UPPSALA
UNIVERSITET

Population Structure and Dispersal of House sparrows (*Passer domesticus*) in an Isolated French Valley

Mohanad Abdel Gadir Ali

Degree project in biology, Master of science (2 years), 2010
Examensarbete i biologi 45 hp till masterexamen, 2010
Biology Education Centre and Department of Population Biology and Conservation Biology, Uppsala University
Supervisors: Peter Halvarsson and Michael Griesser

Abstract

Population structure can be explored by analyzing the genetic diversity and the dispersal rate across the landscape. In this study, the population structure of House sparrows (*Passer domesticus*) was analyzed by investigating the genetic diversity among 922 individuals sampled at fourteen different farms in an isolated valley in Southern France. All individuals were genotyped for 14 microsatellites loci previously developed for the House sparrows and a marker for sexing. I used the basic population genetics and evaluated the level of genetic variation and relatedness between the different localities in the Valley. Using three Bayesian clustering approaches, I inferred the genetic population structure of House sparrow across the landscape. I analyzed the most likely genetic population for individuals and number of migrants to evaluate the dispersal pattern in the Valley. A significance deviation from Hardy-Weinberg was observed in six loci out of fourteen, where two loci displayed significant linkage disequilibria. Among the fourteen loci, null alleles were observed at four of them. The allele frequency distribution indicates low differentiation between the localities. Two clustering methods found nine genetically distinct groups in the study population. The third method indicated only one big group. Pairwise genetic distance between sample sites and relatedness between individuals suggests one big genetic cluster compared to nine obtained from the Bayesian methods. This is likely to be a consequence of dispersal between localities which is confirmed by field observations and allows rejecting the null hypothesis that each individual was born in the population where it was sampled. The weak genetic structuring of the population indicates an ongoing gene flow and substantial connectivity across the Valley. The results reveal a considerable gene flow and it indicates that House sparrows disperse outside localities in the study population despite geographical barriers of a confined valley.

Contents

1. Introduction

Dispersal as a natural process has a strong influence on the dynamics of populations, on both local and regional scales (Johnson and Gaines, 1990). It can also influence the genetic variation within a spatial population (Slatkin, 1987). Dispersal can be affected by landscape heterogeneity and level of geographical barriers (Brooker *et al.*, 1999). The study of genetic structure can reveal how natural populations are structured and is a tool to detect dispersal and gene flow between different populations (Bohonak, 1999). The genetic variation reflects population connectivity (Koenig *et al.*, 1996) and it can also underline the mating system and reveal the evolutionary process behind the geographic variation across populations (Rousset, 2004). If these processes are absent, genetic diversity is expected to be reduced and leading to genetic substructure (Caizergues *et al.*, 2003).

Population studies using genetic markers may give deeper insight on the population structure than field investigations. The use of microsatellite loci as molecular markers has proved to be useful in broad range of population studies (Jarne and Lagoda, 1996). Given their highly polymorphic repetitive DNA sequence, these markers have provided good tool to understand the population structure (Slatkin, 1995). However, the study of the population structure based on microsatellites is posing challenges resulting from the potential presence of null alleles (Dakin and Avise, 2004). Presence of null alleles will bias the allele frequency and decrease observed heterozygosity which bias the pattern of population structure.

The House Sparrow *(Passer domesticus)* is a widely occurring species with a global distribution (Anderson, 2006) and is closely associated with human settlements. Sparrows are gregarious, sexually dimorphic passerine, which are non-migratory and only territorial in the proximity of the breeding site during the reproductive season. In the last decades, many populations in Western Europe have rapidly declined (Hole *et al.,* 2002). A key reason for this decline is the intensification of agricultural land use, which reduces the food supply for house sparrows (Hole *et al.,* 2002) and, changes in the population demography due to increased adult mortality rate (see Ringsby *et al.,* 2006).

From a conservation perspective, information on population structure can be useful when the species occupy fragmented habitats with geographical barriers. The key aim of this study is to measure the level of genetic differentiation and variation considering the variations in

distances between all sample locations among the valley. The genetic markers were used to study the ability of dispersal and detect the patterns of connectivity between sample locations. Here, I will try to answer the question if house sparrows can maintain genetic diversity in such type of habitats assuming that the gene flow decrease with increasing the geographical distances and landscape fragmentation.

2. Materials and methods

2.1. Data and study site background

Data were gathered at 14 different farms in the Lantabat Valley (Fig 1, Table 1) during the period of 2007-2009. In the study area, 11 sampling sites are located within the valley and 3 sites located outside the valley, where the ridge that confines the valley is at its highest elevation (see the map). House Sparrows unlikely fly over the ridge of the valley given that the slopes are not vegetated and 200-300 meters above the bottom of the valley where the sampled farms are located.

All birds were marked using a numbered metal band. The birds were captured during the breeding season (2007) and outside the breeding season during winter (2007-2008) and (2008-2009). During both winter seasons, birds were captured at least twice in each location with at least one month in-between, which allows observing dispersal events both within years and across years. The distances between the localities were ranged from 0.5-10.5km.

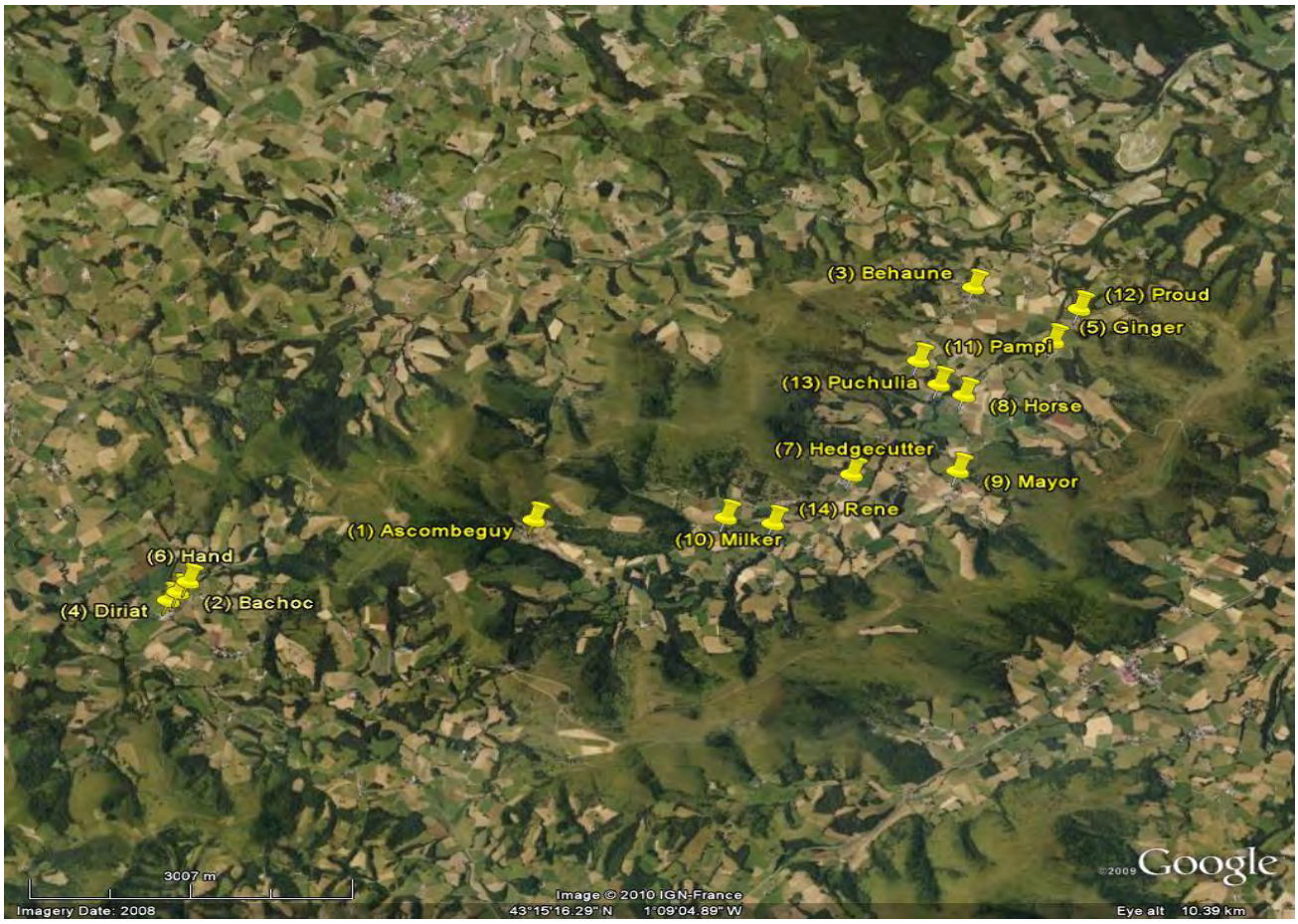


Figure 1. A map showing the sampling sites across the Lantabat Valley. The green belt around Lantabat valley is a mountain ridge with open habitat. The bottom of the valley is characterized by small scale agricultural landscape. Numbers correspond to the order of sampling sites.

