

# The power of allele frequency comparisons to detect the footprint of selection in natural and experimental situations

Carolien Gerda Franciska DE KOVEL\*

Section Theoretical Biology, Institute for Biology Leiden, Universiteit Leiden, the Netherlands

(Received 15 March 2005; accepted 16 September 2005)

**Abstract** – Recently, inter-population comparisons of allele frequencies to detect past selection have gained popularity. Data from genome-wide scans are used to detect the number and position of genes that have responded to unknown selection pressures in natural populations, or known selection pressures in experimental lines. Yet, the limitations and possibilities of these methods have not been well studied. In this paper, the objectives were (1) to investigate the distance over which a signal of directional selection is detectable under various scenarios, and (2) to study the power of the method depending on the properties of the used markers, for both natural populations and experimental set-ups. A combination of recurrence equations and simulations was used. The results show that intermediate strength selection on new mutations can be detected with a marker spacing of about 0.5 cM in large natural populations, 200 to 400 generations after the divergence of subpopulations. In experimental situations, only strong selection will be detectable, while markers can be spaced a few cM apart. Adaptation from standing variation in the base population will be hard to detect, though some solutions are presented for experimental designs.

**artificial selection / hitchhiking / linkage disequilibrium / natural selection / power**

## 1. INTRODUCTION

An important step in understanding adaptation is to identify the number and location of genes that are involved in the process. One option to do this is to examine patterns of variation within populations, *e.g.* by using Tajima's D [18]. Reduced intra-specific variation is the mark of selection sought in this method. Similarly, reduced variation can be established in inter-population comparisons, for example with the lnRV-method [10, 14]. Both of these methods rely on hitchhiking of neutral markers with a selected site. Also based on hitchhiking are methods that compare allele frequencies between populations that have

---

\* Corresponding author: dekov@rulsfb.leidenuniv.nl

presumably experienced different selection regimes or that have come across different mutations to respond to the selection regime. In this approach, divergence of allele frequencies beyond that expected by random processes is the characteristic that is looked for. Though frequencies of a targeted mutation may be compared (*e.g.* [12]), more often the pattern is looked for in neutral markers. Recently, a number of studies have used this second strategy to detect the footprint of selection from a genome-wide distribution of polymorphisms, either in natural populations [3, 4, 13, 19] or in lines under artificial selection [6, 17]. In this paper, I investigate the potential and limitations of the method, and I will discuss them with reference to published data.

Populations that live and evolve in isolation of each other diverge genetically. This is partly due to the random forces of genetic drift and partly to selection that may act in different directions in different lineages. Random genetic drift, under selective neutrality, will affect all loci across the genome in a similar and predictable manner. Natural selection will act at specific loci and can cause detectable deviations from the pattern caused by drift [5]. Directional selection in one of the lineages will cause divergence of allele frequencies to exceed the variation caused by drift. Most of the studies that employ this principle to detect selection use neutral markers and rely on linkage disequilibrium between the markers and the true target of selection.

The hitchhiking of neutral loci with selected loci is a well studied subject [9, 16]. Yet within the current context of population comparisons for allele frequency differences, some issues are left about the required marker spacing and power of the method. In an earlier study, Beaumont and Balding [4] investigated the efficiency of their version of this method when markers are completely linked to the selected locus. In this paper, I describe the behaviour of markers at increasing recombination distance from the selected mutations for a scenario of large, isolated populations over a few hundred generations, and for an experimental scenario in which small laboratory populations experience strong selection during a few dozen generations. In addition to new mutations that experience selection right upon origination, I investigated situations in which selection acts on standing variation. The following questions need answering: What is genetic distance over which the signal extends? What is the power of the method? Does it detect all kinds of mutations equally well? What would be the optimal set-up for comparisons of natural populations or for comparisons in selection experiments? Other issues, such as alternative causes for the pattern of deviating allele frequencies have been discussed in the literature (*e.g.* [2]), and will not be considered here.

During selection, alleles at neutral marker loci hitchhike along with the target. If these alleles are initially in linkage disequilibrium with the favourable mutation, this will result in allele frequency changes at the marker locus. When selection stops, *e.g.* because the mutation has reached fixation, decay of linkage disequilibrium  $D$  proceeds at the well-known rate  $D(1 - r)^t$ . However, though  $D$  between the mutation and the surrounding markers will decline, the allele frequencies of marker alleles that hitchhiked will still be elevated. Unlike in association studies, ongoing recombination does not disturb the pattern of allele frequencies any further. The main threat to detection is that the signal gets ‘drowned’ by random fluctuations in allele frequencies at neutral positions. Since the selected mutation will after some time no longer be associated with the marker allele, further use of these markers in *e.g.* phenotypic comparisons within the population will not be possible, though in an experimental design this will be different.

## 2. METHODS

### 2.1. Hitchhiking dynamics

The expected dynamics of a neutral marker allele (B) hitchhiking with a new, single beneficial mutation (A) can be described by the following recurrence equations:

$$\begin{aligned}\bar{w}(t) &= p_A^2(t) w_{AA} + 2p_A q_A w_{aA} + q_A^2 w_{aa} \\ I(t) &= \frac{p_A(t) - p_{B,l}(t)(1 - \theta)}{1 - p_{B,l}(t)(1 - \theta)} \\ p_B(t) &= p_{B,l}(t) + p_{B,u}(t) \\ p_{(B,u)}(t+1) &= (p_{B,u}(t) + \theta \cdot p_{B,l}(t)) * \left[ \frac{I(t)(p_A(t) w_{AA} + q_A(t) w_{aA}) + (1 - I(t))(p_A(t) w_{aA} + q_A(t) w_{aa})}{\bar{w}} \right] \\ p_{B,l}(t+1) &= p_{B,l}(t)(1 - \theta)(p_A(t) w_{AA} + q_A(t) w_{aA}) / \bar{w}.\end{aligned}\tag{1}$$

With  $p_{B,l}$  the fraction of B-alleles that are still in their original coupling with the A-alleles, and  $p_{B,u}$  the fraction of B-alleles that were not linked or have become unlinked to A, but may have become associated with A secondarily. Furthermore,  $q_A = 1 - p_A$ ,  $w_{ij}$  is the fitness of genotype  $ij$  at the A-locus, and  $\theta$  is the recombination fraction. The intermediate variable  $I$  should be interpreted

as the fraction of unlinked B-alleles that have secondarily become associated with an A-allele.

The behaviour of allele B can be compared with the 95% and 99% confidence boundaries for allele frequencies under random genetic drift, calculated from the variance in allele frequencies between two isolated lineages [8]:

$$\sigma^2 = pq \left( 1 - \left( 1 - \frac{1}{2N} \right)^t \right). \quad (2)$$

Initially, allele frequencies follow a normal distribution. When fixation occurs, this distribution becomes distorted. An approximation for expected allele frequencies can be found in Crow and Kimura [7].

