

ORIGINAL ARTICLE

Large-scale SNP genotyping in crosses between outbred lines: how useful is it?

MC Ledur^{1,2}, N Navarro¹ and M Pérez-Enciso^{1,3}

¹Department Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, Spain; ²Embrapa Suínos e Aves, Concórdia, Santa Catarina, Brazil and ³Institut Català de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Although genome-wide association (GWA) studies are not worth the effort in crosses between inbred lines, many crosses are actually made up of divergent yet outbred populations. Despite its relevance, however, this experimental setting has not been studied at a time when SNP microarrays are available in many species. To assess whether GWA can be useful in this setting, we performed combined coalescence—gene dropping simulations. We studied the influence of marker density, QTL effect and QTL allele frequency on power, false discovery rate (FDR) and accuracy. Our results suggest that GWA in outbred F_2 crosses is useful, especially in large populations. Under these circumstances, accuracy increased and FDR decreased as compared with classical linkage analysis. However, current SNP densities (in the order of 30–60 K

SNPs/genome or equivalent to 10–20 SNPs per cM) may not be much better than linkage analysis and higher SNP densities may be required. SNP ascertainment had an important effect; the best option was to select SNPs as uniformly as possible without setting any restriction on allele frequency. Using only SNPs with fixed alternative alleles in each breed controlled false positive rate but was not useful to detect variability within lines. Finally, the most significant SNP was not necessarily the closest to the causal SNP, although the closest SNPs were usually above the significance threshold; thus, it is prudent to follow-up significant signals located in regions of interest even if they do not correspond to absolute maxima. *Heredity* (2010) **105**, 173–182; doi:10.1038/hdy.2009.149; published online 21 October 2009

Keywords: association mapping; microarray; outbred cross; QTL; SNP

Introduction

Two opposite experimental designs have been extensively studied for complex trait loci (QTL) mapping: either crosses between inbred populations or outbred populations. In the former design, classical linkage QTL analysis has been the method of choice, whereas association mapping with massive genotyping of thousands of SNPs is now being extensively used in the latter design. Nevertheless, these two extreme cases cover only a fraction of experiments. In many species, primarily animals and outcrossing plants, inbred lines are not available although extreme divergent breeds do exist. Thus, many QTL experiments have been performed crossing divergent lines (Abasht *et al.*, 2006; Rothschild *et al.*, 2007). These lines or breeds can be quite dissimilar phenotypically, but still retain considerable amounts of within breed genetic variability. A typical case occurs in animal breeding, where highly successful programmes are carried out *within* breeds. In dairy cattle, heritability of milk production is around 25% even after many years of intense selection applying artificial insemination across countries worldwide. Similarly, the

heritabilities for body weight in broilers are of the same magnitude or even higher than those estimated in early generations of selection programs. In maize, >100 generations of selection for high oil and protein percentage has not exhausted genetic variability, and response to selection continues to be achieved (Moose *et al.*, 2004). These examples show no evidence of loss in variability within experimental or commercial lines.

Extreme breeds are particularly relevant in plant and animal domestic species. They have been primarily the result of intentional breeding by man (for example, dog breeds have been selected for many different objectives including defence, company, hunting, and so on) and also of geographic isolation. The pig, for instance, was domesticated across wide different geographic areas from Europe to Asia starting from the local wild boar populations; the Asian and European wild boar lineages diverged since at least 600 KYA (Larson *et al.*, 2005). Thus, extreme breeds are important resources to investigate the genetic control underlying phenotypic variation (Georges, 2007).

Other examples are studies on genetics of adaptation. With the increasing availability of molecular markers in nonmodel species, the QTL paradigm is also colonizing this area of ecological and evolutionary genetics (Phillips, 2005). Besides species where long-range pedigrees have been recorded (Slate *et al.*, 2002), these studies are also using crosses between natural populations under differential selection regimes (Colosimo *et al.*, 2004; Steiner *et al.*, 2007), which may approximate the

Correspondence: Professor M Pérez-Enciso, Food and Animal Science, Universitat Autònoma Barcelona, Facultat Veterinària, Bellaterra 08193, Spain.

E-mail: miguel.perez@uab.cat

Received 16 January 2009; revised 23 September 2009; accepted 28 September 2009; published online 21 October 2009

'ancestor-descendant' pair. These and other empirical studies show that adaptation can arise from standing genetic variation (Barrett and Schluter, 2008) and that different loci might evolve to the same phenotype among separate populations (Arendt and Reznick, 2008). Therefore, loci responsible for adaptation will be often found segregating both between and within divergent populations.

The difference between crosses of inbred versus outbred lines is subtle yet important. QTL analyses of outbred crosses have traditionally ignored the variability within lines, despite the fact that it can be important, as mentioned. Besides, segregating alleles within line or breed contribute to decrease power under a traditional analysis. Nevertheless, hundreds of QTL for several traits have been mapped using linkage analysis with F_2 resource populations, an evidence that genetic variance between breeds is also relevant (<http://www.animalgenome.org/QTLdb/>). A problem with linkage QTL analysis is that confidence intervals for QTL positioning are very large, in the order of several Mb, making it difficult to select candidate genes. It is a well-known fact that increasing the number of markers does not pay off above a relatively sparse density because of the few recombinants that have been generated in a F_2 cross (Darvasi, 1998). However, this is true for inbred lines but not necessarily so for outbred crosses. Although association can greatly increase accuracy in outbred populations (Risch and Merikangas, 1996), its usefulness in F_2 outbred crosses has not been thoroughly evaluated.

Thus, it is reasonable to ask how powerful—and useful—could be a large-scale association study using SNP microarrays in crosses between divergent outbred populations. Given that many of these crosses have been already generated, with dozens of traits measured in hundreds of individuals, we wish to primarily study whether microarray genotyping will be worth the effort; we are not much concerned here about the optimum design for fine mapping. There are several relevant questions that we wish to address in this work: What are the influences of marker density, QTL effect or QTL allele frequency? Is it better to increase the size of the experiment or the number of markers? How influential is SNP ascertainment bias? that is, the fact that SNPs are discovered in one population but applied to other population(s). We have performed a series of mixed coalescence—gene dropping simulations to address these questions.

Materials and methods

Simulations

Domestic populations are usually highly structured, for example, in different breeds that remain partially isolated. Unequal effective size is also frequent among breeds. For instance, a common practice in QTL animal experiments has been to cross a commercial widely distributed breed with a local breed or with the wild ancestor (Georges, 2007). To mimic this scenario, we simulated two outbred populations, P1 and P2, differing in effective population, $Ne_1 = 600$ and $Ne_2 = 200$, respectively. These founder populations were originated from a unique population of $Ne = 1000$, 2000 generations ago. Five chromosomes of 200 cM length each were simulated

assuming a recombination rate of 1 cM Mb^{-1} . A mutation rate per generation per base pair of 10^{-8} was assumed, as well as a migration rate per generation per individual of $m = 2.5 \times 10^{-4}$. A fixed number of SNPs (10 000) was assigned per chromosome in the F_0 , that is, a total of 50 000 SNPs were simulated. To simulate the whole process, we used a combination of backward (coalescence) and forward approaches (Figure 1). First, parental F_0 genomes were obtained from coalescence simulations using GENOME software (Liang *et al.*, 2007) following the demographic model described. Next, a forward simulation program was used to obtain the F_2 population, conditional on genotypes obtained through the coalescence.