Table 1. The number of individuals sampled from all sites (in alphabetical order) and coordinates for each locality.

Site Name	No. sample	Latitude	Longitude
Ascombeguy	187	43.24673	1.16529
Bachoc	130	43.239075	1.206152
Behaune	124	43.271394	1.114819
Diriat	59	43.238197	1.207254
Ginger	44	43.265645	1.10566
Hand	8	43.24018	1.20501
Hedgecutter	38	43.251561	1.128833
Horse	63	43.259953	1.115932
Mayor	35	43.251931	1.116633
Milker	40	43.246984	1.143286
Pampi	57	43.263613	1.121126
Proud	89	43.269124	1.102635
Puchulia	26	43.261105	1.118805
Rene	22	43.246375	1.137858

2.2. DNA extraction

For DNA extraction, approximate of 30 µl blood were taken from the brachial vein of adults and preserved in 95% Ethanol.

DNA was extracted using a high salt purification protocol (Paxton *et al.,* 1996). A small amount from each blood sample was dissolved into a mixture of 350µl SET solution (0.15M), 12µl Proteinase-K (10 mg/ml), and 5.5µl SDS solution (25%). The samples were incubated at 55°C over night. 300µl of NaCl (0.05M) was added to each sample and centrifuged at 16000 G for 10 minutes. The supernatant was transferred into a new tube and inverted with 150µl Tris-EDTA (0.01M, pH 8.0) and 750µl cold Ethanol 99.5%. Samples were stored at -20°C overnight to allow precipitation of the DNA. The precipitated DNA was centrifuged at 10900 G for 15 minutes and the supernatant were discarded. Pellets were washed with 1µl freezer cold Ethanol 70% and centrifuged at 10900 G for 10 minutes. The pellets were let dry overnight and later dissolved into 100µl TE-Buffer (pH 7.6). DNA samples were stored at -20°C for further laboratory work.

2.3. PCR procedure

The 14 primer pairs were divided into 2 sets (A and B, Table 2). The primers were diluted by mixing a specific volume from each primer and an amount form TE Buffer for both sets A and B (Table 2). The total volume of the primers Set A and B where diluted with 234 µl and 338 µl TE Buffer (pH 7). The both mixes were stored at (-20° C).

The reaction mix was prepared using 5 µl QIAGEN Multiplex PCR Mater Mix, 1µl from the primer Mix, 3µl RNASE free water and 1µl from the DNA following the amount volumes described at QIAGEN® Multiplex PCR Handbook (2008).

The multiplex PCR's were performed on a thermal cycler (Unocycler 2007 VWR® or Applied Biosystem® GeneAmp 2700) using the following thermal program: one cycle of 15 min at 95° C followed by 35 cycles at 94° C for 30 s, 57° C for 90 s, 72° C for 30 s, and a final extension cycle for 10 min at 72° C.

2.4. Genotyping

PCR products were diluted 10 times (*i.e.* 1:10) with RNASE free water. The Megabace size standard was prepared by taking a volume of 0.2µl and 7.8µl RNASE free water. MegaBACE ET550-R size standard was used for PCR Mix A and ET400-R size standard was used for Mix B. 2µl of the diluted PCR product was mixed with the diluted size standard followed by genotyping on a MegaBACE 1000 DNA analyzer (Amersham® life science). The

microsatellites were scored using MegaBACE Genetic Profiler software (Amersham Bioscience® V2.2). Genotypes were assigned for each individual using the size ranges and dyes explained in Table 2.

Table 2. Primer sequences used to amplify the microsatellites from House Sparrow. The table also shows the volumes used to prepare both mixes (A & B) and the size range used to analyze the allele sizes.

Locus	Primer sequence	Volume µl	Mix	Size range	Dye
Pdo1	F: TCTGGGCTGTTGCTATCAGAAGGA R: GCAGGGCTGTCCTTTCAACAAACT	6	A	195-259	HEX
Pdo3	F: CTGTTCATTAACTCACAGGT R: AGTGAAACTTTAATCAGTTG	6	A	155-206	FAM
Pdo4	F: CGATAAGCTTGGATCAGGACTAC R: CTTGGGAAGAGAATGAGTCAGGA	50	A	111-171	NED
Pdo5	F: GATGTTGCAGTGACCTCTCTTG R: GCTGTGTTAATGCTATGAAAATGG	6	A	238-439	NED
Pdo6	F: (CT)GATCATGTGTAGATGTAAGACTGC R: (CA)GATCCTTAAGCAGGAAGTTAGG	6	A	226-262	FAM
Pdo9	F: TGCAGTAACAGAAATAACTG R: CTCTCTTTCATTCTTCTCAC	25	A	296-555	FAM
Ase18	F: ATCCAGTCTTCGCAAAAGCC R: (GTTTCTT)TGCCCCAGAGGGAAGAAG	4	A	377-447	HEX
Pdo10	F: AATGTGAATCCCTCCAGAAAC R: ATGGAGTTTGGGGAATGG	4	B	99-145	FAM
Pdo16	F: GTGTATATGCAAATGACAAGACCAAAGC R: (GTTTCTT)TCACGCTGACCTAGATGCTATCAGAG	10	B	273-301	HEX
Pdo17	F: TATGTCTGATAAACCAATCCCTGCAC R:(GTTTCTT)GATCTGTGGTAAATATGGTAATGGAGAGG	8	B	199-235	HEX
Pdo19	F: TCAGAGAGGGCAGAAGGGATTTC R: GCACCGCAGGAGAGCACTTT	8	B	174-188	HEX
Pdo22	F: CATGGGCACAAGAAATGTGA R: TCAAGAAGAAAATGGGTAATATACTGG	10	B	98-130	HEX
Pdo27	F: TGGCAAGGAAGGAGGAATCG R: AGCAATATAAGGCCCAGGTGCTC	8	B	229-256	FAM
Pdo40	F: TGCCACACTCCACTGGGAAC R: (GTTTCTT)GTGTGGAGCAGGGCAGAGATTAG	8	B	301-335	FAM

2.5. Basic genetics test

As many microsatellites were developed for House Sparrow studies (see Neumann and Wetton, 1996; Griffith *et al.,* 1999), the microsatellites used here were previously tested in a multiplex system (Griffith *et al.* 1999; Griffith *et al.* 2006).

House Sparrow microsatellites were analyzed using basic genetic tests, Hardy-Weinberg Equilibrium (HWE) and Linkage Disequilibrium (LD). The Hardy-Weinberg principle (Hardy, 1908 and Weinberg, 1908), states that the allele frequencies from one generation to another remain constant in the population unless mutations, non-random mating or gene flow occur. Linkage disequilibrium (LD) describes the non-random association between the alleles at different loci, which can be the result of non-random mating, mutations or genetic drift. For each population, the genotypes were analyzed whether there are in Hardy-Weinberg equilibrium and to see if there is evidence of linkage disequilibrium.

The exact test for Hardy-Weinberg was performed using a function implemented in windows-based version of GenePop 4.0 (Rousset 2008). The software uses a Markov Chain Monte Carlo (MCMC) method for simulating direct draws from number of complex distribution. The MCMC parameters were set it to 1000 patches, 5000 iteration per patch and 10000 dememorizations for the both tests. The presence of linkage-disequilibrium for each pair of loci was also performed using the same set of parameters for MCMC method in Genepop 4.0. The parameters were set to obtain as lower as possible standard error.