## 2.2. Simulations

The recurrence equations only describe the average, but for power, we also need the distribution of the marker allele frequencies and we need to include effects of drift. I used individual-based Fisher-Wright simulations to assess the fraction of  $F_{st}$ -values exceeding the neutral expectation under various scenarios. This also allowed me to include scenarios in which the new mutation is not unique, and in complete equilibrium with surrounding markers in the base population, a scenario that may be more common than the new mutation scenario (see also [1]).

In the simulations, populations of organisms with separate sexes were simulated. They possessed diploid genomes, consisting of one functional locus with a beneficial mutation and a wildtype allele, surrounded by 28 bi-allelic marker loci (like SNP). The recombination fraction between neighbouring loci was  $\theta$ , so the whole genome, or actually chromosome fragment, had a length of  $100 \times 28 \times \theta$  cM (without interference). I assumed a selection model in which homozygote females for the mutation have fecundity  $1 + 2s$ , and heterozygotes  $1 + 2hs$ . The mean fitness of the population was calculated from these selection coefficients assuming Hardy-Weinberg equilibrium, and the mean number of offspring per female was adjusted accordingly to keep the population size constant. Population size  $N$ , number of generations, selection coefficient  $s$  and dominance coefficient  $h$ , as well as initial allele frequencies of the marker SNP and the functional locus could be varied. No mutation was assumed at either markers or functional locus.

To quantify population divergence, the Wright fixation index  $F_{st}$  [20] was calculated for each locus. In each population, the allele frequency  $p_{a,1}$  for allele

$a$  at locus  $l$  was counted for in the whole population, after which expected heterozygosities were computed as standard. Then for each locus:

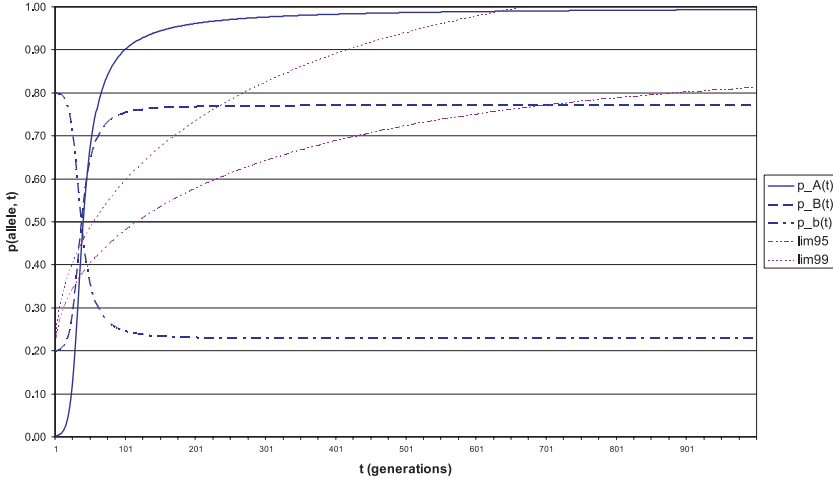
$F_{st} = \frac{H_T - H_S}{H_T}$  where  $H_S$  is the average expected heterozygosity for the separate populations, and  $H_T$  the expected heterozygosity in the combined populations.

The setup was to compare a population experiencing directional selection with a population experiencing no selection at the focal locus, and to study divergence of marker loci between those populations at increasing distances from the focal locus. First, for reference, the distribution of F<sub>st</sub>-values without selection had to be established. Therefore, for each set of conditions, 1000 simulations were run without selection, with each simulation run representing the evolutionary course of a single population. Then pairwise comparisons between the runs without selection were made. Per pairwise comparison, this resulted in 28 F<sub>st</sub>-values, since 28 marker loci were present. As threshold values for the simulations under selection, I used the 99th percentile point of the F<sub>st</sub> distribution without selection. F<sub>st</sub>-values exceeding this threshold were assumed significant. Once the reference was established, selection was included for another 1000 runs with otherwise the same parameters. Pairwise comparison of selected and unselected lines provided F<sub>st</sub>-values with directional selection in one of the populations. For each parameter combination, 1000 runs were done from the same starting conditions, and 1000 pairwise comparisons were made for 28 loci each.

I considered two general scenarios: one was a large population under natural selection; the other was a small population in an experimental design, with short-lasting strong selection.

### 3. RESULTS

Figure 1 shows the typical result of the recurrence equations. It shows that the linked marker allele (B) increases in frequency less strongly than the focal allele (A), while the unlinked marker allele (b) naturally decreases in frequency in proportion. The thin lines represent the 95% and 99% confidence limits over time for the expected allele frequencies without selection, when compared to the initial frequency of allele B, using the inverse of the normal distribution with  $\sigma^2$  as in equation (2). Early after the start of selection, the effect of selection is too weak to be noticed, and the expected allele frequency of the marker locus stays within the neutral expectation. Then selection takes flight and exceeds the effect of random divergence. After a number of generations, in this case about 200, the effect of random divergence (lim99) catches up with

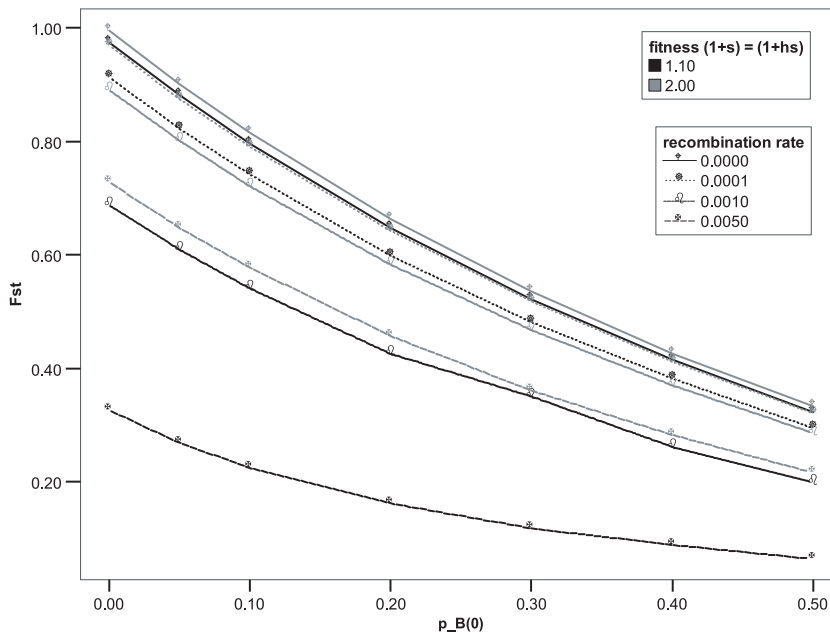


**Figure 1.** Recurrence equation results. The frequencies of the selected mutation (solid line), the hitchhiking marker allele B (dashed) and the alternative marker allele b (dotted) over time. The thin lines represent the upper 95% and 99% boundaries of the expected distribution of allele frequencies under drift (Eq. (2)). In this (unfavourable) scenario, selection can only be detected in the window between 40 and 200 generations.  $N = 250$ ,  $h = 1$ ,  $s = 0.2$ ,  $p_B(0) = 0.20$ ,  $r = 0.005$ .

selection, and the frequency of allele B falls again within the range expected under neutral divergence. To demonstrate the notion of a ‘window of detection’ clearly, I have chosen a quite unfavourable set of parameters, as shown in the figure caption.

