The P28T Mutation in the *GALK1* Gene Accounts for Galactokinase Deficiency in Roma (Gypsy) Patients across Europe

MICHAEL HUNTER, EVELYNE HEYER, FREDERIC AUSTERLITZ, DORA ANGELICHEVA,
VANIA NEDKOVA, PAZ BRIONES, ANNA GATA, ROSARIO DE PABLO, ARANKA LÁSZLÓ,
NILS BOSSHARD, RICHARD GITZELMANN, ATTILA TORDAI, LAJOS KALMAR,
CSABA SZALAI, ISTVAN BALOGH, CONSTANTIN LUPU, AXINIA CORCHES, GABRIELA POPA,
ANNA PEREZ-LEZAUN, AND LUBA V. KALAYDJIEVA

Centre for Human Genetics, Edith Cowan University, Perth, Australia [M.H., D.A., L.V.K.]; Western Australian Institute for Medical Research, Perth, Australia [L.V.K.]; CNRS-Laboratoire d'Anthropologie Biologique, Musée de l'Homme, Paris, France [E.H.]; INRA-Laboratoire de génétique et amélioration des arbres forestiers, Domaine de l'Hermitage, Pierroton-Cestas, France [F.A.]; Department of Pediatrics, Medical School, Pleven, Bulgaria [V.N.]; Institut de Bioquímica Clínica, Barcelona, Spain [P.B., A.G.]; Sección de Immunología, Clínica Puerta de Hierro, Madrid, Spain [R.d.P.]; Department of Pediatrics, Albert Szent-Gyorgyi Medical Centre, Szeged University, Szeged, Hungary [A.L.]; Division of Metabolism and Molecular Pediatrics, University Children's Hospital, Zurich, Switzerland [N.B., R.G.]; Molecular Biology Laboratory, National Institute of Hematology, Budapest, Hungary [A.T., L.K.]; Heim Pal Children's Hospital, Budapest, Hungary [C.S.]; Institute of Laboratory Medicine and Molecular Pathology, Debrecen University, Debrecen, Hungary [I.B.]; Facultat de Ciencies de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona, Spain [A.P.-L.]; Clinical Center of Neuropsychiatry for Children and Adolescents, Timisoara, Romania [C.L., A.C., G.P.]

ABSTRACT

Galactokinase deficiency is an inborn error of metabolism that, if untreated, results in the development of cataracts in the first weeks of life. The disorder is rare worldwide, but has a high incidence among the Roma (Gypsies). In 1999, we reported the founder Romani mutation, P28T, identified in affected families from Bulgaria. Subsequent studies have detected the same mutation in Romani patients from different European countries. The screening of 803 unrelated control individuals of Romani ethnicity from Bulgaria, Hungary, and Spain has shown an overall carrier rate of 1:47 and an expected incidence of affected births about 1:10,000. Using disease haplotype analysis, the age of the P28T mutation was estimated at 750 y, preceding the splits of the proto-Roma into the numerous populations resident in Europe today. The findings suggest that the mutation has spread with the early diaspora of the Roma throughout Europe. Superimposed on this old distribution pattern is the new migration wave of the last decade, with large numbers of Roma moving to Western Europe as a result of the economic changes in the East and the wars in former Yugoslavia. The changing demographic pattern of Romani minorities can be expected to lead to a homogenization of the incidence of "private" Romani disorders and founder mutations. The P28T mutation is thus likely to account for a high proportion of galactokinase deficiency cases across Europe. Mutation-based pilot newborn screening programs would provide current incidence figures and help to design long-term prevention of infantile cataracts due to galactokinase deficiency. (**Pediatr Res 51: 602–606, 2002**)

Abbreviations:
GALK1, galactokinase gene 1

Galactokinase (EC 2.7.1.6) catalyzes the first step of galactose phosphorylation in the Leloir pathway of intermediate metabolism. Galactokinase deficiency (MIM 230200), one of the three autosomal-recessive inborn errors of galactose metabolism leading to hypergalactosemia, is clinically characterized by cataract formation during the first weeks to months of life in homozygous affected children (1, 2). Heterozygosity for this trait has been implicated as a factor predisposing to the development of presenile cataracts (3, 4).