Among other statistics, we investigated the effect of F_2 population size on QTL detection. Either 200 or 2000 F_2 individuals were generated. The small F_2 population consisted of 20 full-sib F_2 families, descendants of five P1 sires and five P2 dams. Each of the five F_1 families was made up of one male and four females. Each F_1 sire was then randomly mated to four F_1 dams producing 10 F_2 offspring per dam. The large F_2 population, consisting of 100 full-sib F_2 families, was generated similarly but mating 10 P1 sires and 10 P2 dams. Each of the 10 F_1 families was made up of 1 male and 10 females. Fifty computer replicates per case were simulated.

The continuous phenotypic trait was controlled by five additive loci, each accounting for 35, 18, 10, 5 and 2% of phenotypic variance, respectively. We considered two extreme linkage scenarios. In the first one, the QTL were located in the first four chromosomes, the fourth chromosome harboured the two QTLs of smallest effect, whereas no QTL was located on chromosome 5. In the second scenario, complete linkage, all QTL were randomly positioned within a chromosome and the rest of the chromosomes were devoid of causal mutations. In general, we focussed on the first scenario as it is in better agreement with the experimental QTL results. Causal SNPs were chosen randomly on each chromosome and removed from the dataset for the QTL discovery analysis. As causal SNPs differed in frequency (p) from replicate to replicate, the absolute effect (a) was adjusted so that QTL heritability (h^2_Q) was as desired: $a = \sqrt{h^2_Q \sigma_y^2 / 2p(1-p)}$, where σ_y^2 is the phenotypic variance. A random normal deviate was added to the genotypic effect to generate the phenotypic value.

A matter of utmost interest is the optimum number of SNPs genotyped, but also how these are chosen to minimize the undesirable effects of ascertainment bias (Nielsen *et al.*, 2004). Five SNP maps with different densities were generated in each replicate. Three maps had a fixed number of SNPs: 500 (0.5 K), 12 500 (12.5 K) and 50 000 (50 K). These SNPs were equally spaced on the genome, that is, every 2, 0.08 and 0.02 cM, respectively. Two additional SNP maps were chosen using information from the allelic frequencies in the parental populations. In MAFP, we retained only the SNPs with a minimum minor allele frequency of 0.20 in population 1. This resembles the usual strategy in building SNP chips, where low frequency markers are discarded and where only a subset of populations is used to uncover SNPs. The last map (FIXP) was made up from SNPs with alternative fixed alleles in each breed. Note that, in this case, the SNPs cannot be used to detect any within breed genetic variation and thus mimics the usual linkage QTL

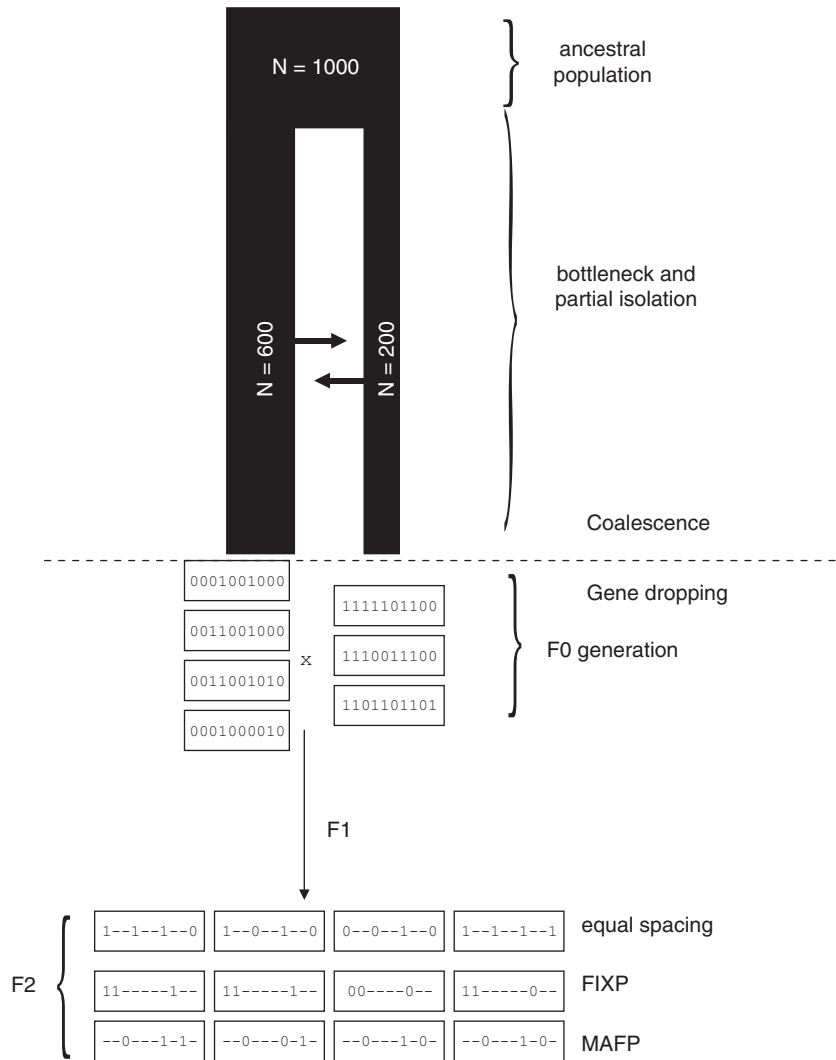


Figure 1 Scheme of combined coalescence and gene dropping simulation strategy, above and below the dashed line, respectively. The top part represents an ancestral population that splits up into two partially isolated subpopulations evolving separately for 2000 generations. In the bottom half of the figure, a sample of 10 SNP chromosomes (the boxes with 0's and 1's) from the F_0 is represented. At the bottom, several genotyping strategies in a sample of four F_2 individuals are shown. In equal spacing strategies, SNPs are selected solely based on position, here every three SNPs. In the fixed allele strategy (FIXP), only SNPs with alternative alleles fixed in each breed are genotyped (SNPs 1, 2 and 8 in the example). In MAFP strategy, SNPs segregating above a certain threshold in population 1 are genotyped, here SNPs 3, 7 and 9 are segregating in population 1. Note that strategies FIXP and MAFP result necessarily in disjoint subsets of SNPs, whereas some SNPs can be shared between equal spacing and either FIXP or MAFP strategies.

analyses, that is, fixed alternative QTL alleles in each breed are assumed. The specific number and density of SNPs in MAFP and FIXP maps varied from replicate to replicate, but consistently ranged from 7000 to 13000. They are thus comparable in terms of genotyping cost to the 12.5-K map.