### 3.1. Unique mutations

If a new favourable mutation arises near an already polymorphic marker, it will initially be linked to one of the alleles at that marker. By chance, this new mutation is more likely to arise near a common marker allele than near a rare marker allele. From a detection point of view, this is unfortunate, because a common allele can only become a bit more common by hitchhiking, while a rare marker allele may become much more common, which increases the strength of the signal that the marker generates. Weak signals (small effects of selection on allele frequency change) are easily drowned in noise. Figure 2 shows the effect of initial allele frequency on  $F_{st}$  in generation 1000 for different values of  $p_B(0)$  and different distances between the mutation and the marker, according to the recurrence equations. It can be seen that when the



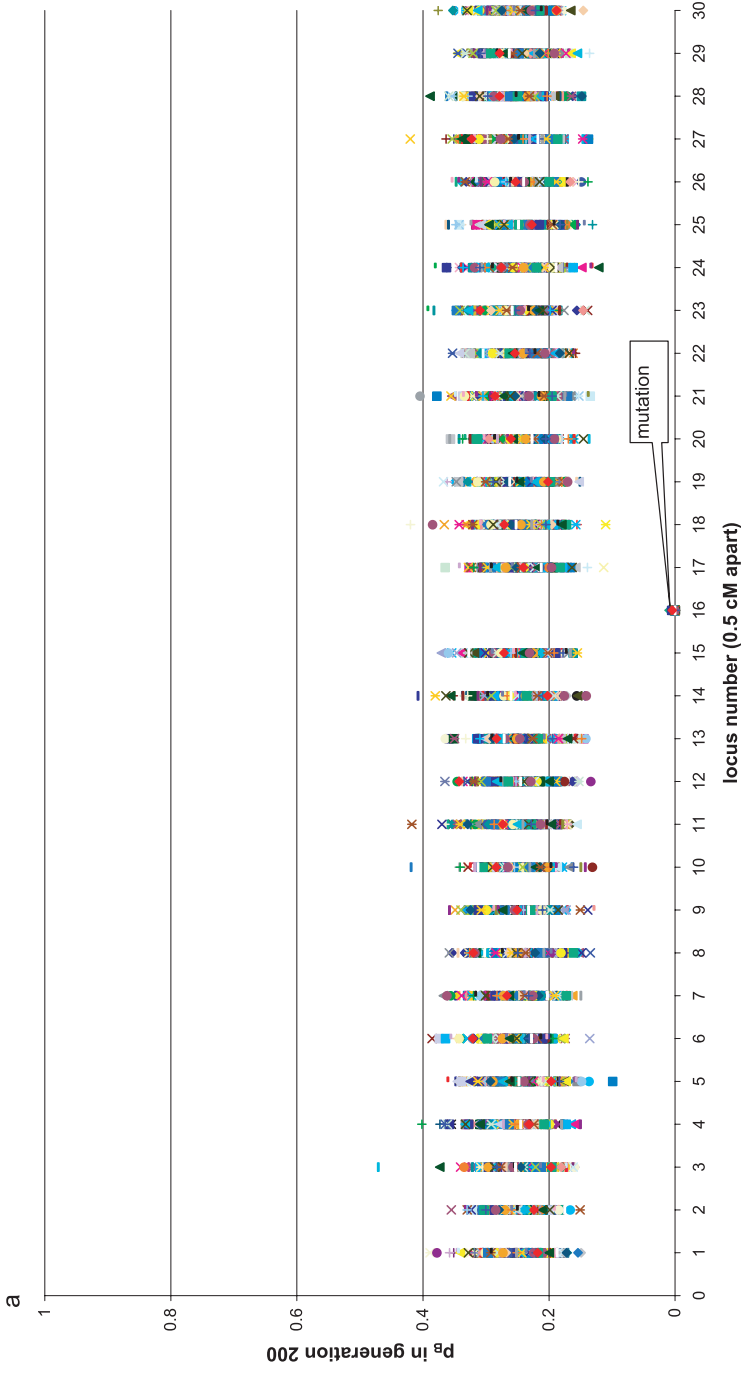
**Figure 2.** Recurrence equation results.  $F_{st}$  in generation 1000 decreases with increasing frequency of the initially linked marker allele  $p_B(0)$ . Markers close to the selected mutation (dot) achieve higher  $F_{st}$ -values than markers further away (triangle), and this effect is stronger for weak selection (black symbols) than for stronger selection (grey symbols). ( $N = 10\,000$ ,  $h = 1.0$ , generations 1000.)

marker allele had  $p_B(0) = 0.5$ , even completely linked markers give only a moderate  $F_{st}$ -value.

The recurrence equations can show the farthest point where the frequency of a hitchhiking allele is expected to exceed the 99% confidence limit of allele distribution without selection. As said before, this should give a rough indication of the limit to detection possibilities. Under the simulation conditions below ( $N = 10\,000$ ,  $s = 0.6$ , generations 200), the detection limit for 0.5:0.5 markers would be 5.0 cM. For 0.25:0.75 markers, it would be 7.6 and 3.0 cM respectively.

In the simulations, the effect of the probability of arising near a rare allele and the strength of the generated signal can be combined, and we can see what the detection probability would be if the markers that are used for the scan had initially certain allele frequencies.

In Figure 3 it can be seen (for parameters Tab. I, nr 12) how allele frequencies of the initially rare marker allele B ( $p_B(0) = 0.25$ ) change when selection acts on a unique mutation. In panel a, the allele frequencies of the marker



**Figure 3.** Simulation results of 200 replicate runs. (a) Allele frequencies of a marker allele with initial frequency  $p_B(0) = 0.25$  after 200 generations without selection. Recombination  $r$  between loci  $= 0.005$ ;  $N = 10\,000$ ;  $s = 0.0$ ;  $h = 1.0$ . The mutation is on locus 16, and was present with one copy in the base population. (b) The same, but now the mutation was under selection:  $s = 0.6$ . Only simulations in which the mutation had attained high frequencies were used.



