's UPBLUE and Fu's *F*s statistic of *S. (P.) gariepinus*

Species		**Langhoogte/Kommagas**	**Holgat River**	**Namibia assemblage**
S. (P.) gariepinus	Tajima's estimate	17.553	18.160	31.179
	Fu's UPBLUE	20.895	1.248	1.172
	UPBLUE/Tajima	1.17	0.038	0.069
	Fu's *F*s	2.654 (ns)	-3.765 (*)	-14.857 (***)

$ ns = non-significant, * = $p < 0.05$, *** = $p < 0.001$

Migration

Migrate showed overall population movement in both a northerly as well as in a southerly direction. Within the Namibian population north and south migration between all the populations appeared to be occurring consistently (Fig. 9).

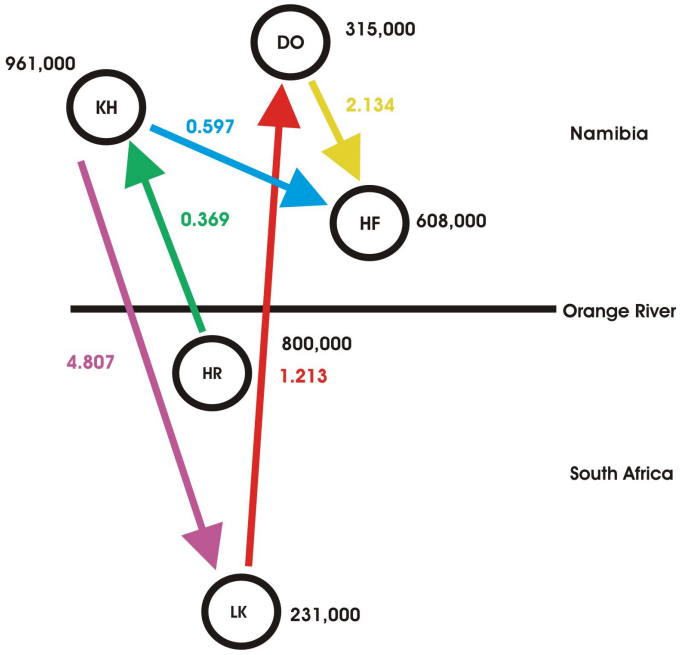


Figure 9. Schematic representation of the migration of individuals between populations of *S. (P.) gariepinus*. The coloured arrows indicate the direction of movement while the numbers in the same colour represent an approximation of the number of individuals moving/generation. The numbers in black represent coalescent times of the populations. (Abbreviations are as follows: LK = Langhoogte/Kommagas, HR = Holgat River, HF = Hohenfels, DO = Daberas to Obib, KH = Klingharts).

Phylogenetic Haplotype Relationships

Using statistical parsimony the maximum number of mutational steps between two haplotypes, excluding homoplasious changes (with a 95 % confidence), was 13 mutational steps. Given these constraints, a minimum spanning tree was constructed (Fig.10). The parsimony network was resolved except for the presence of three reticulations, which were broken. It was not possible to determine a single root for the entire cladogram. There were four major disjointed portions that could not be linked with 95 % confidence, clades 7-1, 9-2, 11-1 and 13-1. Furthermore, haplotypes DO09, DO10, LK04, KH04, KH11, KH12 and HF02

were separated from all other haplotypes by many mutational steps (from 14 to 32). Haplotypes DO09 and DO10 linked to clade 7-1 in 14 mutational steps while KH04 and KH12 linked to clades 2-9 and 8-1 with 15 and 16 mutational steps, respectively. Clade 9-2 and haplotype KH11 linked to clade 9-1 with 18 mutational steps. Haplotypes HF02 and LK04 link to clades 19-1 and 3-1 with 30 and 32 mutational steps, respectively. Clades 11-1 and 13-1 link with 45 mutational steps and the entire network links in 58 mutational steps.

The network shows two distinct clades, 14-1 and 12-1, indicating geographical distinction and supporting the neighbor-joining tree. Clade 14-1 represents the South African populations of *S. (P.) gariepinus*, with 13-1 representing the individuals from Langhoogte/Kommagas and 11-1 those individuals from Holgat River. Clade 12-1 groups all the individuals from the Namibia population together.

The contingency test showed strong geographical association of haplotypes between clades 5-1, 9-1, 10-1, 14-1 and 15-1 (Table 11). The inference chain indicated restricted gene flow with isolation by distance for clade 5-1. Both clades 9-1 and 10-1 indicate inadequate geographic sampling which made it difficult to distinguish between continuous range expansion, long distance colonisation and past fragmentation. Both clades 14-1 and 15-1 indicated allopatric fragmentation as the process responsible for the observed separation between the two South African populations and the populations in South Africa and Namibia. Secondary contact between the populations appears not to have occurred as no shared haplotypes occur between the populations.

The Mantel test revealed a significant association between geographical and genetic distances ($g = 2.055$, $r = 0.6206$, $p < 0.025$), indicating that an increase in distance was related to an increase in the genetic distinctness of the populations.