Galactokinase deficiency was first described in a Roma (Gypsy) patient in Switzerland (1, 2) and a number of subsequent reports, based on clinical observations as well as on newborn screening data (5–8), have suggested that the disorder is common in that ethnic group.

In 1999, we reported the identification of a missense mutation in the *GALK1* gene, P28T, that was shared by six Romani families from Bulgaria, originating from three socially distinct but linguistically and genetically related groups of Vlax Roma, whose historical migrations can be traced back to the northern part of the Balkans (9). Since that initial description, we have detected the P28T mutation in affected individuals from different countries, suggesting that it is the major founder mutation responsible for galactokinase deficiency in Romani patients throughout Europe.

In this report, we present evidence of the common origin of the mutation, an estimate of its age, and data on its frequency in different Romani populations.

SUBJECTS AND METHODS

Subjects

Mutation analysis was conducted in eight galactokinase-deficient subjects, including seven Roma: one from Bulgaria, one from Romania, three from Hungary, and two from Spain (belonging to the same consanguineous family, Fig. 1), as well as one patient of declared Turkish ethnicity, resident in Switzerland.

Haplotype analysis was performed in the affected families from Hungary, Spain, and Switzerland and compared with the original findings in families from Bulgaria (9).

Population screening for the P28T mutation was conducted among 803 unrelated control individuals of Romani ethnicity, including 332 from Bulgaria, 310 from Hungary, and 161 from Spain. Romani group identity was known for the Bulgarian Roma, among whom 263 individuals were Vlax Roma (10, 11), and 69 originated from non-Vlax groups speaking different Balkan dialects of the Romani language. The participating individuals from Hungary and Spain were classified on the basis of area of residence as follows: northeastern Hungary 159, southern Hungary 95, Budapest region 56, Madrid area 77, Zaragoza 46, and Barcelona 38.

Informed consent has been obtained from all individuals included in the investigations. The study complies with the ethics guidelines of the institutions involved.
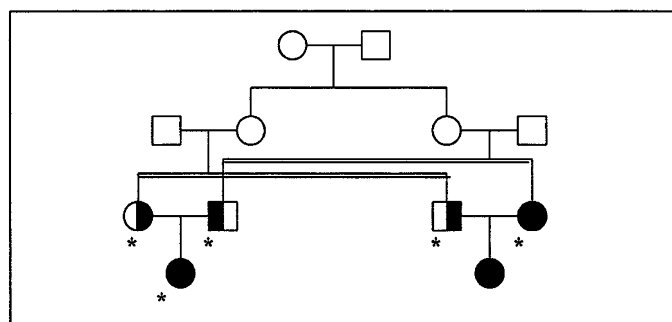
Detection of the P28T Mutation

Screening for the P28T mutation was performed using a PCR-based *Ava*I restriction assay. A 400-bp DNA fragment spanning the region between nucleotides 418 and 818 in the *GALK1* sequence (GenBank accession number L76927) was amplified from genomic DNA and Fitzco-Whatman-FTA cards (Whatman Bioscience, Maidstone, Kent, U.K.) by use of primers 5'-CCCGAGCATCCCGCGCCGAC-3' and 5'-GACAG-GCTGTTCCCCACGT-3'. PCR volumes were made up to 20 μL and consisted of 20 ng of DNA for genomic samples, or a ~1 mm^2 section of FTA card, 1.25 mM of MgCl$_2$, 2 mM of each deoxyribonucleoside triphosphate, 1 $\times$ PCR buffer, 4% DMSO, 0.5 units of *Taq* polymerase and distilled H$_2$O. PCR conditions included initial denaturation at 94°C for 5 min, 35 cycles of 94°C for 30 s, 63°C for 30 s, and 72°C for 40 s, and a final extension step at 72°C for 7 min. PCR products were ethanol-precipitated, resuspended in 5 μL of distilled H$_2$O, and digested with 3 units of *Ava*I, according to manufacturer's recommendations. Products were electrophoresed in 4% agarose gels and visualized by ethidium-bromide staining.

Haplotype Analysis

Genotyping was performed for six microsatellite markers spanning a 9-cM region surrounding the *GALK1* gene. The markers included D17S1797, D17S1602, D17S1864, D17S929, D17S1839, and D17S1817 (Fig. 2). PCR protocols and haplotype construction were performed as previously described (9).

