Human Immunology 71 (2010) 783–788

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Computer-aided human leukocyte antigen association studies: A case study for psoriasis and severe alopecia areata

Daniele Catanzaro [a,*], Marc Andrien [b], Martine Labbé [a], Michel Toungouz-Nevessignsky [b]

[a] Graphs and Mathematical Optimization, Computer Science Department, Université Libre de Bruxelles Brussels, Belgium
[b] Department of Immunology, Erasme Hospital, Université Libre de Bruxelles, Brussels, Belgium

ARTICLE INFO

Article history:
Received 15 January 2010
Accepted 23 April 2010
Available online 4 May 2010

Keywords:
Alopecia areata
Psoriasis
Human leukocyte antigen
Histocompatibility antigens
Disease susceptibility
Computational immunology
Pure parsimony haplotyping
Mixed-integer programming
Combinatorial optimization
Branch-and-bound

ABSTRACT

We present an integer programming model for human leukocyte antigen (HLA) association studies based on the parsimony criterion. The model is simple, compact, easy to implement, and able to handle datasets containing up to 200 phenotypes. Computational experiments carried out on patients affected by psoriasis and severe alopecia areata show that the model is consistent with the experimental haplotype frequencies, showing, for the considered diseases at least, a high reliability of the predictions. These promising results provide perspective on computer-aided association studies and encourage the development of efficient exact computational approaches for haplotype estimation.

1. Introduction

A possible approach to identify genetic risk factors for common diseases or disease conditions such as obesity, cancer, diabetes, and cardiovascular and inflammatory diseases, consists in searching for an association between the variable sites of a specific chromosome region and a disease [1,2]. This approach is generally referred to as an association study and can be performed by exploiting human sequence variation as a genetic marker [3]. In fact, the variable sites causing or associated with the disease are more frequent in a group of affected individuals than in a group of unaffected controls [4]. Hence, by comparing the frequencies of the common patterns of human sequence variation (also called haplotypes) extracted from both groups, the identification of the genetic risk factors becomes possible.

Unfortunately, isolating haplotypes from a group of individuals (phenotypes) is not always an easy task: the process is laborious, cost-prohibitive, requires advanced molecular isolation strategies, and is not even possible when the parental information is missing. For the last reason at least, haplotype estimation via computational methods becomes a valid alternative to the experimental approach.

To rebuild the haplotype map of a population, computational methods have to solve an optimization problem, called the haplotype estimation problem (HEP), consisting of finding a set of haplotypes that, opportunely combined, generate the set of observed phenotypes [5]. Versions of this problem depend on the nature of the criteria used to select a set of haplotypes among plausible alternatives [6–10].

One of the possible criteria proposed in the literature on HEP is the parsimony criterion, which consists of finding the minimal number of distinct haplotypes necessary to explain a given set of phenotypes [11,12]. The parsimony criterion is at the core of several versions of HEP, namely Clark's problem [13], the pure parsimony haplotyping problem [14], the minimum perfect phylogeny problem [9], the minimum recombination haplotype configuration problem [15], the zero recombination haplotype configuration problem [15], and the k-minimum recombination configuration problem [15]. These versions mainly differ from one another either in the nature of the objective function or in the kind of constraints imposed.

In this preliminary article, we investigate the use of the parsimony criterion for human leukocyte antigen (HLA) association studies. Specifically, we introduce a new version of HEP, called the allele-oriented pure parsimony haplotype (AO-PPH) problem and develop a simple and compact integer programming model able to provide exact predictions for datasets containing up to 200 phenotypes.

Computational experiments carried out on patients affected by psoriasis [16] and severe alopecia areata [17] show that the results provided by the model are consistent with the experimental hap-

* Corresponding author.
E-mail address: dacatanz@ulb.ac.be (D. Catanzaro).

0198-8859/10/$32.00 - see front matter © 2010 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.humimm.2010.04.003

lotype frequencies, showing, for the considered diseases at least, a high reliability of the predictions. These promising results give perspective on computer-aided association studies and encourage the development of efficient exact computational approaches to solution of HEP.

1.2. Estimating haplotypes: A mathematical programming model

Assume that a population of n individuals is given and that, for a fixed chromosome region, the corresponding molecular data are known. We denote A as the set of the distinct alleles observed in the population. For example, the set A relative to the three individuals shown in Figure 1a is $\{a_1, a_2, a_3, a_5, a_7, b_1, b_2, b_3, b_8\}$ (Fig. 1b).