Statistical analysis

Although numerous methods exist for QTL detection, the primary goal here was to assess the utility of large-scale genotyping platforms in intercross populations rather than comparing statistical methods. Thus, for simplicity and computational speed, we only considered single marker association analysis. The association analysis was carried out through least-squares fitting an additive QTL model with a custom made C and R

program. We tested the association of each single SNP with the phenotypic trait using an F-test. Genome-wide significance was obtained using 10000 permutations (Churchill and Doerge, 1994). The 5% upper quantiles of the distributions of the minimum P -values were analysed in four extreme combinations of sample sizes (200 and 2000 individuals) and sparse and dense SNP maps (0.5 and 50 K SNPs). As a result, we defined two genome-wide significance thresholds, $P = 3.1 \times 10^{-4}$ and 10^{-5} , based on values obtained for the sparse and dense maps, respectively.

The different scenarios were evaluated according to several criteria, primarily power, the proportion of false positives (that is false discovery rate, FDR) and accuracy in estimating the QTL position. Power was defined as the proportion of replicates where at least one SNP was significant within a maximum distance of 2.5 cM away

from nearest causal position. Note that we did not require the significant SNP to be the *most* significant one in that chromosome. FDR was the average number of replicates where the most significant SNP within a chromosome was at a distance >2.5 cM from the causal position, provided the SNP was significant. Mean accuracy was the average distance between the most significant SNP and the nearest causal position. In contrast to power, accuracy and FDR do depend on the magnitude and position of the most significant SNP.

Results

Genome scan profiles

An appraisal of how SNP selection strategy affects the P -value profile is shown in Figure 2, which portrays two of the replicates for strategies 12.5K, MAFP and FIXP when QTL are in distinct chromosomes. In replicate 1 (left column), the QTL showed extreme allelic frequencies in each parental population, that is, alternative alleles were close to fixation in each breed for all QTL, and thus intermediate frequencies were observed in the F_2 generation. For replicate 2, in contrast, the QTL had similar allelic frequencies in the parental populations, and the five QTL frequencies in the F_2 were 0.89, 0.96, 0.95, 0.91 and 0.01, respectively. Replicate 1 thus mimics the model assumed in most classical analyses, whereas cases similar to replicate 2 are more frequent in crosses between outbred lines.

The P -value profiles were relatively similar among SNP selection strategies although with some important differences (Figure 2). First, the FIXP profile (bottom row) is distinctly smoother than in the remaining strategies. This occurs because the FIXP is equivalent, as mentioned, to using only linkage information, that is, it uses only the meioses having occurred during the F_2 cross. The 12.5K or MAFP P -value profiles were roughly parallel to those of FIXP, but the variability was much larger. We observed this even within nearby markers and for very large effect QTL (chromosome 1). These fluctuations were even larger at extreme QTL allele frequencies (right column in Figure 2). A second, less apparent but highly consistent difference was that FIXP resulted in far fewer false positives than any other option. This can be observed from the profile in chromosome 5, where no QTL resides. When QTL are fixed for alternative alleles (replicate 1), the risk of false positives was very low in any SNP map. However, when the QTL is segregating within breeds, all P -values for chromosome 5 in FIXP were below the threshold (horizontal line), whereas significant P -values, that is, false positives, were frequent in the other maps. In fact, the average P -values in either MAFP or 12.5K maps were similar between chromosomes 4 (two small effect QTL) and 5 (no QTL).

As for complete linkage between QTL, we found overall less impact of QTL allele frequency on P -value profiles. Instead, profiles were more dependent on relative positions between QTL than on allele frequencies. As an example, Figure 3 shows how the most

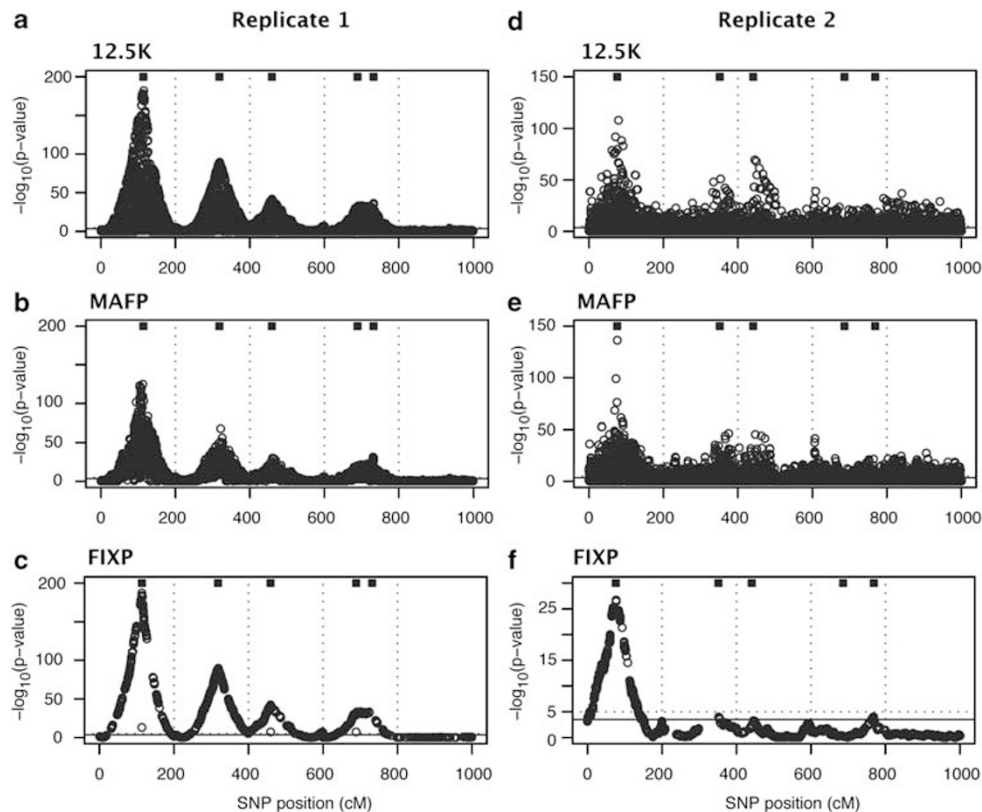


Figure 2 Plots of genome-wide P -value for 12.5K, MAFP and FIXP SNP densities from two replicates of $N=2000$, unlinked QTL. Dashed vertical lines separates chromosomes. Horizontal lines are the thresholds: $P = 3.1 \times 10^{-4}$ (solid line) and 10^{-5} (dashed line). Squares on the top of each figure are the QTL positions. Note that the y-axis scales may differ between plots.

