Phylogenetic Analysis of Mitochondrial DNA in Type 2 Diabetes

Maternal History and Ancient Population Expansion

Emma J. Sherratt, Andrew W. Thomas, Rachael Gill-Randall, and John C. Alcolado

Several studies have suggested a maternal excess in the transmission of type 2 (non-insulin-dependent) diabetes. However, the majority of these reports rely on patients recalling parental disease status and hence are open to criticism. An alternative approach is to study mitochondrial DNA (mtDNA) lineages. The hypervariable region 1 of the rapidly evolving noncoding section of mtDNA is suitable for investigating maternal ancestry and has been used extensively to study the origins of human racial groups. We have sequenced this 347-bp section of mtDNA from leukocytes of subjects with type 2 diabetes (n = 63) and age- and race-matched nondiabetic control subjects (n = 57). Consensus sequences for the two study groups were identical. Pairwise sequence analysis showed unimodal distribution of pairwise differences for both groups, suggesting that both populations had undergone expansion in ancient times. The distributions were significantly different (χ^2 = 180, df = 11, P < 0.001); mean pairwise differences were 4.7 and 3.8 for the diabetic and control subjects, respectively. These data suggest that the diabetic subjects belong to an ancient maternal lineage that expanded before the major expansion observed in the nondiabetic population. Phylogenetic trees constructed using maximum parsimony, neighbor-joining, Fitch-Margolish, or maximum likelihood methods failed to show the clustering of all (or a subset) of the diabetic subjects into one or more distinct lineages. *Diabetes* 48:628–634, 1999

Type 2 (non-insulin-dependent) diabetes represents a common heterogeneous group of disorders defined by hyperglycemia and characterized by impaired insulin release and/or sensitivity (1). Twin (2), family (3), and transracial studies (4) have pointed to the importance of genetic factors in the etiology of the disease but, with a few exceptions (5), type 2 diabetes is not inherited in a simple Mendelian pattern.

Many epidemiologic studies have noted a maternal excess in the transmission of type 2 diabetes (6–12). A major shortcoming of these reports, however, is that they rely on the recollection of patients concerning parental disease status. Usually one or both parents have died and are not available for direct study and, although a maternal excess is not reported in other common disorders such as hypertension (13), it is still possible that the observation is an artifact of some kind. A single study did attempt to perform oral glucose tolerance tests on the parents of patients with type 2 diabetes and found no evidence of maternal transmission (14), although it was possible to investigate both parents in only one-third of the cases (15). If correct, the phenomenon of maternal transmission has profound implications for the study of the genetics of type 2 diabetes (16) and would, for example, support mechanisms such as the role of intrauterine environment or maternally inherited genetic elements. It is now known that mitochondrial DNA (mtDNA) mutations are responsible for at least some cases of maternally inherited diabetes (17–19).

The aim of this study was to investigate the maternal inheritance of diabetes by a technique that would overcome the reliance on patient recall and would not require the availability of parents. mtDNA is exclusively maternally inherited and, unlike nuclear DNA, undergoes little if any recombination (20). Although the majority of the mtDNA sequence is well conserved, a region near the D-loop is highly variable and subject to regular mutation from generation to generation; thus mtDNA successively accumulates mutations that are passed down along the maternal line. This property of mtDNA has made it of immense importance in forensic pathology (21) and the tracing of racial groups (22–24). It has also been possible to postulate an (albeit disputed) sequence of mtDNA that represents the ancestral "Eve" from which divergence took place by the accumulation of a series of mutations in a branch-like manner (25).

Sequence analysis of a population at the D-loop allows definition of the consensus sequence, the sequence shared by the largest number of individuals within the population, which is specific to certain racial groups. In European populations, the consensus sequence in the hypervariable 1 region of the D-loop is identical to the originally published entire mtDNA sequence (the Cambridge sequence). Due to the hypervariability of the D-loop, many subjects within a single racial group will have differences from this consensus, and pairwise analysis gives an indication of the amount of variability within the population. Exponential population growth results in a Poisson distribution of pairwise differences (26), and departures from this pattern may be used to infer details of the nature and timing of population expansion. Some variants from the consensus sequence are relatively common and are used to define race-specific mtDNA haplogroups.

