European Journal of Human Genetics (1999) 7, 188–196

http://www.stockton-press.co.uk/ejhg

ARTICLE

Tracing past population migrations: genealogy of steroid 21-hydroxylase (*CYP21*) gene mutations in Finland

Antti Levo[1], Jarmo Jääskeläinen[2], Pertti Sistonen[1], Marja-Kaisa Sirén[1], Raimo Voutilainen[2] and Jukka Partanen[1]

[1]*Finnish Red Cross Blood Transfusion Service, Tissue Typing Laboratory, Helsinki*
[2]*Kuopio University Hospital, Department of Pediatrics, Kuopio, Finland*

The genealogic origin of steroid 21-hydroxylase gene (*CYP21*) mutations and associated haplotypes was determined in 74 unrelated Finnish families with CYP21 deficiency (congenital adrenal hyperplasia, CAH). These families account for two thirds (85/119) of all diagnosed patients of Finnish descent found in this country. We recently demonstrated that multiple founder mutations each associated with a particular haplotype can be found in Finland. Interestingly, some of the haplotypes were identical to those observed in various European populations, whereas others have not been described elsewhere, indicating a local and perhaps a more recent origin. In the present report we show that each of the major founder haplotypes originates from a particular geographic region of Finland. Thus many local genetic isolates are to be expected in Finland. Our finding is in a clear contrast to the genetic diseases known as the 'Finnish disease heritage', in which only one mutation usually predominates. Some of the CYP21 haplotypes proved very informative for analysis of the history of the Finnish population. For example, the origin of one frequent haplotype was shown to cluster in a region assumed by archaeological data to be a major site of immigration by settlers of either Scandinavian or Baltic origin during the first centuries AD. As this haplotype is frequent in many European patient populations, we provide independent genetic evidence of this Iron Age immigration. On the other hand, another frequent haplotype found solely in Finland reflects a more recent (post 15th century) settlement expansion. Consequently, well characterised and sufficiently frequent autosomal gene markers can provide useful information on migrations both between and within populations.

Keywords: steroid 21-hydroxylase; CYP21; congenital adrenal hyperplasia; major histocompatibility complex; gene mutations; haplotype markers; population migrations; Finnish population history

Introduction

The special features of the Finnish population structure[1–4] provide a good opportunity for studying the origin and distribution of genetic traits. This attribute has proven particularly advantageous in the characterisation of inherited diseases known as the 'Finnish disease heritage'. This concept embodies over 30 mostly autosomal recessive diseases that are more prevalent in Finland than elsewhere.[3–5] Usually a single mutation enriched as a result of founder effect accounts for most, if not all, cases of defective alleles. Only in a few other inherited diseases has the mutation spectrum in Finland been analysed to same extent. Classic congenital adrenal hyperplasia (CAH) due to steroid 21-hydroxylase (CYP21) deficiency.[6,7] shows neither an increased nor a decreased incidence in Finland, but one (1 in 15 000 newborns) consistent with that reported worldwide.[8,9] In clear contrast to the diseases of the Finnish disease heritage, multiple independent founder mutations each associated with a particular haplotype are

Correspondence: Antti Levo or Jukka Partanen, FRC Blood Transfusion Service, Tissue Typing Laboratory, Kivihaantie 7, FIN-00310 Helsinki, Finland. Tel: + 358 9 58011; Fax: + 358 9 5801 429; E-mail: antti.levo@bts.redcross.fi, *or* jukka.partanen@bts.redcross.fi
Received 20 January 1998; revised 13 August 1998; accepted 19 August 1998

found in CAH in Finland.[10] Therefore, CAH could serve as a model for investigating the genetic features underlying a 'classic' inherited disease in Finland. It should be of interest to compare these results to those obtained from analyses of the 'Finnish disease heritage'. In addition, as some – but not all – of the founder mutation–haplotypes we observed in Finland[10] are also found in other European populations, and as the carrier frequency of the CYP21 defect is sufficiently high (~ 1/60), CYP21 mutations could serve as informative markers for the evaluation of genetic relationships between populations. Data obtained from analyses of mitochondrial and Y-chromosomal variation (eg Lahermo *et al*[11]) seem most useful in demonstrating more ancient relationships, whereas certain autosomal markers could provide information of more recent periods.

The gene encoding the steroid 21-hydroxylase enzyme, CYP21, as well as a closely related pseudogene CYP21P,[12,13] are tandemly located in the human major histocompatibility complex (MHC) on chromosome 6p21.3 (Figure 1). With the exception of a few novel sporadic point mutations reported, only two types of mutation are found in CYP21 deficiency:[14] large deletions leading to a nonfunctional CYP21P/CYP21 fusion gene, and gene conversion type of mutation where a part of the functional gene has been replaced by a pseudogene-derived deleterious segment. As the pseudogene carries only a limited number of mutations,[14] identical but obviously independently borne mutations in the CYP21 gene have been detected in all populations studied so far. Thus, haplotype information is essential for analysis of origins of CYP21 mutations. We now present mutation–haplotype data, as well as genealogic data, of more than two thirds of all patients found in Finland. This allowed us to trace the genealogic origin and geographical distribution of the multiple CYP21 mutation–haplotype combinations found in Finland.