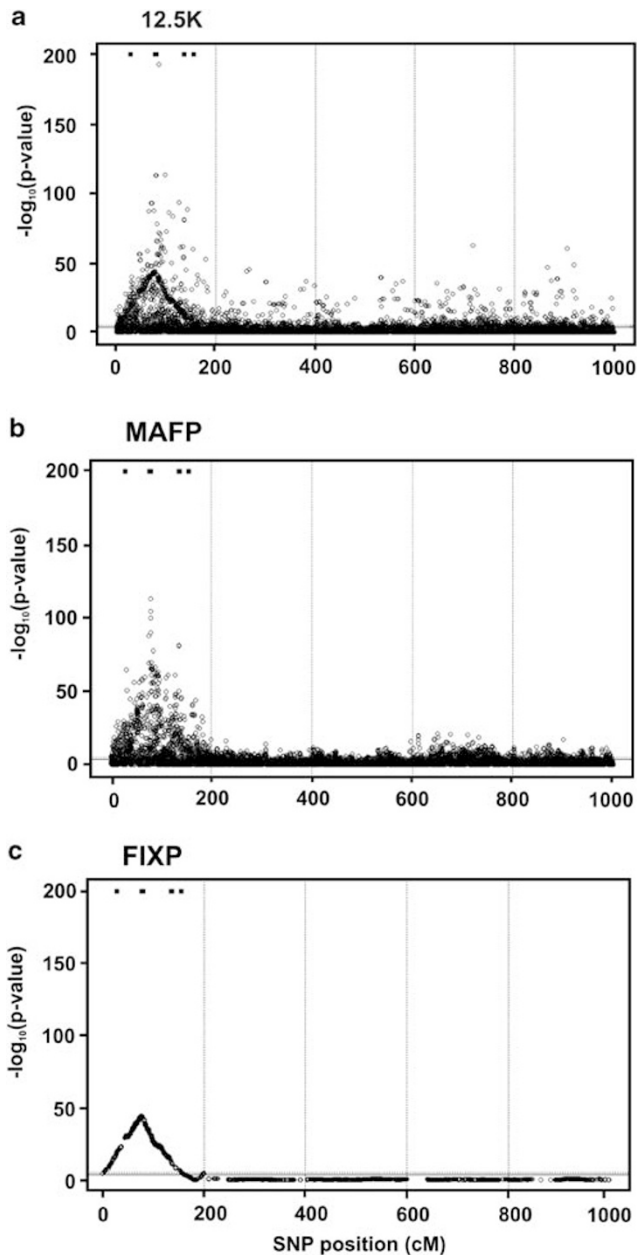


Figure 3 Plots of genome-wide P -value for 12.5K, MAFP and FIXP SNP densities from a replicate with $N=2000$, linked QTL. Dashed vertical lines separates chromosomes. Horizontal lines are the thresholds: $P=3.1 \times 10^{-4}$ (solid line) and 10^{-5} (dashed line). Squares on the top of each figure are the QTL positions.

significant P -value coincides with a cluster of two closely positioned loci. Less significant P -values are scattered in the region of additional loci for either MAFP or 12.5K strategies, whereas the FIXP strategy results in a single maximum.

Impact of QTL frequencies

For each QTL in a given chromosome, its absolute effect was scaled such that the variance explained was constant across replicates (see Materials and methods). Nevertheless, it is of interest to study whether the QTL allele frequency had still an influence *per se*. The distribution of

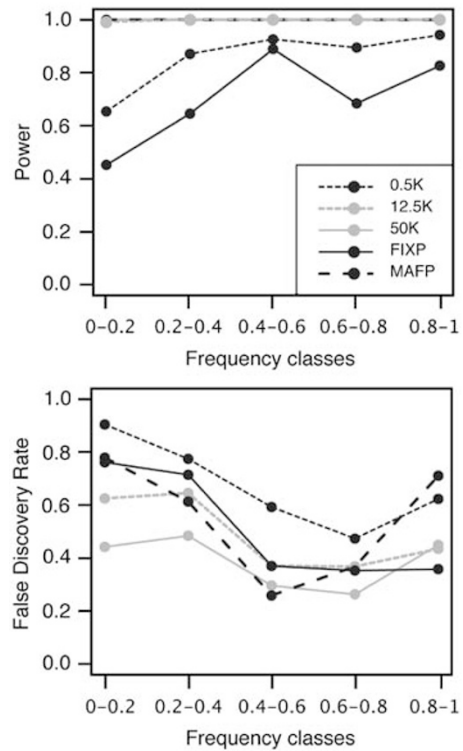


Figure 4 Conditional power and FDR as a function of differences between breeds in QTL allele frequency; $N=2000$, threshold $P=3.1 \times 10^{-4}$.

allelic frequencies in an inbred F_2 cross is narrowly distributed around 0.5, because any locus contributing to the genetic variance has fixed alternative alleles in each of the parental lines. Therefore, absolute QTL effects can be compared across loci. An important difference with outbred line crosses is that the allele frequency spectrum is much broader than in inbred crosses; therefore, the contribution of each QTL to total genetic variance depends both on its absolute effect and on its allelic frequency. To assess this effect, we plotted power and FDR against QTL frequency differences between lines (Δ_{f12}). Figure 4 shows the results when averaging over replicates and QTL for $N=2000$ and the unlinked QTL scenario.

Power increased with the FIXP strategy as Δ_{f12} increased, but the effect was undetectable for $\Delta_{f12} > 0.5$. The same trend was observed at a relatively sparse SNP coverage (0.5K), whereas MAFP and a denser SNP coverage guaranteed maximum power across all settings. Thus, at similar marker density, MAFP or 12.5K are better strategies than FIXP in terms of power. Recall though that we defined power as the percentage of any SNP being significant at a maximum distance of 2.5 cM without regard for the level of significance. As for FDR, it decreased as Δ_{f12} increased, although a minimum occurred at intermediate Δ_{f12} . A sparse coverage map (0.5K) was the worst strategy in terms of FDR, simply because the probability of not finding an SNP nearby the causal QTL increases when SNP density decreases. Overall, the minimum FDR was achieved with the largest SNP coverage (50K) but either MAFP or FIXP performed equally well in some instances. Therefore, when we condition on QTL heritability, the effect of QTL allele frequency on power and FDR depends on the SNP

ascertainment procedure and on SNP density. At similar SNP density, equal spacing (12.5 K) was the safest strategy compared with MAFFP or FIXP.

Power, FDR and accuracy

Table 1 presents the summary statistics for power and FDR across genotyping strategies, population size and QTL effect in the unlinked scenario. Results are presented for both significance thresholds, $P = 3.1 \times 10^{-4}$ and $P = 10^{-5}$. Sparse genotyping (0.5 K) and small population sizes were overall more sensitive to threshold choice than large populations and dense genotyping. Note that there was no advantage in increasing the threshold above a certain level. On the contrary, an increase in the significance threshold reduced power, whereas the rate of false positives was not simultaneously reduced. For instance, for QTL 3 ($h^2_Q = 10\%$), increasing the significance threshold from $P = 3.1 \times 10^{-4}$ to $P = 10^{-5}$ did not pay off. For sparse genotyping (0.5 K) and small F_2 ($N = 200$), power decreased from 0.40 to 0.18, whereas FDR decreased only marginally from 0.69 to 0.65. Similar results are found with other scenarios. Thus, in the following we will focus on the results with threshold $P = 3.1 \times 10^{-4}$.

