4. Adequate sample sizes for accurate detection of population subdivision: a simulation based exploration of summary statistics

Abstract

Accurate estimates of genetic population subdivision are of importance, not only for the understanding of population dynamics in biological systems, but also for identification of unique populations that need to be conserved. Various factors affect the accuracy of these estimations including the sampling scheme employed. Here simulation modelling of microsatellite loci in diploid individuals for a structured population has been used to estimate the optimal number of loci and individuals necessary for accurate estimation of population subdivision. A surprising finding is that more individuals and loci need to be sampled in smaller populations than for larger populations to obtain the same accuracy. I also show that the actual genetic structure of the population does not affect the accuracy of the estimation except when there are very low levels of population subdivision. In these cases fewer individuals are necessary to obtain the same level of accuracy. Lastly, sampling different numbers of loci and individuals showed that a few of each might be adequate to accurately determine population subdivision.

Introduction

Genetic population structure inferences have developed significantly in recent years. It is relatively inexpensive to develop molecular markers for most organisms and obtain data to be used for assigning genetic population structure (Sunnucks, 2000). In addition, analysis of molecular data has become easier through the development of programs that cluster a population into various subpopulations or that test differentiation amongst predefined subpopulations (Luikart & England, 1999; Pearse & Crandall, 2004). Notwithstanding the development of molecular and analysis tools, there is still a shortage of user-friendly experimental design programs. Using the wrong experimental design results in an excess of time and money spent on obtaining redundant samples or genotyping a surplus of loci per sample while the accuracy of the population inferences remains unchanged. Additionally, obtaining high quality DNA often requires invasive sampling of wild animals, while both invasive and non-invasive sampling are labour intensive (Taberlet et al., 1999). Accurate estimation of population structure using the minimum number of individuals and loci is not only beneficial in reducing the cost of excessive data collection but is also important in conservation genetics as these estimates are often used in management and conservation strategies.

Population structure and gene flow estimated from molecular data are routinely used to identify distinct subpopulations (Palsbøll et al., 2006; Waples, 1998; Waples & Gaggiotti, 2006). A classical and often used method to detect population structure is through the use of summary statistics, such as Fst. Fst, as defined by Weir and Cockerham (1984). Although more powerful methods of estimating population substructure are available, summary statistics are still often used as initial indicators of structure, to supplement other indicators or even as a comparative benchmark for other methods (Balloux & Lugon-Moulin, 2002; Neigel, 2002; Pearse & Crandall, 2004). Previous studies found that more loci, rather than more individuals, will increase the accuracy of the estimated population structure (Felsenstein, 2006; Pluzhnikov & Donnelly, 1996). Both Felsenstein's (2006) and Pluzhnikov & Donnelly's (1996) studies used sequence data and showed that 8 individuals (long sequence reads of only one locus) were sufficient for accurate population structure estimation. It is ideal to provide the minimum requirements (i.e. the least number of

loci and individuals necessary, to accurately detect the presence or absence of population structure). These minimum sampling requirements would therefore indicate the optimal sampling strategy by reducing cost and effort in obtaining and typing individuals. They are also likely to be sufficient for more advanced methods (Luikart & England, 1999; Pearse & Crandall, 2004).

I developed the program POPSTAT, which uses individual-based simulation modelling to construct a population from user specifications. The unique feature of POPSTAT, however, is the implementation of repeated sampling from the simulated populations. From the simulated population the repeated sampling schemes may then be tested by the user determine the number of molecular markers and individuals needed to obtain an accurate estimate of population subdivision. In this chapter I describe how different sampling schemes from simulated populations affects the detection of accurate population structure and provide general guidelines for developing a sampling schemes. I found that simulations of different population sizes show that more individuals, or more loci, are necessary for accurate population structure estimates in small populations compared to large populations. This finding is counterintuitive and it is very important to devise appropriate sampling schemes for small populations, as these are often required for conservation purposes of endangered species. In addition there is a trade-off between the minimum number of individuals and the minimum number of loci necessary to obtain accurate estimates of population structure.

General conclusions on optimal sampling schemes, from this study, do not necessarily apply equally to all populations. It would therefore be advisable to estimate adequate sample sizes individually for each study population. The program POPSTAT (version 2) is freely available for this purpose at: http://www.bi.up.ac.za/software/popstatwin.zip (also available in the electronic appendix).

Materials and methods

Model description

The program POPSTAT (version2) was designed to investigate how different sampling schemes affect accurate detections of population structure from two subpopulations. The program is able to simulate the subpopulations as well as importing data from populations simulated by other applications. A detailed description the program can be found in the POPSTAT help file, available in appendix A. However, a brief description of key of the features of the program follows.

The model has two hierarchical levels, namely individual and population. Each individual is diploid and is therefore characterised by two arrays of loci variables, and also an identity number. All loci are unlinked and the user defines the number of loci. The program is a forward-time simulation program that simulates a single, panmictic population for a user-defined number of generations. The population is then separated into two subpopulations and simulated for a number of generations that is also defined by the user (it should be noted that the population homogeneity can be evaluated before it is divided into two subpopulations and the simulation extended to reach equilibrium if necessary, see also appendix A). The population has the following variables that can be defined by the user: population size of each subpopulation, migration rate and mutation rate.

After the simulation of two subpopulations, different sampling schemes can be employed, which allows the user to vary the number of individuals sampled as well as the number of loci used. From these sampling schemes the Gst (Nei, 1987), Fst (Weir & Cockerham, 1984) and Rst (Slatkin, 1995) are calculated. The sampling from each subpopulation is completely random. Any sampling scheme can automatically be repeated from which the accuracy of the summary statistics for that population may be determined.

General parameters

As the aim of this study is to investigate the number of individuals and loci necessary to accurately detect population structure between subpopulations, it is important to create populations with a known structure. For this reason all the individuals and all the loci were used in calculation of the summary statistics after every run. In all the simulations the number of generations since the population split were varied to obtain the desired level of population structure. However, required levels of population structure may also be obtained by changing the migration or mutation rate. In this study all the simulations were performed with no migrants between the two subpopulations, a mutation rate of 0.0001 per gamete per locus per generation (all mutations were stepwise) and 50 loci for each individual in the population. In some cases sampling per subpopulation may indicate a non-integer number of individuals sampled (i.e. 32.5 individuals per subpopulation sampled). This means that a total of 65 individuals were sampled from both populations and one population (chosen at random) contributed 32 while the other contributed 33 individuals to the whole sample.

Experiments

Experiment 4.1: Simulation of different Fst values

To obtain a dataset with increasing Fst values the following simulation parameters were used. A single, random mating population of 20,000 individuals was divided after 5,000 generations into two equal subpopulations and simulated for another 10,000 generations. The subpopulations were sampled at the 5,000-generation interval and at each following 500 generations thereafter.

The following sampling schemes were used: each time 500 re-samplings of 10, 32.5, 55, 77.5 individuals per subpopulation were performed. For each sample the Fst for the population was calculated (using all 50 loci in each case). For the different sample sizes the 500 calculated Fst values were used to determine the standard

deviation at each sample size. In addition, to obtain the actual genetic differentiation for comparison, all the individuals in the two subpopulations were sampled at every 500-generation interval and the actual Fst, Gst and Rst calculated.

Experiment 4.2: Simulation of different population sizes

To investigate the effect of population size on the accuracy of Fst estimation populations of different sizes were simulated. The details of the parameters used for each simulation are given in table 4.1 as well as the Fst calculated using all the individuals and all the loci.

Although the Fst value itself should not have an effect on the accuracy of estimation (see Discussion), all the populations were simulated until a similar Fst value was reached. An exception was however the largest population (population size 100,000) simulated to a final Fst of only 0.2 due to computing time constraints. A high level of subdivision was chosen to reduce the effects produced on the standard deviation when most of the sampled Fst values are 0.

The following sampling scheme was used for all the different population sizes simulated: 500 re-samplings of 5 individuals per subpopulation, increasing stepwise with 5 individuals up to 100 individuals at the final generation. The Fst per sampling was calculated separately for an increasing number of loci from 1 to 50. A standard deviation of the Fst was calculated for each of the 500 repeats per number of individuals sampled per number of loci used. To obtain the actual genetic differentiation, all the individuals were sampled, and all the loci used to calculate the final Fst of the population (table 4.1).

Experiment 4.3: Simulation of a single population for estimates of optimal sampling

A final simulation was performed to demonstrate the effect of number of individuals and loci number sampled on the accuracy of the Fst measure. A total population size of 5000 individuals was simulated as one population for 5000 generations and then as two equal subpopulations for another 1000 generations. The same sampling scheme

Table 4.1. Parameters used for each simulated population in experiment 4.2. The Fst value for all the simulations at the separation of the single population into two sub populations was 0. Final Fst was calculated using all individuals and loci.

Population size	**Number of generations**		**Final Fst**
(total)	**Before separation**	**After separation**	
1000	5000	100	0.323
2000	5000	200	0.334
10000	5000	1200	0.326
20000	10000	10000	0.322
100000	15000	50000	0.200

as for the 5 populations from experiment 4.2 was used. The final Fst for this population was 0.428.