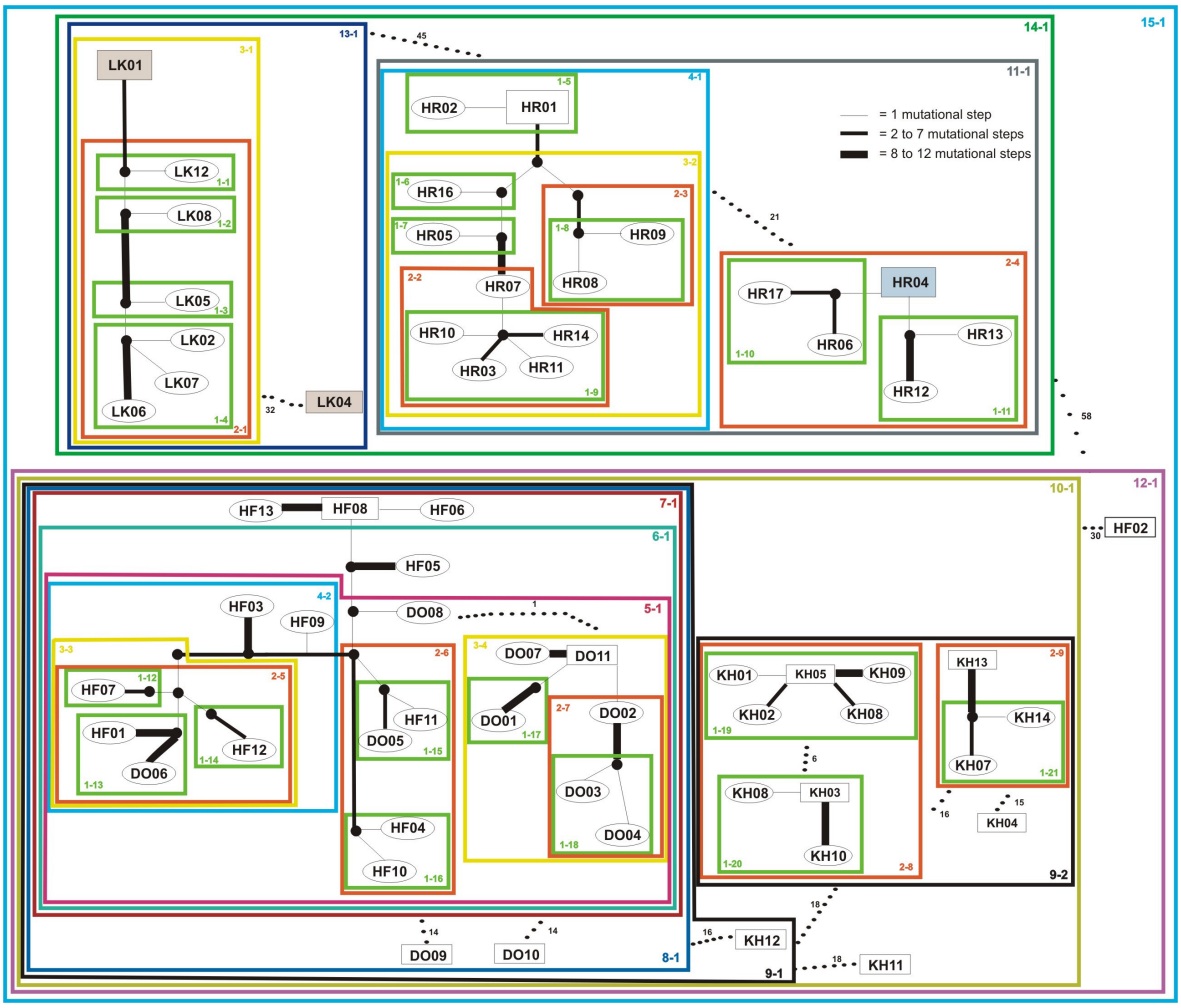


Figure 10. *S. (P.) gariepinus* statistical parsimony network and associated design. Haplotypes are designated by letters representing which population the haplotype came from (as seen in Table 1) and the numbers represent the individual sequenced. The thickness of the connection line indicates the number of mutational steps. Dotted lines represent alternative ambiguous connections. Ancestral haplotypes are represented by squares. Haplotypes with grey background are those that occurred three times and those with a blue background occurred twice.

Table 11. Results of the geographical nested clade analysis for *S. (P.) gariepinus*, inferences based on Templeton (2004)

Haplotype group	χ^2	P-value	Inference chain	
Clades within 5-1	10.644	<0.05	1-no-2-no-11-no-17-yes-4-no	Restricted gene flow with isolation by distance - applicable to lower clade levels only
Clades within 9-1	24.000	<0.05	1-yes-19-yes-20-no	Inadequate geographic sampling
Clades within 10-1	32.926	<0.000	1-no-2-no-11-yes-12-yes-13-no-14-yes	Sampling design inadequate to discriminate between contiguous range expansion, long distance colonisation and past fragmentation
Clades within 14-1	29.000	<0.000	1-yes-19-no	Allopatric fragmentation
Clades within 15-1	67.000	<0.000	1-yes-19-no	Allopatric fragmentation

Discussion

Demographic patterns

Scarabaeus (Pachysoma) gariepinus exhibits a high degree of genetic polymorphism as can be seen by the overall high haplotype diversity. Mitochondrial DNA divergence of *S. (P.) gariepinus* shows strong support for three distinct populations correlating to distinct geographic areas, a Namibian population and two populations in South Africa. The three-population hypothesis shows strong genetic structure as seen by the high Φst value of AMOVA as well as by the overall high sequence divergence (maximum of 10.3 %; Table 7). This is indicative of a long historical separation and could be due to both extrinsic and intrinsic factors.

Extrinsic factors such as environmental barriers contributing to the population structure seen in the present study could be the Orange and Holgat Rivers, which may have separated the populations in South Africa (Holgat River) from each other as well as from the Namibian population (Orange River). Previous studies indicate that sandy pockets at river mouths in Namaqualand (Endrödy-Younga, 1982a), sand accumulations in the lower Orange River (Endrödy-Younga, 1982a; Penrith, 1984) and coastal/littoral dunes (Endrödy-Younga, 1978), specifically in the western Cape (Penrith, 1986) could act as possible areas of origin for various psammophilous taxa. Nested clade analysis distinctly indicates that allopatric fragmentation is the defining factor for the fragmentation between the Langhoogte/Kommagas and Holgat River populations (clade 14-1) as well as between the South African populations and the Namibian population (clade 15-1).

Alternatively, the population structure of *S. (P.) gariepinus* could be maintained by intrinsic factors such as flightlessness, which results in reduced vagility. The Mantel test shows a strong association between geographic and genetic distances, indicating distance as an important factor for influencing the population structure. This is clearly supported by the nested clade analysis, clade 5-1, which includes individuals from two Namibian populations, namely Daberas/Obib and Hohenfels Dunes, where there is restricted gene flow due to isolation by distance. Since the Namibian population occurs on a known dune field continuum, clades 9-1 and 10-1 indicate inadequate sampling hence the need to increase sampling along the complete dune system as opposed to discrete points as was done here. This is important for understanding the apparent lack of structure within the Namibian population.

Both the mismatch distributions and the exponential expansion model indicate a sudden historical increase in all population sizes. Effective female population size differed markedly between the Stepwise and Exponential Expansion Models indicating the importance of using these values as relative indicators and not precise estimates. However, the trends identified across the historical estimation procedures were the same. Fu's *F*s statistics show strong support within the Namibian population for overall recent population growth but the UPBLUE estimate contradicts this. The UPBLUE/Tajima estimate shows the Langhoogte/Kommagas and Holgat River populations to be stable with slight growth in the Langhoogte/Kommagas population. Fu's *F*s statistic indicates a similar trend in the Langhoogte/Kommagas population, in that there is no growth. In contrast to this the *F*s statistic for the Holgat River population is significantly negative indicating recent growth. However, the contrast between the Stepwise and Exponential Expansion Models and Fu's UPBLUE/Tajima estimate and the *F*s statistic is only apparent in that the two expansion models infer population growth from historical/ancient processes whereas Fu's methods infer population parameters based on more recent mutational events.

Migration rates between populations were estimated in an attempt to infer historical movements of the species. Overall it appeared there had been a large amount of movement between populations. Movement occurred in a south-north direction, which is consistent with previous hypotheses that indicate movement with the unidirectional wind regime (Endrödy-Younga, 1982a; Sole *et al*., 2005) as well as in a north-south direction. Two populations appear to have undergone expansion earlier than the others, the Klingharts Mountains and Holgat River populations, 800,000 and 961,000 years ago, respectively. It would, therefore, appear that the ancestor of what is seen today would have invaded the Namib by simply moving from east to west or by remaining at a locality and adapting to the changing climate, thereafter moving and radiating into other favourable habitats (Irish, 1990). The addition of an extra locus (microsatellites/nuclear gene) would possibly allow for a clearer picture. The amount of movement between populations appeared high for a group of flightless individuals. This may show that under conditions of extreme environmental pressure or very favourable periods the beetles will move over long distances. Contradictory to this is the fact that each population has its own set of unique haplotypes indicating that individuals appear to remain in a certain locality.

Species boundaries and conservation issues

One of the most fundamental urges of mankind is to identify and name things (Mayr & Ashlock, 1991). It has been suggested that the taxonomy of a group should be consistent with its evolutionary history (Wiley, 1981; Frost & Hillis, 1990). Every species taxon in nature consists of numerous local populations, which raises the problem of how to treat them taxonomically. Adding dimensions of geography and time poses numerous additional problems (Mayr & Ashlock, 1991). The mitochondrial DNA of *S. (P.) gariepinus* reveals three distinct genetically isolated assemblages, which reflect different demographic histories. All three assemblages are unique in the fact that they do not share haplotypes and have probably been isolated for more than 200,000 years and may therefore warrant distinct taxonomic status under various species concepts. In principle the three assemblages could be defined as separate species based on the Phylogenetic species concept (Nixon & Wheeler, 1990) and Templeton's cohesion species concept (Templeton, 2001), as it appears that there has been no recent gene flow.

Moritz (1994a; b) identified units or targets for conservation by applying the principle of conserving ecological and evolutionary processes in an attempt to conserve biogeography (Moritz, 1999). It may be optimistic to attempt to conserve all the populations of a species therefore one would ideally like to target the populations that will ensure a species remains viable and able to survive in the short-term and diversify in the long-term. Moritz (1994a; 1999) describes these units or targets as Evolutionary Significant Units (ESUs) and Management Units (MUs). ESUs he defines as having to be reciprocally monophyletic for mtDNA alleles and to have shown significant divergence of allele frequencies at nuclear loci. MUs are recognised as populations with significant divergence of allele frequencies at nuclear or mitochondrial loci, regardless of phylogenetic distinctiveness of the alleles and are the units used for population monitoring and demographic based studies (Moritz, 1994a). According to these definitions the populations of *S. (P.) gariepinus* could be described as MUs. These populations are connected by low levels of historical gene flow but are functionally independent and would therefore need to be managed as individual entities, forming part of an inclusive species.