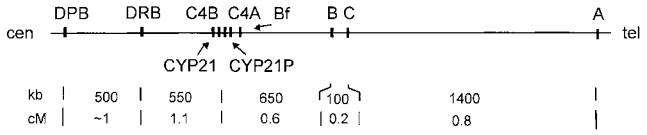

Figure 1 *Haplotype marker loci in the human MHC region.*

Schematic organisation of the polymorphic marker loci in the human major histocompatibility complex (MHC) on chromosome 6p21.3 used for haplotype construction; centromere to the left, telomere to the right. Their approximate physical and genetic distances are taken from references.[17,29] The functional steroid 21-hydroxylase gene (*CYP21*) and 21-hydroxylase pseudogene (CYP21P) are duplicated in tandem with the complement C4 genes (*C4A* and *C4B*). kb = kilobase, cM = centi Morgan, Bf = complement factor B gene.

Materials and Methods

Patients

Seventy-four unrelated Finnish families with CAH (85 patients) were analysed. 148 unrelated chromosomes plus two: in one family, both the mother and son were affected. In another family, due to a *de novo* mutation both maternal alleles were considered affected. The CYP21 mutations and clinical phenotypes of 78 of them have been reported previously.[8] In terms of sex ratio, clinical spectrum, and regional distribution, the patients analysed here are a representative sample of the total Finnish patient population. In our recent screening, altogether 120 patients with CAH due to steroid 21-hydroxylase deficiency could be identified in Finland, resulting in an incidence of 1 in 15 000 live births.[8]

CYP21 *Genotypes*

Genotyping of disease-causing mutations in the *CYP21* gene as well as detailed molecular analysis of the most common affected haplotypes, ie mapping of deletion and conversion breakpoints in *CYP21*, was performed by using allele-specific PCR, direct DNA sequencing, PCR-RFLP, standard genomic RFLP followed by Southern blotting, and pulsed-field gel electrophoresis (PFGE) as described in detail previously.[8,10,15,16]

Haplotype Markers

Seven polymorphic markers (HLA-A, -B, -Cw, -DRB1, C4A, C4B and Bf), in the human MHC region[17] encompassing approximately 2.7 Mbp of DNA flanking the *CYP21* gene were typed (Figure 1). In some cases, HLA-DPB1 typings were also performed, expanding the typed genomic region to approximately 3.2 Mbp.

In addition to standard HLA typing methods,[10,16] we utilised a PCR-SSP 'low-resolution' kit (Dynal AS, Oslo, Norway) for the determination of the HLA-A, -B and -Cw alleles of 18 patients not previously typed. Also, all samples serologically typed as HLA-B40 were retyped at the DNA level. Accordingly, the haplotype marked as B48 S01 DR8 in the present study refers to haplotype B40 S01 DR8 of our previous reports.[10,15,16] Some further samples were retyped to resolve serological ambiguities.

Mutation–haplotype Combination Construction

The CYP21 mutations and haplotypes in 51 of the 74 unrelated patients have been reported previously in detail.[10] In 38 families both parents, and in six families one parent, were available for segregation analysis and hence accurate mutation–haplotype combination could be determined. Four additional samples were homozygous for the haplotype markers and CYP21 mutations. For the remaining 26 patients, the criteria for the haplotype reconstruction were as follows: a previously described mutation–haplotype combination[10] was assumed if, firstly, an identical CYP21 mutation was observed, and, secondly, the sample had the characteristic alleles in a minimum of five loci (ie HLA-B, Bf, C4A, C4B, and DRB1) flanking the *CYP21* gene. These criteria, though

arbitrary, were considered to give a reliable outcome, since each of the frequent haplotypes carried at least one marker that is rare in the general population. More allelic variation was allowed in certain families if informative members were available for segregation analysis. Also, two patients were included, for whom only the haplotype results were available, but no DNA for CYP21 mutation determination. In both cases, however, DNA samples of both parents could be analysed, thus allowing an accurate definition of their mutation–haplotype combinations.

Genealogic Analyses

Ancestors of patients were traced initially by questionnaires sent to all families enquiring the full names, dates, and places of birth of the patients' parents and grandparents. An answer was received from 49 families (66.2%). In other cases, information on the pedigrees was obtained from public registers. Distant ancestors were identified using church parish registries, which provide nationwide information on births, deaths, and marriages extending typically some 300 years back (about 12–15 generations). To some extent, the data are accessible also on the World Wide Web (http://www.genealogia.org/historia/fi). However, any information obtained was confirmed by local church registries. In general, the genealogy was traced back at least three generations, but even up to 13 generations in some families.

To investigate the geographic distribution of a mutation, the birthplaces of grandparents of patients carrying a particular mutation–haplotype combination were dotted on a map of Finland of the beginning of the present century, before two recent major population admixtures took place in the country – the resettlement of some 430 000 (> 10% of the population) Karelian evacuees in 1939–44, and the migration from rural regions to urban areas since the second world war.

Haplotype Markers in the Present General Population

HLA antigen frequencies in Finland, as well as their regional variation, were recently reported by Sirén *et al*.[18] The data were obtained from 10 000 voluntary Finnish donors registered with the Finnish Bone Marrow Donor Registry.

