Coll. Antropol. **26** (2002) 1: 77–84
UDC 575.17:616.831-056.7
Original scientific paper

Bayesian Approach to Searching for Susceptibility Genes: Gc2 and EsD1 Alleles and Multiple Sclerosis

M. Di Bacco[1], D. Luiselli[2], M. L. Manca[3] and G. Siciliano[3]

[1] Department of Statistical Sciences, University of Bologna, Bologna, Italy
[2] Department of Experimental Evolutionary Biology, University of Bologna, Bologna, Italy
[3] Department of Neurosciences, University of Pisa, Pisa, Italy

A B S T R A C T

Multiple sclerosis (MS) is one of the most common causes of neurological disability in early adulthood. The current literature is interested in identifying biological or DNA markers associated with genetic susceptibility to MS. The aim of this study is to investigate, by means of Bayesian statistical inference, whether the presence of Gc2 (Gc = group-specific component) and/or EsD1 (EsD = esterase D) alleles affects MS susceptibility. Gc and EsD are two classical genetic markers, being the first a serum protein polymorphism, the latter an isoenzyme polymorphism. The interest of the proposed statistical approach of searching for MS susceptibility genes relies on the analysis of two different functions, one function being inferred from our results on 56 unrelated patients from central Italy affected by MS, the other one from Italian and worldwide epidemiological data. The graphical analysis suggests that MS susceptibility is influenced by both Gc2 and EsD1 alleles; and EsD1 allele is more informative than Gc2. These results point out the advantages of the Bayesian approach in searching for susceptibility genes. Furthermore, the significant association between the considered alleles and the susceptibility to MS suggests possible hypotheses about the pathogenesis of the disease.

Introduction

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system, of unknown aetiology, characterised by T-cell mediated myelin destruction. An etiological role for genetic factors was proposed about 100 years ago by Eichorst who labelled MS as an inherited, transmissible disease[1]. Genetic susceptibility to autoimmune demyelinisation has been postulated and several studies, using both population association and family linkage methods, have

Received for publication July 10, 2001

been carried out to identify and localise genes responsible for this susceptibility[2]. Genetic contribution to MS aetiology has been investigated by means of classical association and linkage methods and by clinical concordance rates between monozygotic and dizygotic twins. These latter studies have shown a 25–30% concordance rate in monozygotic but a much lower rate (2–5%) in dizygotic twins[3] and in non-twin siblings[4]. Furthermore, a higher disease incidence in women that in men (2:1) has been observed.

Investigators have focused on genetic loci recognised to play a major role in the immune response, that is human leukocyte antigen genes[5,6] and T-cell receptor genes[7,8]. Minor areas of study concern immunoglobulin[10,11] and myelin basic protein genes[12]. Furthermore, previous studies, by our group, on serum proteins and enzyme polymorhisms have shown remarkable differences between patients and healthy controls concerning Gc2 (χ^2 = 13.15; p<0.005; d.f.=2) and EsD1 (χ^2 = 9.43; p<0.01; d.f.=2) alleles of the two polymorphic systems Gc (group-specific component) and EsD (esterase D). A relative risk (r.r.) of 3.23 in developing MS was calculated for the Gc2 phenotype and of 4.53 for the EsD1 phenotype[13,14].

The Gc gene, which encodes the vitamin D-binding protein (DBP), maps to chromosome 4q12. Polymorphisms in this gene give rise to three major electrophoretic variants of the DBP serum glycoprotein, which differ by amino acid substitutions as well as attached polysaccharide structures[15]. These variants of DBP are termed Gc1 fast (Gc1f), Gc1 slow (Gc1s), and Gc2, because before the identification of its function as a transporter of vitamin D[16], this plasma protein was known as group-specific component (Gc). The erythrocyte isoenzyme EsD, with unknown function, is codified by two isovalent alleles (EsD1 and EsD2) into a locus of the chromosome 13q14. Although the region on chromosome 13q14 is supposed to be implicated in bipolar affective disorder, schizophrenia[17], obesity[18], no association between genetic markers and MS has been yet reported for this chromosome.