Power should augment with increasing QTL effect size, SNP density and population size. However, these three parameters do not behave additively. Overall, it is better to increase the population size rather than SNP density. Logically, this holds provided phenotyping is not too expensive. The current expectation is, however, that genotyping costs will continue to decrease dramatically, whereas the phenotyping costs should increase. In addition, many experiments have already been devel-

oped and population size is already fixed. Therefore, it is more realistic to consider that SNP density or ascertainment procedure can be modified rather than population size. Increasing SNP density was important to reduce FDR, especially in large populations. In QTL 2 ($h^2 = 18\%$), FDR decreased from 0.77 to 0.50 ($N = 200$) and from 0.68 to 0.16 ($N = 2000$) when SNP density increased from 0.5 to 50 K, respectively. Although power increased with population size, note that the rate of false positives was very high for small effect QTL, even with $N = 2000$ and 50 K SNPs. Thus, even if a nearby SNP is significant, it will not necessarily be the most significant SNP when QTL effect decreases. Power close to one for all QTL effect sizes was achieved when $N = 2000$ except with FIXP or very sparse 0.5 maps. For FIXP and 0.5 K SNP densities, power increased with the increase of QTL effect, especially for loci with small-to-medium effect size. When the locus effect is large, power was very high in either $N = 200$ or $N = 2000$ F_2 populations.

A different matter is deciding the most reasonable strategy for choosing SNPs, that is, what is best among FIXP, MAFFP or 12.5 K maps, which all contain approximately the same number of SNPs. Table 1 shows that MAFFP outperformed FIXP in terms of power but FIXP resulted in lower FDR than MAFFP. Interestingly, allowing for a uniform SNP coverage without regard for allele frequency (12.5 K) was usually the best strategy, both in terms of power and of FDR. This agrees well with the data in Figure 2, where a spacing between SNPs as uniform as possible was the most robust strategy.

Accuracy is probably the single most relevant property for fine mapping studies. Average accuracies and their s.d. across genotyping strategies, population size and QTL effect in the unlinked scenario are shown in Table 2.

Table 1 Conditional power and false discovery rate (FDR) of single marker association analysis over 50 replicates when the three largest effect loci are located in separate chromosomes (unlinked scenario)

Parameter	Individuals	QTL (h^2_Q) ^a	SNP density—threshold $P = 3.1 \times 10^{-4}$					SNP density—threshold $P = 10^{-5}$				
			0.5 K	12.5 K	50 K	FIXP	MAFFP	0.5 K	12.5 K	50 K	FIXP	MAFFP
Power ^b	200 F_2	1 (0.35)	0.78	0.98	1.00	0.74	0.96	0.70	0.96	1.00	0.64	0.90
		2 (0.18)	0.50	0.98	1.00	0.52	0.92	0.40	0.88	0.96	0.40	0.76
		3 (0.10)	0.40	0.84	0.90	0.38	0.80	0.18	0.66	0.76	0.20	0.58
		4 (0.05)	0.24	0.56	0.68	0.22	0.46	0.10	0.30	0.40	0.08	0.30
		5 (0.02)	0.10	0.46	0.64	0.14	0.44	0.02	0.20	0.30	0.02	0.20
	2000 F_2	1 (0.35)	0.88	1.00	1.00	0.70	1.00	0.86	1.00	1.00	0.58	1.00
		2 (0.18)	0.92	1.00	1.00	0.84	0.98	0.84	1.00	1.00	0.78	0.98
		3 (0.10)	0.78	1.00	1.00	0.66	1.00	0.70	1.00	1.00	0.66	0.98
		4 (0.05)	0.78	1.00	1.00	0.54	1.00	0.68	1.00	1.00	0.46	0.96
		5 (0.02)	0.68	1.00	1.00	0.48	1.00	0.60	0.96	1.00	0.40	0.96
200 F_2	1 (0.35)	0.66	0.38	0.40	0.30	0.61	0.64	0.38	0.40	0.29	0.61	
	2 (0.18)	0.77	0.62	0.50	0.66	0.66	0.70	0.62	0.50	0.58	0.66	
	3 (0.10)	0.69	0.67	0.72	0.59	0.63	0.65	0.64	0.70	0.50	0.62	
	4 (0.05)	0.89	0.93	0.93	0.71	0.91	0.84	0.92	0.92	0.75	0.91	
	5 (0.02)	0.97	0.98	0.98	1.00	0.96	0.95	0.97	0.97	1.00	0.97	
2000 F_2	1 (0.35)	0.70	0.24	0.08	0.49	0.46	0.70	0.24	0.08	0.44	0.46	
	2 (0.18)	0.68	0.36	0.16	0.35	0.62	0.68	0.36	0.16	0.35	0.62	
	3 (0.10)	0.68	0.38	0.32	0.41	0.58	0.67	0.38	0.32	0.39	0.58	
	4 (0.05)	0.74	0.66	0.58	0.65	0.66	0.74	0.66	0.58	0.64	0.66	
	5 (0.02)	0.92	1.00	0.96	0.86	0.94	0.92	1.00	0.96	0.85	0.94	

^a h^2_Q , QTL heritability.

^bPower, proportion of replicates where at least one SNP was significant at a maximum distance of 2.5 cM away from the causal position.

^cFDR, average number of replicates where the most significant SNP within a chromosome was at a distance >2.5 cM from the causal position, provided the SNP significance was above the threshold.

Table 2 Mean accuracy in cM and its standard deviation in parenthesis, over 50 replicates when the three largest effect loci are located in separate chromosomes (unlinked scenario)

Individuals	QTL (h_{Q}^2) ^a	SNP density—threshold $P = 3.1 \times 10^{-4}$				
		0.5 K	12.5 K	50 K	FIXP	MAFP
200 F ₂	1 (0.35)	16.3 (37.2)	11.7 (36.1)	5.1 (8.3)	8.0 (29.8)	24.7 (51.5)
	2 (0.18)	28.6 (46.4)	10.2 (13.6)	10.9 (26.2)	13.7 (36.4)	19.4 (41.5)
	3 (0.10)	36.5 (55.4)	22.8 (41.2)	20.0 (31.5)	30.3 (60.8)	23.1 (40.6)
	4 (0.05)	50.2 (52.2)	58.4 (52.5)	56.2 (50.6)	36.7 (54.2)	49.8 (51.4)
	5 (0.02)	63.2 (48.5)	66.8 (57.2)	62.9 (52.8)	72.5 (55.1)	66.3 (50.6)
2000 F ₂	1 (0.35)	17.6 (33.9)	3.0 (6.7)	1.1 (2.8)	5.0 (7.2)	17.0 (43.3)
	2 (0.18)	20.9 (43.0)	3.9 (7.0)	1.4 (2.0)	7.3 (29.2)	18.8 (39.8)
	3 (0.10)	26.8 (45.4)	9.2 (27.9)	8.6 (29.7)	13.2 (39.6)	11.8 (28.3)
	4 (0.05)	44.9 (48.4)	39.1 (53.9)	33.7 (52.1)	33.3 (47.0)	23.6 (38.1)
	5 (0.02)	66.6 (52.4)	59.4 (47.8)	61.1 (49.5)	52.1 (53.4)	56.6 (46.1)

