Pro Gradu -tutkielma

Heterozygosity-fitness correlations and the maintenance of genetic variation in small populations in *Drosophila littoralis*

Marja Heikkinen

Jyväskylän yliopisto

Bio- ja ympäristötieteiden laitos

Evoluutiogenetiikka

16.8.2010

JYVÄSKYLÄN YLIOPISTO, Matemaattis-luonnontieteellinen tiedekunta

Bio- ja ympäristötieteiden laitos
Evoluutiogenetiikka

HEIKKINEN, M.: Heterozygosity-fitness correlations and the maintenance of genetic variation in small populations in *Drosophila littoralis*

Pro Gradu –tutkielma: 16 s.
Työn ohjaajat: FT Mikael Puurtinen, PhD K. Emily Knott, FM Nina Pekkala
Tarkastajat: Prof. Janne Kotiaho, Prof. Mikko Mönkkönen
Elokuu 2010

Hakusanat: kokeellinen populaatiogenetiikka, mikrosatelliitit, populaatiokoko, sukusiitos

TIIVISTELMÄ

Heterotsygotia-fitnesskorrelaatio voidaan määritellä korrelaatioksi markkerilokuksessa mitatun heterotsygotian ja kelpoisuuteen vaikuttavan ominaisuuden välillä. Näitä korrelaatioita selittämään on esitetty kolme hypoteesia. Ensimmäinen niistä on suoran vaikutuksen hypoteesi, joka perustuu toiminnalliseen ylidominanssiin kelpoisuuteen vaikuttavassa lokuksessa. Toinen hypoteesi, ns. paikallisen vaikutuksen hypoteesi edellyttää kytkentäepätasapainoa markkerilokuksen ja kelpoisuuteen vaikuttavan lokuksen välillä. Kolmas ja viimeinen hypoteesi on yleisen vaikutuksen hypoteesi, jonka perustana on osittaisesta sukusiitoksesta johtuva vaihtelu yksilöiden heterotsygotiassa koko genomin alueella. Useissa tutkimuksissa havaitut korrelaatiot mikrosatelliittiheterotsygotian ja kelpoisuuden välillä johtuvat luultavasti paikallisista tai yleisistä vaikutuksista, sillä mikrosatelliiteilla ei oleteta olevan suoria kelpoisuusvaikutuksia. Sukusiitosheikkoutta pidetään yleisesti korrelaatioiden aiheuttajana ja on esitetty, että mikrosatelliittilokuksissa mitattua multilokusheterotsygotiaa voitaisiin käyttää sukusiitoksen mittarina luonnonpopulaatioissa. Jos heterotsygotia-fitnesskorrelaatiot ovat todellisia, valinnan oletetaan suosivan heterotsygotiaa ja säilyttävän näin geneettistä vaihtelua. Korrelaatioiden taustalla olevat tekijät, kuten lisääntynyt sukusiitos ja lisääntynyt kytkentäepätasapaino, ovat luontaisia pienille populaatioille. Sukusiitos, yhdessä geneettisen pullonkaulan ja satunnaisajautumisen kanssa, lisää kytkentäepätasapainoa. Tutkimuksessani selvitin korreloiko mikrosatelliiteissa havaittu heterotsygotia kelpoisuuteen vaikuttavien ominaisuuksien kanssa *Drosophila littoralis* -kärpäsellä ja mikä mekanismi mahdollisia korrelaatioita selittää. Lisäksi tutkin mitä geneettiselle vaihtelulle tapahtuu pienissä populaatioissa ajan kuluessa. Tutkimuksessa mitattiin eri sukusiitoskertoimen omaaville naaraille jälkeläistuotto ja määritettiin niiden heterotsygotian aste yhdeksässä mikrosatelliittilokuksessa. Jälkeläistuoton ja yksilöllisen heterotsygotian välille laskettiin korrelaatiokerroin. Osalle naaraista laskettiin munien ja jälkeläisten määrä sekä jälkeläisten selviytyminen, mutta osalle vain munien määrä. Lisäksi heterotsygotian vähenemistä populaatiossa verrattiin ennustettuun heterotsygotian vähenemiseen neutraaleissa lokuksissa. Yksittäisissä populaatioissa havaittiin joitakin heterotsygotia-fitnesskorrelaatioita, mutta tarkasteltaessa kaikkia populaatioita korrelaatiot olivat heikkoja ja ei-merkitseviä. Heterotsygotia väheni erikokoisissa populaatiossa neutraaleille markkereille ennustetulla tavalla.

UNIVERSITY OF JYVÄSKYLÄ, Faculty of Mathematics and Science

Department of Biological and Environmental Science

Evolutionary genetics

HEIKKINEN, M.: Heterozygosity-fitness correlations in and the maintenance of genetic variation in small populations in *Drosophila littoralis*

Master of Science Thesis: 16 p.

Supervisors: PhD Mikael Puurtinen, PhD K. Emily Knott, MSc Nina Pekkala

Inspectors: Prof. Janne Kotiaho, Prof. Mikko Mönkkönen

August 2010

Key Words: experimental population genetics, inbreeding, microsatellites, population size