GenePop 4.0 was also used to detect null allele frequencies and to calculate the P- values of Hardy-Weinberg equilibrium test. The exact test for genotypic differentiations based on the allele frequency was calculated and the P-values were estimated using the Fisher's method implemented in Genepop 4.0.

To analyze correlations of individual genotypes among and within populations, the Wright fixation indices F_{it}, F_{is} and F_{st} (Wright, 1931) were calculated using Weir and Cockerham (1984) formula implemented in GenePop 4.0. The test for genotypes correlation based on the allele size, R_{st}, were also calculated. Observed and expected heterozygosity were calculated using GENETIX 4.05 (Belkhir *et al.,* 2004). The pairwise relationship as described by Queller and Goodnight (1989) was calculated using SPAGeDi-1.3a (Hardy and Vekemans, 2002). Genetic distances between different populations were also estimated using the function in SPAGeDi-1.3a. Both relations were tested and graphed in SPSS 11.5 (SPSS Inc., Chicago, IL,

USA) to evaluate the effect of distance in creating such relations. All the basic genetic analyses were calculated based on the sample locations and not the genetic populations obtained from the Bayesian methods.

2.6. Genetic clustering

To infer the population genetic structure, a Bayesian clustering analysis was performed using an approach implemented in STRUCTURE 2.3.1 (Pritchard *et al.*, 2000). STRUCTURE uses MCMC methods to simulate genotypes under the assumption of HWE and LD. The software assumes that there are different genetic clusters in the genotypes and each cluster corresponds to a 'K' value. Multiple runs for the K-values were performed using the admixture model with correlated allele's frequency model. For each K, a burn in period of 30000 (MCMC) were performed followed by 10000 iteration and 5 runs for each K, where the K range was 1 to 20. The output result from STRUCTURE was uploaded into web-based software "Structure Harvester" v0.56.3. Dec 2009 (Pritchard *et al.,* 2000; Evanno *et al.,* 2005; Jakobsson and Rosenberg, 2007) to analyze the K values. Structure Harvester analyzes the outfiles of Structure to visualize the most likely number of genetic clusters (K) by calculating the mean likelihood (Ln) and the variance in likelihood (see Evanno *et al.,* 2005). Evanno *et al.,* (2005) explained that each ΔK represents the uppermost level of the population structure and STRCTURE works to find the posterior probability value of K by calculating lnP(D) which corresponds to the exact probability for a given K. The estimated membership coefficient for each individual in each cluster was obtained from the run of most likely K.

Another Bayesian clustering method was used to analyze the spatial population genetics using TESS 2.3 (Durand *et al.,* 2009). TESS uses MCMC methods to assign distinct genetically groups, based on the genotype and geographical location of each individual. TESS can predict the capability for each run using a statistical method called Deviance Information Criterion (DIC) and sorts them according to DIC values. The method gives the relationships between individuals by creating a network structure. The individual genotypes and the coordinates were simulated with the admixture model using total of 12000 sweeps, 2000 burn-in sweeps with 20 clusters (K) and 5 runs per each K. Other runs were simulated with different values of K where K set it to K from 2 to 10 and k from 2 to 20. The runs where sorted according to the DIC value, where the run with lower value was considered as the best run. The hard clustering assignment of individuals and the neighborhood system based on the genotypes and the geographical information was obtained from the best run.

The third Bayesian method applied in this study was implemented in BAPS 4.14 (Corander *et al.,* 2003; Corander *et al.,* 2004). BAPS is software based on MCMC method for inferring the genetic structure of population using the allele frequencies in the population as reference. BAPS able to detect the number of clusters (*K*) and creates spatial genetic cluster using admixture and mixture models. As pointed by Corander *et al,* (2004), the *K* value must be equal or greater than the reference clusters; the maximum value of *K* was set it to 20. Using population mixture analysis, the clustering of individuals was performed using multiple values of *K* (where *K max =20*) to infer the number of genetic clusters within the population. The mixture clustering graph and the spatial clustering module were obtained as tessellations graph based on the geographical information of each individual.

2.7. First generation migrants

To evaluate the dispersal patterns between sample sites and first generation migrants GeneClass2 (Piry *et al.,* 2004) was used. GeneClass2 calculates the likelihood of an individual belonging to a certain population (L) following the method described by Paetkau *et al.,* 2004. The likelihood populations were computed from the population where the individual was sampled (L_home) over the highest likelihood value among all population samples (L_max). The Bayesian method of Rannala and Mountain, (1997) was used as computation method combined with Monte-Carlo re-sampling method and a simulation algorithm. The alpha value, which corresponds to type I error in the data set, was set to 0.01 as default value and the assignment threshold was set to 0.05 as lower value (see Paetkau *et al.,* 2004). The frequencies-based method (Paetkau *et al.,* 2004) was used as computation method with the same combination of parameters used in the Bayesian method. The simulation parameter was 1000 individuals, all loci and all sampled populations. The method of Rannala and Mountain was used to analyze the likelihood estimation and, the method of Paetkau was applied to reject our null hypothesis that each individual was born in the population where it was sampled (Paetkau *et al,* 2004; Boessenkool *et al.,* 2009).

3. Results

3.1. General results

The locus *Pdo17* showed very low amplification success, and subsequently the 922 individuals were analyzed using 13 loci. Using the sexing marker, 432 males and 337 females were correctly identified among the different localities while the marker failed to identifiy153 individuals. The number of alleles ranged from 4 at Pdo19 to 38 alleles at Pdo4. Null allele's frequencies were observed from the simulation (Appendix 1). Six loci (Ase18, Pdo4, Pdo5, Pdo6, Pdo9, and Pdo22) deviated significantly from the HWE where one pair of loci (Pdo4 and Pdo9) showed significant evidence of LD among the populations (Table 3). The heterozygote deficiency within individuals was observed at 6 loci (Table 3).

The overall pairwise *Fst* value across the populations and loci was small but positive, statistically significant and different from zero (Appendix 2-3). The analysis of genotypic differentiations based on the allele's frequencies indicated that most of populations are distinct from the other observed from P-values (Table 3). There was low but significant variation based on the values of *Rst* (Appendix 3).

Since there was clear presence of null alleles in the data set, I discarded those loci from the Bayesian simulation to avoid a bias in the result. This bias was clearly observed when the clustering results were compared with and without those loci. The clustering results showed here were obtained after the null alleles were removed.

Table 3. Characteristics of the 13 microsatellites loci in House Sparrow in 14 sampling sites in the French Valley. H(e), expected heterozygosity; H(o), observed heterozygosity , P values of Hardy-Weinberg *(HWE)* and for genotypic differentiations *(GD)*. * Departure from HWE and ** Evidence of Linkage Disequilibrium estimated by Genepop4.0.

Locus	**No. alleles**	**Size range *bp***	**H(e)**	**H(o)**	**P-value for *HWE***	**P-value for GD**
As18	11	195-259	0.86	0.74*	1.000	0
Pdo1	10	155-206	0.84	0.80	0.996	0
Pdo3	11	111-171	0.88	0.85	0.999	0.0002
Pdo4**	38	238-439	0.87	0.67*	1.000	0
Pdo5	12	226-262	0.83	0.60*	1.000	0
Pdo6	35	296-555	0.96	0.69*	1.000	0
Pdo9**	12	377-447	0.81	0.64*	1.000	0.0068
Pdo10	11	99-145	0.84	0.79	0.999	0.0157
Pdo16	11	273-301	0.87	0.88	0.799	0.0208
Pdo19	4	174-188	0.65	0.66	0.180	0.0342
Pdo22	9	98-130	0.77	0.46*	1.000	0.1479
Pdo27	10	229-256	0.81	0.73	1.000	0.0031
Pdo40	12	301-335	0.90	0.89	0.580	0.0367

3.2. Clustering

3.2.1. STRUCTURE

The most likely *K* (i.e. assumed genetic cluster) was confirmed by observing the highest value of *lnP(X/K)* obtained from different runs. Based on the first run result obtained from structure harvester (where *K* ranged from 1 to 20) and since the null alleles effect were observed on the graph, set of supplementary runs were performed using the same set of MCMC parameters with different values of *K* (5 to 15, 8 to10 and 6 to 12). The most likely *K* observed from the runs was 9: *K*=9 explained by the highest value of *lnP(X/K)* and *ΔK* (Fig 3). Based on the graphs obtained, the likelihood started from lowest value at *K=1* and increased rapidly to reach maximum value at *K=9* then decreased to *K=20*. The supplementary runs as well showing the rapid increase in *lnP(X/K)* until maximum value at *K=9*. From *lnP(X/K)* and *ΔK* graphs (Fig 2) k=9 tend to attract the most genetic structure for house sparrow.