^a h_{Q}^2 , QTL heritability.**Table 3** Conditional power, false discovery rate (FDR) and mean accuracy in cM over 50 replicates, with $N = 2000$ F₂, when all QTL are located in a single chromosome

Parameter	QTL (h_{Q}^2) ^a	SNP density—threshold $P = 3.1 \times 10^{-4}$				
		0.5 K	12.5 K	50 K	FIXP	MAFP
Power	1 (0.35)	0.90	1.00	1.00	0.76	1.00
	2 (0.18)	0.92	1.00	1.00	0.72	1.00
	3 (0.10)	0.90	1.00	1.00	0.68	1.00
	4 (0.05)	0.92	1.00	1.00	0.72	1.00
	5 (0.02)	0.88	1.00	1.00	0.66	1.00
FDR	All	0.64	0.34	0.22	0.31	0.32
Accuracy	All	8.53	1.75	1.48	3.69	2.43

^a h_{Q}^2 , QTL heritability.

As expected, accuracy increased with QTL effect and population sizes. In terms of accuracy, FIXP was a better strategy than MAFP for high-to-moderate QTL effect sizes. Interestingly, evenly spaced markers (12.5 K) was again the best compromise across all settings studied. In agreement with the classical results that has shown the limits of linkage analysis (Darvasi, 1998), no much improvement in accuracy was achieved by increasing the sample size in MAFP. However, accuracy improved by increasing the number of genotyped markers in the experimental design studied here. Importantly, a relevant observation was that the variance of accuracy decreased dramatically with a high marker density. For instance, for $h_{Q}^2 = 0.18$ and $N = 2000$, the s.d. of accuracy decreased from 43 cM (0.5 K) to 2 cM (50 K). This is an important consideration, because the variation in accuracy was very large in general. Finally, the two last QTL in chromosome 4 merit a especial attention because they were positioned in the same chromosome. In each chromosome, we retained only the most highly significant P -value and thus increasing accuracy in one QTL is opposed to accuracy for the additional QTL, unless they were proximal, which is not too likely, the approximate probability being $5/200 = 2.5\%$.

The above-mentioned results are, overall, also valid when a single chromosome harbours many QTL (complete linkage scenario, Table 3). As for FDR and accuracy, note that a single value is reported because we

considered only a single maximum per chromosome. FDR was somewhat reduced with respect to the unlinked scenario, probably because the average P -value remains significant for longer stretches on the chromosome than when QTL are unlinked. As for accuracy, it was also improved but recall once more that we refer only to one locus. As there are more causal loci in the chromosome, the chances of having a causal locus nearby are increased with respect to the unlinked scenario.

Distribution of P -values

The usual approach to infer the QTL position is to choose the SNP with the most significant P -value. Nevertheless, many P -values are usually above the significance cut-off in any large-scale association study. The behaviour of these extreme P -values is also of interest, especially comparing different SNP—selection strategies. To investigate this further, we ascertained the sets of contiguous SNPs (chromosome segments) where all markers were significant. Within each segment, we retained the most significant P -value. We did that separately for each chromosome to distinguish between QTL magnitudes. Figure 5 plots the densities of these extreme $-\log_{10}(P\text{-values})$ in the unlinked scenario. Some interesting results appear. For instance, under the null hypothesis, that is, in chromosome 5 that harbours no QTL, there was no chromosome segment above the significance threshold for the FIXP option. In contrast, other SNP choices resulted in P -value distributions, with a considerable mass above the cut-off ($P = 3.1 \times 10^{-4}$). In other words, a fraction of the chromosome harbours significant markers, that is, false positives. The distribution of P -values is dramatically different for FIXP between chromosome 4, which contains two small QTL, and chromosome 5, without any QTL. In contrast, those of MAFP or 12.5 K were indistinguishable between chromosomes 4 and 5. Finally, note that the mass of the distribution is shifted towards larger $-\log_{10}(P\text{-values})$ for chromosome 1, simply a result of a much larger QTL effect. Again, the distribution in FIXP was very different from the rest of the SNP maps. FIXP P -value profiles were shifted towards the right because FIXP collects only the recombinants that have appeared in the F₂ pedigree and thus the strong disequilibrium causes more regions harbouring strongly significant P -values.