Chapter IV (c)

Genetic structure, phylogeography and demography of *S. (P.) denticollis* based on inferences from Cytochrome Oxidase I

Introduction

Phylogeographic structure is important for organisms with extensive ranges and complex geographical patterns. When integrated with data on geographical distribution of morphology and ecological variation such inferences can be used to test hypotheses of speciation processes (Nice *et al*., 2005).

Many formerly continuous areas of natural habitats have been subdivided into smaller habitat islands surrounded by human-altered environments (van Dongen *et al*., 1998). Artificially divided populations often have a limited number of individuals interchanging between sub-populations. The fewer individuals moving between populations the greater the effect of genetic drift in that genetic diversity decreases within and increases between sub-populations (van Dongen et al., 1998; Driscoll & Hardy, 2005). One would intuitively assume that a species occurring within fragmented habitats would show decreased genetic variation as opposed to those occurring over a continuous habitat.

Scarabaeus (Pachysoma) denticollis are restricted to the coastal and inland dunes of the central Namib dune sea (see Figure 1), and are conserved within the Namib Naukluft Park. The species occurs from Luderitz (S26°41' - E15°15') to Walvis Bay (S22°55' – E14°28') and populations of this species exhibit individuals with elytral colours ranging from orange to black with some showing a mix of the two colours (Scholtz, pers. obs.). The individuals with the black elytra were previously described as a subspecies of *S. (P.) denticollis, P. denticollis penrithae* (Harrison *et al*., 2003), but were synonomised, based on morphology, with *S. (P.) denticollis sensu stricto* by Holm & Scholtz (1979).

Materials and Methods

See main body of chapter. Details of specimen collecting sites can be seen in Figure 11.

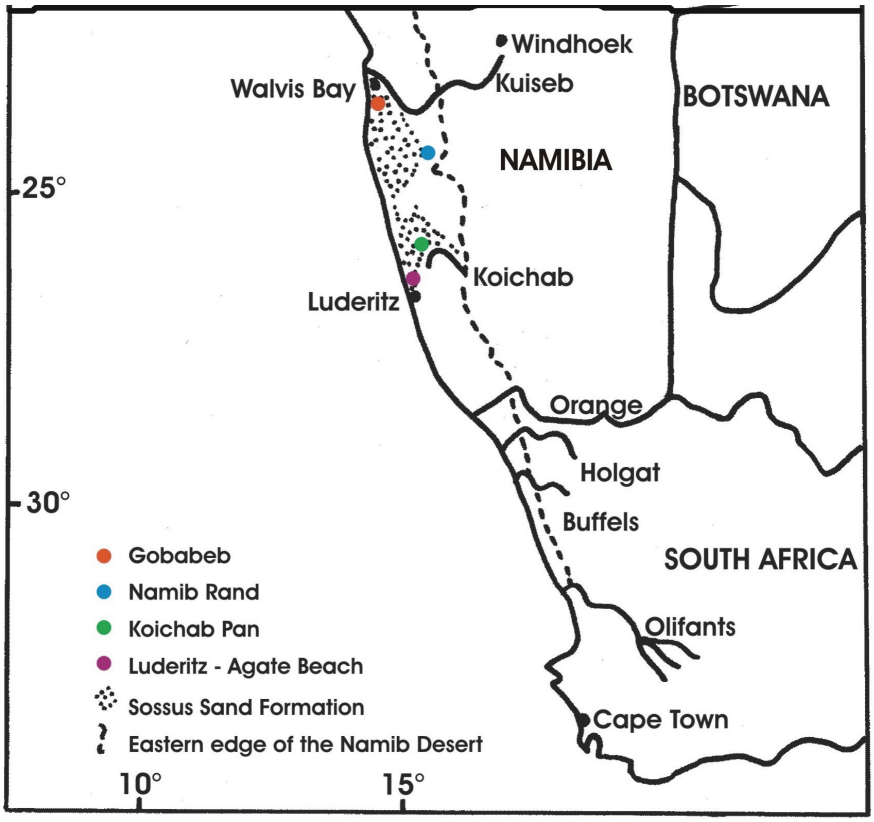


Figure 11. Localities in Namibia where *S. (P.) denticollis* were collected for this study.

Results

Phylogenetic and Molecular Diversity

Population Statistics

Thirty-two individuals were used for molecular characterisation (Genbank Accession numbers AY965207 – AY965238). The sequences exhibited an A/T bias of 69 %, which is the same as that observed for the other two species investigated in this chapter. Of the 32 individuals studied 29 represented unique haplotypes with no evidence found for a haplotype being specific to any geographic region (Table 12; which includes general molecular diversity statistics). Accordingly, overall haplotype diversity expressed was high (H = 0.992 +/- 0.011), and intermediate to that within *S. (P.) hippocrates* and *S. (P.) gariepinus*.

Gamma distribution for the data was estimated at 0.3413 (which is intermediate to that of the previous two species) with the proportion of invariable sites being 0.6791. The transition/transversion ratio was 1.9 (estimated in MEGA) and the model best fitting the data selected by Modeltest was Tamura-Nei. The neighbor-joining tree of 32 individuals can be seen in Figure 12 and indicates two assemblages, namely 1 and 2. Assemblage 1 consists of

individuals from all four localities and even though it appears that individuals from collecting localities group together, the support thereof is poor allowing for no distinct population designation within this assemblage. Assemblage 2 shows strong support for individuals from two localities (namely Koichab Pan and Agate Beach). However, individuals from these two localities also occur within assemblage 1. This as well as the fact that there are no shared haplotypes between the collecting sites of *S. (P.) denticollis* may indicate that incomplete lineage sorting has occurred i.e. speciation is in the process of occurring. *S. (P.) denticollis* is therefore treated as a single population throughout this sub-chapter (Parsimony and Maximum Likelihood trees exhibited similar topologies, data not shown).

Genetic differentiation among populations

Mean nucleotide diversity was calculated across the three collecting sites (Table 13). The results of AMOVA revealed that 22.16 % of the variance resulted from the differences among the three collecting sites, while 77.85 % of the variance resulted from differences within collecting sites. The fixation index value ($\Phi st = 0.222$) was low but significantly so ($p < 0.001$) indicating weak genetic structure between collecting sites.