ABSTRACT

A heterozygosity-fitness correlation (HFC) can be defined as a correlation between heterozygosity, measured at a marker locus or at a set of marker loci, and a fitness-related trait. Three hypotheses have been proposed to explain HFCs. The first one is the direct effect hypothesis which relies on functional overdominance in fitness-related loci. Secondly, linkage disequilibrium between neutral markers and fitness-related loci could explain HFCs. This is the so called local effects hypothesis. The third hypothesis is the general effect hypothesis which is based on identity disequilibria resulting from partial inbreeding (variation in heterozygosity among individuals). Many studies have shown correlations between microsatellite heterozygosity and fitness. These could be explained either with local or general effects because microsatellites are putatively free from selection. HFCs are usually thought to be caused by inbreeding depression and it has been proposed that multilocus heterozygosity in microsatellites could be used as an indicator of inbreeding in natural populations. If HFCs are real, it could be assumed that heterozygosity is favoured by selection and genetic variation is thus preserved. The possible causes of HFCs are intrinsic to small populations such as increased inbreeding and increased linkage disequilibrium. Inbreeding, together with genetic bottleneck and drift increase linkage disequilibrium. The purpose of my study was to determine, if there are heterozygosity-fitness correlations in the fly *Drosophila littoralis* and what the mechanism behind possible correlations is. Furthermore, I studied what happens to genetic variation in populations of small size. Reproductive output was measured for females of different inbreeding coefficients and their degree of heterozygosity in nine microsatellite loci was determined. A correlation coefficient was calculated between individual heterozygosity and reproductive output. For some of the females, number of eggs and offspring was counted, but for the others, only the number of eggs was counted. In addition, the observed loss of heterozygosity in the experimental populations was compared to the expected loss of heterozygosity assuming neutrality of the markers. The study gave some indication of heterozygosity-fitness correlations but overall the correlations were weak and non-significant. Loss of heterozygosity in small populations agreed with the theoretical expectations about the rate at which heterozygosity is lost from populations.

Contents

1. INTRODUCTION

A heterozygosity-fitness correlation (HFC) can be defined as a correlation between heterozygosity, measured at a marker locus or at a set of marker loci, and a fitness-related trait (David 1998). HFCs can be explained with overdominance, meaning that a heterozygote has higher fitness than either of the homozygotes. Alternatively, directional dominance could be an explanation, based on the assumption that in heterozygous individuals the deleterious effects of recessive alleles are overshadowed by the effects of dominant alleles (Charlesworth & Charlesworth 1987). With these in mind, three hypotheses have been proposed to explain HFCs. These hypotheses are called direct effect, local effect and general effect hypothesis.

The idea of the direct effect hypothesis is based on functional overdominance at the studied loci, meaning that heterozygotes have higher fitness than homozygotes, which leads to heterozygote advantage that may result in multilocus heterozygosity-fitness correlation (Hansson & Westerberg 2002). In other words, heterozygous individuals have higher fitness compared to homozygous individuals, and the loci in question are under selection (David 1998, Hansson & Westerberg 2002).

The local effect hypothesis is the second plausible explanation for HFCs. According to this hypothesis, linkage disequilibria between neutral markers and fitness related loci are needed (Bierne et al. 1998, Tiira et al. 2006). Linkage disequilibrium means that alleles at different loci are not randomly associated in gametes (Hartl & Clark 1997). This linkage between neutral marker and fitness-related loci leads to associative overdominance of fitness loci and the neutral markers (Pamilo & Pálsson 1998). Associative overdominance means that fitness differences between heterozygotes and homozygotes in neutral locus are due to association of neutral locus with fitness-related loci (David 1998). Consequently, scored neutral loci appear to exhibit overdominance and result in positive heterozygosity-fitness correlation (David 1998, Hansson & Westerberg 2002).

Finally, the general effect hypothesis can also explain HFCs. A prerequisite for the general effect hypothesis is identity disequilibria. Identity disequilibria mainly results from partial inbreeding (Hansson & Westerberg 2002) and depicts a situation where the observed number of multiheterozygotes and multihomozygotes is larger than what is expected from single locus heterozygosities (David 1998). Simply speaking, individuals differ in their degree of heterozygosity due to partial inbreeding and thus they have different inbreeding coefficients (Hansson & Westerberg 2002). The idea of the general effect hypothesis is that heterozygosity in non-coding markers reflects genome-wide heterozygosity and; thus, individuals with homozygous marker loci are likely to also show homozygosity in fitness-related loci (David 1998, Hansson & Westerberg 2002, Slate et al. 2004). As a result, inbred individuals are possible to distinguish from outbred individuals.

In recent years many studies have shown positive correlations between microsatellite heterozygosity and fitness (see Hansson & Westerberg 2002 and David 1998 for review) along with the numerous positive relationships between heterozygosity in allozyme loci and fitness, which have been acknowledged for a long time (see Mitton & Grant 1984 for review). Microsatellites, however, are non-coding markers and are considered to be selectively neutral (Queller et al. 1993, Jarne & Lagoda; 1996, but see also Kashi et al. 1997). Thus, the observed correlations between microsatellite heterozygosity and fitness-related traits are most likely to be due to local effects or general effects, instead of direct effects as in the case of association between allozyme heterozygosity and fitness.

HFCs with microsatellites have usually assumed to be due to general effects (i.e., inbreeding depression) (Pemberton, 2004). The degree of inbreeding has traditionally been

expressed as the inbreeding coefficient, f, which is the probability that the two alleles of a gene in an individual are identical by descent (autozygous) and it measures the reduction of heterozygosity (Hartl 1988). Since a growing body of evidence has shown relationships between heterozygosity in non-coding markers and fitness-related traits (i.e. Coltman et al. 1998, DaSilva et al. 2009), it has been proposed that multilocus heterozygosity (MLH) in microsatellites could be used as an indicator of inbreeding in natural populations where pedigrees for calculation of inbreeding coefficient are difficult to obtain (Coulson et al. 1998, Marshall et al. 2002).

It can be assumed that if heterozygosity and fitness correlate with each other, heterozygosity is favoured by selection. Consequently populations will show a higher degree of heterozygosity than would be expected with increasing inbreeding. One possible mechanism for how loss of genetic variation could be impeded and variation maintained is through balancing selection. Balancing selection favours heterozygotes at the selected loci. The type of balancing selection that could maintain heterozygosity in the case of HFCs is heterozygote advantage also known as overdominance (Frankham et al. 2002, Charlesworth 2006, Hedrick 2007) which is the basis for the direct effect hypothesis. Associative overdominance is also a plausible mechanism for maintenance of genetic variation in small populations (Rumball et all. 1994, Pamilo & Pálsson 1998) as in the case of the local effect hypothesis. However, it should not be forgotten that selection does not work very efficiently in small populations, but its effects are overridden by drift (Frankham et al. 2002).