Results

Seventy-four unrelated Finnish families with CAH due to CYP21 deficiency (150 unrelated affected chromosomes) were analysed. They accounted for 71% (85/119) of all diagnosed patients of Finnish descent found in this country, ie all patients from whom blood samples for genotyping could be obtained. Nineteen apparently independent founder mutation–haplotypes totalling 83% (124/150) of all affected chromosomes could be identified. The individual frequencies and the associated markers of the four most frequent haplotypes are shown in Figure 2. Essentially, the haplotypes and their frequencies are identical to those recently reported.[10] The remaining fifth of the chromosomes occurred in single families only. Their overall geographic distribution in comparison to the summary of the founder haplotypes is presented in Figure 3A.

Geographic Origin of Haplotypes

No obvious close consanguinity was initially observed between the parents or grandparents of the patients. Figure 2A–C shows the birthplaces of the grandparents of carriers of the three most frequent mutation–haplotype combinations observed. Each was found in more than 10 unrelated families. Together, these haplotypes accounted for nearly half (44.7%; 67/150) of all the affected chromosomes. Map D in Figure 2 shows an example of the origin of one of the 16 lower-frequency haplotypes (2–6 unrelated families for a given haplotype; the 15 other maps are available from the authors on request). In all cases, the average date of birth of the grandparents was well before the second world war, that is before the major population movements in Finland.

It is evident that the three high-frequency haplotypes all have a rather limited geographic origin. The most prevalent haplotype (Figure 2A) shows three clear clusters. The most abundant is around the coastal town of Oulu in north-western Finland (Figure 2). Another, closely connected to the Oulu cluster, is situated just 100 kilometres to the south of it and extends along the Kalajoki river valley to the inland region of Savo. The third cluster is near the coastal towns in the Vakka–Suomi region in south-western Finland (Figure 2). Interestingly, the families of Vakka–Suomi have a recombinant haplotype carrying a different telomeric segment (A2 Cw2 B61 instead of A24 Cw8 B48). The origin of the ancestors of families carrying the second most common haplotype (Figure 2B) forms a continuum extending along the Kokemäki river from the coastal areas of Satakunta eastwards via Häme to Savo and Karelia in eastern Finland. The birthplaces in Figure 2C are more scattered, but it should be noted that they actually represent four distinct recombinant haplotypes, all demonstrated to have originated from the A3 Cw6 B47 F10 DR7 haplotype.[19] By studying each recombinant separately, clear clustering can be observed. For example, those with B18 follow the route of the Lestijoki river (Figure 2C).

The majority of haplotypes with a lower frequency also showed strong clusters in their origins. For example, the fourth most frequent haplotype (Figure 2D) has a strikingly restricted origin in the region of Southern Ostrobothnia. Nevertheless, some of the lowest frequency haplotypes (maps not shown) may actually represent separate mutational events.

Pedigrees and Consanguinity

In general, there was remarkably little close consanguinity within the Finnish CAH families. Only two marriages between first cousins were found in the 74 pedigrees analysed. However, the parents of the patients were often found to share distant ancestors. Figure 4A represents the family tree of one patient homozygous for the haplotype B7 S31 DR15 (I172N), whose pedigree includes multiple remote consanguineous marriages. Figure 4B presents a case of shared ancestors, dating back to the late 1700s, of two additional patients with the same haplotype.

Some additional patients, carrying identical mutation–haplotype combinations, were demonstrated to be remotely consanguineous. A good example of a complex pedigree is depicted in Figure 4C, which shows the joint family tree of four unrelated patients, one of whom was homozygous, who all carried the most frequent haplotype (Figure 2A) They were found to have common ancestors, born in the region of the Iijoki and Haukipudas parishes (located in the region of Oulu, Figure 2A) in the late 1600s, and the lineages merge on multiple occasions. In other words, there are several possible lines of shared descent.

A common ancestor was also found for certain of the more rare haplotypes; for example, for the only two patients with a B17 S02 DR13 (Q318X) haplotype (Figure 4D; map not shown). Finally, ancestors could often be traced back to a single parish and to the same or adjacent villages. Occasionally, they even lived on the same farm, but no actual link demonstrating direct consanguinity was readily discoverable. This was the case, for example, in haplotypes B40 S30 DR1 (delA) and B47 F10 DR7 (delB).