The diploid nature of human genome implies that, for a specific locus of a chromosome region, one allele is inherited from the father and one from the mother. Hence, if p is the number of loci in the interested chromosome region, the phenotype of an individual in the population can be thought as a sequence of length 2p over the alphabet A. For example, the first sequence $f_1 = [a_1, a_2, b_1, b_2]$ in Fig. 1a denotes a phenotype relative to a chromosome region made up of two loci, in which the first locus consists of the pair of alleles (a_1, a_2) and the second locus of alleles (b_1, b_2). To our ends, the order of the alleles at the ith locus is not important, *i.e.*, the pairs, *e.g.*, (a_1, a_2) and (a_2, a_1) are considered as synonymous.

Consider a phenotype f having length 2p over an alphabet A. Then a possible *haplotype* h of f can be thought as any sequence of length p over the alphabet A such that the allele at ith locus in h is one of the two alleles at ith locus in f. For example, the sequence $h = [a_1, b_1]$ denotes a haplotype for phenotype f_1 in Fig. 1a. Haplotyping a set of phenotypes means finding a set H of haplotypes such that, for each phenotype f in the population, there exits a pair of haplotypes in H, say h_r and h_s, whose combination generates f. For example, the combination of haplotypes $_{h1} = [a_1, b_1]$ and $_{h2} = [a_2, b_2]$ generates phenotype f_1 in Fig. 1a, in symbols $f_1 = h_1 \in h_2$.

It is worth noting that, given phenotype f relative to a chromosome region of p loci, there exist 2^p possible haplotypes generating f. For example, phenotype f_1 in Fig. 1a can be generated by combining appropriately either the pair of haplotypes $[a_1, b_1]$, $[a_2, b_2]$ or the pair $[a_1, b_2]$, $[a_2, b_1]$. This insight entails the use of a criterion to select pairs of haplotypes among plausible alternatives. The parsimony criterion may help in this circumstance. Specifically, under the parsimony criterion, a haplotype set H is defined to be optimal (or the most parsimonious) for the given phenotypes, if its cardinality (*i.e.*, the number of haplotypes in H) is minimal [5,18,19]. Hence the problem of haplotyping a population of individuals under the parsimony criterion can be formally stated in terms of the following optimization problem:

Problem. *Allele-oriented pure parsimony haplotype (AO-PPH)*
Given a set G of n phenotypes having p loci each, find the minimum set H of haplotypes such that for each phenotype f_k in G there exists a pair of haplotypes $\{h_r, h_s\}$ in H whose combination generates f_k.

As an example, Fig. 1 shows an instance of AO-PPH consisting of 3 phenotypes, the associated set of alleles, the corresponding (optimal) solution, and the relative combinations generating the input population.

Finding an optimal solution to AO-PPH is fundamental to validate the parsimony criterion, *i.e.*, to verify whether, for a given instance of AO-PPH, the results predicted by the criterion match the experimental ones. Unfortunately, it is possible to prove that AO-PPH belongs to the class of the *NP*-Hard problems, *i.e.*, the class of those optimization problems for which no polynomial time solution algorithm is known so far in the literature. An *NP*-Hard problem can be solved by enumerating all of its own solutions, but this approach usually proves computationally very intensive and may become quickly intractable even for small instances of the problem. For example, finding an optimal solution to AO-PPH would imply enumerating all of the possible 2^p haplotypes associated with each phenotype in G, a task that would take years or centuries for values of $p \geq 50$. The need for providing predictions of practical use for association studies justifies the development of alternative approaches to the brute force enumeration. In this context, mathematical programming constitutes a powerful tool for developing exact algorithms able to provide solutions to *NP*-Hard problems in

Input data: The phenotype set

Phenotypes	Locus a		Locus b	
	Alleles		Alleles	
Phenotype 1	a1	a2	b1	b2
Phenotype 2	a1	a7	b1	b3
Phenotype 3	a3	a5	b2	b8

(a)

Corresponding set of alleles								
a1	a2	a3	a5	a7	b1	b2	b3	b8

Allele	Allele occurrence in the population
a1	2
a2	1
a3	1
a5	1
a7	1
b1	2
b2	2
b3	1
b8	1

(b)

Solution: The haplotype set

Haplotypes	Locus a	Locus b
	Allele	Allele
Haplotype 1	a1	b1
Haplotype 2	a2	b2
Haplotype 3	a7	b3
Haplotype 4	a3	b2
Haplotype 5	a5	b8

(c)

Combination

Phenotype 1	=	Haplotype 1	⊕	Haplotype 2
Phenotype 2	=	Haplotype 1	⊕	Haplotype 3
Phenotype 3	=	Haplotype 4	⊕	Haplotype 5

(d)

Fig. 1. (a) Population of three diploid individuals (phenotypes), constituting an instance of AO-PPH. (b) Associated set of alleles and their occurrence in each input phenotype. (c) Minimal set of haplotypes whose combination explains the input phenotypes. (d) Combinations of haplotypes generating each input phenotype.

reasonable computing times. Hence, we shall describe now a possible mathematical programming model for solving AO-PPH.

