Abstract

CRISSMAN, JONATHAN ROSS. Population Genetics of the German Cockroach, *Blattella germanica* L., in Low-Income Urban Apartments. (Under the direction of Dr. Coby Schal and Dr. Edward L. Vargo.)

The goal of the current study was to explore German cockroach population differentiation and dispersal in urban residences at several spatial scales. Little differentiation was expected within buildings due to frequent active dispersal of the German cockroach. However, differentiation was expected between spatially separated buildings because the German cockroach cannot survive beyond human structures and dispersal must be human-mediated. Differentiation was expected to increase with increasing spatial distance as human-mediated dispersal becomes successively less frequent.

Cockroaches were collected in a geographical hierarchy, with samples at the scale of single apartments, apartment buildings, and apartment complexes within Raleigh, North Carolina in addition to 16 other cities in the United States and 6 Eurasian cities.

All individuals were genotyped at 10 microsatellite loci. Microsatellite diversity was generally high across our samples, with average alleles per locus as high as 9.11, and observed heterozygosity in excess of 0.6 in most populations. However, global differentiation was lower than expected, with a global F_{ST} of 0.05 across Raleigh, 0.10 across the U.S. with Raleigh excluded, and 0.16 across global samples with Raleigh excluded.

No departure from panmixia was found between German cockroach samples within apartments by pairwise *G*-tests ($P > 0.05$). *G*-tests did detect highly significant departures from panmixia between all but one pair of apartments within Raleigh (all $P < 0.001$). These tests indicate that each separate apartment represents an individual population.

Global F_{ST} values suggested that dispersal and gene flow occurred more often within buildings than between them and more within apartment complexes than between them. Similarly, significant Mantel tests of isolation by distance within Raleigh ($P < 0.001$), across the U.S. ($P = 0.029$), and across global samples ($P < 0.005$) indicated a significant positive correlation between genetic distance and geographic distance. These tests suggested that German cockroach dispersal and gene flow occurred less frequently at successively larger spatial scales.

However, phylogenetic trees constructed from genetic distances for each spatial scale were unable to cluster apartments according to geographic proximity, except for apartments within a single building. The individual-based clustering program *Structure* was also unable to detect any informative underlying structure beyond a single building when no *a priori* population information was used. A similar lack of geographic pattern was shown in *B. germanica* mitochondrial 16S diversity. Only 10 unique 16S haplotypes were found for 66 individuals across 40 global populations, and haplotypes varied greatly in their frequency and geographic ranges.

Beyond the scale of a single building, gene flow by passive dispersal seems to have formed no spatial clustering of genetic similarity, even at the global scale. This may be due to non-equilibrium conditions and non-linear evolution caused by dramatic declines in population size due to insecticides or large fluctuations in temperature, humidity, and food availability. These population bottlenecks are known to rapidly change allele frequencies and affect various evolutionary estimates, including genetic distance. The challenge for future studies will be to find markers conserved enough to be robust to these non-equilibrium conditions but variable enough for fine scale analysis.

Keywords: *Blattella germanica*; microsatellites; population genetics; human-mediated dispersal; commensal; non-equilibrium

Population Genetics of the German Cockroach, *Blattella germanica* L., in Low-Income Urban Apartments

by
Jonathan Ross Crissman

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Entomology

Raleigh, North Carolina

2008

APPROVED BY:

____
David Tarpy

Coby Schal	Edward Vargo
Co-chair of Advisory Committee	Co-chair of Advisory Committee

Biography

Jonathan Ross Crissman was born in Virginia in January 1981. He lived in Gloucester, Virginia until his graduation from Gloucester High School in 1999. After an abortive attempt to attain a business degree he completed a degree in Biology from Virginia Commonwealth University in 2005. After graduation, Jonathan worked in an entomology lab for several months, took some time to travel, then came to NC State University to pursue a Master of Science in Entomology. Now that he has earned that degree, he does not see himself extracting cockroach DNA again for the foreseeable future.

Table of Contents

List of Tables

List of Figures

Chapter I: Literature Review

Molecular ecology and genetic markers

The field of molecular ecology arose from the broader discipline of population genetics, founded in the early twentieth century by Sewell Wright, J.B.S. Haldane, and Ronald Fisher. Population genetics is concerned with the study of genetic variation within and between populations as it is shaped by the combined forces of mutation, drift, selection, and gene flow (Hedrick 2005b). Molecular ecologists apply the principles of population genetics with the use of molecular genetic markers, which allow the researcher to assign alleles to individuals, and allele frequencies to populations (Carvalho 1998). Several comprehensive summaries of the history and theory of population genetics have been written, including the recommended volumes by Carvalho (1998), Avise (2004) and Hedrick (2005a).

The ideal marker for quantifying and comparing changes in genetic variability due to various evolutionary forces would be inherited in a predictable fashion, would possess discrete and identifiable states, and would be sufficiently polymorphic to distinguish individuals, populations, and species (Carvalho 1998).

Allozymes. The first markers to be used in population genetics were variable enzyme products known as allozymes (Harris 1966; Hubby & Lewontin 1966; Lewontin & Hubby 1966). Allozymes can be found in all species and are inherited in Mendelian patterns (Sunnucks 2000). Their variability allows inferences on the forces acting on populations such as selection and drift, as well as gene flow between populations (Schlötterer 2004). They are technically advantageous as well, requiring little equipment or training. The enzymes are easily extracted and then separated by use of gel electrophoresis.

The large amount of protein required for electrophoresis is a disadvantage because samples must be destroyed for protein extraction (Carvalho 1998). Difficulties finding variable loci can be another constraint because some species are monomorphic for up to half of all loci; and most loci do not have more than three alleles in any population and may be monomorphic in some (Parker et al. 1998). These drawbacks make allozymes less attractive for detecting fine scale population structure. Also, being a phenotypic character means that allozymes may be subject to selection, and some allozyme loci have displayed evidence of selective pressure (e.g., Berry & Kreitman 1993). The presence of selection would violate the assumptions of many population genetics models.

RFLPs. The discovery of bacterial restriction enzymes made possible restriction fragment length polymorphisms (RFLPs), the first DNA approach in molecular ecology (Avise 2004). Restriction enzymes recognize a specific DNA sequence and cut the DNA at each occurrence of this sequence, resulting in a large number of DNA pieces which can be separated by electrophoresis. The specific pattern in the gel can be used to distinguish individuals. Because of the large number of bands that are produced by restriction cutting of nuclear DNA, the much smaller mitochondrial DNA (mtDNA) molecule is often used instead (Hedrick 2005b). Even with the use of the smaller molecule, however, the number of bands needed to be reduced through the use of specific hybridization probes and Southern blotting techniques, which greatly increase the amount of time and labor as well as the cost of processing large numbers of samples (Parker et al. 1998).

RFLPs have a number of analytical concerns as well. RFLPs are known as multilocus markers because each restriction enzyme cuts multiple places in the genome making the assignment of allele frequencies to specific loci impossible (Avise 2004). RFLPs also show dominant inheritance, meaning that each individual can only be scored for presence or absence of a specific site, rather than for a specific allele. Multilocus dominant markers typically produce banding patterns that can only be compared between individuals within a single study (Sunnucks 2000).

The polymerase chain reaction (PCR, Mullis et al. 1986) for the amplification of DNA was a significant discovery for the study of molecular ecology. Millions of copies of specific sequences can be amplified from a few nanograms of DNA, meaning that small organisms may be studied and the tissue requirement is very small (Carvalho et al. 1998).

AFLPs. The application of PCR to RFLP products led to the development of amplified fragment length polymorphisms (AFLPs; Vos et al. 1995). By attaching linkers to the ends of the fragments after restriction cutting, a reduced number of specific fragments can be amplified by PCR without the use of hybridization or Southern blotting. AFLPs share all of the disadvantages of RFLPs with regard to multilocus amplification and dominant inheritance.

RAPDs. Another early application of PCR in molecular markers was the use of randomly amplified polymorphic DNA (RAPDs; Williams et al. 1990). RAPD primers are typically 10 base pairs in length and only one PCR primer is used at a time rather than the usual two. As their name suggests, they recognize and amplify sequences complementary to the primer sequence at random areas of the genome, therefore primers are designed with no prior knowledge of DNA sequence (Welsh & McClelland 1990). RAPD products can be obtained in a single PCR step, unlike AFLPs. Like AFLPs, however, RAPDs are codominant and multilocus markers and heterozygotes are distinguishable from homozygotes only by band intensity (Jarne & Lagoda 1996). RAPDs have also been shown to be unreliable and too infrequently reproducible (Schierwater & Ender 1993), and are discouraged from use by some journals.

VNTRs. Variable number tandem repeats (VNTRs) are segments of genomic DNA which contain 2–64 base pair DNA sequences repeated in tandem. Because they can be highly variable for the number of repeat units, they have become highly useful in fine scale molecular ecology analysis (Estoup & Angers 1998). VNTRs are typically divided into microsatellites (also known as short tandem repeats, STRs), which consist of repeat units of 2–6 base pairs, and minisatellites, which consist of larger repeat units.

Minisatellites. Minisatellites were first used in a multilocus approach similar to that used for RFLPs, which requires the use of Southern blotting and radiolabeled probes (Jeffreys et al. 1985a, b). The banding patterns produced by this method are so complex that they are usually unique to each individual, leading this method to be termed 'fingerprinting'. Despite their effectiveness in forensic analysis (Gill et al. 1985), population level analysis with minisatellites has been difficult because restriction cutting produces multilocus banding patterns (Carvalho 1998).

Single locus amplification of minisatellite regions was later developed (Armour et al. 1990). The length of minisatellite arrays makes obtaining a reliable PCR product difficult and the size difference between fragments is typically small relative to the length of the array, making it difficult to resolve them by gel electrophoresis (Estoup & Angers 1998).

mtDNA. Mitochondrial DNA (mtDNA) consists of a 15–20 kilobase circular genome, nearly the entire sequence consisting of its 37 coding genes. The rise of mtDNA as a marker corresponded to its use in the field of phylogeography, which involves creating gene genealogies for the study of intraspecific populations (Avise et al. 1987). One reason for the utility of mtDNA for genealogies is the fact that it is only maternally transmitted and does not recombine (Sunnucks 2000) and therefore provides a clear phylogenetic signal. The maternal transmission and lack of recombination also mean that it has a ¼ effective population size, so it tends to become diagnostic of a certain lineage much more quickly than nuclear loci due to increased genetic drift (Hedrick 2005b). Another benefit of animal mtDNA is that it mutates at a much greater rate than coding nuclear DNA (Parker et al. 1998).

As mentioned above, much of the early studies using mtDNA were done with the use of restriction enzymes, but direct sequencing and PCR approaches are now more common. One potential setback to the use of these PCR-based approaches is the discovery of mitochondrial 'pseudogenes', copies of mitochondrial genes found in the nuclear genome (Zhang & Hewitt 2003). The theoretical basis of the use of mtDNA genealogy in phylogeography has also been questioned (discussed below). In short, statistical evidence has been mounting that suggests that single locus gene trees are insufficient for the inference of the evolutionary and demographic history of populations (Brumfield et al. 2003).

SNPs. One of the most promising markers for population level analysis to emerge in recent years is single nucleotide polymorphisms (SNPs), which are single base variations between individuals. SNPs have a number of advantages which may make them superior to microsatellites and mtDNA for population analysis. They exhibit low variability in mutation rates between loci, their mutation is well understood, and they are less likely to display homoplasy (Brumfield et al. 2003).

Some barriers still exist to their widespread use, however. SNPs are subject to 'ascertainment bias', which results from selecting only the most polymorphic loci and from screening for polymorphism from a limited sample of individuals (Nielsen 2000). Fortunately, statistical corrections now exist for most forms of ascertainment bias. Also, screening for SNPs is a laborious process, usually involving cloning of overlapping genome sequences (Brumfield et al. 2003). In addition, since SNPs are usually biallelic markers they inherently provide less information than microsatellites, and as many as 2–4 times as many SNPs may be necessary for fine scale analysis (Schlötterer 2004). In the years to come the potential ability to quickly assay hundreds of SNPs may make them the marker of choice for population genetic analysis.

DNA Sequencing. While many of the DNA markers mentioned above provide a measure of genetic variation, they are still indirect and incomplete, at best (Avise 2004). Although not typically considered a genetic 'marker', direct sequencing offers the highest possible resolution of genetic variation, with no bias (Schlötterer 2004), and thus warrants a mention here. Most current sequencing is still derived from the PCR and electrophoresis based method of Sanger (1977), but new methods may soon make direct sequencing cheap and efficient enough to assay large numbers of samples for fine-scale analysis (Parker et al. 1998).

Use and applications of microsatellites

Microsatellites were the first widespread PCR-based marker (Tautz 1989; Weber & May 1989). They have since become the most used molecular marker for population genetic analysis, having been discovered to occur frequently in the genomes of organisms from nearly every taxonomic group (Estoup & Angers 1998). Besides their ubiquity, microsatellites have a number of characteristics that are favorable to researchers. Microsatellite loci typically exhibit high levels of genetic variability, with heterozygosity proportions above 0.5 and usually much higher (Jarne & Lagoda 1996). The high level of microsatellite variability means that individual genotypes are often unique, making possible studies of kinship, breeding systems, paternity analysis, and forensics (Sunnucks 2000). Microsatellites are also single locus markers with codominant inheritance, allowing assignments of allele frequencies for population level analysis (Estoup & Angers 1998).

In contrast to minisatellites, total microsatellite repeat arrays are typically in the range of only 10–200 base pairs in length because of their smaller repeat length, making them much more amenable to amplification by PCR and separation by electrophoresis than minisatellites (Schlötterer 2004). In addition, techniques have been developed which allow for many loci to be analyzed at once (multiplexing), which includes conducting multiple PCR reactions on a single DNA sample simultaneously, and loading multiple sets of samples onto an electrophoretic gel simultaneously (Carvalho 1998).

Although the single locus nature of microsatellites is an advantage for analysis, initial development of individual PCR primers for each locus is a major disadvantage (Jarne & Lagoda 1996). Discovery of polymorphic repeat regions is typically a labor intensive and time consuming process, involving probing of genomic DNA and bacterial cloning (Estoup & Angers 1998). Despite this initial difficulty, large numbers of samples can be processed very quickly once the primers have been developed. In addition, primer sequences are typically published in public databases for use by other researchers, and their specificity makes direct comparison between studies possible (Sunnucks 2000). Primers can usually be used across populations without issue, but mutations in the primer region of a specific allele can cause a PCR failure, a problem known as a null allele. If null alleles are sufficiently widespread and uncorrected they can lead to severe overestimates of differentiation and genetic distance (Chapuis & Estoup 2006). Inefficient PCR amplification may also lead to multiple products with varying numbers of repeat units, resulting in multiple consecutive bands on a gel, known as stutter bands (Chambers & MacAvoy 2000). They are particularly common in dinucleotide repeats, and introduce ambiguity as to the identity of the real allele.

Although microsatellites are assumed to be selectively neutral, they can be found within expressed DNA regions and some have been shown to have functional impacts, including disease generation (Sutherland & Richards 1995). Since microsatellites are randomly cloned and most occur in unexpressed DNA, it is assumed that most microsatellites will be selectively neutral and tests exist to test for non-neutrality (Sunnucks 2000). Even non-functional microsatellites may be subject to selection. It has been suggested that selection against large microsatellite alleles may contribute to allele size constraints (Garze et al. 1995). In addition, microsatellites may be subject to hitchhiking by linkage to loci under selective pressure (Charlesworth et al. 1993).

Microsatellite primers may also work for species very closely related to the species for which they were developed, but often with decreased diversity and a greater presence of null alleles (Estoup & Angers 1998). Finally, one of the greatest drawbacks of microsatellites for population genetic analysis is the uncertainty surrounding their mutation, which is discussed in greater detail below.

Genetic variation

Genetic variation is the fundamental component of the study of population genetics, and the theoretical basis to the study of genetic variation is given by the Hardy-Weinberg principle, which defines the relationship of genotype frequencies to allele frequencies in a population which is in equilibrium (Hardy 1908; Weinberg 1908). Hardy-Weinberg equilibrium assumes that a population experiences random mating and is free from selection, mutation, or gene flow. Therefore, any significant departure from the genotypic frequencies indicates that one or more of the assumptions of Hardy-Weinberg equilibrium has been violated. This theory underlies the most widely used measure of genetic variation, heterozygosity (H), which applies to the calculation of the proportion of heterozygotes observed (H_o) in the population at a particular locus, as well as the proportion which would be expected (H_e) under Hardy-Weinberg equilibrium (Hedrick 2005b). When conducting a study using multiple loci, heterozygosity can be averaged across them. Nei (1986) suggested variance in the heterozygosity calculation is more effectively reduced by using more loci than by sampling more individuals per population.

Other measures of genetic variation include the proportion of polymorphic loci (*P*), and allelic richness (*A*; also known as allele diversity or number of alleles). Proportion of polymorphic loci gives the fraction of loci in a given population which have multiple alleles. This measure is rarely used in modern studies because monomorphic markers are rarely discovered or are discarded before the study begins (Avise 2004). Allelic richness is simply a count of the alleles at a given locus in a given population. This measure is often strongly influenced by sample size and can be influenced by stochastic population processes (Nei et al. 1975). Variance in allelic richness due to sample size may be corrected by standardizing to the smallest sample size (El Mousadik & Petit 1996).

Evolutionary changes to genetic variation

The forces that act to change the amount and pattern of genetic variation have differing effects. Genetic drift and inbreeding will always decrease variation within a population, mutation always increases variation, while selection and gene flow may work to either increase or decrease variation (Hedrick 2005b).

Selection. Natural selection can either increase the level of genetic variation at some loci and, unlike genetic drift, inbreeding or gene flow, will not have the same average effect on all loci (Beaumont & Balding 2004). Selection is assumed to be absent under the Hardy-Weinberg principle and many genetic markers such as microsatellites are assumed to be neutral under many population genetic models. Therefore, if this assumption is violated inferences about population structure and gene flow may be misleading because selection can potentially be as important to patterns of variation as drift or gene flow (Hare 2001). For example, distinct differentiation between two populations that appears to be due to isolation and drift may in fact be due to local adaptation if the locus is under selection (Slatkin 1987).

As mentioned, selection acts only on certain loci and many tests to detect selection rely on differential amounts of variation between loci or between populations. The commonly used method of Tajima (1989) compares the number of sequence differences within populations, and Beaumont & Nichols (1996) developed a coalescent-based approach for detecting outlier loci which may not be selectively neutral.

Even loci which are not directly under selective pressure may be affected by selection if their position in the genome is close to a locus under selection. "Selective sweeps" occur when positive selection increases a particular allele and removes variation at adjacent loci, a process also known as genetic "hitchhiking" (Avise 2004). A similar phenomenon, "background selection", occurs when selection against deleterious alleles removes variation in adjacent loci (Charlesworth et al. 1993).

Linkage disequilibrium. The effect of selection at one locus that affects allele variation at a nearby locus is known as linkage disequilibrium, the nonrandom association between alleles at different loci. Generally this occurs by hitchhiking, when loci are located too closely together on the same chromosome for recombination to frequently separate them (Marjoram & Tavaré 2006). Population genetic parameters which depend on calculation of allele frequencies typically assume independent allele frequencies for each locus, but this assumption is invalid for loci which are in linkage disequilibrium (Avise 2004).

Linkage disequilibrium may also be generated by admixture, or gene flow between differentiated populations (Slatkin 1987). If drift has been acting separately on each population, individuals will continue to show correlations between loci because of the great difference in their ancestral allele frequencies relative to the population to which they have migrated (Falush et al. 2003).

Mutation. Mutation is the original source of all genetic variation. Each individual mutation occurs within a single individual who may pass it on to offspring. Even if a mutant allele is passed on, most are expected to be lost due to genetic drift, even those with a slight selective advantage. Since probability of fixation is equal to the initial frequency in the population, the smaller the population the greater the chance of fixation (Hedrick 2005b).

Various models have been proposed to account for the way in which mutations arise and accumulate within populations assuming selective neutrality. The first of these was the infinite allele model (IAM, Kimura & Crow 1964), which assumes that every new mutation results in a new allele not previously present in the population. The IAM also assumes that this new allele may be of any size, without limit. The alternative to this hypothesis was given by the stepwise mutation model (SMM, Ohta & Kimura 1973) which assumes that each mutation creates an allele one step away from the allele from which it mutated. Unlike the IAM, the SMM allows for the mutation of alleles toward preexisting states. This accounts for the possibility of homoplasy, alleles that are identical in state rather that identical due to common ancestry. Size homoplasy is not accounted for under the TPM or IAM but has been detected in empirical studies (Estoup et al. 1995b, Garza & Freimer 1996). A revision of the IAM which allowed for homoplasy was also suggested. Called the K-allele model (KAM, Kimura 1968), this model allows for K possible allele states, with each allele being equally probable to mutate to any of the other states, including those already present in the population.

The two extremes of the SMM and the IAM can give vastly different estimates for various population parameters. Expected heterozygosity would be much lower under SMM than the IAM because the SMM assumes that some independent mutations will create identical alleles (Balloux & Lugon-Moulin 2002). The difference between the two models can also have a large effect on the estimate of effective population size and time since divergence of two populations.

Microsatellite mutation. Microsatellites mutate at a very high rate, thought to be in the range of 10^{-5} to 10^{-2} (Jarne & Lagoda 1996), most involving slippage during replication (Primmer et al. 1996). This high mutation rate gives microsatellites the favorable attribute of very high variation. But it also means that the choice of mutation model is important for inferences of long term evolutionary history, including effective population size, time since differentiation, and migration rates (Estoup & Angers 1998).

Selection of a mutation model is complicated by the observation that mutation processes do not appear to be uniform amongst loci (Ellegren 2000a). While the SMM has traditionally been thought to best explain mutational variation in most microsatellites (Primmer et al. 1996), direct sequencing of alleles has shown that repeat number does not reliably correlate to phylogenetic relationships between populations according to the SMM (Ortí et al. 1997). Contrary to SMM expectations, Huang et al. (2002) found that 63% of mutations in humans were greater than one step, with an average of 2.6 steps per mutation. In addition, alleles consisting of multiple sizes of repeat unit (compound microsatellites) may conform better to the IAM (Estoup et al. 1995a). A two phase model (TPM, Di Rienzo et al. 1994), which allows for given proportions of both one-step and multi-step mutations, was specifically created to model microsatellite mutation. Studies have shown that the TPM may be the most realistic model so far for generalizing the mutation of microsatellites (Ellegren 2000b).

Rate of mutation appears to vary between microsatellite loci as well. Mutation rate rises with repeat motif length (Primmer et al. 1996; Schlötterer 1998), and may cease completely if the number of repeats drops below a threshold. Mutation rates also seem to be negatively correlated with repeat unit size, with dinucleotide repeats mutating as much as twice as fast as tetranucleotide repeats (Chakraborty et al. 1997).

Finally, microsatellite mutation appears to exhibit a directional bias. Evidence suggests that there is an upper constraint on repeat number. If all mutation is truly neutral then extremely high repeat number should be possible for highly conserved loci, but repeat number greater than 60 is rarely observed (Goldstein & Pollock 1997). This can be explained in part by the fact that longer motifs are biased towards reduction in repeat number (Huang et al. 2002; Xu et al. 2000), while short repeat motifs show bias towards increase in allele size (Huang et al. 2002). Selection against large alleles may also play a role in limiting the maximum allele size (Garza et al. 1995; Nauta & Weissing 1996).

Simulations have shown that allele size constraints greatly increase the occurrence of size homoplasy, particularly when population sizes are large and mutation rates are high (Estoup et al. 2002). Also, loci under severe allele size constraints are expected to conform more to a KAM specific to their mutation rate than a simple SMM (Rousset 1996).

Genetic drift. Genetic drift is the change in allele frequencies due to the random inheritance of alleles from generation to generation. The change in frequency of any given allele is random, but drift always leads to fixation of a single allele at each locus with the extinction of all others, with the probability of an allele becoming fixed being proportional to its initial frequency in the population (Hedrick 2005b). Drift occurs much more rapidly in small populations.

Bottlenecks and founder effects. Other stochastic processes such as bottlenecks or founder effects which reduce population size can put populations at greater danger of further loss of genetic diversity due to drift. A bottleneck is a sudden decrease in population size, which is expected to severely reduce overall heterozygosity and many alleles are likely to have gone extinct. Following a bottleneck, the number of alleles per locus is expected to recover much more rapidly than heterozygosity as the population grows (Nei et al. 1975). This principle has been shown to be useful for detecting the occurrence of recent population bottlenecks (Cornuet & Luikart 1996; Luikart & Cornuet 1998). A similar genetic phenomenon, called a founder effect, occurs when a small group of individuals is separated from a larger population and begins a new population (Hedrick 2005b). Like a bottleneck, the founding individuals represent a small subset of the genetic diversity present in the source population, with a reduction in the number of rare alleles. One consequence of the decrease in genetic diversity due to bottlenecks and founder effects is that genetic distance will be quickly increased between the affected population and all other populations (Paetkau et al. 1997). This will distort evolutionary estimates between recently diverged populations because allele frequencies have rapidly changed.