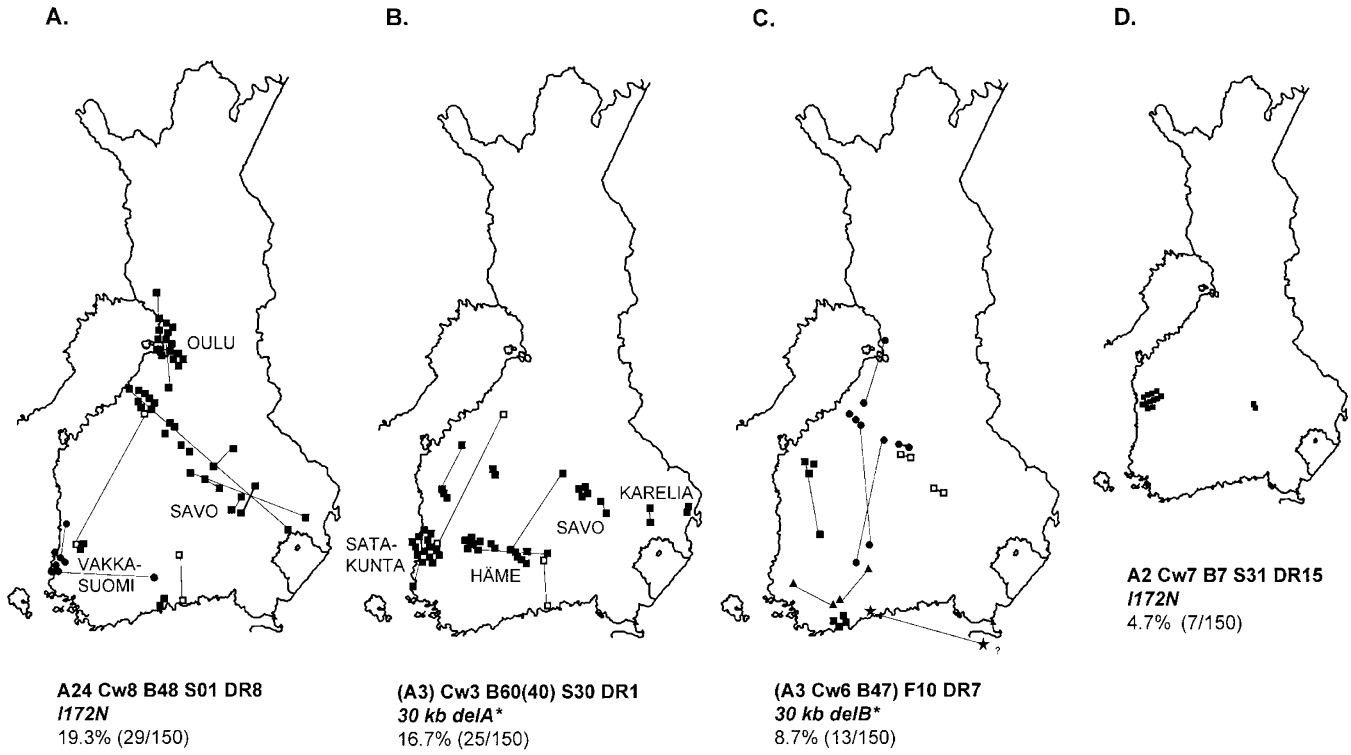

Figure 2A–D *Distribution of the four common CYP21 mutations in Finland. Geographic distribution of the four most frequent CYP21-deficient haplotypes as shown by the birthplaces of grandparents of patients. The mutation and haplotype, as well as the observed frequency in patients, are given below each map. The maps show the territory of Finland in the beginning of the present century, consistent with the grandparents' mean year of birth (1912 in A; 1914 in B; 1915 in C; 1898 in D). As the grandparents were not available for genetic analyses, the birthplaces of both grandparents either on the maternal or the paternal side are shown. Hence, it should be noted that in fact only half of the points presented are true.*
Married couples are connected with lines. In Figures 2A and C, the recombinants were included in the same map and are marked by different symbols. In Figure 2A, ● indicates the A2 Cw2 B61 haplotype instead of A24 Cw8 B48. In Figure 2C: ● (A3 Cw7 B18), ▲ (A3 Cw3 B62), ★ (A11 Cw4 B35); ■ the original A3 Cw6 B47. Open symbols for all four grandparents indicate that the parents could not be studied. * *delA* and *delB* denote distinct large deletions with different breakpoints in the *CYP21* gene.[10]

Markers for Disease Haplotypes in the Present Population

We utilised the databank of the Finnish Bone Marrow Donor Registry (n = 10 000 individuals) to check whether any trace of the clusters could be seen in the present Finnish population. We selected only three haplotypes for this study, since we reasoned that in order to be able to interpret the results the haplotypes tested ought to have a readily identifiable tag, ie a rare HLA allele. We looked for the current place of residence of those individuals positive for HLA B48 together with A24 and DR8, the haplotype in Figure 2A; for B47 with DR7, the haplotype in Figure 2C; and for alleles B22 with DR4 (the haplotype B22 S45 DR4 with mutation I172N + P453S, map not shown). No definitive conclusions could be drawn from the B22 or B47 searches. On the other hand, the clear majority of B48-positive haplotypes was found to occur also today in the same general region (data not shown), whence the CYP21-defective A24 Cw8 B48 S01 DR8 haplotype originated (Figure 2A).

Discussion

The two major findings of the present study are that each of the frequent mutation–haplotype combinations has a clear restricted geographic origin in Finland, and that some of the haplotypes proved very informative for studies of the population history of the country. To our knowledge, this report is the first attempt to trace the origin of CYP21-deficient chromosomes in a human population.

Many single-gene defects in Finland, particularly those belonging to the 'Finnish disease heritage', have been shown to be due to just one major mutation.[3,4] CYP21 deficiency appears to be an exception: 19 founder mutations and a number of sporadic mutations exist in Finland (Levo[10] and the present report). An explanation for this heterogeneity could be the high frequency of CYP21 deficiency. The worldwide carrier frequency has been estimated to be about 1/60,[9] from which Finland seems not to deviate.[8] Assuming that there has been no substantial change in the carrier rate