Denote x_i, for all $i \in H$, as a decision variable equal to 1 if haplotype h_r belongs to the solution and 0 otherwise. Let $H_k \in H$ be the set of haplotypes generating phenotype $f_k \in G$ and denote y_{ij}^k, for all $k \in G$, $i, j \in H_k$, as a decision variable equal to 1 if phenotype f_k is generated by haplotypes h_r and h_s, and 0 otherwise. Then, finding the optimal solution to AO-PPH means solving the integer programming model described below.

1.3. Formulation: Haplotype estimator model

$$\min \sum_{i \in H} x_i \tag{1}$$

$$s.t.\ y_{ij}^k \leq x_i \qquad \forall k \in G,\ \ i, j \in H_k \tag{2}$$

$$y_{ij}^k \leq x_j \qquad \forall k \in G,\ \ i, j \in H_k \tag{3}$$

$$\sum_{ij \in H_k} y_{ij}^k = 1 \qquad \forall k \in G \tag{4}$$

$$x_i \in \{0, 1\} \qquad \forall i \in H \tag{5}$$

$$y_{ij}^k \in \{0, 1\} \qquad \forall k \in G,\ i,j \in H_k. \tag{6}$$

The objective function (1) represents the number of distinct haplotypes or equivalently the cardinality of H. Constraints (2 and 3) impose that haplotypes h_r and h_s generate phenotype f_k only if h_r and h_s belongs to the solution. Finally, constraint (4) imposes that exactly one pair of haplotypes h_r and h_s in H_k can generate phenotype f_k.

The haplotype estimator model (HEM) is compact, characterized by a polynomial number of variables and constraints, relatively easy to implement, and solvable with any standard solver for mixed integer programming (MIP). Specifically, the solution found by the solvers is usually obtained by means of implicit enumeration techniques (*i.e.*, branch-and-bound) and its optimality is ensured by the fundamental theorems at the core of mathematical programming.

2. Subjects and methods

For validating the predictions obtained from the resolution of HEM, we considered two clinical settings previously studied at the Department of Immunology, Hematology and Transfusion of the Erasmus Hospital, Université Libre de Bruxelles, Belgium. Specifically, the first clinical setting consists of 88 unrelated patients affected by severe alopecia areata (27 male and 61 female patients; age at time of the study, 30.7 ± 2.3 years; range, 4–78 years; age at onset of alopecia areata, 24 ± 2.1 years). The second clinical setting consists of 99 unrelated patients affected by psoriasis of type I and II (54 male and 45 female patients; age at time of the study, 54.7 ± 3.67 years; range, 1–91 years; age at onset of psoriasis, 30.96 ± 4.5 years). We tested both settings versus a control group consisting of 100 healthy, unrelated, random-sampled, Belgian individuals [20]. To obtain a high level of resolution, we genotyped the samples for HLA-A*, B*, Cw*, DRB1*, DQA1*, and DQB1* by oligonucleotide probe hybridization, after PCR sequence specific amplification (PCR-SSO oligotyping) for low-resolution typing, followed by polymerase chain reaction sequence-specific primer (PCR-SSP) for high-resolution typing.

We implemented HEM in Mosel 2.0 of Xpress-MP, Optimizer version 18, running on a Pentium 4, 3.2 GHz, equipped with 2 GB of RAM and operating system Gentoo release 7 (kernel linux 2.6.17). During runtime we activated Xpress Optimizer automatic cuts, Xpress presolving strategy, and used Xpress primal heuristic to generate the first upper bound to the problem. We removed duplicate phenotypes in the input clinical settings through a preprocessing step. This reduction does neither interfere with the value of the optimal solution nor reduce the gap (*i.e.*, the difference between the value of the optimal solution and the linear programming relaxation at root node divided by the value of the optimal solution [21]). The MIP solver processed both clinical settings in less than 1-minute computing time and found the respective optimal solutions without branching. Finally, in both clinical settings the gaps were inferior to 0.05% showing that HEM is a tight formulation for the considered instances. Finally, coherently with Schmitt-Egenolf et al. [16] and Marques Da Costa et al. [17], we used the Wilcoxon test to check the statistical significance of the haplotype frequency variations in patients and controls (95% confidence level). No multiple testing (*e.g.*, Bonferroni correction) has been accounted for.