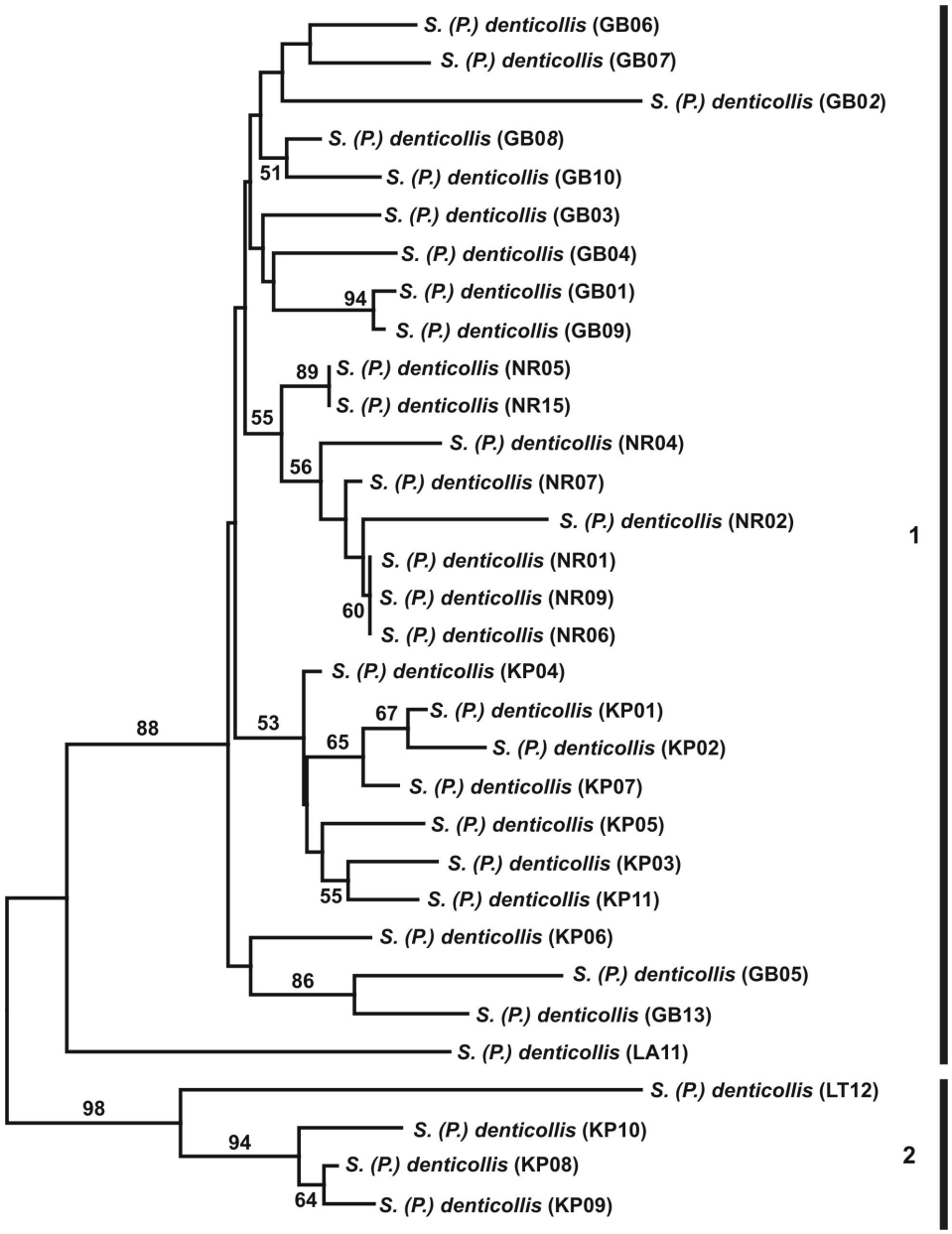


Figure 12. Mid-point rooted Neighbor-joining tree for the COI sequences of *S. (P.) denticollis*. Bootstrap values below 50 % were removed.

Table 12. Summary statistics of general nucleotide diversity over the 960 bp of *S. (P.) denticollis*

Species	Assemblage	*N*	Number of haplotypes	Haplotype diversity	Nucleotide diversity	% Pairwise divergence	Variable sites (V)	Parsimoniously Informative Sites (PI)	Singletons (S)
S. (P.) denticollis	Koichab Pan	13	13	1.000 (0.030)	0.020 (0.011)	0.003 - 0.035			
	Namib Rand	8	5	0.857 (0.108)	0.005 (0.003)	0.001 - 0.011			
	Gobabeb	11	11	1.000 (0.039)	0.013 (0.007)	0.004 - 0.024			
Total		32	29	0.992 (0.011)	0.016 (0.008)	0.001 - 0.019	90 (9.375%)	48 (5%)	42 (4.375%)

[$] V, PI and S were only estimated for the overall dataset

Table 13. Summary of Fst statistics calculated by AMOVA (Excoffier *et al*., 1992) for *S. (P.) denticollis*

Species		Φst	%	P
S. (P.) denticollis	Among collecting site variation		22.16	< 0.001
	Within collecting site variation		77.84	< 0.001
	Fixation index	0.222		< 0.001

[b] P values were determined from 10000 random permutations.

Demographic patterns based on the Stepwise and Exponential Expansion Models

Stepwise Expansion Model

The tree topology (Fig. 12) and the mismatch distributions (Fig. 13), for the individual collecting sites as well as for the entire population, indicate recent sudden demographic expansion for the species as a whole, hence only the mismatch distribution for the population as a whole is presented. These patterns suggest recent expansion in population size and geographic range. The variance (sum of the squared deviation - SSD) and Harpendings Raggedness Index (HRI) for the mismatch distributions were not significant. Under a model of population expansion, the observed and simulated curves were not significantly different from one another.

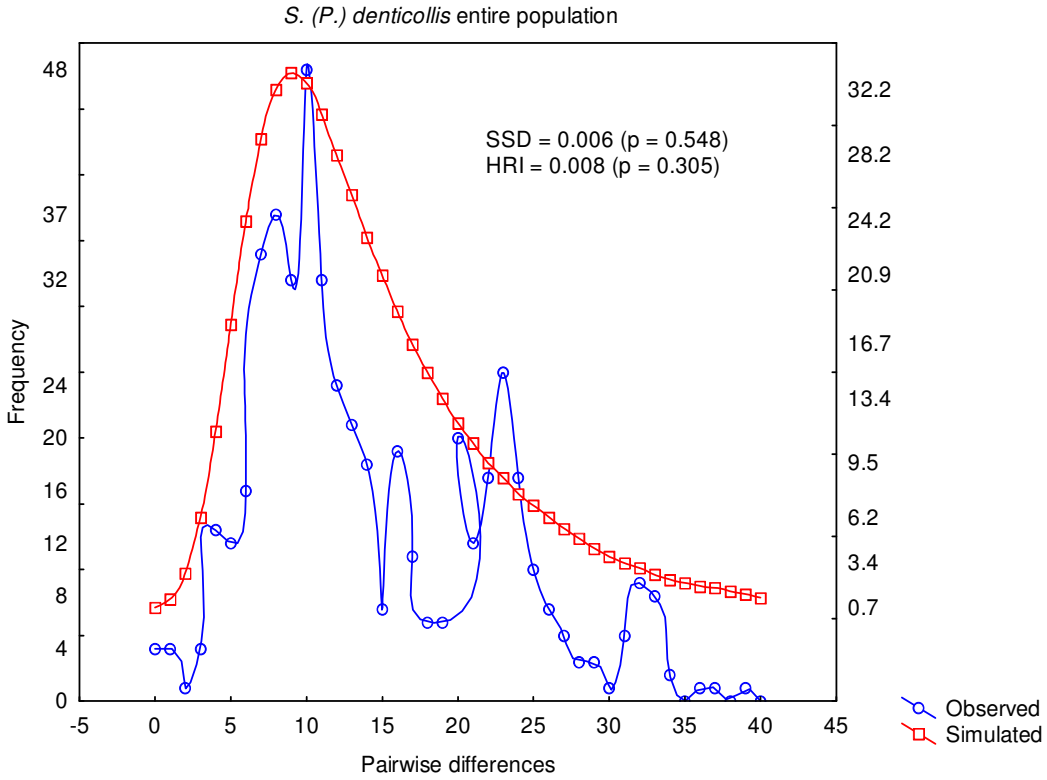


Figure 13. Mismatch frequency distribution of pairwise nucleotide differences for *S. (P.) denticollis* as a single population with sum of the squared deviation (SSD) and Harpendings Raggedness Index (HRI) represented on the graph.

Time of divergence

The average number of nucleotide substitutions per site (*d*) was calculated to be 0.06. Divergence time between *S. (P.) denticollis* and its sister species, *S. (P.) rotundigenus* (Sole *et al*., 2005; chapter 2) was estimated at approximately 2.6 million years ago. This gives the estimate of nucleotide substitutions per site, per lineage, per year (γ) to be 0.06/(2 x 2,600,000) = 1.1 x 10^{-8} with the mutation rate per nucleotide site, per generation (μ) being 1.1 x 10^{-8} and the mutation rate per haplotype (v) being 1.05 x 10^{-5}. The coalescence time in generations, for each population, was calculated using the τ values estimated by Arlequin (Table 14a).

