Genome-Wide SNP Association Study Identifies Regions Of Interest Associated With Osteochondrosis In French Trotters

S. Teyssèdre,* M.C. Dupuis,† J.M. Elsen *, G. Guérin ‡, L. Schibler ‡, J.M. Denoix† and A. Ricard ‡

Introduction

Osteochondrosis (OC), which is a disorder of cartilage in growth affecting foals, can be a factor of low performance in racing horses. There is a variety of lesions and localizations of OC, but they are mainly located in fetlock, hock and stifle. The heritability of osteochondrosis was estimated around 0.22 in French trotters. Two previous studies have been performed to detect quantitative trait loci (QTLs) for hock and fetlock OC. Collectively, these studies identified 23 large QTL regions using about 200 microsatellites (Dierks et al., 2007; Wittwer et al., 2007). Here we report the preliminary results of a genome-wide association study taking advantage of the equine 54k SNP chip on a large sample of French Trotters. This work represents a first step toward the fine mapping of candidate regions and the later identification of genes responsible for equine OC.

Material and methods

Data. The data are from the French ANR funded GENEQUIN program that aims to obtain a sire sib case-control design. The study relies on a population of 623 French trotters. DNA was obtained from 98 sires with 458 offspring (mean 4.7 ± 3.6) and 67 offspring without genotyped sires. About 82% of horses were less than 4 years old at inspection time (mean 2.8 ± 1.9). Phenotypes were recorded on the offspring by the study of 13 radiographic sites. For different skeleton sites, a local score was assigned based on osteochondrosis type anomaly and severity. Thus, a horse had a global score (GS) corresponding to the sum of these local scores (log(1+GS) : mean 0.76 ± 0.85). An ordered categorical outcome (GCC) was defined from GS where controls were defined with a GS of 0, cases with a GS ≥ 2; otherwise it was intermediate (but treated as controls). A case-control set on fetlock (FCC) and hock (HCC) were also defined. Table 1 shows the distribution among cases and controls by measurements. No fixed effects were found significant (pvalue> 0.2 for sex, pvalue > 0.13 for age).

Preliminary analyses. Illumina Bead-Chip EquineSNP50 was used for genotyping. This array includes 54602 evenly distributed SNPs throughout the genome (mean 0.043 ± 0.055 Mb between two SNPs). Figure 1a shows the distribution of SNPs on the $31 + X$ chromosomes.

*INRA, UR 631, SAGA, 31326 Castanet Tolosan, France
†ENVA, CIRALE, 14430 Dozule, France
‡INRA, UMR 1313, GABI, 78350 Jouy-en-Josas, France

A combination of quality criterion was used to discard poor quality markers. The filter was based on the percentage of samples genotyped for a given SNP (call freq > 0.8), the minor allele frequency (MAF $> 5\%$) and the result of an Hardy Weinberg test (pvalue $> 10^{-8}$). The X chromosome was not included in the analysis. Finally, 41249 SNPs were analysed. Linkage disequilibrium (LD) within each chromosome was estimated on 267 genotyped horses. Average r^2 was obtained using a range of distance from 0 to 1.6 Mb by windows of 0.02 Mb. The expected extent of LD assuming an effective size of population with 150 and 1000 individuals was also estimated using the Tenesa et al. (2007) method.

Statistical analyses. The data were analyzed using four methods:

1. Discrete measurements, linkage disequilibrium (LD) : Following Sasieni (1997), we used the genotype Cochran-Armitage trend test (CA) :

$$T = \frac{N(N\sum_{i=1}^{3} r_i x_i - R\sum_{i=1}^{3} n_i x_i)}{R(N-R)(N\sum_{i=1}^{3} n_i x_i^2 - (\sum_{i=1}^{3} n_i x_i)^2)} \sim \chi_1^2$$

   with n_i (resp. r_i) the number of horses (resp. of cases) with genotype i, $\sum_{i=1}^{3} n_i = N$. We choose $x = (0, \lambda, 1)$ ($\lambda = \{0, 0.5, 1\}$ for recessif, additif and dominant) on the measurement GCC (CA-R,CA-A,CA-D), and only $\lambda = 0.5$ for the measurements HCC (CA-HCC) and FCC (CA-FCC).
2. Quantitative measurement, LD : We used the polygenic-SNP mixed model (PS) $log(1 + GS) = \mu + SNP + Zu + e$, with $u \sim N(0, A\sigma_a^2)$ the polygenic effect with 2796 horses in the pedigree, $e \sim N(0, I\sigma_e^2)$ the residual and SNP the genotype coded $\{0, 1, 2\}$. We used the ASReml software to compute this model (Gilmour et al., 2002).
3. Quantitative measurement, linkage analysis (LA) : We used the model : (H0) $log(1 + GS_{ij}) = s_i + e_{ij}$ against (H1) $log(1 + GS_{ij}) = s_i \times x_{ij}\beta_i + e_{ij}$ where s_i correspond to the sire i, β_i the SNP effect of the family i and e_{ij} the error. Only heterozygous sires with at least 2 offspring were used. $x_{ij} = \{1, 0, -1\}$ if the progeny genotype is $\{11, 12, 22\}$ and otherwise $x_{ij} = 0$.
4. Quantitative measurement, LDLA : We kept horses from sire-families with at least 2 offsprings and used the model :

$$log(1 + GS_{ij}) = \mu + (P_i - \bar{P})B_b + Z_{ij}B_w + e_{ij}$$

   with $P_i = \{-1, 0, 1\}$ for sire genotype $\{11, 12, 22\}$ and $Z_{ij} = \{-1, 0, 1\}$ for sire genotype 12 and offspring genotype $\{11, 12, 22\}$ or else $Z_{ij} = 0$. We test (H0) $B_w = 0$ against (H1) $B_w \neq 0$. This model is similar to the QTDT of Abecasis et al. (2000) but adapted to half-sib families.