3. Results

3.1. Predictions on the alopecia areata clinical settings

HLA association studies performed on large groups of patients affected by alopecia areata revealed an association with HLA-DRB1*11 (or DR11 antigen), for which DRB1*1104 allele appears to play a central role [22–24]. More recently, Marques Da Costa et al. [17] also observed that HLA-DRB1*1104 allele is more correlated with alopecia areata onset rather than extension considering patchy alopecia areata and alopecia totalis/alopecia universalis. HEM predictions on the alopecia areata clinical setting showed that HLA-DRB1*11-DQB1*03 is the most frequent haplotype in patients, affecting 33 individuals (37.5%) versus 21 controls (21%). The significant variation ofHLA-DRB1*11-DQB1*03 frequency in both patients and controls (p = 5.367e-05) is therefore a confirmation of the findings of Colombe et al. [22], Price and Colombe [23], and Colombe et al. [24] and marks the association between the haplotype and the disease.

A finer analysis on HLA-DRB1*11-DQB1*03 showed (Figure 2) that HLA-DRB1*1101-DQB1*03 is characterized by the highest frequency, being present in 18 patients (20.45%) versus 14 controls

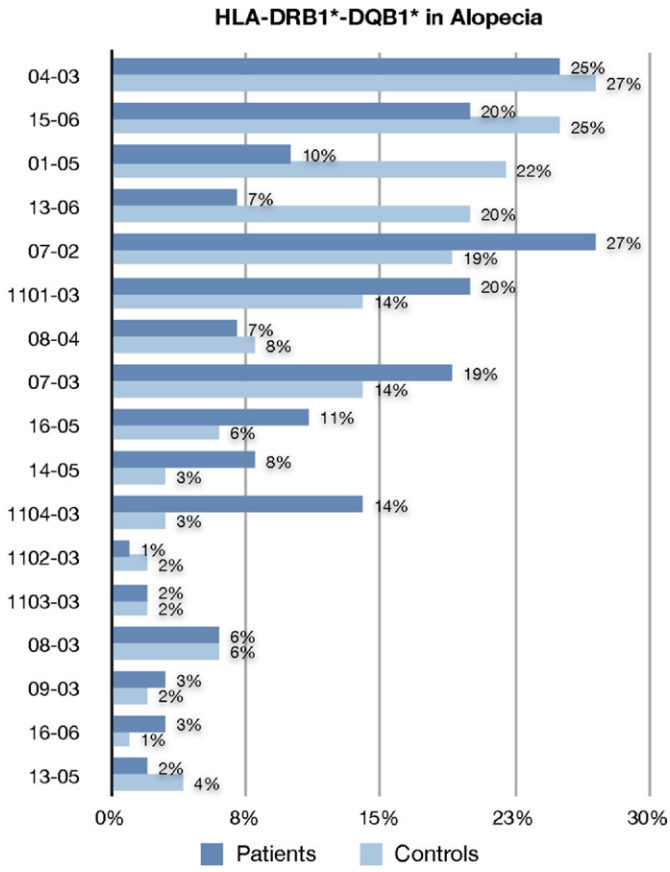

Fig. 2. Predicted haplotype frequencies in controls and patients affected by alopecia areata.

(14%), followed by HLA-DRB1*1104-DQB1* 03 present in 12 patients (14.63%) versus three controls (3%), and by the less frequent HLA-DRB1*1103-DQB1*03 and HLA-DRB1*1102-DQB1*03 present in one patient (1%) versus two controls (2%) and two patients (2%) versus two controls (2%), respectively. Similarly to Marques Da Costa et al. [17], the Wilcoxon test evidenced no significant variation of the HLA-DRB1*1101-DQB1*03 frequency in both groups ($p = 0.089\,93$), and possibly more important, a significant correlation between age of onset of alopecia areata and HLA-DRB1*11-DQB1*03 haplotype. Specifically, HEM predictions evidenced (data not shown) that HLA-DRB1*11-DQB1*03 positive patients developed the first manifestation of the disease already by the age of $16.77 \in 2.5$ years versus $27.4 \in 2.2$ years in negative patients ($p = 0.002\,97$), and that by the age of 20 years, 73.2% of the positive patients had their first disease manifestation versus 40% of the negative ones ($p = 0.0035$). Moreover, the age range further decreases when performing a similar analysis on HLA-DRB1*1104-DQB1*03 positive patients. In fact, HEM predictions evidenced that in these patients the first manifestation of the disease were already present by the age of $12.8 \in 2.5$ years versus $19 \in 8.8$ years in negative patients ($p = 1.311\text{e-}05$), a result that confirms another insight of Marques Da Costa et al. [17], *i.e.*, the earlier occurrence of severe alopecia areata in HLA-DRB1*1104-DQB1*03–positive patients.

