

HAPLOTYPE RECONSTRUCTION FROM UNPHASED GENOTYPE DATA AT THE BOVINE PRKAG3 GENE

DETERMINAZIONE INDIRETTA DEGLI APLOTIPI DEL GENE PRKAG3 BOVINO A PARTIRE DA DATI GENOTIPICI

ROBERTA CIAMPOLINI ⁽¹⁾, ELENA CIANI ⁽²⁾, MATTHIEU ROUX ⁽³⁾,
FRANCESCA CECCHI ⁽¹⁾, ELISA MAZZANTI ⁽¹⁾,
MARIELLA TANCREDI ⁽⁴⁾, ELISABETTA CASTELLANA ⁽²⁾,
SILVANO PRESCIUTTINI ⁽⁵⁾, VALERIE AMARGER ⁽⁶⁾

SUMMARY

Haplotype analysis has become an area of intense research, both for population genetics studies and for molecular dissection of complex phenotypes. Haplotypes provide increased informativeness with respect to single nucleotide polymorphisms and allow to condense information on genomic variation, with a sensible gain of power in association studies.

Direct haplotyping via molecular analysis provides more exact information per individual, but is much more expensive and labour-intensive than indirectly inferring haplotypes from genotypes. Several methods have been proposed so far to indirectly reconstruct haplotypes from unphased genotypes. In the present study, we adopted four different approaches (implemented in the computer programs ARLEQUIN, HELIXTREE, HAP and PHASE) to infer phase information from genotypic data on 197 subjects at 14 polymorphic sites of the bovine PRKAG3 gene. In the whole, twelve different haplotypes had been inferred by all the four different methods, although at slightly varying frequencies. Other three haplotypes were inferred by at least two different approaches and some haplotypes were unique to a single method. No dramatic differences among the four selected approaches were observed for the considered genomic target. This is probably due to the linkage-disequilibrium structure of the analysed region and to the moderate amount of missing genotype data.

Keywords: haplotype reconstruction; bovine; PRKAG3 gene.

RIASSUNTO

L'ambito di analisi degli aplotipi evidenzia un interesse crescente da parte della ricerca, sia in termini di studi di genetica delle popolazioni che per la comprensione delle basi molecolari

⁽¹⁾ Dipartimento di Produzioni Animali, Direttore Prof. Paolo Verità.

⁽²⁾ Dipartimento di Fisiologia Generale ed Ambientale, Direttore Prof. Giovanna Valenti, Università di Bari.

⁽³⁾ Unità di Genetica Molecolare Animale, UMR1061 INRA, Direttore Prof. Hubert Leveziel, Università di Limoges, Francia.

⁽⁴⁾ Collaboratore esterno.

⁽⁵⁾ Dipartimento di Patologia Sperimentale, Biotecnologie Mediche, Infettivologia ed Epidemiologia, Centro di Genetica Statistica, Direttore Prof. Alessandro Casini.

⁽⁶⁾ Fisiologia dell'Adattamento Nutrizionale, INRA, Direttore Prof. Dominique Darmaun, Nantes, Francia. Ricerca finanziata da Istituto Zooprofilattico Sperimentale dell'Umbria e delle Marche.

di caratteri complessi. Gli aplotipi forniscono un livello maggiore di informatività rispetto alle mutazioni a carico di singole basi nucleotidiche e consentono di condensare l'informazione relativa alla variabilità genetica, permettendo, altresì, di acquisire un incremento nella potenza di analisi in studi di associazione.

I metodi diretti di ricostruzione dell'aplotipo basati sull'analisi molecolare forniscono un livello di informazione più accurato ma presentano il grosso inconveniente di essere estremamente costosi e laboriosi rispetto ai metodi indiretti di inferenza. Molteplici approcci sono stati proposti ad oggi per la ricostruzione indiretta degli aplotipi a partire da dati genotipici a fase gametica ignota. Nel presente lavoro, sono stati adottati quattro diversi metodi (implementati nei programmi ARLEQUIN, HELIXTREE, HAP e PHASE) per l'inferenza della fase gametica a partire dai dati genotipici di 197 soggetti relativi a 14 siti polimorfici del gene PRKAG3 bovino. Complessivamente, dodici diversi aplotipi sono stati inferiti da tutti i quattro metodi adottati, sebbene con frequenze relativamente divergenti. Altri tre aplotipi sono stati inferiti da almeno due dei quattro diversi approcci mentre alcuni aplotipi sono stati inferiti esclusivamente da uno dei quattro metodi. Complessivamente, non sono state osservate differenze sostanziali tra i diversi approcci adottati relativamente alla regione genomica di interesse. Ciò è presumibilmente ascrivibile al profilo di linkage disequilibrium della regione analizzata ed alla presenza molto contenuta di dati genotipici mancanti.