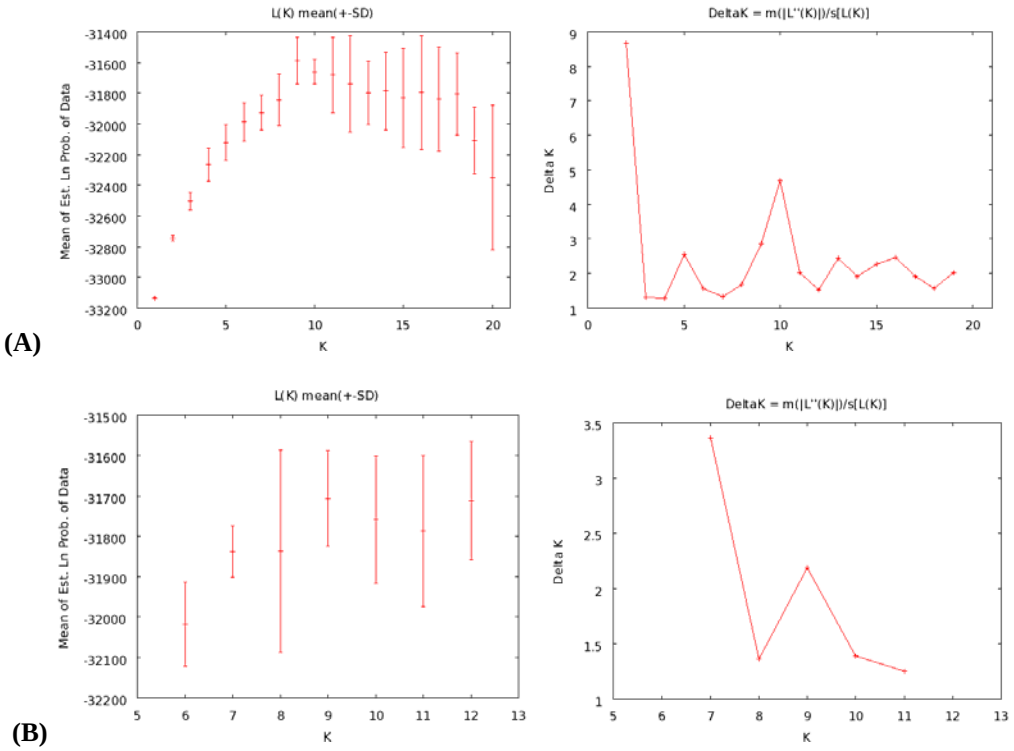


Figure 2. Number of clusters (*K*) estimated using Structure2.3.1. (A) Showing the likelihood of the data given K, lnP(X/K), when K was set it from 1 to 20. Delta *K* (*ΔK*) represents the order rate of change in *lnP(X/K)*. (B) Showing *lnP(X/K)* and *ΔK* when K was set from 6 to12.

From the plots of ancestry estimates, the estimated membership coefficient for each individual in each cluster were obtained from the runs where k=9. The average proportion membership 'Q', of which describes the individual's estimated membership fraction in each *K* cluster, indicated nine genetic clusters corresponding to nine different colors (Fig 3). From the

both estimates, an admixture ancestry was observed. Each individual in this population tend to share one or more ancestors with the other individuals.

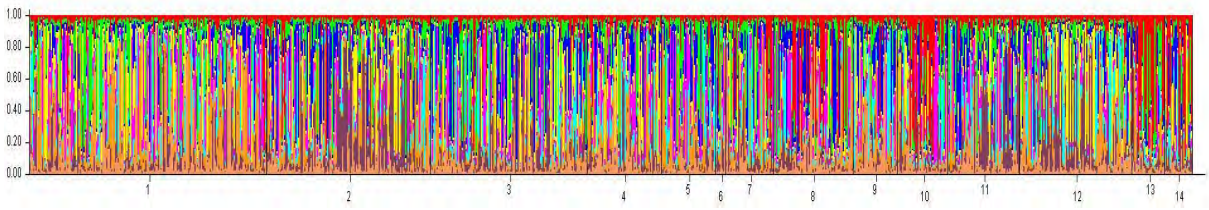
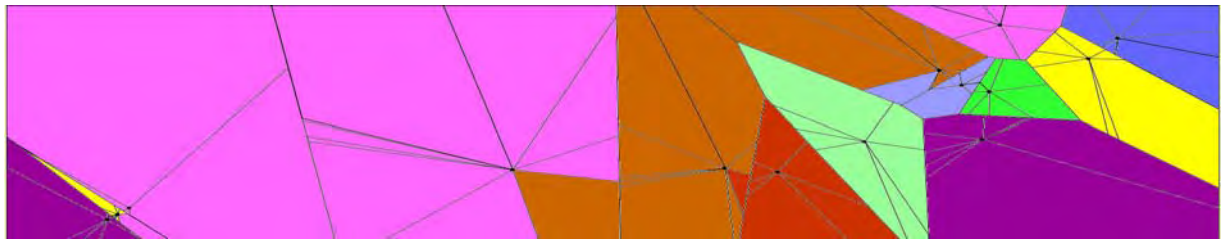


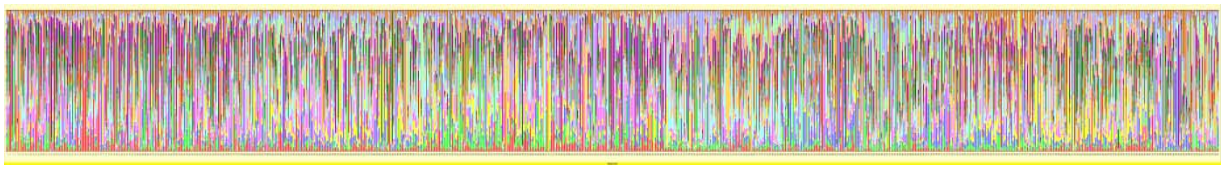
Figure 3. Plot of ancestry estimates where K=9 obtained from Structure 2.3.1. The plot showing proportion membership at K=9 sorted by population ID. Each vertical line represents an individual and the numbers are corresponding to population ID.

3.2.2. TESS

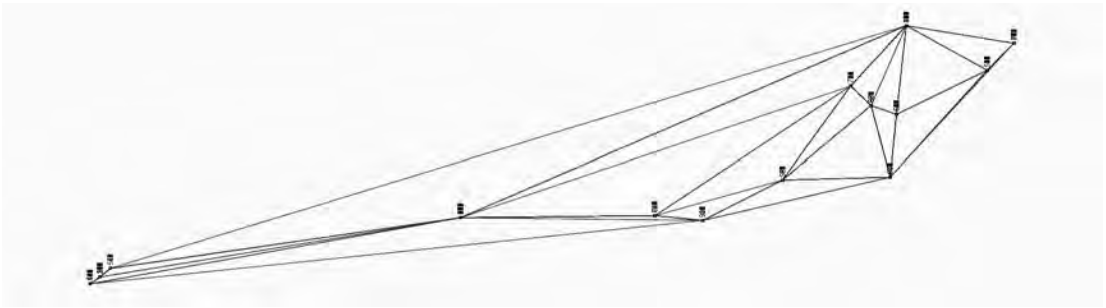
The hard clustering diagram obtained from TESS shows nine different genetic clusters within the valley (Fig 4 A). As the best run corresponds to the lowest value of DIC, the best run indicated nine clusters (See Appendix 4 for DIC values). The lowest DIC values obtained from the other runs for K (K from 2 to 10 and 2 to 20) also indicated nine genetic clusters. Assignment probabilities for individuals (as in Fig 4 B) indicated a clear admixture where each individual has membership with one or more individuals in the population. The neighborhood system diagram (Fig 4 C) show that the existence of geographical barriers did not affect the spatial neighborhood between individuals since there is clear spatial connection despite those outside the ridge of the valley.