Dating of the P28T Mutation

Standard Luria-Delbruck estimation. The age of the mutation was estimated using the formula P = $(1 - \theta)^g \approx e^{-\theta g}$, which implies g = $-\log(P)/\theta$, where θ is the length of haplotype and P is the proportion of the most frequent haplotype for this size. This value of g (g_{app}) has to be corrected according to the model of Luria and Delbruck (12). The basic principle is that when we evaluate g using the formula above, we omit a number g_0 of generations during which one recombinant is

Fig. 1. Pedigree of consanguineous Spanish Gypsy family from Barcelona. Asterisks indicate individuals available for mutation and haplotype analysis. Despite consanguinity, three different haplotypes were observed in this family (see Fig. 2).
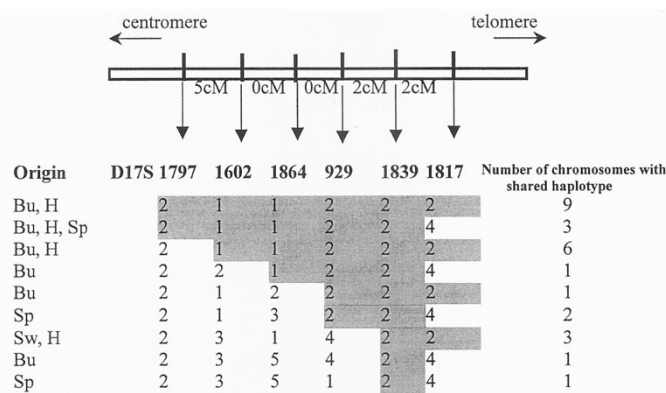

Fig. 2. Haplotypes in the 17q24 region of disease chromosomes in patients from Bulgaria (*Bu*), Hungary (*H*), Spain (*Sp*), and Switzerland (*Sw*). Allele 2 of marker D17S1839 is conserved in all chromosomes carrying the P28T founder mutation. Shaded regions indicate conserved haplotype.

Origin	D17S 1797	1602	1864	929	1839	1817	Number of chromosomes with shared haplotype
Bu, H	2	1	1	2	2	2	9
Bu, H, Sp	2	1	1	2	2	4	3
Bu, H	2	1	1	2	2	2	6
Bu	2	2	1	2	2	4	1
Bu	2	1	2	2	2	2	1
Sp	2	1	3	2	2	4	2
Sw, H	2	3	1	4	2	2	3
Bu	2	3	5	4	2	4	1
Sp	2	3	5	1	2	4	1

expected. This number of generations is the time necessary for one recombination to occur. For a given θ, the number of meioses (M) such that only one recombination is expected is $1/\theta$. When the population is growing at a rate d, the number of meioses is

$$M = \sum N_i = \sum e^{di} = (e^{dg} - 1)e^d/(e^d - 1). \quad (1)$$

Labuda *et al.* (12) made a simplification: $e^{dg} - 1 \approx e^{dg}$. This simplification is accurate in the case of a fast-growing population, like the one they studied. To be more general, we do not make this simplification. Then

$$g_0 = -1 + \frac{\log[r + (r - 1/\theta)]}{\log(r)}, \quad (2)$$

with $r = e^d$. Thus, the time since the introduction of the mutation in the population becomes $g = g_0 + g_{app}$

Multipoint Luria-Delbruck estimation. We designed a new method, which we refer to as multipoint Luria-Delbruck estimation, that takes into account the whole information on haplotypes, not only the longest one. The basic principle is to calculate the likelihood of the whole set of data, taking into account the Luria-Delbruck correction. We then estimate g by maximum likelihood, and provide the confidence interval by the standard Max-2 rule.

RESULTS

P28T analysis in galactokinase-deficient subjects and controls. All eight affected individuals, referred for analysis from different European countries, were homozygous for the C→A transversion in exon 1 of the *GALK1* gene, resulting in the threonine for proline substitution at amino acid position 28 of the galactokinase protein.

Screening for the P28T mutation (Table 1) among 803 control Roma individuals from Bulgaria, Hungary, and Spain identified 17 carriers, giving an overall carrier rate of 1:47 and an expected incidence of affected births about 1:10,000.