Parole chiave: ricostruzione aplotipica; bovini; gene PRKAG3.

INTRODUCTION

The protein kinase adenosine monophosphate-activated γ 3-subunit (PRKAG3) gene encodes the muscle specific isoform of the regulatory γ subunit of adenosine monophosphate activated protein kinase (AMPK), which plays a key role in regulating energy homeostasis in eukaryotes (Carling, 2004). In pigs, the PRKAG3 gene have been shown to affect muscle glycogen content and hence meat quality traits like ultimate pH, meat colour, water-holding capacity, drip loss, tenderness and cooking loss (Milan et al., 2000; Ciobanu et al., 2001).

Single nucleotide polymorphisms (SNPs) analysis is a well-established tool for the identification of genes associated with traits of economic interest in livestock populations (Yu et al., 1995; Page et al., 2002; Ge et al., 2003; Schenkel et al., 2005; Allan et al., 2007). More recently, haplotype analysis has become an area of intense research for complex genetic phenotypes (Ciobanu et al., 2004; Stone et al., 2005; Khatib et al., 2006). An haplotype can be defined as a set of SNPs on a single chromosome that are closely linked and inherited as a unit. The knowledge of haplotype for several SNPs in one gene could provide more information about genotype-phenotype associations than individual underlying SNPs. Crucial for such studies is the need for phase-known information. Direct and indirect approaches have been proposed to infer phase information. Molecular haplotyping (for a detailed reference list, see Konfortov et al., 2007) provides more exact information per individual, but is much more expensive and labour-intensive than indirectly inferring haplotypes from genotypes. This could severely constrain the size of the analysed sample. It has been shown that haplotype-inference methods such as parsimony

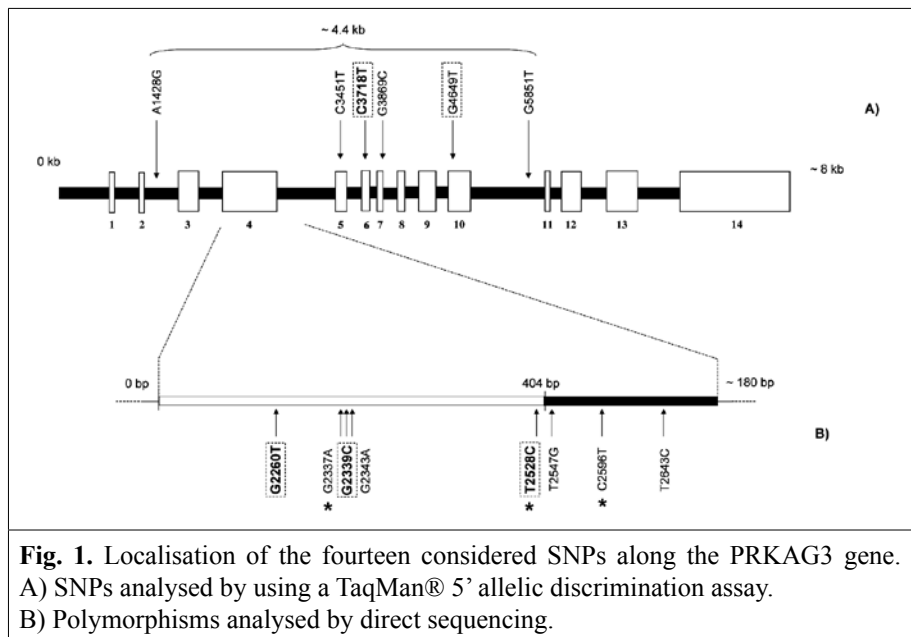
(Clark, 1990), expectation-maximization (Excoffier & Slatkin, 1995), and Bayesian approaches (Stephens et al., 2001) are fairly effective and economical alternatives to molecular-haplotyping methods (Niu et al., 2002).

In the present study, we explored four indirect approaches to infer phase information from genotypic data at the bovine PRKAG3 gene.