From conservation genetics' point of view, HFCs are significant because possible causes of HFCs are intrinsic to small and threatened populations, such as increased inbreeding and increased linkage disequilibrium (Grueber et al. 2008). Inbreeding leads to loss of heterozygosity in offspring, which may reduce the fitness of the individual (Hartl & Clark 1997, Bijlsma et al. 2000) since recessive deleterious alleles are expressed in homozygote individuals (Bijlsma et al 1999) and heterozygote advantage is lost from overdominant loci (Charlesworth & Charlesworth 1987). In addition to expression of deleterious alleles and increased homozygosity in inbred populations, and consequently inbreeding depression, the ability of small populations to adapt to changing environmental conditions, also described as the population's evolutionary potential, is threatened due to loss of genetic variation (Reed & Frankham 2003).

As for the disequilibrium between loci, it can be caused by inbreeding but also by other factors. The others are genetic bottleneck, population admixture and drift (Ardlie et al. 2002). The reduction of population size creates a genetic bottleneck (Young et al. 1996) and followed by this initial loss of variation, small and isolated populations continue losing genetic variation due to random genetic drift (Hedrick & Kalinowski 2000). These circumstances enable linkage disequilibrium to arise in small populations.

In this study I concentrate on how heterozygosity correlates with fitness in the fly *Drosophila littoralis*; I investigate whether population size affects the strength of correlations and consider the mechanism for possible correlations. I will also make some conclusions about how quickly heterozygosity disappears from populations of limited size.

2. MATERIALS AND METHODS

2.1 Study population

The flies used in this study originated from 256 wild-caught individuals that were collected near River Tourujoki in Jyväskylä, Finland in 2006. For five generations (P-F4) the flies were mated in random (but avoiding mating within a family) in small plastic vials

(diameter 23.5 mm, height 75.0 mm). Population size was increased to approximately 400 couples in F2. After generation F4 (F5-F6), the flies were reared at population size of 500 couples in plastic bottles where the flies may mate freely (with separate generations, 14-20 couples in a bottle). In the seventh generation (F7) the flies were assigned to smaller subpopulations. From here on, the flies that constituted the first generation after subdivision are signed the F1 generation. 16 subpopulations were started with 10 individuals in every generation, and 12 subpopulations were started with 40 individuals in every generation (called N10 and N40). The control population was started with 500 individuals in every generation. The sex ratio in all populations was 1:1. The flies were cultured on malt medium after Lakovaara (1969). They were kept in plastic bottles in density of five pairs in each bottle, in constant light, with air temperature at 19ºC and humidity at 60 %. For N40, in order to start the next generation, offspring of all subpopulation bottles were mixed and then randomly assigned to new bottles. This ensured that population size was indeed 40 although the flies were maintained under similar densities as flies in the N10 experimental populations. The flies were killed with carbon dioxide and preserved in ethanol.

2.2 Fitness measures

Before the division to subpopulations, lifetime reproductive success (LRS) was measured for 75 female flies. These females are from the 3rd lab generation, four generations before the subdivision. I will refer them as LRS females from now on. In order to measure fitness, one female (age 5 days) and one male (age 13-22 days) were placed into a plastic vial (diameter 23.5 mm, height 75.0 mm) with malt medium. Every second day the flies were moved to a new vial. The male was replaced with a new one (age 13-22 days) every second week, or when found dead or escaped from the vial. Egg and offspring production were measured from each vial, and egg-to-adult survival of the offspring was calculated from these measures.

To measure the fitness of females of the subpopulations, 1 to 6 females from each subpopulation were mated to males of the control population and allowed to lay eggs for a 10 day period. The number of eggs was counted for the last six days in that same time frame. This should give a good estimate of the lifetime reproductive success of these females (Pekkala et al., unpublished results). The age of the flies at the time of mating was 14 to 21 days. The plastic vials were replaced with a new one every second day.

2.3 Microsatellite analysis

Genomic DNA was extracted from EtOH preserved flies. Whole individuals were first air-dried to remove traces of EtOH and then crushed in a microcentrifuge tube with a hand-held pestle. Qiagen DNeasy Tissue Kit reagents were used for extraction following the manufacturer's protocol modified for use with the Kingfisher magnetic particle processor (Thermo Scientific). PCR was done for 11 microsatellite loci (Table 1) that have been proven to be polymorphic in *D. littoralis* (Routtu et al. 2007). PCR reactions were carried out in a volume of 10,5 µl. The reaction mix contained 1X Buffer, 200 µM dNTP's, 10 µM R-primer, 10 µM F-primer, 50 µM $MgCl_2$, 1 unit of Taq DNA Polymerase (Biotools) and 1 µl template DNA. The PCR was done with the following conditions: initial denaturation at 94°C for 3 min followed by 30 cycles of denaturation at 94°C, annealing at 55-58,5°C (see Table 1) and extension at 72°C, and a final extension at 72°C for 10 min using Bio-Rad thermocyclers (C1000 or S1000). The PCR products were denaturated with formamide together with GeneScan™ 500 LIZ™ Size Standard (Applied Biosystems) and then separated with an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems). GeneMapper software version 4.0 was used to score the alleles. Number of individuals

genotyped and their inbreeding coefficients calculated from the pedigrees are shown in Table 2. Only female flies were genotyped.

Table 1 Microsatellite loci, repeat unit, primer sequences and annealing temperature, T_a, for each primer pair.