(A)



(B)

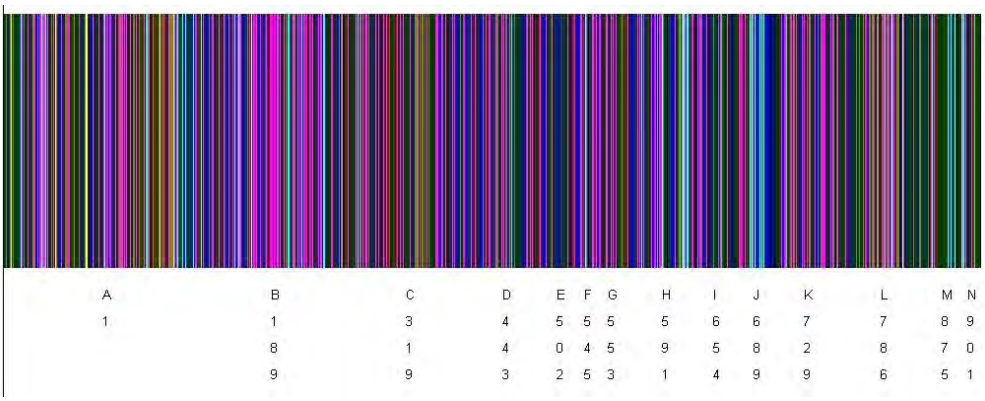


(C)

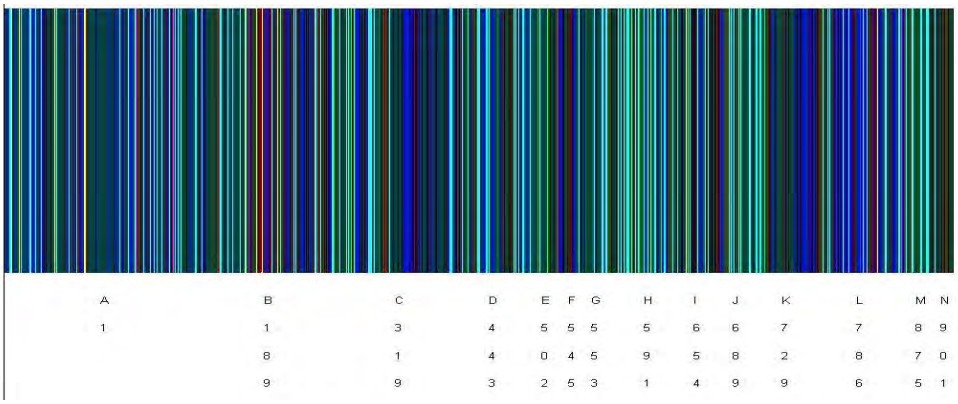
Figue 4. Diagrams obtained from TESS using Bayesian admixture model where K=9 (A) Showing the hard clustering tessellation where each color corresponds to one cluster.(B) Assignment probabilities for individuals; each vertical line represents one individual and (C) Neighborhood system diagram showing the spatial connection within the Valley.

3.2.3. BAPS

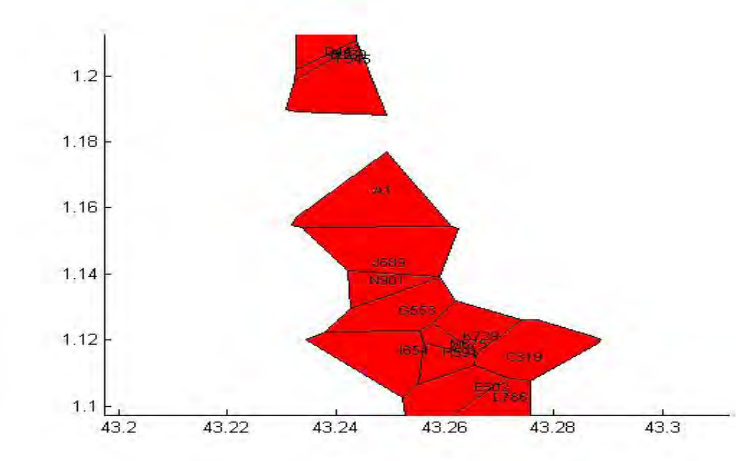
The clustering graphical output from BAPS shows membership for each individual represented by a vertical line. Surprisingly, the output results from BAPS indicated one big genetic cluster exist in the valley. This cluster can be observed in the tessellation at the spatial clustering module (Fig 5).



(A)



(B)

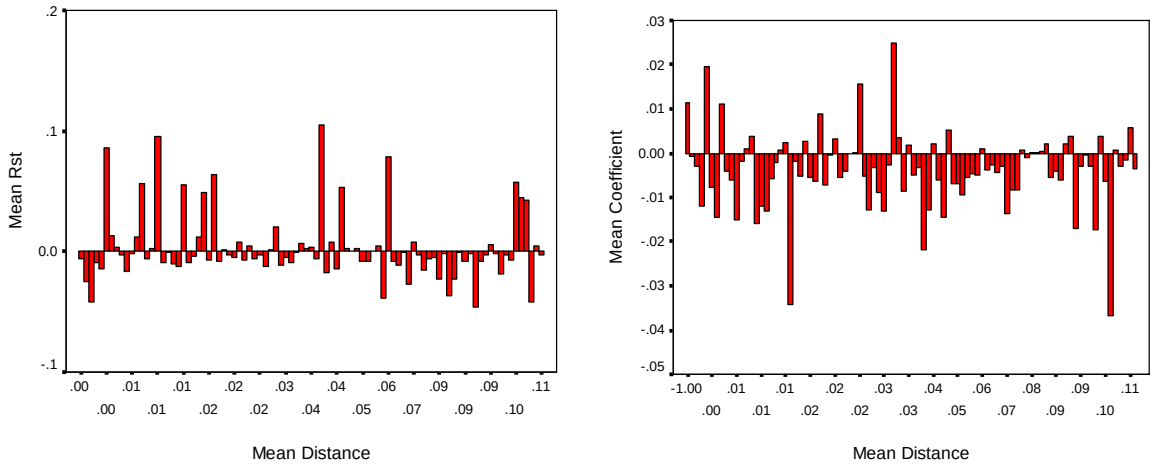


(C)

Figure 5. Mixture clustering graphs in both mixture (A) and admixture (B) models where each vertical line represents one individual. (C) A tessellation spatial clustering graph based on the geographical information of sample sites showing one big genetic cluster indicated in red color. X and Y axes refer to latitudes and longitudes coordinates.

3.3. Pairwise genetic distance and relatedness

Pairwise genetic distances between different populations as obtained from SPAGeDi-1.3a show that distance has no clear effect in creating variation between populations (Fig 6A). The relatedness coefficient of Queller and Goodnight, (1989) indicated admixture relations between individuals regardless the distances between each pair of them (Fig 6B). Both pairwise genetic distance and relatedness showed that no effect on population and individual levels in the valley.



(A) **(B)**

Figure 6. (A) Pairwise genetic distance between sample sites using mean values for the distance and the pairwise genetic coefficient. (B) Mean relatedness coefficient with the mean distances between individuals.

3.4. Dispersal ability

The results from *GeneClass2* indicate a considerable dispersal between sample sites. The likelihood estimation from Rannala & Mountain (1997) method indicated 693 individuals with a probability (*i.e.* probability to belong to the sample site) below the threshold ($p<0.05$). The simulation following the Paetkau *et al.* (2004) method, explained those individuals belong genetically to other populations rather than those where they were sampled from, rejecting the null hypothesis that they were born in the sampled population. The likelihood results obtained from GeneClass2 were combined with the sampled populations, recaptured data and individual sexes to analyze the pattern of dispersal between the localities. Obtained results show that House sparrows keep dispersing regardless the geographical distances between localities (Appendix 5).

4. Discussion

4.1. Null alleles and genetic differentiation

The results from population genetic studies based on the microsatellites are generally reproducible compared to other genetic markers like minisatellites (see Jarne and Logoda, 1996). However, such kinds of studies are harmed in the presence of null alleles. Although the presence of null alleles is quite common in microsatellites (Neumann and Wetton, 1996), little has been written on their effects in population genetic analyses (Dakin and Aves, 2004). The potential cause of null alleles, as described by Kwok *et al.,* (1990), arise from the divergence of the sequence in one or both flanking primers which cause a poor primer annealing. This mutation at the end side of the primer may lead to poor PCR amplification. Null allele presence could also result from an error during the genotype scoring (Dewoody *et al.,* 2006). As a result, the genotype-based statistics may then be biased in the presence of null alleles.