MATERIALS AND METHODS

Blood samples from 97 Chianina and 100 Holstein non-consanguineous animals were collected. Genomic DNA was isolated from whole blood following the procedure of Rouzaud et al. (2000) and Jeanpierre (1987) for the Holstein and the Chianina samples, respectively. PCR primer pairs and conditions are available upon request.

In the whole, 14 single nucleotide polymorphisms were considered, spanning a region of 4.4 kb in the bovine PRKAG3 gene (Fig. 1). A TaqMan® 5' allelic discrimination assay (Applied Biosystems) and a sequencing approach using Big Dye Terminator chemistry on a ABI PRISM 310 cycle sequencing were adopted to obtain genotypic data.



Haplotype frequency estimates were obtained by using the software ARLEQUIN v. 2.000 (Schneider et al., 2000) and the HAP web-server (Halperin & Eskin, 2004), available at <http://research.calit2.net/hap/WebServer.htm>. In addition, the computer program PHASE v. 2.1 (Stephens et al., 2003) and the HELIXTREE software package (Golden Helix, Inc., Bozeman, MT) were also used to infer frequency estimates.

RESULTS AND DISCUSSION

In the present study, four different approaches were adopted to infer phase information from genotypic data at the PRKAG3 gene on cattle.

The ARLEQUIN software implements the Expectation-Maximization algorithm by Excoffier and Slatkin (1995), which is based on an iterative process aiming at obtaining the maximum-likelihood estimates of haplotype frequencies. Unlike the other three computer programs, ARLEQUIN do not provide, for each individual animal in the population sample, a list of the most probable diplotypes, i.e. both chromosomal haplotypes. This is a serious limitation, mainly when haplotype reconstruction is performed in order to carry out association studies.

HELIXTREE is a fast, easy-handled software which also implements the Expectation-Maximization algorithm by Excoffier and Slatkin (1995); unlike than with ARLEQUIN, we adopted here the “composite haplotype method” (CHM) to estimate haplotype probabilities.

The computer program PHASE has the main advantage of allowing for recombination and decay of Linkage Disequilibrium (LD) with distance; in some cases, this could result in more accurate haplotype estimates.

The software HAP implements a method for resolving haplotypes within blocks of limited diversity based on the perfect phylogeny model of Gusfield (2002). This model assumes an infinite site mutation model and allows no recombinations. Since these assumptions are not realistic, HAP presents a relaxed phylogeny model that allows for a certain number of recurrent mutations and recombinations within a block. The input file only requires the first allele at a single site to be indicated (if the second allele is a different one, user must only specify “heterozygous”). Thus, if the second allele is not present in the homozygous state within the analysed population, the software is unable to resolve it and it will indicate the allele with a “M”.

Results are summarised in Tab. I. In the whole, twelve identical haplotypes had been inferred by all the four different methods, although at slightly varying frequencies. Other three haplotypes were inferred by at least two different approaches. Fifteen haplotypes (bottom part of Tab. I) were instead unique to a single method. Anyway, it should be noted that nine out of ten haplotypes unique to the ARLEQUIN software could be easily ascribed to corresponding “shared” haplotypes (top part of Tab. I) by only substituting missing variant sites (indicated by “?”). In this manner, results obtained with the ARLEQUIN software would be very similar to those obtained by the HELIXTREE program (data not shown), which is not surprising considering that they adopt almost the same algorithm.

CONCLUSION

Our results show that, in this particular case, there is no dramatic difference among the four selected approaches in reconstructing haplotypes from unphased genotypes. This is probably due to the linkage-disequilibrium structure of the analysed region and to the moderate amount of missing genotype data.

Slight discrepancies among the available haplotyping computer programs could be a minor worry when haplotype inference is needed in order to carry out population genetics studies. However, accurate phase-inference algorithms are a necessary prerequisite in haplotype-based association analysis. Thus, considerable attention is currently being given to haplotyping related issues.

Tab. I. List of haplotypes inferred using four different approaches.