From the Department of Medicine (E.J.S., R.G.-R., J.C.A.), University of Wales College of Medicine; and the Department of BioMedical Sciences (A.W.T.), University of Wales Institute Cardiff, Cardiff, Wales, U.K.

Address correspondence and reprint requests to Dr. J.C. Alcolado, Department of Medicine, University of Wales College of Medicine, Heath Park, Cardiff, Wales, U.K. CF4 4XN. E-mail: alcolado@cardiff.ac.uk.

Received for publication 16 June 1998 and accepted in revised form 10 November 1998.

mtDNA, mitochondrial DNA; PCR, polymerase chain reaction.

The thrifty genotype hypothesis of type 2 diabetes suggests that current susceptibility to the disease is programmed by the presence of gene variants that in ancient times provided a survival advantage (27). It is thought that certain genetic polymorphisms were beneficial in the hunter-gatherer environment and became established in the population only to predispose individuals to type 2 diabetes in the modern era of high caloric intake. Despite this widely invoked mechanism, there is currently little firm evidence to support the hypothesis or the differential expansion of an ancestral diabetes-prone population (28). Analysis of hypervariable mtDNA sequences in patients with type 2 diabetes will provide important information on two counts. First, if the disease is maternally inherited, we might expect that, as a group, affected individuals might be more closely maternally related than the general nondiabetic population. This hypothesis can be tested by examining consensus sequences and parsimony trees. Second, evidence to support the thrifty genotype hypothesis would arise from significant departures of pairwise difference plots between subjects with and without diabetes.

In this study, we sequenced the 347-bp hypervariable region 1 within the D-loop of mtDNA in 63 subjects with type 2 diabetes and 57 race-matched control subjects. The resultant data were analyzed by consensus sequence generation, pairwise difference, and phylogenetic analysis.

RESEARCH DESIGN AND METHODS

Patient selection. Clinical data of patients attending diabetes outpatient clinics at the University Hospital of Wales, Cardiff, are routinely entered into a computerized database from which were selected at random all individuals who had a positive family history of the disease. To concentrate our study on subjects in whom maternal transmission might be important, subjects were recruited if they reported an affected mother on subsequent direct inquiry. In accordance with local ethics board approval, 10 ml venous blood was drawn from consenting patients after the purpose of the study had been explained. All subjects (n = 63) had type 2 diabetes, were European in origin, and had a Welsh maternal grandmother. An age- and race-matched control group (n = 57, mean age 68 years [range 40–83], all with Welsh grandmothers) was recruited from patients attending nondiabetes clinics at the University Hospital of Wales who had no personal or family history of diabetes.

Amplification and sequencing of mtDNA. mtDNA was extracted from EDTA blood samples using a Nucleon DNA extraction kit (Scotlab, Glasgow, Scotland, U.K.). A 501-bp section of the mitochondrial control region was amplified using a forward primer spanning nucleotides 15,971–15,990 and a reverse primer spanning nucleotides 16,471–16,452. Reactions were carried out in final volumes of 25 µl, which contained 400 ng DNA, 200 µmol/l of each dNTP, 50 ng of each primer, 10 mmol/l Tris-HCl (pH 9.0), 50 mmol/l KCl, 0.1% Triton-X100, 1.5 mmol/l $MgCl_2$, and 0.625 U *Taq* polymerase. A total of 31 cycles—45 s at 94°C for denaturation, 1 min at 50°C for annealing, and 3 min at 72°C for elongation—were performed, followed by an extension step of 72°C for 10 min, using an Omnigene thermocycler (Hybaid, Teddington, U.K.). The quality of the polymerase chain reaction (PCR) was determined by electrophoresis of 5 µl of the reaction mixture on gels containing 2% UltraPure agarose (Life Technologies, Paisley, U.K.) in TBE buffer. PCR products were purified using Wizard PCR Prep kit (Promega, Southampton, U.K.). The mtDNA hypervariable region 1 sequence (16,042–16,388 nt) was determined using 30 ng DNA in an Applied Biosystems *Taq* DyeDeoxy Terminator cycle sequencing reaction (Perkin-Elmer/Applied Biosystems Division [PE/ABI], Warrington, U.K.), containing either the forward or reverse primer described. Electrophoresis and sequence analyses were performed with a Model 373A Automated DNA Sequencing System (PE/ABI).