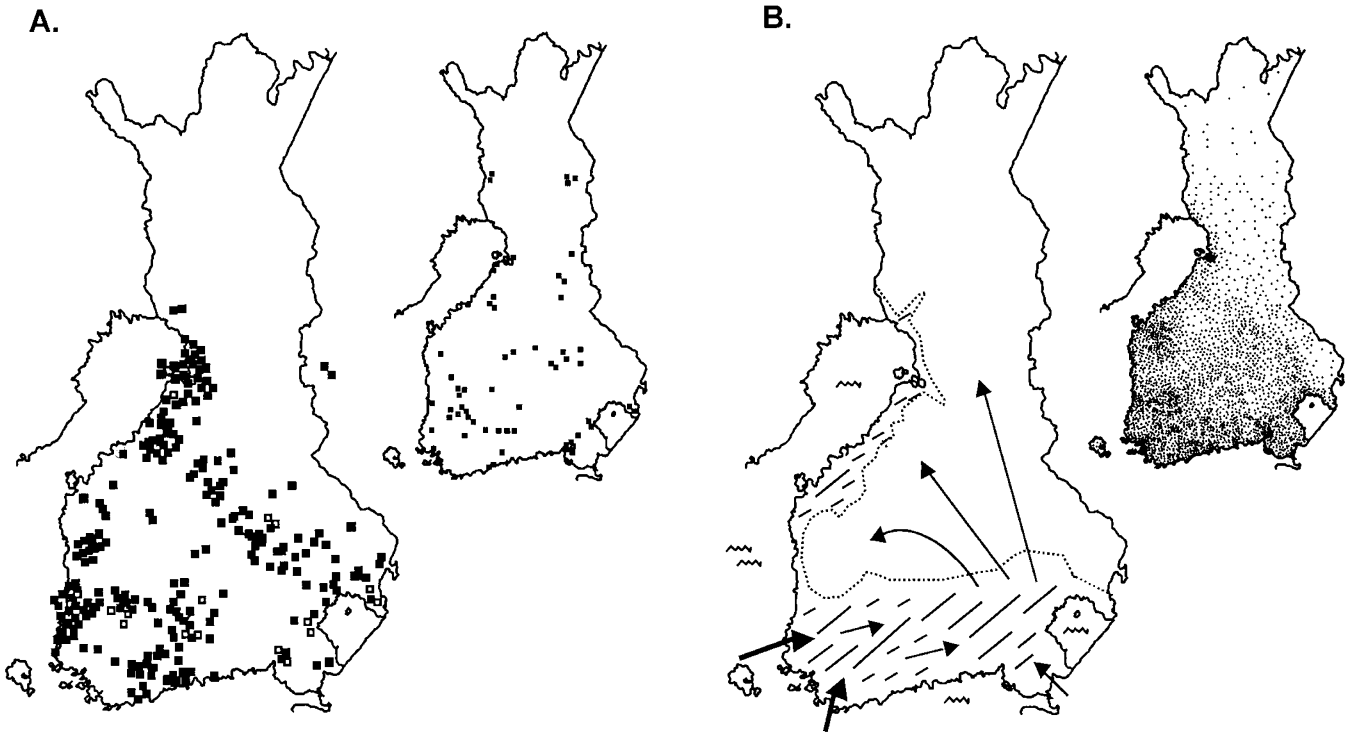

Figure 3 *Overall distribution of CYP21 mutations in Finland, main lines of settlement and distribution of population.* **A** *Geographic distribution of birthplaces of ancestors of all analysed Finnish patients with CYP21 defect carrying any one of the 19 founder haplotypes (left) as compared with the distribution of haplotypes occurring only in single families (right).* **B** *A schematic picture of the major population migrations in Finland (left); the arrows denote the main directions. Since the end of the glacial period (c. 9000 years ago) there has been gradual but continuous migration to Finland mainly from the south and southwest. The ruled area indicates permanent settlements before 1500 AD. On the right: distribution of rural population in Finland at the turn of the century (1908); each dot represents 1000 inhabitants. Adapted from references.*[1,22–24]