Results

Experiment 4.1: Effect of the Fst values on the standard deviation of the Fst

The actual Fst value of the population did not have an influence on the standard deviation of the Fst (figure 4.1). An exception of course was when the Fst is 0 or very small (figure 4.1). The Fst values (as well as the Gst and Rst) increased in successive generations between the two separated subpopulations (figure 4.2). This can be seen at generation 5000 where the population was still un-separated and the Fst was 0, and to some degree at generation 5500 (Fst 0.0933, compare figures 4.1 and 4.2). The Fst values for the repeated samplings were not normally distributed at very low levels of population structure and as most values were 0 or close to 0, the standard deviation was also lower. A similar effect will be observable at Fst values that are 1 or very close to it.

It could also be observed that there was less variation in the standard deviation of larger sample sizes as well as a lower standard deviation in general than for small sample sizes (figure 4.1).

From figure 4.2 it was interesting to note that there was a marked difference between the different summary statistics as the two subpopulations became more different. Both the Gst and Fst plateau earlier and at a lower level than the Rst even after long separation times. All the summary statistics at their respective plateaus in this experiment were however indicative of genetic structure in the subpopulations.

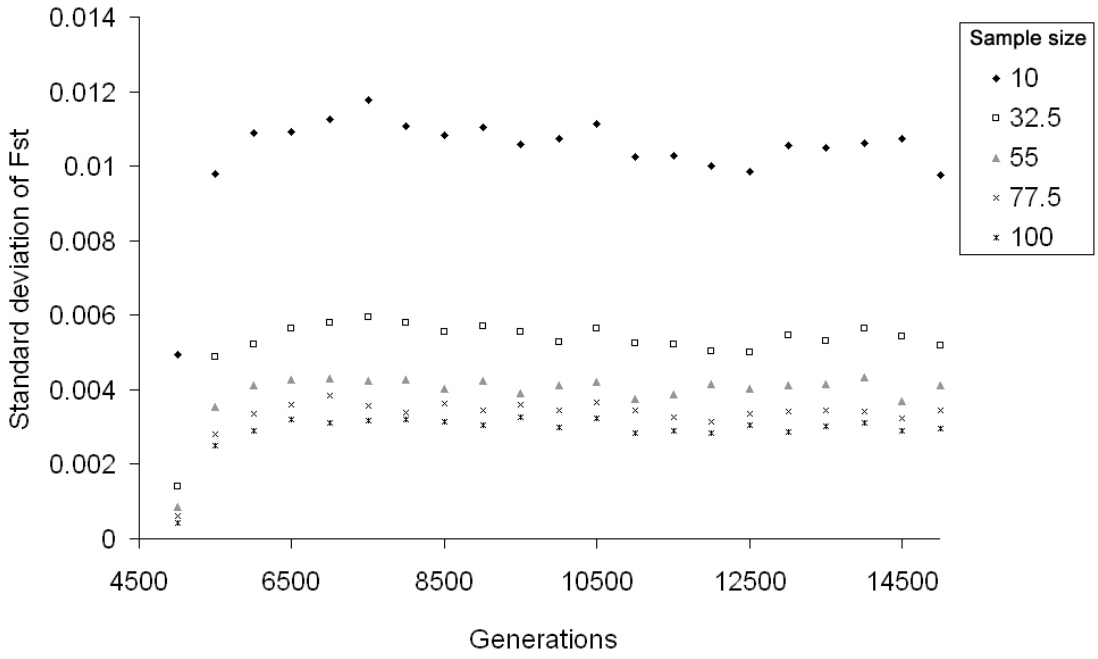

Figure 4.1. Standard deviation of Fst over a number of generations (calculated using repeated sampling of different sample sizes, from 10 to 100). It can be seen that the variation over generations was very small except at generation 5000, where the populations were separated (also figure 4.2 for actual Fst values).

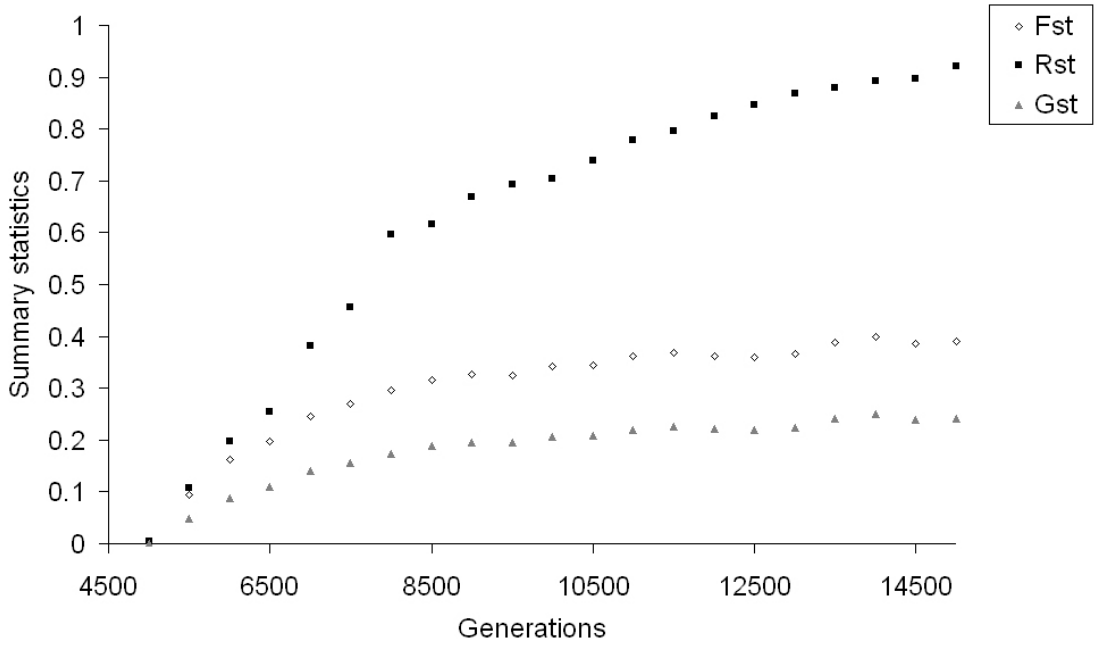

Figure 4.2. The increase in summary statistics when a population was separated into two subpopulations. The summary statistics were calculated using all the loci from all the individuals at different generations and a clear increase in the population structure can be seen.

Experiment 4.2: Effect of population size on the standard deviation of the Fst

The standard deviation of the Fst decreased as the population size increased, as observed in figure 4.3 (reported only a subset of the data, i.e. for 5, 10, 25 and 50 loci). This trend was observed at all sample sizes and different numbers of loci genotyped except in a few cases when only 1 or 2 loci were used in the determination of the Fst value. From this we can see that accurately estimating the Fst in large populations require smaller sample sizes, while small populations require larger sample sizes to obtain the same accuracy (figure 4.3).

Experiment 4.3: Effect of sample size vs. number of loci on the accuracy of the Fst

The accuracy of the estimated Fst increased with an increase in either the sample size or number of loci used (figure 4.4). It is clear that every additional locus used in the determination of the Fst decreased the standard deviation of the Fst and that this decrease was larger when sample sizes were very small (figure 4.4). Similarly, a large decrease in the standard deviation of the Fst was seen with every additional sample used, with a larger decrease when only a few loci were typed (figure 4.4). Using the standard deviation of the Fst, plotted against the number of loci and individuals sampled (as in figure 4.4) provided a quick and easy way to determine the optimal sampling scheme. For example, if an adequate level of accuracy is where the standard deviation of the Fst is less than 0.02 (figure 4.4) then for 20 loci typed, one can read from the graph that 15 individuals per subpopulation would be sufficient to get that accuracy.