The present study aims to reanalyse our previous results[13,14] to investigate the advantages of a strictly Bayesian approach in searching for susceptibility genes. In particular, it has been tested the selectivity of the Gc2 and EsD1 alleles for the susceptibility to MS. The interest of the proposed statistical approach relies on the analysis of two functions, one function being deducted from unrelated patients from central Italy affected by MS, the other one from Italian and worldwide epidemiological data. In fact, the Bayesian approach does not require the use of a group genetically homogeneous to the investigated sample of MS patients.

Materials and Methods

Materials

Our statistical inference is based on two sets of empirical data. The first one is a random sample from our case history, ethnically and geographically homogeneous, of unrelated 56 patients from the region Abruzzi in Central Italy[13,14], affected by MS diagnosed according to the Poser criteria[19] (Table 1). We detected EsD and Gc phenotypes in these patients by cellulose acetate electrophoresis[13,14]. The second set of data summarizes the

TABLE 1
OUR CASE HISTORY VS PRESENCE AND ABSENCE OF THE GC2 AND ESD1 ALLELES

	Gc2 present	Gc2 absent	Total
EsD1 present	34	21	55
EsD1 absent	1	0	1
Total	35	21	56

TABLE 2
FREQUENCIES OF THE ESD1 AND GC2 ALLELES AND MS INCIDENCE IN DIFFERENT ITALIAN REGIONS AND PROVINCES

Region/Province	EsD1 frequency (τ_i)	Gc2 frequency (γ_i)	MS incidence ($\times$ 100.000 inhabitants)
Trentino	0.855	0.278	38
Emilia	0.861	0.281	39
Ferrara	0.865	0.266	69
Bologna	0.854	0.329	39
L'Aquila	0.850	0.428	53
Palermo	0.845	0.281	43
Sicilia	0.841	0.225	53
Lombardia	0.865	0.262	24
Veneto	0.856	0.247	20
Padova	0.858	0.257	21
Toscana	0.857	0.280	29
Campania	0.854	0.246	72
Liguria	0.862	0.275	29
Puglia	0.891	0.435	31
Basilicata	0.846	0.494	31
Sardegna	0.883	0.319	150
Italia (without Sardegna)	0.883	0.319	150

available Gc2 and EsD1 allele frequencies and MS incidence in 18 Italian regions and provinces (Table 2). The detailed frequency values for the different Italian regions and provinces were reported by Piazza et al.[20] and in reviews of the world distribution of classic markers[21,22]. The MS incidence values for the same populations were drawn from epidemiological studies[23–25].

Object and method of the Bayesian inference

In order to evaluate the possible association between the susceptibility to MS and the presence of: 1) the Gc2 allele regardless of the EsD system; 2) the EsD1 allele regardless of the Gc system; 3) both Gc2 and EsD1 alleles, we have considered the probability distribution $F(\mu,\lambda)$ for the random vector (μ,λ). The components of μ were the conditional probabilities *a, b, r* which measure the chance that a person chosen at random is affected by MS if it is known that he/she has: 1) the Gc2 allele without knowledge of the EsD system, 2) the EsD1 allele without knowledge of the Gc system, 3) both Gc2 and EsD1 alleles, respectively. The components of λ are *s, g, t* which denote the chance that one was a MS patient, one had the Gc2 allele, one had the EsD allele, respectively. The object of our inference was μ.

The problem was solved by $dF(\mu/\text{sample}) = (1/18) * \Sigma_i\, dG(\mu/\lambda_i, \text{sample})$, where $dG(\mu/\lambda_i, \text{sample})$ was obtained by the Bayes' theorem:

$$dG(\mu/\lambda_i, \text{sample}) = \frac{dG(\mu/\lambda_i) * V(\mu/\lambda_i, \text{sample})}{\int dG(\mu) * V(\mu/\lambda_i, \text{sample})}$$

$i = 1,..., 18$; $dG(\mu,\lambda_i)$ was the prior and $V(\lambda_i/\mu, \text{sample})$ the Likelihood. The prior