The earliest expansion event appeared to have occurred at the Gobabeb collecting site, which was estimated at around 343,000 generations/years ago while the Koichab Pan site appeared to have undergone recent expansion approximately 144,000 generations/years ago. This is based on a τ value of 7.235 and 3.054 (Table 14a), respectively, and a mutation rate per haplotype (v) of 1.06 x 10^{-5} per COI (as calculated above). Estimated effective population size after expansion (N_1) was an order of magnitude higher than before expansion (N_0).

Exponential Expansion Model

All three collecting sites have high and positive 'g' values for the Exponential Expansion Model. Koichab Pan and Gobabeb have the highest 'g' values while Namib Rand is slightly lower. However, they all indicate historical population expansion. The effective female population sizes were large (in the millions) for both Koichab Pan and Gobabeb while an order of magnitude smaller for the Namib Rand collecting site (Table 14b)

Table 14. Estimated parameters for (a) Stepwise and (b) Exponential Expansion Models for *S. (P.) denticollis*.

(a) Stepwise Expansion Model

		Stepwise Expansion Model			
Species	**Collecting sites**	τ	$\theta_0 = 2\mu N_0$	$\theta_1 = 2\mu N_1$	$t = \tau/2\nu$
S. (P.) denticollis	Koichab Pan	3.054	21.785	4645.000	144,000
	Namib Rand	5.961	0	10.013	281,000
	Gobabeb	7.235	4.175	5712.500	343,000
	Population as a whole	6.462	8.674	810.625	306,000

(b) Exponential Expansion Model

		Exponential Expansion Model		
Species	**Collecting sites**	$\theta = 2\mu Nf$	**g**	N_f
S. (P.) denticollis	Koichab Pan	0.0873	180.209	4,000,000
	Namib Rand	0.0123	405.484	560,000
	Gobabeb	0.3837	638.549	17,000,000
	Population as a whole	0.229	301.807	10,000,000

The UPBLUE/Tajima estimate shows a two-fold increase in recent population size (Table 15). Fu's *Fs* statistic for the population as a whole was significantly negative, indicating that the population is undergoing expansion, which is in direct contrast to the other two species (Table 15).

Table 15. Summary of estimations of Tajima's estimate, Fu's UPBLUE and Fu's *Fs* statistic of *S. (P.) denticollis*

Species		Complete population
S. (P.) denticollis	Tajima's estimate	22.422
	Fu's UPBLUE	60.695
	UPBLUE/Tajima	2.718
	Fu's *Fs*	-7.315 (**)

$ ** = p < 0.01

Migration

MIGRATE indicated ancestral movement was strongly in a northerly direction with no evident movement to the south (Fig. 14).

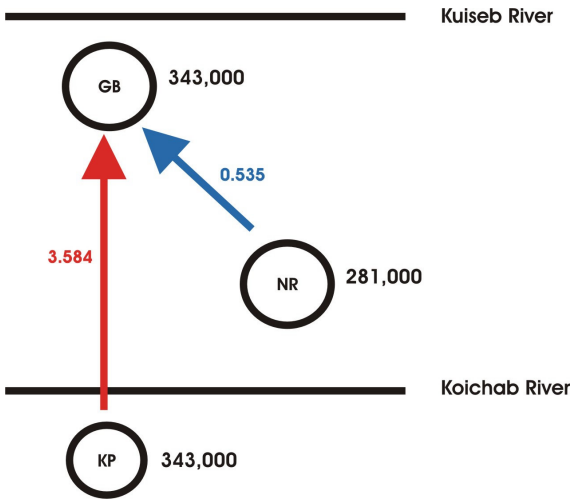


Figure 14. A schematic representation of the migration of individuals between collecting sites of *S. (P.) denticollis*. The coloured arrows indicate the direction of movement while the numbers in the same colour represent an approximation of the number of individuals moving/generation. The numbers in black represent coalescent times of the populations. (Abbreviations are as follows: KP = Koichab Pan, NR = Namib Rand, GB = Gobabeb).

Phylogenetic haplotype relationships

A 95 % parsimony cladogram was estimated for the mtDNA COI data (Fig 15) of *S. (P.) denticollis*, with the maximum number of mutational steps between two haplotypes being 13. The parsimony network was resolved except for the presence of four reticulations, which were broken. There was, however, one disjoint portion within the network, represented by clade 3-1 that could not be linked with 95 % confidence. In addition, haplotypes GO02, LA11 and LT12 were separated from the network by 14, 19 and 20 mutational steps respectively. To simplify the nested diagram not all nesting levels were represented. Haplotypes were collapsed for drawing the cladogram but not in the nesting structure that was used to calculate the contingency test values. GO02 linked to clade 14-1 in 14 mutational steps and LA11 and LT12 linked to clades 19-1 and 20-1 in 19 and 20 mutational steps, respectively. Clade 3-1 was linked by 25 mutational steps to clade 25-1 i.e. the entire cladogram.

The network shows two distinct clades in 3-1 and 25-1. Clade 3-1 consists of three individuals from Koichab Pan, while clade 25-1 consists of the balance of the individuals from all the sites sampled, including others from Koichab Pan. This supports the neighbor-joining tree in that *S. (P.) denticollis* appears as a single population on a dune continuum. The contingency test showed significant geographical association of haplotypes within clades 22-1, 25-1 and 26-1. The inference chain (Table 16) indicates restricted gene flow due to isolation by distance for clades 22-1 and 25-1. The inference for clade 26-1 indicates that the population structure could be attributed to restricted gene flow or dispersal, with some long distance dispersal over intermediate areas not occupied presently by the species, or that there was past gene flow which has been followed by the extinction of intermediate populations.

Due to the small number of populations the Mantel test could not calculate whether there were significant differences or not between the geographic and genetic distances (Liedloff pers. comm.).