Inbreeding. Another way in which genetic diversity can be lost in a population is by inbreeding, the mating between individuals in a population which are more related than would be predicted by chance. Inbreeding causes a decrease in heterozygosity from Hardy-Weinberg proportions as individuals are likely to inherit the same allele from both parents because both share common ancestry (Avise 2004). Classical *F*-statistics (Wright 1951; Weir & Cockerham 1984) provide measures of inbreeding called inbreeding coefficients, which are meant to measure this increase in homozygosity.

Effective population size. Decrease in genetic diversity by drift or inbreeding underlies the concept of effective population size, N_e. The effective population size estimates the number of individuals in a hypothetical population undergoing random mating and in Hardy-Weinberg equilibrium which would give the same amount of genetic variance in the sampled population (Wright 1931). Because nearly all populations differ from ideal Hardy-Weinberg assumptions, and because not all individuals are reproductively active, this effective population size is nearly always smaller than the true population size. This is especially true in inbreeding populations or after bottlenecks or founder events which lead to a reduction in heterozygosity (Avise 2004).

The use of highly variable, selectively neutral markers such as microsatellites can increase the accuracy of N_e estimates, particularly for small populations (Estoup & Angers 1998). The classical methods of estimating N_e rely on Wright's (1940) island model of migration. This model makes assumptions – such as low rates of gene flow, mutation, and migration – that cannot often be met in natural populations (Whitlock & McCauley 1999). The method of Waples (1989) uses changes in allele frequencies over time, and recently developed methods have made use of likelihood-based statistics (e.g., Beaumont 2003). Another positive advance has been the use of models which relax some of the assumptions inherent in previous models, such as allowing for the presence of gene flow (Wang & Whitlock 2003), and lack of Hardy-Weinberg equilibrium (Beerli & Felsenstein 1999, 2001). These relaxed methods can give widely differing estimates of N_e compared to previous approaches, and the method which gives the most accurate estimate may depend on the particular system being studied (Fraser et al. 2007).

Gene flow. Gene flow is the movement of alleles from one population to another, and tends to homogenize allele frequencies between populations which share migrants (Hedrick 2005b). In this way gene flow can act to counterbalance the effect of natural selection and genetic drift by inhibiting the fixation of certain alleles that would occur under these forces (Slatkin 1987).

This balance between genetic drift and gene flow has been integral to the concept of population differentiation since the earliest formulations of population genetic theory

(Wright 1931). It is no surprise then that measures of gene flow have traditionally been derived from the classical measure of population differentiation, F_{ST}, derived by Wright (1951). This approach involves characterizing the extent of differentiation between two populations as the interplay between the effective population size, N_e, and migration, m (Fraser et al. 2007). The use of F_{ST} in studies of population differentiation, and the criticisms of its use for gene flow are discussed below.

As with estimates of N_e, new models employing coalescent theory and likelihood statistics have been introduced in recent years (e.g., Beerli & Felsenstein 2001; Nielsen & Wakeley 2001; Wilson & Rannala 2003). These methods can often estimate multiple parameters at once and relax certain assumptions, such as Hardy-Weinberg equilibrium or migration-drift equilibrium.

Population differentiation

Relationships between populations may be very dynamic. Populations may actively split by dispersal or be divided into smaller populations by environmental factors. They may be isolated from some populations and may exchange individuals with other populations. Wright (1931) recognized differentiation between populations as being mainly a balance between drift and gene flow. Each population experiences its own rate of mutation and drift, which tends to differentiate populations from one another, while gene flow between them makes them more similar (Hey & Machado 2003). The balance between selection and drift is independent of population size (Slatkin 1987). The effects of both gene flow and drift are strong in small populations and weak in large populations. Several models have been designed to simplify the balance between these forces to more easily study their effects on the genetic structure of several populations.

Migration models. The island model of migration assumes a large ‘mainland’ population which exchanges migrants with smaller ‘island’ populations (Wright 1940). The size of each population determines the rate at which it will diverge from the others by genetic drift, while rates of gene flow determine levels of genetic similarity. While island models assume equal migration to all islands, stepping-stone models take into account the spatial distribution of populations, with gene flow only occurring between adjacent populations (Kimura & Weiss 1964). The simplest stepping-stone model has populations aligned in one dimension, but they may be two dimensional or three dimensional as well. The restriction of gene flow due to the stepping stone model leads to an isolation by distance model, under which more geographically proximate populations are more closely related due to increased gene flow (Wright 1943). Tests of isolation by distance involve plotting the relationship between genetic distance and geographic distance (Slatkin 1993; Rousset 1997), with a linear relationship between the two measures suggesting that isolation by distance is present. Testing isolation by distance at multiple meaningful geographic regions or scales rather than a whole study area can help tease apart the relative effects of gene flow and drift at each region (Hutchinson &

Templeton 1999). The statistical significance of the relationship between genetic distance and geographic distance is often tested using a Mantel permutation test (Mantel 1967).

More recently, attempts have been made to model widely distributed populations, called metapopulations, which consist of several distinct subpopulations (Wade & McCauley 1988). These models attempt to account for the fact that metapopulation structure is shaped by the formation and extinction of subpopulations as well as differing levels of gene flow between the various subpopulations.

Indirect measures of differentiation. The classical approach for measuring differentiation between divided populations is the use of *F*-statistics, pioneered by Wright (1951). Such methods are called 'indirect' because they rely on estimates of population-wide allele frequencies. Wright divided total sampled genetic variation, F_{IT}, into the amount due to differences between subpopulations, F_{ST}, and that due to differences between individuals within subpopulations, F_{IS}. Therefore, F_{ST} is a measure of differentiation between subpopulations, and F_{IT} and F_{IS} can detect a heterozygote deficiency according to Hardy-Weinberg expectations. The most frequently used analogue of F_{ST}, Θ, was developed by Weir & Cockerham (1984) using an analysis of variance (ANOVA) framework. Nei (1973) developed G_{ST}, which is calculated from expected, observed, and total heterozygosities averaged over loci.

Indirect measures and microsatellites. Most F_{ST} analogues underestimate differentiation when marker variation is high, such as when microsatellites are used

(Charlesworth 1998; Balloux et al. 2000). For example, because G_{ST} is calculated using heterozygosities, it is particularly sensitive to high levels of genetic variation (Hedrick 1999). As a result, G_{ST} estimates can be very close to zero when microsatellites are used, and a measure of G_{ST} standardized to total genetic diversity has been proposed to compensate for the effect of high diversity markers (Hedrick 2005a).

Another concern when using microsatellites for the computation of F-statistics is the IAM mutation model implicit in the calculation of F_{ST} as measured by Wright (1951) or Weir & Cockerham (1984). Slatkin (1995) developed R_{ST} specifically for microsatellites under the SMM, which is though to better represent their mutation. A measure adapting R_{ST} to ANOVA methods has also been developed (ρ_{ST}, Rousset 1996). Goodman (1997) suggested a method for standardizing R_{ST} with microsatellite data from differing sample sizes and differing variances among loci.

The fundamental difference between the two approaches is that R_{ST} takes into account allele size differences for the calculation of differentiation, while F_{ST} only considers the identity of alleles. A method has recently been developed for comparing F_{ST} and R_{ST} to determine whether allele size differences are informative to population differentiation (Hardy et al. 2003). The values of F_{ST} and R_{ST} are expected to be similar when mutation is not important, such as between populations that are recently diverged or are linked by high gene flow. However, F_{ST} has been shown to be higher than R_{ST} in weakly differentiated populations, possible due to the higher variance of R_{ST} (Balloux &

Goudet 2002). Conversely, R_{ST} may be significantly higher than F_{ST} when populations have not experienced recent contact, such as those separated by large geographic distances (Estoup et al. 1998). Using simulations, Balloux & Goudet (2002) found that R_{ST} performed better than F_{ST} for measuring differentiation only when migration rate and number of populations was small, even under a strict SMM. The standardized method of Goodman (1997) performed better at intermediate levels of gene flow, while F_{ST} was best at high levels of gene flow.

Departures from panmixia. When a population is continuously distributed and linked by continuous gene flow such that no subpopulation is differentiated from any others, that population is said to be in panmixia. Departures from panmixia may be achieved by testing for $F_{ST} = 0$, with a rejection zone defined by resampling techniques (Weir & Cockerham 1984). Various goodness-of-fit statistics can also be used to test for departures from panmixia, the most common of which are the χ^2 test, the *G*-test, and the Fisher exact probability test carried out using a Markov Chain (Raymond & Rousset 1995a). The probability test and *G*-test are more powerful than F_{ST} at detecting significant differentiation, especially when performing tests on allelic distribution rather than genotypic distribution (Goudet et al 1996). Waples & Gaggiotti (2006) showed the high power of the probability test of Raymond & Rousset (1995a) to detect the true number of populations when testing allelic distributions in simulated populations. Both F_{ST}-based and probability tests are more likely to detect departures from panmixia when genetic markers show high levels of diversity (Estoup & Angers 1998), but it is important to ask whether statistical significance necessarily indicates biological importance in these situations (Hedrick 1999).

Problems with indirect measures. Several authors have debated the continued use of F_{ST} and analogous indirect measures of differentiation because their limitations are now well known, particularly for estimating levels of migration and gene flow. Because these methods rely simply on current population gene frequencies, F_{ST} cannot distinguish between the effects of historical gene flow, contemporary gene flow, and drift (Pearse & Crandall 2004). The problem lies in the fact that F_{ST} estimates population size and migration only as the product of effective population size, N_e, and migration, m, and cannot tease apart the two (Fraser et al. 2007). For example, very low F_{ST} values cannot distinguish whether two populations have very recently diverged, have been long diverged but experience little drift due to high N_e, or have been long diverged populations and have recently exchanged gene flow (Bossart & Prowell 1998).

In addition, the models used to determine differentiation levels by *F*-statistics often contain unrealistic assumptions such as constant population size and equal rates of dispersal which can make them inaccurate for estimates of migration rates (Bossart & Prowell 1998; Whitlock & McCauley 1999). Violation of the standard model assumption of negligible mutation rates will also adversely affect estimates of migration (Rousset 1996). When mutation rates are high and migration rates are low, estimates of F_{ST} and

R_{ST} will be severely downwardly biased (Rousset 1996; Neigel 1997). This leads to an upward bias in migration rates, particularly when using microsatellites (Gaggiotti et al. 1999). Similarly, when the effect of drift is small, even low migration rates can severely reduce F_{ST} estimates (Neigel 2002).

It is clear that most populations likely experience a more complex combination of gene flow and genetic drift than is generally represented in models (Neigel 1997). Indirect measures of population subdivision should not be used as literal interpretations of demographic history, but they often offer a reliable first estimate of biological processes (Bohonak et al. 1999; Bossart & Prowell 1998). Also, because F_{ST} estimates have been obtained for thousands of species under thousands of conditions, they will remain a valuable tool for comparison among studies (Neigel 2002).

Phylogenetics

Phylogenetic analysis involves visually portraying the inferred evolutionary relationships between groups in a tree diagram (Edwards & Cavalli-Sforza 1967). The groups of interest, or Operative Taxonomic Units (OTUs), are traditionally species or higher taxonomic groups, but may be individuals, populations, or even genes (Avise 2004). Many methods of inferring evolutionary relationships and choosing the most representative tree have been developed in the preceding 40 years, some of which are discussed below, but all phylogenetic trees share some common features. They may be rooted, with a particular group presumed to be the common ancestor of all the others, and they may be scaled, with each branch length proportional to the amount of time or evolutionary change (Felsenstein 2004). The variance in each node of the tree is often assessed by bootstrapping (Felsenstein 1985), in which a hypothetical data set is created from the sample. This is often replicated several hundred times, with the proportion of replicates in which each node is observed being given as a measure of its strength.

Distance matrix methods. Distance matrix methods start with a matrix of pairwise genetic distances between groups, which represent their degree of genetic similarity. Genetic distances may be calculated by any of several distance measures then assembled into a tree by some algorithm, most often UPGMA or Neighbor-joining (Hedrick 2005b). Distance matrix methods are among the most commonly used for gene frequency data such as microsatellites.

UPGMA. Unweighted pair grouping with arithmetic mean (UPGMA; Sokal & Michener 1958) begins by clustering the groups with the smallest distance between them in the matrix, then averaging them to compute the distance to all remaining groups. This process proceeds stepwise until all groups have been added. The averaging may have the favorable effect of removing the stochastic variance inherent in distance calculations (Nei et al. 1983). Conversely, averaging across clusters means that each group contributes equally, and the UPGMA method is sensitive to incorrect grouping when trees are highly unbalanced (Hedrick 2005b). Another problem with UPGMA is its implicit assumption that rates of evolution are even among all branches, which may not be a reasonable assumption for distantly related groups (Felsenstein 2004).

Neighbor-joining. Another common distance matrix algorithm, neighbor-joining (NJ; Saitou & Nei 1987) is also based on clustering, and is probably the most common in use today (Avise 2004). Rather than adding groups one at a time, it begins with all groups on an unresolved star shaped tree, and chooses pairings and adjusts branch lengths at each step (Hedrick 2005b). Neighbor-joining performs well because it does not assume an equal rate of evolution among branches, but instead tries to minimize the discrepancy between the matrix distances and tree distances based on branch length (Avise 2004). In fact, NJ is guaranteed to recover the true tree if it is perfectly reflected by the distance matrix (Felsenstein 2004).

Fitch-Margoliash. The Fitch-Margoliash method (Fitch & Margoliash 1967), like NJ, attempts to minimize the difference between the distance matrix distances and the inferred tree distances. It accomplishes this by applying the statistical concept of least squares. Fitch-Margoliash is less commonly used today than UPGMA or NJ, but it has been shown to perform slightly better than NJ (Kuhner & Felsenstein 1994).

Distance measures. Tree construction algorithms for distance matrix methods are very reliable for finding the true tree if the distance matrix is correct (Felsenstein 2004). This depends, of course, on the distance measure that is used. There has been a vast literature on distance measures, with each measure attempting to better model certain evolutionary processes. Among the most commonly used distance measures are Nei's (1972) standard genetic distance, the chord-distance of Cavalli-Sforza & Edwards (1967), and the F_{ST}-based distance measure of Reynolds et al. (1983). Cavalli-Sforza & Edwards (1967) apply no evolutionary model, assuming divergence by genetic drift alone, while Nei (1972) and Reynolds et al. (1983) predict that distance arises linearly with time according to a certain rate of neutral mutation.

Microsatellites and distance measures. For microsatellites, the uncertainty inherent in their mutation becomes problematic when using these classical distance measures. Allele size constraints and varying rates of evolution between loci mean that genetic distance is unlikely to arise linearly between distantly related populations (Estoup & Angers 1998). In addition, the IAM and equal rates of neutral mutation assumed by Nei (1972) and Reynolds et al. (1983) may not be appropriate for microsatellites. These measures are also expected to perform poorly in recently diverged populations when no appreciable mutation has occurred (Felsenstein 1981b).

Goldstein et al. (1995) attempted to control for the problem of size constraints in microsatellites by creating a distance measure which takes into account allele size differences by assuming that most mutations are one step under the SMM. This model is hindered by size constraints to a lesser degree than classical distance measures, but cannot be correct because it still predicts both negative and infinitely large allele sizes across evolutionary time (Felsenstein 2004). In a simulation study, classical distance measures which assume no mutation, such as those of Cavalli-Sforza & Edwards (1967) and Nei & Takezaki (1983), have been shown to perform better than the method of Goldstein et al. (1995) when populations are recently diverged, due to the lower variation of these measures (Takezaki & Nei 1996). By contrast, the method of Goldstein et al. (1995) performed better between distantly related populations, when the assumption of linearity with time is no longer acceptable (Goldstein & Pollock 1997). However, an empirical study on microsatellite data from humans showed that the Goldstein et al. (1995) measure performed poorly, possibly due to varying rates of evolution between loci (Takezaki & Nei 2008).

A more complex model allowing for allele size constraints of microsatellites was suggested by Pollock et al. (1998), but the requirement that more loci be used and that each locus be analyzed individually makes its use impractical in most circumstances. Kruglyak et al. (1998) have suggested that size constraints may not need to be considered at all, suggesting instead a model balancing slippage mutations and point mutations. This model did well at predicting equilibrium allele sizes but had problems estimating across species (Calabrese et al. 2001). Microsatellites are likely inappropriate at the interspecific level altogether, because all current models overestimate observed allele size differences and homoplasy may be important at large evolutionary scales (Jarne & Lagoda 1996).

Despite these difficulties, distance matrix methods are generally very accurate when distance measures are used appropriately. Genetic distance measures summarize genetic diversity in a single manageable parameter and facilitate easy comparison of multiple populations. Applying a single distance measure loses all historical factors and higher order relationships between loci, but this actually discards far less information than might be expected and does not affect their performance (Felsenstein 2004). Hillis et al. (1992) tested UPGMA and NJ distance matrix algorithms as well as parsimony methods (see below) using a known phylogeny of T7 generated in the lab. They found that all methods recovered the correct tree, but differed in branch lengths. The discrepancy in branch length points is the one major disadvantage to distance matrix methods, which is their inability to utilize information on the varying rates of evolution at different loci (Felsenstein 2004).

Parsimony. The method of parsimony was first suggested by Edwards & Cavalli-Sforza (1963), who stated that the best phylogenetic tree had the fewest total evolutionary changes. Maximum parsimony methods group taxa by the minimum number of evolutionary steps between them. There are now many algorithms for computing parsimony, but each goes through each character in the data set counting the number of changes at each one (Hedrick 2005b). The most parsimonious tree or group of trees has the fewest total number of steps across all branches.

The biggest impediment to the use of parsimony is that it does not perform better as more information is added (Felsenstein 2004). Also, parsimony methods are not amenable to continuous data such as microsatellite allele frequencies. Due to the ambiguity in microsatellite evolution, they are not readily converted to discrete data to take advantage of character weighting based on frequency of change or number of possible states (Farris 1969). Unweighted parsimony will perform well when rates of evolution are low, but often the number of changes is too great to make this justification (Felsenstein 1981a).

Likelihood-based methods. Maximum likelihood is a powerful statistical approach which attempts to find a hypothesis which most likely explains the observed data. Likelihood functions are often very computationally intensive because of the large number parameters involved in evolutionary models, but likelihood-based approaches have received most of the focus in phylogenetics since modern computers have removed the barriers to their use (Marjoram & Tavaré 2006). Unlike parsimony methods, likelihood methods are guaranteed to produce better estimates when they are provided with more information and exhibit less variance about the true parameter value (Hedrick 2005b).

Allele frequencies and maximum likelihood. Maximum likelihood can be employed for the phylogenetic analysis of allele frequency data such as microsatellites. In fact, the first application of likelihood for phylogenies was on gene frequency data (Edwards & Cavalli-Sforza 1964). However, as with parsimony, mutation models cannot be applied to this type of continuous data. The classical Wright-Fisher model of genetic drift also proved intractable to likelihood methods and Edwards & Cavalli-Sforza (1964) applied a Brownian motion model to approximate the evolution of gene frequencies. Distance matrix methods may be preferable to maximum likelihood for use with allele frequencies, particularly when mutation is expected to have occurred. However, Wiens (2000) found that maximum likelihood under a Brownian motion model performed at least as well as the distance measures of Cavalli-Sforza & Edwards (1967) and Nei (1972), even in the presence of mutation. Nevertheless, these comparisons were made on data from allozyme markers, which experience comparatively much less mutation than microsatellites.

MCMC. Many likelihood methods are made possible by a class of computational methods known as Markov chain Monte Carlo (MCMC), which allows the study of complex probability distributions which cannot be easily studied by traditional methods (Beaumont & Rannala 2004). For models containing many interdependent parameters it is computationally impractical to explore the entire probability for all combined parameters. Instead, a Markov chain moves randomly through this distribution of parameters, accepting each new step only if it has higher likelihood than the step before (Excoffier & Heckel 2006). Ideally, after an initial "Burn-in" period of several thousand steps, the chain will settle on a space of highest likelihood and samples can then be taken only from this optimal region (Marjoram & Tavaré 2006). The success of MCMC methods depends on the starting point of the chain, the length of the burn-in period, and the way in which the chain moves from one step to the next (Excoffier & Heckel 2006).

Bayesian analysis. Bayesian analysis is a type of likelihood-based statistics related to maximum likelihood, but it differs in the use of a prior distribution of the quantity being inferred. A prior is essentially an expectation for a certain parameter or set of parameters put in place by the researcher based on additional knowledge. While maximum likelihood depends entirely on manipulating multiple parameters to maximize a likelihood function, Bayesian priors essentially constrain certain parameters and reduce the computational power needed (Beaumont & Rannala 2004). Bayesian analysis returns a posterior distribution, which is the distribution of the parameters after considering the data as well as the prior distribution input by the user. Bayesian analysis is very attractive because this posterior distribution depends not only on the likelihoods of the parameters under the data, but also on the prior odds placed on them beforehand (Felsenstein 2004). Bayesian approaches often rely on MCMC to find posterior probabilities, which take into account the prior distribution as well as the observed data (Marjoram & Tavaré 2006).

Phylogeography

Phylogeography is the study of the evolutionary relationships between populations taking into account spatial distributions as well as contemporary and historical forces. One lasting and important contribution of phylogeography has been the combination of the macro-evolutionary field of phylogenetics and micro-evolutionary field of population genetics (Avise 1998). The name and theoretical basis of the field lie in the pioneering work of Avise et al. (1987) on examining the history of related populations based on a tree built from DNA sequence data. Their work and nearly all phylogeographic analyses since have utilized mtDNA because its lack of recombination means it shows a clear phylogenetic record (Avise et al. 1987). Also, its smaller effective population size and higher mutation rate mean that it is more likely to show divergence between populations. So far population subdivision according to mtDNA genealogy structure has been shown in a huge range of organisms, and the number of organisms showing low variability is small (Avise 2000).

Gene trees and the coalescent. Phylogeography based on gene genealogy begins with the use of phylogenetic tools to create trees to describe patterns of gene ancestry, or gene trees. The initial gene trees of Avise et al. (1987) were created using a rather informal framework, and Templeton (1988) soon developed the more statistical method of nested clade analysis for mtDNA haplotypes. However, most of the current statistical framework for creating the structure of gene trees is based on the theory of the coalescent (Kingman 1982), which provides a genealogy of alleles at a particular locus. The basis of coalescent theory is that all present alleles have lineages which trace back through time and eventually "coalesce" at the most recent common ancestor (MRCA). This makes it possible to compute for any two lineages the time to MRCA as a function of evolutionary relationships, population size, and the number of gene copies in the population (Hey & Machado 2003).

Gene trees vs. population trees. Although coalescent theory is rarely disputed, it has become clear that gene trees from a single non-recombining gene as was pioneered by Avise et al. (1987) and Templeton (1988) could be very misleading for population-level analysis (Knowles & Madison 2002). The primary reason for the discrepancy between gene trees and population trees is stochastic variance, which leads to vastly different genealogical histories that can be experienced by different alleles, even within the same population (Hey & Machado 2003). This means that the coalescent tree based on any one gene may show a vastly different branching pattern from the true population tree such that evolutionary relationships between populations could be greatly oversimplified by single locus gene trees (Zhang & Hewitt 2003).

The stochastic variance in gene trees was known to early phylogeography researchers. Neigel & Avise (1986) recognized that genealogies could exhibit "polyphyly" or "paraphyly" in which populations contain some alleles which are more related between populations than within them. Although sampling more than one individual per population will improve the gene tree, it has long been recognized that the best improvement would be reached with the addition of more loci (Pamilo & Nei 1988).

Coalescent approaches for population parameters

Since each locus within an individual may have a different coalescent tree, and mtDNA transmits along a single genealogy, it becomes desirable to add additional gene trees by analyzing nuclear loci as well (Hare 2001). The emergence of likelihood-based methods made them the most promising tools for extending coalescent theory to a more general approach allowing multiple populations, multiple individuals per population, and multiple loci. This problem was made more difficult by the use of diploid markers such as microsatellites for which the gametic phase often was not known, and which undergo recombination. Recombination causes genes within the same individual to exhibit different genealogies depending on their degree of linkage (Rosenberg & Nordborg 2002).

The approach of Nielson (1998) was the first to come close to a general approach, allowing only two populations and one locus but multiple individuals per population. In the subsequent years, there has been an explosion of coalescent based methods which use likelihood and MCMC methods to consider multiple gene trees (Hey & Machado 2003). This has allowed for genealogical models with a large number of parameters. For example, the Nielson (1998) model was later extended and improved to allow multiple loci and migration between the two isolated populations (Nielson & Wakeley 2001; Hey & Nielson 2004; Hey & Nielson 2007), and can estimate probability distributions of six parameters at once.

Coalescent approaches employing MCMC have now been developed to estimate various population demographic parameters such as effective population size (e.g., Beerli & Felsenstein 2001; Beaumont 2003), migration rate (e.g., Bahlo & Griffiths 2000; Beerli & Felsenstein 2001), population growth or decline (e.g., Beaumont 1999) and the presence of natural selection (e.g., Beaumont & Balding 1999). Many of these methods are now available in user-friendly programs with sophisticated graphical interfaces, and rapid processing times. However, many models also contain explicit assumptions regarding Hardy-Weinberg equilibrium, population growth rates, migration model, et cetera, which should be understood and considered by the researcher before use (Excoffier & Heckel 2006).