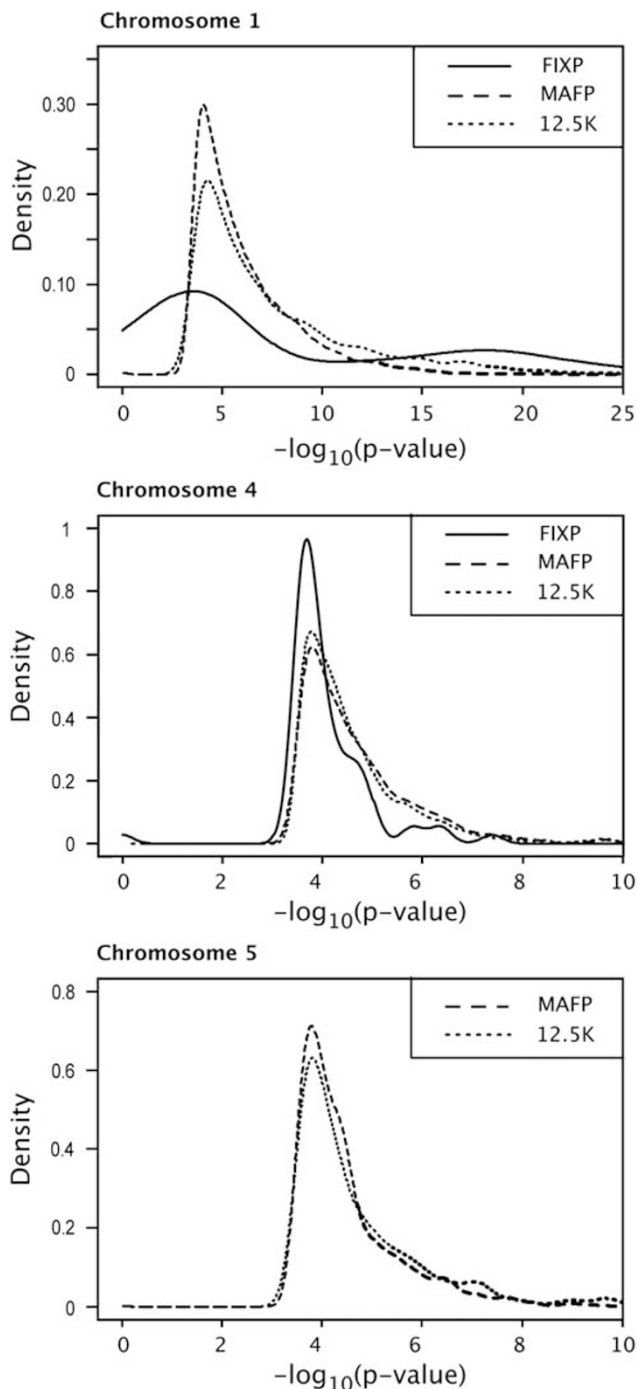


Figure 5 Smoothed distributions of maximum segment $-\log_{10}$ (P -values) for chromosomes 1 (largest QTL effects), 4 (two smallest effect QTL) and 5 (no QTL), for FIXP, MAFP and 12.5K SNP densities; $N = 200$.

Discussion

Crosses with outbred lines are complex

Currently, genome-wide association (GWA) studies have emerged as the method of choice for fine mapping complex trait genes. This is because microarray have made large-scale genotyping affordable and because of the advantages of association versus linkage in terms of

accuracy for QTL positioning. Certainly, this approach is not worth the effort in crosses between inbred lines. However, many crosses in domestic species are actually made up of divergent, yet outbred, populations. Despite its relevance, however, this experimental setting has been neglected at a time that SNP microarrays are becoming commercially available in several species. Moreover, next-generation sequencing technologies have the promise to bring such genomic resources also in ecological and evolutionary model organisms (Hudson, 2008).

Broadly, two simulation strategies are available in Genetics, backward (coalescence) and forward methods. The coalescence traces back the ancestors of a given sample until the most recent common ancestor is found. This method is very efficient because only the sequences that contributed to the current sample are simulated. In contrast, the forward strategy simulates the entire population from the past to present and is much slower computationally. Nevertheless, forward strategies can accommodate any structure or selection process and are becoming fashionable again due to new algorithms and better computer performance (Carvajal-Rodriguez, 2008). Here, we used a combination of both methods to generate a realistic nucleotide polymorphism pattern. We also compared different SNP selection strategies that use the two existing levels of disequilibrium in this experimental design. The F_2 disequilibrium is captured by the FIXP map, whereas the within breed disequilibrium is used primarily by MAFP.

As we have shown in this work, the presence of disequilibrium at the between and within breed levels can be properly used to reduce FDR and improve location accuracy as compared with classical linkage, provided a sufficiently dense genotyping is performed. In contrast to inbred crosses, a much broader allele frequency spectrum can be observed in crosses between outbred lines. This difference is also relevant because there exists an effect of the allele frequency difference between breeds on power and FDR, even for QTL explaining the same amount of variance (Figure 4). Allele frequency affected power only at sparse genotyping (Figure 4, top). FDR, however, was still sensitive to QTL allele differences at 50 K SNP density, increasing at the extremes of the distribution. FDR was especially high in option MAFP, which mimics a common criterion to choose SNPs for microarray platforms, that is, a cut-off on allele frequency.

The 12.5-K map, that is, one SNP every 0.08 cM, is roughly equivalent to the density of chips commercially available in livestock or other species, except human or mouse where much larger genotyping panels exist. Unless much denser genotyping are carried out, our results indicate that this density may not be enough to gain all advantages from association studies and is not necessarily a much better option than usual linkage analysis (FIXP map), especially considering the presence of ascertainment bias. Besides, considering that most available F_2 resource populations are made up of about 400 to 1000 individuals, it is likely that higher SNP densities are needed.

SNP ascertainment bias

At equal SNP density, how does SNP choice affect the results? or equivalently, what is the best strategy to select

genotyped SNPs? SNP ascertainment is a matter of critical importance in association studies and has received considerable attention (Nielsen *et al.*, 2004; Clark *et al.*, 2005). In most commercially available microarrays, SNPs have been selected according to informativity across one or more breeds, that is, there is a strong bias towards intermediate frequency markers, biasing the allele frequency spectrum. In addition, microarrays are used in populations that have not been used in the discovery process. We mimicked the consequences of this strategy in the MAFF option, whereby only SNPs with MAF above 0.20 in population 1 were selected. In the simplest strategy (12.5K), SNPs were chosen simply according to their position, without regard to frequency. SNP ascertainment bias is maximum in FIXP, whereby only the markers with most divergent allele frequency are selected. This strategy in turn is equivalent to classical linkage analyses because only breed origin can be traced with this kind of markers.

A comparison between these extreme SNP choice strategies is thus illuminating. MAFF improved conditional power over FIXP, especially for large populations (Table 1). However, these results are somewhat misleading because accuracy, that is, the distance from the most significant *P*-value to actual QTL position, was worse with MAFF and had larger variances (Table 2), particularly for large and intermediate effect QTL. For instance, mean accuracies for the largest effect QTL were 5 cM versus 17 in FIXP and MAFF maps for $N=2000$, respectively. Importantly, the best option was random SNP choice (12.5K), probably because it averages over all possible QTL allele and SNP frequencies. In terms of FDR, FIXP and MAFF were comparable, although FIXP performed slightly better. Again, uniformly spaced SNP was the best option. For large N and QTL effect, FDR was halved: 0.24 versus 0.49 with 12.5 K and FIXP, respectively. These results evidence a large impact of the SNP ascertainment process, and suggest that SNPs should be chosen to be uniformly distributed along the genome but without setting any restriction on allele frequency. Note that, although several methods have been developed for correcting ascertainment bias (Nielsen *et al.*, 2004), these help to alleviate bias in population parameter estimates but are of little help for association studies, where the aim is to compute a correlation between a phenotype and a genotype.

The optimum choice

The optimum allocation of experimental resources is a difficult topic in an association study and we do not intend to provide a general answer here, but rather some general guidelines. Although linkage analysis does not benefit from an increase in marker density above a modest level, this is not the case of association—linkage disequilibrium only—approaches. Before association studies were so widespread, it was already shown that SNP coverage in the human genome should be much higher than anticipated (Kruglyak, 1999). More recent work within the Wellcome Trust Case Control Consortium initiative (WTCCC, 2007) has also underlined the importance of genotyping a very large population, in the order of thousands at least. All this is a consequence of the highly stochastic nature of disequilibrium, a phenomenon that has been known for quite

some time (Hein *et al.*, 2005), but whose practical implications we are now encountering.