As noted by Chakraborty *et al.,* (1992), an observed heterozygote deficiency can also indicate the presence of null alleles. The reason for that, those heterozygous individuals in the presence of null alleles will be scored as homozygous. The allele frequencies will be biased in the presence of null alleles and the observed heterozygosity will decrease and the level of inbreeding will then increased.

The deficiency observed can be also explained considering the PCR procedures. According to Wattier *et al.,* (1998), alleles with large size tend to amplify lower than those with small size

due to competitive nature of PCR. This situation can be observed in the case of Pdo4 and Pdo6 where the size range is 238-439 for Pdo4 and 296-555 for Pdo6. This also can explain the observed deficiency of heterozygosity in Pdo4, Pdo6, and Pdo9.

The false observations of excess homozygotes might lead to deviation from Hardy-Weinberg equilibrium observed at the six loci. However the deviation may indicate problems with genotyping (Wigginton *et al.,* 2005); the one observed here can be explained in the term of null allele's presence. When some loci are deviating from HWE and some are showing equilibrium that might be clear indication for random mating and admixture between the individuals.

Wright's *F-statistics* which was used to detect level of genetic differentiations and to infer gene flow in House sparrow population was expected to be less informative. Weak but significant genetic differentiations were detected within the valley. In a large landscape and relatively equally connected populations, *F-statistics* is assumed to be more informative (Manel *et al.,* 2003). Accordingly, the *F-statistics* for such highly distributed species occupying small heterogeneous landscape will be less helpful. The reason is that such type of connectedness between localities will create a complex pattern of gene flow and then the differentiations might be uneasy to detect.

The *F-statistics* measures were noted to be affected in the presence of null alleles whereas the measure of spatial genetic structure tends to be less affected (Dakin and Eves, 2004). The reason for this is that the *F-statistics* is quantifying the gene flow to infer the population structure; the presence of null alleles might reduce the sensitivity of this measure.

Potential presence of null alleles in the genotypes was constrained the ability of conducting deep analysis and to evaluate the pattern of differentiations in both individual and population levels. Therefore, such constrain should be kept in mind when interpreting the result.

4.2. Migration and dispersal

There was large variation in the combined result from GeneClass2 and observational data which support frequently occurring dispersal events. For example, some individuals were sampled in one place and where genetically assigned to another place and were recaptured in a third place. Another example shows that individuals were sampled in one site far from the

recaptured one (*e.g.* a female sampled in Proud and recaptured at Diriat where the distance between the two localities is 9.05 Km). The assigned site for this individual was Pampi a close locality to the sample site (see Appendix 5 for some examples). The likelihood results showed that Mayor, Ginger and Diriat tend to be the most likely genetic populations for major number of individuals. Since these localities are far from each other that could explain the great dispersal within the valley. Hand showed very low frequency as an assigned sample site and can therefore be regarded to be part of the nearby locations (Bachoc and Diriat). Some individuals expressed residency where they were sampled, assigned and recaptured at the same site (Appendix 5). The combined result form GeneClass2 and field data showed that females tend to disperse more widely whereas the majority of males were assigned to their sampled population. This phenomenon was mentioned by Fleischer *et al.,* (1984) who showed that small female House sparrows disperse long distances when the social interaction force them to leave the site before been able to establish themselves in a new flock. Male sparrows disperse less if the survival prospects are high (Altwegg *et al.,* 2000). Large numbers of individuals (*i.e.* those are below threshold) were assigned to other population indicate large dispersal and random mating occurring between the different localities.

4.3. Pairwise genetic distance and relatedness

Weak isolation by distance was observed in this study at scale up to 10km. Since there was no clear effect of distance in population pairwise genetic difference and considering the high dispersal rate, the dynamic of House Sparrow population in the valley seems to act as one big population. As well, and since the relatedness between individuals were not affected by distance, most of house sparrow individuals are close related despite the geographical barriers. The low genetic differentiations between sample sites is a result from the complex pattern of gene flow (see *F-statistics* values) which is also reflected in the patterns of relatedness and pairwise genetic distance. As pointed out by Wagner *et al.* (2006), presence of null alleles may skew the result of genetic relatedness. Their study highlighted the importance of correcting for null alleles to avoid the bias associated with relatedness estimates. Accordingly, relatedness observed here might be inaccurate in presence of null alleles.

4.4. Population genetic structure

Considering the small scale among the House sparrows sample sites, I found weak genetic structure. Surprisingly, and based on the Bayesian analysis from Structure 2.3.1, the population of house sparrows was subdivided into nine genetic clusters. The membership

coefficients indicated each individual has membership in one or more localities. The admixture observed from Structure clearly indicates random mating.
Unexpectedly, the three populations outside the valley (Bachoc, Diriat and Hand) were not assigned as genetically distinct populations but, as two small patches from those in the bottom the valley (Fig 4). Taking the result from GeneClass2 into account, comparing it to TESS hard clustering, there are two populations instead of three laying out the ridge of the valley. The third population, as seen from GeneClass2, expected to be a small extension from the two other populations. The hard clustering indicated a strong gene flow between the sample sites at the bottom of the valley and those outside the ridge.

Isolation by distance show that House sparrows have mixed ancestors with the other localities (from GeneClass2 result).Thus, the result from Bayesian methods (mainly Structure and TESS) is not giving a true picture of the genetic population structure. The mixed ancestry may affect the value of *K* and then the clusters may not correspond to real population. As noted by Pritchard *et al.,* (2000) and considering the high migration rate between the localities in the valley, the *K* value in this case is not biologically interesting. Pritchard *et al.,* (2000) explained that the population will not act as one structured unit in the presence of migration. Consequently, the high level of migration observed between the different localities will be enough to assure there is no independent population. The procedure of inferring *K* is *ad hoc* procedure and obtained clusters may not have clear biological interception. Considering the low genetic differentiation observed in this study, Evanno *et al.,* (2005) explained that ΔK is performing better in population with high genetic differentiation and not with weak differentiation compare to the method of Pritchard *et al.,* (2000) to estimate the genetic cluster. Thus, the ΔK observed here must be taken in highly speculative as result of the low genetic differentiation.
Deviation from HWE and the presence of LD at different loci might also affect the value of *K* obtained since Structure simulates the genotypes assuming there are in HWE and there is no LD. In other words, each locus is an independent piece of information represents the genetic history of the population. If the loci are not independent (*i.e.* genetically linked), the result will be biased since the loci information will be over represented. Therefore, result expected to produce linkage groups depending on the linkage between the loci. As Pritchard *et al*., (2000) noted the presence of these factors, even in the absence of population structure, my lead to weak statistical signals (STRUCTURE user manual page 13). As well, deviation from HWE in large sample sizes may increase the false statistical signal.

Taking into account the problems associated with Structure I mentioned before (see also Pritchard *et al.,* 2000; Evanno *et al.,* 2005), the valley population of House sparrows is comprised one big cluster instead of nine. Consequentially, the population structure output from BAPS is expected to be more realistic than Structure and TESS. In way to study the effective Bayesian method for estimating number of populations, Waples and Gaggiotti, (2006) found that BAPS tend to be more reliable and conservative method. BAPS uses conservative calculations to avoid creating additional populations and can estimate the number of population despite presence of migrants. Since migration can create differences in population structure, clustering individuals in BAPS perform better and reliable compare to the same function in Structure.

Population structure identified by the distance and in the massive migration event was less qualitative. Without prior information about sampling sites, the identified populations were also less informative. In way to improve the study of genetic structure and its correlation to geographical distances, landscape structure should be included.

5. Conclusion

The result suggests that House sparrows are good dispersers across the distances between the sampled sites including those outside Lanabat Valley. The low pairwise differentiations observed within the 14 localities in consideration to the small scale of the valley may describe the sedentary nature of House sparrow. The gene flow is to be related to the level of connection between sites. Methods used for clustering individuals, particularly Structure and TESS, had a limited power in the presence of high gene flow and therefore, more detailed analysis are needed. Ongoing levels of migration within the valley clearly indicate that the House sparrows in the valley comprises of one genetic group. The high genetic connection and the gene flow observed in this study might be sufficient for the House sparrow to maintain genetic variability and to avoid inbreeding between different populations in Lantabat Valley. From a genetic point of view, such connected populations can be managed as one genetic unit for conservation implications.