The distribution of the mutation was nonrandom: 14/17 carriers belonged to the Bulgarian sample and, among those, 12 individuals were Vlax Roma and 2 originated from non-Vlax groups; 3/17 were Spanish Gypsies; no carriers were identified in the Hungarian screening sample. Carrier rates can therefore

be estimated at about 1:22 among the Vlax Roma and 1:35 in non-Vlax groups in Bulgaria, and 1:54 in Spain.

Haplotype analysis and age of the P28T mutation. Polymorphic haplotypes on disease chromosomes were studied in the affected families from Spain, Hungary, and Switzerland. The data were compared with those obtained in the original study of galactokinase-deficient Romani families from Bulgaria (9). The analysis revealed complete homozygosity for allele 2 of marker D17S1839 (Fig. 2), previously shown to be the conserved allele in strong linkage disequilibrium with the P28T mutation (9). Five disease chromosomes (four Hungarian and one Spanish) carried haplotypes that were closely related to the full ancestral haplotype found in the Bulgarian families (Fig. 2) (9). The remaining haplotypes provided evidence of different historical recombinations that have occurred in the divergent Romani groups. In general, haplotype diversity in the new sample was greater than that observed among disease chromosomes in Bulgaria.

Estimates of the age of the P28T mutation, based on the diversity of polymorphic haplotypes observed in the Bulgarian families and in the families sampled in other European countries, are shown in Table 2. The younger age calculated for the Bulgarian Roma is the result of the recent splits of the proto-Romani population, where diversity within each group is less than that in the entire population. The dating method incorporates the growth rate of the population, and the values obtained for the European Roma range between 18 generations (the lower confidence limit for a fast-growing population) and 54 generations (the upper confidence limit for a slow-growing population). If we consider 17,000 the population size in the 14th century [a reasonable approximation given the available information on the historical demography of the Roma (10)] and the current population of 8 million Romani in Europe (13), the overall growth rate, for a generation time of 25 y, is 1.32. The estimated time of the origin of the P28T mutation is then around 30 generations or 750 y ago.

DISCUSSION

The P28T mutation was identified in a study of Romani families with galactokinase deficiency from Bulgaria (9). The results of the present study suggest that it is shared by galactokinase-deficient patients of Romani descent throughout Europe. A common origin of the mutation and its inheritance, identical by descent, is strongly suggested by the closely

Table 1. *Population screening for P28T and carrier frequency*

Romani population	No.	Carriers	Carrier (%)
Bulgaria	332	14	4.2
Vlax	263	12	4.6
Non-Vlax	69	2	2.9
Hungary	310	0	0
Northeast	159	0	0
Central (Budapest)	56	0	0
South (Pécs)	95	0	0
Spain	161	3	1.9
Madrid	77	1	1.3
Zaragoza	46	1	2.2
Barcelona	38	1	2.6
Total	803	17	2.1

Table 2. *Multipoint estimates of the age of the P28T mutation in Europe*

r	European Roma	Bulgarian Roma
1.1	37.54 (27.14, 53.62)	23.67 (19.81, 30.50)
1.15	34.39 (23.91, 50.51)	20.64 (16.52, 27.65)
1.2	32.46 (21.93, 48.60)	18.84 (14.55, 25.95)
1.25	31.14 (20.58, 47.28)	17.64 (13.23, 24.82)
1.3	30.16 (19.59, 46.32)	16.77 (12.29, 23.99)
1.35	29.42 (18.82, 45.57)	16.12 (11.57, 23.36)
1.4	28.82 (18.22, 44.98)	15.60 (11.01, 22.86)

r, Demographic growth rate.

Mutation age estimates are in generations, with the confidence interval in brackets.

related polymorphic haplotypes carried by all disease chromosomes and by the complete homozygosity for a single conserved allele of marker D17S839, which was shown in our previous study to be at a small physical distance from the *GALK1* gene (9).

The limited number of affected families where haplotype analysis has been performed, and the great variety observed in these families, means that we have sampled only a fraction of the haplotype diversity that may exist in Romani populations across Europe. The age of the P28T mutation, based on the current observations, is therefore likely to be an underestimate. Nonetheless, even the calculated 750 y since the origin of P28T are sufficient to explain its wide geographical distribution: comparative linguistic studies of Romani dialects suggest that the major splits occurred around the 13th century, after the arrival of the proto-Roma in the Balkans (14). An older age, compatible with our present data, could mean that the mutation originated in the ancestral population before the exodus from India, around 1000 y ago.