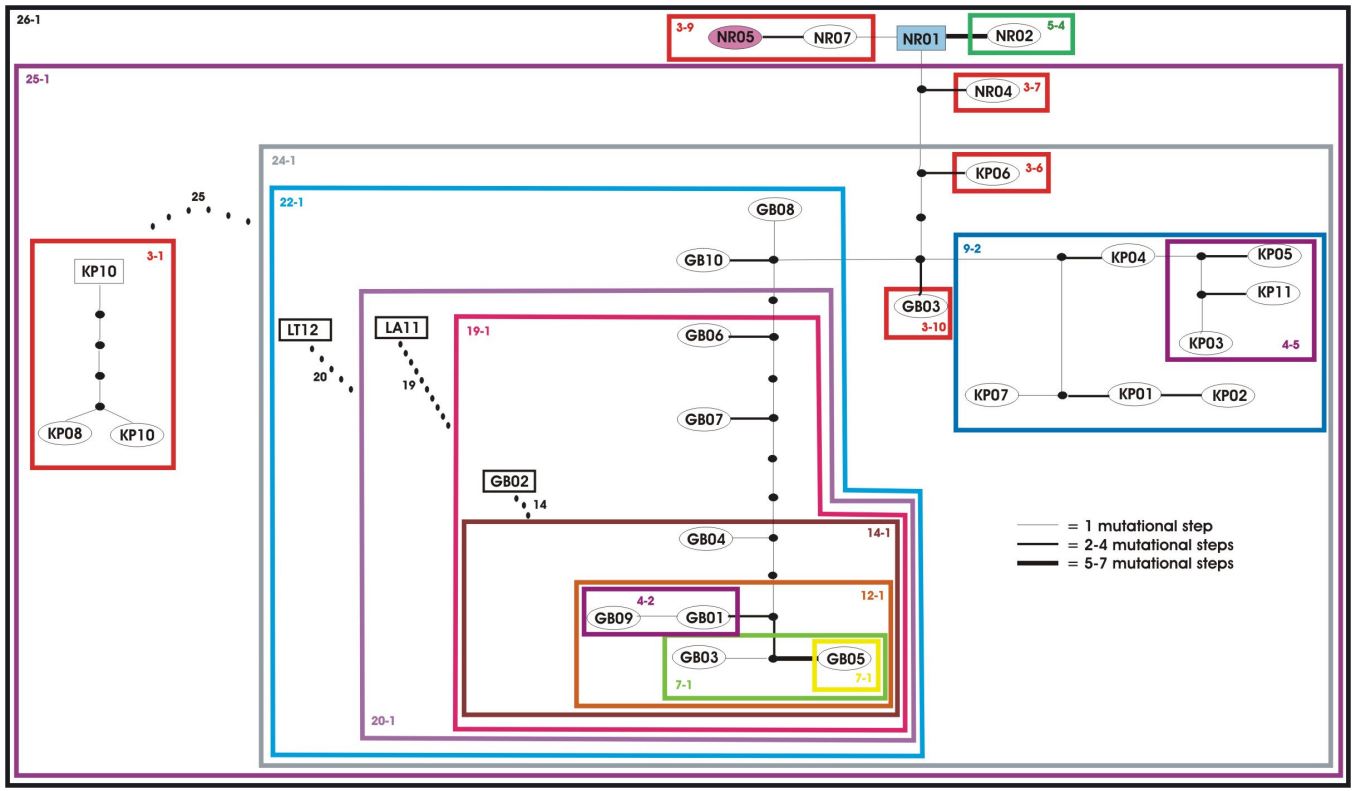


Figure 15. *S. (P.) denticollis* statistical parsimony network and associated design. Haplotypes are designated by letters, which represent the population from where the haplotype came (as seen in Table 1) and numbers representing the individual sequenced. The thickness of the connection line indicates the number of mutational steps. Dotted lines represent alternative ambiguous connections. Ancestral haplotypes are represented by squares. Haplotypes with blue background occurred three times and those with a pink background occurred twice.

Table 16. Geographical nested clade analysis of *S. (P.) denticollis*, inferences based on Templeton (2004)

Haplotype group	χ^2	P-value	Inference chain	
Clades within 22-1	20.303	<0.001	1-19-yes-20-yes-2-yes-3-no-4-no	Restricted gene flow with isolation by distance
Clades within 25-1	29.221	<0.05	1-2-yes-3-no-4-no	Restricted gene flow with isolation by distance
Clades within 26-1	26.880	<0.05	1-no-2-yes-3-yes-5-no-6-no-7-no-8-yes	Restricted gene flow/dispersal but with some long distance dispersal over intermediate areas not occupied by the species; or past gene flow followed by extinction of intermediate populations

Discussion

Demographic patterns of S. (P.) denticollis as compared with S. (P.) hippocrates and S. (P.) gariepinus

The partial mtDNA COI gene phylogeny of *S. (P.) denticollis* shows weak support for the designation of distinct populations. Poor population structure is supported by a low fixation index value, low within population variation as well as low range of within- and between-population percent pairwise distances (Table 12; 0.4% – 2.3 %). This is in contrast *to S (P.) hippocrates* and *S. (P.) gariepinus* which show strong population structure and have a larger percent pairwise divergence range (1 – 12.3 %: Table 3 Chapter 4a and 0.1 – 10.3 % Table 7 Chapter 4b) as well as higher overall nucleotide diversity. Comparison of the percent pairwise divergence and nucleotide diversity between the Namibian population of *S. (P.) gariepinus* and *S. (P.) denticollis,* which both represent a dune field continuum population, reveals that *S. (P.) denticollis* has much lower values for both parameters which is counter-intuitive. Therefore, does landscape fragmentation increase genetic differentiation due to an isolation effect? This study as well as previous ones indicates that rivers, towns, agricultural fields, roads etc. present barriers to species movement by thinning the inhabited area, causing isolation by distance. This study clearly demonstrates that genetic differentiation is higher in species occurring within fragmented landscapes as opposed to those within a continuous landscape. These results are in qualitative agreement with population genetics theory and support the results seen by Knutsen *et al.* (2000) (Order: Coleoptera – Tenebrionidae), Driscoll & Hardy (2005) (Order: Squamata – Agamid Lizards) and van Dongen *et al.* (1998) (Order Lepidoptera – Geometridae). Population fragmentation affects the genetic structure of a species and represents a potential threat to those species with reduced dispersal capabilities. However, it has been argued that some degree of genetic isolation may be advantageous for the conservation of genetic variation and that genetic diversity may be maintained if a population is subdivided into sub-populations. Within a sub-population genetic variation will decrease due to genetic drift but overall population genetic variation will be maintained as different alleles will be preserved in different sub-populations (van Dongen *et al.*, 1998). Although elevated genetic variation is observed in two of the species there are implications in that with reduced local and global populations effective population sizes and loss of advantageous alleles or fixation of disadvantageous alleles could result in the ultimate extinction of a species. This highlights the importance of understanding the patterns and processes acting on and within a species. Both demographic and fine-scale genetic factors need to be examined to reveal likely evolutionary processes acting on a population or sub-

population and will provide a strong guide for conservation management decisions (Driscoll & Hardy, 2005).

Recent versus historical population trends

S. (P.) denticollis appears on a continuum of dune fields from Koichab Pan all the way up the west coast of Namibia to Gobabeb, with both recent and historical estimates showing an increase in population size. There is historical movement in a northerly direction with the Gobabeb collecting site having received individuals from both the Namib Rand as well as the Koichab Pan collecting sites. This provides support for the hypothesis that individuals within a species are moving with their substratum, the barchan dune, in conjunction with the unidirectional wind regime. The Namibian assemblage of *S. (P.) gariepinus* shows similar trends. The nested clade analysis indicates two processes that may have shaped the *S. (P.) denticollis* population. Firstly it appears that *S. (P.) denticollis* has experienced restricted gene flow due to isolation by distance, indicating that there is a minimum amount of recent gene flow between collecting sites. The fact that this species is flightless coupled with their large distributional range (extending over 400 km) would support the fact that their movement between suitable habitats has been at a minimum. Absence of shared haplotypes between the collecting sites may also be an indication that reduced gene flow is occurring. In contrast to this, large population sizes, as indicated by both the Exponential and Stepwise Expansion Models, could have reduced the probability of collecting individuals with overlapping haplotypes. Secondly, it appears that when looking at the entire population, extinction of intermediate populations may have occurred. Extinction of intermediate populations could possibly be attributed to sub-standard habitat quality in intermediate areas, environmental barriers and human induced changes occurring within their habitat. The Agate Beach collecting site, which occurs near Luderitz, is clearly affected by recent mining activities and an encroaching town. The genetic patterns of *S. (P.) denticollis* suggest that the species is still expanding into new or formally occupied habitats - as seen in both Fu's *F*s and UPBLUE statistics - followed by a period of stasis, during which isolation by distance and intermediate population extinction are the causal factors attributed to species phylogeography.

Incomplete lineage sorting vs. clinal variation

Slight morphological differences are visible between the individuals occurring in the most southern and northern distribution of *S. (P.) denticollis* (Harrison *et al*., 2003). Elytral colour

differentiation is also noted within this species, with some individuals having black elytra, others a mix between black and orange and some individuals have orange elytra (Harrison *et al*., 2003). The colour and morphological variation is clinal and is probably a response to a selective environmental gradient (Barrowclough *et al*., 2005). The mtDNA variation within *S. (P.) denticollis* is therefore inconsistent with the morphological clinal variation. Concordance among different datasets often occurs over a long period of time. However, where rapid and recent divergence (within the last 200,000 years for *S. (P.) denticollis*) has occurred retardation of lineage sorting (i.e. incomplete lineage sorting) leading to the identification of ESU's or MU's (Moritz, 1994a & b; Nice *et al*., 2005) becomes difficult. If the traits used to define clinal variation were under selection, surveys of neutral variation would fail to detect distinctive evolutionary lineages where adaptive differences already exist. Differences in life history traits, ecological requirements, morphology and demographic characters would constitute evolutionary significance of the individuals from the different collecting sites. Therefore, ecological non-exchangeability could provide sufficient evidence for the designation of the individual collecting sites as distinct evolutionary units under the strategy posed by Crandall *et al*. (2000).