Locus	Microsatellite repeat	Primer sequences 5´-3´	T_a
Vir4ms	GT	F: ttgcaatattcccccatttc R: gcggcagaaatgacattgac	55°C
Vir11ms	TA	F: gcaaaacatgaataatgcgaac R: ctttgacaatggcaccacac	55°C
Vir32ms	CA	F: gggtgttgatgtcgagtgtg R: aagaagtccaaagcgctcaa	55°C
Vir38ms	CA	F: gaattcgcaatgcacgtaaa R: cgacgtatctgtgagccact	55°C
Vir90ms	GT	F: caattaaaaggaccgcctga R: tcattatgcggaaatgctga	55°C
Vir93ms	GA	F: acgtggtccaagcaatttgt R: tgagctcccgaccagtttag	55°C
Vir99ms	CA	F: acaatgcttgcacaatgacg R: ccatgcaaattgtgaactgc	55°C
Mon6ms	GTT	F: gtccgaaccacgcaataact R: gctgttgatgatgatgaggc	58,5°C
Mon17ams	GT	F: atatctgtgcagaggcaggg R: tgaaattcaagtgcagcgac	55°C
Mon20ms	TG	F: gcagcagccacaatatcaaa R: ggctgctgttgttaaaggct	58,5°C
Mon26ms	TG	F: gagtggcagacacaacctca R: gccaacagtgcacgtaatttt	52°C

Table 2 Populations, number of sampled individuals and their expected inbreeding coefficients.

Population	Sample size	*f*
LRS	75	0
N500	100	0
N10 F11	56*	0,14
N10 F14	29*	0,25
N10 F22	43*	0,50
N40 F20	46*	0,25

*See Table 5 for number of individuals in each subpopulation.

2.4 Statistical and genetic analyses

Five different estimates of individual heterozygosity (IH) were calculated using GENHET program (Coulon 2010). The first of the five estimates calculated was proportion of heterozygous loci (PHt). This is a very simple estimate but it may not be appropriate if the number of markers is small, if the loci differ in number and frequency of alleles or when all the individuals are not typed for the same set of markers (Aparicio et al. 2006). The second estimate used was standardized heterozygosity based on the mean expected heterozygosity (Hs_exp, Coltman et al. 1999). The disadvantage of this estimate is that it gives equal weight to all loci in spite of their allelic frequencies. In addition, the standardization may underestimate the effect of variable loci if the relationship between locus-specific heterozygosity and number of alleles is nonlinear (Aparicio et al 2006). Standardized heterozygosity is based on the mean observed heterozygosity (Hs_obs, Amos 2005). Internal relatedness (IR), proposed by Amos et al. (2001), is based on allele sharing giving more weight to sharing rare alleles than to sharing more common alleles. The

distribution of values is between 1 and -1, with negative values indicating higher heterozygosity and positive values indicating higher homozygosity. However, the problem with this estimate is that its distribution is asymmetrical. The maximum value of 1 can be obtained with all loci being homozygous despite the allelic frequencies but the minimum value can be reached only when there are only two alleles per locus and the individual is heterozygous for all of them. Moreover, IR gives higher values to individuals with rare alleles overestimating their homozygosity (Aparicio et al. 2006). To overcome the problems of IR, Aparicio et al. (2006) have proposed the estimate homozygosity by locus (HL). Since some loci are more informative than others, HL takes into an account the contribution of each locus instead of the contribution of each allele giving more weight to informative loci. In this case, the informative loci are those that have more alleles and those with alleles that are more evenly frequent. This index gives values from 0 to 1, completely heterozygous individuals having value of 0 and completely homozygous individuals having value of 1. It is still noteworthy that indexes that are based on allelic frequencies like IR may be more efficient estimates of genome-wide heterozygosity in populations with high inbreeding, whereas HL may work better with more open populations with migration, admixture of founders or other processes that increase genetic variability.

Correlations between heterozygosity calculated with all five IH measures and all fitness measures were calculated with SPSS 15.0 Program using the Pearson correlation coefficient and two-tailed test of significance. A meta-analysis was done for small populations with several generations of N10 and N40 populations. Only subpopulations with at least three individuals were included in the meta-analysis since correlation coefficients could not be calculated for a fewer number of individuals. For the meta-analysis the correlation coefficient, r, for each subpopulation was computed with SPSS 15.0 Program. These values of r were summed and divided by the number of studies combined to get the Fisher transformed r (Rosenthal 1991). Then, the Fisher Zr was inverted back to r. The variance was calculated Var = 1/(N-3) where N is the number of subpopulations included in the analysis. The square root of the variance was taken to get the standard error. Finally, the Zr was divided by the square root of the variance (Shadish & Haddock 1994) and the P value matching to Z was compared to statistical tables (Zar 1996).

Population genetic parameters including allelic richness and observed heterozygosity were calculated using Fstat Version 2.9.3.2 (Goudet 2001). I used Nei's Gst as an estimate of differentiation of populations. Nei's Gst is an equivalent of Wright's Fst which measures gene differentiation between subpopulations (Nei 1973). Fstat was also used to detect possible linkage disequilibria between loci and possible deviations from Hardy-Weinberg equilibrium. Allelic richness was calculated for the subpopulations that had at least one individual typed for each locus. There were a few occasions where no genetic data could be achieved for a subpopulation i.e. the subpopulation had only one individual typed and some of the loci did not amplify in the PCR.