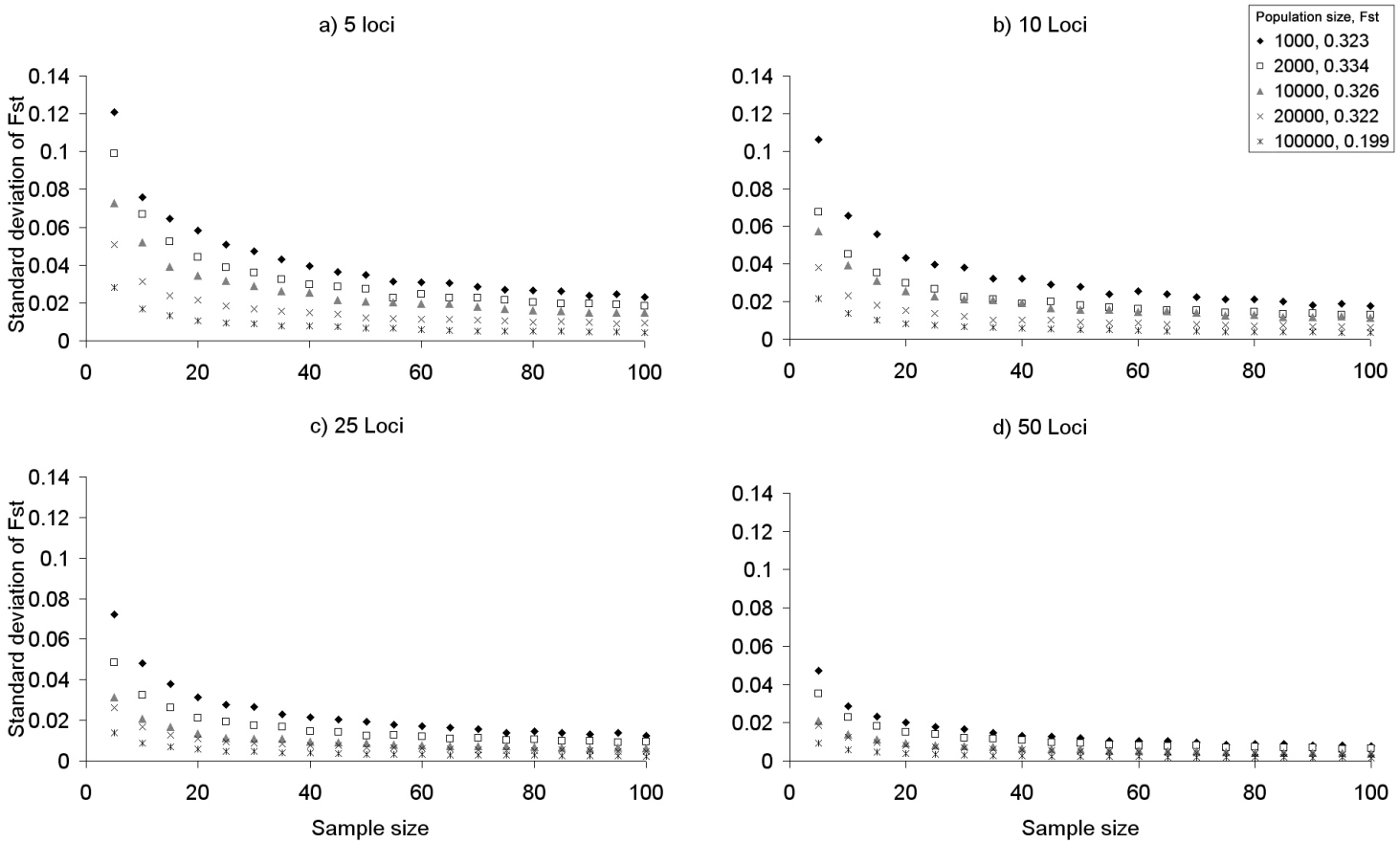

Figure 4.3. Decrease in standard deviation of the Fst as the number of individuals sampled increases for an increasing number of loci sampled for different sized populations (subpopulation size indicated in the legend from 1000 to 10000 with the Fst of the population when all individuals and loci are used). Larger populations have a lower standard deviation for all combinations of loci and individuals sampled.

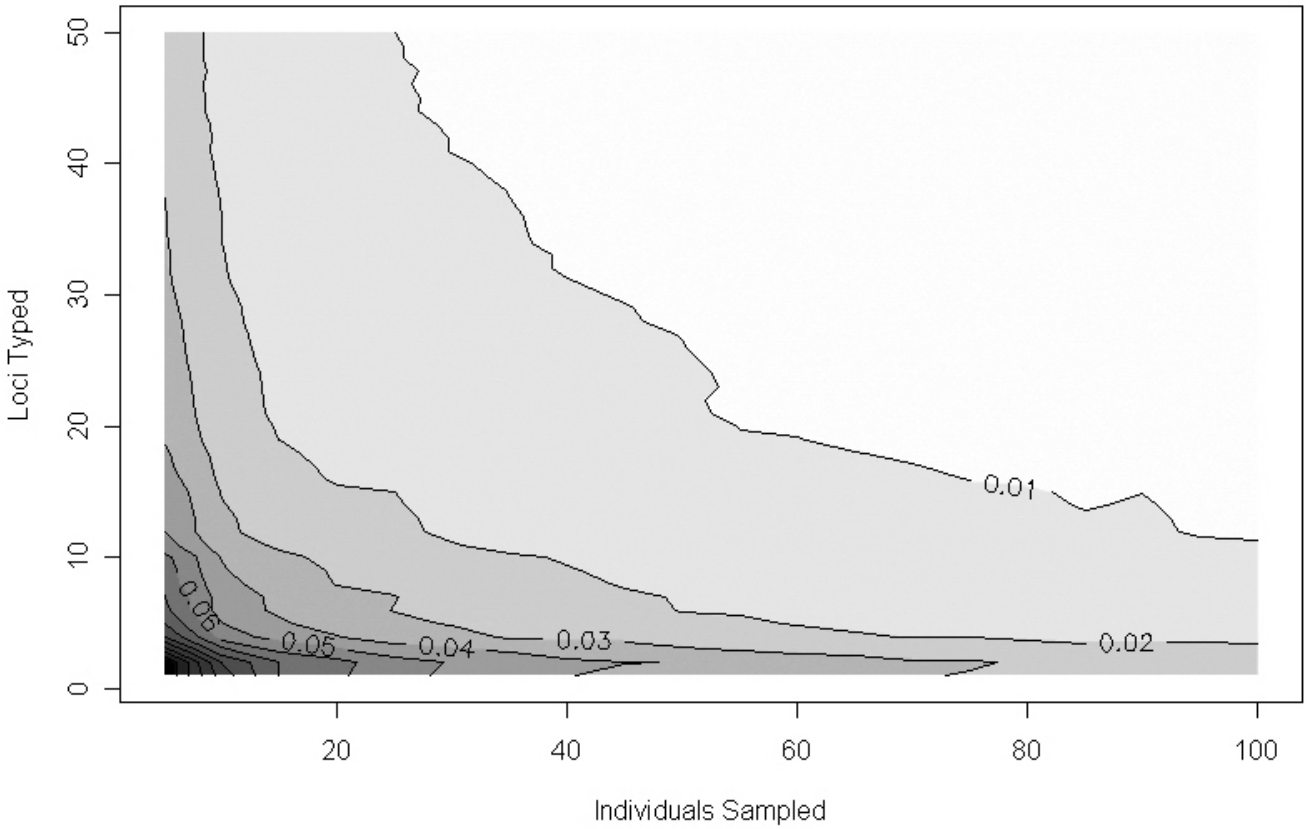

Figure 4.4. Contour plot of the standard deviation of the Fst. The standard deviation decreases when either the number of loci genotyped or the number of individuals sampled increases.

Discussion

Investigation of the simulated data led to inferences of four factors that may affect the accuracy of the Fst value. First, smaller populations require larger sample sizes to have the same accuracy as large populations. Second, the degree of structure itself (or the actual value of the Fst) does not seem to influence the accuracy of the inferred structure. The last two factors are the number of loci and the number of individuals sampled. Increasing both these factors increases the accuracy of the estimated population structure.

Effect of population size on accuracy

A general trend in statistics is to increase sample size as the study population size increases. However, from figure 4.3 we see that the standard deviation of the Fst is smaller in large populations at the same sample size used for small populations. This leads to the counterintuitive argument that we need fewer individuals from large populations to accurately detect population structure.

The reason for the inverse in adequate sample size vs. population size is because there is a larger probability of sampling unique genotypes when the population and sample size is small (see also Paetkau et al., 2004). Because big subpopulations have many individuals, who we expect to be randomly distributed, average genotypes are expected to be sampled more often. If this is not the case the subpopulation may need to be sub-structured even further and the sampling scheme re-evaluated. The resampling experiments in this chapter support this result (figure 4.3). This has the following implications for once-off sampling schemes in field experiments: small populations need more samples than large populations to get an accurate estimate of population substructure.

In the simulations the subpopulations were always equal in size, but the same reasoning as above can provide some guidelines to unequal subpopulations. Sampling from a small subpopulation may provide a skewed view of the uniqueness of the subpopulation, and enough samples should be obtained to circumvent this. The same

number of individuals, sampled from the larger subpopulation (assuming the large subpopulation to be truly panmictic), should then be sufficient to accurately detect population subdivision.

All the arguments used for sample sizes also hold for the number of loci used in calculating the Fst value, i.e. smaller populations need more loci per individual to reduce the bias that may arise from, by chance, sampling only the unique loci in the population.

Effect of the Fst on the accuracy

We do not expect the true population Fst to affect the accuracy of its estimation (figure 4.1). It is however important to verify this as general inferences on the sampling scheme may be affected if there is an effect. From figure 4.1 we see that the standard deviation of the Fst was lower only when the Fst value itself was still very small (also shown by Ryman et al., 2006). This effect was seen because the subpopulations were still very similar and even with repeated sampling the calculated Fst values will mostly be very small or 0. However, as soon as the Fst value approaches 0.1, the level at which the application STRUCTURE (Pritchard et al., 2000) can detect population subdivision (personal observation), the standard deviation stabilises.

The Fst value or actual subdivision of the population therefore has a number of implications on the sampling scheme. If the population is structured, large sample sizes will detect this structure, irrespective of the degree of structuring in the population. If, on the other hand, the population is not structured, fewer samples are necessary to accurately predict the Fst. It is, however, extremely important to have enough samples to accurately estimate high population subdivision even if the population is suspected to be unstructured, otherwise hidden structure may remain undetected.

Mention must be made of the different rate of change between some of the summary statistics as observed in figure 4.2. All three statistics are indicating population subdivision at the end of the simulation, even though the respective values are different. To put this in context however, the Gst and Fst values are both calculated through the genotype frequencies in the population while the Rst also takes

into account the actual genotype. This specifically refers to stepwise mutation of microsatellite loci, (Slatkin, 1995). The Rst values are therefore more sensitive to any changes in the genotypes and will increase quicker than frequency-based methods, such as Gst and Fst. Furthermore, the Fst is almost always double the Gst in all the simulations performed (personal observation). Inferences of population subdivision should be made by also taking into account the method used to detect this structure. For instance, if an Fst value of 0.3 will be used as an indication of strong population structure, the corresponding Rst value will be much higher. Rst, however has the disadvantage of being specifically applicable only to microsatellites under the generalised stepwise mutation model (Slatkin, 1995). It is therefore advisable to use various methods to infer population structure.