An uneven distribution of the mutation frequency among different Romani groups, and thus the existence of groups at particularly high risk, has been suggested by our previous observations (9) and is supported by the present results. Within Bulgaria, carrier rates among the Vlax Roma are twice those among other groups. In Hungary, where the occurrence of P28T is clearly documented by the analysis of affected families, the screening of 310 Gypsy Romani control individuals failed to detect a single carrier, suggesting that a high-risk Romani group may not have been included in the screening sample. The reasons for such nonrandom distribution of the galactokinase deficiency, as well as other mutations (15–18) among the Roma, are not clear. The early historical descriptions of small Gypsy groups traveling in different parts of Europe (13) suggest that the differences observed today could easily result from a founder effect/genetic drift scenario, leading to high frequencies in some groups and loss of the mutant allele in others. In Spain, whose Romani population is considered genetically homogeneous (de Pablo R, unpublished data), the average carrier rate is 1:54.

Galactokinase deficiency is a rare disorder, with a reported incidence of affected births around 1 per million in America, Japan, and western Europe (19; Gitzelmann R, Steinmann B, unpublished data). Predictably, out of 21 mutations in the *GALK1* gene identified in these populations (20–23), 20 are private defects confined to individual families, and one, Q382X, is common to affected families of Costa Rican descent (22).

A high incidence of the disorder among the Roma has been suggested by a number of reports (5–8), as well as by the gradient in the incidence of galactokinase deficiency in Europe, from the low figures reported in the West, to 1:150,000 in central Europe, and 1:52,000 in Bulgaria (9, 24), coincident with the distribution of Romani minorities in the different parts of Europe (13, 14). However, the traditional demographic pattern has undergone significant changes during the last decade, as a result of the political and economic changes in eastern and central Europe and the wars in former Yugoslavia. The shift in the demographic profile, with significant numbers

of Romani refugees moving to the West (25, 26), can be expected to result in an equalization of the frequencies of "private" Romani disorders between different countries.

Our studies have shown that galactokinase deficiency among the Roma is genetically homogeneous and, similar to other mendelian disorders (15–18), is caused by a single founder mutation. The mutation, P28T, is found in patients with galactokinase deficiency along the historical migration routes of the Roma, from the Balkans to the Atlantic, and our estimates of the age of the mutation are compatible with its occurrence among socially divergent and geographically dispersed Romani groups. Superimposed on this historical spread will be the migration wave of the last decade. P28T is thus likely to account for a very high proportion of cases of galactokinase deficiency across Europe. The low cost, high specificity, and relatively high sensitivity of the P28T detection assay appear to justify its inclusion in pilot screening programs, to determine the mutation frequency among the newborn populations of different European countries and design the longer-term public health strategy for prevention.