Analysis of data. Hypervariable region 1 sequences were aligned, and pairwise sequence comparisons were made using the PILEUP and DISTANCES functions of the Genetics Computer Group sequence analysis package. In essence, this involves comparing the sequence from each member of the diabetic or control group with all other members of the same group. Pairwise differences for each group were then plotted graphically (frequency versus number of differences). Construction of phylogenetic trees was performed using the PHYLIP package. Both software packages are available via the HGMP Resource Centre worldwide web site. Maximum parsimony and neighbor-joining analyses were performed using DNAPARS or NEIGHBOR with SEQBOOT and CONSENSE to obtain bootstrapped data, and Fitch-Margolish analysis was achieved using FITCH. Maximum likelihood analysis (DNAML) was performed on only a subset of sequences from both study groups, due to processor time constraints. Transition/transversion ratios were estimated by quartet puzzling (29), based on the evolutionary model of Hasegawa et al. (30). Phylogenetic trees were rooted by a theoretical African ancestor (25).

RESULTS

The 120 individuals yielded 81 different sequences for the hypervariable region 1 (Fig. 1). Of these, 44 sequences were exclusive to diabetic subjects, 31 were exclusive to control subjects, and 6 were detected in both diabetic and control subjects. The most common sequence, present in 11 diabetic and 10 control subjects, was identical to the Cambridge reference sequence (31). This represents the consensus sequence of both groups. The variable positions in the consensus sequences differed between the two groups, however: 26 positions were variable in diabetic subjects only, 19 variable in control subjects only, and 31 variable in both study groups.

It is possible to classify individuals into some of the previously defined European mtDNA haplogroups (32) by comparing the hypervariable sequence results with those of other studies in European populations. Analysis of the current results places a similar proportion of patients and control subjects in mtDNA haplogroups J, K, T, and I. The most common haplogroup (H), however, is defined by the absence of an *Alu*I cutting site at nt 7,025 and has no unique sequences in hypervariable region 1. PCR–restriction fragment length polymorphism (RFLP) analysis of our samples (Table 1) also demonstrates that a similar proportion of diabetic subjects and control subjects fall into haplogroup H. These results support the close racial matching of our groups.

The frequency distributions of sequence differences between all possible pairs of individuals within each population were examined. Both diabetic and control groups exhibited a unimodal distribution of pairwise differences (Fig. 2), which suggests that the populations underwent an expansion in ancient times. Theoretical analyses of pairwise differences indicate that a population that underwent a continuous exponential expansion would result in a distribution of pairwise differences which would conform to a Poisson distribution (26). However, distributions for both diabetic and control groups differed significantly from Poisson (diabetic subjects $\chi^2 = 126.61$, $P < 0.00001$; control subjects: $\chi^2 = 49.48$, $P < 0.00001$), showing that neither group came from a population that expanded uniformly with time. Comparison of the two distributions showed that they differed significantly from each other ($\chi^2 = 180$, df = 11, $P < 0.0001$). The statistical significance was unaffected by removing the diabetic subjects with 11, 12, and 14 pairwise differences (that is, the right-tail outliers) from the calculations. The peak for the control population is further to the left than that for the diabetic population: mean pairwise differences were 3.8 and 4.7 for control and diabetic subjects, respectively.

Phylogenetic trees were constructed from the sequence data by maximum parsimony (Fig. 3), neighbor-joining, Fitch-Margolish, and, on a subset of the data, maximum likelihood methods (data not shown). In all analyses, the diabetic and control subjects were distributed equally between all the branches of the trees, rather than the diabetic subjects (or a subset of diabetic subjects) clustering into one or more distinct lineages.