Effect of sample size and loci number on accuracy

When formulating a sampling scheme it is important to consider what would be a sufficient level of accuracy. There are various ways to test the accuracy of different statistical methods or sampling schemes. For this study I chose to use the standard deviation of the Fst from repeated sampling as a measurement of accuracy as it is straightforward to compare and can be readily determined from the data generated by the simulation output. The required level of accuracy of Fst may of course be vary between different species or different populations and sampling schemes with more accuracy can be devised using the same principles and software if necessary.

I find that an increase in either the number of loci or the number of individuals sampled will increase the accuracy of the Fst (figure 4.4). Figure 4.4 also reveals that after the rapid decrease in the Fst, when loci and sample numbers increase, there is a stabilization of the standard deviation. Once the standard deviation has stabilised in this way, increasing the number of loci or individuals would contribute very little to the accuracy of the estimates. The optimal number of individuals used in any sampling scheme is, however, mostly dependent on the number of loci used and *vice versa*. As these two factors are inversely correlated, both need to be estimated simultaneously as in figure 4.4. From this figure it is easy to determine the optimal number of individuals to sample once the loci to be used are developed to obtain

accurate estimates of population subdivision. It is however important to estimate this specifically for the different population sizes that will be sampled.

Guidelines on sampling

According to the results obtained from the simulations in this study, it is important to keep the following in mind when planning a sampling scheme. The program POPSTAT simulates only two populations from which sampling are performed and inferences made. In field studies there are often more than two subpopulations and to reconcile this with the simulation data, the two most diverged subpopulations need to be used for an estimation of the sample size and the number of loci that are necessary to obtain accurate results. Using the sub-populations that are assumed to be most divergent prevents sample size estimates based on an unstructured population, which is the only case where the actual Fst values have an effect on the accuracy of the estimates. Pair-wise comparison of subpopulations is routinely performed in population studies (Kitada et al., 2007) and there are a number of programs used to estimate pair-wise Fst values from genotypic data (Excoffier et al., 2005; Goudet, 1995; Raymond & Rousset, 1995). Starting with a complex simulation model with more than two subpopulations require more assumptions about the population such as differential migration rates and sub population sizes. Moreover the minimum adequate sample size required to differentiate between the most diverged subpopulations would be adequate to reveal any other substructure regardless of the complexity of the population structure. Once the minimum number of samples is determined to accurately detect population structure for the two most unrelated subpopulations, the same number of samples from each of the other assumed subpopulations should be obtained (see Goudet et al., 1996 on balanced sampling). This would provide accurate estimates for differentiation between all pair-wise comparisons of the subpopulations.

In conclusion, when sampling from different populations, the actual population structure does not play a role in the accuracy of the estimations, except when the population is almost completely unstructured. However, smaller sample sizes are required for these populations and the same sampling scheme as for more structured populations would therefore be sufficient. If the population is small, slightly more loci and individuals should be used in determining population structure. The number of

loci to be used depends on the number of individuals that would be sampled and *vice versa*. As a general rule (even for small populations, e.g. 5000 individuals), 20 loci (which are easily obtainable with current molecular tools) and 15 individuals per subpopulation are sufficient to accurately determine population subdivision. It is still extremely important to know the population biology of the study organism. This will determine the assumed population divergence (from isolation by distance or other environmental factors) that will be used in the simulation, the accuracy of simulations as well as the resultant sampling schemes.

5. Inbreeding depression does not promote mixed mating and dispersal in a male pollinating fig wasp, *Platyscapa awekei*

Abstract

Theory predicts that high levels of inbreeding depression or kin competition will promote the evolution of dispersal. Species with mixed mating systems (sibmating as well as random mating) provide strong evidence for the importance of inbreeding depression in determining the mating system. Using the pollinating wasps *Platyscapa awekei* as a model system we investigate the recent evolution of male dispersal in relation to their level of inbreeding depression. Previous work suggested that *P. awekei* suffers from inbreeding depression but possibly also from outbreeding depression. With a much larger sample size we show that *P. awekei* females have low levels of inbreeding depression. We used the program SibMate to determine if this low inbreeding depression could be sufficient to select for male dispersal and mixed mating. We find that the level of inbreeding depression alone is not severe enough in *P. awekei* to cause male dispersal and that other factors, such as the high proportion of male-only broods, kin competition and sex ratio, may be necessary to bring about the evolution of male dispersal.

Introduction

When an individual is faced with the option of mating with a relative, a number of factors play a role in the mating decision. Ultimately, the mating strategy depends on the number of genes (identical to your own), which are successfully transferred to the next generation. In nature, a multitude of different mating strategies are found (Keller & Waller, 2002; Thornhill, 1993) and include inbreeding, outbreeding or a strategy that lies somewhere in-between. Focusing on mixed mating, where individuals frequently outbreed with unrelated individuals, as well as, inbreed with close relatives, leads to the following question: What would cause individuals to stay close to their relatives and mate with them and at other times cause them to migrate and mate with non-relatives?

Due to the chronic inbreeding that occurs in many haplodiploid species (Godfray & Cook, 1997; Hardy, 1994; Werren, 1993), including pollinating fig wasps (Greeff, 2002; Greeff et al., 2003; Herre et al., 1997; Molbo et al., 2002; Molbo et al., 2004; Zavodna et al., 2005), alongside purging in the haploid males (Bruckner, 1978; Werren, 1993), it is expected that haplodiploids suffer less from inbreeding depression (Antolin, 1999; Henter, 2003; Werren, 1993). Despite this expectation, considerable inbreeding depression in haplodiploids has been observed (Antolin, 1999; Henter, 2003).

Inbreeding depression may lead to the evolution of male dispersal (Bengtsson, 1978; Gandon, 1999; Motro, 1991; Perrin & Mazalov, 2000; Ronce, 2007; Waser et al., 1986). Male dispersal, in turn, could lead to mixed mating, if intermediately related individuals are not available, and if the inbreeding depression is not too severe (chapter 2).

In addition to inbreeding depression, a number of other factors may induce male dispersal: Dispersal is often a solution to high levels of local mate competition (Clobert et al., 2001; Frank, 1986; Hamilton & May, 1977; Perrin & Mazalov, 2000; Ronce, 2007; Van Valen, 1971). Moore et al. (2006), have proposed this as the cause of dispersal in *Platyscapa awekei* males. It has also been shown that the sex ratio in *P. awekei* (which is less female biased than those of other pollinating wasps) could trigger male dispersal (Moore et al., 2006; Nelson & Greeff, 2009; but see also chapter 3).

The presence of mixed mating as a breeding strategy is difficult to explain as factors causing the inbreeding would normally oppose outbreeding and *vice versa*. In addition to factors that cause male dispersal, a number of factors that cause inbreeding have been identified. An obvious motive for inbreeding is outbreeding depression, which has been observed in *P. awekei* (Greeff et al., 2009). However, the study by Greeff et al. (2009) was based on a small sample and the results were somewhat ambiguous and failed to explain the mixed mating system. Additional factors that might reduce male dispersal and therefore mixed mating include the high cost of dispersal (Clobert et al., 2001; Frank, 1986; Gandon, 1999; Hamilton & May, 1977; Motro, 1983; Motro, 1991; Taylor, 1988) and the kin advantage of inbreeding (Bateson, 1983; Bengtsson, 1978; Fisher, 1941; Kokko & Ots, 2006; Parker, 1979; Pusey & Wolf, 1996; Waller, 1993; Waser et al., 1986; Wolf, 2000; see also chapter 2).

The life history of pollinating fig wasps provides us with an opportunity to test hypotheses that could explain the evolution of mixed mating. The relevant life history details, of the study species *P. awekei,* are as follows: A single or a few mated females enter receptive figs, lay their eggs and die. The larvae develop by feeding on the galled flowers inside the fig. It has been suggested that the success of development depends on the quality of the egg and gall, both of which may be determined by the mother's genotype (Greeff et al., 2009). After development, the males eclose first and mate with the females, who are receptive only whilst in their respective galls. *Platyscapa awekei* males regularly engage in contest competition (Greeff et al., 2003; Nelson, 2005) and dispersal to other figs (Greeff et al., 2003), often before the depletion of receptive females in their native fig (Moore et al., 2006). Once a female ecloses from her gall she leaves the fig in order to start the cycle anew. This species enable us to investigate three key aspects that may affect the mating strategy: the level of inbreeding depression, the kin advantage to sibmating, and the advantages to male dispersal.

The aim of this study is to determine if there is an optimum between inbreeding depression and outbreeding depression in *P. awekei,* which may lead to mixed mating (Bateson, 1983; Greeff et al., 2009; Price & Waser, 1979; chapter 2). This is done by: 1) Determining the effect of the mother's inbredness on her ability to produce mature offspring. We find, by using a larger sample than Greeff et al. (2009), that only inbreeding depression is present in *P. awekei*. 2) Using the program SibMate (chapter

2), we simulate the empirical results for *P. awekei* and predict the optimal mating strategy. Our results indicate that inbreeding depression is too low to cause outbreeding and male dispersal is probably not an adaptation to reduce inbreeding depression.