Individual-level analysis – assignment and clustering

A similar and somewhat overlapping class of methods to the coalescent approaches focuses on individual level analysis by considering each sampled individual's genotype separately. Like the gene tree methods above, they also typically employ Bayesian or maximum-likelihood approaches with the use of MCMC (Excoffier & Heckel 2006). These methods are widely known as 'assignment methods' and can either assign individuals to populations defined by the user or to populations defined by the program as the optimal number of genetic 'clusters' (Pearse & Crandall 2004).

The class of methods which use predefined populations input by the user can be useful for the detection of immigrants and for determining the population of origin of each sampled individual (e.g., Wilson & Rannala 2003; Piry et al. 2004). Because these methods may be misleading if an individual's true population of origin is not among those sampled by the user, Cornuet et al. (1999) defined a method for determining whether a population can be excluded as the source of an individual. Studies have shown that assignment and exclusion methods can be very effective when used in conjunction (Primmer et al. 2000; Manel et al. 2002).

By contrast, clustering methods attempt to define the most likely number of groups without a priori information on sampled populations (e.g., Pritchard et al. 2000; Corander et al. 2004). This can be a much more challenging problem because population allele frequencies must be individually estimated, but clustering methods can be extremely useful when population boundaries are ambiguous or when cryptic structure may exist (Manel et al. 2005). Direct comparison of clustering solutions to sampled populations can also provide a measure of population differentiation and departures from panmixia. However, the accuracy of assignment methods when populations are not defined a priori depends heavily on sample size and number of loci, but especially on migration rates and marker variability (Waples & Gaggiotti 2006).

As with coalescent-based population demographic programs, individual level models often contain assumptions such as Hardy-Weinberg equilibrium, linkage equilibrium, and the sampling of all source populations (Excoffier & Heckel 2006). Use of the model without consideration of its assumptions and limitations may lead to misleading results.

Chapter II: Population Genetics of the German Cockroach, *Blattella germanica* L.

The German cockroach, *Blattella germanica* L.

The German cockroach, *Blattella germanica* L., is widely considered the most prominent of the household pest cockroaches, and is a ubiquitous pest of human-built structures in cosmopolitan regions on every continent (Appel 1995). It is particularly well-adapted to the indoor environment and is not known to exist in natural environments anywhere in its distribution (Roth 1985). Human structures provide the German cockroach with numerous harborages, a stable environment with regard to temperature and humidity, as well as access to plentiful food and water. The specificity of German cockroach adaptation to human structures means that this commensal can be difficult to control, and is an economically and medically important pest. In addition, their inability to survive beyond human structures means that they must invade new structures primarily by human transport.

Control of the German cockroach would be furthered by an understanding of the degree of both active movement and gene flow within buildings as well as the patterns of passive human-mediated transport across wider geographic ranges. However, much is still lacking in our knowledge of German cockroach movement at both of these spatial levels. The objective of the current study is to employ current population genetic tools to study population differentiation and gene flow of *B. germanica* in the urban environment, at spatial distances ranging from within a single apartment to the global scale.

Pest significance of *B. germanica*

The adaptation of *B. germanica* to human structures can mean that its population levels can become high very quickly and can be difficult to completely eradicate. Besides the pervasive distaste for the German cockroach, there are a number of direct and indirect health risks associated with its infestation. Pesticide exposure is the primary indirect threat, as widespread non-targeted use of insecticides without knowledge of cockroach population distributions increases the chance of human and pet contact (Koehler et al. 1995). In addition, frequent use of a single insecticide can lead to the development and spread of insecticide resistance (Cochran 1987).

Pathogen transmission. The greatest human health risks associated with *B. germanica* infestations are presented by the insects themselves. German cockroaches have been shown to be capable vectors of various pathogens. Studies have found that often greater than 90% of cockroaches caught in homes and hospitals are carrying bacteria known to be pathogenic to humans, many of them exhibiting multiple antibiotic resistance (e.g., Fotedar et al. 1991; Pai et al. 2005; Elgderi et al. 2006). As many as 30 species of bacteria have been isolated from *B. germanica* caught in low income urban apartments (Cloarec et al. 1992). German cockroaches have also been implicated at vectors of fungi (Fotedar et al. 1992) and protozoa (Graczyk et al. 2005; Pai et al. 2003).

Allergens. German cockroaches have also been revealed to play a large role in the production and dissemination of household allergens. *B. germanica* is a major source of indoor allergens – particularly in urban environments – producing at least seven known allergens (Gore & Schal 2007). There is a positive correlation between the level of cockroach allergen exposure in the home and allergenic sensitization (Eggleston et al. 1998), and asthmatic inner-city children who are sensitized and exposed to large levels of cockroach allergens experience significantly greater asthma morbidity (Rosenstreich et al. 1997). Cockroach allergens are more likely to occur at this sensitization level in urban areas and in high-rise or duplex and triplex buildings, especially in households where annual income is below $20,000 (Cohn et al. 2006).

Morgan et al. (2004) demonstrated that reduction of allergen levels in the home can reduce asthma morbidity. The authors employed an intensive approach which combined tenant education, cockroach control, and thorough cleaning. Arbes et al. (2004) showed that cockroach control alone, when guided by continued monitoring of population levels, was sufficient to significantly reduce cockroach allergen levels. A clearer understanding the degree of German cockroach movement within a building would be a significant step towards the control of cockroaches for the mitigation of household allergens and other potential health concerns.

Active dispersal of *B. germanica* within buildings

Possible factors in dispersal within apartments. Although German cockroaches undoubtedly have the ability to disperse within and between apartments in a building, there are few data on the extent of this dispersal. Many factors have been suggested to play a role in the active dispersal of *B. germanica*, with availability of food resources the most frequently cited (Schal & Hamilton 1990). The highest population levels within an infested apartment usually exist in the kitchen where food resources would be most readily available (Keil 1981; Owens & Bennett 1982), and a founding group of cockroaches may first establish in food preparation areas within apartments, then distribute more widely as the population grows (Brenner 1995). That hypothesis may be supported by the observation that population levels are also greater in more unsanitary apartments (Akers & Robinson 1981; Schal 1988) where food resources may be available throughout. Therefore, although all cockroaches within an apartment may represent a single human introduction, once cockroaches become established in several rooms they may not disperse throughout an apartment if food and water are available and foraging is not required.

A few studies have attempted to quantify German cockroach dispersal with regard to foraging and resource availability. When food is provided at varying distances, cockroaches most often utilize the closest food resource (Bret & Ross 1985; Rivault & Cloarec 1991; Durier & Rivault 2001). However, when no proximate resource is available, cockroach populations were found to feed as often on a single resource located in a separate compartment of an arena as those populations located nearer to the food source (Silverman 1986). These studies seem to suggest that cockroaches disperse no further than necessary, but may exploit a distant resource while retaining their initial harborage.

Other factors which may reduce cockroach exploratory behavior have been suggested by laboratory studies. Dispersal has been shown to decrease due to unfamiliarity with an area (Durier & Rivault 2003) and loyalty to a chosen refuge (Denzer et al. 1988), as well as to propensity of cockroaches to aggregate in the fewest shelters that will accommodate a population (Amé et al. 2006). These studies suggest a limited role of overcrowding in promoting dispersal. However, these studies were conducted in closed laboratory arenas without food or water, and the decrease in exploratory activity did not take into account the effect of resource availability on dispersal and harborage choice. Therefore, their relevance to apartment populations is unclear.

Dispersal between apartments. It is widely believed that cockroaches move frequently between apartments in multi-unit buildings, giving cockroaches refugia from treated apartments and allowing infestations to spread rapidly (Robinson *et al*. 1980). Several studies have attempted to quantify cockroach movement in a field environment. Rivault (1990) captured and marked cockroaches in several rooms of a public swimming facility. With weekly monitoring, only 15% were recaptured after 1 year, with 85% of those being trapped in the area of their original capture.

Owens & Bennett (1982) found that up to 31.8% of recaptured *B. germanica* had moved from another room within an apartment over a 9 day period. Over the same period, there was up to 30.9% movement of recaptured cockroaches between adjacent apartments in one building, but only a maximum of 9.4% movement between apartments in another building. The authors noted the presence of shared plumbing between the apartments with greatest movement. A subsequent study (Runstrom & Bennett 1984) found a significantly greater recapture rate after 3 weeks between adjacent apartments with shared plumbing than those without. A similar follow-up (Runstrom & Bennett 1990) found that 13% of recaptured *B. germanica* were immigrants from a neighboring apartment, with a greater percentage being from apartments sharing kitchen plumbing. While all of the above studies found that most recaptured cockroaches had not left their original capture area, they suggest that cockroach movement between apartments may occur with sufficient frequency to homogenize them.

Data concerning movement beyond adjacent apartments is lacking. Runstrom & Bennett (1984) found a maximum of 1.1% of recaptured individuals having moved through more than one apartment in a four-apartment building, in agreement with the previous finding of Akers & Robinson (1983) in a six-apartment building.

Limitations of existing studies. All of the above studies suffered from a low recapture rate and low survival rate. The difficulty lies in the mark-recapture or laboratory observation approaches used in these studies. Such ‘direct’ methods of inference into population structure and dispersal have several disadvantages. Recapture rates are typically very low and drop rapidly with distance from the point of release (Pocock et al. 2005), making studies beyond neighboring apartments impractical. In addition, because *B. germanica* cannot actively disperse between buildings, the study of movement beyond a single building is impossible using traditional mark-recapture methods. Even at scales where recapture is feasible, these methods allow no inferences to be made as to the actual amount of gene flow that is occurring (Slatkin 1987). This problem is complicated further by the impossibility of monitoring several possible dispersal locations at once with the hopes of recapturing marked individuals.

Direct observation studies usually also suffer from a very short period of observation. Direct observation based on sampling from a limited number of locations at a particular point in time cannot provide insight into events occurring before the period of observation (Slatkin 1987).

Clearly, more versatile methods are required for inferences on population interactions and relationships over large geographic and temporal scales. Ideal techniques would allow for inferences regarding historical and present relationships between populations based on groups of randomly sampled individuals. These techniques are provided by the application of molecular markers and 'indirect' measures based on allele frequencies within the expansive field of molecular ecology.

Passive dispersal of *B. germanica* beyond buildings

Few studies have applied molecular markers to the study of *B. germanica*. Hampson & Steiner (1982) found differentiation between two U.S. cities using electrophoresis of five enzyme loci. Cloarec et al. (1999) used eight enzyme loci in a study of 31 *B. germanica* populations from two cities in France. Despite a separation of approximately 900 km, no genetic divergence was found between the cities and a phylogenetic clustering method did not clearly differentiate them. These results may be explained in part by the extremely low genetic diversity present at the loci used. A single locus was found to contribute most of the structure present in the clustering tree, being monomorphic for one of two alleles in each population. A follow-up study (Jobet et al. 2000) was conducted on two *B. germanica* populations from each of the same two French cities using 10 randomly amplified polymorphic DNA (RAPD) markers. Significant genetic differentiation was found between the two cities, though it was much lower than the differentiation shown within and between populations in the same city. Two-dimensional canonical correspondence analysis explained a significant amount of the genetic variance, but did not separate the four populations according to city. This surprising lack of genetic structure between cities relative to within cities is contrary to the hypothesis that greater gene flow would occur more frequently between geographically proximate populations. Though low marker diversity may explain the results of Cloarec et al. (1999), no such diversity statistics were provided by Jobet et al. (2000). Nonetheless, RAPD markers have been shown to be unreliable and poorly reproducible (Schierwater & Ender 1993), calling into question the results of Jobet et al. (2000).

Mukha et al. (2007) applied restriction fragment length polymorphism (RFLP) analysis of the non-transcribed spacer (NTS) region within the ribosomal genes of *B. germanica*. High polymorphism was present within three populations sampled from three swine farms in North Carolina. Although all population pairs were significantly differentiated, greater genetic divergence was observed between the geographically proximate farms (~15 km) than between farms separated by greater distance (>100 km). The authors noted that the two geographically distant populations were owned by the same company and suggested that the genetic similarity between the roaches at these farms may be due to movement within a company's supply chain. Although the sample size was small, these results suggest an important role of passive dispersal that warrants further investigation.

Commensalism and dispersal

The unnatural ecosystem of urbanized areas tends to favor only those few species which are capable of inhabiting these disturbed environments. Some species are simply robust to urbanization and are able to utilize the fragmented natural resources of urban areas (McIntyre 2000). Others, such as the German cockroach, are specifically adapted to urban settings and are actually more abundant in association with humans. These synanthropic species are often invasive and are regularly distributed in multiple urban environments across huge geographic scales. The sharing of synanthropic species between numerous cities is likely explained by the two factors of habitat and transportation (McKinney 2006). The first factor, habitat, is especially important for species such as the German cockroach which are dependent on humans to provide food resources. Transportation is also clearly a factor in the global urban distribution of the German cockroach. Like many synanthropic species, it is frequently spread by human-mediated 'jump' dispersal (Suarez et al. 2001).

Synanthropic species may be further characterized by whether they spend most of their lives outside or inside of buildings (McKinney 2006). The German cockroach represents an extreme case of the latter category, spending its entire life indoors. Free living colonies are not known to occur, even in the epicenter of *Blattella* diversity (Roth 1985). Therefore, the natural environment represents an effective barrier to dispersal and gene flow and any movement beyond buildings requires human activity.

Other commensal systems. Few studies have investigated genetic differentiation in species which spend much of their life indoors. In one such study, microsatellite genetic diversity was lower and genetic differentiation higher in a commensal rat than in a free living congener (Brouat et al. 2007). While such results were consistent with expectations, an unexpected isolation by distance pattern was found in the commensal species. This pattern may have been the result of increased human-mediated transport between neighboring villages relative to more distant ones. In commensal house mouse, Dallas et al. (1998) found substantial microsatellite genetic differentiation between all pairs of populations, even at a distance of only 0.5 kilometers. No isolation by distance was observed at distances of 8–60 km, suggesting human-mediated dispersal is more important than active dispersal at this range and is likely haphazard in nature, rather than correlated with distance. Like the German cockroach and commensal rat species, house mice have become invasive on nearly every continent and several oceanic islands by human transport (Boursot *et al*., 1993). Both studies suggest that reduced active dispersal due to commensalism contributes to genetic differentiation even at short geographic distances. But while Brouat et al. (2007) found that passive dispersal promotes isolation by distance between towns in a commensal rat, Dallas et al. (1998) found no isolation by distance at this scale in the house mouse suggesting random human mediated dispersal in the latter species.

The above studies focused on rodent species whose movement is not exclusively limited to the indoors. In contrast, Schäfer et al. (2001) studied genetic differentiation in the cellar spider *Pholcus phalangioides*, at spatial scales ranging from within a single building to 917 km. No significant differentiation was found within buildings, suggesting that spiders may actively disperse between rooms. However, significant genetic differentiation was found between all pairs of samples both within and between cities. Isolation by distance was observed over the entire range area studied, but not within distances of 70 kilometers. Similar to the pattern seen in the house mouse (Dallas et al. 1998), the pattern seen in the cellar spider suggests that human mediated dispersal may occur haphazardly at distances within 70 km but plays little role at greater distances for human-associated spider species. However, Schäfer et al. (2001) also acknowledged the limitations presented by their use of small sample sizes and low diversity allozyme markers.

The present study

With the current study, our goal was to explore German cockroach population differentiation and dispersal in the urban environment at several spatial scales. We employed 10 highly polymorphic microsatellite loci recently identified for the German cockroach (Booth *et al*. 2007). The high diversity present in microsatellites makes them ideal for studies of population structure (Schlötterer 2004), and the use of microsatellites will allow for a higher resolution exploration of German cockroach population genetics than markers used in previous studies (Hampson & Steiner 1982; Cloarec et al. 1999; Jobet et al. 2000). This resolution allows for the study of gene flow mediated both by active dispersal within a building as well as by passive dispersal at larger geographic scales.

We expected levels of differentiation to be limited within apartments and within buildings because of the dispersal capacity of the German cockroach at these scales (e.g. Runstrom & Bennett 1984, 1990). However, significant genetic structure should exist beyond buildings because human-mediated dispersal is likely infrequent enough to allow populations to differentiate. Our hypothesis was that genetic structure is hierarchical, with genetic similarity greater within complexes than between them, greater within Raleigh that between cities, and greater within continents than between them. However, we expect genetic distance to correlate with geographic distance only between continents due to the reduced role of passive dispersal at these scales, but not within continents or within a city due to the haphazard nature of human-mediated transport.

Materials & Methods

Cockroach collection

In order to address the issue of gene flow and differentiation at several spatial scales, cockroach samples were collected in a geographical hierarchy. Samples at the scale of single apartments, apartment buildings, and apartment complexes within a city were collected from low-income apartments within Raleigh, North Carolina. All Raleigh apartments were within duplex, triplex, or low-rise buildings located in urban apartment complexes. This type of housing commonly experiences high levels of German cockroach infestations and high cockroach allergen levels (Gore & Schal 2007), and previous studies have revealed several apartment complexes in Raleigh with chronic infestations (e.g., Arbes et al. 2004).

Cockroaches were collected from apartment complexes within Raleigh from October to November 2006, May 2007, and December 2007 to January 2008. Infested homes were identified by a survey of residents and cockroaches were captured using unbaited sticky traps left for 24 or 48 hours. A total of 888 adult individuals from 18 apartments were selected for genetic analysis and stored at -20^{o}C in 95% ethanol.

To investigate gene flow and differentiation between rooms of an apartment, multiple traps were placed overnight in the kitchen, living room, bedroom, and bathroom of each apartment. From these traps, four samples of 30 adult *B. germanica* were selected for genetic analysis: two from the kitchen, one from the living room, and one from the bedroom or bathroom. Each sample was collected from a single trap whenever possible so that multiple traps were combined only if no trap within a room contained 30 adult individuals. Also, to test the hypothesis that aggregations within a kitchen may be differentiated, we attempted to sample from the two most spatially separated traps in the kitchen. A fifth sample of live individuals was collected near one of these kitchen traps using a modified vacuum. This sample allowed us to verify that trapped cockroaches were genetically representative of local aggregations. This multiple-room sampling was replicated in three apartment complexes (FR, DAC, and RC), and these apartments were labeled apartment 'A' for each apartment complex. These complexes were spatially separated as follows: FR to DAC = 2.5 km; FR to RC = 3.4 km; DAC to RC = 1.3 km.

To investigate differentiation between buildings in an apartment complex, two additional apartments from separate buildings were sampled from apartment complexes FR, DAC, and RC. Samples were taken from sticky traps placed in the kitchen of each apartment. These apartments were labeled 'B' and 'C'.

For inference on population structure between apartments in a single building, three apartments were sampled from each of the low-rise apartment complexes CW, CS, and CC. Samples of 30 individuals were taken from sticky traps placed in the kitchen of each apartment. Apartments were chosen such that two apartments were adjacent and shared a wall, and the third was located elsewhere in the building. Adjacent apartments in each building were labeled 'X' and 'Y', and the nonadjacent labeled 'Z'. These complexes were spatially separated as follows: CW to CS = 6.5 km; CW to CC = 1.6 km; CS to CC = 5.5 km.

Additional samples of *B. germanica* were provided by colleagues from another 18 cities in the United States as well as six international cities. The range of sampling dates and procedures differed, but each sample represented a single collection from a single apartment at a single point in time.

DNA extraction and genotyping

DNA was extracted from the legs and thorax of each German cockroach using the Gentra Puregene DNA extraction kit (Qiagen, Germantown, Maryland). Each cockroach was amplified for 10 microsatellite loci (*Bg*-1D5, *Bg*-A7, *Bg*-B12, *Bg*-CO4, *Bg*-D05, *Bg*-D9, *Bg*-F7, *Bg*-G7, *Bg*-wb1 A, *Bg*-wb2 A) according to the polymerase chain reaction (PCR) procedures outlined by Booth et al. (2006). Amplified products were labeled with M13 IRDye infrared dyes (Li-Cor Biosciences, Lincoln, Nebraska) and run on 6.5% polyacrylamide gels using a Li-Cor 4300 sequencer (Li-Cor Biosciences). Allele sizes were scored by comparison to IRDye sizing standard (Li-Cor Biosciences) or Microstep 20a sizing standard (The Gel company, San Francisco, California) using Gene Profiler software version 4.05 (Scanalytics, Inc., Rockville, Maryland).

Two individuals per population were also selected for mitochondrial DNA (mtDNA) sequence analysis. Extracted DNA was amplified for mtDNA 16S gene using primers LR-J-13017 and LR-N-13398 (Nagaraja et al. 2004) following the protocol of Xiong & Kocher (1991). Resulting products were purified using QIAquick PCR purification kit (Qiagen, Valencia, California) and labeled for sequencing reactions using BigDye Terminator cycle sequencing kit (Applied Biosystems, Foster City, California). Sequences were edited and forward and reverse sequences for each individual were assembled using the ContigExpress utility within Vector NTI Advance 10.3.0 (Invitrogen

Co., Carlsbad, California). Contigs were aligned using the AlignX component of Vector NTI Advance.

Genetic diversity and summary statistics

Mean number of alleles per locus, observed heterozygosity, and unbiased heterozygosity (Nei 1987) were calculated for each sample using the Microsatellite Toolkit software (Park 2001). Departures from Hardy-Weinberg equilibrium for each sample across all loci were tested in Genepop version 4.0 (Raymond & Rousset 1995b; Rousset 2008). A global test of genotypic linkage disequilibrium for each pair of loci across populations was also conducted in Genepop 4.0.

At each spatial scale, differentiation was assessed by global and pairwise F_{ST} values (Weir & Cockerham 1984) computed using Fstat version 2.9.3.2 (Goudet 1995). Confidence intervals for global F_{ST} values were computed by Fstat by bootstrapping over loci. Because microsatellite mutation is complex, mutation model may be important to estimates of population differentiation, especially when migration is expected to be minimal (Balloux & Lugon-Moulin 2002). Therefore, the stepwise mutation based R_{ST} (Slatkin 1995) may give different results under these circumstances. A test of significance between F_{ST} and R_{ST} values was conducted using the program SPAGeDi 1.2 (Hardy & Vekemens 2002). SPAGeDi implements a permutation of alleles to determine whether stepwise mutation significantly contributes to estimates of differentiation (Hardy et al. 2003).

Departures from panmixia between samples were determined by tests of pairwise genotypic differentiation implemented in Genepop. Genepop compares the genotype distributions between samples at each locus using a log-likelihood based *G*-test (Goudet et al. 1996), then obtains a *P*-value using Fisher's method for combining across loci. Departures from panmixia were used to group samples into populations such that samples were considered to be part of the same population when pairwise genotypic *G*-tests were insignificant. Following the method proposed by Waples & Gaggiotti (2006), populations ultimately consisted of all samples which could be connected by a chain of nonsignificant tests.

Isolation by distance was tested at the within city, between city and between continent scales by regression analysis of pairwise genetic and geographic distances using the Mantel test (Mantel 1967) implemented in Genepop. Geographic distances were Euclidian distances between sample cities. Arbitrary geographic distance values were applied for each spatial scale within apartment complexes for the isolation by distance test within Raleigh. Neighboring apartments (X and Y) were assigned a distance of 1 m, non-neighboring apartments within the same building (Z) were assigned a distance of 50 m, and apartments in separate buildings (A, B, C) were given distances of 100 m. Geographic distances were $\log_e$ transformed and genetic distances were transformed to $F_{ST}/(1-F_{ST})$, following Rousset (1997).

Phylogenetic trees

Phylogenetic trees were constructed from genetic distances computed in the program Powermarker (Liu & Muse 2005). Distances were calculated according to the chord distance (D_C) of Cavalli-Sforza & Edwards (1967) the standard genetic distance (D_S) of Nei (1972), and the coancestry coefficient (Θ_W) of Reynolds et al. (1983). Chord distance is computed employing no evolutionary model while standard genetic distance assumes a constant rate of neutral IAM mutation, and Reynolds' distance assumes linear divergence by drift. Trees were constructed by Powermarker using both the neighbor-joining (NJ; Saitou & Nei 1987) and UPGMA (Sokal & Michener 1958) algorithms with 1000 bootstrap replicates. Consensus trees were calculated using MEGA version 4.1 (Tamura et al. 2007; Kumar et al. 2008) and all nodes with less than 50% bootstrap support were collapsed.

Assignment methods

Inferred clustering. To detect underlying genetic structure at each spatial scale, the Bayesian clustering program *Structure* 2.2 was used (Pritchard et al. 2000). *Structure* estimates allele frequencies for a given number of inferred clusters, *K*, and simultaneously assigns fractions of each individual's genome to one or more clusters without a priori information about their sampling origin. This allows *Structure* to find the highest likelihood number of clusters that explain the genetic variation observed in the data. Preliminary runs at each spatial scale were performed with a burn-in of 100,000 replicates, followed by a sampling period of 100,000 steps. Runs found to be informative were rerun with a burn-in of 500,000 replicates, followed by a sampling period of 500,000 steps for each value of *K*. The Markov Chain was found to sufficiently converge for runs at this length. *Structure* analysis was done using the admixture model, which allows individuals to share ancestry in multiple populations, and the correlated alleles model, which assumes some correlation of alleles within populations.