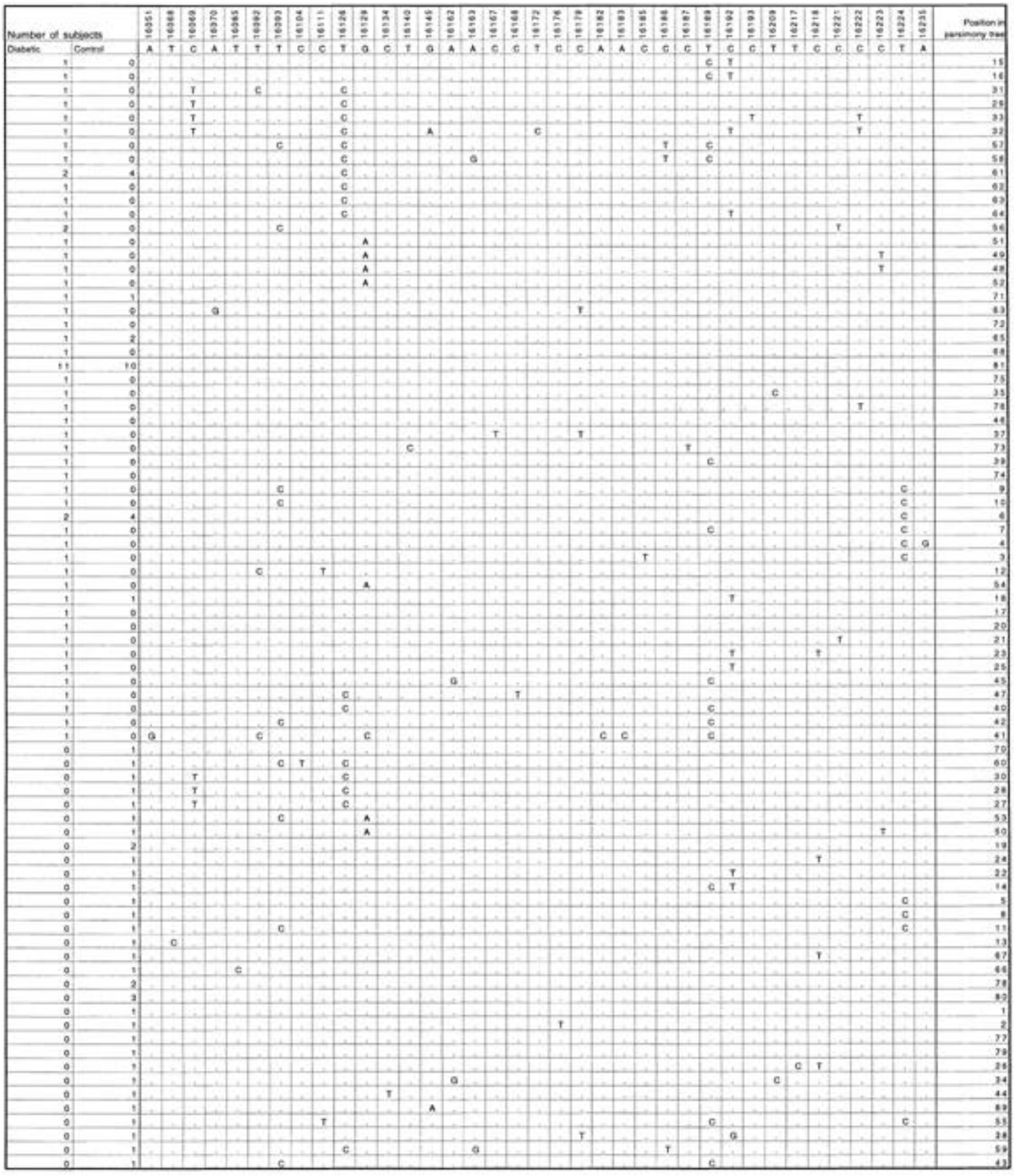

FIG. 1. Polymorphic sites in the hypervariable region 1 of the mitochondrial control sequence, shown as differences from the Cambridge (consensus) sequence (32). Nomenclature is in accordance with this sequence; dots (.) indicate homology with the reference sequence and variations are indicated by the base change and the number of diabetic and control samples found to harbor the polymorphism.