Materials and Methods

Sample collection

Sampling was performed at the National Botanical Gardens in Pretoria during the summer (February and December) over 2 years (table 5.1). We carried out single foundress introductions of the pollinating fig wasp *Platyscapa awekei* into their natural host *Ficus salicifolia.* Single foundress introductions ensure that the data were from a single mother. This permitted the reconstruction of the mother's genotype from her offspring. In addition, the clutch-sizes of each mother, which is equal to her total lifetime reproductive success, could be used as a good proxy for her fitness.

Material bags were placed around figs 2-3 weeks prior to becoming receptive for pollination. The bags prevented non-experimental pollinator females from entering the figs as they developed. Once the bagged figs became receptive single wasps were allowed to enter each fig after which the figs were again bagged and allowed to develop. Females used for introductions were obtained one day prior to introduction by collecting releasing figs (from the botanical gardens in Pretoria) and placing each fig in a single glass vial stopped with cotton wool. The next morning females released during the night were placed on a fig (only one female from each vial) and left to enter (manual introduction). Alternatively, bags were removed from the experimental figs as they became receptive and female wasps, arriving naturally, were allowed to enter the figs (one female per fig only, natural introduction). This occurred when trees with releasing figs were hard to obtain. Again bags were placed around the figs and left to develop. The type of introduction (manual or natural) for each tree is indicated in table 5.1.

Once the figs were almost ready to release, the bags were removed and traps were placed on each fig. The traps were constructed from eppendorf tubes with both

Table 5.1. Population source data with averages and standard deviation (SD) for the clutch size and number of underdeveloped individuals as well as the percentage of virgin females (i.e. male only broods) and average homozygosity per tree.

Population	Introduction date	Introduction method	# Genotypes reconstructed	Clutch size (SD)	Underdeveloped (SD)	% Unmated	Average loci homozygous
Tree 1	February 2006	manual	37	46.946 (10.724)	0.541 (1.070)	29.730	2.135
Tree 2	December 2006	manual	22	61.045 (9.105)	0.409 (1.333)	4.454	1.864
Tree 3	December 2006	natural	50	54.48 (13.380)	0.100 (0.303)	4.000	1.940
Tree 4	February 2007	natural	29	43.172 (12.048)	0.517 (1.122)	24.138	2.517
Tree 5	February 2007	mixed	42	51.095 (12.060)	0.952 (1.710)	0.000	2.571

ends removed. This formed a hollow, tapering, tube and a small piece of mesh material was melted to one side (figure 5.1a). A whole range of different sized traps was obtained, depending on where an eppendorf tube was cut (figure 5.1a), to suit the common fig size range of the host species. As the figs ripen, they expand, fitting tightly into the traps and none of the released offspring could escape (the species *P. awekei* only emerge through the ostiole and traps were placed to cover this area). Traps were checked once a day and the figs were picked if any offspring were observed. Each fig was left for two days to release while the offspring were continuously collected and placed in 96% ethanol. After 2 days the remaining wasps were dissected from each fig with the aid of a dissecting microscope. Some of the dissected individuals were considered as under-developed. The criteria for denoting an individual as under-developed were that these individuals were in the larval stage, had less developed features such as reduced wings or legs, or were, in general, too immature to live independent from their galls. Data on the clutch-size, the number of under-developed wasps and mating status of the mother were recorded for each fig (table 5.1). (Note that if only sons were produced the mother was assumed to be unmated as unfertilised eggs develop into males).

Genotype reconstruction

The mothers' genotypes were reconstructed from the genotypes of her offspring. At least 7 males were typed to reconstruct their mother's genotype. As the males are haploid they receive their genotype solely from their mother. If the allele for each male (per locus) was the same, the mother was deemed homozygous for that locus. When there were two alleles, the mother's genotype consisted of the two alleles observed in her sons. The probability of wrongly assigning a mother as homozygous when she is in fact heterozygous is $(1/2)^n$, where n is the number of males typed. The probability of a wrong reconstruction of the mother's genotype, using 7 males, is therefore 0.0078. In cases where clutches contained less than 7 males, females were used instead of the males. Here a mother was considered homozygous if all the genotyped daughters had at least one copy of the allele found in their brothers and provided the same confidence.

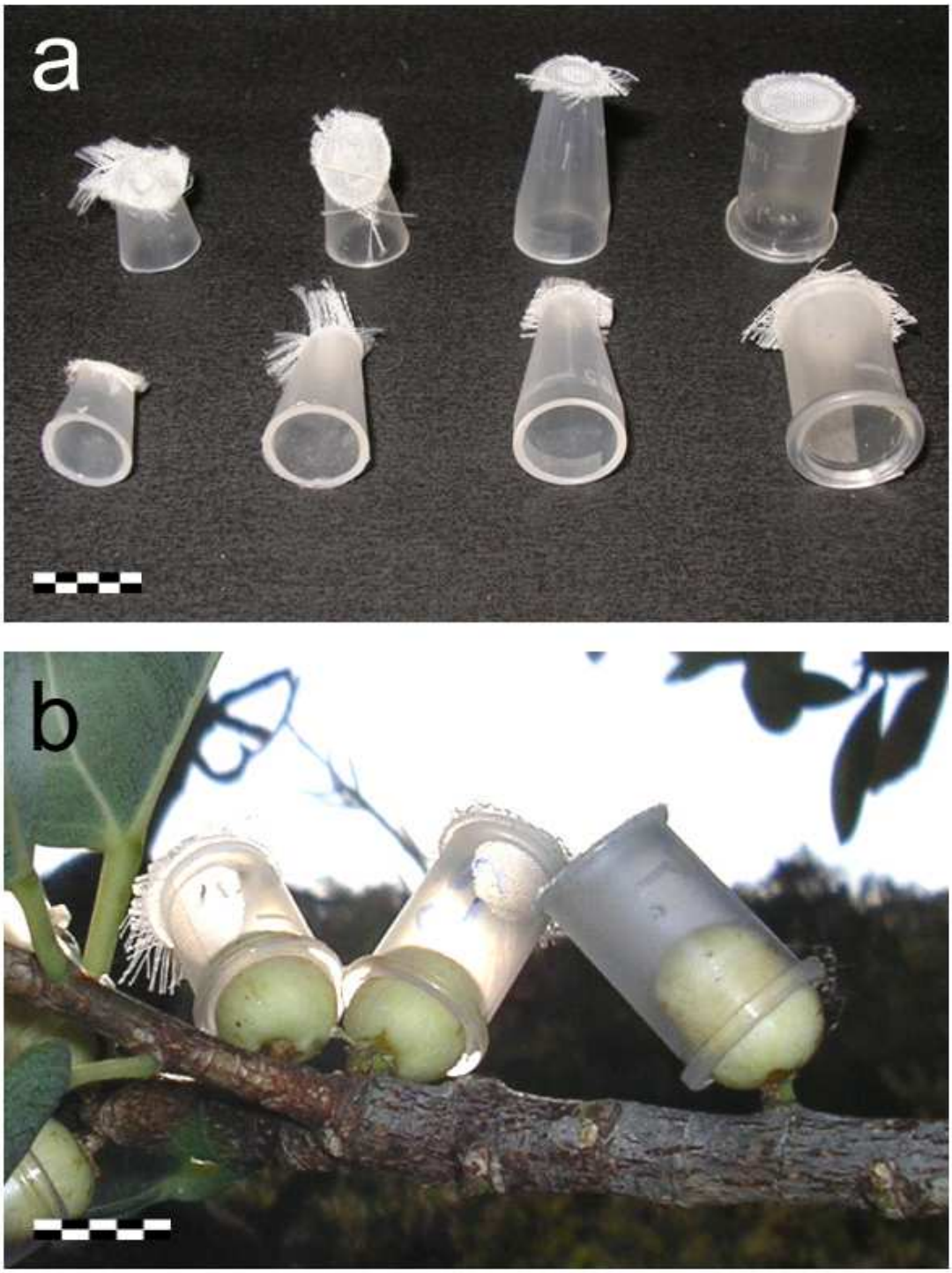

Figure 5.1. a) Range of traps, constructed from eppendorf tubes, used on different sized fruits. b) Traps fitted on figs of Tree 2. Scale = 1 cm.

Genotyping

DNA extractions of all the individuals to be genotyped were performed using the protocol as described by Estoup et al. (1996). Two males (from one clutch) or one female were placed in an eppendorf tube and grounded using a sterile pestle. 500µl Chelex (10% Chelex in Sabax water), at 60°C was added to the grounded wasps using a wide bore tip. The samples were incubated at 100°C for 15 minutes after which 7.5µl Proteinase K (Fermentas, 20mg/µl) was added. The samples were incubated for 1.5 hours at 55°C (shaken at 15-minute intervals). Incubating the samples at 100°C for 15 minutes stopped protein digestion. PCR reactions were performed directly from these solutions.

Each individual was genotyped at 6 polymorphic microsatellites designed for *P. awekei* (Jansen van Vuuren, 2005), using fluorescently labelled forward primers (table 5.2). (Note that, no null alleles were observed in this study or from a previous study where wasps were individually typed at the six microsatellite loci developed for this species (Jansen van Vuuren, 2005; Newman, 2007)). Table 5.3 provides the PCR conditions for the four reactions necessary to amplify the 6 microsatellite loci. Product amplification was confirmed on 1% agarose gels. PCR products were diluted 1:10 and mixed in the following ratios 2:2:1:1 from pool 1, pool 2, pool 3 and pool 4 respectively. LIZ^{TM} size standard was diluted with formamide (1:10). For each sample, 1µl of the pooled and diluted PCR product was added to 10µl of the formamide-size standard solution and denatured at 95˚C for 5 minutes. Samples were subsequently run on a Genetic Analyzer 3100 (Applied Biosystems) and results analysed with GeneMapper v3.5 (Applied Biosystems).