Replicate runs of *Structure* were compiled using the program *CLUMPP* (Jakobsson & Rosenberg 2007). *CLUMPP* attempts to align the replicate runs by maximizing similarities between each individual's membership assignments across clusters. For each data set, 10 *Structure* runs for highest likelihood *K* were aligned using the *Greedy* algorithm in *CLUMPP* with 10,000 random input orders. *CLUMPP* outputs the optimal membership matrix averaged across all replicates. This matrix was graphically visualized using the software *Distruct* (Rosenberg 2004).

City-scale population assignment

Structure was also run using predefined population information to determine how well individuals sampled from within the 18 Raleigh apartment populations corresponded to the population from which they were sampled. The number of clusters, *K*, was set to the number of geographical samples, and *Structure* estimated likelihood for each individual belonging to its geographic location of origin as well as likelihood of being a migrant from each of the other sampled locations. An admixture model was again used, with runs of 750,000 steps preceded by a 750,000 step burn-in period. Individuals were considered to be members of their sampled population if they were assigned there with greater than 95% probability.

Structure assigns German cockroach individuals to populations under the assumption that all potential source populations are included in the data set. We know this is inaccurate for our *B. germanica* populations since we sampled only three apartments per complex. Therefore, the genetic exclusion method described in Cornuet et al. (1999) was also employed using the program *Geneclass* (Piry et al. 2004). This is a less powerful but more practical method than the *Structure* method when all source populations are not thought to have been sampled, and the combination of the two programs has been suggested when this is the case (Manel et al. 2002). The *Geneclass2* exclusion criteria for populations were based on the Bayesian method of Rannala & Mountain (1997), and probability computations were conducted using simulated populations of 100,000 individuals. Simulated populations were generated by resampling from the source populations according to the Monte Carlo procedure of Paetkau et al. (2004).

Geneclass was used to assign individuals to their sampled population by two separate criteria: (1) Individuals fitting their sampled population with greater than 95% probability, and (2) individuals not excluded from their sampled population by having greater than 5% probability, but being excluded from all other populations.

mtDNA statistical analysis

Aligned sequences for all global German cockroach populations were analyzed using *DnaSP* version 4.50.3 (Rozas et al. 2003) to calculate the number of haplotypes, haplotype diversity (H_d, Nei 1987), and average number of nucleotide differences between sequences (K, Tajima 1983). Haplotype networks for all populations were created using TCS (Clement et al. 2000). The maximum number of allowable steps between haplotypes within the network was set to a minimum parsimony probability of 95%.

Results

Within apartments

Diversity was high within each *B. germanica* sample from the three multi-sample apartments, DAC-A, RC-A, and FR-A, representing three apartment complexes, with observed heterozygosity ranging from 0.62 to 0.74 and mean number of alleles per locus from 6.67 to 9.11 (Table 1). However, only one of the 15 samples tested was within Hardy-Weinberg equilibrium at $\alpha = 0.05$. This may suggest that German cockroach populations are highly dynamic and could adversely affect later analyses which assume genetic equilibrium. Although the kitchen is often the most highly infested room in an apartment, no Hardy-Weinberg equilibrium or greater genetic diversity were observed in the kitchen versus other rooms within the same apartment.

Sticky traps intercept cockroaches during night-time foraging. To validate that sticky traps represent local aggregations, we vacuum-collected cockroaches in day-time from aggregations that were nearest to a single kitchen sticky trap in one apartment in each of three apartment complexes. Each pair of samples within an apartment was found to be part of the same panmictic population based on nonsignificant *G*-tests (Table 2a).

No departure from panmixia was found between German cockroach room samples by pairwise *G*-tests in any of the three apartments, with a string of nonsignificant *G*-tests connecting all five samples in each apartment (Table 2b). Similarly, global F_{ST} values were not significantly different from zero at any of the apartments as determined by overlap of 95% confidence intervals. Because of the lack of differentiation between samples, we determined that each of the apartments DAC-A, RC-A and FR-A represented a single German cockroach population. The individual apartment was chosen as the smallest unit of sampling for all analyses done at greater spatial scales and the K1 sample from each apartment was chosen to represent these apartments for all further analysis.

Within and between buildings within apartment complexes

In addition to the three apartments from the within-apartment analysis, there were 15 additional apartments sampled, for a total of 18 apartments (K1 samples only) representing six apartment complexes. Observed heterozygosity at these apartments ranged from 0.62 to 0.76 with mean number of alleles per locus from 7.11 to 9.11 (Table 3). Most apartment samples, with only two exceptions, exhibited significant departures from Hardy-Weinberg equilibrium.

Global F_{ST} values for German cockroach apartment samples in a single building were 0.017, 0.029 and 0.019 for apartment complexes CW, CC and CS, respectively. These were much higher F_{ST} values than within-apartment values (with no overlap in 95% CI), but generally lower than the F_{ST} values for apartments in different buildings within an apartment complex (Table 4). These results were consistent with expectations that populations within the same building might be within the possible active dispersal range of the German cockroach.

Pairwise *G*-tests were highly significant for all apartment pairs within the same building, even between neighboring apartments X and Y, with the exception of complex CS where X and Y were not significantly differentiated (Table 4). These results indicated that even for adjacent apartments, the individual apartment represents the largest sampling unit which may be considered a single German cockroach population. This conclusion, paired with our conclusion from the within-apartment analysis, led us to choose the individual apartment as the basic unit of sampling for spatial analysis at larger scales.

As expected, significant departure from panmixia between German cockroach populations in separate apartment buildings was found between all pairs of apartments (K1 samples only) at all three tested apartment complexes by pairwise *G*-tests ($P < 0.001$; Table 4). Global F_{ST} values for the three apartments at each complex were similar, with values from 0.027 to 0.030.

Within a city

Analyses at the city spatial scale were conducted using all of the 18 German cockroach populations analyzed at the building and complex scales. As expected, all *G*-tests between apartments from different apartment complexes were highly statistically significant (results not shown). Global F_{ST} across all Raleigh apartments remained low, with a value of 0.046 (Table 5). Pairwise F_{ST} values between Raleigh populations were also low, with most being below 0.1 (Table 6). Allele permutation tests conducted according to Hardy et al. (2003) on the 18 apartments within Raleigh found that the observed global R_{ST} was significantly greater than the permuted R_{ST} ($P = 0.047$; Table 5), suggesting that stepwise mutations do contribute to differentiation between Raleigh German cockroach populations. These results are surprising, since any differentiation at this spatial scale is expected to occur primarily due to genetic drift, with mutation playing a minor role.

A test of isolation by distance was conducted to determine if genetic distance was positively correlated with geographic distance within Raleigh, as would be expected if

German cockroach dispersal and gene flow occurred more often within apartment complexes than between them. The slope of the correlation between genetic distance and geographic distance was low, with a value of 0.0052, but a Mantel test showed this to be highly significant ($P < 0.001$; Table 7). The R^2 value for the linear regression analysis was 0.1528 (Fig. 1). These results indicate that F_{ST} values between apartment complexes increase with distance between complexes, at least up to the maximum of 9.1 km analyzed within Raleigh.

Phylogenetic trees were computed for the 18 German cockroach populations within Raleigh with 1000 bootstraps for three distance measures and two tree-building algorithms, for a total of six trees. Little genetic structure was identified within Raleigh by bootstrap values greater than 50% in any of the six trees beyond the scale of a single building. All clusters that were present in the UPGMA tree for a given distance measure were present in the neighbor-joining tree also, differing only in bootstrap values. The sole exception was a cluster consisting of FR-A and RC-B present in the UPGMA tree constructed from Reynolds et al. (1983) genetic distances (not shown). This cluster was not present in any of the other trees. By contrast, UPGMA trees for a given distance measure were frequently missing clusters present in the neighbor-joining tree. For this reason, only neighbor-joining trees are shown (Fig. 2). The three German cockroach populations from apartment complex CW were clustered together with greater than 50% bootstrap support in all six trees, as were the three populations from complex CC. A cluster consisting of the three apartments from complex CS was present in both the neighbor-joining (Fig. 2) and UPGMA trees constructed by Chord distance (Cavalli-Sforza & Edwards 1967). Only the *B. germanica* populations from CS-Y and CS-Z were clustered together in the neighbor-joining trees from Nei's (1972) standard genetic distance (Fig. 3) and Reynolds et al. (1983) genetic distance (Fig. 4), while these clusters were absent altogether from the UPGMA trees (not shown). The only additional cluster supported by greater than 50% bootstraps in multiple trees was a cluster consisting of DAC-C and FR-B (Fig. 2, 3, 4). Because allele permutations indicated that stepwise mutations contributed significantly to genetic structure within Raleigh, an additional tree was calculated using the distance measure of Goldstein et al. (1995). This tree displayed no further genetic structure (not shown).

Of the various potential number of clusters tested using *Structure*, $K = 4$ was found to be the highest likelihood. The best match between 10 replicate runs of $K = 4$ as permuted using CLUMPP (Jakobsson & Rosenberg 2007) is shown in Fig. 5, alongside single runs of $K = 2$–6. Only the three populations from CW form a distinct cluster at $K = 2$, with many individuals in all other apartments being split between the two clusters. Adding another cluster at $K = 3$ places apartment complex CC into another distinct cluster, and CS generally conforms to its own distinct cluster as well when $K = 4$. No apartment in apartment complexes DAC, RC, or FR groups distinctly into its own cluster, instead being made up of individual cockroaches divided between multiple clusters. The distinct clusters begin to break down at $K \geq 5$. Although pilot analysis was done on higher K values, no further genetic structure was revealed. The results of the *Structure* analysis are in accordance with the phylogenetic tree data in that no city-wide pattern of *B. germanica* genetic structure was detected beyond the single building apartments of apartment complexes CW, CC, and CS.

Agreement of individual genetic identity with sampling locations and detection of migrants was assessed using two assignment methods using the programs *Structure* (Pritchard et al. 2000) and *Geneclass2* (Piry et al. 2004). The number of *B. germanica* individuals being assigned to their sampled population by *Structure* with greater than 95% probability varied from 14 (CS-X) to 28 (CC-X) per apartment (Table 8). No apartment complex had consistently higher or lower assignment values, though CS had the lowest number of assignments and CC exhibited the greatest number of sampled population assignments. These results agree in part with the clustering results performed in *Structure* which showed the populations from CC forming the first distinct cluster at $K = 2$. Our *Structure* results show that many German cockroach individuals from the 18 Raleigh apartments share more genetic similarity with their sampled population than with any of the other populations, at least when all potential immigrant source populations are assumed.

The individual assignment/exclusion method implemented in *Geneclass2* does not assume that all potential source populations have been sampled and assesses probabilities for each individual for each population individually. *Geneclass2* assigned very few individuals to their sampled population with greater than 95% certainty, and also excluded very few sample populations with lower than 5% certainty. From zero (CW-Y, CW-Z, CS-Z) to four (RC-C) German cockroach individuals were assigned to their sampled population among the 18 Raleigh apartments (Table 8). Assignment of individuals to their sampled population by the exclusion of all other populations was also very low, with zero (DAC-A, DAC-B, DAC-C, RC-A, RC-C, CS-Z) to 10 (CC-Y) individuals assigned. Only two individuals were excluded from their sampled populations (DAC-B, and RC-A). These surprising results suggest that most German cockroaches are not genetically similar to their sampled populations, perhaps due to dynamic, non-equilibrium conditions or frequent immigration.

Between cities within the U.S.

An additional 14 urban German cockroach samples from the continental United States, as well as samples from Hawaii and Puerto Rico were added to the 18 Raleigh populations for the between-city analysis (Table 9). These samples contained from eight to 30 individuals and levels of genetic diversity were highly variable among the samples with expected heterozygosities from 0.49 to 0.75 and mean number of alleles per locus ranging from 3.44 to 9.11. Although a majority of populations were significantly out of Hardy-Weinberg equilibrium, many samples were not, in contrast to the Raleigh populations.

Global F_{ST} for all United States German cockroach populations was much lower than expected at such large geographic distance, with a value of only 0.071 (Table 5). Even after removing the Raleigh samples from consideration, which make up more than half of the total samples, the global F_{ST} value was only 0.099. This global F_{ST} value for the entire United States is exceeded by several of the pairwise F_{ST} values between apartment populations within Raleigh (Table 6). Allele permutation tests conducted according to Hardy et al. (2003) were not significant, suggesting that stepwise mutations did not contribute to genetic structure between all U.S. samples although these German cockroach populations that are separated by hundreds to thousands of kilometers and were expected to have exchanged little human-mediated gene flow and diverged sufficiently long ago for many mutations to have accumulated.

Although traditional models of isolation by distance as a measure of active dispersal are not expected to apply to the German cockroach, a Mantel test was performed to determine whether genetic distance and geographic distance were correlated between populations across the U.S. The Mantel test included all Raleigh populations but excluded the Hawaii and Puerto Rico populations, with geographic distances varying from zero within Raleigh apartment complexes to a maximum of 3789 km. Although the slope of the regression was very small at 0.0062 (Fig. 6), the Mantel test was significant, with $P = 0.029$ (Table 7). To determine the impact of the Raleigh populations on these results a separate Mantel test excluding the Raleigh populations was conducted as well.

This test showed no significant isolation by distance across the remaining 14 U.S. populations (P = 0.696; Fig. 6). Although one distance was 15.5 km, most were in excess of 500 km. As expected, passive dispersal of *B. germanica* appears to play little role in genetic structure at this geographic range and the correlation between genetic and geographic distances deteriorates.

As with city-level analysis, phylogenetic trees were uninformative, detecting no pattern of genetic structure within the U.S. samples. The only clusters consistently with greater than 50% bootstrap support in several of the six trees were the clusters corresponding to apartment complexes CC, CW, and CS. The three samples HNC, OCA, and HAW clustered together in all UPGMA trees (not shown). This was likely due to the sensitivity of the UPGMA method to levels of genetic diversity, since these samples had the lowest average number of alleles per locus and among the lowest observed heterozygosities (Table 9).

Trees were expected to cluster German cockroach populations according to close geographic proximity. No such cluster existed for the 18 Raleigh populations despite the significant isolation by distance detected when Raleigh populations were included. The proximate populations of LCA and CCA did cluster together in the neighbor-joining tree constructed from Chord distances (Cavalli-Sforza & Edwards 1967) (Fig. 7). These two populations were separated by a distance of only 15.5 km and were the most distantly separated from all other samples within the continental U.S. since they represented two of only three West coast samples. This represents the only evidence from these trees of a correlation between geographic and genetic distances. However, the third West Coast sample, OCA, did not cluster with LCA and CCA, the other samples from California.

Preliminary results from *Structure* analysis found that the program was unable to detect any underlying structure within the U.S. populations (results not shown). Consistent with the phylogenetic tree data, no cluster representing the Raleigh samples was displayed, and few populations fell clearly into distinct clusters. Instead, most individuals shared genetic identity in several clusters and no high likelihood K value was found. The failure of *Structure* to uncover underlying genetic structure is likely due to low levels of differentiation between U.S. populations and a lack of Hardy-Weinberg equilibrium within most samples.

Between cities: international populations

In addition to the 34 North American populations, six *B. germanica* populations were added from cities in Eastern Europe and Asia (Table 10). Genetic diversity in these populations was very low, with only Singapore having a level of heterozygosity and average alleles per locus comparable to the U.S. populations. In all other populations, observed heterozygosity and alleles per locus were as low as 0.43 and 3.0, respectively. Only Kiev, the international sample with the lowest average number of alleles per locus, was not significantly out of Hardy-Weinberg equilibrium. All other populations tested highly significant ($P < 0.001$) for departure from equilibrium.

The global F_{ST} for the international samples was 0.266, much higher than the level seen over all U.S. samples (Table 5). This large value may be due to the increased average distance between the international populations or because of the reduced genetic diversity of these populations. Across the geographic range, a negative correlation with a regression slope of -0.0785 was found between genetic distance and geographic distance (Fig. 8). This pattern is inconsistent with a pattern of isolation by distance, as confirmed by the nonsignificant Mantel test (Table 7). Most geographic distances between these Eurasian populations are in excess of 2000 km, well beyond the range at which passive dispersal of German cockroaches is expected to shape genetic structure. An allele permutation test suggests that mutation does contribute to genetic structure ($P = 0.023$; Table 5). Complementary to the isolation by distance test, low levels of dispersal and gene flow at the large geographic distances between the international samples allows mutations to accumulate within populations.

No underlying genetic structure within the Eurasian German cockroach populations was uncovered by running *Structure*. The highest likelihood number of clusters was $K = 6$, with each population adhering closely to its own cluster (not shown). The large genetic differentiation between international populations most likely contributed to this result.

Global

Global analyses consisted of all 40 U.S. and international populations. The addition of the international samples to the total U.S. samples only raised the global F_{ST} value to 0.106, nearly equal to the U.S. samples with Raleigh excluded (Table 5). Removing Raleigh from the global samples elevated F_{ST} to 0.158. This value is still lower than expected given that many of the geographic distances between U.S. and international populations were in excess of 10,000 km.

The Mantel test performed across all global German cockroach populations was highly significant ($P < 0.001$; Table 7), although the slope of the regression was only 0.0193. A separate Mantel test was performed excluding the 18 Raleigh samples. This test also showed a significant isolation by distance ($P = 0.005$). Although there was much variance in both data sets, the slope of the regression with Raleigh excluded was higher (0.0439; Fig. 9) and the significant Mantel test indicated that there was greater genetic similarity between *B. germanica* populations within continents.

An allele permutation test conducted across all global samples was significant ($P = 0.039$; Table 5), indicating that mutation did play a role in genetic structure as was expected at this scale. However, when Raleigh was excluded, the allele permutation test was not significant. This result is in contrast to the expectation that populations on separate continents would have acquired different mutations which would increase in the populations due to the action of genetic drift.

Phylogenetic trees constructed from the microsatellite data for the 40 global populations were unsuccessful at determining relationships between German cockroach populations. Trees constructed from Reynolds et al. (1983) genetic distances and Nei's (1972) standard genetic distance were not informative, so only the neighbor-joining and UPGMA trees constructed from Chord distance (Cavalli-Sforza & Edwards 1967) are displayed (Fig. 10). These trees are nearly identical to the trees constructed from U.S. populations with the addition of a separate cluster consisting of the two Russian and two Ukrainian samples. Consistent with our hypothesis, these samples formed a group on the tree indicating their divergence from the U.S. populations, although bootstrap support was only 50% in the neighbor-joining tree. The addition to this cluster of the Hickory (NC), Orange (CA), Hawaii, and Tehran populations in the UPGMA tree again suggests that this method is sensitive to varying levels of genetic diversity (Felsenstein 2004) since these populations all had low mean numbers of alleles.

Similarly, preliminary runs of *Structure* placed only the lowest diversity populations from Hickory (NC), Orange (CA), Moscow, Tomsk, Kiev, Crimea, and Singapore into distinct clusters (not shown). The exceptions were the populations from apartment complex CC which formed a discrete cluster, and the population from Tehran which also fell cleanly into its own cluster.

Of the 66 successfully sequenced 16S mtDNA individuals, only 10 unique haplotypes were found (Table 11). Haplotype diversity, H_d (Nei 1987), across all populations was high, with a value of H_d = 0.737. However, the average number of nucleotide differences, *K* (Tajima 1983), between sequences was only 2.63 (0.68% sequence divergence). A parsimony haplotype network computed in TCS (Clement et al. 2000) found a maximum of eight connection steps within a probability limit of 95% (P = 0.954), with Haplotype 1 not connecting to the others within this limit (Fig. 11). Haplotypes varied greatly in their frequency and geographic ranges, and no geographic pattern of *B. germanica* mitochondrial diversity was detectable with this network.

Discussion

The current study sought to add new understanding of the dispersal and population structure of the German cockroach across a wide range of geographic scales. At the two extremes of spatial scale the German cockroach is alternately a broadly distributed and highly restricted species. Although on a global scale it has been transported to nearly every human populated region (Appel 1995), it is restricted at the local scale by its inability to survive and disperse outside of human-made structures. Understanding the extent to which cockroaches move between residences and between buildings is integral to controlling infestations and preventing their spread (Schal & Hamilton 1990). Beyond a single building, the strict commensal relationship between the German cockroach and humans makes *B. germanica* a unique and ideal model for studying the extent to which human-mediated dispersal occurs and its role in shaping population structure—from buildings separated by a few meters to cities separated by several thousand km. Our goal was both to extend previous studies at smaller spatial scales with more modern molecular markers and techniques, and larger sample sizes, as well as to provide insight into *B. germanica* population structure at larger scales never before studied. Some of our analyses resulted in paradoxical results, which served to highlight the relative strengths and weaknesses of the tools and techniques we utilized, as well as the unusual dynamic nature of German cockroach populations. Despite these constraints, however, some important conclusions may be drawn on the movement and population history of *B. germanica*.

Genetic diversity and marker analysis

The 10 microsatellite markers we used were highly diverse across most of the 40 *B. germanica* populations. The numbers of observed alleles for many of the markers were generally lower than those reported by Booth et al. (2006) for 12 German cockroach populations. These authors reported an average of 11.9 ± 1.95 (SEM) alleles per microsatellite locus (range 4 to 22), whereas we found 8.1 ± 0.14 alleles per locus in 18 Raleigh apartments, 6.9 ± 0.42 in 16 North American populations (including Hawaii and Puerto Rico), and only 4.9 ± 0.78 alleles per locus in six Eurasian populations. Observed heterozygosity was similarly high in most populations, although often much lower than expected heterozygosity. The overall amount of genetic diversity in the German cockroach populations was surprisingly high given the population size fluctuations expected to occur. German cockroaches are highly sensitive to changes in ambient temperature and humidity, as well as levels of food and water availability (Schal & Hamilton 1990). These environmental factors, as well as frequent insecticide treatments and tenant turnover, would be expected to disrupt German cockroach populations and impose frequent bottlenecks.

Despite the high diversity, nearly all populations were significantly out of Hardy-Weinberg equilibrium due to heterozygote deficiencies. No loci were significantly out of Hardy-Weinberg equilibrium in all populations, but locus *Bg*-D9 (repeat motif $[GAA]^{20}$) frequently tested significant. Departure from Hardy-Weinberg equilibrium may be due to a number of factors. Non-random mating and genetic drift due to small population size, as might be expected in isolated aggregations of the German cockroach, would be expected to have the same average effect on all loci (Beaumont & Balding 2004), a state not seen in any of our *B. germanica* populations. For example, inbreeding can be assessed by F_{IS} values, which were not large but were often significantly greater than zero at all spatial scales beyond a single apartment (Table 2a, 4, 5). However, variance was often high for these estimates, and they were strongly influenced by large values for locus *Bg*-D9 rather than being consistent across loci.

Departures from Hardy-Weinberg equilibrium may also be caused at particular loci by the presence of selection (Slatkin 1987). While microsatellites are assumed to be selectively neutral (Estoup & Angers 1998), loci may be affected by selection at linked loci (Marjoram & Tavaré 2006). German cockroaches are frequently subjected to great selective pressure when exposed to insecticides (Cochran 1987) and other household chemicals.

Null alleles would also be expected to cause an apparent heterozygote deficiency and subsequent departure from Hardy-Weinberg equilibrium (Chapuis & Estoup 2006). Locus *Bg*-wb1a (repeat motif $[GAA]^{10}$) was removed from analysis due to the probability of null alleles. A null frequency test performed in Genepop (Raymond & Rousset 1995) according to the EM algorithm (Dempster et al. 1977) found that only locus *Bg*-D9 frequently had estimates of null allele frequencies in excess of 10% (data not shown).

Of the 36 pairs of loci, very few were found to be in significant departure from linkage equilibrium (Table 1,3,9,10), and no pair was consistently found to be significantly out of linkage equilibrium in a majority of populations, so this is not expected to adversely affect our analyses.

Differentiation within a building

Most early studies of *B. germanica* movement focused on movement within a building and utilized mark-recapture techniques (e.g., Akers & Robinson 1983; Runstrom & Bennett 1984, 1990; Rivault 1990). These studies suffered from low recapture rates and generally lacked the ability to estimate historical movement. Our study used molecular markers which overcome these constraints. Microsatellites have become one the most widely used markers for the study of fine scale population structure (Jarne & Lagoda 1996). The high amounts of diversity observed in our microsatellites makes them ideal for analysis of *B. germanica* dispersal at spatial scales as small as individual aggregations within apartments.

We validated our use of sticky trap sampling of cockroaches by showing that cockroaches trapped on a single sticky trap in the kitchen were not genetically differentiated from cockroaches vacuum sampled from the nearest aggregation (Table 2a). As hypothesized, no genetic differentiation also was observed within apartment kitchens and between rooms within any of the three apartments in three apartment complexes for which we obtained multiple sampled, DAC-A, RC-A, or FR-A (Table 2b). This genetic homogenization may either be due to frequent movement and gene flow between rooms or to very recent colonization of new rooms by cockroaches from a single introduction.

