Imprint of selection in pedigrees of modern bread wheat varieties

Horvath A, Koenig J, Didier A, Charmet G, Balfourier F

INRA, UMR 1095 Genetics, Diversity and Ecophysiology of Cereals, F-63100 Clermont-Ferrand, France

ABSTRACT

Signature of selection has been studied on a sample regrouping of wild emmer and bread wheat genotypes. Progenitors representing the whole pedigree of seven elite bread wheat lines were compared and contrasted to the wild ancestor from the point of view of haplotype diversity, linkage disequilibrium patterns and temporal changes of allele frequencies along the chromosome 3B. Three chromosomic regions of 3B are suspected to be influenced by human selection. When comparing wild and cultivated genotypes, one chromosomic region seems to have been the subject of a domestication event. After this approach, focusing on one chromosome only, we want to extend our studies to the whole genome using the same methodology.

INTRODUCTION

Hexaploid bread wheat (*Triticum aestivum* L.) has been cultivated since the Neolithic period. First steps of human selection began in the middle of the 19th century. This selection has been boosted thanks to hybridisation techniques and to the introduction of new exotic material during the 20th century, resulting in high-yield varieties with good baking quality.

In the present work we aim to detect the imprint of selection by tracing back haplotypes in modern French bread wheat varieties, using pedigree information. It was postulated that regions of the genomes showing a positive selective value present a strongly reduced allelic diversity and a longer range of local linkage disequilibrium (LD) pattern due to hitch-hiking effects[1]. According to Goldringer and Bataillon[2], temporary changes of allele frequencies (Fc) between different generations at a particular locus also indicate an effect of selection. Based on these facts, we make an attempt to describe haplotypic variability within pedigrees of elite breeding lines and to compare local LD and Fc patterns, first along chromosome 3B, then to extend the analysis to the whole genome.

MATERIAL AND METHODS

Pedigrees of a selected set of seven modern varieties have been reconstructed back to the landraces with the help of the Wheat Pedigree Online Database (http://genbank.vurv.cz/wheat/pedigree/). Analyses have been obtained for 242 lines in total. The oldest cultivars date from 1830, so 10-12 generations of selection are recovered in this material.

Progenitors have been split into 3 categories: 1) landraces and old cultivars (N=74), 2) semi-modern cultivars (N=93) and 3) modern cultivars (N=75), representing 3 periods of selection history.

31 genotypes of B-genome ancestor wild emmer (*Triticum turgidum* ssp. *dicoccoides*) have been chosen in order to compare cultivated and wild gene pools.

Genotyping was firstly performed with 47 SSR (microsatellite) markers mapped on the chromosome 3B, then DArT (Diversity Array Technology) markers were added. DArT fingerprinting was undertaken at Diversity Arrays Technology Pty Limited (Canberra, Australia, http://triticarte.com.au) with the panel Wheat *Pst*I(*Taq*I) v2.3.

Basic statistics of genetic diversity were calculated with GENETIX software[3]. Polymorphic Information Content (PIC) values were estimated according to Nei[4].

Haplotype constitution, frequencies and haplotype diversity index ($H=1-\sum p_i^2$) were estimated with haplo.em algorithm of haplo.stats package[5] of R language and environment for statistical computing[6] using a sliding window (N=4 markers) along the chromosome.

Temporal changes of allele frequencies (Fc) were analysed with NeEstimator software[7] using the method of Waples[8].

LD patterns were calculated with TASSEL software[9] using the DArT concensus map[10].

RESULTS AND DISCUSSION

Allele number and mean Polymorphic Information Content (PIC) values decreased from the wild gene pool to cultivated material (Table 1). This tendency is not surprising. Thus, despite the unbalanced sample sizes, the most important difference is observed between wild emmer genotypes and old cultivars, due to evolution history and domestication syndrome. Modern cultivars are characterised by the lowest allele number and PIC value.

	Mean allele number	Mean PIC value
Wild emmer (N=31)	9.383	0.715
Old wheat cultivars (N=74)	6.382	0.515
Semi-modern cultivars (N=93)	6.595	0.534
Modern cultivars (N=75)	5.553	0.4873