DISCUSSION

The results of this study confirm the highly polymorphic nature of the hypervariable region of mtDNA in the Welsh population. As expected, the most common sequence in both patients with diabetes and control subjects was identical to the Cambridge reference sequence. Fifty-two patients with diabetes and 47 control subjects had sequences that varied at one or more positions from this reference sequence, but because the consensus sequences for both groups were identical, it is unlikely that a single mtDNA variant in the hypervariable

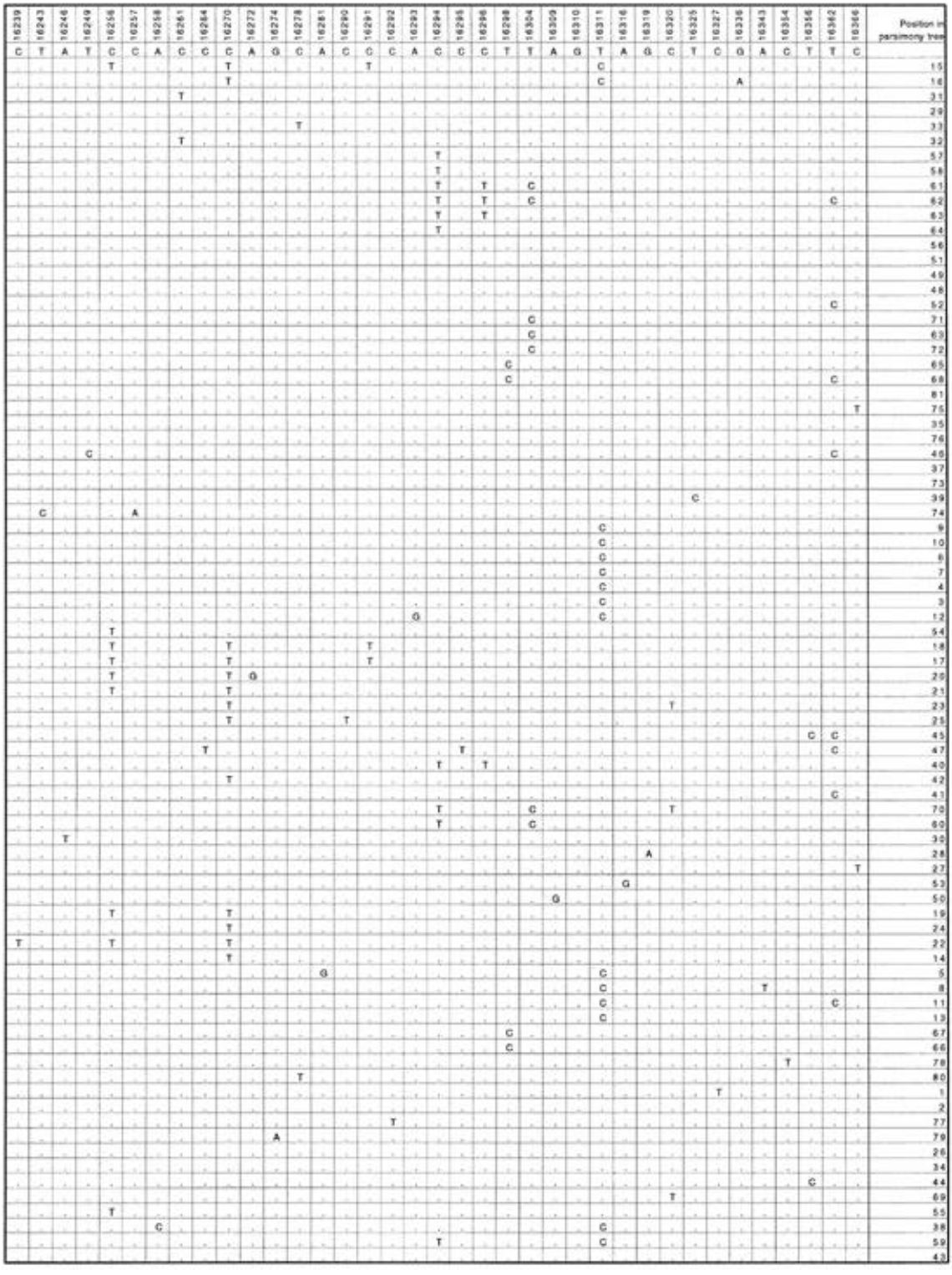

FIG. 1 continued.

region is associated with development of the disease. Although the variable positions in the consensus sequences were different in the two groups, there is no evidence that variation at one particular site is statistically more common in patients with diabetes.