TABLE 3
FREQUENCIES OF THE ESD1 AND GC2 ALLELES AND MS INCIDENCE IN 26 DIFFERENT COUNTRIES IN THE WORLD

Country	EsD1 frequency (τ_i)	Gc2 frequency (γ_i)	MS incidence (× 100.000 inhabitants)
Denmark	0.923	0.217	33
Finland	0.889	0.255	30
France	0.878	0.282	62
Germany	0.877	0.283	57
Ireland	0.845	0.212	16
Italy	0.858	0.292	58
Holland	0.880	0.264	36
Norway	0.918	0.343	32
Poland	0.929	0.256	21
Sweden	0.867	0.296	51
Swiss	0.883	0.269	61
Great Britain	0.917	0.332	24
Russia	0.855	0.218	8
Israel	0.788	0.172	3
Iraq	0.776	0.257	1
India	0.683	0.157	1
Taiwan	0.642	0.267	2
Japan	0.680	0.292	3
Korea	0.869	0.257	3
Turkey	0.914	0.102	3
South Africa	0.731	0.239	34
Canada	0.866	0.257	7
Texas	0.906	0.260	22
United States	0.616	0.230	5
Chile	0.791	0.139	2
Venezuela	0.901	0.271	74

was estimated by utilising the allele frequencies and relative incidence of MS in different Italian regions and provinces reported in literature. The Likelihood was calculated by considering the sample as the outcome of 56 random choice from the population of our MS patients classified according to Present – Absent with regard to Gc2 and EsD1 (multinomial distribution).

DF(μ, sample) was not analytically explicitable; therefore, it was only possible to do numerical tabulations of the prior and likelihood functions for each of the triplets (s_i, g_i, t_i), i = 1,..., 18. Tabulation of dG(μ, λi, sample) was then still obtained with $s = s_i, g = g_i, t = t_i$.; finally we deduced dF(μ, sample) by averaging such tabulations. By means of three numerical integration of dF(μ, sample) with respect to b,

r or *a*, *r*, or *a*, *b* we were able to have an accurate graphical representation of the marginal probability distributions for random numbers *a*, *b*, *r*, respectively.

Automatic procedure and sensitivity analysis

The numerical problem had been graphically solved by means of a dedicated computerised program in MATLAB environment (Mathworks, Inc.).

Taking into account that our statistical inference is based on two sets of heterogeneous data we have investigated the sensitivity of the results to the choice of the prior by using the same automatic procedure. In particular, we have used Gc2 andEsD1 allele frequencies[21,22] and epidemiological data[24,25] from 26 worldwide countries (Table 3).

Results

Figure 1 shows the probability distributions dF_a (. /), dF_b (. /), dF_r (. /) obtained by the marginalization of dF (μ / sample). They represent the degree of confidence on the values of probability of MS susceptibility if 1) allele Gc2 is present; 2) allele EsD1 is present; 3) both alleles are present, respectively.

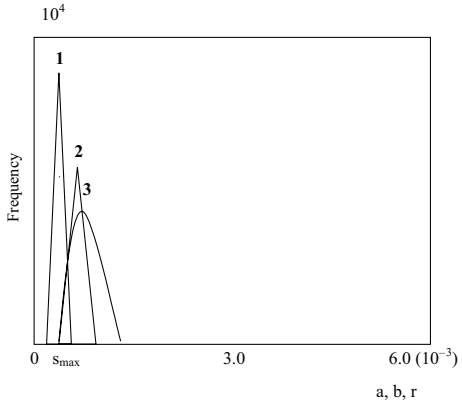

Fig. 1. Italian regions and provinces (n=18). 1 = FDP dF_a (. /); 2 = FDP dF_b (. /); 3 = FDP dF_r (. /)

The graphical analysis suggests that the susceptibility to MS is influenced by both the Gc2 and EsD1 alleles. This is evident because the maximum probability s_{max} (Figure 1) is left-shifted respect to the modal points of the distribution. Furthermore, we observe that EsD1 allele is more informative than Gc2 allele and that co-presence of both alleles is even more informative.

The usefulness of the Bayesian approach is confirmed by sensitivity analysis carried out on allele frequencies and epidemiological data from world wide distributed populations. In fact obtained results show a fully comparable pattern (Figures 1 and 2).