HELIXTREE		HAP		ARLEQUIN		PHASE	
	freq.		freq.		freq.		freq.
AGGGGTGCTCCGGG	0,521	AGGGGTGCTCCGGG	0,505	AGGGGTGCTCCGGG	0,474	AGGGGTGCTCCGGG	0,497
GTGGGTCCCCGGG	0,298	GTGGGTCCCCGGG	0,312	GTGGGTCCCCGGG	0,299	GTGGGTCCCCGGG	0,315
GGGGATGCCTCCGT	0,081	GGGGATGCCTCCGT	0,079	GGGGATGCCTCCGT	0,074	GGGGATGCCTCCGT	0,079
GGAGGCGTCCGGG	0,024	GGMGMGM/TCCGGG	0,028	GGAGGCGTCCGGG	0,010	GGAGGCGTCCGGG	0,010
GGGGGTGCCCGGG	0,019	GGGGGTGCCCGGG	0,020	GGGGGTGCCCGGG	0,021	GGGGGTGCCCGGG	0,025
AGGGGTGCTCTGGG	0,017	AGGGGTGCTC MGGG	0,018	AGGGGTGCTCTGGG	0,016	AGGGGTGCTCTGGG	0,023
AGGGGTGCCCGGG	0,008	AGGGGTGCCCGGG	0,008	AGGGGTGCCCGGG	0,008	AGGGGTGCCCGGG	0,008
GGAGGCGTCCGGG	0,006	GGMGMGM/CCCGGG	0,003	GGAGGCGTCCGGG	0,003	GGAGGCGTCCGGG	0,020
AGGGGTGCTCCCGG	0,005	AGGGGTGCTCCCGG	0,005	AGGGGTGCTCCCGG	0,005	AGGGGTGCTCCCGG	0,005
AGGGATGCCCGGG	0,005	AGGGATGCCCGGG	0,005	AGGGATGCCCGGG	0,005	AGGGATGCCCGGG	0,005
GGGGGTGCCCTGGG	0,005	GGGGGTGCC MGGG	0,005	GGGGGTGCCCTGGG	0,004	-	-
AGGGGTGCTTCGGG	0,003	AGGGGTGCTTCGGG	0,003	AGGGGTGCTTCGGG	0,003	AGGGGTGCTTCGGG	0,003
GGGGATGCTCCGG	0,003	-	-	GGGGATGCTCCGG	0,003	-	-
GTGGGTGCCCGGG	0,003	GTGGGTGCCCGGG	0,003	GTGGGTGCCCGGG	0,003	GTGGGTGCCCGGG	0,003
GTGCGTCCCCGGG	0,003	GTGMGTCCCCGGG	0,003	GTGCGTCCCCGGG	0,003	-	-
		GGGGATCCTCCGG	0,003	AGGGGTGC?CCGGG	0,015	GGGGATGCTCCGG	0,003
		GTGGGTGCTCCGGG	0,003	GGAGGCGT?CCGGG	0,015	GTGGGTGCTCCGGG	0,003
				GTGGGTTC?CCGGG	0,013	AGGCGTGCTCCGGG	0,003
				A?GGGTGCTCC?G?	0,005		
				AGG?TTCCTCCGGG	0,005		
				AGGGGTGCT?CCGGG	0,005		
				GGGGATGC?TCCGT	0,005		
				?GGGTGCT?CCGGG	0,003		
				?TAGGCTT?CCGGG	0,003		
				AGGGGTGC?CTGGG	0,003		

REFERENCES

- ALLAN M. F., THALLMAN R. M., CUSHMAN R. A., ECHTERNKAMP S. E., WHITE S. N., KUEHN L. A., CASAS E., SMITH T. P. L. (2007). Association of a single nucleotide polymorphism in *spp1* with growth traits and twinning in a cattle population selected for twinning rate. *J. Anim. Sci.*, 85: 341-347.
- CARLING D. (2004). The AMP-activated protein kinase cascade-a unifying system for energy control. *Trends Biochem. Sci.*, 29: 18-24.
- CIOBANU D., BASTIAANSEN J., MALEK M., HELM J., WOOLLARD J., PLASTOW G., ROTHSCHILD M. (2001). Evidence for new alleles in the protein kinase adenosine monophosphate-activated gamma(3)-subunit gene associated with low glycogen content in pig skeletal muscle and improved meat quality. *Genetics*, 159 (3): 1151-1162.
- CIOBANU D. C., BASTIAANSEN J. W. M., LONERGAN S. M., THOMSEN H., DEKKERS J. C. M., PLASTOW G. S., ROTHSCHILD M. F. (2004). New alleles in calpastatin gene are associated with meat quality traits in pigs. *J. Anim. Sci.*, 82: 2829-2839.
- CLARK A. G. (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol. Biol. Evol.*, 7: 111-122.
- EXCOFFIER L., SLATKIN M. (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, 12 (5): 921-927.
- GE W., DAVIS M. E., HINES H. C., IRVIN K. M., SIMMEN R. C. M. (2003). Association of single nucleotide polymorphisms in the growth hormone and growth hormone receptor