Table 5.2. Primer sequences, label colour, size and number of alleles of the 6 microsatellites loci used to determine foundress genotypes. Reaction pools indicate the combinations of primers amplified together, with their respective annealing temperatures.

Primer name	Primer sequence (5'-3')	Label colour	Size (bp)	# Alleles	Reaction pool	Annealing temperature
Pa 1	F: GTA GCG CCG TAT CAA ATT GCA A	Green	215-266	30	1	50
	R: GGG AGGG CTT GGG ATC TTT AAC GA					
Pa 4	F: GGG TGT TGT CGG TTT GTG AGA	Yellow	188-238	27	2	65
	R: GGC AAA CAT CCA TCG GAG TGA					
Pa 7	F: CTG CCG GTC AGA GGA GGA G	Blue	277-347	45	3	60
	R: TAT GAC GTC ATC GGT TTG GCA A					
Pa 8	F: GAG GAA GTC CGA TGA ATG AAC GA	Blue	190-225	17	3	
	R: GCG AAC AGG AGA CAA AGA CAG A					
Pa 21	F: GCT GTC GAG GCG AAA CAC A	Green	147-215	32	3	
	R: GCG CGA GGC ATT GGC AA					
Pa 32	F: CGG TGT TCA ATT GCC AAG TGA	Yellow	107-192	32	4	60
	R: TCG TGT TCT TCG TAA TCG CGT A					

Table 5.3. PCR conditions, and reagents. The denaturing, annealing and elongation steps were repeated 30 times. *quantity for each primer in the pool 3.

	Conditions			
Step	**Temperature ˚C**	**Time**		
Hotstart	95	10 min		
Denature	95	40 sec		
Annealing	see table 2	1 min		
Elongation	72	2 min		
Hold	4	-		
	Quantity			
Reagent	**Pool 1**	**Pool 2**	**Pool 3**	**Pool 4**
Genomic DNA	1 µl	0.5 µl	0.5 µl	0.5 µl
Buffer	1 mM	1 mM	1 mM	1 mM
$MgCl_2$	2 mM	1.8 mM	2 mM	1.6 mM
Primer	0.6 µl	0.3 µl	0.3 µl*	0.3 µl
dNTPs	0.16 mM	0.16 mM	0.16 mM	0.16 mM
Taq DNA Polymerase	0.1 u/µl	0.05 u/µl	0.05 u/µl	0.05 u/µl
Reaction volume	10 µl	10 µl	10 µl	10 µl

Statistics

All statistics were performed in R, version 2.4.1. Linear regressions were used to test the effect of the following variables on the clutch-size, mating status, number of under-developed wasps, tree, number of homozygous loci and number of homozygous loci squared. All models were reduced by deleting non-significant terms, starting with the interactions.

We also tested if any of the measured variables explained the absolute number under-developed individuals with a linear model. Additionally we tested the measured variables against the proportion of under-developed individuals per clutch with a generalised liner model with binomial distribution.

Lastly, we used the parameter estimated from *model 5.1* (see below) as input to the program SibMate (see chapter 2). We also determined the 95% confidence intervals for the *model 5.1* and derived the maximum and minimum inbreeding depression slopes as input. For each set of parameter estimates we ran 10 simulations to determine the optimal sibmating rate (in 5 simulations the females could decide to sibmate while in the other 5 the males could choose). It should be noted that each extra homozygous locus does not indicate an additional sibmating event directly. To be precise, more than a single locus extra should be homologous to indicate a sibmating event, and we assigned the fitness values in the simulations accordingly (Greeff et al., 2009). Table 5.4, contains the complete parameter sets of all the simulations.

Results

In total, 1260 individuals were successfully typed, and the genotypes of 180 foundress females reconstructed. The average clutch-size, percentage females unmated, average number of offspring under-developed and average heterozygosity per crop is summarised in table 5.1. From the table it is clear that many females are unmated and the average for the whole dataset is 13.207%.

Table 5.4. Parameters used for simulations. The maximum and minimum fitness values are from the 95% confidence intervals of *model 5.1*. Each parameter set was simulated 5 times. The number of homozygous loci for each successive sibmating was determined from Greeff et al. (2009).

Times Sibmated	Fitness values			# loci homozygous
	model 5.1	Maximum	Minimum	
0	55	59	51	0-1
1	55	59	51	2
2	53	57	50	3
3	52	54	50	4
4	52	54	50	4
5	50	52	49	5
6	50	52	49	5
7	50	52	49	5
8	50	52	49	5
≥ 9	49	49	49	6

General parameters

Ploidy	Haplodiploid
Population size	10000
Generations	10000
Mutation rate	0.0001
# alleles	101

All inbreeding was via sibmating

All simulations were repeated separately with both male and female choice

Clutch-size was best explained, with the following minimum adequate linear model, as a function of tree and the number of homozygous loci (see table 5.5, figure 5.2):

Clutch-size ~ tree + homozygous loci [*model 5.1*]

There was a significant difference in the number of under-developed offspring between the different trees (table 5.6), as well as the proportion of under-developed offspring per tree (table 5.7). However, none of the other variables or any of the interactions had an effect on the number or the proportion of under-developed offspring (deleted terms in tables 5.6 and 5.7).

Simulating the inbreeding depression from *model 5.1* and the maximum and minimum inbreeding depression from the 95% confidence intervals, indicated that sibmating would not evolve when males could decide the mating strategy (table 5.8). However, when the females are able to choose the mating strategy, low levels of sibmating (α) may evolve for the higher levels of inbreeding depression (the slope of *model 5.1* and the maximum inbreeding depression from the 95% confidence interval, table 5.8).

Discussion

We found that a mother's ability to produce mature offspring decreased significantly as she became more inbred (for every additional pair of homozygous loci 1.5 or 3% fewer offspring are produced). Our results also indicated that the host tree was a major determinant in the number of offspring produced by a mother, accounting for the largest part of the variation in *model 5.1*. However, we showed that the inbreeding depression in *P. awekei* is probably not severe enough to cause mixed mating and male dispersal, which are common behaviours for this species. An alternative reason for male dispersal may be the high proportion of virgin mothers (table 5.1) that produce male-only clutches. These males will have no access to females if they are not able to disperse and could be a large driving factor in the evolution of male dispersal.

Table 5.5. ANOVA table and model estimates for *model 5.1* (Adjusted $R^2 = 0.188$)

		Estimates	***df***	***F***	**P**
Retained terms					
Homozygous loci		-1.496	1	8.115	0.005
Tree			4	8.580	<0.001
constant:	Tree 1	50.141			
	Tree 2	63.834			
	Tree 3	57.383			
	Tree 4	46.939			
	Tree 5	54.943			
Deleted terms					
Mating status			1	0.051	0.822
Under-developed			1	0.378	0.540
(Homozygous loci)2			1	0.182	0.670
Tree:homozygous loci			4	0.944	0.440
Tree:under-developed			4	2.233	0.068

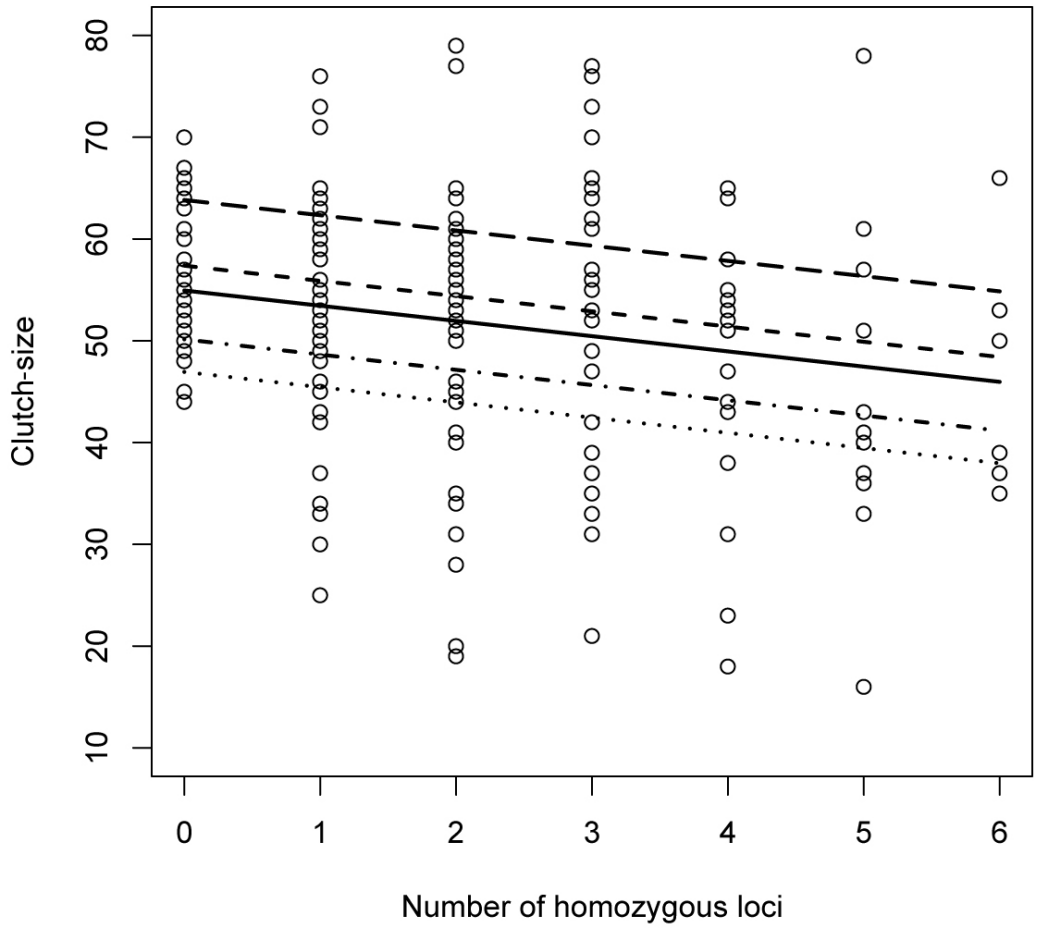

Figure 5.2. Relationship between clutch size and number of homozygous loci for each tree (*model 5.1*). Tree 1: dash-dot line, Tree 2: Long-dash line, Tree 3: short-dash line, Tree 4: dotted line, Tree 5: solid line.