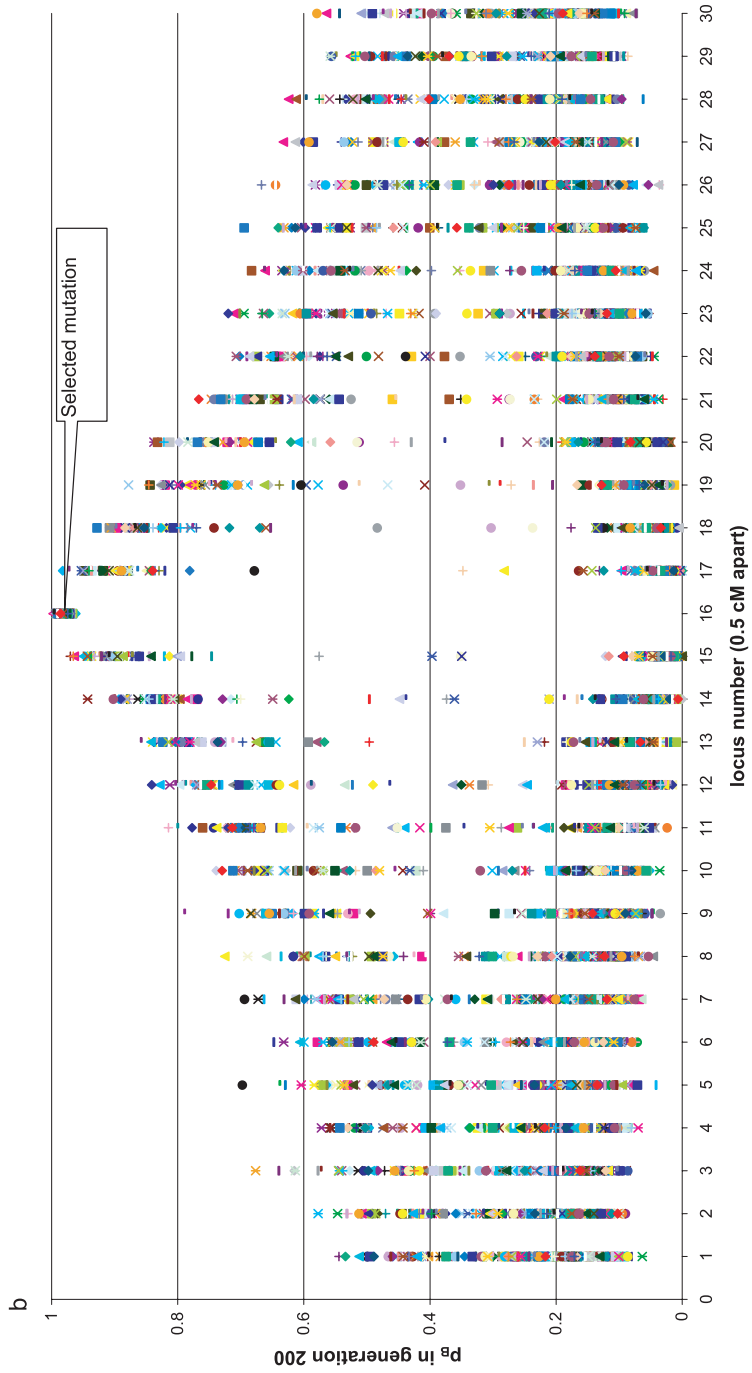


Figure 3. Continued.

**Table I.** Simulation results for natural population scenarios. The frequencies of runs (replicates = 1000) with  $F_{st}$ -values exceeding the 99th percentile point of  $F_{st}$ -values under neutral evolution, and the maximum obtained  $F_{st}$ -value for markers at various distances from the selected mutation for different sets of parameters. Scenarios vary with respect to the dominance of the mutation ( $h$ ), the selection coefficient ( $s$ ), the initial frequency of the favoured mutation ( $p_A(0)$ ), and of the rarer marker allele ( $p_B(0)$ ), and the number of generations. Simulations were run with different inter-marker distances ( $\theta$ ), either 0.005 between adjacent loci or 0.001. As a consequence, some parameter sets miss entries for 0.001–0.004, while others miss entries for locus distances 0.015–0.05.

nr	$h$	$s$	Population size	$p_B(0)$	$p_A(0)$	gen	$\theta =$	$\theta =$	$\theta =$	$F_{st}$ (max)
							0.001–0.004	0.005–0.01	0.015–0.05	
1	1.0	0.6	10000	0.5	Unique	200	-	98	69	0.424
2	1.0	0.6	10000	0.1	0.0005	50	75	75	47	0.335
3	1.0	0.6	10000	0.1	0.0005	100	70	50	17	0.787
4	1.0	0.6	2000	0.1	0.0025	100	40	36	-	0.607
5	1.0	0.6	10000	0.1	0.0005	200	-	47	19	0.624
6	0.5	0.6	10000	0.1	0.0005	200	-	48	19	0.591
7	1.0	0.6	10000	0.1	0.05	200	2	2	-	0.097
8	0.0	0.6	10000	0.1	0.20	200	2	2	-	0.083
9	1.0	0.6	10000	0.1	0.0005	500	-	24	12	0.808
10	1.0	0.6	10000	0.25	0.0005	50	-	44	23	0.265
11	1.0	0.6	10000	0.25	0.005	50	-	32	17	0.107
12	1.0	0.6	10000	0.25	Unique	200	-	73	41	0.620
13	1.0	0.6	2000	0.25	Unique	200	-	38	19	0.933
14	1.0	0.6	10000	0.25	0.0005	200	-	45	22	0.406
15	1.0	0.6	10000	0.25	0.05	200	3	2	-	0.119
16	1.0	0.6	10000	0.5	0.0005	200	-	53	30	0.381
17	1.0	1.2	10000	0.1	0.01	50	-	11	10	0.059
18	1.0	1.8	10000	0.1	0.0005	200	-	42	24	0.289
19	1.0	0.6	2000	0.5	Unique	200	-	65	52	0.674

alleles are shown after 200 generations without selection. In panel b it can be seen that marker alleles that are in close linkage with the selected mutation in the centre hitchhike along, either up (if the mutation was linked to allele B) or down (if the mutation was linked to the alternative allele b). Overlying the panels shows the significant deviations, mainly close to the mutation, and mainly the alleles that have hitchhiked up, and therefore when the rare marker allele was originally linked to the mutation. Only very few marker frequencies are left in the centre. These are the cases when the mutation has recombined off its haplotype early.