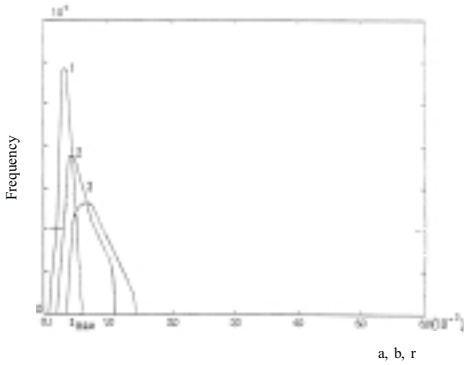

Fig. 2. World countries (n=26). 1 = FDP dF_a (. /); 2 = FDP dF_b (. /); 3 = FDP dF_r (. /)

Discussion

Recently, there is an increasing interest in the application of the Bayesian statistics in many genetic areas[26,27] such as linkage mapping[28]. In fact, Bayesian analysis is simple to interpret. Each quantity of interest is described by a posterior distribution which summarises the researcher knowledge in light of the empirical evidence, and these hypotheses of interest are evaluated by interpreting appro-

priate features of the posterior distribution.

In the present case, the quantities of interest are the probabilities $\mu = (a, b, r)$, which were chosen to evaluate the association between MS and Gc2 and EsD1 alleles. We built the joint posterior distribution of a, b, r using the empirical data reported in Tables 1 and 2. The hypothesis about the association between MS and the alleles was then evaluated by graphical examination of some characteristics of the posterior distribution. The computational problem was solved with appropriate computer implementation.

More specifically, our procedure involved the following steps:

a) to assign the prior distribution connected with the vector $\mu = (a, b, r)$, which was modelled by using data on EsD and Gc allele frequencies and MS incidence rates;

b) to built up the likelihood function utilising a classic multinomial model;

c) to calculate the marginal probability distribution function for a, b, r, by using the posterior probability distribution functions obtained by the Bayes' theorem.

Results support our hypothesis that in presence of Gc2, or EsD1, or Gc2 and EsD1, the probability of susceptibility to MS is higher than that evaluated in the absence of such information about the respective presences. The maximum probability of MS susceptibility, assessed with the population data in Table 2, is left-shifted with respect to the posterior distributions relative to a, b, r.

Our results, graphically represented in Figure 1, also demonstrate that the alleles together have a greater influence on the susceptibility to MS than each allele considered separately. In fact, the mode of the posterior distribution for Gc2 and EsD1 together was to the left of those of the posterior distribution for Gc2 and EsD1. Furthermore, if we consider the two respective modes, the EsD1 allele showed a greater influence than the Gc2 allele. However, the fact that Gc and EsD data together better predict Ms does not necessarily indicate that there is an interactive effect of these genes on MS.

It is not easy to clarify the causes of the possible associations between 1) Gc2 and MS, and 2) EsD1 and MS, suggested by our Bayesian analysis. An interesting hypothesis could be related to the study by Petrini and co-workers[29], which revealed the presence of Gc in membrane receptors of T lymphocytes. More precisely, this research localised Gc in the receptor complex for the Fc portion of the G immunoglobulin. These data could be significant in view of the pathogenetic role of T lymphocytes in MS[30], as indicated by their presence in acute myelin lesions in humans[31], or in the most accredited animal model of MS, experimental allergic encephalomyelitis. Another hypothesis about the possible role of Gc2 in MS susceptibility is based on the fact that the Gc system is a protein vector of vitamin D, a factor able to hinder the onset of EAE[32]. An altered metabolism of vitamin D could interfere with the activation of T lymphocytes[33], which represent a direct target for calcitriol[34].

It is even more difficult to interpret the possible link between EsD1 and MS, since the function of EsD is still unknown. However, EsD has been identified with S-formylglutathione hydrolase[35,36], and therefore is possibly involved in the function of the cell antioxidant system. Hence, our results could be in agreement with the hypothesis suggested by Calabrese and co-workers[37], that a compromise of the antioxidant mechanisms is involved in the pathogenesis of MS. This pathogenetic interpretation could be supported by epidemiological data, such as: 1) the increased incidence of MS in populations that consume high proportion of animal fat[38], 2) the decreased level of

linoleic acid[39], and 3) the increased concentration of malondialdehyde in the blood of MS patients[40].