REFERENCES

1. Gitzelmann R 1965 Deficiency of erythrocyte galactokinase in a patient with galactose diabetes. Lancet 2:670–671
2. Gitzelmann R 1967 Hereditary galactokinase deficiency, a newly recognized cause of juvenile cataracts. Pediatr Res 1:14–23
3. Stambolian D, Scarpino-Myers V, Eagle RC, Hodes B, Harris H 1986 Cataracts in patients heterozygous for galactokinase deficiency. Invest Ophthalmol Vis Sci 27:429–433
4. Elman MJ, Miller MT, Matalon R 1986 Galactokinase activity in patients with idiopathic cataracts. Ophthalmology 93:210–215
5. Thalhammer O, Gitzelmann R, Pantlitschko M, 1968 Hypergalactosemia and galactosuria due to galactokinase deficiency in a newborn. Pediatrics 42:441–445
6. Bolgiani MP, Gallenca M, Barocelli PC 1984 Su un caso di galattosemia da deficit di galattochinasi. Pediatr Med Chir 6:333–336
7. Gitzelmann R 1987 Hereditary galactokinase deficiency: this week's citation classic. Curr Contents Clin Med 30:14
8. Kalaydjieva L, Kremensky I 1992 Screening for phenylketonuria in a totalitarian state. J Med Genet 29:656–658
9. Kalaydjieva L, Perez-Lezaun A, Angelicheva D, Onengut S, Dye D, Bosshard NU, Jordanova A, Savov A, Yanakiev P, Kremensky I, Radeva B, Hallmayer J, Markov A, Nedkova V, Tournev I, Aneva L, Gitzelmann R 1999 A founder mutation in the GALK1 gene is responsible for galactokinase deficiency in Roma (Gypsies). Am J Hum Genet 65:1299–1307
10. Marushiakova E, Popov V 1997 Gypsies (Roma) in Bulgaria. Peter Lang, Frankfurt am Main
11. Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, Underhill P, Tournev I, Marushiakova E, Popov V 2001 Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. Eur J Hum Genet 9:97–104
12. Labuda M, Labuda D, Korab-Laskowska M, Cole DE, Zietkiewicz E, Weissenbach J, Popowska E, Pronicka E, Root AW, Glorieux FH 1996 Linkage disequilibrium analysis in young populations: pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians. Am J Hum Genet 59:633–643
13. Liegeois JP 1994 Roma, Gypsies, Travellers. Strasbourg, France, Council of Europe Press
14. Fraser A 1992 The Gypsies. Oxford, Blackwell Publishers
15. Kalaydjieva L, Gresham D, Gooding R, Heather L, Baas F, de Jonge R, Blechschmidt K, Angelicheva D, Chandler D, Worsley P, Rosenthal A, King RHM, Thomas PK 2000 N-myc downstream regulated gene 1 is mutated in hereditary motor and sensory neuropathy–Lom. Am J Hum Genet 67:47–58
16. Rogers T, Chandler D, Angelicheva D, Thomas PK, Youl B, Tournev I, Gergelcheva V, Kalaydjieva L 2000 A novel locus for autosomal recessive peripheral neuropathy in the EGR2 region on 10q23. Am J Hum Genet 67:664–671

17. Angelicheva D, Turnev I, Dye D, Chandler D, Thomas PK, Kalaydjieva L 1999 Congenital cataracts facial dysmorphism neuropathy syndrome: a novel developmental disorder in Gypsies maps to 18qter. Eur J Hum Genet 7:560–566

18. Plašilová M, Stoilov I, Sarfarazi M, Kadasi L, Feráková E, Ferák V 1999 Identification of a single ancestral CYP1B1 mutation in Slovak Gypsies (Roms) affected with primary congenital glaucoma. J Med Genet 36:290–294

19. Levy HL 1980 Screening for galactosemia. In: Burman D, Holton JB, Pennock CA (eds) Inherited Disorders of Carbohydrate Metabolism. Lancaster, U.K., MTP Press Limited, pp 133–140

20. Stambolian D, Ai Y, Sidjanin D, Nesburn K, Sathe G, Rosenberg M, Bergsma DJ 1995 Cloning the galactokinase cDNA and identification of mutations in two families with cataracts. Nat Genet 10:307–317

21. Asada M, Okano Y, Imamura T, Suyama I, Hase Y, Isshiki G 1999 Molecular characterization of galactokinase deficiency in Japanese patients. J Hum Genet 44:377–382

22. Kolosha K, Anoia E, de Cespedes C, Gitzelmann R, Shih L, Casco T, Saborio M, Trejos R, Buist N, Tedesco T, Skach W, Mittlemann O, Ledel D, Huang K, Stambolian D 2000 Novel mutations in 13 probands with galactokinase deficiency. Hum Mutat 15:447–453

23. Hunter M, Angelicheva D, Levy HL, Pueschel SM, Kalaydjieva L 2001 Novel mutations in the GALK1 gene in patients with galactokinase deficiency. Hum Mutat 17:77–78

24. Gitzelmann R, Hansen RG 1980 Galactose metabolism, hereditary defects and their clinical significance. In: Burman D, Holton JB, Pennock CA (eds) Inherited Disorders of Carbohydrate Metabolism. Lancaster, U.K., MTP Press Limited, pp 61–101

25. Reyniers A 1995 Gypsy populations and their movements within central and eastern Europe and towards some OECD countries. In: International Migration and Labour Market Policies: Occasional Papers No. 1. Organisation for Economic Co-operation and Development, Paris

26. Romani east-west migrations: strangers in anybody's land. Cambridge Review of International Affairs, Issue 13-2: Spring-Summer 2000