In Table I (nr 1) it is shown that for bi-allelic markers with initial allele frequency 0.5:0.5, 98% of the markers at a distance between 0.5–1.0 cM from the mutation give a detectable signal ( $N = 10\,000$ ,  $s = 0.6$ , generation 200), and 69% of the markers 1.5–5.0 cM still do so. This is because it does not matter with which marker allele the mutation is linked originally. Nearer to the limit of 5 cM calculated from the recurrences, the random noise has some runs disappear below the detection threshold. If the marker initially had alleles with frequencies 0.25:0.75 (Tab. I, nr 12), 73% of the marker loci between 0.5–1.0 cM give a signal, and 41% of the loci between 1.5–5.0 cM do so. This shows the difference between the full simulations and the expectation from the recurrence equations that up to 3 cM all markers would respond and up to 7 cM 25%. The results show how the strength of the F<sub>st</sub>-signal not only depends on  $s$  (and population parameters), but also on the initial allele frequency at the marker locus.

In an experimental scenario ( $N = 250$ ,  $s = 3.0$ , 25 generations), detection, according to the equations, is possible up to no further than 0.5 cM when  $p_B(0) = 0.75$ , but to 7 cM if  $p_B(0) = 0.50$ , and up to 14 cM if  $p_B(0) = 0.25$ . Therefore, for 0.25:0.75 markers, only the rare allele (25% of the loci) is likely to give some signal beyond 0.5 cM, but will do so over quite some distance. In the simulations (Tab. II), it is shown that for 0.50:0.50 markers, 67% of the marker loci between 0.5 and 1.0 cM gave a signal, and 45% of the markers between 1.5–5.0 (nr 12). For 0.25:0.75 markers, 45% and 33% of the markers at 0.5–1.0 and 1.5–5.0 respectively indicated the presence of selection (Tab. II, nr 8).

The correspondence between recurrence equations and simulations is quite good. The simulations give a better idea of power, but in general, the recurrences can give an approximation of the distance over which the signal extends on average.

### 3.2. Standing genetic variation

In the previous section, I investigated the probability of being able to demonstrate selection with various types of markers at increasing distances from a mutation, if this mutation were unique to the population under selection. What, however, if selection acts on some mutation that is present in linkage equilibrium in the original population, before sub-populations become isolated, and which subsequently becomes favourable in one of the populations? It may be a mutation that was originally effectively neutral or in mutation-selection balance.

**Table II.** Simulation results; experimental scenarios (see Tab. I for explanation).

nr	$h$	$s$	Population size	$p_B(0)$	$p_A(0)$	gen	$\theta =$	$\theta =$	$\theta =$	$F_{st}$ (max)
							0.001–0.004	0.005–0.01	0.015–0.05	
1	1.0	0.6	250	0.10	0.02	25	-	8	2	0.599
2	0.5	0.6	250	0.10	0.02	25	-	6	6	0.705
3	1.0	0.6	250	0.10	0.10	25	-	2	2	0.328
4	1.0	0.6	250	0.25	0.02	25	9	7	-	0.604
5	1.0	1.2	250	0.10	0.02	25	-	6	5	0.696
6	1.0	2.0	250	0.25	Unique	25	-	41	30	0.913
7	1.0	3.0	250	0.10	0.01	25	10	9	-	0.956
8	1.0	3.0	250	0.25	Unique	25	-	45	33	0.863
9	1.0	3.0	250	0.25	0.01	25	25	23	-	0.851
10	1.0	3.0	250	0.25	0.1	25	-	5	3	0.490
12	1.0	3.0	250	0.5	Unique	25	-	67	45	0.859
13	1.0	3.0	250	0.5	0.01	25	33	32	-	0.714
14	1.0	3.0	250	0.5	0.10	25	4	4	3	0.656
15	1.0	3.0	500	0.25	Unique	25	-	66	50	0.829
16	1.0	3.0	250	0.25	Unique	40	-	30	20	1.000
17	1.0	3.0	250	0.25	Unique	15	-	44	38	0.614

If the favourable mutation is rare, with 10 copies in a population of 10 000 individuals (20 000 alleles), does detection become more difficult than for unique mutations? Table I shows various sets of conditions for  $p_A(0) = 0.0005$ . If we compare nr 16 with nr 1, with the only difference the frequency of the mutation, we see that the number of runs in which a marker indicates selection has been halved. For a different marker type, compare Table I nr 12 and nr 14, and again the power has been reduced by 40–50%. Similar results are found for 0.10:0.90 markers (nr 5).

In Figure 4 (Tab. I, nr 14), allele frequencies of a marker allele with initial frequency  $p_B(0) = 0.25$  are shown after 200 generations without (panel a) or with (panel b) selection on the focal locus. In contrast to the case with a unique mutation (Fig. 3), many allele frequencies are stuck in the middle. These are mainly cases in which both copies of the mutation that were linked to  $B$ , and copies that were linked to  $b$  have been retained and have responded to selection.

In the experimental setup, increasing the favourable mutation from 1 to 5 copies also halves the detection probability, as can be seen by comparing Table II nrs 8 and 9 or nrs 12 and 13. Further increases in allele frequency of the mutation causes dramatic drops in detection (Tabs. I and II).

### 3.3. Effects of population properties

#### 3.3.1. Population size

The larger the populations are, the smaller is the effect of drift on driving populations apart. This makes detection of selection easier in large than in small populations. In Table I, a five-fold decrease in population size from 10 000 to 2000 (nrs 12 and 13, nrs 1 and 19) reduced detection. When originally ten mutations were present, the detection dropped with 30 to 40% (nrs 3 and 4), with the exact numbers depending on the remaining parameters.

In the experimental set-up, lowering the population size to 100 dramatically decreased detection power (not shown). The detection of selection in populations of 500 individuals was about 50% more than in populations of 250.