The experimental design studied here is somewhat intermediate between inbred crosses and within outbred populations, and one would expect that a high SNP density may not be so important. In fact, the experimental design studied is a continuum between both classical extremes. It is a matter of concern, therefore, how general our conclusions could be. In the specific population settings analysed here, the number of fixed SNPs was comparable to the number of SNPs segregating at intermediate frequencies in one of the parental populations. This number was $\sim 12\,000$ SNPs in our simulation scenario. Therefore, it is reasonable to believe that uniform spacing will be a better strategy that either MAFF or FIXP. The main difference across experiments will lie in the relative amounts of between and within breed linkage disequilibrium, or equivalently, in how divergent and inbred are the parental lines. This parameter in turn will govern how many fixed and segregating SNPs are per cM. If within breed disequilibrium is very high, most SNPs will be fixed between breeds, and a small number of SNPs will suffice. If, on the contrary, there is little divergence between breeds, a much larger number of SNPs should be genotyped to maximize accuracy.

Increasing population size is, in general, the best option to improve power. However, increasing sample size does not reduce FDR nor increases accuracy when SNP density is low, 0.5K or equivalent to a marker spacing of 2 cM. Increasing the number of SNPs had a positive effect on all parameters studied. Therefore, increasing both SNP density and population size is important, the exact balance depending on the parameter of interest and on the QTL effect size. In principle, accuracy can be greatly improved for large and medium effect QTL through association analysis when the sample size is large ($N=2000$) and genotyping is very dense, say the 50-K map, that is, one SNP every 0.02 cM. However, if one has to choose between small N with large SNP number (50K) and large N and lower SNP number (12.5K) the latter option is better both in terms of accuracy and of FDR. The user should not forget that high-density genotyping is no substitute for a large experimental size. In any case, uniform spacing was the optimum strategy for SNP choice. In this study, we simulated an extreme scenario with a high heritability trait (0.7) controlled by a small number of loci (5) with additive effects only. Although such large heritabilities are not unusual in crosses between highly divergent breeds, our genetic model is rather simplistic and should be taken as a starting point to compare with more complex situations, such as traits of low-to-medium heritabilities, controlled by larger number of loci, including dominance, epistasis and sex by QTL interaction.

The apparent discrepancy between high conditional power for MAFF as opposed to relatively lower accuracy compared with 12.5K or FIXP (Tables 1 and 2) is explained by the fact that conditional power was defined as *any* significant SNP within the window being higher than the threshold, whereas accuracy was based on the most significant *P*-value, which was not necessarily the closest to the QTL position. An interesting follow-up is that regions containing significantly associated SNP should not be discarded as potential candidate regions even if the minimum *P*-value is located in another

genome region. The FDR will always be high for loci with very small effect size, unless a very large sample size could be provided, a well-known limit from this kind of studies. This limit is nevertheless extremely relevant in many quantitative trait applications, where a large number of loci with small effect sizes could be the rule rather than the exception as, for instance, recent large-scale GWA studies on human height have shown (Visscher, 2008). Theoretical models based on Fisher's geometric model predicts such distribution of QTL effect sizes when adaptation involves new mutations and stable optimum (Orr, 2005). Nevertheless, this does not necessarily apply when adaptation involves standing genetic variation. Finally, a more sophisticated statistical framework incorporating uncertainties such as Bayesian or resampling approaches, as well as model averaging, could be used to improve on current QTL mapping strategies.

Conclusions

Large-scale genotyping in F_2 crosses are useful for outbred populations, especially for large N and SNP number. Under these circumstances, accuracy is increased and the rate of false positives decreased compared with classical linkage analysis. But importantly, current SNP densities in the order of 30–60 K SNPs (~ 10 –20 SNPs per cM) may not be much better than linkage analysis. In addition, the rate of false positives can still be high. Therefore, expectations about power increase may not be fulfilled. Some simple recommendations would be (1) not to increase significance threshold beyond sensible levels, for example, 5% genome-wide significance; being more strict decreases power whilst not decreasing FDR; (2) be aware that FIXP results in a smaller number of false positives but that, in turn, is not useful to detect variability within lines; (3) the optimum strategy is to select SNPs based on uniform position distribution rather than on any frequency selection criteria; and (4) it will be prudent to follow-up significant signals located in regions of interest even if they do not correspond to absolute maxima.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

MCL was supported by a grant from BID/Agrofuturo (Brazil); NN by a Beatriu de Pinós post doc grant from the Generalitat de Catalunya (Spain). This work was funded by MEC grant AGL2007-65563-C02-01/GAN (Spain) to MPE.

References

- Abasht B, Dekkers JCM, Lamont SJ (2006). Review of quantitative trait loci identified in the chicken. *Poult Sci* **85**: 2079–2096.
- Arendt J, Reznick D (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol* **23**: 26–32.
- Barrett RDH, Schluter D (2008). Adaptation from standing genetic variation. *Trends Ecol Evol* **23**: 38–44.
- Carvajal-Rodriguez A (2008). Simulation of genomes: a review. *Curr Genomics* **9**: 155–159.
- Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005). Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res* **15**: 1496–1502.
- Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D *et al.* (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol* **2**: e109.
- Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.
- Darvasi A (1998). Experimental strategies for the genetic dissection of complex traits in animal models. *Nat Genet* **18**: 19–24.
- Georges M (2007). Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Ann Rev Genomics Hum Genet* **8**: 131–162.
- Hein J, Schierup MK, Wiuf C (2005). *Gene Genealogies, Variation and Evolution*. Oxford University Press: Oxford
- Hudson ME (2008). Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Res* **8**: 3–17.
- Kruglyak L (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**: 139–144.
- Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J *et al.* (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**: 1618–1621.
- Liang L, Zollner S, Abecasis GR (2007). GENOME: a rapid coalescent-based whole genome simulator. *Bioinformatics* **23**: 1565–1567.
- Moose SP, Dudley JW, Rocheford TR (2004). Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends Plant Sci* **9**: 358–364.
- Nielsen R, Hubisz MJ, Clark AG (2004). Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics* **168**: 2373–2382.
- Orr HA (2005). The genetic theory of adaptation: a brief history. *Nat Rev Genet* **6**: 119–127.
- Phillips PC (2005). Testing hypotheses regarding the genetics of adaptation. *Genetica* **123**: 15–24.
- Risch N, Merikangas K (1996). The future of genetic studies of complex human diseases. *Science* **273**: 1516–1517.
- Rothschild MF, Hu ZL, Jiang Z (2007). Advances in QTL mapping in pigs. *Int J Biol Sci* **3**: 192–197.
- Slate J, Visscher PM, MacGregor S, Stevens D, Tate ML, Pemberton JM (2002). A genome scan for quantitative trait loci in a wild population of red deer (*Cervus elaphus*). *Genetics* **162**: 1863–1873.
- Steiner CC, Weber JN, Hoekstra HE (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol* **5**: e219.
- Visscher PM (2008). Sizing up human height variation. *Nat Genet* **40**: 489–490.
- WTCCC (2007). Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* **447**: 661–678.