These results are consistent with mark-recapture studies which showed that German cockroaches migrate freely between rooms (Owens & Bennett 1982; Rivault 1990). Our results also support the hypothesis that cockroach populations likely first become established in the kitchen and subsequently spread to other rooms of the residence as the population becomes larger (Brenner 1995). Although the kitchen was the most highly infested room of our sample apartments, there was movement to the living room and bed room, as evidenced by genetic panmixia within our *B. germanica* apartment populations. Although previous studies suggested that cockroaches may disperse only the minimum distance necessary, because they are highly gregarious (Amé et al. 2006), exhibit refuge loyalty (Denzer et al. 1988), utilize the most closely available food source (Bret & Ross 1985; Silverman 1986; Rivault & Cloarec 1991; Durier & Rivault 2001), and tend to avoid unfamiliar areas (Durier & Rivault 2003), these traits did not promote isolation and differentiation between rooms within apartments, probably because food was readily available in these rooms.

These results have important implications to cockroach and allergen mitigation efforts. The spatial coverage of most pest control treatments generally target only the kitchen and bathroom, and only recent interventions with baits have treated the whole home. The latter have resulted in unexpectedly efficacious reductions in both cockroaches and allergen concentrations throughout the home (Arbes et al. 2003, 2004; Sever et al 2007). Because of extensive dispersal and gene flow between rooms within a residence, it is not surprising that pest control limited only to the kitchen would leave untreated cockroach aggregations in other rooms, which subsequently can disperse back to the kitchen. Indeed, our preliminary comparison of whole apartment pest control with treatments of the kitchen and bathroom only, have shown large and highly significant differences in suppression of cockroaches and allergen reductions in the living rooms and bedrooms. Thus, lack of genetic differentiation between rooms within apartments suggests extensive active dispersal of cockroaches between rooms, and highlights the importance of comprehensive treatment of the whole home to suppress cockroaches.

Dispersal between apartments was found to be rare in previous mark-recapture studies (Akers & Robinson 1983; Runstrom & Bennett 1984), but these studies were limited in their observations of dispersal during the study period. Our use of genetic markers allowed us to make inferences about the recent history of the sampled cockroach populations. The significant differentiation observed between two out of three pairs of adjacent apartments and six out of six pairs of non-adjacent apartments within the same building (Table 4) suggests that movement between non-neighboring apartments within a building is infrequent. Movement between adjacent apartments (e.g., CS-X and CS-Y) most certainly occurs with more frequency than between apartments separated by greater distance within the building, but gene flow may not be prevalent enough to genetically link those populations in many cases. Although some features of home construction, such as shared plumbing, may play a role in the extent of dispersal between apartments (Runstrom & Bennett 1984, 1990; Schal & Hamilton 1990), no differences were observed in construction features between apartment complex CS and complexes CW and CC.

Differentiation beyond buildings: within and between apartment complexes

Movement of *B. germanica* between buildings must be passive dispersal mediated by humans because *B. germanica* is unable to survive out of doors. For this reason, movement between buildings within a complex is expected to occur less frequently than movement within a building. Our results showed that overall differentiation, as expressed by global F_{ST} values, was higher in the multiple-building complexes of DAC, RC, and FR than in the single-building complexes of CW, CC, and CS (Table 4). We also expected that dispersal would occur more frequently within complexes than between complexes due to interactions between neighbors and the movement of property within complexes. However, none of the six phylogenetic trees constructed from three distance measures was able to group the separate buildings of DAC, RC, or FR onto single branches with great support (Fig. 2, 3, 4). Only German cockroach populations from CW, CC, and CS consistently clustered on a single branch with greater than 50% bootstrap support. Even then, the 'Z' apartment consistently clustered on a branch with one of the other two non-adjacent apartments, rather than the two neighboring apartments clustering together, as expected. This was the case even in complex CS, where neighboring apartments did not significantly differ from panmixia. Because stepwise mutations were found to significantly contribute to genetic structure within Raleigh, an additional tree was constructed using the $(\delta\mu)^2$ distance (Goldstein et al. 1997). This measure failed to detect any further structure between Raleigh populations, and it may also be less accurate than other measures among weakly differentiated populations due to its higher variation (Takezaki & Nei 1996).

Lack of tree resolution is likely not a weakness of the tree building algorithm, as both UPGMA and neighbor-joining trees produce accurate trees if distance measures are representative of the true tree. In particular, neighbor-joining is guaranteed to recover the true tree if it is perfectly reflected by the distance matrix (Felsenstein 2004). In the case of our study the problem likely lies in a weakness of ‘indirect’ genetic distance measures in general. The difficulty in distinguishing clear relationships between German cockroach populations in the separate buildings of complexes DAC, RC, and FR lies in the fact that indirect genetic distances are calculated from current allele frequencies, which means that historical information is lost (Pearse & Crandall 2004). These distance measures are unable to tease apart the relative effects of gene flow and drift (Fraser et al. 2007).

Overall diversity across Raleigh was very low in our study, with a global F_{ST} of 0.047 and most pairwise F_{ST} values lying below 0.1. This low level of diversity in the current study and lack of higher differentiation between apartment complexes relative to within them may be because German cockroach populations experience frequent passive, human-mediated dispersal and gene flow both within and between apartment complexes. Alternatively, this lack of a differentiation-by-distance may be because all Raleigh populations have similar allele frequencies due to shared ancestry and have diverged randomly due to drift. It is clear in either case that human-mediated dispersal of *B. germanica* occurs no more frequently between buildings within an apartment complex than between apartment complexes. Similarly, the strong clustering of the single building apartments of CW, CC, and CS suggests that German cockroaches are able to disperse throughout a building and do so more frequently than they are moved between buildings.

Analysis by the program *Structure* produced similar results. The highest likelihood number of clusters was 4, with only the apartments of CW, CC, and CS forming their own distinct clusters (Fig. 5). Running *Structure* under the correlated allele frequency model causes *Structure* to assume that allele frequencies are correlated within populations (Falush et al. 2003). The *Structure* results suggest that the greatest genetic similarity occurs between cockroach populations within a single building. This conclusion is supported by the lower global F_{ST} values for German cockroach populations within apartment complexes CW, and CS (but not CC) relative to DAC, RC, and FR.

It is unclear, then, why all individual buildings did not fall into distinct clusters. Since the apartments of complexes DAC, RC, and FR are from individual buildings, the total number of buildings sampled within Raleigh was 12. Yet *Structure* results for K = 12 were uninformative, with most individuals sharing membership in a large number of clusters (data not shown). If high levels of gene flow between the apartments of DAC, RC, and FR have caused *Structure* to be unable to clearly differentiate them, then the same should be true of apartments CW, CC and CS.

The discrepancy in the *Structure* results is likely due to abnormalities in the data. *Structure* assigns individuals to clusters according to allele frequencies assuming Hardy-

Weinberg equilibrium within clusters (Pritchard et al. 2000). Running *Structure* on populations out of Hardy-Weinberg equilibrium may lead to misleading and inconclusive results (Excoffier & Heckel 2006). However, German cockroach populations were significantly out of Hardy-Weinberg equilibrium in all apartment complexes, including CW, CC, and CS.

The fact that *Structure* did not assign *B. germanica* individuals from complexes DAC, RC, and FR to clusters corresponding to their individual building may mean that individual allele frequencies and population allele frequencies are not highly correlated. This hypothesis was tested by use of assignment methods. *Structure* was run using population information model of *Structure*, which allows the program to simultaneously assign allele frequencies to each of the 18 populations and define assignment probabilities of each individual to each population. Over all 18 populations, *Structure* assigned only 70.8% of individuals to the population from which they had been sampled with greater than 95% probability (Table 8). This assignment rate is lower than expected since the population information model of *Structure* initially assumes that individuals belong to their sampled population (Pritchard et al. 2000). Individuals must be largely different from their sampled populations to be assigned there with low certainty. One problem in using *Structure* with population information to test the fit of *B. germanica* individuals to their respective populations is that *Structure* assumes that individuals either belong to their sampled population or are immigrants from one of the other populations.

Probabilities for the 18 populations sum to 100% for each individual. *Structure* chooses the most likely population of origin, assuming that all populations which may have possibly contributed immigrants have been sampled. Since we know that there are many German cockroach populations within Raleigh which may have contributed immigrants to our 18 apartments, *Structure* results cannot be assumed to be entirely accurate (Manel et al. 2002). The true number of individuals belonging to their sampled population may actually be lower than the 70.8% predicted by *Structure*.

Geneclass was used to create 18 simulated populations of 10,000 individuals from the genotypes of the 18 German cockroach populations. This method is useful for our study because it does not assume that all possible source populations have been sampled, instead calculating probabilities for each population individually. Over all 18 populations, only 6.3% of individuals were assigned to their sampled population with over 95% probability. Individuals having less than 5% probability in all populations except their sampled population was only 7.3%. *Geneclass* has been shown to be less powerful than *Structure* (Manel et al. 2002); however, its ability to assign or exclude individuals from each population independently makes it important for our analysis. Populations of *B. germanica* were not recognized as representative of sampled individuals by *Geneclass*, implying that populations may be made up of many recent immigrants.

Structure has also been shown by Waples & Gaggiotti (2006) to perform poorly at estimating the true number of populations when gene flow is high or differentiation is low, particularly when the number of markers was below 20 and number of individuals sampled was below 50, as they were in our study. The authors found that the most powerful method for estimating the true number of populations was the pairwise *G*-test. Therefore, the most reliable answer may be the 17 differentiated populations identified by the pairwise *G*-tests, with only apartments CS-X and CS-Y not being differentiated. These results suggest that human-mediated dispersal of German cockroaches between buildings within a complex or between apartment complexes within a city occurs infrequently enough to be considered negligible on the whole. Passive dispersal of German cockroaches may occur as random, isolated events rather than as frequent incidences. However, German cockroach populations may have experienced frequent mixing before recent history or may have originated from a single introduction, evidenced by low F_{ST} values and poor assignment performance within Raleigh populations.

Differentiation between cities

The structure of *B. germanica* genetic variation at great distances was studied between an additional 16 cities in the United States, and also between six international cities within Eurasia. At these distances we expected differentiation to be high, since gene flow between cities would be limited to random infrequent jumps or to slow sequential movements across continents. Either type of movement might be mediated by the movement of property or products, such as by moving trucks, shipping trucks, or cargo ships or planes. Overall differentiation across the 16 cities within the United States was low as indicated by a global F_{ST} of 0.099 (Table 5). This value is lower than some of the pairwise F_{ST} values observed within Raleigh (Table 6). By contrast, global F_{ST} was much higher across the international German cockroach populations, at 0.266. This discrepancy may be due to the greater average distance between the sampled Eurasian populations relative to the sampled cities within the United States. It might also be explained by a greater amount of movement of people, property, and freight across the United States relative to the movement among our sampled Eurasian cities. We expected both long distance jumps and short sequential movements of German cockroaches to be frequent. However gene flow between any two given cities should be low in the recent history detectable by microsatellite allele frequencies, relative to movement within a city.

This hypothesis was not supported by phylogenetic trees constructed for the German cockroach populations from across the United States. As with the Raleigh populations, there was a general lack of structure within the trees. Only the apartments of CW, CC, and CS were consistently clustered on the same branch with greater than 50% bootstrap support (Fig. 7). The two California samples, LCA and CCA, also clustered together on the neighbor-joining tree constructed from Chord distance (Cavalli-Sforza & Edwards 1967); however, the other California sample, OCA, did not cluster with them. The 18 populations from Raleigh did not group together on a single branch as expected.

These trees did not allow us to infer that dispersal of German cockroaches occurs more frequently within a city than among cities.

Analysis of the United States populations with *Structure* (not shown) also gave similar results to the Raleigh analysis. The apartments of CW, CC, and CS were consistently placed in their own clusters, as were some of the least genetically diverse populations (HNC, OCA and HAW). This is likely due to a greater correlation of allele frequencies when the average number of alleles per locus is low.

As with the Raleigh populations, inability to separate gene flow from drift with indirect distance measures makes inferences about German cockroach dispersal across the United States difficult. The low differentiation and lack of local structure could suggest that dispersal and gene flow between German cockroach populations occurs often, even at distances of several hundred km. It is difficult to imagine, however, that German cockroaches experience gene flow just as frequently between cities separated by several hundred km as between populations within the same city. The results are similar to those shown by Jobet et al. (2000), who were unable to clearly distinguish German cockroach populations according to city, despite the presence of genetic differentiation between them. The authors also found no greater differentiation between cities relative to within cites, as we observed. Cloarec et al. (1999) collected many samples of German cockroaches in each of two cities and also found that a phylogenetic tree did not cluster populations according to city.

At the great range of geographic and temporal distances separating the German cockroach populations across the United States, inaccuracies with indirect distance measures and tree building algorithms could be expected to play a role in the lack of tree structure. The most likely problem, however, is a general lack of differentiation across *B. germanica* populations from throughout the United States. The low microsatellite differentiation between geographically distant populations agrees with the previously published lack of both allozyme diversity (Cloarec et al. 1999), and RAPD diversity (Jobet et al. 2000).

Differentiation across the* B. germanica *species range

The total populations from the United States and Eurasia were combined for analysis of total global differentiation of *B. germanica* populations. Overall diversity was higher for the combined data set, but still returned a relatively low F_{ST} value of 0.158 with Raleigh populations excluded (Table 5). Phylogenetic trees were created to explore the relationships between all 40 populations. Lower levels of German cockroach dispersal was expected to occur between continents than within them, a relationship likely to be reflected in tree structure. However, the addition of the Eurasian samples to the United States samples did little to resolve historical relationships between German cockroach populations. The four populations from Russia and Ukraine did cluster together on the two trees constructed from Chord distance (Fig. 10), as well as the UPGMA tree constructed from Reynolds' et al. (1983) distance measure (not shown). However, in the

UPGMA trees, the populations of HNC, OCA, and HAW were added to the Russian and Ukrainian samples on a single branch. Also, the Asian sample, SIN, was placed within the United States samples in the Chord distance neighbor-joining tree, while TIR was placed with the Russian and Ukrainian populations in the UPGMA trees.

These trees, constructed from the total global populations, may most starkly illustrate the weaknesses of distance matrix tree construction. The tree construction algorithm of UPGMA contains implicit assumptions which are likely violated by our data. UPGMA trees assume a molecular clock, meaning that the rate of evolution is assumed to be the same among all branches (Felsenstein 2004). This is not expected to be true when a wide range of long-diverged populations are investigated. The populations MRU, TRU, KUA, CUA, TIR, HNC, OCA, and HAW are clustered together on the UPGMA trees constructed from global populations. Since these populations are also the least diverse populations, they are placed on the same branch under an assumption of equal evolution. Neighbor-joining trees do not assume an equal rate of evolution and attempt to adjust the fit of the tree after the addition of each cluster (Avise 2004). This is a more accurate method than UPGMA, but only if the correct tree is reflected in the distance matrix (Felsenstein 2004).

However, many distance measures have implicit assumptions and evolutionary models as well, which may affect their performance. The Chord distance of Cavalli-Sforza & Edwards (1967) assumes no evolutionary model other than divergence solely by genetic drift. This distance measure seems to have been the most informative with regards to *B. germanica* structure, although it often did not cluster populations as expected according to geographic proximity. By contrast, the standard genetic distance of Nei (1972) contains a number of assumptions, including mutation by the infinite allele model, mutation-drift equilibrium, linear mutation rates, and constant population size equal across populations. The distance of Reynolds et al. (1983) similarly assumes a constant linear rate of divergence by drift. This assumption, as well as the assumptions of constant population size and mutation-drift equilibrium, is especially unlikely for *B. germanica*, which is likely to endure periodic stressors and reduction of population size. These periodic population disturbances may also lead to changes in allele frequency and distance measurements, violating the assumption of linear evolution in the affected populations.

In general, distance measures which attempt to model mutation, such as the model of Nei (1972), have been shown to be less accurate than those which assume no evolutionary model, such as Cavalli-Sforza & Edwards' (1967) chord distance (Felsenstein 1981b). This is highlighted by the lack of clustering of even CW, CC, and CS apartments in the trees from Nei's standard genetic distance constructed from all 40 global populations (not shown). Further, the infinite allele model assumed by Nei's (1972) distance is likely to be inaccurate for microsatellites, which are thought to undergo mostly stepwise mutations (Primmer et al. 1996). The $(\delta\mu)^2$ distance of Goldstein et al.

(1995) was specifically created to model stepwise mutation but has been shown to perform poorly between long-separated populations, possibly due to varying rates of evolution between loci (Takezaki & Nei 2008). It is not preferable when populations are recently diverged either, due to its higher variation relative to measures such as chord distance (Takezaki & Nei 1996). This may explain why $(\delta\mu)^2$ was unable to detect *B. germanica* population structure within Raleigh, even within a single building.

It is possible also that microsatellites may be an inappropriate tool to study *B. germanica* genetic diversity at the global scale. Mutations leading to null alleles are also more likely to occur between long-diverged populations. If null alleles do indeed occur in German cockroach populations and they are sufficiently widespread they can lead to severe overestimates of differentiation and genetic distance (Chapuis & Estoup 2006). In addition, allele size constraints (Goldstein & Pollock 1997) and varying rates of evolution between loci (Ellegren 2000a,b) mean that genetic distance is unlikely to arise linearly between distantly related populations (Estoup & Angers 1998). Therefore, mutation model is important for inferences of long term evolutionary relationships such as those between cities and between continents.

All 40 global *B. germanica* populations were also sequenced for the mitochondrial 16S gene. Mitochondrial gene sequences may be more highly conserved across closely related populations, although mtDNA point mutations occur at a much greater rate than in coding nuclear DNA (Parker et al. 1998). Also, since mtDNA is maternally transmitted and does not recombine it has one quarter the effective population size of nuclear DNA and tends to become diagnostic of a certain lineage much more quickly than nuclear loci due to increased genetic drift (Hedrick 2005b). However, of the 66 sequenced individuals from 36 global populations, only 10 haplotypes were found and average nucleotide difference (Tajima 1983) between populations was only 2.63. A haplotype network was unable to clearly detect lineages specific to apartment complex, city, or even continent.

One potential reason for the lack of pattern of mtDNA differentiation is that this single locus may not be informative for the divergence history of our populations. Trees based on a single gene are likely to be far less informative than those based on multiple independent genes because of the large variance in genealogical histories experienced by different genes (Hey & Machado 2003).

Another possibility may be that the one or two individuals that we sequenced for 16S per population may also have been too few to detect any pattern of differentiation by haplotype network even if such a pattern exists. Variance within a single locus such as mtDNA can lead to vastly different genealogical trees from different alleles, even within the same population (Hey & Machado 2003). For this reason, the low number of German cockroaches that we sequenced per population may not encompass all of the diversity within a population. Sequencing additional individuals or additional loci may create a more accurate network of global relationships between *B. germanica* populations.

Genetic distance as a function of geographic distance

Pairwise F_{ST} values and phylogenetic trees from genetic distance data have shown that German cockroach populations do not share greater genetic similarity with geographically closer populations than with distant populations at some spatial scales. Testing the correlation between genetic distance and geographic distance allows for a formal test of significance. Also, allele permutation tests allow us to determine whether mutations significantly contribute to genetic differentiation between German cockroach populations relative to drift.

Although phylogenetic trees and pairwise F_{ST} values within Raleigh did not reveal German cockroach populations to be more related within apartment complexes, a Mantel test for all individuals within Raleigh was significant. On the whole across Raleigh, genetic distance between *B. germanica* populations is significantly correlated with geographic distance. However, since distances within any apartment complex were set to zero, these results do not provide an estimate of the relationship between German cockroach populations in separate apartments within a complex. Also, since Mantel tests only test the overall trend of correlation between genetic distance and geographic distance the test can be made significant by a few pairs of strongly correlated apartments. For example, there would be a very strong correlation between apartments within complex CW where overall differentiation was small and genetic distance would be zero. Regardless, the test shows a surprising correlation which is in contrast to the phylogenetic data. It suggests that gene flow between German cockroaches within Raleigh occurs more frequently within apartment complexes than between them, and between more proximate apartment complexes than between distant complexes.

As previously discussed, indirect genetic distance measures, such as F_{ST}, are unable to separate the relative effects of mutation and gene flow which contribute to genetic differentiation. Allele permutation tests were performed at each spatial scale to determine whether stepwise mutations significantly contribute to population structure between German cockroach populations at a given range of distances. This test determines whether the stepwise differentiation measure R_{ST} is significantly different from F_{ST}. The result of the test for the 18 Raleigh German cockroach populations was significant (Table 5). This is also a surprising result because it suggests that mutation plays a part in shaping the differentiation across Raleigh. We expected that the populations would be linked by passive dispersal frequently enough that new mutations would be shared among populations. The test was only weakly significant, and since the permuted R_{ST} has a greater variance than F_{ST} (Hardy & Vekemans 2002) this is a possible reason for the significance. Another possibility is that although microsatellites are assumed by this test to undergo stepwise mutation, larger mutations have been shown to occur (e.g., Huang et al. 2002). This would violate the assumptions of the test and may lead to a spurious significant result, although the test is thought to be robust to small numbers of non-stepwise mutations (Hardy et al. 2003).

Results of the Mantel test were significant across all United States populations, but were not significant when Raleigh was excluded (Fig. 6). Similarly, the test for Eurasian samples was also non-significant, with a negative correlation between genetic distance and geographic distance. The Mantel test for the United States with the Raleigh populations excluded clearly shows that differentiation is lower within Raleigh than at larger geographic distance. This is the expected result, but is in direct contrast to the phylogenentic trees, which showed no clustering of the Raleigh populations. The results for the Mantel test with Raleigh populations excluded and for the Eurasian samples was likely not significant because the geographic distances involved were outside the range of those expected to follow consistent patterns of isolation-by-distance. Many pairwise distances were in excess of 1000 km, and German cockroaches would likely be carried by random long-distance jumps at this range.

Results of the allele permutation tests were not significant, even with the Raleigh populations excluded (Table 5). However, the test for Eurasian samples was also highly significant. Only the Eurasian results are consistent with expectations since populations of *B. germanica* are not likely to exchange frequent gene flow at this distance, and new mutations should contribute to differentiation between distant populations. However, the U.S. result is consistent with the lack of structure within phylogenetic trees for U.S. samples, as well as the global F_{ST} within the U.S., which indicated very low divergence overall.

The Mantel test for all global populations was significant both with Raleigh included and excluded (Table 7). Although relationships between German cockroach populations were not expected to follow strict isolation-by-distance patterns at this scale, the Mantel tests indicate the relative greater frequency of *B. germanica* dispersal within continents than between them. However, only the neighbor-joining tree from chord distances was able to clearly cluster some of the Eurasian samples on a separate branch (Fig. 10).

The allele permutation test for global populations was not significant with Raleigh populations excluded (Table 5). The lack of mutations differentiating U.S. samples from Eurasian samples is surprising, and it is unclear why this test was significant when the geographically neighboring samples from Raleigh were included.

Conclusions and future directions

The general lack of a clear pattern of structure and low differentiation among *B. germanica* populations, even on a global scale, does not allow us to make many inferences about the processes of dispersal, gene flow, and differentiation. Populations which have experienced recent gene flow would be expected to share common alleles and allele frequencies, as was detected within and between apartments within a single building. However, the genetic distances used to construct phylogenetic trees were not smaller (a) between apartments in an apartment complex than between apartment complexes, (b) within a city than across a continent, or (c) within a continent than between continents. We found little evidence of ongoing gene flow beyond the scale of a single building.

By contrast, if populations have been long separated, new mutations would occur in each population and become widespread due to drift. This trend would differentiate populations which have not experienced recent gene flow. However, genetic distances and *Structure* clustering were unable to clearly discriminate between different apartment complexes, different cities, or even different continents. Further, no significant mutation was found to contribute to genetic differentiation of German cockroach populations across the entire United States or at a global scale.

No such pattern of differentiation was seen across *B. germanica* in our study. In fact, overall differentiation was fairly low even at a global scale. The reason for this lack of diversity is likely the frequent disruption that German cockroach populations likely endure. Insecticides, large fluctuations in temperature, humidity, and food availability, and physical changes in their habitat (e.g., tenant turnover, renovations) could all cause highly localized and dramatic declines in population size. These population bottlenecks are known to rapidly change allele frequencies and affect various evolutionary estimates, including genetic distance (Paetkau et al. 1997). Since new mutations are initially rare in the population they would most likely be lost in a population bottleneck (Hedrick 2005b). This may explain the lack of a mutation effect in our *B. germanica* populations.