Table 5.6. ANOVA table and estimates for the linear model: Under-developed ~ tree (Adjusted $R^2 = 0.045$).

		Estimates	***df***	***F***	**P**
Retained terms					
Tree			4	3.094	0.017
constant:	Tree 1	0.517			
	Tree 2	0.409			
	Tree 3	0.100			
	Tree 4	0.517			
	Tree 5	0.952			
Deleted terms					
Homozygous loci			1	2.985	0.568
Mating status			1	0.145	0.704
Clutch-size			1	0.252	0.617
(Homozygous loci)^2			1	0.967	0.327
Tree:homozygous loci			4	0.473	0.755

Table 5.7. ANOVA table and estimates for the generalised linear model: Proportion under-developed per clutch ~ tree (data transformed with logit transformation)

		Estimates	*df*	P
Retained terms				
Tree			4	<0.001
constant:	Tree 1	-4.464		
	Tree 2	-5.006		
	Tree 3	-6.300		
	Tree 4	-4.425		
	Tree 5	-3.983		
Deleted terms				
Homozygous loci			1	0.960
Mating status			1	0.509
(Homozygous loci)2			1	0.077
Tree:homozygous loci			4	0.058

Table 5.8. Probability of sibmating from simulated parameter estimates (see table 5.4).

	Probability of sibmating (α)	
	Females (SD)	**Males (SD)**
Model 5.1	0.951 (0.009)	1.000 (0.000)
Maximum	0.832(0.017)	0.995(0.003)
Minimum	1.000(0.000)	0.999(0.001)

A study by Greeff et al. (2009), showed that the number of homozygous loci in *P. awekei* was a good proxy for their inbreeding level, due to the high level of sibmating in this species (see also: Balloux et al., 2004; Pemberton, 2004). They found that *P. awekei* suffer from outbreeding depression and possibly inbreeding depression. However, their results (from 57 mothers and 3 trees), indicated that the effect of inbreeding depression was ambiguous, and they suggested that more data was required to resolve this uncertainty. Using a much larger dataset we clearly show that there is inbreeding depression in *P. awekei*, but in contrast to the study by Greeff et al., no outbreeding depression.

In our results we observed low levels of inbreeding depression in *P. awekei* as the clutch size significantly decreased with every one to two successive sibmatings. Persistence of low levels of inbreeding depression in the face of inbreeding is not uncommon (Antolin, 1999; Henter, 2003). However, to accurately detect low levels of inbreeding depression requires large sample sizes.

The absolute number as well as the proportion of under-developed offspring were significantly different between the different trees. This, in addition to the results from *model 5.1*, showed that the host tree played an important role in how many offspring reach maturity. This means that the tree (i.e. the quality of the fruit) is able to directly affect the development of the wasps, irrespective of the mothers inbreeding level. Additionally, the size of the fruit may also be an indication of the number of flowers within (ovipositing sites), and smaller fruits may therefore reduce the amount of offspring a mother is able to produce. We therefore predict that inbreeding depression in *P. awekei* is more prominent when environmental conditions are harsher while the mortality of developing individuals increase as tree condition decreases.

As an example, it could be mentioned that Tree 2, which had the largest clutch sizes, had the largest fruits (only the largest traps fitted on these figs, see figure 5.1a). In addition, the general status of the tree during wasp development was noted to be excellent, and probably led to the high quality fruit. (*F. salicifolia* are often shrub-like and vary widely in amount of foliage, cycles of fig production, amount of figs, fig size and even fig colour (Burrows & Burrows, 2003). Tree 2 is more tree-like and was covered with leaves and figs during the study (figure 5.3). Our sampling took place during the summer months as more trees produced figs during this season and access to fruiting trees and therefore pollinating wasps was easier. This may actually lead to an underestimate of inbreeding depression.

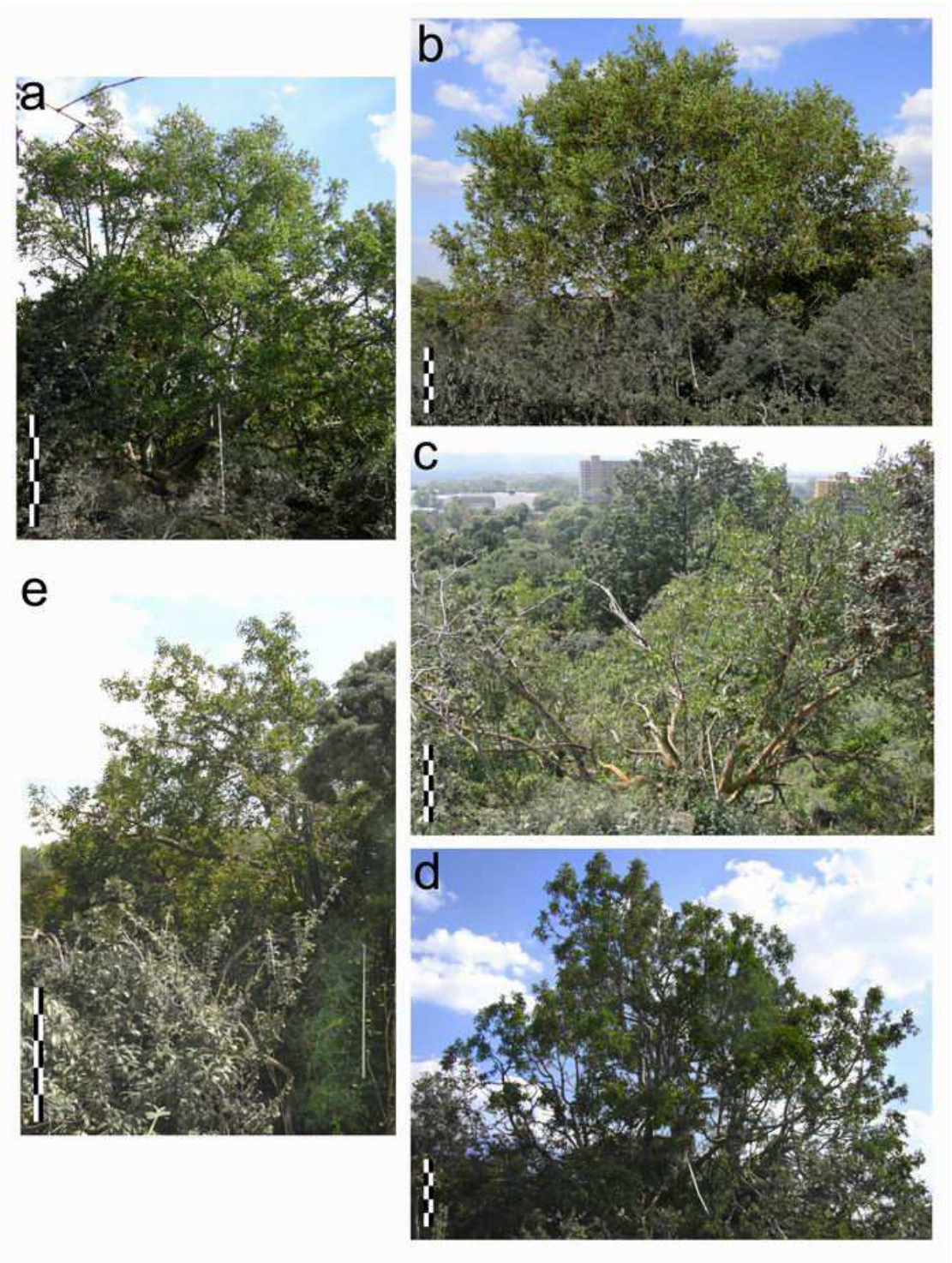

Figure 5.3. Different trees used in study: a) Tree 1; b) Tree 2; c) Tree 3; d) Tree 4 and e) Tree 5. Scale = 1 meter, measured from marker attached to each tree.