Because of the small number of HLA-DRB1*1102-DQB1*03 and HLA-DRB1*1103-DQB1*03–positive patients and controls in the considered setting, we did not investigate their frequencies any further. According to the experiments carried out by Frentz et al. [25] and Welsh et al. [26], the HLA-DRB1*04-DQB1*03 haplotype is prone to be associated with alopecia areata. This result is not confirmed by HEM predictions, which showed no significant variation in the analyzed setting ($p = 0.2744$). Moreover, no significant variation was found for HLA-DRB1*15-DQB1*06 ($p = 0.3889$), HLA-DRB1*07-DQB1*02 ($p = 0.05632$), and the remaining haplotypes characterized by smaller frequencies. On the contrary, HEM predictions showed a statistically significant variation for HLA-DRB1*01-DQB1*05 and HLA-DRB1*13-DQB1*06 ($p = 0.0011$ and $p = 0.0012$, respectively). The high incidence of these haplotypes in controls could be interpreted as a protection factor from the considered disease. However, we prefer to be cautious and suggest that possibly only the use of a larger setting (out of the scope of the present article) could prove useful to investigate the incidence of these haplotypes in the disease.

3.2. Predictions on the psoriasis clinical settings

HLA association studies performed on large groups of patients affected by psoriasis type I and II revealed an association with HLA class I antigens for which HLA-Cw6, HLA-B13, and HLA-B57 alleles appear to play a central role in early onset of the disease [27–30]. This insight was further investigated more recently by Schmitt-Egenolf et al. [16], who confirmed the early occurrence of psoriasis in HLA-B57-Cw6–positive patients, and observed that the Caucasian extended haplotype HLA-A1-B57-Cw6-DRB1*0701-DQB1*0303 is also highly correlated with psoriasis onset.

Also, in this case, the predictions obtained from HEM on the psoriasis clinical setting confirmed the results from the literature. Specifically, the predictions showed (Figure 3) that HLA-B*5701-Cw*0602 is significantly the most frequent haplotype in patients ($p = 2.219\text{e-}09$), affecting 20 individuals (20.20%) versus four controls (4%), immediately followed by HLA-B*13-Cw*0602 ($p = 1.521\text{e-}11$), affecting 17 individuals (17.17%) versus two controls (2%). HEM predictions also evidenced that HLA-B*5701-Cw*0602 positive patients developed the first manifestation of the disease already by the age of 17.69 ± 9.19 years versus 35.85 ± 18.26 in negative patients ($p = 3.945\text{e-}06$), confirming the results from the literature.

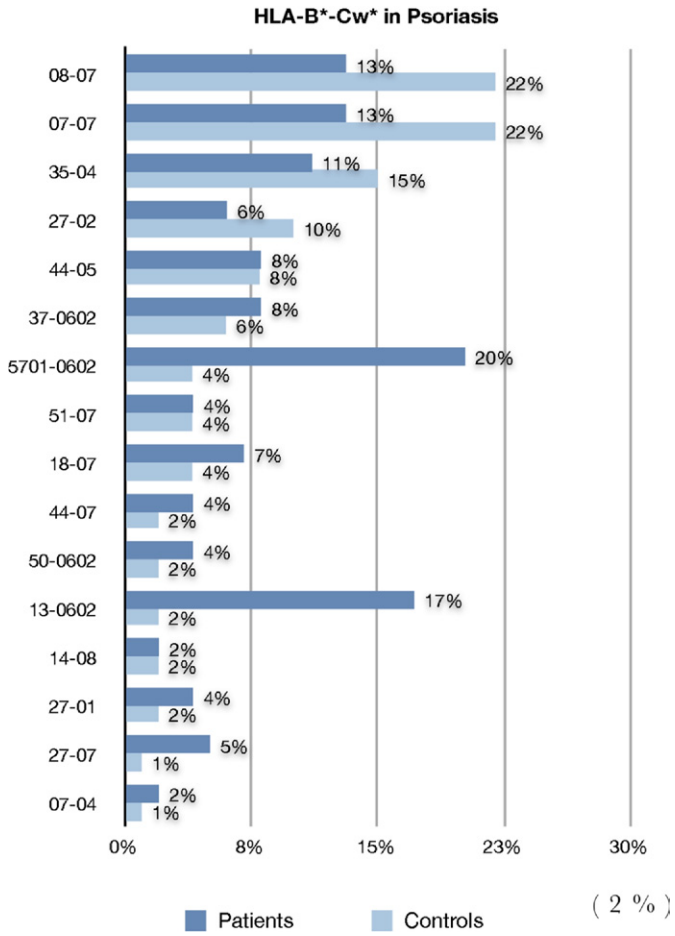

Fig. 3. Predicted HLA-B*-Cw* haplotype frequencies in controls and patients affected by psoriasis.