- genes with blood serum insulin-like growth factor I concentration and growth traits in Angus cattle. *J. Anim. Sci.*, 81: 641-648.
- GUSFIELD D. (2002). Haplotyping as perfect phylogeny: conceptual framework and efficient solutions. In *Proceedings of RECOMB 2002: The 6th Annual International Conference on Computational Biology*, 166-175.
- HALPERIN E., ESKIN E. (2004). Haplotype reconstruction from genotype data using imperfect phylogeny. *Bioinformatics*, 20(12): 1842-1849.
- JEANPIERRE M. (1987). A rapid method for the purification of DNA from blood. *Nucleic Acids Res.*, 15: 9611.
- KHATIB H., LEONARD S. D., SCHUTZKUS V., LUO W., CHANG Y. M. (2006). Association of the *olr1* gene with milk composition in Holstein dairy cattle. *J. Dairy Sci.*, 89: 1753-1760.
- KONFORTOV B.A., BANKIER A.T., DEAR P.H. (2007). An efficient method for multi-locus molecular haplotyping. *Nucleic Acids Res.*, 35(1) e6 doi:10.1093/nar/gkl742.
- MILAN D., JEON J.T., LOOFT C., AMARGER V., ROBIC A., THELANDER M., ROGEL-GAILLARD C., PAUL S., IANNUCELLI N., RONNE H., LUNDSTROM K., REINSCH N., GELLIN J., KALME E., LE ROY P., CHARDON P., ANDERSSON L. (2000). A mutation in *PRKAG3* associated with excess glycogen content in pig skeletal muscle. *Science*, 288: 1248-1251.
- NIU T., QIN Z.S., XU X., LIU J.S. (2002). Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am. J. Hum. Genet.*, 70:157-169.
- PAGE B. T., CASAS E., HEATON M. P., CULLEN N. G., HYNDMAN D. L., MORRIS C. A., CRAWFORD A. M., WHEELER T. L., KOOHMARAIE M., KEELE J. W., SMITH T. P. L. (2002). Evaluation of single-nucleotide polymorphisms in *capn1* for association with meat tenderness in cattle. *J. Anim. Sci.*, 80: 3077-3085.
- ROUZAUD F., MARTIN J., GALLET P.F., DELOURME D. (2000). A first genotyping assay of French cattle breeds based on a new allele of the extension gene encoding the melanocortin-1 receptor (*Mc1r*). *Genet. Sel. Evol.*, 32: 511-520.
- SCHENKEL F. S., MILLER S. P., YE X., MOORE S. S., NKRUMAH J. D., LI C., YU J., MANDELL I. B., WILTON J. W., WILLIAMS J. L. (2005). Association of single nucleotide polymorphisms in the leptin gene with carcass and meat quality traits of beef cattle. *J. Anim. Sci.*, 83(9): 2009-2020.
- SCHNEIDER S., ROESSLI D., EXCOFFIER Y L. (2000). ARLEQUIN: a software for population genetics data analysis, version 2.000. Genetics and Biometry Laboratory. University of Geneva, Switzerland.
- STEPHENS M., DONNELLY P. (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.*, 73: 1162-1169.
- STEPHENS M., SMITH N., DONNELLY P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, 68: 978-989.
- STONE R. T., CASAS E., SMITH T. P. L., KEELE J. W., HARHAY G., BENNETT G. L., KOOHMARAIE M., WHEELER T. L. S., SHACKELFORD D., SNELLING W. M. (2005). Identification of genetic markers for fat deposition and meat tenderness on bovine chromosome 5: development of a low-density single nucleotide polymorphism map. *J. Anim. Sci.*, 83: 2280-2288.
- YU T. P., TUGGLE C. K., SCHMITZ C. B., ROTHSCHILD M. F. (1995). Association of *pit1* polymorphisms with growth and carcass traits in pigs. *J. Anim. Sci.*, 73: 1282-1288.