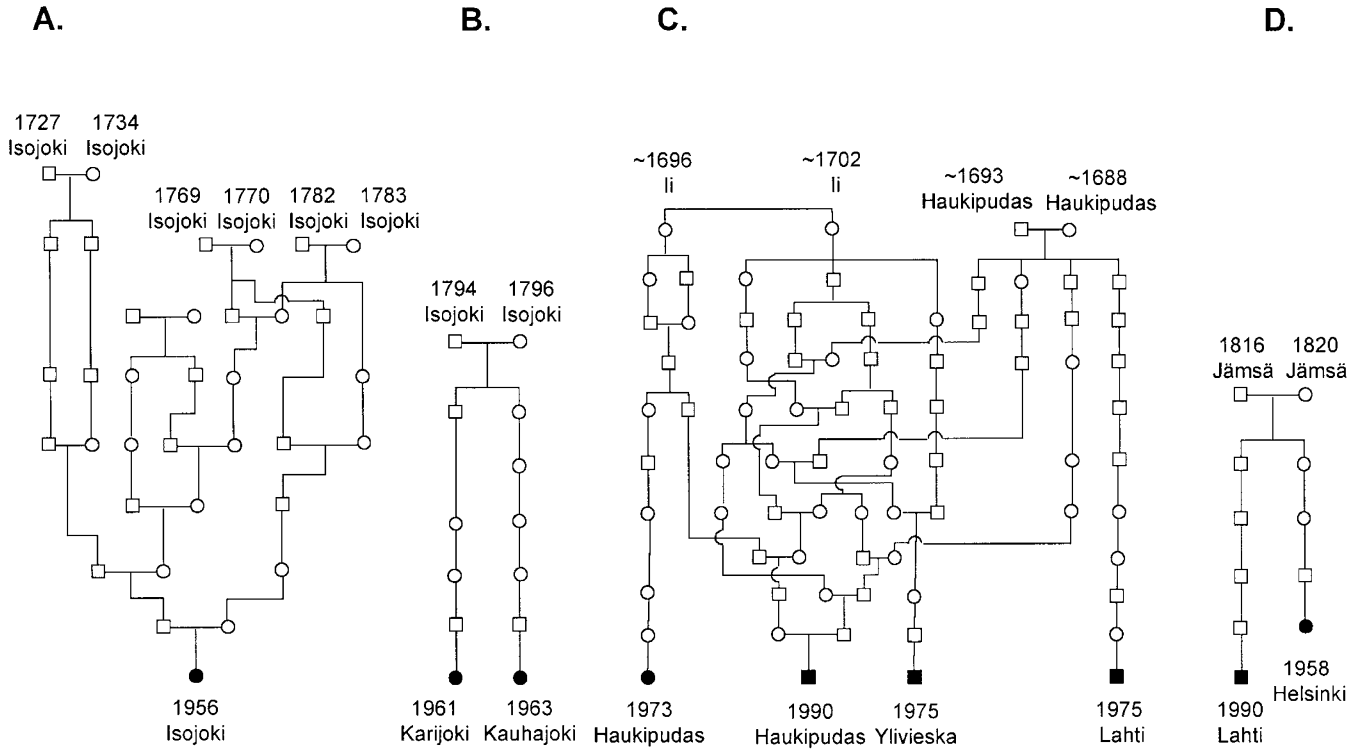

Figure 4 *Typical pedigrees of Finnish CAH families* **A** *The ancestry of one patient homozygous for the B7 S31 DR15 (I172N) mutation showing multiple consanguinity.* **B** *A pedigree demonstrating direct consanguinity between two carriers of the B7 S31 DR15 (I172N) mutation.* **C** *A joint family tree of four patients (one homozygous) carrying the B48 S01 DR8 (I172N) mutation. The tree demonstrates common ancestors and how the lineages merge on multiple occasions.* **D** *A pedigree demonstrating direct consanguinity between the two carriers of the B17 S02 DR13 (Q318X) mutation.*

in the past, many of the early settlers of Finland may have carried a defective *CYP21* gene. In addition, the frequency of *de novo* mutations in CYP21 is high (1–2% of affected alleles),[14] increasing further the mutational spectrum. Our finding of genetic heterogeneity in CYP21 deficiency is of general importance for genetic studies of common disorders. The susceptibility loci for these diseases are usually thought to be frequent polymorphisms or mutations, rather than rare defective alleles. The genetic heterogeneity of these loci in the general Finnish population can similarly be expected to be rather high. However, in local subpopulations or clearly defined clinical subgroups one can still expect founder mutations. Indeed, recent studies of the genetics of asthma[20] and multiple sclerosis[21] in Finland have focused on local isolates. If we consider each founder CYP21 mutation separately, many of them in fact are very similar to the findings of the Finnish disease heritage.[2–5] The overall geographic distribution of the CYP21 founder mutations in Finland is somewhat scattered (Figure 3A left), demonstrating that there is no one particular area from which the majority of cases originate. This fits with the idea of a multiple founder effect.

Unexpectedly, we found evidence that even today some of the founder haplotypes are still mainly restricted to the same region where they are thought to have originated; the majority of individuals positive for the A24, B48, and DR8 alleles live in the very same region as shown in Figure 2A. These findings are consistent with the concept of 'isolation by population density', introduced by Nevanlinna a quarter of a century ago[1] to interpret the deviations in gene frequencies of local subpopulations. Thus the population admixture in the more rural areas is still low and some present-day subpopulations are good representations of the genetic features of the population which occupied the region in the past.

Although a clear geographic origin could be inferred for many of the mutations, a common ancestor was identified only rarely. However, it should be noted that in most cases we actually did not even attempt to trace back all relatives; our aim was merely to show evidence of shared origin. In addition, many shared lines that

indicate multiple consanguinity were evident in more complex pedigrees, like the one shown in Figure 4C. Therefore it cannot be taken for granted that any of them actually represents the shared origin of this particular gene mutation. Although inbreeding in Finland has been thought to be low, for example marriages between first cousins were prohibited by law until 1872, the remote consanguinity found in the pedigrees may be a common feature in rural areas of Finland. This effectively hampers definitive identification of a shared ancestor based on pedigrees only.

The genealogic and geographic data collected in the present paper are interesting in connection with the known population history of Finland.[1,22–24] Archaeological evidence clearly indicates that Finland has been inhabitated since the end of the glacial period some 9000 years ago, but these earliest residents presumably have only a minor influence on today's gene pool. There has been gradual, rather low in number but continuous, migration to Finland, mainly from the south and southwest across the Gulf of Finland and the Baltic Sea throughout prehistoric times (Figure 3B). The total number of these settlers, who were predominantly of Baltic, Germanic, and Scandinavian origin must have been small, tens at a time rather than thousands. Additional immigration to Finland has been sparse throughout historical times. Only the southwestern and southern coastal parts of Finland were inhabited rapidly and early (Figure 3B). The vast northern and northwestern wildernesses were settled permanently as late as in the 16th and 17th centuries, only 15–20 generations ago, mainly by a relatively small number of families from the Savo region.