No recent population bottleneck is likely to have affected most of our German cockroach populations, as many of them have high levels of genetic diversity, although almost none of them were in Hardy-Weinberg equilibrium. Another possibility is that recent gene flow amongst German cockroaches within buildings has restored diversity. High gene flow within buildings was suggested by F_{ST} values, *Structure*, and phylogenetic tree results for the apartments from CW, CC, and CS. If neighboring apartments had contributed many immigrants to our sampled populations then our samples would not be representative of cockroaches from several apartments. This may be the case with the poor performance of the assignment tests. The accidental sampling of many subpopulations may also lead to an apparent heterozygote deficiency and cause departures from Hardy-Weinberg equilibrium, known as a Wahlund effect (Wahlund 1928).

Overall, our study shows that active dispersal of German cockroaches within a building occurs frequently. However, beyond the scale of a single building, gene flow by passive dispersal seems to have formed no spatial clustering of genetic similarity, even at the global scale. Any gene flow of *B. germanica* that does occur by passive human-mediated dispersal would likely go undetected by the microsatellite markers used in the current study, as there was a global lack of marker variation. It seems from the results of the current study that the unstable environment in which German cockroaches live contributes to the dynamic, non-equilibrium nature of their populations.

A few lessons may benefit future studies of German cockroach dispersal and differentiation. It would be beneficial to use markers which are sufficiently variable to detect intra-specific variation but less subject to the drift and non-equilibrium conditions that are likely to affect German cockroach populations. If microsatellites are to be used, a greater number of loci may be required to overcome the loss of diversity from these conditions. Gene sequencing may be the preferred option since it captures the highest possible resolution of genetic variation with no bias (Schlötterer 2004). It is unclear whether mtDNA will be a feasible option, since only 10 haplotypes were found across our 40 global populations. There may prove to be too little variation in the mitochondrial genome of *B. germanica* to investigate levels of gene flow and differentiation. The challenge will be to find a nuclear or mitochondrial gene which has enough variation for population-level analysis but does not require so many sequences to be obtained per population so as to be prohibitively expensive.

Another change in future studies might be to the sampling scheme. If German cockroaches move frequently between apartments in a building, as is suggested here, future studies may sample many more apartments per building to achieve a clearer understanding of the range of active dispersal. Similarly, passive dispersal may occur between apartments within a complex, though only 3 apartments were sampled in the current study. Only more intensive sampling would be able to detect less frequent movement, although this may be prohibitively difficult and expensive. Finally, future studies would benefit from sampling within buildings and between buildings within the same apartment complex. This change would make the two spatial scales more directly comparable.

Table 1. *B. germanica* within apartment sample diversity statistics across nine loci. Sample location, sample size (n), Nei's (1987) unbiased gene diversity (H_E), observed heterozygosity (H_O), and mean number of alleles per locus (N_A), results for Hardy-Weinberg equilibrium tests (HWE), and linkage disequilibrium tests (LD)

Sample location[a]	N	H_E	H_O	N_A	HWE[b]	LD[c]
DAC-A K1	30	0.72	0.65	8.56	**	2
DAC-A K2	30	0.74	0.65	8.56	**	ns
DAC-A vK2	30	0.75	0.74	8.56	*	1
DAC-A LR	30	0.72	0.73	7.78	ns	1
DAC-A BR	30	0.70	0.62	8.44	***	2
Across all rooms:					***	*ns*
RC-A K1	30	0.73	0.71	7.56	**	5
RC-A K2	30	0.72	0.66	7.44	**	2
RC-A vK2	30	0.73	0.69	7.67	**	6
RC-A LR	30	0.72	0.68	7.11	**	3
RC-A BR	21	0.73	0.67	6.67	*	2
Across all rooms:					***	*5*
FR-A K1	30	0.76	0.70	8.67	**	2
FR-A K2	29	0.76	0.65	8.00	***	ns
FR-A vK1	30	0.78	0.71	9.11	***	1
FR-A LR	30	0.74	0.70	8.22	**	2
FR-A BR	30	0.76	0.69	8.44	***	1
Across all rooms:					***	*ns*

[a] DAC, RC, and FR are three different apartment complexes; A represents an apartment within each complex; K = kitchen, vK = kitchen vacuum sample, LR = living room, BR = bedroom; Numeral refers to location of trap or vacuum sample within the kitchen.

[b] Significance of Hardy-Weinberg tests (Raymond & Rousset 1995): * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; ns, not significant.

[c] LD tests: number of pairwise comparisons between loci (of 36) significant at $P < 0.05$.

Table 2. Results for exact *G*-test for departures from panmixia per pair of *B. germanica* samples within each apartment, global F_{IS} for each apartment, 95% confidence interval for F_{IS} (95% CI), global F_{ST} for each apartment, 95% confidence interval for F_{ST} (95% CI)

a.

Apartment[a]	**Sample Pair**[b]		***G*-test**[c]
DAC-A	K2	vK2	ns
RC-A	K2	vK2	ns
FR-A	K2	vK2	ns

b.

Apartment[a]	**Sample Pair**[b]		***G*-test**[c]	F_{IS}	F_{IS} **95% CI**	F_{ST}	F_{ST} **95% CI**
DAC-A				0.067	-0.003, 0.144	0.001	-0.003, 0.006
	K1	K2	ns				
	K1	LR	ns				
	K1	BR	ns				
	K2	LR	ns				
	K2	BR	ns				
	LR	BR	*				
RC-A				0.058	-0.013, 0.151	0.002	-0.007, 0.003
	K1	K2	ns				
	K1	LR	*				
	K1	BR	ns				
	K2	LR	ns				
	K2	BR	ns				
	LR	BR	ns				
FR-A				0.089	-0.007, 0.221	0.002	-0.003, 0.008
	K1	K2	ns				
	K1	LR	ns				
	K1	BR	ns				
	K2	LR	ns				
	K2	BR	ns				
	LR	BR	ns				

[a] DAC, RC, and FR are three different apartment complexes; A represents an apartment within each complex.[b] K = kitchen, vK = kitchen vacuum sample, LR = living room, BR = bedroom; Numeral refers to location of trap or vacuum sample within the kitchen.

[c] Significance of exact *G*-test (Goudet et al. 1996): * $P < 0.05$; ns, not significant.

Table 3. Raleigh apartment *B. germanica* diversity statistics across nine loci: Apartment, sample size (n), Nei's (1987) unbiased gene diversity (H_E), observed heterozygosity (H_O), and mean number of alleles per locus (N_A), results for Hardy-Weinberg equilibrium tests (HWE), and linkage disequilibrium tests (LD).

Apartment[a]	**n**	H_E	H_O	N_A	**HWE**[b]	**LD**[c]
DAC-A	30	0.72	0.65	8.56	**	2
DAC-B	30	0.74	0.68	8.44	**	2
DAC-C	30	0.71	0.67	8.22	**	1
RC-A	30	0.73	0.71	7.56	**	5
RC-B	30	0.75	0.70	8.00	***	4
RC-C	30	0.71	0.72	7.22	Ns	8
FR-A	30	0.76	0.70	8.67	**	2
FR-B	30	0.71	0.63	7.22	***	5
FR-C	30	0.70	0.61	8.33	***	3
CW-X	29	0.76	0.72	8.00	***	2
CW-Y	30	0.77	0.76	8.22	*	6
CW-Z	30	0.76	0.67	7.78	***	ns
CC-X	30	0.76	0.66	8.56	***	14
CC-Y	30	0.76	0.68	8.33	***	11
CC-Z	29	0.77	0.75	8.44	**	4
CS-X	30	0.76	0.70	9.11	Ns	4
CS-Y	29	0.77	0.73	8.89	**	2
CS-Z	30	0.71	0.62	7.11	*	2

[a] DAC, RC, FR, CW, CC and CS are six different apartment complexes; A, B, and C represent individual apartments within each individual buildings within the complex; X and Y represent adjacent apartments sharing a wall and Z represents a distant apartment within the same building.

[b] Significance of Hardy-Weinberg tests (Raymond & Rousset 1995): * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; ns, not significant.

[c] LD tests: number of pairwise comparisons between loci (of 36) significant at $P < 0.05$.

Table 4. Results for exact *G*-test for departures from panmixia per pair of *B. germanica* apartment samples, global F_{IS} for each complex, 95% confidence interval for F_{IS} (95% CI)global F_{ST} for each complex, and 95% confidence interval for F_{ST} (95% CI)

Apartment complex	Apartment pair[a]	*G*-test[b]	F_{IS}	F_{IS} 95% CI	F_{ST}	F_{ST} 95% CI
Within buildings						
CW			0.062	-0.007, 0.132	0.017	0.009, 0.025
	X & Y	***				
	X & Z	***				
	Y & Z	***				
CC			0.088	0.018, 0.167	0.029	0.018, 0.038
	X & Y	***				
	X & Z	***				
	Y & Z	***				
CS			0.080	0.019, 0.152	0.019	0.007, 0.039
	X & Y	ns				
	X & Z	***				
	Y & Z	***				
Between buildings						
DAC			0.083	0.025, 0.150	0.028	0.011, 0.045
	A & B	***				
	A & C	***				
	B & C	***				
RC			0.058	-0.013, 0.151	0.030	0.018, 0.042
	A & B	***				
	A & C	***				
	B & C	***				
FR			0.112	0.010, 0.227	0.027	0.012, 0.046
	A & B	***				
	A & C	***				
	B & C	***				

[a] X and Y represent adjacent apartments sharing a wall and Z represents a distant apartment within the same building; A, B, and C represent individual apartments within each of three individual buildings, respectively, within the apartment complex.

[b] Significance of exact *G*-test (Goudet et al. 1996): *** $P < 0.001$; ns, not significant.

Table 5. Global F_{IS} for *B. germanica* populations at the city, U.S., international and global spatial scales, 95% confidence interval for F_{IS} (95% CI), Global F_{ST} at each spatial scales, 95% confidence interval for F_{ST} (95% CI) and results for allele permutations tests on *B. germanica* populations at several spatial scales: Global R_{ST}, global permuted R_{ST} (pR_{ST}), and significance for the comparison test

Spatial scale	F_{IS}	**F_{IS} 95% CI**	F_{ST}	**F_{ST} 95% CI**	R_{ST}	pR_{ST}	***P*-value**[a]
Raleigh	0.075	0.014, 0.149	0.046	0.029, 0.078	0.056	0.038	0.047
U.S.	0.078	0.026, 0.138	0.071	0.053, 0.098	0.077	0.068	ns
-Raleigh excluded	0.081	0.048, 0.118	0.099	0.074, 0.130	0.100	0.105	ns
International	0.140	0.037, 0.266	0.266	0.223, 0.311	0.349	0.239	0.023
Global	0.086	0.028, 0.155	0.106	0.088, 0.131	0.127	0.100	0.039
-Raleigh excluded	0.097	0.041, 0.161	0.158	0.135, 0.184	0.181	0.151	ns

[a] *P*-value for test comparing R_{ST} and pR_{ST} with H_1: $R_{ST} > pR_{ST}$

ns, not significant at α = 0.05

Table 6. F_{ST} values for each pair of *B. germanica* apartment populations within Raleigh, NC

	DAC-A	DAC-B	DAC-C	RC-A	RC-B	RC-C	FR-A	FR-B	FR-C	CW-X	CW-Y	CW-Z	CC-X	CC-Y	CC-Z	CS-X	CS-Y
DAC-A																	
DAC-B	0.023																
DAC-C	0.013	0.046															
RC-A	0.020	0.028	0.020														
RC-B	0.015	0.026	0.023	0.013													
RC-C	0.046	0.034	0.049	0.038	0.038												
FR-A	0.031	0.027	0.035	0.017	0.007	0.032											
FR-B	0.030	0.041	0.011	0.024	0.019	0.054	0.025										
FR-C	0.014	0.040	0.017	0.019	0.018	0.048	0.027	0.030									
CW-X	0.054	0.047	0.058	0.039	0.021	0.058	0.021	0.043	0.051								
CW-Y	0.049	0.053	0.036	0.032	0.030	0.060	0.034	0.038	0.039	0.018							
CW-Z	0.051	0.060	0.047	0.046	0.034	0.066	0.042	0.061	0.053	0.022	0.011						
CC-X	0.083	0.107	0.099	0.080	0.079	0.095	0.072	0.101	0.103	0.095	0.093	0.094					
CC-Y	0.078	0.073	0.079	0.058	0.056	0.061	0.047	0.078	0.076	0.066	0.064	0.079	0.036				
CC-Z	0.063	0.069	0.072	0.065	0.051	0.064	0.050	0.067	0.079	0.056	0.063	0.077	0.026	0.026			
CS-X	0.015	0.029	0.028	0.013	0.020	0.036	0.020	0.037	0.018	0.033	0.040	0.041	0.082	0.065	0.066		
CS-Y	0.020	0.024	0.020	0.010	0.014	0.031	0.012	0.031	0.024	0.030	0.024	0.032	0.079	0.046	0.048	0.008	
CS-Z	0.051	0.052	0.051	0.033	0.040	0.057	0.040	0.062	0.061	0.058	0.040	0.052	0.097	0.055	0.073	0.038	0.011

Table 7. Results for isolation by distance tests for *B. germanica* populations at various spatial scales: slope of the regression, and *P*-value for Mantel test

Spatial scale	Slope	*P*-value[a]
Raleigh	0.0052	<0.001
Continental U.S.	0.0062	0.029
-Raleigh excluded	-0.0060	ns
International	-0.0785	ns
Global	0.0193	<0.001
-Raleigh excluded	0.0439	0.005

[a] for significance of Mantel test: ns, not significant at $\alpha = 0.05$

Table 8. Percentage of *B. germanica* individuals assigned to the population from which they were sampled using two assignment programs: *Structure* and by two separate criteria for *Geneclass*

		Structure[a]	Geneclass[b]	
Population	n	% assigned	% assigned	% by exclusion
DAC-A	30	56.7	3.3	0
DAC-B	30	76.7	10.0	0
DAC-C	30	73.3	10.0	0
RC-A	30	66.7	10.0	0
RC-B	30	73.3	3.3	6.7
RC-C	30	80.0	13.3	0
FR-A	30	56.7	6.7	10.0
FR-B	30	80.0	6.7	6.7
FR-C	30	66.7	3.3	3.3
CW-X	29	82.8	10.3	6.9
CW-Y	30	73.3	0	10.0
CW-Z	30	66.7	0	3.3
CC-X	30	93.3	6.7	30.0
CC-Y	30	73.3	6.7	33.3
CC-Z	29	79.3	6.9	6.9
CS-X	30	46.7	6.7	6.7
CS-Y	29	58.6	10.3	6.9
CS-Z	30	70.0	0	0
Mean:		**70.8**	**6.3**	**7.3**

[a] Structure assignment criteria greater than 95% probability

[b] Geneclass assigned by 95% probability; or all populations excluded (less than 5% probability) except sampled population

Table 9. *B. germanica* diversity statistics for U.S. apartments across nine loci: Apartment, sample size (n), Nei's (1987) unbiased gene diversity (H_E), observed heterozygosity (H_O), and mean number of alleles per locus (N_A), results for Hardy-Weinberg equilibrium tests (HWE), and linkage disequilibrium tests (LD)

Location	Code	n	H_E	H_O	N_A	HWE[a]	LD[b]
Norfolk, Virginia	NVA	30	0.72	0.74	6.67	ns	2
Richmond, Virginia	RVA	30	0.69	0.65	6.44	ns	2
Hickory, North Carolina	HNC	30	0.56	0.49	4.89	***	6
Griffin, Georgia	GGA	29	0.70	0.64	7.00	ns	3
Bryan, Texas	BTX	30	0.73	0.69	8.22	**	3
Baton Rouge, Louisiana	BRL	22	0.70	0.57	7.33	***	1
White Eagle, Oklahoma	WOK	29	0.74	0.75	8.44	ns	2
Gary, Indiana	GIN	30	0.70	0.61	8.00	***	ns
Orange, California	OCA	8	0.56	0.57	3.44	ns	ns
Los Angeles, California	LCA	13	0.73	0.69	6.33	ns	1
Compton, California	CCA	29	0.74	0.64	8.11	***	2
Cleveland, Ohio	COH	27	0.65	0.62	6.22	***	6
Gainesville, Florida	GFL	28	0.75	0.73	9.11	*	ns
Minneapolis, Minnesota	MMN	30	0.73	0.54	8.33	***	6
Hawaii	HAW	10	0.61	0.59	3.78	ns	1
Puerto Rico	PR	30	0.75	0.69	8.44	***	2

[a] Significance of Hardy-Weinberg tests (Raymond & Rousset 1995): * $P < 0.05$, **$P < 0.01$, ***$P < 0.001$; ns, not significant

[b] LD tests - number of pairwise comparisons between loci (of 36) significant at $P < 0.05$

Table 10. *B. germanica* diversity statistics for international apartments across nine loci: Apartment, sample size (n), Nei's (1987) unbiased gene diversity (H_E), observed heterozygosity (H_O), and mean number of alleles per locus (N_A), results for Hardy-Weinberg equilibrium tests (HWE), and linkage disequilibrium tests (LD)

Location	Code	n	H_E	H_O	N_A	HWE[a]	LD[b]
Moscow, Russia	MRU	30	0.57	0.45	4.11	***	3
Tomsk, Russia	TRU	30	0.61	0.51	6.00	***	2
Kiev, Ukraine	KUA	30	0.45	0.44	3.00	ns	2
Crimea, Ukraine	CUA	30	0.52	0.47	3.67	***	8
Tehran, Iran	TIR	30	0.52	0.43	4.33	***	2
Singapore	SIN	30	0.78	0.67	8.22	***	9

[a] Significance of Hardy-Weinberg tests (Raymond & Rousset 1995): * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; ns, not significant

[b] LD tests - number of pairwise comparisons between loci (of 36) significant at $P < 0.05$

Table 11. Summary of *B. germanica* mitochondrial 16S haplotypes for 66 individuals; percent frequency of haplotype and samples for each haplotype

Haplotype[a]	% frequency	Samples		
1	7.6	BRL 1	RC-C 2	RVA 1
		DAC-C 2	TIR 1	
2	39.4	CC-Z 1	CS-Z 2	HAW 2
		CC-Z 2	CW-Z 1	HNC 2
		CC-X 1	DAC-A 1	LCA 2
		CCA 1	DAC-A 2	PR 1
		COH 1	FR-A 1	TIR 2
		COH 2	GFL 1	NVA 2
		CS-X 1	GFL 2	WOK 1
		CS-Y 2	GGA 2	WOK 2
		CS-Z 1	HAW 1	
3	1.5	GIN 2		
4	1.5	GGA 1		
5	1.5	NVA 1		
6	27.3	CS-X 2	OCA 1	TRU 1
		CW-Y 1	OCA 2	TRU 3
		DAC-B 1	RC-B 1	KUA 1
		HNC 1	RC-B 2	KUA 2
		LCA 1	MRU 1	CUA 1
		MMN 1	MRU 2	CUA 2
7	3.0	CC-Y 1	CW-Z 2	SIN 1
8	15.2	CCA 2	FR-A 2	SIN 2
		CS-Y 1	GIN 1	
		DAC-C 1	MMN 2	
		DAC-B 2	RC-C 1	
9	1.5	CC-X 2		
10	1.5	FR-B 1		

[a] Haplotype number corresponds to haplotype network in Figure 11

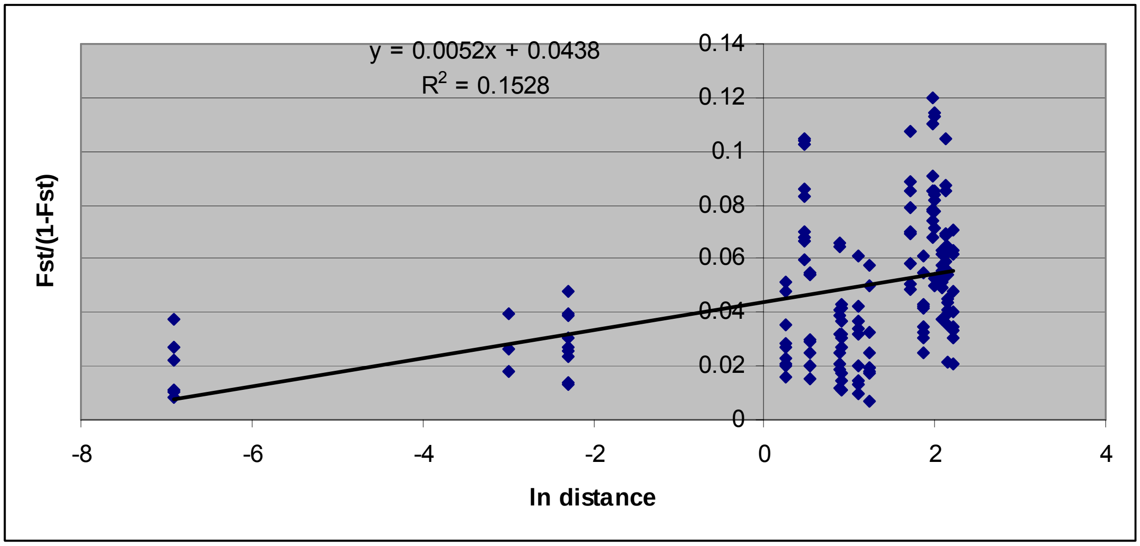


Figure 1. Isolation by distance analysis for *B. germanica* between apartments within Raleigh. The equation for the regression analysis and R^2 value are given. The results for the Mantel test of significance are given in Table 7.

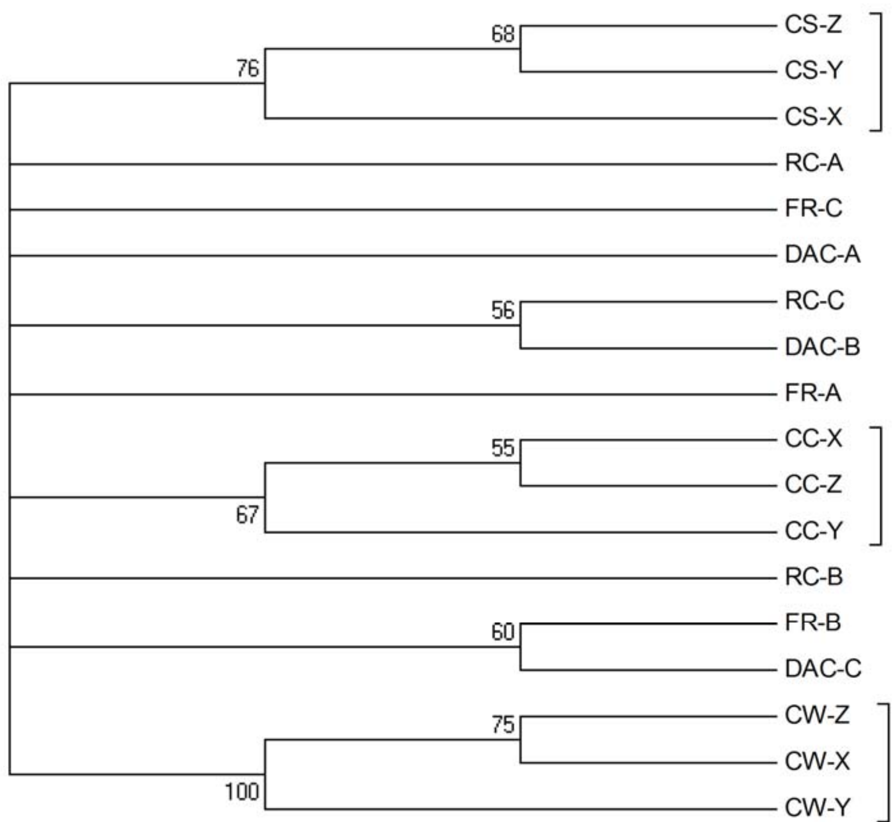


Figure 2. Neighbor-joining tree from chord distances (Cavalli-Sforza & Edwards 1967) for *B. germanica* apartment populations in Raleigh. Branch support is given as a percentage of 1000 bootstraps, nodes with less than 50% support have been collapsed. Brackets indicate clusters consisting of apartments from the same apartment complex.