Implications for conservation of Scarabaeus (Pachysoma) species

Three distinct species of *Scarabaeus (Pachysoma)* have been studied here, all exhibiting different population demographics with population demographic overlap seen in areas of geographic similarity (as seen in the Namibian population of *S. (P.) gariepinus* and *S. (P.) denticollis* as well as the South African populations of *S. (P.) gariepinus* and *S. (P.) hippocrates*). The patterns of gene flow within the presented phylogeographic regions suggest that the three species were each a single continuous population, possessing a relatively high level of dispersal capabilities. This pattern suggests that the observed phylogeographic patterns were probably due to the extinction of intermediate populations causing fragmentation of the entire population. Extinction of intermediate populations could have been caused by anthropogenic and environmental factors, as mentioned throughout the sub-chapters. Physical barriers appear to have had an increased effect on the population structure seen in *S. (P.) gariepinus* while anthropogenic factors appear to be affecting *S. (P.) hippocrates* to a greater degree.

All three of these species were chosen as they exhibit south-north morphological clinal variation and it has clearly been shown that extensive genetic variation occurs within two of the species, *S. (P.) hippocrates* and *S. (P.) gariepinus*. There is strong evidence to

suggest that selective changes are taking place and having an effect on population structure as a whole.

The spatial scale of genetic differences indicates the scale at which conservation should occur. If the species decline over areas spanning the distance between the populations substantial genetic variation will be lost. A series of adequately sized reserves spanning the collecting sites of each population would sample most of the genetic structure. However, would these reserves sustain population perpetuity? In addition to conserving genetic diversity, an important goal of conservation should be to maintain evolutionary processes. Evolutionary processes, within all three species, appear to have been maintained by population and range expansions followed by isolation and fragmentation leading to subsequent divergence. A carefully maintained meta-population strategy may be required to prevent biodiversity loss (Driscoll & Hardy, 2005).

Summary

All three species reported on in this chapter allow for contrasting as well as similar inferences to be made. *S. (P.) hippocrates* and *S. (P.) gariepinus* exhibit strong population structure, supported by AMOVA and high sequence divergence. *S. (P.) denticollis* shows poor phylogenetic structure, as seen by the significantly low AMOVA and the low sequence divergence. All three species exhibit high haplotype diversities with no overlap of haplotypes between populations or collecting sites. Areas of refugia could therefore not be speculated upon.

All three species show strong historical population expansion as seen by the Stepwise and Exponential Expansion Models. Fu's UPBLUE and *F*s statistic indicates that *S. (P.) hippocrates* and *S. (P.) gariepinus* are not undergoing present day expansion which is in line with current species trends in that overall numbers are declining. However, as no present day census data are available, this is difficult to substantiate. Recent events are therefore masked by past events and the genetic signal observed could be misleading. *S. (P.) denticollis* shows a strong trend towards recent range expansions after which a period of stasis occurred.

Extrinsic factors such as rivers and anthropogenic influences have affected all the species in some way. The major rivers act as barriers causing fragmentation leading to allopatric fragmentation with strong support obtained from the NPCA. Anthropogenic factors affecting population structure include agriculture, town encroachment and mining activities which all remove large tracts of suitable habitat leading to fragmentation of a species and in some instances extinction of a population. Since *Scarabaeus (Pachysoma)* species are

flightless and therefore exhibit reduced vagility this appears to have contributed to species structure. NPCA analysis indicates that isolation by distance is a factor contributing to species structure and this could be directly related to low or poor vagility. In contrast to this the results of *S. (P.) gariepinus* and *S. (P.) denticollis* indicate that individuals have historically moved between populations or collecting localities. The fact that the beetles have clearly been shown to move in a south-north direction with the barchan dunes may be the factor underlying the strong movement over large geographic distances. Coalescence of the species is shown to have occurred during the Pleistocene era coincident with the onset of hyper-aridity and the formation of advective fog, which is wind blown up to 50 km inland. The formation of the fog would have allowed for a consistent source of water permitting the species to inhabit previously inhospitable areas.

S. (P.) hippocrates and *S. (P.) gariepinus* show far higher genetic divergence as opposed to *S. (P.) denticollis*. However counter-intuitive this may appear it is in line with genetic theory, in that fragmentation of a landscape, and in turn a species' populations increases genetic variation. This has implications for conservation strategies being implemented; as the variation in populations represents genetic material which if lost could result in imminent extinction.

Acknowledgements

Financial support received from the South African National Research Foundation (NRF) and the University of Pretoria is gratefully acknowledged. NAMDEB, in Namibia, and De Beers, in South Africa, are thanked for letting CS and CHS complete field work in restricted areas. Adam Liedloff wrote the Mantel Nonparametric Test program and is thanked for his help regarding data analysis of *S. (P.) denticollis*.

References

Avise JC, Arnold J, Martin Ball R, Birmingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific Phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecological Systematics*, **18**, 489-522.

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

Barber PH (1999) Patterns of gene flow and population genetic structure in the canyon treefrog, *Hyla arenicolor* (Cope). *Molecular Ecology*, **8**, 563-579.

Barrowclough GF, Groth JG, Mertz LA, Gutiérrez RJ (2005) Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls. *Molecular Ecology*, **14**, 1109-1120.

Beerli P, Felsenstein J (1999) Maximum-Likelihood estimation of migration rates and effective population numbers in two populations using a coalscent 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 of the USA*, **98**, 4563-4568.

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

Carisio L, Cervalla P, Palestrini C, DelPero M, Rolando A (2004) Biogeographical patterns of genetic differentiation in dung beetles of the genus *Trypocopris* (Coleoptera, Geotrupidae) inferred from mtDNA and AFLP analysis. *Journal of Biogeography* , **31**, 1149-1162.

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

Colville J, Picker M.D, Cowling RM (2002) Species turnover of monkey beetles (Scarabaeidae: Hopliini) along environmental and disturbance gradients in the Namaqualand region of the succulent Karoo, South Africa. *Biodiversity and Conservation*, **11**, 243-264.

Cowling RM, Esler KJ, Rundel PW (1999) Namaqualand, South Africa – an overview of a unique winter-rainfall desert ecosystem. *Plant Ecology*, **142**, 3-21.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959-969.

Crandall KA, Bininda-Edmonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, **15**, 290-295.

Driscoll DA, Hardy CM (2005) Dispersal and phylogeography of the agamid lizard *Amphibolurus nobbi* in fragmented and continuous habitat. *Molecular Ecology*, **14**, 1613-1629.

Endrödy-Younga S (1982a) Dispersion and translocation of dune specialist Tenebrionids in the Namib area. *Cimbebasia (A)*, **5**, 257-271.

Endrödy-Younga S (1982b) The evidence of Coleoptera in dating the Namib Desert re-examined. In: *Palaeoecology of Africa and the surrounding islands,* **15**. (eds Coetzee JA, van Zinden Bakker EM), pp. 217-223. AA Balkema, Rotterdam.

Endrödy-Younga S (1978) Coleoptera. In: *Biogeography and ecology of Southern Africa*. (eds Werger MJA), pp. 797-821. W Junk, The Hague.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-491.

Felsenstein J (1973) Maximum likelihood and minimum-steps methods for estimating evolutionary treesfrom data on discrete characters. *Systematic Zoology*, **22**, 240-249.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368-376.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783-791.

Frost DR, Hillis DM (1990) Species in concept and practice: herpetological considerations. *Herpetologia*, **46**, 87-104.

Fu Y-X (1994a) A phylogenetic estimator of effective population size or mutation rate. *Genetics*, **136**, 685-692.

Fu Y-X (1994b) Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. *Genetics*, **138**, 1375-1386.

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

Graur D, Martin W (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics*, **20**, 80-86.

Harrison JduG (1999) Systematics of the endemic south-west African dung beetle genus *Pachysoma* MacLeay (Scarabaeidae: Scarabaeinae). MSc Thesis, University of Pretoria.