The distribution of the two most frequent CYP21-deficient haplotypes (Figures 2A and Figure 2B) is notably consistent with two major population migrations documented in Finland (Figure 3B). The geographic distribution of ancestors of the second most common haplotype, B40 S30 DR1 with a large deletion (Figure 2B), presents a clear continuum in the southwestern region stretching from west to east. The southwestern region is considered an area of the most ancient inhabitation.[1,22,23] This pattern parallels an archaeologically well documented population movement in Finland.[22] There has been a staged spread of permanent inhabitation (beginning in the Iron Age in approximately 200 AD) from the southwestern coastal regions along the Kokemäenjoki river valley to Häme (about 600–700 AD) extending gradually to the eastern parts of Finland (about 800 AD). Moreover, archaeological findings in these areas are continuous and similar to those found in southern Scandinavia and in the areas east of the Baltic Sea.[1,22] As the identical B40 S30 DR1 haplotype is also consistently found in patients in many other populations in Europe,[10] for example its frequency in Irish patients is 26%,[25] this mutation was probably introduced into the Finnish population by immigration and began to spread nearly 2000 years ago. The fact that we found no consanguinity between any families is in line with this haplotype's believed ancient origin. We have also analysed one patient of Estonian ancestry (unpublished results), who was homozygous for this very same haplotype. Unfortunately there are no haplotype data on Swedish or other Scandinavian patients, so we are not able to make a better guess of its origin.

On the other hand, the geographic distribution of the most frequent haplotype, B48 S01 DR8, which seems to be restricted to Finland,[10] is in agreement with another major population migration within the country. The expansion of people from Savo from the late 15th century onward is well documented.[24] It started in southern Savo with one of its main streams proceeding rapidly northwestward, including the Kalajoki river and Oulu areas (Figure 3B). Indeed, based on parish records, it is known that the first permanent settlers in the upper reaches of the Kalajoki river in the early 1500s came from Savo.[26] Consequently, we can speculate that this particular mutation spread into the population at least some 400–500 years ago. Thereafter it was enriched by random drift in the various isolated rural subpopulations. This interpretation fits with our genealogic analyses, which demonstrate that the family lineages merge several times, their most recent common ancestors dating from the 17th century. However, the observation of a third cluster in the old southwestern settlement area implies that this haplotype with A2 Cw2 B61 alleles may actually represent the 'original' mutated haplotype possibly introduced into the Finnish population by immigration at a much earlier point in time, whereas the B48 chromosome might be a recombinant that has become greatly enriched over the centuries in the more recently inhabited northern regions. Yet another interpretation is that the two haplotypes are actually independent.

The origin of the third most common haplotype, B47 F10 DR7 carrying a large deletion, is more obscure due to the lower number of cases. It also shows more haplotypic variation. The scattered geographic distribution of birthplaces (Figure 2C) suggests that it may have

more than one single origin. It should be noted that this deletion haplotype is widespread in most patient groups of European descent.[10] Interestingly, the overall distribution of the B47 haplotypes shows similarity with the clustering of the cystic fibrosis mutation (Δ)F508 in Finland,[27] another widespread and ancient human mutation.

The number of individuals carrying one of the lower frequency haplotypes was small, only 2–6 families per haplotype, but they also showed tight geographic clustering. For example, the fourth most frequent haplotype (Figure 2D) has a strikingly narrow origin. Unlike the three most common defective haplotypes, whose HLA allele combinations are extremely rare in the general Finnish population, the B7 S31 DR15 combination has a frequency of 6% in Finland,[28] making it one of the most frequent haplotypes. Many facts strongly suggest a common and fairly recent origin for this mutation: firstly, virtually no allelic variation was seen in the patients' affected haplotypes; secondly, two of the five patients were homozygous for it; and finally, all but one patient's ancestors map to three neighbouring parishes (Figure 2D). The most recent common ancestors of two of the patients with this haplotype were born in the late 18th century (Figure 4B). We can therefore speculate that the mutation occurred at least 200 years ago. A few possible exceptions for the limited origin were also observed, for instance, the haplotypes B7 S31 DR13 and B62 S42 DR4 (maps not shown), both with the 12 splice mutation.

In summary, the CYP21 mutations in Finland appeared very disparate in age and origin. We were able to provide independent, though speculative, genetic evidence of the two major population migrations that previously have been documented by archaeological and historical findings. Consequently, well characterised and sufficiently frequent autosomal gene markers can provide useful information on migrations both between and within populations. The data from various autosomal, mitochondrial and Y-chromosomal markers may each enlighten different perspectives of human population histories.

Acknowledgements

We are grateful to all the Finnish CAH families who kindly provided family data for this study. We also want to express our sincere gratitude to Dr Matti Huurre, Dr Markku Mäkivuoti, and Professor Jouko Vahtola for their expert advice in the fascinating fields of archaeology and history. We also thank the personnel of church registry offices throughout the country for kindly contributing genealogic data. This work was financially supported by the Foundation for Pediatric Research, the Paulo Foundation, the Instrumentarium Scientific Fund, the Maud Kuistila Memorial Foundation, the Jalmari and Rauha Ahokas Foundation, and the Finnish Cultural Foundation.