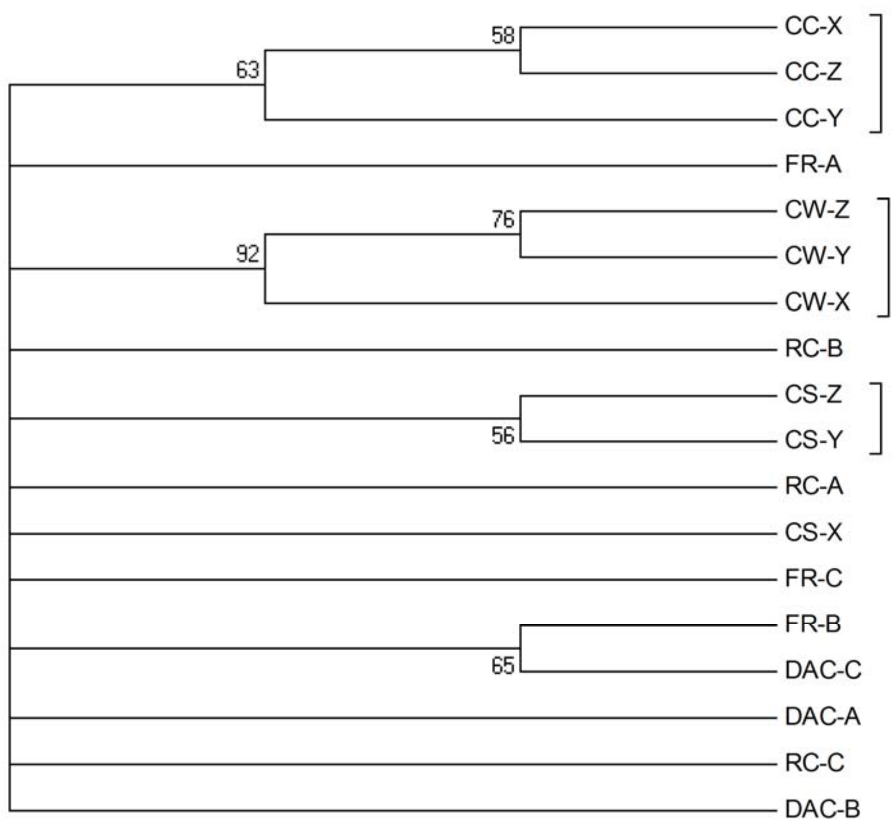


Figure 3. Neighbor-joining tree from standard genetic distances (Nei 1972) for *B. germanica* apartment populations in Raleigh. Branch support is given as a percentage of 1000 bootstraps, nodes with less than 50% support have been collapsed. Brackets indicate clusters consisting of apartments from the same apartment complex.

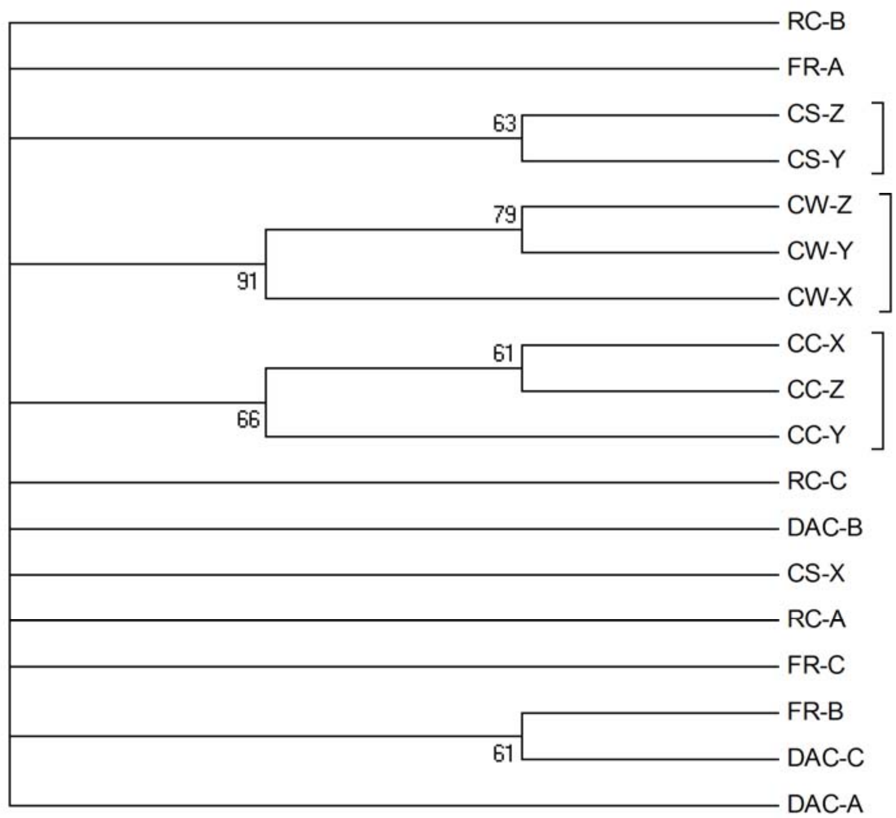


Figure 4. Neighbor-joining tree from coancestry coefficients (Reynolds et al. 1983) for *B. germanica* apartment populations in Raleigh. Branch support is given as a percentage of 1000 bootstraps, nodes with less than 50% support have been collapsed. Brackets indicate clusters consisting of apartments from the same apartment complex.

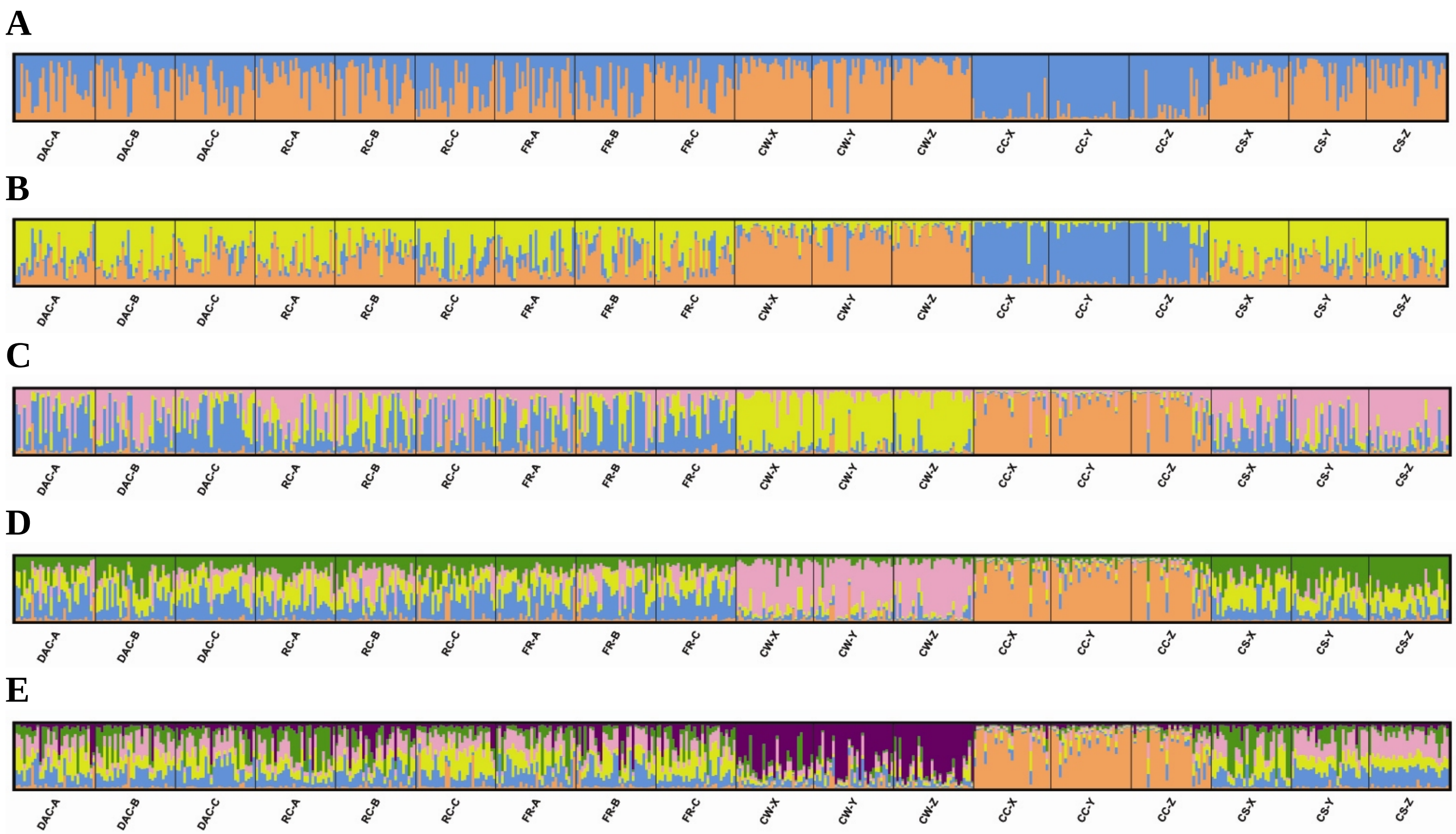


Figure 5. *Structure* clustering results of Raleigh *B. germanica* apartment populations, for *K* values from 2–6 (A–E). The plot for the optimal number of clusters, *K* = 4 (C), represents 10 sequential *Structure* runs aligned using CLUMPP (Jakobsson & Rosenberg 2007). All other plots represent single *Structure* runs.

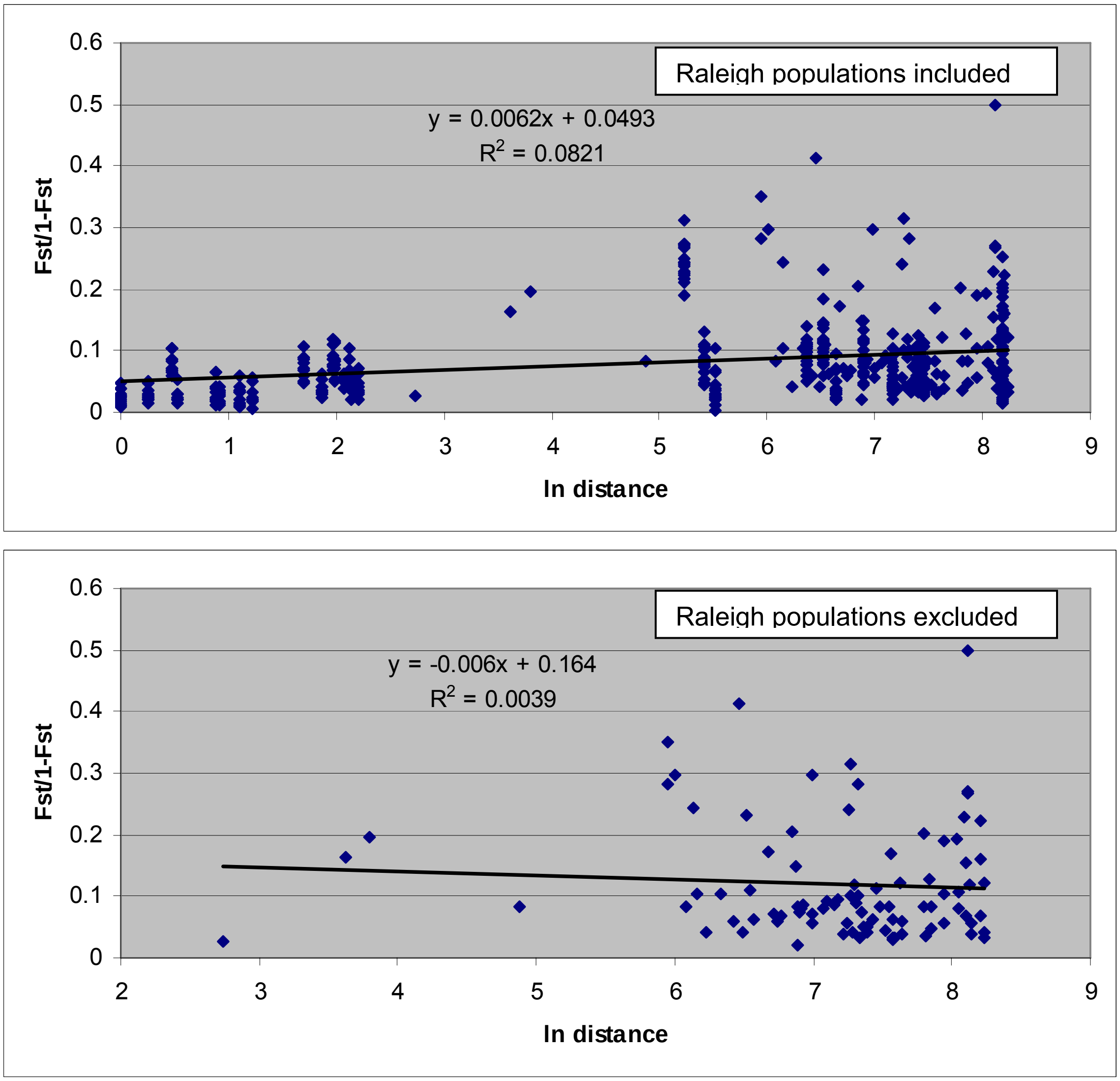


Figure 6. Isolation by distance analysis for *B. germanica* across all U.S. populations with and without Raleigh samples included. The equation for the regression analysis and R^2 value are given. The results for the Mantel test of significance are given in Table 7.

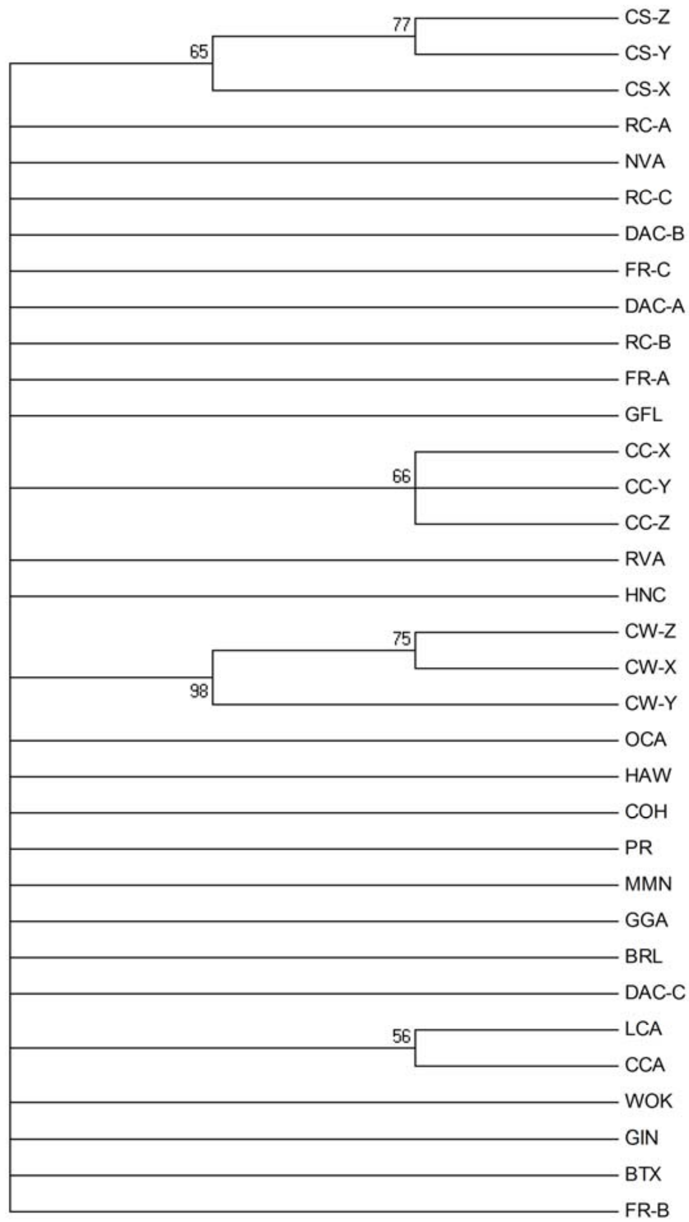


Figure 7. Neighbor-joining tree from chord distances (Cavalli-Sforza & Edwards 1967) for *B. germanica* apartment populations in Raleigh. Branch support is given as a percentage of 1000 bootstraps, nodes with less than 50% support have been collapsed.

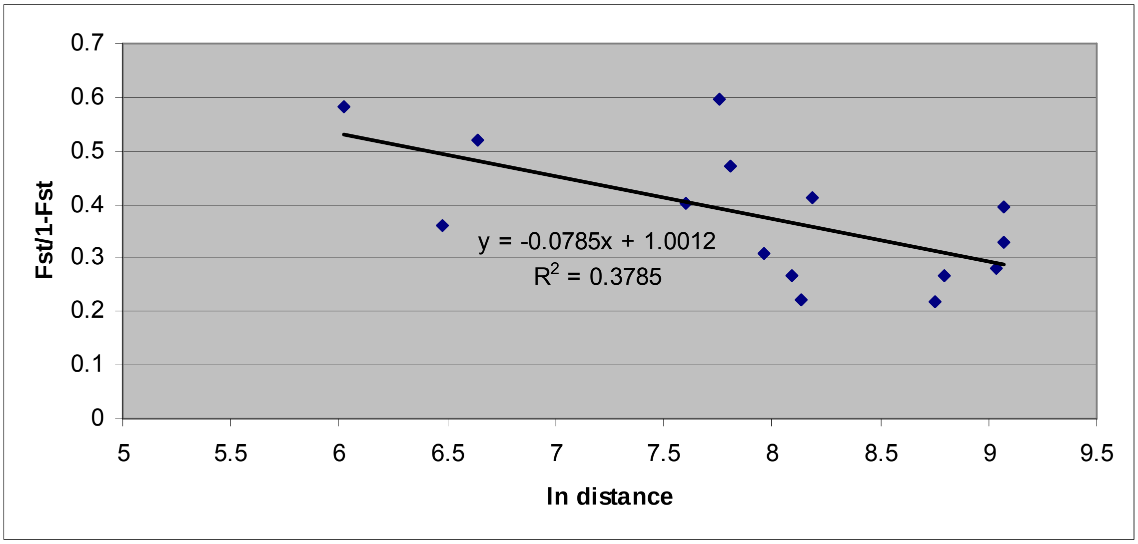
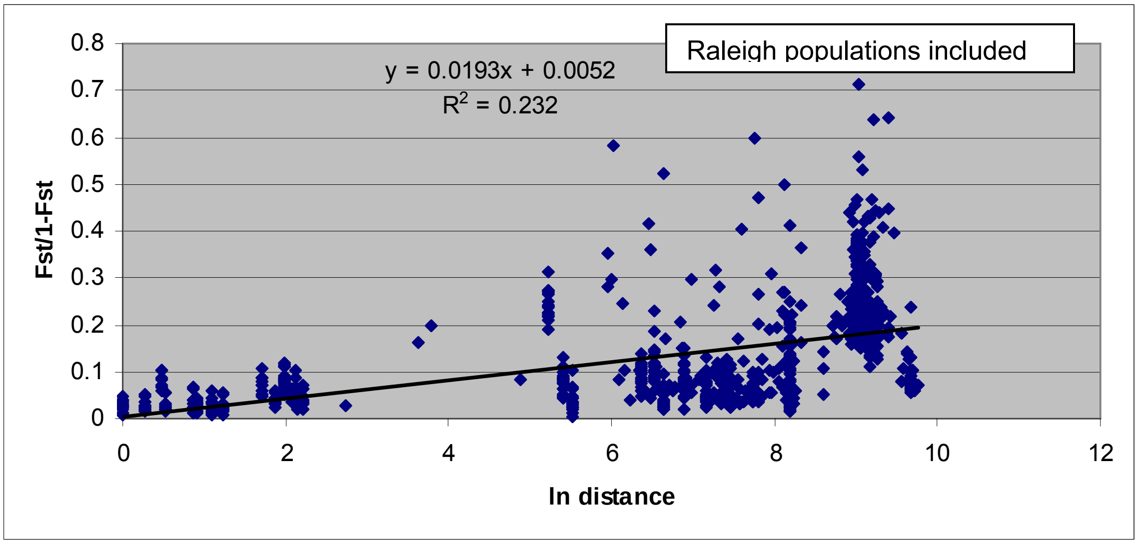


Figure 8. Isolation by distance analysis for Eurasian *B. germanica* populations. The equation for the regression analysis and R^2 value are given. The results for the Mantel test of significance are given in Table 7.



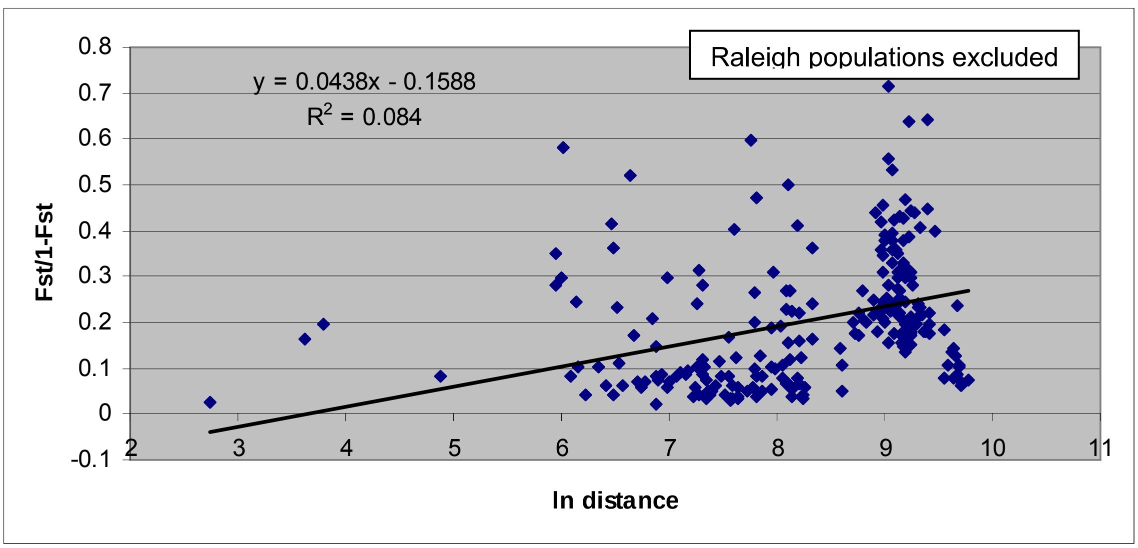


Figure 9. Isolation by distance analysis for *B. germanica* across all global populations with and without Raleigh samples excluded. The equation for the regression analysis and R^2 value are given. The results for the Mantel test of significance are given in Table 7.

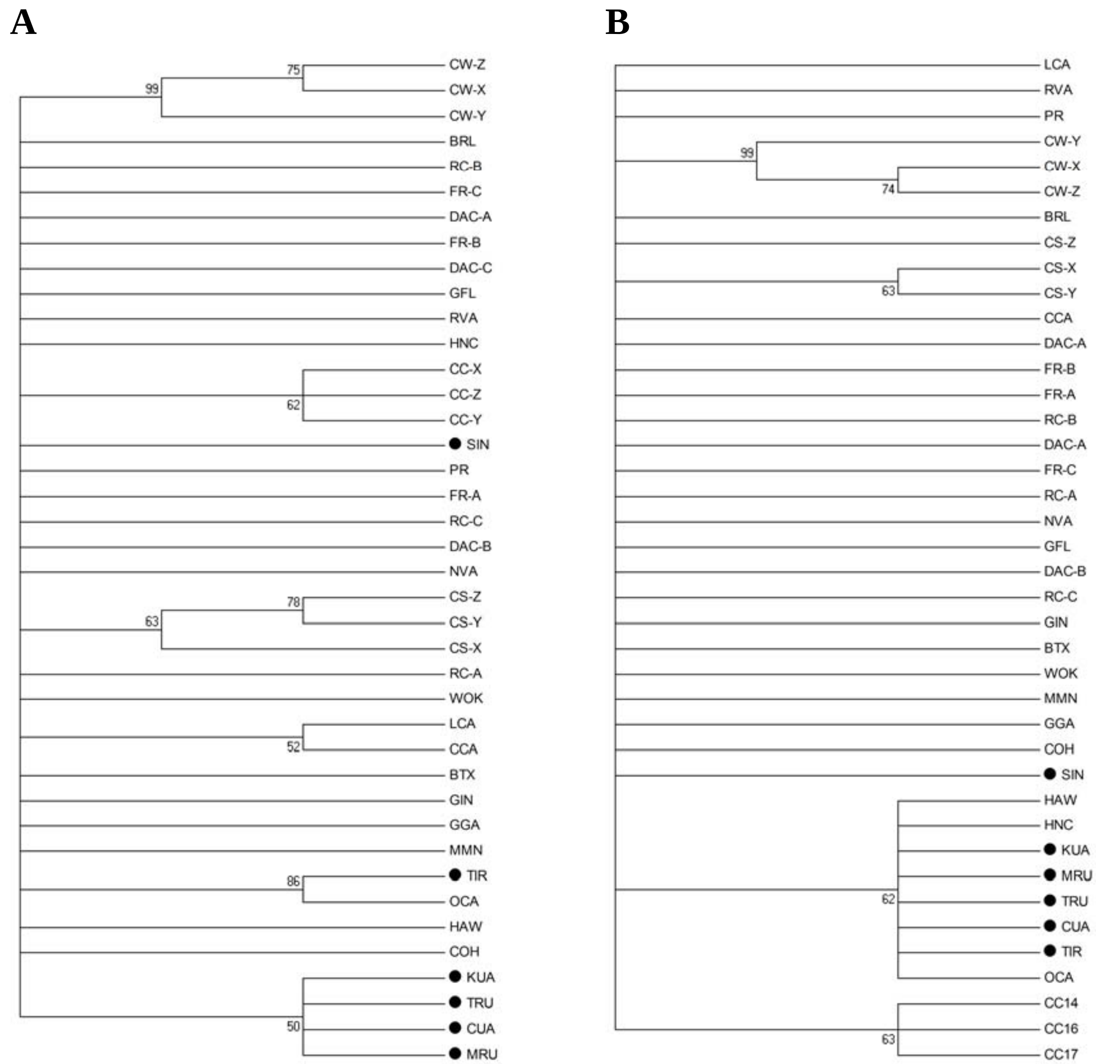


Figure 10. (A) Neighbor-joining and (B) UPGMA trees from chord distances (Cavalli-Sforza & Edwards 1967) for *B. germanica* apartment populations in Raleigh. Branch support is given as a percentage of 1000 bootstraps, nodes with less than 50% support have been collapsed. Filled circles denote samples from outside the U.S.