Table 1: Mean allele number and mean PIC values

Haplotype diversity was compared in three different groups: wild emmer, old cultivars and modern elite lines. Figure 1 presents haplotype diversity index (H) patterns in sliding windows along the chromosome 3B. Wild emmer genotypes show stable haplotype diversity along the whole chromosome, with a value about H=0.96. On the other hand, modern varieties present reduced haplotype diversity compared to the old cultivars. This loss of allelic richness between landraces and registered varieties has already been reported by Roussel et al[11]. In our study, four important decreases are observed in sliding windows 6, 9-11, 14 and 18, called region 1, 2, 3 and 4, respectively. This decrease in diversity separates modern elite cultivars from their progenitors in regions 1, 2 and 3. Nevertheless, in region 4 both cultivated groups present an identical profile, which indicates a spectacular contrast with the wild ancestor ssp. *dicoccoides*.
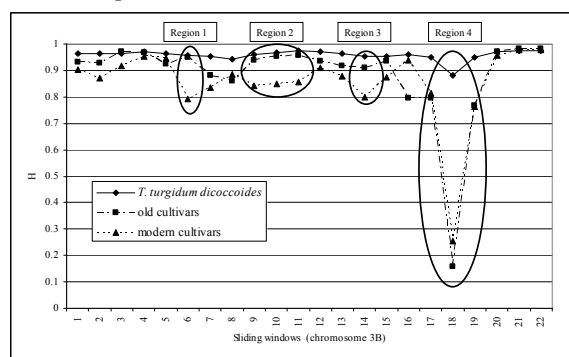

Figure 1: Haplotype diversity patterns along the chromosome 3B
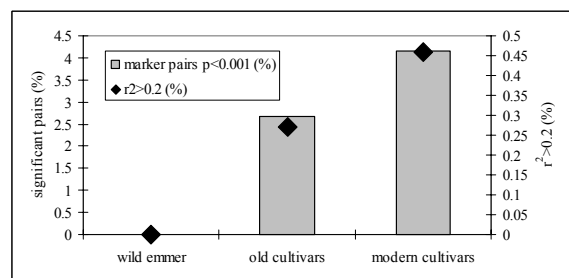

Figure 2: LD increase tendency in three groups of genotypes.

Based on this first observation, we can develop our working hypothesis concerning the presence of signature of selection in regions 1-3. According to Rafalski and Morgante[1], reduction of diversity due to the selection effect is accompanied by a larger linkage disequilibrium pattern. The above-mentioned three groups were compared for pair-wise LD pattern between SSR markers along the chromosome 3B. Globally, the number of significant (p<0.001) pairs and r^2 values increase from wild ancestor to modern elite lines (Figure 2). In the wild emmer group there is no important linkage disequilibrium. In old cultivars, LD is observed in region 2 (r^2=0.2, p<0.0001). In this region the LD is higher in modern elite lines and it extends on more

markers, with r^2 values 0.2-0.4 (p<0.0001). In addition to this, in modern cultivars highly significant, but low-level LD is observed along the region 1. There is no LD observed in regions 3 and 4.

Temporary changes of allele frequencies (Fc values) were first calculated between wild emmer and old cultivars, and then old and modern cultivars were compared. Globally, Fc values change similarly, with higher values when comparing wild ancestor to old bread wheat cultivars (Figure 3). All of the regions suspected to be under selection show local temporary changes of allele frequencies. Fc values are the highest (Fc=1.2) in region 4 in the analysis between wild emmer and old bread wheat cultivars.
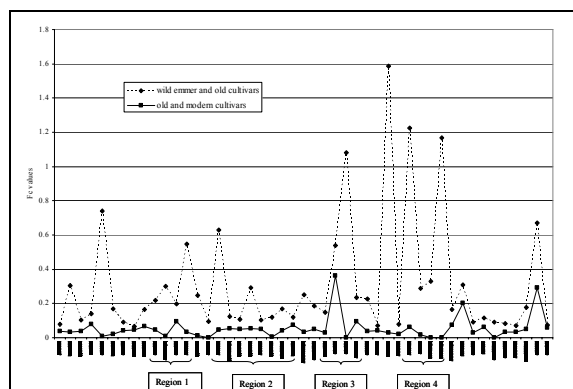