Acknowledgement

I would like to express my gratitude and appreciation to my supervisors Peter Halvarsson and Michael Griesser for allowing me to work in such interesting data and for their extreme and kindly help during the project work. Thanks to Peter for his guides and valuable tips in the lab and during the data analysis; Gunilla Engström for answering endless questions, facilitating everything and for her nice personality in the lab. Also, I would like to thank Reija Dufva from Animal Ecology Department for allowing me to use some of her facilities and for been helpful in the lab. Thanks to Martin Carlsson for spending his time to help me handling Structure 2.3.1 and for giving valuable tips and comments. Many thanks to Magnus Johansson for his help in solving the input file problems and for his friendly mode. Thanks to the nice mode of people and the wonderful climate at the Department of Population Biology and Conservation Biology of Uppsala University and for anyone provided me with such valuable suggestions and comments.

Thanks to all supports form Wildlife Research Center to conduct my master study and all colleagues, especially Dr. Elfatih Eltagi, for their warm support. Great thanks to my brother Moawia and my sister Sanna for their extreme and endless support and for make me always feel comfort and warm.

Thanks to the inhabitants of Lantabat for allowing us to collect data on their farms, and the family Bachoc at Etchemendigaraya as well as Graciella in Ascombeguy for their hospitality.

This study was supported by the Swedish Natural Research Council, the Stiftelsen för Zoologisk Forskning, C. F. Lilljewalchs Resestipendium, the Swedish Royal Academy of Sciences and Helge Ax:son Johnson Stiftelsen (all to M. G.).

Blood samples were taken under the ethics license from the DIREN Aquitaine in Bordeaux, and birds were ringed under the license of the CRBPO in Paris.

References

Altwegg, R., Ringsby, T.H., and Saether, B.E., (2000). Phenotypic correlates and consequences of dispersal in a metapopulation of house sparrows Passer domesticus. Journal of Animal Ecology, 69:762–77

Anderson, Ted R., (2006). Biology of the ubiquitous house sparrow: from genes to populations. Oxford University Press, Pages 21-28. Found at: http://books.google.com/books?id=aMqo_gS_xIcC&printsec=frontcover&dq=house+saprrow&cd=1#v=onepage&q&f=false

Belkhir, K., Borsa, P., Chikhi L., Laboratoire Génome, Populations, Interactions, CNRS UMR 5000. University of Montpellier II, Montpellier, (2004). GENETIX 4.05, Windows software for population genetics.

Boessenkool, S., Star, B., Waters, J.M., Seddon P.J., (2009). Multilocus assignment analyses reveal multiple units and rare migration events in the recently expanded yellow-eyed penguin (Megadyptes antipodes). Mol Ecol 18:2390–2400

Brooker, L., M. Brooker, and P. Cale. (1999). Animal dispersal in fragmented habitat: measuring habitat connectivity, corridor use, and dispersal mortality. Conservation Ecology, 3(1):4. Online version: http://www.consecol.org/vol3/iss1/art4/

Caizergues A., Ratti O., Helle P., Rotelli L., Ellison L., and Rasplus J. (2003) Population genetic structure of male black grouse (Tetrao tetrix L.) in fragmented vs. continuous landscapes. Mol Ecol, 12:2297–2305

Chakraborty, R., De Andrade, M.; Daiger, S.P., and Budowle B., (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann Hum Genet, 56: 45–57.

Corander, J. Waldmann, P. and Sillanpää, M.J. (2003).Bayesian analysis of genetic differentiation between populations. Genetics, 163: 367-374.

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

Dakin, E.E., and Avies, J.C., (2004). Microsatellites null alleles in parentage analysis. Heredity, 93:504-509.

Dewoody J.D., Nason J.D., Hipkins V.D., (2006). Mitigating scoring errors in microsatellites data from wild populations. Molecular Ecology Notes, 6:951–957.

Durand, E.F., Jay, O.E., Gaggiotti, O. François (2009). Spatial inference of admixture proportions and secondary contact zones, Molecular Biology and Evolution, 26:1963-1973.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol., 14:2611-2620.

Griffith, S.C., Stewart, I.R.K., Dawson, D.A., Owens, I.P.F., and Burke T., (1999). Contrasting levels of extra-pair paternity in mainland and island populations of the house sparrow (Passer domesticus): is there an 'island effect'? Biological Journal of the Linnean Society, 68:303–316.

Hardy, H.G., (1908). Mendelian proportions in a mixed population. Science 28:49–50

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

Hole, D.G., Whittingham M. J., Bradbury R.B, Anderson G.Q.A., Lee P.L.M, Wilson J.D and Krebs J.R, (2002). Widespread local house-sparrow extinctions—agricultural intensification is blamed for the plummeting populations of these birds. Nature, 418, 931–932.

Jakobsson, M. and Rosenberg, N. A., (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. and Lagoda, P. J. L., (1996). Microsatellites, from molecules to populations and back. Trends Ecol Evol, 11: 424–429.

Johnson M.L., and M.S., Gaines (1990). Evolution of dispersal: theoretical models and empirical tests using birds and mammals, Anna. Rev. Ecol. Syst, 21:449–480.

Koenig W.D., Van Vuren D., Hooge P.N., (1996). Delectability, philopatry, and the distribution of dispersal distances in vertebrates. Trends Ecol Evol, 11:514–517.

Manel S., Schwartz M.K., Luikart G., Taberlet P., (2003). Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol, 18:189–197.

Kwok S, Kellog D.E., McKinney N., Spasic D., Goda L., Levenson C., (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus 1 model studies. Nucleic Acids Res 18: 999–1005.

Neumann, K., and Wetton, J.H., (1996). Highly polymorphic microsatellites in the house sparrow Passer domesticus. Molecular Ecology, 5:307-309.

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

Paxton, R. J., Thoren, P. A., Tengo, J., Estorup, A. And Pamilio, P. (1996). Mating structure and the nestmate relatedness in a communal bee, Andrena jacobi (Hymenoptera, Andrenidae), using microsatellites. Mol Ecol, 5:511-519.

Piry S, Alapetite A, Cornuet, J.M., Paetkau D, Baudouin, L., Estoup, A., (2004). GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. Journal of Heredity, 95:536-539.

Pritchard, J. K., Stephens, M., and Donnelly, P., (2000). Inference of population structure using multilocus genotype data. Genetics, 155:945–959.

QIAGEN® Multiplex PCR Handbook, February (2008). Can be found at: www1.qiagen.com/HB/MultiplexPCR
Rannala B, Mountain J.L., (1997). Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences USA, 94:9197-9201.

Ringsby, T.H., Saether, B.E., Jensen, H. & Engen, S., (2006). Demographic characteristics of extinction in a small insular population of House Sparrows in Northern Norway. Conservation Biology, 20:1761–1767.

Rousset, F., (2004). Genetic structure and selection in subdivided population. Princeton University Press. Pages 1-3 Found at: http://books.google.com/books?id=rB7W9PE1-DEC&printsec=frontcover&dq=genetic+structure&cd=1#v=onepage&q&f=false

Rousset, F., (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resources, 8: 103-106.

Slatkin, M., (1987). Gene flow and the geographic structure of natural populations. Science, 236:787-92.

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

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in Gracilaria gracilis (Rhodophyta). Mol Ecol, 7: 1569–1573.

Weinberg, W. (1908). On the evidence of heredity in humans. Annual Reviews of the Association for patriotic Naturkunde in Württemberg, 64: 368-38

Weir, B.S., and Cockerham, C.C., (1984). Estimating F-Statistics for the analysis of population structure. Evolution, 38: 1358–1370.

Wigginton J.E., Cutler D.J. and Abecasis G.R., (2005) A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet, 76:887–893.