Similarly, HLA-B*13-Cw*0602–positive patients developed the first manifestation of the disease by the age of 26.82 ± 17.89 ($p = 2.142\text{e-}11$); however there is no statistical evidence to assess the predominance of HLA-B*5701-Cw*0602 respect to HLA-B*13-Cw*0602 for early onset of psoriasis ($p = 0.041\,41$). Finally, no significant variation was found for the remaining haplotypes.

A comparison of the allele frequencies relative to HLA-A*, -B*, -Cw*, -DRB1*, and -DQB1* (Fig. 4) also confirmed the findings of Schmitt-Egenolf et al. [16], assessing the overrepresentation of B*5701, Cw*0602, DRB1* 0701, and DQB1*0303 alleles in psoriatic individuals. Specifically, HEM predictions evidenced the following: HLA-B*5701 was present in 24% of patients versus 4% of controls ($p = 9.12\text{e-}13$); HLA-Cw*0602 was present in 51% of patients versus 15% of controls ($p = 2.2\text{e-}16$); HLA-DRB1*0701 was present in 39% of patients versus 19% of controls ($p = 2.269\text{e-}06$); and HLA-DQB1*0303 was present in 17% of patients versus 5% of controls ($p = 8.208\text{e-}06$). Moreover, similarly to Schmitt-Egenolf et al. [16], the predictions also showed no significant variation in HLA-A*01, present in 46% of patients versus 33% of controls ($p = 0.005\,354$).

The analysis of the extended haplotype HLA-A*-B*-Cw*-DRB1*-DQB1* revealed that HLA-A*01-B*5701-Cw*0602-DRB1* 0701-DQB1*0303 is the most frequent haplotype in patients (Fig. 5), confirming, also in this case, the Schmitt-Egenolf et al. [16] findings. Specifically, the haplotype was found in 7% of patients versus 2% of controls, and the positive patients developed the first manifestation of the disease already by the age of 18 ± 8.89 years versus 39.91 ± 23.27 years in negative patients, showing a high correlation with early onset of psoriasis ($p = 2.2\text{e-}16$).

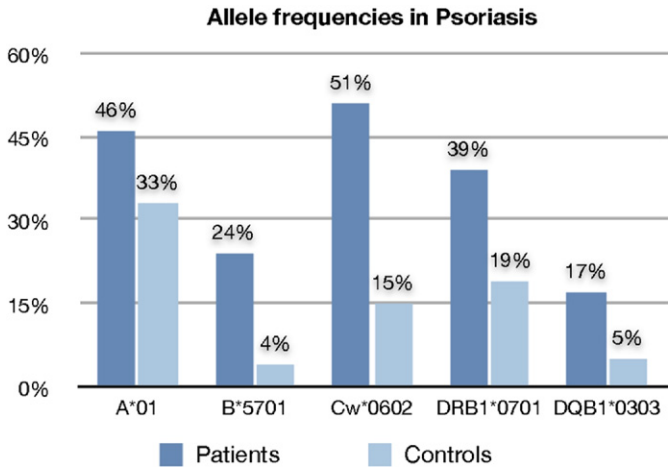

Fig. 4. Predicted HLA-A*01, B*5701, Cw*0602, DRB1* 0701, and DQB1*0303 allele frequencies in controls and patients affected by psoriasis.

Interestingly, we also performed an analysis focused on the HLA-DRB1* and HLA-DQB1* alleles in both patients and controls. The resulting predictions are summarized in Fig. 6 and show that HLA-DRB1*0701-DQB1*0303 is the most frequent haplotype in patients ($p = 0.0001653$), affecting 16 individuals (15.84%) versus 4 controls (4%). No significant variation was found for the remaining haplotypes, with the exception of HLA-DRB1*03-DQB1*02, HLA-DRB1*01-DQB1*05, and HLA-DRB1*13-DQB1*06, whose significant variation ($p = 8.33e\text{-}10$, $p = 0.000\,192\,3$, and $p = 0.001\,396$) can be interpreted as a protection factor from the considered disease.

4. Discussion

The parsimony criterion is one of the possible criteria for haplotype estimation proposed in the literature on HEP. It consists in finding the minimal number of distinct haplotypes necessary to explain a given set of phenotypes [11,12]. In this study, we investigated the use of the parsimony criterion for HLA association studies. Specifically, we developed a simple, compact, and polynomial-sized integer programming model able to provide exact predictions for clinical settings containing up to 200 phenotypes. The model is relatively easy to implement and solvable with any standard solver for mixed-integer programming. The model can be further improved, *e.g.*, by adding strengthening valid inequalities to tackle settings containing large phenotypes, or by, including support for missing or ambiguous data.