Harrison JduG, Phillips TK (2003) Phylogeny of Scarabaeus (Pachysoma MacLeay) sta. nov., and related flightless Scarabaeini (Scarabaeidae: Scarabaeinae). *Annals of the Transvaal Museum*, **40**, 47-71.

Harrison JduG, Scholtz CH., Chown SL (2003) A revision of the endemic south-western African dung beetle subgenus *Scarabaeus (Pacgysoma)* MacLeay, including notes on other flightless Scarabaeini (Scarabaeidae: Scarabaeinae). *Journal of Natural History*, **37**, 305-355.

Harpending RC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591-600.

Holm E, Scholtz CH (1979) A revision of the genus *Pachysoma* M'Leay with an evaluation of the subtribe Pachysomina Ferreira and its genera (Coleoptera: Scarabaeidae). *Journal of the Entomological Society of South Africa*, **42**, 225-244.

Irish J (1990) Namib Biogeography, as exemplified mainly by the Lepmismatidae (Thysanura: Insecta). In: *Namib Ecology: 25 years of Namib Research.* (eds Seely MK), pp. 61-66. Transvaal Museum Monograph No. 7. Transvaal Museum, Pretoria.

Juan C, Ibrahim KM, Oromi P, Hewitt GM (1998) The phylogeography of the darkling beetle, *Hegeter politus*, in the eastern Canary Islands. *The Royal Society of London*, **265**, 135-140.

Kirchman JJ, Whittingham LA, Sheldon FH (2000) Relationships among cave swallow populations (*Petrochelidon fulva*) determined by comparisons of microsatellite and cytochrome *b* data. *Molecular Phylogenetics and Evolution*, **14**, 107-121.

Kluge AG, Farris JS (1969) Quantitative phylogenetics and the evolution of anurans. *Systematic Zoology*, **18**, 1-32.

Knutsen H, Rukke BA, Jorde PE, Ims RA (2000) Genetic differentiation among populations of the beetle *Bolitophagus reticulates* (Coleoptera: Tennebrionidae) in a fragmented and a continuous landscape. *Heredity*, **84**, 667-676.

Kuhner MK, Yamato J, Beerli P *et al*. (2004) *LAMARC v 1.2.1*. University of Washington, http://evolution.gs.washington.edu/lamarc.html.

Kumar S, Tamura K, Jacobsen IB, Nei M (2001) Molecular Evolutionary Genetic Analysis Software. *Bioinformatics*, **17**, 1244-1245.

Lavery S, Moritz C, Fielder DR (1996) Genetic patterns suggest exponential population growth in a declining species. *Molecular Biology and Evolution*, **13**, 1106-1113.

Liedloff A (1999) Mantel nonparametric test calculator for Windows Version 2.00. School of Natural Resource Sciences, Queensland University of Technology.

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

Mayr E, Ashlock PD (1991) *Principles of Systematic Zoology*. McGraw Hill, Inc.

Moritz C (1994a) Defining evolutionary significant units for conservation. *Trends in Ecology and Evolution*, **9**, 373-375.

Moritz C (1994b) Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, **3**, 401-411.

Moritz C (1999) Conservation units and translocations: strategies for conserving evolutionary processes. *Hereditas*, **130**, 217-228.

Moya Ó, Contreras-Díaz HG, Oromí P, Juan C (2004) Genetic structure, phylogeography and demography of two ground-beetle species endemic to Tenerife laurel forest (Canary Islands). *Molecular Ecology*, **13**, 3153-3167.

Nahum LA, Perreira SL, de Campos Fernandes FM, Matioli SR, Wajntal A (2003) Diversification of Ramphastinae (Aves, Ramphastidae) prior to the Cretaceous/Tertiary boundary as shown by molecular clock of mtDNA sequences. *Genetics and Molecular Biology*, **26**, 411-418.

Nice CC, Anthony N, Gelembiuk G, Raterman D, Ffrench-Constant R (2005) The history and geography of diversification within the butterfly genus Lycaeides in North America. *Molecular Ecology*, **14**, 1741-1754.

Nixon KC, Wheeler QD (1990) An amplification of the phylogenetic species concept. *Cladistics*, **6**, 211-223.

Penrith M-L (1984) New taxa of *Onymacris* Allard, and relationships within the genus (Coleoptera: Tenebrionidae). *Annals of the Transvaal Museum*, **33**, 511-533.

Penrith M-L (1986) Revision of the Zophosini (Coleoptera: Tenebrionidae). Part 10. Key to the subgenera, supplement, evolution and biogeography of the tribe, and catalogue. *Cimbebasia (A)*, **6**, 417-502.

Posada D, Crandall KA (1998) Modeltest 3.0: testing the model of DNA substitution. *Bioinformatics,* **14**, 817-818.

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

Posada D, Crandall KA, Templeton AR (2000) GeoDis: A program for the Cladistic Nested Analysis of the Geographical Distribution of Genetic Haplotypes. *Molecular Ecology,* **9**, 487-488.

Rogers AR (1995) Population forecasting: do simple models outperform complex models? *Mathematical Population Studies*, **5**, 187-202.

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

Rooney AP, Honeycutt RL, Derr JN (2001) Historical population size change of Bowhead whales inferred from DNA sequence polymorphism data. *Evolution*, **55**, 1678-1685.

Satou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenic trees. *Molecular Biology and Evolution*, **4**, 406-425.

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

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

Seely MK (eds) (1990) *Namib Ecology 25 years of Namib Research*. Transvaal Museum Monograph No. 7, Transvaal Museum, Pretoria.

Simon C, Frati, F, Benckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Soceity of America*, **87**, 652-701.

Smith CI, Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate change. *Molecular Ecology*, **14**, 1025-1044.

Sole CL, Scholtz CH, Bastos ADS (2005) Phylogeography of the Namib Desert dung beetles *Scarabaeus (Pachysoma)* MacLeay (Coleoptera:Scarabaeidae). *Journal of Biogeography*, **32**, 75-84.

Sole CL (2005) The Phylogeography of *Scarabaeus (Pachysoma)*. PhD Thesis, The University of Pretoria.

Su B, Fu Y, Wang Y, Jin L, Chakraborty R (2001) Genetic diversity and population history of the Red Panda (*Ailurus fulgens*) as inferred from mitochondrial DNA sequence variations. *Molecular Biology and Evolution*, **13**, 1070-1076.

Swofford DL (1998) PAUP*. *Phylogentic Analysis using Parsimony, Beta version 4.0b1*. Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics*, **105**, 437-460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesisy DNA polymorphism. *Genetics*, **123**, 597-601.

Tankard AJ, Rogers J (1978) Late Cenozoic palaeoenvironments on the west coast of southern Africa. *Journal of Biogeography*, **5**, 319-337.

Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*, **10**, 779-791.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 1-23.

Templeton AR, Crandall KA, Sing CF (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 AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analysis with cladogram uncertainty and recombination. *Genetics*, **134**, 659-669.

Templeton AR, Routman E, Phillips CA (1995) Separating populations structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in Tiger Salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767-782.

Van Dongen S, Backeljau T, Matthysen E, Dhondt AA (1998) Genetic population structure of the winter moth (*Operophtera brumata* L.) (Lepidoptera, Geometridae) in a fragmented landscape. *Heredity*, **80**, 92-100.

Van Zinderen Bakker EM (1975) The origin and palaeoenvironment of the Namib Desert biome. *Journal of Biogeography*, **2**, 65-73.

Wiley J (1981) *Phylogenetics: The theory and practise of phylogenetic systematics.* New York: John Wiley and Sons.

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

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323-354.

Wright S (1965) The interpretation of population structure by F – statistics with special regards to systems of mating. *Evolution*, **19**, 395-420.

Yang Z, Goldman N, Friday A (1994) Comparison of models from nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Molecular Biology and Evolution*, **11**, 316-324.

Zheng X, Arbogast BS, Kenagy GJ (2003) Historical demography and genetic structure of sister species: deermice (*Peromyscus*) in the North American temperate rain forest. *Molecular Ecology*, **12**, 711-724.