Figure 3: Temporary changes of allele frequencies along the chromosome 3B

A complementary study has been initiated in order to fit analysis methodology developed on one chromosome model to the whole genome. For the time being we can report results concerning genomes A and B of bread wheat. After DArT genotyping, 548 markers of the panel revealed polymorphism. Markers have been positioned according to the consensus map[10] and then redundancies identified. On the whole, we analysed 246 markers divided onto chromosomes of A and B genomes (except for chromosome 5A), with a mean of 18.92 markers/chromosome. Haplotype diversity analysis of landrace progenitors and modern elite cultivars reveals a global tendency for diversity reduction in modern cultivars. When comparing LD patterns, modern varieties present (almost in every case) higher r^2 values than their progenitors, particularly at chromosomes 1A, 2A, 3A and 4A (Figure 4). In future studies we will make an attempt to describe the extension of LD patterns between adjacent locus pairs chromosome per chromosome, as Somers[12] et al. did with SSR markers, in order to discriminate modern elite lines and their progenitors at the LD extension level.
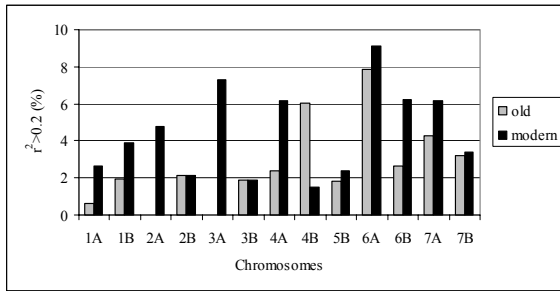
Figure 4: Significant (p<0.001) LD values (r^2) of A and B genomes chromosomes in elite inbred cultivars and their genitors

CONCLUSION

In conclusion, in regions 1 and 2 the loss of haplotype diversity is accompanied by higher linkage disequilibrium and local changes of allele frequencies. The presence of selection imprint in these regions seems to be confirmed by known QTLs of time to maturity and test weight[13] localised in region 1. Similarly, QTLs of grain protein yield and total N amount[14] have been identified in region 2. In region 3 the reduction of haplotype diversity and local change of allele frequencies can be observed, but no significant LD is detected; thus we cannot reach conclusions about this chromosomic region. Region 4 presents a huge difference of haplotype diversity between wild and cultivated gene pools and an important local increase of allele frequencies when comparing wild ancestor and old varieties. LD is not detected, but these results suggest an ancient event, probably related to the evolution and domestication of bread wheat.

The preliminary genome-wide study suggests that the analysis methodology developed for one-chromosome approach can be successfully adapted to the whole genome.

ACKNOWLEDGEMENTS

REFERENCES

1 Rafalski, A.; Morgante, M. 2004. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. TRENDS in Genetics 20: 103-111.

2 Goldringer, I.; Bataillon, T. 2004. On the distribution of temporal variations in allele frequency: consequences for the estimation of effective population size and the detection of loci undergoing selection. Genetics 168: 563-568.

3 Belkhir, K.; Borsa, P.; Chikhi, L.; Raufaste, N.; Bonhomme, F. 2004. Genetix 4.05.2, Logiciel sous Windows™ pour la Génétique des Populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

4 Nei, M. 1973. Analysis of gene diversity in subdivised populations. PNAS 70: 3321-3323.

5 Sinnwell, J.P.; Schaid, D.J.; Yu, Z. 2007. The haplo.stats package: Statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous, Version 1.3.1. Mayo Foundation for Medical Education and Research.

6 R Development Core Team 2006. R: A language and environment for statistical computing, Version 2.4.0. R Foundation for Statistical Computing, Vienna, Austria.

7 Peel, D.; Ovenden, J.R.; Peel, S.L. 2004. NeEstimator: software for estimating effective population size. Queensland Government, Department of Primary Industries and Fisheries, www.2.dpi.qld.gov.au/fishweb/13887.html

8 Waples, R.S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency Genetics. 121:379-391.

9 Zhang, Z.; Bradbury, P.J.; Kroon, D.E.; Casstevens, T.M.; Buckler, E.S. 2006. TASSEL 2.0: a software package for association and diversity analyses in plants and animals (www.maizegenetics.net). Plant & Animal Genomes XIV Conference, Poster P956/CP012, San Diego, USA.

10 Crossa, J.; Burgueno, J.; Dreisigacker, S.; Vargas, M.; Herrera-Foessel, S.A.; Lillemo, M.; Singh, R.P.; Trethowan, R.; Warburton, M.; Franco, J.; Reynolds, M.; Crouch, J.H.; Ortiz, R. 2007. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177: 1889-1913.

11 Roussel, V.; Koenig, J.; Beckert, M.; Balfourier, F. 2004. Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. TAG 108: 920-930.

12 Somers, DJ.; Banks, T.; DePauw, R.; Fox, S.; Clarke, J.; Pozniak, C.; McCartney, C. 2007. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50:557-567.

13 McCartney, C.A.; Somers, D.J.; Humphreys, D.G.; Lukow, O.; Ames, N.; Cloutier, S.; McCallum, B.D. 2005. Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 x 'AC Domain'. Genome 48: 870-883.

14 Laperche, A.; Brancourt-Hulmel, M.; Heumez, E.; Gardet, O.; Hanocq, E.; Devienne-Barret, F.; Le Gouis, J. 2007. Using genotype x nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. TAG 115: 399-415.