The knowledge of the most parsimonious prediction (*i.e.*, the optimal solution from the model) enabled us to validate the parsimony criterion on two clinical settings constituted by patients affected by severe alopecia areata and psoriasis. Specifically, computational experiments showed that the results provided by the

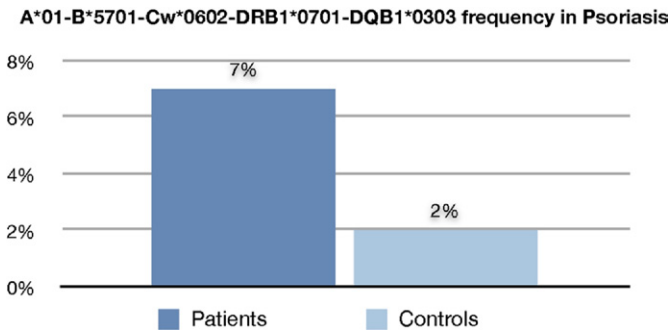

Fig. 5. Predicted HLA-A*01-B*5701-Cw*0602-DRB1* 0701-DQB1*0303 extended haplotype frequencies in controls and patients affected by psoriasis.

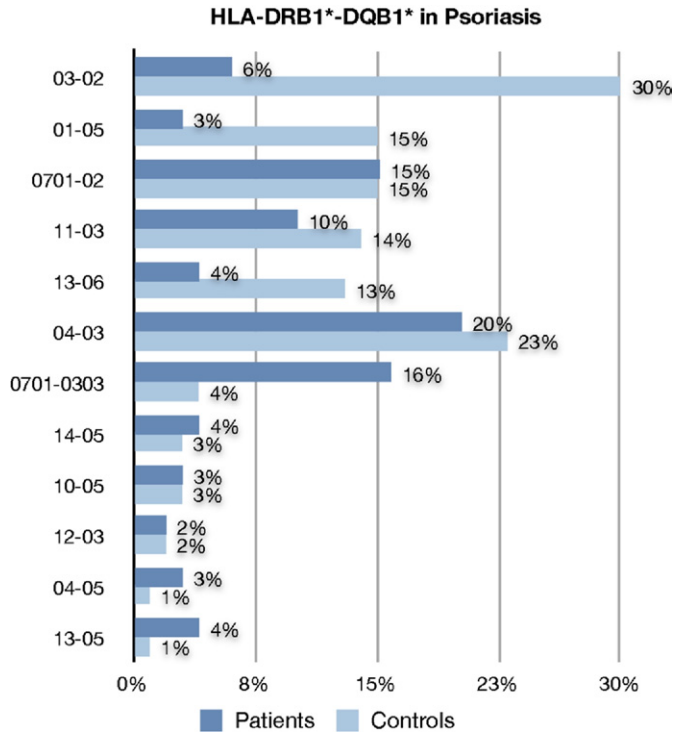

Fig. 6. PredictedDRB1*-DQB1* haplotype frequencies in controls and patients affected by psoriasis.

parsimony criterion are consistent with the experimental haplotype frequencies described in Schmitt-Egenolf et al. [16] and Marques Da Costa et al. [17], showing, for the considered clinical settings at least, a high reliability of the model predictions. However, although these preliminary results are encouraging, it is worth noting that at present it is not clear whether the use of the parsimony criterion could over-amplify the difference among patient and control populations, as it may create strong genetic structures among populations (see, [31,32]). Possibly, only a systematic analysis (out of the scope of the present article) could provide answers to the above questions that deserve for sure future research efforts.

Acknowledgments

The authors acknowledge support from COST Action BM0803 (HLA-NET), the Belgian National Fund for Scientific Research, and the Communauté Française de Belgique Actions de Recherche Concertées. The authors also thank Alicia Sanchez-Mazas and José Nunes for helpful comments and discussions.

References

[1] International HapMap Consortium. The international *HapMap* project. Nature 2003;426:789–96.
[2] International HapMap Consortium. A *second* generation human haplotype map of over. 3.1 Million SNPs. Nature 2007;449:851–61.
[3] Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.
[4] Li WH, Sadler LA. Low nucleotide diversity in *man*. Genetics 1991;129:513–23.
[5] Lancia G, Pinotti MC, Rizzi R. *Haplotyping* populations by pure parsimony: Complexity of exact and approximate algorithms. Informs J Comput 2004;16:348–59.
[6] Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 1995;12:921–7.
[7] Bonizzoni P, Della Vedova G, Dondi R, Jing L. The *haplotyping* problem: A view of computational models and solutions. Int J Comput Sci Technol 2003;18:675–88.
[8] Zhang XS, Wang RS, Wu LY, Chen L. Models and algorithms for *haplotyping* problem. Curr Bioinformatics 2006;1:105–14.
[9] Gusfield D, Orzack SH. Haplotype inference. In: Aluru S (ed), Handbook on Bioinformatics. Boca Raton, CRC Press, 2005, pp 1–28.