The expected inbreeding coefficients of the experimental populations were calculated after Crow & Kimura (1970): $f_t = f_{t-1} + (1-2f_{t-1} + f_{t-2})/2N$. Firstly, we assume that two uniting gametes in a finite population cannot come from the same parent. Secondly, the inbreeding coefficient at time t, f_t, and the coefficient of consanguinity of two different randomly chosen individuals at time t, g_t, are introduced to the equation. The inbreeding coefficient at time t equals the coefficient of consanguinity at time t-1, since mating is at random. In order to get the consanguinity coefficient in generation t, it must be noted that the two chosen genes have to come from the same individual in the previous generation with the probability 1/N and from different individuals with probability 1-1/N. If the genes

are from the same parent the probability of identity is $(1+f_{t-1})/2$ and if they are from different parents g_{t-1}. Putting them together we get

$$f_t = g_{t-1},$$

$$g_t = \frac{1}{2N}(1 + f_{t-1}) + \frac{N-1}{N} g_{t-1},$$

substitution of the first one with the second one gives

$$f_{t-1} = \frac{1}{2N}(1 + f_{t-1}) + \frac{N-1}{N} f_t,$$

after this, we go back one generation and rearrange the equation, which is now

$$f_t = f_{t-1} + (1 - 2f_{t-1} + f_{t-2})/2N.$$

3. RESULTS

LRS females did not show any correlations between (IH) and the fitness measures number of eggs, number of offspring, and offspring survival, no matter which estimates of IH were used (Table 3).

However, when the HFCs were examined at the level of subpopulations, one subpopulation in both N10 F22 and N40 F20 populations showed significant correlations, both positive and negative, between fitness and individual heterozygosity, P<0,05 (Table 4), with every estimate of individual heterozygosity. In addition to this, there was one subpopulation in N10 F14 population that showed significant negative correlation between internal relatedness (IR) and fitness (Table 4). Despite this, no significant correlations were found between the number of eggs laid and IH when the meta-analysis of subpopulation correlation coefficients was done (Table 5).

Table 3 Pearson correlations, r, between IH and fitness for LRS females.

IH estimate	Number of eggs		Number of offspring		Offspring survival	
	r	P-value	r	P-value	r	P-value
PHt	-0,001	0,992	-0,026	0,826	-0,085	0,476
Hs_obs	-0,013	0,912	-0,036	0,762	-0,073	0,541
Hs_exp	-0,007	0,951	-0,037	0,755	-0,077	0,518
IR	0,009	0,937	0,025	0,831	0,080	0,499
HL	0,041	0,730	0,056	0,636	0,084	0,478

Table 4 Subpopulation sizes, Pearson correlation coefficients and P-values for different estimates of IH and fitness of small populations.

	Subpopulation	N	PHt		Hs_obs		Hs_exp		IR		HL	
			r	P	r	P	r	P	r	P	r	P
N10 F11	OP00	6	-.004	.994	.075	.887	.071	.894	-.050	.924	-.083	.875
	OP01	5	.638	.246	.662	.223	.666	.219	-.684	.203	-.653	.233
	OP02	4	.586	.414	.635	.365	.576	.424	-.563	.437	-.424	.576
	OP03	4	.878	.122	.901	.099	.916	.084	-.901	.099	-.854	.146
	OP05	4	.706	.294	.765	.235	.779	.221	-.519	.481	-.227	.773
	OP06	5	.625	.259	.727	.164	.709	.180	-.726	.165	-.569	.316
	OP07	5	-.400	.504	-.336	.580	-.318	.602	.232	.707	.380	.528
	OP08	4	.038	.962	-.036	.964	.012	.988	-.004	.996	-.035	.965
	OP09	4	-.853	.147	-.896	.104	-.897	.103	.910	.090	.939	.061
	OP11	5	.109	.862	.145	.817	.146	.815	-.093	.882	.081	.897
	OP12	6	-.264	.614	-.255	.626	-.257	.623	.346	.502	.368	.472
	OP97	4	.797	.203	.429	.571	.536	.464	-.414	.586	-.375	.625
N10 F14	OP00	3	-.995	.064	-.995	.064	-.995	.064	.980	.128	.952	.198
	OP01	3	-.445	.707	-.294	.810	-.323	.791	.340	.779	.384	.749
	OP02	4	.713	.287	.716	.284	.724	.276	-.720	.280	-.743	.257
	OP03	4	.904	.096	.902	.098	.894	.106	-.973	**.027***	-.766	.234
	OP04	4	-.645	.355	-.647	.353	-.653	.347	.586	.414	.579	.421
	OP08	3	.477	.683	.488	.675	.522	.650	-.586	.601	-.063	.960
	OP12	4	-.862	.138	-.862	.138	-.864	.136	.809	.191	.865	.135
	OP98	4	-.545	.455	-.545	.455	-.545	.455	.556	.444	.647	.353
N10 F22	OP00	4	.971	**.029***	.962	**.038***	.959	**.041***	-.964	**.036***	-.974	**.026***
	OP02	5	-.453	.443	-.479	.414	-.484	.409	.480	.413	.492	.399
	OP08	6	-.208	.693	-.208	.693	-.208	.693	.208	.693	.208	.693
	OP11	6	.228	.664	.160	.762	.210	.689	-.166	.754	-.100	.850
	OP12	6	.348	.499	.367	.475	.370	.470	-.392	.442	-.384	.452
	OP97	6	.120	.821	.190	.719	.126	.813	-.066	.901	-.224	.670
	OP98	4	.432	.568	.459	.541	.456	.544	-.322	.678	-.434	.566
	OP99	6	-.641	.170	-.714	.111	-.598	.210	.633	.178	.696	.125
N40 F20	OP13	5	-.946	**.015***	-.951	**.013***	-.951	**.013***	.930	**.022***	.911	**.031***
	OP14	5	.479	.414	.490	.402	.490	.402	-.486	.407	-.638	.246
	OP15	3	-.945	.212	-.926	.246	-.896	.293	.992	.078	.836	.370
	OP16	3	.489	.674	.489	.674	.489	.674	-.758	.452	-.633	.564
	OP17	5	.122	.846	.100	.873	.101	.872	.293	.632	-.248	.687
	OP18	4	-.462	.538	-.405	.595	-.437	.563	.631	.369	.759	.241
	OP19	4	.468	.532	.481	.519	.480	.520	-.520	.480	-.558	.442
	OP21	5	-.579	.307	-.590	.295	-.589	.296	.647	.238	.816	.092
	OP22	4	-.737	.263	-.585	.415	-.573	.427	.321	.679	.545	.455
	OP23	5	.259	.675	.263	.669	.273	.656	-.406	.498	-.748	.146
	OP24	3	-.178	.886	-.469	.689	-.447	.705	.139	.911	.205	.868

* P<0,05

Table 5 Meta-analysis of correlation coefficients, r, between IH and number of eggs laid for experimental populations.