Criteria for selecting regions of interest. Initially and for each method, we selected the maximum by windows of 7 Mb in which $-log10(pvalue) > 3.5$. From these maxima, we defined two criteria : (crit1) at least 5 SNP at ± 5 Mb of the max have a $-log10(pvalue) > 2$ and (crit2) $pvalue(Max) < pvalue(Bonferroni)$ (chromosome-wide Bonferroni).

Table 1: Distribution for global (GCC), hock (HCC) and fetlock (FCC) measurements

	Measurements		
	GCC	FCC	HCC
Number of cases	241	168	116
Number of controls	268 (16)[a]	357	409
Total	525	525	525

[a] In parenthesis, intermediate horses treated as controls

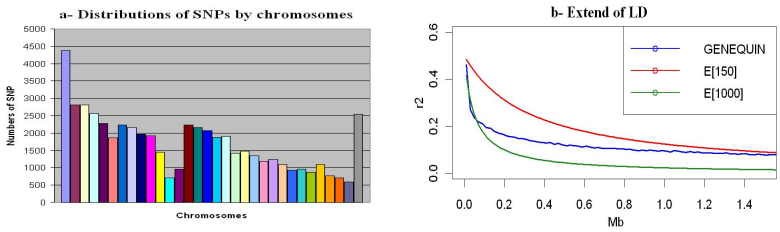

Figure 1: (a) Distribution of SNPs by chromosomes ranged from 1 to 31 and the last is X (b) Extent of LD (r^2) along a distance of 1.6 Mb

Results and discussion

Extent of LD. Figure 1b shows the extent of LD obtained in a genotyped subsample of individuals. The strong decrease of LD for short distances indicates that the effective size of the population was quite high in the past. Conversely, we can see a slight decrease in the long distance which indicates a recent effective size of population much lower.

Detection of regions of interest. Table 2 shows that Crit 1 gives various numbers of regions according to the method (4 to 26). The use of the conservative Crit 2 reduced these numbers to about 75%. LD methods (CA-A, PS) provide many regions (17 and 18) while this number decreased with the use of LA. The use of discrete or quantitative measurements (CA-A and PS), shows roughly the same number of regions. The dominant and recessive hypotheses seem to show less locations than additive methods. There are more regions on hock (26), then on the global measurements (17 and 18) and finally on fetlock (12). There are 11 common regions between LD methods presented in Table 3. These regions are often associated with a region of interest on fetlock or hock but not both. LD and LDLA methods have 2 common regions (ECA 13 and 14). One region is the same for HCC and FCC (ECA 22). CA-A shows 17 regions of interest, and 4 of these are associated with a region on fetlock and 6 on hock.

Discussion. The large number of regions detected on hock suggests that this site is especially informative for OC. The global study of OC gives us several regions of interest, often associated with fetlock or hock, but not both. This suggests that the global analysis could be considered as an OC indicator, but this suggests also that regions of interest are different between sites as hock and fetlock. LD methods have many common regions and therefore the use of discrete or quantitative measurements seems consistent. Methods with LA, used here, are not really adapted to our data because the sire families are small.

Table 2: Number of regions of interest by measurements, methods and criteria

	Methods							
	CA-A	CA-FCC	CA-HCC	CA-D	CA-R	PS	LDLA	LA
Crit 1	17	12	26	10	10	18	9	4
Crit 2	5	3	6	5	2	3	2	0

Table 3: Common regions between LD methods and significance for HCC and/or FCC

	pvalue		
ECA	CA-A	PS	HCC/FCC
2	2.10^{-5}	10^{-4}	FCC
3	10^{-4}	10^{-4}	-
3	2.10^{-4}	10^{-4}	HCC
5	7.10^{-6}	10^{-4}	FCC
11	2.10^{-4}	2.10^{-4}	HCC
13	8.10^{-5}	10^{-4}	-
13	9.10^{-5}	10^{-5}	FCC
14	2.10^{-4}	2.10^{-5}	HCC
15	10^{-4}	3.10^{-4}	-
15	9.10^{-6}	5.10^{-5}	FCC
28	2.10^{-6}	7.10^{-5}	-

Conclusion

This is the first time that a whole-genome association study is made on osteochondrosis with a equine chip SNP 54k. Analysis of grouped sites with the polygenic-SNP model shows many regions of interest and also seems to capture some regions found in the hock and fetlock analysis. This paves the way for a selection against osteochondrosis susceptibility and a fine mapping of regions of interest. The results also show that osteochondrosis on hock and fetlock do not seem to be controlled by the same regions. Others measurements will be studied and haplotypes or global methods, which test all SNPs simultaneously, will be used in the future.

Acknowledgments

The GENEQUIN program was financially supported by INRA, ANR, Haras-Nationaux, Fonds Eperon and Région basse-Normandie.

References

Abecasis, G., Cardon, L., and Cookson, W. (2000). *Am. J. Hum. Genet.*, 66:279–292.
Dierks, C., Löhring, K., Lampe, V., et al. (2007). *Mamm. Genome*, 18:739–747.
Gilmour, A., Gogel, B., Cullis, B., et al. (2002). *ASReml User Guide Release 1.0*.
Sasieni, P. (1997). *Biometrics*, 53:1253–1261.
Tenesa, A., Navarro, P., Hayes, B., et al. (2007). *Genome Research*, 17:520–526.
Wittwer, C., Löhring, K., Drögemüller, C., et al. (2007). *Animal Genetics*, 38:350–357.