Once we established that *P. awekei* suffers from low levels of inbreeding depression, the obvious question was: does this inbreeding depression cause mixed mating and male dispersal in this species? When the measured inbreeding depression was simulated, we found that the females would benefit from low levels of outbreeding (~5%). However, it is the males of this species that decide whether to mate with the females in their native fig or to disperse. When we repeated the simulation with male choice, exclusive inbreeding evolved. Simulating the maximum and minimum inbreeding depression in the 95% confidence interval gave the same result for the males, as they would prefer to inbreed rather than disperse. The females, however, would prefer more outbreeding (~17%) when inbreeding depression increases but when the minimum inbreeding depression was simulated they too benefited more from inbreeding. Because *P. awekei* males (rather than the females) have the choice of sibmating and exclusive inbreeding is optimal under most simulated conditions, the level of inbreeding depression in *P. awekei* is probably not the main force driving the evolution of mixed mating and male dispersal.

From the data we see that a factor that might influence the evolution of male dispersal is the high number of virgin females that produced male-only broods. Our results are varied between the different trees but are on average much higher than the published results for most other pollinating wasp species (often between 0-2% but higher numbers have been recorded, see Godfray, 1988; West et al., 1997). When males find themselves in a patch with no females we expect that this would be an incentive for them to disperse and if this occurs frequently may cause the evolution of dispersal.

In conclusion, we show that for the pollinating fig wasp, *P. awekei,* maternal effects cause low levels of inbreeding depression, which is probably more prominent under poor environmental conditions. While inbreeding depression could play a role in the evolution of mixed mating and male dispersal, our simulation results show that the kin advantage to inbreeding, often outweighs the observed inbreeding depression. The level of inbreeding depression in *P. awekei* is on the border of making mixed mating a viable strategy. Factors that may influence the evolution of male dispersal (and provide the necessary additional incentives) are the high proportion of clutches that have male-only offspring, to reduce local mate competition (Moore et al., 2006), higher and varied sex ratios relative to other pollinating wasp species (Nelson &

Greeff, 2009; chapter 3), or as a syndrome coupled with fighting behaviour (Greeff et al., 2003).

6. Conclusions

In this thesis I addressed a number of mating theories with the aid of simulation modelling. With this approach I explained some of the factors that may drive inbreeding or dispersal. Additionally, I investigated how detection of population structure is affected by different sampling schemes. Lastly, I used molecular data from the pollinating fig wasp *P. awekei* to determine how their level of inbreeding affects their mating system. This combinational approach together with analytical models and ecological data help to unravel systems with inbreeding and dispersal, and shows that individual-based simulation modelling plays an important role in evolutionary studies.

Mixed mating

The enigma of mixed mating, where individuals inbreed with close relatives or outbreed with unrelated individuals has been examined in a number of theoretical models (Campbell, 1986; Charlesworth & Charlesworth, 1987; Damgaard et al., 1992; Feldman & Christianson, 1984; Holsinger, 1988; Holsinger et al., 1984; Lande & Schemske, 1985; Latta & Ritland, 1993; Lloyd, 1979; Maynard Smith, 1977; Maynard Smith, 1978a; Taylor & Getz, 1994; Uyenoyama & Waller, 1991a; Uyenoyama & Waller, 1991b; Waser et al., 1986). Using individual-based simulation modelling and assigning different fitness values for different levels of inbreeding, it was possible to predict the optimal mating strategies for a number of genetic systems. I showed that mixed mating is stable when there are low levels of inbreeding depression and when intermediate inbred individuals have the highest fitness. A requirement for the evolution of mixed mating in these models is, however, the presence of at least three inbreeding classes. The evolutionary stable mating strategies for all the fitness levels investigated were stable and were not affected by the initial mating strategy of the population, except for situations that are comparable to purging in natural systems.

The importance of inclusive fitness in the evolution of mating systems is clearly seen in chapter 2. When the fitness of inbred and outbred offspring are equal

exclusive inbreeding is expected to be the prevalent strategy due to the inclusive fitness advantage to inbreeding (Bateson, 1983; Bengtsson, 1978; Fisher, 1941; Kokko & Ots, 2006; Parker, 1979; Pusey & Wolf, 1996; Waller, 1993; Waser et al., 1986; Wolf, 2000). This was found, as well as, exclusive inbreeding in situations where low levels of inbreeding depression are present, similar to previous studies (Maynard Smith, 1977; Taylor & Getz, 1994). In addition, as the relatedness of the mating partners increased, the benefit of inbreeding increased. This was seen when comparing different mating systems and is important to keep in mind when investigating fitness related traits. Here, however, simulation modelling has an advantage over analytical models, as the kin benefit is an emerging property of individual-based simulations. It was therefore possible to extend the model to include sibmating into the simulation model, in addition to selfing, with relative ease.

It was also possible, using the simulation model, to use empirical data as input and predict when mixed mating will evolve. The results indicated that mixed mating found in a number of species may indeed be due to different fitness levels of serial inbred individuals.

Preventing the co-evolution of fitness values with mating strategies allowed the investigation of possible mating strategies given fitness values while ignoring the effect of the fitness response on the mating strategies. This study is an initial step, using simulation modelling, to unravel how mating strategies evolve when optimal mating partners (i.e. intermediate relatives) are unavailable. An interesting follow-up study would be to compare the results in this thesis with a simulation model where co-evolution of mating strategies and fitness values are allowed.

Dispersal

The occurrence of mixed mating and outbreeding often requires individuals to disperse for them to reach unrelated breeding partners. It can therefore be concluded that various levels of inbreeding depression may drive the evolution of dispersal to attain the optimal mating strategy (Bengtsson, 1978; Clobert et al., 2001; Perrin & Mazalov, 2000; Ronce, 2007). However, theory predicts that dispersal also evolves to reduce the competition of relatives competing for matings while it is inhibited if there is a cost to dispersing (Clobert et al., 2001; Frank, 1986; Gandon, 1999; Hamilton &

May, 1977; Motro, 1983; Motro, 1991; Taylor, 1988; Van Valen, 1971). The same results were found using simulation modelling, shown in chapter 3.

Theory was further corroborated when using the simulation model for situations where dispersal rates were lower when offspring, rather than their parents, determined the rate of dispersal (Gandon, 1999; Hamilton & May, 1977; Motro, 1983; Taylor, 1988). In other words, the parent-offspring conflict was revealed by the unwillingness of the sons to disperse as much as their mothers would want them to.

The main conclusion from chapter 3, however, was that realistic population parameters yielded different results from analytical models, which require simplifying assumptions to make them tractable. It is not surprising that dispersal increased when the sex ratio was unfavourable for obtaining mating opportunities locally while other patches may have had readily available mating partners. However, this was only found when stochastic population dynamics were included in the simulation and sex ratios differed from patch to patch. Similarly, when realistic clutch sizes were used, effects that were found in natural situations emerged. For example, the threat that all the males disperse from a patch, leaving the females unmated decreased dispersal in natural conditions, as well as, in the simulation model, when the number of males per patch decreased.

These results point to the necessity to extend current models by increasing their complexity or reducing their simplifying assumptions to obtain a better understanding of forces driving the evolution of certain traits.

Sampling

After any simulation has completed, the complete genotypes of all the individuals in the population (and if necessary, all the individuals from all the previous generations) are available. Unfortunately, this is not the case in empirical studies, and sampling schemes need to be developed to ensure sufficient individuals are sampled to obtain a valid signal. In chapter 4, I take advantage of the readily available data generated during simulation modelling to compare different sampling schemes.

Large or structured populations require fewer samples than small or unstructured populations. This is important for investigations where there are constraints on the amount of samples obtainable. Current programs that calculate

population subdivision do not require prior information on population structure (Falush et al., 2003; Pritchard et al., 2000). However, when only limited sampling is possible it is easy to miss a population when the sampling scheme is not planned correctly. The application and guidelines developed in chapter 4 show that prior information on biology, ecology and population structure together with summary statistics is still valuable when sampling and typing of individuals are difficult or expensive. Even when sampling and typing of individuals is easy and cheap, careful planning may improve the results from current programs and reduce redundant work.

Platyscapa awekei

The molecular data from chapter 5 indicated that the pollinating fig wasp species, *P. awekei*, suffers from low levels of inbreeding depression. It is common to suppose that inbreeding depression leads to dispersal and outbreeding, which is often seen in this species (Bengtsson, 1978; Clobert et al., 2001; Perrin & Mazalov, 2000; Ronce, 2007). However, before this conclusion is made, it is important to investigate the level of inbreeding while taking into account the kin advantage of inbreeding. By comparing the number of offspring a female was able to produce to her level of homozygosity it is shown that, *P. awekei* suffers from low levels of inbreeding depression. For every additional homozygous locus a mother is able to produce 1.5 fewer offspring. Using this empirical data for *P. awekei*, as input in the simulation model, SibMate, it was found that the level of inbreeding depression is not severe enough to cause dispersal and facultative outbreeding (although a slight increase in inbreeding depression may cause mixed mating). On the other hand, the simulation model in chapter 3 indicated that variation in sex ratio between patches might also drive dispersal. It is therefore possible that low levels of inbreeding depression together with the high local mate competition, a high level of male-only broods found in this species and variation in sex ratio between the patches caused *P. awekei* males to evolve the ability to disperse.

The combinational approach used in this thesis to unravel factors causing inbreeding and dispersal, indicated that various methods can be used to study evolution. I focused on the role of individual-based simulation modelling, and how it may be used to extend analytical models, many of which form the basis of our current

understanding of evolutionary processes. I also showed how empirical data could be used in combination with simulation modelling to clarify processes driving the evolution of traits. There is still much scope for refinement, and a cross-disciplinary approach may not only help but may also be essential to obtain a deeper understanding of natural processes.