	N10 F11		N10 F14		N10 F22		N40 F20	
	r	P-value	r	P-value	r	P-value	r	P-value
PHt	0.297338	0.085165	-0.38363	0.182984	0.224839	0.159362	-0.34103	0.100112
Hs_obs	0.286797	0.093483	-0.36477	0.19627	0.1942	0.195616	-0.33195	0.106752
Hs_exp	0.301542	0.081992	-0.36765	0.194218	0.212224	0.173779	-0.3159	0.119136
IR	-0.23807	0.138851	0.167489	0.352686	-0.1848	0.20757	0.365153	0.083751
HL	-0.10793	0.313997	0.354761	0.203464	-0.21477	0.170806	0.21003	0.221032

All the loci were polymorphic. The number of alleles varied from 3 to 14 depending on the locus (see Table 6 for allelic richness). There were 17 alleles in the Vir93 locus but it was discarded from the analyses because of difficulties in scoring the alleles correctly. Mon20 was also discarded for the same reason. However Mon20 showed only 13 putative alleles. No significant deviations from Hardy-Weinberg equilibrium were detected for any of the experimental populations with nominal level 5 %, although Fstat implied that there could be something wrong with loci Vir38 and Vir90 in N500 F8 population because the proportion of randomisations that gave a larger F_{is} than the observed was close to significant. Since no other population showed deviations from the Hardy-Weinberg equilibrium for those loci, I included them in further analyses. No significant linkage disequilibrium was found with 5 % nominal level.

The heterozygosity decreased with a rate that was comparable to expected rate of loss (Table 7).

Table 6 Allelic richnesses

	Vir4	Vir11	Vir32	Vir38	Vir90	Vir99	Mon6	Mon17a	Mon26
LRS	6,754	8,668	5,651	3,796	12,413	8,313	4,735	2,000	8,622
N500	8,702	8.304	4.953	3.547	12.169	5.086	4.253	2.806	8.779
N40	1.546	1.832	1.382	1.429	1.849	1.730	1.358	1.255	1.756
N10 F11	1.686	1.816	1.473	1.387	1.841	1.631	1.514	1.295	1.758
N10 F14	1.393	1.822	1.470	1.447	1.845	1.660	1.483	1.300	1.721
N10 F22	1.450	1.829	1.206	1.262	1.756	1.531	1.553	1.056	1.778

Table 7 Populations and their expected and observed loss of heterozygosities.

Population	He	Ho
N10 F11	0,14	0,149
N10 F14	0,25	0,331
N10 F22	0,50	0,485
N40 F20	0,14	0,201

4. DISCUSSION

The main purpose of this study was to explore the presence of association between neutral genetic variation in microsatellites and fitness in the fly *Drosophila littoralis*. Other studies have shown such correlations between microsatellites and fitness-related traits in many species (i.e. Coltman et al. 1998, von Hardenberg et al. 2007, Charpentier et al 2008) but the subject is still very controversial.

The LRS females in this study did not show any correlations between individual heterozygosity and fitness. As mentioned before, in case of HFCs between microsatellite markers and fitness, the most likely explanation would be due to either local or general effects (Hansson & Westerberg 2002). If correlations were found with LRS females, they

most likely would have been due to local effects since LRS females were from a presumably randomly mating population with no inbreeding. In order to reflect genome-wide effects, identity disequilibria created by partial inbreeding would have been needed (Hansson & Westerberg 2002). When it comes to local effects, similarly to identity disequilibria, linkage disequilibrium can be caused by inbreeding but also by various demographic processes such as bottlenecks, population admixture and drift (Ardlie et al. 2002). LRS females in this study did originate from a wild natural population and it can be questioned if the artificial demographic bottleneck that occurred when the flies were brought into the lab was severe enough to cause real linkage disequilibrium between loci.

In comparison to the LRS females, there was one subpopulation in both in N10 F22 and N40 F20 populations that showed significant correlations between heterozygosity and fitness with every estimate of individual heterozygosity. N10 F22 showed negative correlations between IR and fitness and between HL and fitness, but rest of the estimates of individual heterozygosity showed positive associations with fitness. In contrast to N10 F22 population, N40 F20 population showed positive correlations between IR and fitness and between HL and fitness, but rest of the correlations were negative. In addition, N10 F14 population showed significant negative correlation between IR and fitness. There were also some subpopulations that showed almost significant correlations, P-values being <0,10. When different estimates of individual heterozygosity and their association to fitness are compared, IR has more significant or almost significant P-values than the others. This makes sense since, as mentioned before, estimates based on allele frequencies like IR are likely to give better estimates of genome-wide heterozygosity in inbred populations compared to e.g. HL (Aparicio et al. 2006). However, when the meta-analysis was done for the small populations, no significant correlations were found. It is quite intriguing why two subpopulations showed HFC's but the meta-analysis did not find such correlations. The closest to significance were correlations between PHt and fitness, Hs_obs and fitness and Hs_exp and fitness in N10 F11 population. Approximately the same level of significance was found between IR and fitness in the N40 F22 population. It could be expected that with the amount of inbreeding that the small populations have gone through, HFCs caused by genome-wide effects would come forward more clearly. On the other hand, if individuals do not differ enough in heterozygosity, HFCs cannot be detected. In the case of these small inbred populations, linkage disequilibrium caused by both inbreeding and bottlenecks could account for HFCs. As it was not tested, it is difficult to say if the few correlations are due to local or general effects. However, Szulkin et al. (2010) have argued in their recently published paper that linkage disequilibrium is not really an alternative for inbreeding, although, it can result from inbreeding. Thus, identity disequilibria are really needed to create the observed correlations between fitness-related traits and marker heterozygosity.