Wright, S., (1931). Evolution in mendelian populations. Genetics, 10:97-159

Appendix

Appendix 1. Null allele frequencies as estimated by GenePop4.0

	Ascombeguy	Bachoc	Behaune	Diriat	Ginger	Hand	Hedgecutter	Horse	Mayor	Milker	Pampi	Proud	Puchulia	Rene
As18	0.1077	0.0944	0.0421	0.176	0.0514	0	0.1305	0.0516	0.0999	0.1179	0.0546	0.1737	0	0.0639
pdo1	0.0687	0.1105	0.0902	0.0431	0.1366	0.0822	0.0702	0.0545	0.2712	0.1299	0.1927	0.0887	0.1163	0.161
pdo3	0.0319	0.0157	0.0515	0.0732	0	0.1553	0.0036	0.0073	0.019	0.0208	0.0105	0.1039	0.0449	0.0039
pdo4	0.0589	0.1252	0.0387	0.0902	0.0983	0.0878	0.1052	0.1276	0.1037	0.1495	0.0913	0.1464	0.0531	0.0275
pdo5	0.1111	0.199	0.1528	0.1056	0.2363	0.1221	0.1698	0.1293	0.2555	0.0764	0.1963	0.1713	0.1843	0.1861
pdo6	0.1088	0.1343	0.0825	0.123	0.1202	0.2857	0.0782	0.1078	0.0998	0.167	0.1816	0.1429	0.1492	0.1106
pdo9	0.0581	0.1336	0.0564	0.1465	0.1211	0	0.0254	0.1242	0.1404	0.0962	0.1456	0.1214	0.1098	0
pdo10	0.2342	0.2551	0.2476	0.1778	0.2345	0.4119	0.2695	0.1909	0.1857	0.2162	0.2325	0.2896	0.2141	0.2476
pdo16	0.0112	0.0102	0.0331	0	0	0	0.0803	0.0109	0.0181	0.0025	0.0456	0.014	0	0
pdo19	0	0.0215	0.0034	0.0251	0.0783	0	0.0153	0.0205	0	0.041	0.0992	0	0	0.0093
pdo22	0.1543	0.1438	0.1259	0.1645	0.1195	0.291	0.1783	0.134	0.1333	0.0974	0.148	0.1682	0.1865	0.2299
pdo27	0.0232	0.0171	0.0633	0.0591	0.1055	0.0892	0.1112	0.0547	0	0	0.0045	0.0072	0.0272	0.0799
pdo40	0.0077	0.0112	0.0103	0	0.0661	0	0	0.0168	0.0064	0	0	0	0	0

Appendix 2. Pairwise Fst values among 14 sampling sites and 922 individuals using 13 microsatellites loci as calculated by GenePop4.0

	Ascombeguy	Bachoc	Behaune	Diriat	Ginger	Hand	Hedgecutter	Horse	Mayor	Milker	Pampi	Proud	Puchulia	Rene
Ascombeguy														
Bachoc	0.006343													
Behaune	0.008882	0.005077												
Diriat	0.000877	0.004866	0.007568											
Ginger	0.01339	0.01178	0.008995	0.01092										
Hand	0.01828	0.01087	0.01959	0.01413	0.03169									
Hedgecutter	0.00717	0.005739	0.007075	0.001222	0.01327	0.004429								
Horse	0.009065	0.007751	0.0131	0.005316	0.01886	0.005327	0.002883							
Mayor	0.01037	0.007308	0.00511	0.0076	0.005093	0.0112	0.008763	0.01142						
Milker	0.008911	0.008725	0.008539	0.006901	0.01907	0.009328	0.007042	0.003492	0.008329					
Pampi	0.008748	0.004414	0.005695	0.004107	0.01195	0.006471	0.00342	0.006069	0.004307	0.005715				
Proud	0.008296	0.003194	0.007403	0.006086	0.01889	0.006526	0.009625	0.01011	0.007162	0.01045	0.00399			
Puchulia	0.02378	0.02282	0.02699	0.02488	0.04096	0.01403	0.01461	0.00578	0.02696	0.003137	0.01997	0.0222		
Rene	0.01832	0.01596	0.009833	0.01429	0.02803	0.02098	0.008454	0.01718	0.0148	0.004397	0.009101	0.0139	0.01421	

Appendix 3. Fstat (A) and Rstat (B) values per each locus calculated by GenePop4.0.

(A)

Locus	**F(is)**	**F(st)**	**F(it)**
As18	0.1062	0.009	0.1142
pdo1	0.0432	0.0189	0.0613
pdo3	0.0253	0.0058	0.0309
pdo4	0.2064	0.0454	0.2424
pdo5	0.2623	0.0143	0.2729
pdo6	0.2546	0.0026	0.2565
pdo9	0.2174	0.0034	0.2201
pdo10	0.0492	0.0014	0.0505
pdo16	0.0105	0.0005	0.011
pdo19	-0.0296	0.0031	-0.0265
pdo22	0.3672	0.004	0.3697
pdo27	0.0831	0.0052	0.0879
pdo40	0.0132	0.0028	0.0159
	ALL		
	0.1226	0.009	0.1305

(B)

Locus	**R(is)**	**R(st)**	**R(it)**
As18	0.0961	0.0091	0.1043
pdo1	0.0158	0.0041	0.0197
pdo3	-0.0158	0	-0.0158
pdo4	0.2452	0.0032	0.2476
pdo5	0.2485	0.0099	0.256
pdo6	0.4107	0.0109	0.4171
pdo9	0.1342	0.0103	0.1431
pdo10	0.0553	0.0069	0.0619
pdo16	-0.0061	-0.0032	-0.0093
pdo19	0.0038	0.0028	0.0066
pdo22	0.1999	-0.0039	0.1968
pdo27	0.0699	0.0107	0.0798
pdo40	-0.0106	-0.0055	-0.0161
	ALL		
	0.2642	0.0063	0.2689

Appendix 4. Values of DIC obtained from TESS simulations for different ranges of K. Bold raw indicts the lowest DIC which corresponds to the best run.

Label	Kmax	DIC	Trend_Degree	Model
housesparrow10_RUN_000001	02	65074.1	1	BYM
housesparrow10_RUN_000002	03	63707.9	1	BYM
housesparrow10_RUN_000003	04	63385.9	1	BYM
housesparrow10_RUN_000004	05	62883.2	1	BYM
housesparrow10_RUN_000005	06	62962.4	1	BYM
housesparrow10_RUN_000006	07	63084.9	1	BYM
housesparrow10_RUN_000007	08	62952.9	1	BYM
housesparrow10_RUN_000008	**09**	**62678.8**	**1**	**BYM**
housesparrow10_RUN_000009	10	62753.6	1	BYM

Appendix 5. Table showing some examples from the combined result from GeneClass2 and field data.

Individual ID	Sampled site	Sex	Likelihood Pop. from GeneClass2	recapture site	number of recaptures	distance between banding and recap site (km)
SC133656	Puchulia	male	Horse	Horse	3	0.5
SC133638	Puchulia	male	Horse	Horse	3	0.5
SC115829	Proud	female	Pampi	Diriat	2	9.5
SC115901	Proud	female	Bachoc	Proud	2	0
SB21709	Hedgecutter	male	Pampi	Hedgecutter	2	0
SC133780	Behaune	female	Proud	Ascombeguy	2	4.9
SC133778	Behaune	female	Rene	Ascombeguy	2	4.9
SC115848	Bachoc	male	Ginger	Diriat	2	0.5
SC133749	Ascombeguy	male	Puchulia	Ascombeguy	2	0
SC133740	Ascombeguy	male	Proud	Ascombeguy	2	0
SC115753	Ascombeguy	male	Rene	Ascombeguy	2	0
SC115741	Ascombeguy	female	Pampi	Ascombeguy	2	0
SB66894	Ascombeguy	male	Ginger	Milker	2	2.5
SB21785	Mayor	male	Behaune	Mayor	2	0
SB21783	Mayor	male	Mayor	Mayor	2	0
SC133795	Horse	male	Horse	Horse	2	0
SC133790	Horse	male	Horse	Horse	2	0
SC115996	Horse	female	Puchulia	Horse	2	0
SC115985	Horse	female	Milker	Horse	2	0
SC115979	Horse	female	Puchulia	Horse	2	0
SC115972	Horse	female	Puchulia	Horse	2	0
SC115961	Horse	female	Ascombeguy	Horse	2	0
SC133630	Hand	female	Mayor	Hand	2	0
SC133576	Milker	male	Milker	Milker	2	0
SB21793	Mayor	male	Mayor	Mayor	2	0
SC115997	Horse	female	Horse	Horse	2	0
SC115965	Horse	male	Horse	Horse	2	0
SC115964	Horse	male	Horse	Horse	2	0