In conclusion, the obtained results indicate that the Bayesian approach in searching for susceptibility genes has revealed useful. The significant association between the Gc2 and EsD1 alleles and the susceptibility to MS suggests possible hypotheses about the pathogenesis of the disease. Our results could confirm the hypothesis that MS is a polygenic and multifactorial disease, taking into account the possible role of vitamine D and oxidative stress in its phenotypic expression.

REFERENCES

1. EICHORST, H., Virchows Arch., 146 (1986) 173. — 2. BELL, J. I., G. M. LATHROP, Nat. Genet., 13 (1996) 377. — 3. MUMFORD, C. J., N. W. WOOD, H. E. KELLAR-WOOD, J. THORPE, J. MILLER, D. A. S COMPSTON, Neurology, 44 (1994) 11. — 4. EBERS, G. C., D. E. BULMAN, A. D. SADOVNICK, N. Engl. J. Med., 315 (1986) 1638. — 5. HILLERT, J., O. OLERUP, Neurology, 43 (1993) 163. — 6. SAWCER, S., H. B. JONES, R. FEAKES, J. GRAY, N. SMALDON, J. CHATAWAY, N. ROBERTSON, D. CLAYTON, P. N. GOODFELLOW, A. COMPSTON, Nat. Genet., 13 (1996) 464. — 7. COMPSTON, D. A. S., H. KELLAR-WOOD, N. ROBERTSON, S. SAWCER, N. W. WOOD, Acta Neurol. Scand., 161 (1995) 43. — 8. MARTINEZ-NAVES, E., M. VICTORIA-GUTIERREZ, C. LOPEZ-LARREA, J. Neuroimmunol., 47 (1993) 9. — 9. EOLI, M., N. W. WOOD, H. F. KELLAR-WOOD, J. Neurol. Sci., 124 (1994) 32. — 10. HILLERT, J., J. Neuroimmunol., 43 (1993) 9. — 11. WALTER, M. A., W. T. GIBSON, G. C. EBERS, D. W. COX, J. Clin. Invest., 87 (1991) 1266. — 12. WOOD, N. W., P. HOLMANS, D. CLAYTON, N. P. ROBERTSON, D. A. S. COMPSTON, J. Neurol. Neurosurg. Psych., 57 (1994) 1191. — 13. LUISELLI, D., S. MARCIATORI, Rivista di Antropol., 74 (1997) 157. — 14. TARABORELLI, T., F. D'ANDREA, D. LUISELLI, A. MANCINI, G. GRUPPIONI, M. PRENCIPE, Rivista di Antropol., 74 (1997) 167. — 15. BRAUN, A., A. KOFLER, S. MORAWEITZ, H. CLEVE, Biochim. Biophys. Acta, 1216 (1993) 385. — 16. DAIGER, S. P., M. S. SCHANFIELD, L. L. CAVALLI-SFORZA, Proc. Natl. Acad. Sci. USA, 72 (1975) 2076. — 17. BADENHOP, R. F., M. J. MOSES, A. SCIMONE, P. B. MITCHELL, K. R. EWEN, A. ROSSO, J. A. DONALD, L. J. ADAMS, P. R. SCHOFIELD, Mol. Psychiatry, 6 (2001) 396. — 18. FEITOSA, M. F., I. B. BORECKI, S. S. RICH, D. K. ARNETT, P. SHOLINSKY, R. H. MYERS, M. LEPPERT, M. A. PROVINCE, Am. J. Hum. Genet., 70 (2002) 72. — 19. POSER, C. M. D. W. PATY, L. SCHEINBERG, W. I. MC DONALD, F. A. DAVIS, G. C. EBERS, K. P. JOHNSON, D. H. SIBLEJ, W. W. TOURTELLETTE, Ann. Neurol., 13 (1983) 227. — 20. PIAZZA, A., E. OLIVETTI, O. CARBONARA, M. BARGAGNA, F. PECORI, P. BENCIOLINI, P. CORTIVO, F. BREDA, R. DOMENICI, JAYAKAR S.: La distribuzione di alcuni polimorfismi genetici in Italia. (Il Ponte, Milano, 1982). — 21. TILLS, D., KOPEC A. C.: The distribution of the human blood groups and other polymorphism. (Oxford University Press, Oxford, 1983). — 22. ROYCHOUDHURY, ARUN K., MASATOSHI N.: Human polymorphic gene: World distribution. (Oxford University Press, Oxford, 1988). — 23. GRANIERI, E., I. CASETTA, M. R. TOLA, Neurology, 49 Suppl. 2 (1997) 33. — 24. KURTZKE, J. F.: Epidemiology of Multiple Sclerosis: Handbook of clinical neurology, demyelinating diseases. (Elsevier Sciences Publisher, 1985) — 25. SWINGLER, R. J., D. A. S. COMPSTON, J. Neurol. Neurosurg. Psych., 51 (1988) 1520. — 26. SHOEMAKER, J. S., I. S. PAINTER, B. S. WEIR, Trends Genet., 15 (1999) 354. — 27. DI BACCO, M., Proc. World Meet. ISBA. (1994). — 28. OTT, J.: Analysis of human linkage. (John Hopkins University Press, 1991). — 29. PETRINI, M., D. L. EMERSON, R. M. GALBRAITH, Nature, 306 (1983). — 30. STEINMAN, L, Springer Semin. Immunopathol., 14 (1992) 79. — 31. LEE, S. C., Neuropathol. Appl. Neurobiol., 17 (1990) 265. — 32. LEMIRE, J. M., D. C. ARCHER, J. Clin. Invest., 87 (1991) 1103. — 33. TOKUDA, N., N. MIZUKI, M. KASAHARA, R. B. LEVY, Immunol., 75 (1992) 349. — 34. VANHAM, G., J. L. CEUPPENS, R. BOUILLON, Cell Immunol., 124 (1989) 320. — 35. APESHIOTIS, F., BENDER, K., Hum. Genet., 74 (1986) 176. — 36. EIBERG, H., J. MOHR, Hum. Genet., 74 (1986) 174. — 37. CALABRESE, V., R. RAFFAELE, E. COSENTINO, V. RIZZA, Int. J. Clin. Pharmacol. Res., 14 (1994) 119. — 38. SWANL, R. L., Am. J. Med. Sci., 220 (1985) 421. — 39. THOMPSON, R. H. S.: Unsatured fatty acids in multiple sclerosis. In: DAVISON, A. N. (Ed.): Multiple sclerosis research. (London, 1975). — 40. ROGOVINA, N., A. P. KOKLOV, Zh. Nevorpotal. Paikhiatr. Jm., 80 (1980) 696.