Correlations calculated only for a few individuals (four individuals in N10 populations, five in N40 population, see Table 4) showed significant correlations in this study, but Coltman & Slate 2003 have argued whether a few dozens of individuals is enough to detect significant associations between multilocus heterozygosity and fitness. In addition, it is likely that a panel of 10-20 microsatellite markers can generate correlations between heterozygosity and fitness in very limited circumstances only (Balloux et al. 2004). Simulations done by Balloux et al. (2004) showed that heterozygosity in a panel of 10-20 markers correlates too poorly with the inbreeding coefficient to create a secondary correlation with fitness. I used only nine microsatellite loci in my study and in this light it can be questioned if the correlations detected in this study are real or some artefact. In addition, heterozygosity-fitness correlations are usually very weak (Britten 1996, David

1998). Coltman & Slate (2003) have also shown that an apparent publication bias exists in the literature in favour of significant results.

In contrast to Balloux et al (2004), the observed loss of heterozygosity followed quite nicely the expected inbreeding coefficients that were calculated from the pedigrees. N10 F14 and N40 F20 populations showed somewhat higher inbreeding than what was expected but this could be due to sampling since observed heterozygosity is more likely to be affected by sample size than expected heterozygosity (Frankham et al. 2002). The loss of heterozygosity agrees with the absence of HFCs. It can be assumed that if heterozygote individuals had higher fitness, the loss of heterozygosity would have been slowed down by selection. Rumball et al. (1994), for example, showed that heterozygosity decreased at a much slower pace than was expected under full-sib and first-cousin mating. This was due to selection favouring heterozygote individuals in inbred lines. Selection is also considered to be responsible for the increase in heterozygosity in an insular mouflon population that was founded with only one pair of individuals (Kaeuffer et al. 2007). On the other hand, it should not be forgotten that selection is not very efficient in small populations (Frankham et al. 2002).

Although, it was not statistically tested, there seemed to be a remarkable drop in allele numbers between LRS females and small populations and between the control population and small populations, which goes hand in hand with the observed loss of heterozygosity. Two factors could explain the reduction both in heterozygosity and allelic diversity (Keller et al. 2001). Firstly, the initial loss of allelic diversity happened when the populations went through a demographic bottleneck, after the subdivision of large population. Secondly, the initial loss was further induced by drift.

In conclusion, this study left many open questions about the existence of heterozygosity-fitness correlations. All in all, due to small sampling size it was not likely that strong correlations could be found. On the other hand, it was positive that the loss of heterozygosity was in accordance with the theoretical expectations. It also makes sense that heterozygosity decreased according to theoretical expectations since no strong correlations were found.

ACKNOWLEDGEMENTS

I thank my supervisors Emily, Nina and Mikael. I am grateful for the people in the DNA lab who helped me when I couldn't find what I was looking for or needed help with the machinery. I also appreciate the financial support from Vanamo ry, Kuopion Luonnon Ystäväin Yhdistys; Betty Väänänen's Fund, Olvi Foundation and the Centre of Excellence in Evolutionary Biology. Last but not least, I thank my family and friends for their support during this project.

LITERATURE

Amos, W. (2005). IR macroN3. IR Manual. http://www.zoo.cam.ac.uk/zoostaff/amos/#ComputerPrograms (19.5.2010)

Amos, W., Worthington Wilmer, J., Fullard, K., Burg, T.M., Croxall, J.P., Bloch, D. & Coulson, T. (2001). The influence of parental relatedness on reproductive success. *Proc. R. Soc. Lond. B* 268:2021-2027.

Aparicio, J.M., Ortego J. & Cordero, P.J. (2006). What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology* 15:4659–4665.

Ardlie, K.G., Kruglyak, L. & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-309.

Balloux, F., Amos, W. & Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* 13:3021-3031.

Bierne, N., Launey, S., Naciri-Graven, Y. & Bonhomme, F. (1998). Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. *Genetics* 148:1893–1906.

Bijlsma, R., Bundgaard, J. & Boerema, A.C. (2000). Does inbreeding affect the extinction risk of small populations?: predictions from *Drosophila*. *J. Evol. Biol*. 13:502-514.

Bijlsma, R., Bundgaard J. & Van Putten W.F. (1999). Environmental dependence of inbreeding depression and purging in *Drosophila melanogaster*. *J. Evol. Biol*. 12:1125-1137.

Britten, H.B. (1996). Meta-analyses of the association between multilocus heterozygosity and fitness. *Evolution* 6:2158-2164.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* 2:e64.

Charlesworth, D. & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Ann. Rev. Ecol. Syst*. 18:237-268.

Charpentier, M.J.E., Williams, C.V. & Drea, C.M. (2008). Inbreeding depression in ring-tailed lemurs (*Lemur catta*): genetic diversity predicts parasitism, immunocompetence, and survivorship. *Conserv. Genet*. 9:1605–1615.

Coltman, D.W., Bowen, W.D. & Wright, J.M. (1998). Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites. *Proc. R. Soc. Lond. B* 265:803-809.

Coltman, D.W., Pilkington, J.G., Smith, J.A. & Pemberton. J.M. (1999). Parasite-mediated selection against inbred soya sheep in a free-living, island population. *Evolution* 53:1259-1267.