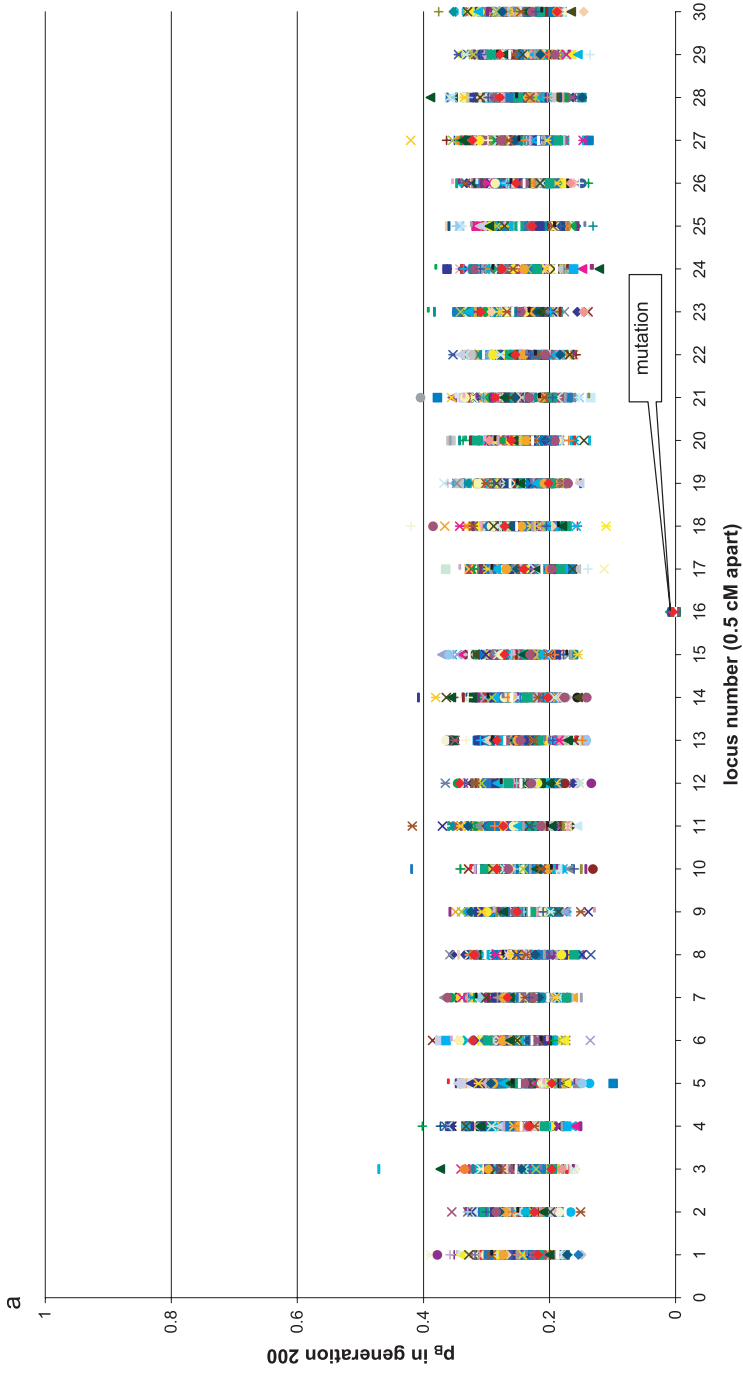
#### 3.3.2. Number of generations

As shown in Figure 1, looking too soon can lead to low detection probability, although this situation will rarely occur in natural populations. Looking too late, may have the effect that random allele frequency changes have caught up with the selection signal. Compare *e.g.* Table I, nrs 1, 2, 4 and 29 at large distances. The effect is, however, not yet visible in the comparison between nr 10, and nr 14, where a different type of marker was used.

In the experimental set-up, fewer generations may mean that, depending on selection strength, the mutation has not yet reached high enough frequencies, so neither have the linked marker alleles. The recurrence equation will show, however, that mutations that have a strong enough effect to be detected pass through this phase rather quickly. The probability of missing mutations because drift effects have drowned them is much higher, especially in small populations (Tab. II, nrs 8, vs. 16 and 17). Since effect size of mutation(s) will be unknown in advance, probably the best moment to start genotyping would be as soon as selection response has reached a plateau.

### 3.4. Effects of selection strength

Obviously, the stronger a mutation is selected, the faster it increases in frequency, and the less possibility recombination has to break down the association between the mutation and the surrounding marker alleles. This means that mutations that confer a stronger positive effect on the bearer, especially on the heterozygote, will be easier to detect. This effect will be stronger for markers



**Figure 4.** Simulation results. As in Figure 3, but here the mutation was present with a frequency of 0.0005 initially in linkage equilibrium with surrounding markers. Note that there is no open space in the centre in comparison to Figure 3.

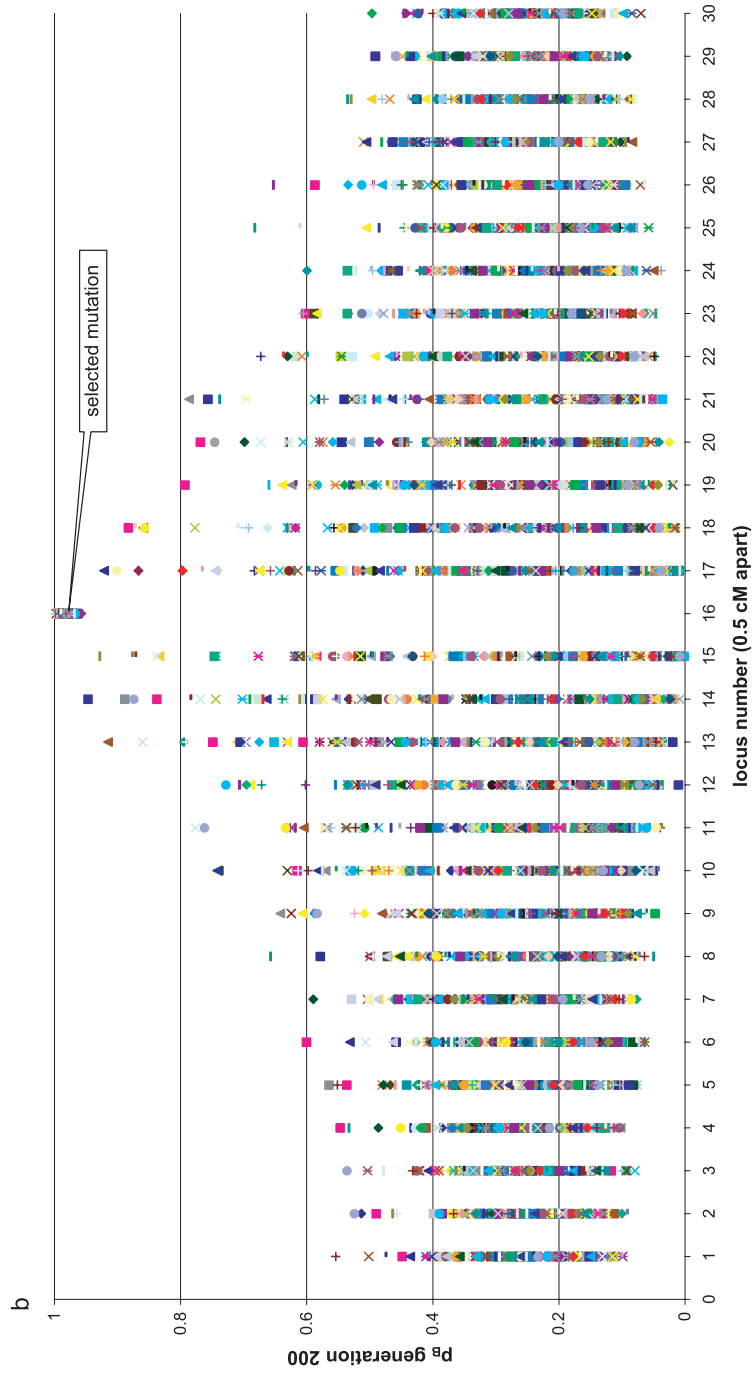
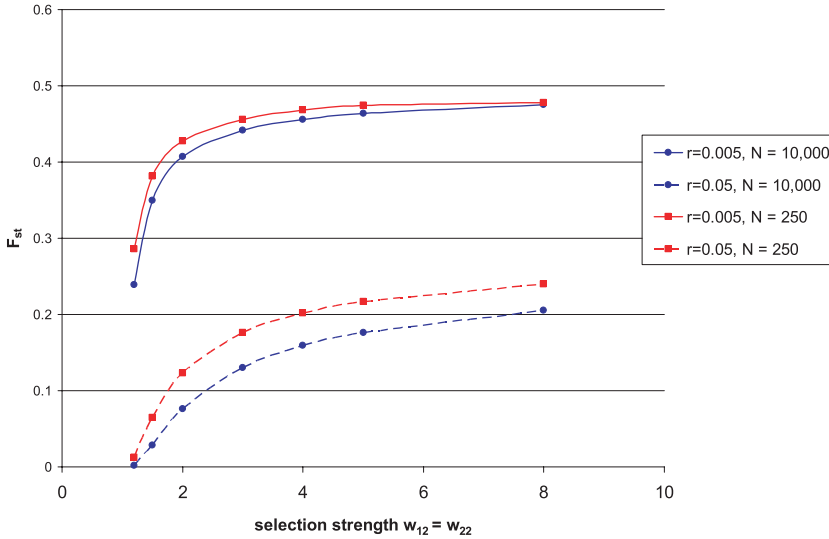