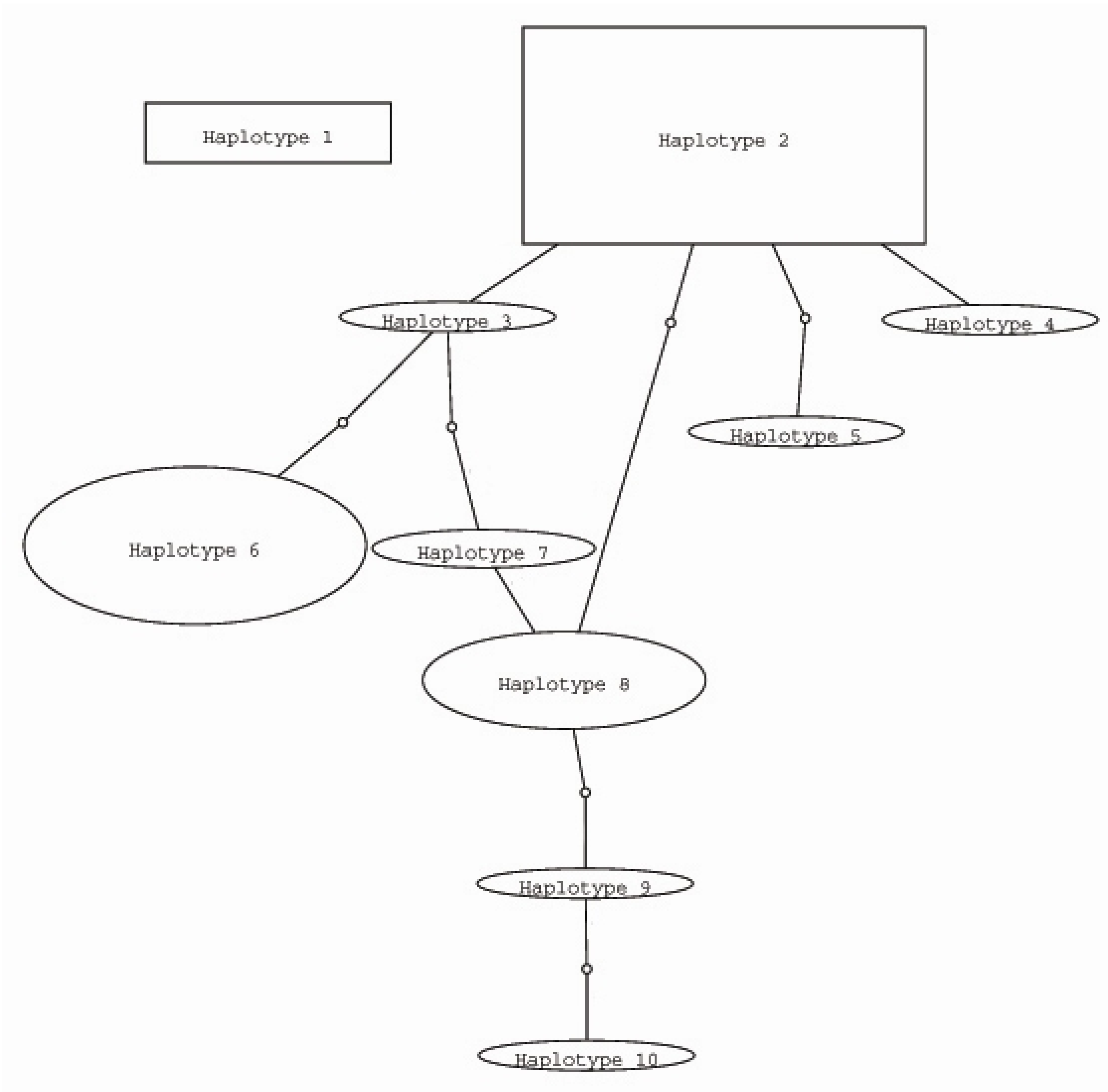


Figure 11. Haplotype network generated by *TCS* (Clement et al. 2000) from 66 *B. germanica* 16S sequences. A maximum of eight steps were possible within a parsimony probability of 95%. Haplotypes in rectangles have the greatest outgroup probability, size of ovals is proportional to haplotype frequency. Open circles represent proposed unsampled haplotypes. Frequency and samples contained in each haplotype are given in Table 11.unsampled haplotypes.

References

Akers RC, Robinson WH (1983) Comparison of movement behavior of 3 strains of German cockroach, *Blattella germanica*. *Entomologia Experimentalis Et Applicata* **34**, 143-147.

Ame JM, Halloy J, Rivault C, Detrain C, Deneubourg JL (2006) Collegial decision making based on social amplification leads to optimal group formation. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5835-5840.

Appel AG (1995) *Blattella* and related species. In: *Understanding and Controlling the German Cockroach* (eds. Rust MK, Owens JM, Reierson DA). Oxford University Press, New York.

Arbes SJ, Sever M, Archer J*, et al.* (2003) Abatement of cockroach allergen (Bla g 1) in low-income, urban housing: A randomized controlled trial. *Journal of Allergy and Clinical Immunology* **112**, 339-345.

Arbes SJ, Sever M, Mehta J*, et al.* (2004) Abatement of cockroach allergens (Bla g 1 and Bla g 2) in low-income, urban housing: Month 12 continuation results. *Journal of Allergy and Clinical Immunology* **113**, 109-114.

Armour JAL, Povey S, Jeremiah S, Jeffreys AJ (1990) Systematic cloning of human minisatellites from ordered array charomid libraries. *Genomics* **8**, 501-512.

Avise JC (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology* **7**, 371-379.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates, Inc., Sunderland, MA.

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

Bahlo M, Griffiths RC (2000) Inference from gene trees in a subdivided population. *Theoretical Population Biology* **57**, 79-95.

Balloux F, Brunner H, Lugon-Moulin N, Hausser J, Goudet J (2000) Microsatellites can be misleading: An empirical and simulation study. *Evolution* **54**, 1414-1422.

Balloux F, Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology* **11**, 771-783.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**, 155-165.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013-2029.

Beaumont MA (2003) Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**, 1139-1160.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences* **263**, 1619-1626.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics* **5**, 251-261.

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

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

Berry A, Kreitman M (1993) Molecular analysis of an allozyme cline: Alcohol-dehydrogenase in *Drosophila melanogaster* on the East coast of North America. *Genetics* **134**, 869-893.

Bohonak AJ (1999) Dispersal, gene flow, and population structure. *Quarterly Review of*

Biology **74**, 21-45.

Booth W, Bogdanowicz SM, Prodohl PA*, et al.* (2007) Identification and characterization of 10 polymorphic microsatellite loci in the German cockroach, *Blattella germanica*. *Molecular Ecology Notes* **7**, 648-650.

Bossart JL, Prowell DP (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology & Evolution* **13**, 202-206.

Boursot P, Auffray JC, Brittondavidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics* **24**, 119-152.

Bret BL, Ross MH (1985) A laboratory study of German cockroach dispersal (Dictyoptera, Blattellidae). *Proceedings of the Entomological Society of Washington* **87**, 448-455.

Brouat C, Loiseau A, Kane M, Ba K, Duplantier JM (2007) Population genetic structure of two ecologically distinct multimammate rats: the commensal *Mastomys natalensis* and the wild *Mastomys erythroleucus* in southeastern Senegal. *Molecular Ecology* **16**, 2985-2997.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* **18**, 249-256.

Calabrese PP, Durrett RT, Aquadro CF (2001) Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics* **159**, 839-852.

Carvalho GR (1998) Molecular ecology: origins and approach. In: *Advances in Molecular Ecology* (ed. Carvalho GR), pp. 1-23. IOS Press; Ohmsha, Amsterdam; Washington, D.C.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: Models and estimation procedures. *Evolution* **21**, 550-570.

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 1041-1046.

Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **126**, 455-476.

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

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**, 538-543.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-1303.

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

Cloarec A, Rivault C, Cariou ML (1999) Genetic population structure of the German cockroach, *Blattella germanica*: Absence of geographical variation. *Entomologia Experimentalis Et Applicata* **92**, 311-319.

Cloarec A, Rivault C, Fontaine F, Leguyader A (1992) cockroaches as carriers of bacteria in multifamily dwellings. *Epidemiology and Infection* **109**, 483-490.

Cochran DG (1987) Selection for pyrethroid resistance in the German cockroach (Dictyoptera, Blattellidae). *Journal of Economic Entomology* **80**, 1117-1121.

Cohn RD, Arbes SJ, Jaramillo R, Reid LH, Zeldin DC (2006) National prevalence and exposure risk for cockroach allergen in US households. *Environmental Health Perspectives* **114**, 522-526.

Corander J, Waldmann P, Marttinen P, Sillanpaa MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**, 2363-2369.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.

Dallas JF, Bonhomme F, Boursot P, Britton-Davidian J, Bauchau V (1998) Population genetic structure in a Robertsonian race of house mice: evidence from microsatellite polymorphism. *Heredity* **80**, 70-77.

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

Denzer DJ, Fuchs MEA, Stein G (1988) Behavioral studies on *Blattella germanica* L: Radius of action and loyalty to the refuge. *Journal of Applied Entomology-Zeitschrift Fur Angewandte Entomologie* **105**, 330-343.

Dirienzo A, Peterson AC, Garza JC*, et al.* (1994) Mutational processes of simple sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 3166-3170.

Durier V, Rivault C (2001) Effects of spatial knowledge and feeding experience on foraging choices in German cockroaches. *Animal Behaviour* **62**, 681-688.

Durier V, Rivault C (2003) Exploitation of home range and spatial distribution of resources in German cockroaches (Dietyoptera : Blattellidae). *Journal of Economic Entomology* **96**, 1832-1837.

Edwards AWF, Cavalli-Sforza LL (1963) Reconstruction of evolution. *Heredity* **18**, 553-570.

Edwards AWF, Cavalli-Sforza LL (1964) Reconstruction of evolutionary trees. In: *Phenetic and Phylogenetic Classification* (eds. Heywood VH, McNeill J), pp. 67-76. Systematics Association, London.

Eggleston PA, Rosenstreich D, Lynn M*, et al.* (1998) Relationship of indoor allergen exposure to skin test sensitivity in inner-city children with asthma. *Journal of Allergy and Clinical Immunology* **102**, 563-570.

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

Elgderi RM, Ghenghesh KS, Berbash N (2006) Carriage by the German cockroach (*Blattella germanica*) of multiple-antibiotic-resistant bacteria that are potentially pathogenic to humans, in hospitals and households in Tripoli, Libya. *Annals of Tropical Medicine and Parasitology* **100**, 55-62.

Ellegren H (2000a) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics* **24**, 400-402.

Ellegren H (2000b) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551-558.

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: Theoretical and empirical considerations. In: *Advances in Molecular Ecology* (ed. Carvalho GR), pp. 55-86. IOS Press; Ohmsha, Amsterdam; Washington, D.C.

Estoup A, Garnery L, Solignac M, Cornuet JM (1995a) microsatellite variation in honey bee (*Apis mellifera* L) populations: Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* **140**, 679-695.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**, 1591-1604.

Estoup A, Rousset F, Michalakis Y*, et al.* (1998) Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). *Molecular Ecology* **7**, 339-353.

Estoup A, Tailliez C, Cornuet JM, Solignac M (1995b) Size homoplasy and mutational processes of interrupted microsatellites in 2 bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Molecular Biology and Evolution* **12**, 1074-1084.

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics* **7**, 745-758.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Farris JS (1969) A successive approximations approach to character weighting.

Systematic Zoology **18**, 374-380.

Felsenstein J (1981a) A likelihood approach to character weighting and what it tells us about parsimony and compatability. *Biological Journal of the Linnean Society* **16**, 183-196.

Felsenstein J (1981b) Evolutionary trees from gene frequencies and quantitative characters: Finding maximum-likelihood estimates. *Evolution* **35**, 1229-1242.

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

Felsenstein J (2004) *Inferring Phylogenies* Sinauer Associates, Inc., Sunderland, MA.

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. *Science* **155**, 279-83.

Fotedar R, Banerjee U (1992) Nosocomial fungal infections: Study of the possible role of cockroaches (*Blattella germanica*) as vectors. *Acta Tropica* **50**, 339-343.

Fotedar R, Shriniwas UB, Verma A (1991) Cockroaches (*Blattella germanica*) as carriers of microorganisms of medical importance in hospitals. *Epidemiology and Infection* **107**, 181-187.

Fraser DJ, Hansen MM, Ostergaard S, *et al.* (2007) Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Molecular Ecology* **16**, 3866-3889.

Gaggiotti OE, Lange O, Rassmann K, Gliddon C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* **8**, 1513-1520.

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. *Genome Research* **6**, 211-217.

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution* **12**, 594-603.

Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA fingerprints. *Nature* **318**, 577-579.

Goldstein DB, Linares AR, Cavallisforza LL, Feldman MW (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 6723-6727.

Goldstein DB, Pollock DD (1997) Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* **88**, 335-342.

Goodman SJ (1997) R-ST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* **6**, 881-885.

Gore JC, Schal C (2007) Cockroach allergen biology and mitigation in the indoor environment. *Annual Review of Entomology* **52**, 439-463.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Goudet J, Raymond M, deMeeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics* **144**, 1933-1940.

Graczyk TK, Knight R, Tamang L (2005) Mechanical transmission of human protozoan parasites by insects. *Clinical Microbiology Reviews* **18**, 128-132.

Hampson BC, Steiner WWM (1982) An electrophoretic analysis of population structure and gene diversity in the German cockroach. In: *Recent Developments in the Genetics of Insect Disease Vectors*, pp. 648-663. Stipes Publishing, Champaign, IL.

Hardy GH (1908) Mendelian proportions in a mixed population. *Science* **28**, 49-50.

Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. *Genetics* **163**, 1467-1482.

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

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution* **16**, 700-706.

Harris H (1966) Enzyme polymorphisms in man. *Proceedings of the Royal Society Series B-Biological Sciences* **164**, 298-306.

Hedrick PW (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313-318.

Hedrick PW (2005a) A standardized genetic differentiation measure. *Evolution* **59**, 1633-1638.

Hedrick PW (2005b) *Genetics of Populations*, 3rd edn. Jones and Bartlett Publishers, Boston.

Hey J, Machado CA (2003) The study of structured populations - New hope for a difficult and divided science. *Nature Reviews Genetics* **4**, 535-543.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2785-2790.

Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ (1992) Experimental phylogenetics: Generation of a known phylogeny. *Science* **255**, 589-592.

Huang QY, Xu FH, Shen H*, et al.* (2002) Mutation patterns at dinucleotide microsatellite loci in humans. *American Journal of Human Genetics* **70**, 625-634.

Hubby JL, Lewontin RC (1966) A molecular approach to the study of genic heterozygosity in natural populations. I. Number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**, 577-590.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**, 1898-1914.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* **11**, 424-429.

Jeffreys AJ, Wilson V, Thein SL (1985a) Hypervariable minisatellite regions in human DNA. *Nature* **314**, 67-73.

Jeffreys AJ, Wilson V, Thein SL (1985b) Individual-specific fingerprints of human DNA. *Nature* **316**, 76-79.

Jobet E, Durand P, Langand J*, et al.* (2000) Comparative genetic diversity of parasites and their hosts: population structure of an urban cockroach and its haplo-diploid parasite (oxyuroid nematode). *Molecular Ecology* **9**, 481-486.

Keil CB (1981) Structure and estimation of shipboard German cockroach (*Blattella germanica* (Dictyoptera, Blattellidae) populations. *Environmental Entomology* **10**, 534-542.

Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research* **11**, 247-257.

Kimura M, Crow JF (1964) the number of alleles that can be maintained in a finite population. *Genetics* **49**, 725-37.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561-70.

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

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

Koehler PG, Patterson RS, Owens JM (1995) Chemical systems approach to pest control. In: *Understanding and Controlling the German Cockroach* (eds. Rust MK, Owens JM, Reierson DA). Oxford University Press, New York.

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10774-10778.

Kuhner MK, Felsenstein J (1994) Simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* **11**, 459-468.

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* **9**, 299-306.

Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595-610.

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128-2129.

Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12**, 228-237.

Manel S, Berthier P, Luikart G (2002) Detecting wildlife poaching: Identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conservation Biology* **16**, 650-659.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions techniques with appropriate. *Trends in Ecology & Evolution* **20**, 136-142.

Mantel N (1967) Detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-215.

Marjoram P, Tavare S (2006) Modern computational approaches for analysing molecular genetic variation data. *Nature Reviews Genetics* **7**, 759-770.

McIntyre NE (2000) Ecology of urban arthropods: A review and a call to action. *Annals of the Entomological Society of America* **93**, 825-835.

McKinney ML (2006) Urbanization as a major cause of biotic homogenization. *Biological Conservation* **127**, 247-260.

Morgan WJ, Crain EF, Gruchalla RS*, et al.* (2004) Results of a home-based environmental intervention among urban children with asthma. *New England Journal of Medicine* **351**, 1068-1080.

Mukha DV, Kagramanova AS, Lazebnaya IV*, et al.* (2007) Intraspecific variation and population structure of the German cockroach, *Blattella germanica*, revealed with RFLP analysis of the non-transcribed spacer region of ribosomal DNA. *Medical and Veterinary Entomology* **21**, 132-140.

Mullis K, Faloona F, Scharf S*, et al.* (1986) Specific enzymatic amplification of DNA invitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51**, 263-273.

Nagaraja, Nagaraju J, Ranganath HA (2004) Molecular phylogeny of the nasuta subgroup of *Drosophila* based on 12S rRNA, 16S rRNA and CoI mlitochondrial genes, RAPD and ISSR polymorphisms. *Genes & Genetic Systems* **79**, 293-299.

Nauta MJ, Weissing FJ (1996) Constraints on Allele size at microsatellite loci: Implications for genetic differentiation. *Genetics* **143**, 1021-1032.

Nei M (1972) Genetic distance between populations. *American Naturalist* **106**, 283-290.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 3321-3323.

Nei M (1986) Definition and estimation of fixation indexes. *Evolution* **40**, 643-645.

Nei M (1987) *Molecular evolutionary genetics* Columbia University Press, New York.

Nei M, Maruyama T, Chakraborty R (1975) Bottleneck effect and genetic variability in populations. *Evolution* **29**, 1-10.

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

170.

Nei M, Takezaki N (1983) Estimation of genetic distances and phylogenetic trees from DNA analysis **21**, 405-412.

Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Annual Review of Ecology and Systematics* **28**, 105-128.

Neigel JE (2002) Is F-ST obsolete? *Conservation Genetics* **3**, 167-173.

Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: *Evolutionary Processes and Theory* (eds. Nevo E, Karlin S), pp. 515-534. Academic Press, New York.

Nielsen R (1998) Maximum likelihood estimation of population divergence times and population phylogenies under the infinite sites model. *Theoretical Population Biology* **53**, 143-151.

Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* **154**, 931-942.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**, 885-896.

Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10745-10749.

Owens JM, Bennett GW (1982) German cockroach (Orthoptera, Blattellidae) movement within and between urban apartments. *Journal of Economic Entomology* **75**, 570-573.

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* **147**, 1943-1957.

Pai HH, Chen WC, Peng CF (2005) Isolation of bacteria with antibiotic resistance from household cockroaches (Periplaneta americana and *Blattella germanica*). *Acta Tropica* **93**, 259-265.

Pai HH, Ko YC, Chen ER (2003) Cockroaches (*Periplaneta americana* and *Blattella germanica*) as potential mechanical disseminators of *Entamoeba histolytica*. *Acta Tropica* **87**, 355-359.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**, 568-583.

Park SDE (2001) *Trypanotolerance in West African cattle and the population genetic effects of selection*, University of Dublin.

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA (1998) What molecules can tell us about populations: Choosing and using a molecular marker. *Ecology* **79**, 361-382.

Pearse DE, Crandall KA (2004) Beyond F-ST: Analysis of population genetic data for conservation. *Conservation Genetics* **5**, 585-602.

Piry S, Alapetite A, Cornuet JM*, et al.* (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536-539.

Pocock MJO, Hauffe HC, Searle JB (2005) Dispersal in house mice. *Biological Journal of the Linnean Society* **84**, 565-583.

Pollock DD, Bergman A, Feldman MW, Goldstein DB (1998) Microsatellite behavior with range constraints: Parameter estimation and improved distances for use in phylogenetic reconstruction. *Theoretical Population Biology* **53**, 256-271.

Primmer CR, Ellegren H, Saino N, Moller AP (1996) Directional evolution in germline microsatellite mutations. *Nature Genetics* **13**, 391-393.

Primmer CR, Koskinen MT, Piironen J (2000) The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267**, 1699-1704.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9197-9201.

Raymond M, Rousset F (1995a) An exact test for population differentiation. *Evolution* **49**, 1280-1283.

Raymond M, Rousset F (1995b) Genepop (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* **105**, 767-779.

Rivault C (1990) Distribution dynamics of *Blattella germanica* in a closed urban environment. *Entomologia Experimentalis Et Applicata* **57**, 85-91.

Rivault C, Cloarec A (1991) Exploitation of food resources by the German cockroach, *Blattella germanica*, in an urban habitat. *Entomologia Experimentalis Et Applicata* **61**, 149-158.

Robinson WH, Akers RC, Powell PK (1980) German cockroaches in urban apartment buildings. *Pest Control* **48**, 18-20.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3**, 380-390.

Rosenstreich DL, Eggleston P, Kattan M*, et al.* (1997) The role of cockroach allergy and exposure to cockroach allergen in causing morbidity among inner-city children with asthma. *New England Journal of Medicine* **336**, 1356-1363.

Roth LM (1985) A taxonomic revision of the genus *Blattella* Caudell (Dictyoptera, Blattaria, Blattellidae). *Entomologica Scandinavica*, 1-221.

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* **142**, 1357-1362.

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

Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Runstrom ES, Bennett GW (1984) Movement of German cockroaches (Orthoptera, Blattellidae) as influenced by structural features of low-income apartments. *Journal of Economic Entomology* **77**, 407-411.

Runstrom ES, Bennett GW (1990) Distribution and movement patterns of German cockroaches (Dictyoptera, Blattellidae) within apartment buildings. *Journal of Medical Entomology* **27**, 515-518.

Saitou N, Nei M (1987) The Neighbor-joining method: A new method for constructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463-5467.

Schäfer MA, Hille A, Uhl GB (2001) Geographical patterns of genetic subdivision in the cellar spider *Pholcus phalangioides* (Araneae). *Heredity* **86**, 94-102.

Schal C (1988) Relation among efficacy of insecticides, resistance levels, and sanitation in control of the German cockroach (Dictyoptera, Blattellidae). *Journal of Economic Entomology* **81**, 536-544.

Schal C, Hamilton RL (1990) Integrated suppression of synanthropic cockroaches. *Annual Review of Entomology* **35**, 521-551.

Schierwater B, Ender A (1993) Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Research* **21**, 4647-4648.

Schlötterer C (2004) The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics* **5**, 63-69.

Sever ML, Arbes SJ, Gore JC*, et al.* (2007) Cockroach allergen reduction by cockroach control alone in low-income urban homes: A randomized control trial. *Journal of Allergy and Clinical Immunology* **120**, 849-855.

Silverman J (1986) Adult German cockroach (Orthoptera, Blattellidae) feeding and drinking behavior as a function of density and harborage-to-resource distance. *Environmental Entomology* **15**, 198-204.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**, 264-279.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457-462.

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* **38**, 1409-1438.

Suarez AV, Holway DA, Case TJ (2001) Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from Argentine ants. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 1095-1100.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**, 199-203.

Sutherland GR, Richards RI (1995) Simple tandem DNA repeats and human genetic disease. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 3636-3641.

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

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

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

Takezaki N, Nei M (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics* **178**, 385-392.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**, 6463-6471.

Templeton AR, Sing CF, Kessling A, Humphries S (1988) A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. *Genetics* **120**, 1145-1154.

Vos P, Hogers R, Bleeker M*, et al.* (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-4414.

Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**, 995-1005.

Wahlund S (1928) Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* **11**, 65-106.

Wang JL, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* **163**, 429-446.

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379-391.

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

Weber JL, May PE (1989) An abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**, 388-396.

Weinberg W (1908) Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* **64**, 368-382.

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

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**, 7213-7218.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). *Heredity* **82**, 117-125.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531-6535.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 0097-0159.

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

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

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

Xiong B, Kocher TD (1991) Comparison of mithochondrial DNA sequences of 7 morphospecies of black flies (Diptera, Simuliidae). *Genome* **34**, 306-311.

Xu X, Peng M, Fang Z, Xu XP (2000) The direction of microsatellite mutations is dependent upon allele length. *Nature Genetics* **24**, 396-399.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**, 563-584.