D. Luiselli

Department of Experimental Evolutionary Biology, Unit of Anthropology, University of Bologna, Via Selmi, 3, 40126 Bologna, Italy

BAYESOV PRISTUP POTRAZI ZA PREDISPOZICIJSKIM GENIMA: GC1 I ESD1 ALELI I MULTIPLA SKLEROZA

S A Ž E T A K

Multipla skleroza (MS) jedan je od najčešćih uzroka neurološkog invaliditeta u ranoj odrasloj dobi. Današnja istraživanja usmjerena su identifikaciji ili bioloških ili DNK markera koji su povezani s genetskom predispozicijom za MS. Cilj je ovog istraživanja ustanoviti putem Bayesove statistike je li prisustvo alela Gc2 (Gc = komponenta specifična za skupinu) i/ili EsD1 (EsD = D esteraza) predisponira razvoj MS-a. Gc i EsD su dva klasična genska markera, prvi ukazuje na polimorfizam serumskog proteina a drugi na polimorfizam izoenzima. U potrazi za genima koji predisponiraju razvoj MS-a predloženi statistički pristup zasniva se na analizi dvije različite funkcije. Prva funkcija proizišla je iz naših prethodnih rezultata dobivenih na 56 nesrodnih bolesnika iz središnje Italije oboljelih od MS, dok druga proizlazi iz epidemioloških podataka koji se odnose na Italiju te drugih populacija širom svijeta. Grafička analiza sugerira da je predispozicija za MS pod utjecajem i Gc2 i EsD1 alela, pri čemu je EsD1 alel informativniji od Gc2. Ovi rezultati ukazuju na prednost Bayesovog pristupa u potrazi za predispozicijskim genima. Štoviše, značajna povezanost između razmatranih alela i MS-a sugerira moguće hipoteze o patogenezi ove bolesti.