Figure 4. Continued.



**Figure 5.** Recurrence equation results.  $F_{st}$  in generation 1000 for two population sizes and two recombination frequencies as a function of selection strength acting on the linked mutation.

at a greater distance from the mutation (Fig. 2). In general, recessive mutations will not be detected (Tab. I, nr 8).

Expected  $F_{st}$ -values increase in a decelerating way with increasing selection strength (Fig. 5). Once selection is strong enough to pull the  $F_{st}$ -values well past the threshold value, higher selection strength has little effect on detection probabilities, and increasingly smaller effects on absolute  $F_{st}$ -values. In small populations, the detection threshold lies higher, because of larger drift effects, and therefore only more strongly selected mutations can be detected in small populations. The same is true if many generations separate the populations (compare Tab. I nrs 4 and 9).

### 3.5. Comparison to published data

In the study by Akey *et al.*, [3] 26 530 SNP were used, with the majority having a minor allele frequency  $\geq 0.20$  (at the time of study) in at least one of the three populations – Asian, European, Afro-American – under study (average frequency 0.4). Average marker spacing was about 0.1 cM. Average pairwise  $F_{st}$  (recalculated from raw data) was 0.07, a value reflecting a longer time since divergence (or smaller population sizes) than computer time would allow for the simulations. Consequently, they used  $F_{st} = 0.45$  as their threshold, which



corresponded to an empirical significance level of  $\alpha = 0.026$ . They detected 174 candidate genes, with on average 1.5 SNP per candidate. The highest pairwise F<sub>st</sub> Akey *et al.* found was 1.0, meaning their samples were fixed for alternative alleles (in samples of 42 individuals per population). Though F<sub>st</sub> was weakly correlated between neighbouring sites up to about 200 kb or roughly 0.2 cM (their Fig. 4), significantly diverged markers were often flanked by markers with very low F<sub>st</sub>-values (*e.g.* their Fig. 5a).

Pairwise F<sub>st</sub>-values were quite often similar in two out of the three inter-population comparisons, with one pairwise comparison deviating from the others. This suggests that one of the populations was selected away from the two others. A new mutation at some moment appearing in the diverging population could very well have caused such a pattern. However, patterns with three different pairwise F<sub>st</sub>-values were present as well. If these truly reflect selection rather than drift, a possible scenario would be that the same mutation, segregating in the base population before splitting, was selected at different intensities in two of the populations. An equally likely explanation, though, is that, owing to random forces, such a mutation showed different degrees of linkage disequilibrium with the surrounding marker alleles in the different sub-populations, while selection pressures acted the same in two or three of the populations. Different mutations in the same or close-by genes, appearing in different populations represent a third possibility. A scenario of two-way diversifying selection on an already common mutation would be improbable to generate the ‘footprint of selection’, since it would require the functional mutation to be so common that useful LD becomes unlikely. Still, (near) fixation for alternative alleles was found at least three times. The easiest explanation here would be linkage disequilibrium between a new beneficial mutation and a very rare marker allele. Without selection, this allele went to extinction or apparent extinction in some of the sub-populations. The appearance of a new marker allele near a selected site in one sub-population is also a possibility, though it would require a rather fortuitous timing of events.

For the experimental scenario, the noise due to small population size makes detection of selection less probable per marker. However, the signal extends over a larger distance. With a relatively low density of markers, it will still be possible to detect signals of selection, though the position cannot be determined accurately. That may require fine-mapping in later generations. In an experimental set-up, which I followed more or less in the simulations, Colson [6] used the allele frequency method to detect selection for tolerance to octanoic acid in *Drosophila*. The main differences between her work and this paper were that she used multi-allelic microsatellites and that she started from a mixture of

inbred lines rather than from an outbred population. An important difference was also that she used three selection lines *vs.* three controls, where I used pairwise comparisons. Because the experiment was started from inbred lines, it is quite likely that the beneficial mutations were uniquely associated with particular haplotypes deriving from specific lines, and that they were present in multiple copies at outset. Clearly, this, as well as the use of multi-allelic markers, increases the probabilities of detection much above the situations I described in this paper.

#### 4. DISCUSSION

I investigated the probability of detecting past diversifying selection between two populations by the method of comparing allele frequencies of neutral marker alleles, with one of the populations experiencing directional selection, while the other does not. I used both recurrence equations and simulations to assess the probability of detecting selection. In general, correspondence between the methods was fine, and the recurrences can be used to get an idea of a study's ability to detect selection on *new* mutations. The 'F<sub>st</sub>-method' can detect selection on a new or initially very rare beneficial mutant, but is very likely to miss selection that acts on a mutation that was segregating in the populations in slightly higher numbers at linkage equilibrium, and suddenly became favoured in one population, because of *e.g.* a changing environment. Functional alleles that were neutral until an environmental change will probably be in linkage equilibrium with markers at useful distances in the population and therefore will often be missed in a genome scan comparison, even when they contribute substantially to adaptation. Mutations that existed before an environmental change, at lower frequencies, in mutation-selection balance may be detectable in some cases.