[10] Catanzaro D, Labbé M. The pure parsimony *haplotyping* problem: Overview and computational advances. Int Transact Operat Res 2009;16:561–84.
[11] Gusfield D. Inference of haplotypes from samples of diploid populations: Complexity and algorithms. J Comput Biol 2001;8:305–24.
[12] Wang L, Xu Y. Haplotype inference by maximum parsimony. Bioinformatics 2003;19:1773–80.
[13] Clark AG. Inference of haplotypes from PRC-amplified samples of diploid populations. Mol Biol Evol 1990;7:111–22.
[14] Catanzaro D. The minimum evolution problem: Overview and classification. Networks 2009;53:112–25.
[15] Li J, Jiang T. Efficient inference of haplotype from genotype on a pedigree. J Bioinform Comput Biol 2003;1:41–69.
[16] Schmitt-Egenolf M, Eiermann TH, Boehncke WH, Ständer M, Sterry W. Familial juvenile onset psoriasis is associated with the human leukocyte antigen (HLA) class I side of the extended haplotype Cw6-B57-DRB1*0701-DQA1*0201-DQB1*0303: A population and family-based study. J Invest Dermatol 1996; 106:711–4.
[17] Marques Da Costa C, Dupont E, Van der Cruys M, Andrien M, Hidajat M, Song M, Stene JJ. Earlier occurrence of severe alopecia areata in HLA-DRB1*11-positive patients. Dermatol 2006;213:12–4.
[18] Lancia G, Rizzi R. A polynomial case of the parsimony haplotyping problem. Oper Res Lett 2006;34:289–95.
[19] Gusfield D. Haplotype inference by pure parsimony. In: Lecture Note in Computer Science (ed), Annual Symposium in Combinatorial Pattern Matching, volume 2676. Berlin, Germany, Springer-Verlag, 2003, pp 144–55.
[20] Hidajat M, Andrien M, De Canck I, Dupont E. Caucasian belgian normal. In: Gjertson D, Terasaki PI (eds), HLA 1998. American Society for Histocompatibility and Immunogenetics, 1998, pp 119–21.
[21] Catanzaro D, Godi A, Labbé M. A class representative model for pure parsimony *haplotyping*. Informs J Comput 2009, in press.
[22] Colombe BW, Price VH, Khoury EL, Garavoy MR, Lou CD. *HLA* class II antigens associations help to define two types of alopecia areata. J Am Acad Dermatol 1995;33:757–64.
[23] Price VH, Colombe BW. Heritable factors distinguish two types of alopecia areata. Clin Dermatol 1996;14:679–89.
[24] Colombe BW, Lou CD, Price VH. The genetic basis of alopecia areata: *HLA* associations with patchy alopecia areata *versus* alopecia totalis and alopecia universalis. J Invest Dermatol 1999;4:216–9.
[25] Frentz G, Thomsen K, Jakobsen BK, Svejgaard A. *HLA*-DR4 in alopecia areata. J Am Accademy Dermatol 1986;14:129–30.
[26] Welsh EA, Clarck HH, Epstein SZ, Duvic M. Human leukocyte antigen DQB1* *03* alleles are associated with alopecia areata. J Invest Dermatol 1994;103: 758–63.
[27] Christopher E, Henseler T. Patient subgroups and the inflammatory pattern in psoriasis. Acta Dermatol Venereol 1989;69:88–92.
[28] Hanson IM, Poustka A, Trowsdale J. New genes in the class II region of the human major histocompatibility complex. Genomics 1991;10:417–24.
[29] Sargent CA, Dunham I, Campbell RD. Identification of multiple HTF-Island associated genes in the human major histocompatibility complex class III region. EMBO J 1989;8:2305–12.
[30] Trowsdale J, Ragoussis J, Campbell RD. Map of the human *MHC*. Immunol Today 1991;12:443–6.
[31] Dugoujon JM, Hazout S, Loirat F, Mourrieras B, Crouau-Roy B, Sanchez-Mazas A. *GM* haplotype diversity of 82 populations over the world suggests a centrifugal model of human migrations. Am J Phys Anthropol 2004;125:175–92.
[32] Sanchez-Mazas A. An appointment of human *HLA* diversity. Tissue Antigens 2007;69:198–202.