References

1 Nevanlinna HR: The Finnish population structure. A genetic and genealogic study. *Hereditas* 1972; **71**: 195–236.
2 Norio R, Nevanlinna HR, Perheentupa J: Hereditary diseases in Finland; rare flora in rare soil. *Ann Clin Res* 1973; **5**: 109–141.
3 de la Chapelle A: Disease gene mapping in isolated human populations: the example of Finland. *J Med Genet* 1993; **30**: 857–865.
4 Peltonen L, Pekkarinen P, Aaltonen J: Messages from an isolate: lessons from the Finnish gene pool. *Biol Chem* 1995; **376**: 697–704.
5 Norio R: The diseases of Finland and Scandinavia. In: Rothschild HR (ed). *Biocultural Aspects of Disease*. Academic Press: New York, 1981, pp 359–415.
6 White PC, New MI, Dupont B: Congenital adrenal hyperplasia. *N Engl J Med* 1987; **316**: 1519–1524.
7 Morel Y, Miller WL: Clinical and molecular genetics of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Adv Hum Genet* 1991; **20**: 1–68.
8 Jääskeläinen J, Levo A, Voutilainen R, Partanen J: Population-wide evaluation of disease manifestation in relation to molecular genotype in steroid 21-hydroxylase (CYP21) deficiency: good correlation in a well-defined population. *J Clin Endocrinol Metab* 1997; **82**: 3293–3297.
9 Pang S, Wallace MA, Hofman L *et al*: Worldwide experience in newborn screening for classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Pediatrics* 1988; **81**: 866–874.
10 Levo A, Partanen J: Mutation–haplotype analysis of steroid 21-hydroxylase (CYP21) deficiency in Finland. Implications for the population history of defective alleles. *Hum Genet* 1997; **99**: 488–497.
11 Lahermo P, Sajantila A, Sistonen P *et al*: The genetic relationship between Finns and the Finnish Saami (Lapps): Analysis of nuclear DNA and mtDNA. *Am J Hum Genet* 1996; **58**: 1309–1322.
12 Higashi Y, Yoshioka H, Yamane M, Gotoh O, Fujii-Kuriyama Y: Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. *Proc Natl Acad Sci USA* 1986; **83**: 2841–2845.
13 White PC, New MI, Dupont B: Structure of human steroid 21-hydroxylase genes. *Proc Natl Acad Sci USA* 1986; **83**: 5111–5115.
14 White PC, Tusié-Luna M-T, New MI, Speiser PW: Mutations in steroid 21-hydroxylase (CYP21). *Hum Mutat* 1994; **3**: 373–378.

15 Partanen J, Kere J, Wessberg S, Koskimies S: Determination of deletion sizes in the MHC-linked complement C4 and steroid 21-hydroxylase genes by pulsed-field gel electrophoresis. *Genomics* 1989; **5**: 345–349.
16 Partanen J, Koskimies S, Sipilä I, Lipsanen V: Major-histocompatibility-complex gene markers and RFLP analysis of steroid 21-hydroxylase (CYP21) and complement C4 genes in classical congenital adrenal hyperplasia patients in a single population. *Am J Hum Genet* 1989; **44**: 660–670.
17 Campbell RD, Trowsdale J: Map of the human MHC. *Immunol Today* 1993; **14**: 349–352.
18 Sirén MK, Sareneva H, Lokki ML, Koskimies S: Unique HLA antigen frequencies in the Finnish population. *Tissue Antigens* 1996; **48**: 703–707.
19 Levo A, Westman P, Partanen J: An approach to mapping haplotype-specific recombination sites in human MHC class III. *Immunogenetics* 1996; **43**: 136–140.
20 Laitinen T, Kauppi P, Ignatius J *et al*: Genetic control of serum IgE levels and asthma: linkage disequilibrium studies in an isolated population. *Hum Mol Genet* 1997; **6**: 2069–2076.
21 Kuokkanen S, Sundvall M, Terwilliger JD *et al*: A putative vulnerability locus to multiple sclerosis maps to 5p14–p12 in a region synthetic to the murine locus Eae2. *Nat Genet* 1996; **13**: 477–480.
22 Huurre M: *9000 vuotta Suomen esihistoriaa*. 5th edn. Keuruu, Kustannusosakeyhtiö Otava, 1995.
23 Jutikkala E: *Atlas of Finnish History*. 2nd edn, revised. Werner Söderström Oy: Porvoo, 1959.
24 Vahtola J: The main phases of colonization in northern Finland. FARAVID, *Acta Soc Hist Finland Septent* 1993; **16**: 141–151.
25 Sinnot PJ, Costigan C, Dyer PA, Harris R, Strachan T: Extended MHC haplotypes and CYP21/C4 gene organisation in Irish 21-hydroxylase deficiency families. *Hum Genet* 1991; **87**: 361–366.
26 Huurre M, Virrankoski P, Vilkuna K: *Suur-Kalajoen historia I*. Keski-Pohjanmaan Kirjapaino Oy: Kokkola, 1956.
27 Kere J, Estivill X, Chillón M *et al*: Cystic fibrosis in a low-incidence population: two major mutations in Finland. *Hum Genet* 1994; **93**: 162–166.
28 Ikäheimo I, Silvennoinen-Kassinen S, Tiilikainen A: HLA five-locus haplotypes in Finns. *Eur J Immunogenet* 1996; **23**: 321–328.
29 Thomsen M, Barmada M, Bois M *et al*. MHC recombinant families. In: Charron D (ed). *Genetic diversity of HLA – Functional and Medical Implication*. Proceedings of the Twelfth International Histocompatibility Workshop and Conference. EDK Medical and Scientific International Publisher: Paris 1997, vol 1, pp 188–195.