I discussed the dependence of power on the initial frequency of the marker alleles. This, of course, will be unknown in a natural population, but can be something to take into account in experimental set-ups. The main point is that any marker close to a selected site may not give a signal if the mutation originated near a common allele. Like with association studies, no signal does not mean no selection, and to discover selection any site needs to be covered by multiple markers. The choice for multi-allelic markers should be considered.

As a threshold, I used the 99-percentile point of the distribution under the null-hypothesis of no selection. This is quite arbitrary, and in real experiments multiple-testing issues may need to be considered.

#### 4.1. Applicability to experimental selection

The method of allele frequency comparisons seems suitable for experimental set-ups, provided some recommendations are kept in mind. In experiments lasting only a few dozen generations, most adaptation will be from the standing genetic variation. Mutations that are in complete linkage equilibrium in the population can only be detected when they are initially quite rare (*e.g.* five copies in a population of 250). Using a mixture of inbred lines as a base population, as done by Colson [6], presents therefore a much better situation, since favoured alleles will usually be in LD with their surrounding markers. Because selection is usually strong in experiments, the signal extends over longer distances, which makes sparse markers feasible. However, since small populations generate much background noise, only initially rare marker alleles can generate strong enough signals to be detected. This means that several markers need to cover each site or that multi-allelic markers must be used. Loci that experience only weak selection on the heterozygote have no chance of being detected by this method. The optimal set-up, if feasible, would be replicated selection lines, populations of at least 250, a starting population composed of a combination of inbred lines, and multi-allelic markers placed about every 3–5 cM.

#### 4.2. Applicability to natural populations

In natural populations, population sizes are probably larger, selection strength is probably weaker, and time since divergence is probably longer than in selection lines. Because of larger populations and longer periods, adaptation will not only use standing genetic variation, but also mutations that originate in one of the subpopulations after the start of divergence. These are the mutations that are most easily detected, since they will be in LD at the moment they arise. Still, the association must be with a rare marker allele, especially when many generations separate the sub-populations or when population sizes are small, so multiple markers covering each site are necessary. This is similar to association studies [15], as is probably the higher efficiency of multi-allelic markers (not tested in this study). Under the best conditions, in particular early after divergence, markers up to 1 cM had a detection probability more than 60% for new mutations under favourable conditions, and 40% when the mutation was present at low frequencies in the base population. Under these conditions, two to three markers per cM would have been enough to cover the whole genome with overlap. This would even allow detection of selection on rare, but

not unique alleles present in the base population. However, weaker selection, smaller populations, or longer time since divergence (which is very likely to be the case) will necessitate ever-denser spacing. In humans, the anticipated 100K SNP-chip should be able to pick up most of the stronger selection events separating populations from different continents.

The problem of detecting selection on standing variation is shared with other methods [11], and needs attention.

## ACKNOWLEDGEMENTS

This work was funded by the Netherlands Organisation for Scientific Research (NWO - ALW, project number: 813.04.001).

## REFERENCES

- [1] Abdallah J.M., Goffinet B., Cierco-Ayrolles C., Perez-Enciso M., Linkage disequilibrium fine mapping of quantitative trait loci: A simulation study, *Genet. Sel. Evol.* 35 (2003) 513–532.
- [2] Akey J.M., Eberle M.A., Rieder M.J., Carlson C.S., Shriver M.D., Nickerson D.A., Kruglyak L., Population history and natural selection shape patterns of genetic variation in 132 genes, *PLoS Biol.* 2 (2004) 1591–1599.
- [3] Akey J.M., Zhang G., Zhang K., Jin L., Shriver M.D., Interrogating a high-density SNP map for signatures of natural selection, *Genome Res.* 12 (2002) 1805–1814.
- [4] Beaumont M.A., Balding D.J., Identifying adaptive genetic divergence among populations from genome scans, *Mol. Ecol.* 13 (2004) 969–980.
- [5] Cavalli-Sforza L.L., Population structure and human evolution, *Proc. Royal Soc. London Series B.* 164 (1966) 362–379.
- [6] Colson I., *Drosophila simulans*' response to laboratory selection for tolerance to a toxic food source used by its sister species *D-sechellia*, *Evol. Ecol.* 18 (2004) 15–28.
- [7] Crow J.F., Kimura M., *An Introduction to Population Genetics Theory*, Harper & Row, 1970.
- [8] Falconer D.S., Mackay T.F.C., *Introduction to Quantitative Genetics*, 4th edn., Pearson Education, Harlow, 1996.
- [9] Fay J.C., Wu C.I., Hitchhiking under positive Darwinian selection, *Genetics* 155 (2000) 1405–1413.
- [10] Harr B., Kauer, M., Schlotterer C., Hitchhiking mapping: A population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*, *Proc. Natl. Acad. Sci. USA* 99 (2002) 12949–12954.
- [11] Hermisson J., Pennings P.S., Soft sweeps – molecular population genetics of adaptation from standing genetic variation, *Genetics* 169 (2005) 2335–2352.

- [12] Hughes A.L., Packer B., Welch R., Bergen A.W., Chanock S.J., Yeager M., Effects of natural selection on interpopulation divergence at polymorphic sites in human protein-coding loci, *Genetics* 170 (2005) 1181–1187.
- [13] Kayser M., Brauer S., Stoneking M., A genome scan to detect candidate regions influenced by local natural selection in human populations, *Mol. Biol. Evol.* 20 (2003) 893–900.
- [14] Schlotterer C., Hitchhiking mapping – functional genomics from the population genetics perspective, *Trends Genet.* 19 (2003) 32–38.
- [15] Sham P.C., Zhao J.H., Curtis D., The effect of marker characteristics on the power to detect linkage disequilibrium due to single or multiple ancestral mutations, *Ann. Hum. Genet.* 64 (2000) 161–169.
- [16] Smith J.M., Haigh J., Hitch-hiking effect of a favorable gene, *Genet. Res.* 23 (1974) 23–35.
- [17] Stuber C.W., Moll R.H., Goodman M.M., Schaffer H.E., Weir B.S., Allozyme frequency changes associated with selection for increased grain-yield in Maize (*Zea mays* L.), *Genetics* 95 (1980) 225–236.
- [18] Tajima F., Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism, *Genetics* 123 (1989) 585–595.
- [19] Vitalis R., Dawson K., Boursot P., Interpretation of variation across marker loci as evidence of selection, *Genetics* 158 (2001) 1811–1823.
- [20] Wright S., The genetical structure of populations, *Ann. Eugen.* 15 (1951) 323–354.