Coltman, D.W. & Slate, J. (2003). Microsatellite measures of inbreeding: a meta-analysis. *Evolution* 57:971-983.

Coulon, A. (2010). GENHET: an easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources* 10:167-169.

Coulson, T.N., Pemberton, J.M., Albon, S.D., Beaumont, M., Marshall, T.C., Slate, J., Guinness, F.E. and Clutton-Brock, T.H. (1998). Microsatellites reveal heterosis in red deer. *Proc. R. Soc. Lond. B* 265:489-495.

Crow, J.F. & Kimura, M. (1970). *An introduction to population genetics theory*. Harper & Row, New York, 591 p.

Da Silva, A., Gaillard, J.-M. Yoccoz, N.G., Hewison, A.J.M., Galan, M., Coulson, T., Allainé, D., Vial, L., Delorme, D., Van Laere, G., Klein, F. & Luikart, G. (2009). Heterozygosity-fitness correlations revealed by neutral and candidate gene markers in roe deer from a long-term study. *Evolution* 63:403–417.

David, P. (1998). Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* 80:531-537.

Frankham, R., Ballou, J.D. & Briscoe, D.A. (2002). *Introduction to conservation genetics*. Cambridge university press. Cambridge, 617 p.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at: http://www2.unil.ch/popgen/softwares/fstat.htm.

Grueber, C.E., Wallis, G.P. & Jamieson, I.G. (2008). Heterozygosity-fitness correlations and their relevance to studies on inbreeding depression in threatened species. *Molecular Ecology* 17:3978-3984.

Hansson, B. & Westerberg, L. (2002). On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* 11:2467-2474.

Hartl, D.L. (1988). *Primer of population genetics*. Sinauer Associates, Inc., Sunderland, 305 p.

Hartl D.L. & Clark A.G. (1997). *Principles of population genetics*. Sinauer Associates, Inc., Sunderland, 542 p.

Hedrick, P.W. (2007). Balancing selection. *Current Biology* 7:R230.

Hedrick, P.W. & Kalinowski, S.T. (2000). Inbreeding depression in conservation biology. *Annu. Rev. Ecol. Syst*. 31:139-62.

Jarne, P. & Lagoda, P.J.L. (1996). Microsatellites, from molecules to populations and back. *Trends Ecol. Evol*. 11:424-429.

Kaeuffer, R., Coltman, D.W., Chapuis, J.-L., Pontier, D. & Réale, D. (2007). Unexpected heterozygosity in an island mouflon population founded by a single pair of individuals. *Proc. R. Soc.* B 274:527-533.

Kashi, Y., King, D. & Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics* 13:74-78.

Keller, L.F., Jeffery, K.J., Arcese, P., Beaumont, M.A., Hochachka, W.M., Smith, J.N.M. Bruford, M.W. (2001). Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. *Proc. R. Soc. Lond.* B 268:1387-1394.

Lakovaara, S. (1969). Malt as a culture medium for Drosophila species. *Dros. Inf. Serv*. 44: 128.

Marshall, T.C., Coltman, D.W., Pemberton, J.M., Slate, J., Spalton, J.A., Guinness, F.E., Smith, J.A., Pilkington, J.G. & Clutton-Brock, T.H. (2002). Estimating the prevalence of inbreeding from incomplete pedigrees. *Proc. R. Soc. Lond*. B. 269:1533-1539.

Mitton, J.B. & Grant, M.C. (1984). Association among protein heterozygosity, growth rate, and developmental homeostasis. *Ann. Rev. Ecol. Syst.* 15:479-499.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci. USA* 70:3321-3323.

Pamilo, P. & Pálsson, S. (1998). Associative overdominance, heterozygosity and fitness. *Heredity* 81:381-389.

Pemberton, J. (2004). Measuring inbreeding depression in the wild: the old ways are the best. TREE 19:613-615.

Queller, D.C., Strassmann, J.E. & Hughes, C.R. (1993). Microsatellites and kinship. *Trends Ecol. Evol.* 8:285−288.

Reed, D.H. & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conserv. Biol.* 17:230–37.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Sage Publications, Newbury Park, 168 p.

Routtu, J., Hoikkala, A. & Kankare, M. (2007). Microsatellite-based species identification method for Drosophila virilis group species. *Hereditas* 144:213-221.

Rumball, W., Franklin, I.R., Frankham, R. & Sheldon, B.L. (1994). Decline in Heterozygosity under full-sib and double first-cousin inbreeding in *Drosophila melanogaster*. *Genetics* 136:1039-1049.

Shadish, W.R. & Haddock, C.K. (1994). Combining estimates of effect size. In *The handbook of research synthesis* (Editors Cooper, H. & Hedges, L.V.) Russell Sage Foundation, New York, 573 p.

Slate, J., David, P., Dodds, K.G., Veenvliet, B.A., Glass, B.C., Broad, T.E. & McEwan, J.C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* 94:255-265.

Szulkin, M., Bierne, N. & David, P. (2010). Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* 64:1202-1217.

Tiira, K., Laurila, A., Enberg, K., Piironen, J., Aikio, S., Ranta, E. & Primmer, C.R. (2006). Do dominants have higher heterozygosity? Social status and genetic variation in brown trout, *Salmo trutta*. *Behav. Ecol. Sociobiol*. 59: 657–665.

von Hardenberg, A., Bassano, B., Festa-Bianchet, M. Luikart, G. Lafranchi, P. & Coltman, D. (2007). Age-dependent genetic effects on a secondary sexual trait in male Alpine ibex, *Capra ibex*. *Molecular Ecology* 16:1969-1980.

Young, A., Boyle. T. & Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. *TREE* 11:413-418.

Zar, J.H. (1996). *Biostatistical analysis*. Prentice Hall, Upper Saddle River, 662 p.