Pairwise differences did not conform to a Poisson distribution in either the diabetic or control subjects. This implies that the ancestral populations did not expand uniformly with time. Such a phenomenon has been reported in other control racial groups and implies that the environmental and biological factors involved in establishing population fertility and survival have varied throughout history.

The key observation is that the group of patients with diabetes had a wider distribution of pairwise differences com-

TABLE 1
Mitochondrial DNA haplogroup distributions in diabetic patients and control subjects

Haplogroup	Diabetic patients	Control subjects
H	29 (47)	26 (46)
U	18 (29)	14 (25)
K	5 (8)	4 (7)
T	4 (6)	4 (7)
J	4 (6)	3 (5)
I	1 (2)	1 (2)
X	2 (3)	1 (2)

Data are *n* (%). Michochondrial DNA haplogroup distributions in diabetic patients and control subjects. Haplogroup letter designations correspond to the caucasoid haplogroups described by Torroni et al. (32). I, J, K, and T were assigned with reference to the hypervariable region 1 sequence data. H, U, and X were assigned by PCR-RFLP analysis.

pared with the nondiabetic population, as indicated by the large statistically significant difference between the resultant pairwise distributions. The studies of Rogers and Harpending (33) suggest that mean pairwise differences increase in number with increasing antiquity of population expansion, since there is more time to develop and fix polymorphisms within the mitochondrial D-loop. Our data therefore suggest that the diabetic subjects belong to a population that underwent expansion earlier in history than the nondiabetic population. In this scenario, we would hypothesize that in ancestral times, a population had enough advantage to expand rapidly, and the offspring of that population are now predisposed to develop type 2 diabetes. In a later period, there was a more rapid expansion of a group whose descendants are now found within the Welsh nondiabetic community.

Our data, for the first time, provide support for the separate expansion of a population destined to develop diabetes in the modern era. This would support the contentions of the thrifty genotype hypothesis, but we cannot speculate regarding the timing of this population growth on the basis of data from a single racial group. The higher mean pairwise differences in the diabetic population compared with the control subjects could have other potential explanations, including the presence of a higher mtDNA mutation (or mutation fixation) rate in the diabetic lineages. It should be noted, however, that these mechanisms would have to be functional at the germ cell level for them to be carried from generation to generation.

Phylogenetic trees constructed using a variety of statistical approaches failed to show clustering of patients with diabetes into one or more distinct maternal lineages. This finding implies that the maternal excess reported in epidemiologic studies is not due to the inheritance of type 2 diabetes from a small number of ancestral maternal lineages that have remained constant throughout history. Although this in turn suggests that the previously reported maternal excess may be spurious, it is not possible to exclude the possibility that the heterogeneous nature of type 2 diabetes has impaired the ability of phylogenetic analysis to define precise maternal relationships.

A number of heteroplasmic mtDNA point mutations and rearrangements have been described in diabetes. Their prevalence in the Welsh population is <1% (34) and therefore too small to have been reflected in clustering within the phylogenetic trees in the current study. More controversially, several homoplasmic mtDNA variants have recently been associated with diabetes (35). The lack of phylogenetic clustering suggests that a major homoplasmic variant has not been transmitted down a maternal line over many generations in the Welsh community.

We conclude that mtDNA analysis of the Welsh population suggests that patients with type 2 diabetes are descendants of a cohort that underwent expansion in ancient times before the rapid expansion of ancestors of the nondiabetic community. However, we could not demonstrate the presence of any unique maternal lineages predisposing to the disorder.

ACKNOWLEDGMENTS

This work was funded by the British Diabetes Association and Welsh Scheme for Health and Social Research.

REFERENCES

1. De Fronzo RA: Lilly Lecture 1987. The triumvirate: beta-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes* 37:667–687, 1988
2. Kaprio J, Tuomilheto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, Stengard J, Kesaniemi YA: Concordance for type 1 (insulin dependent) and type

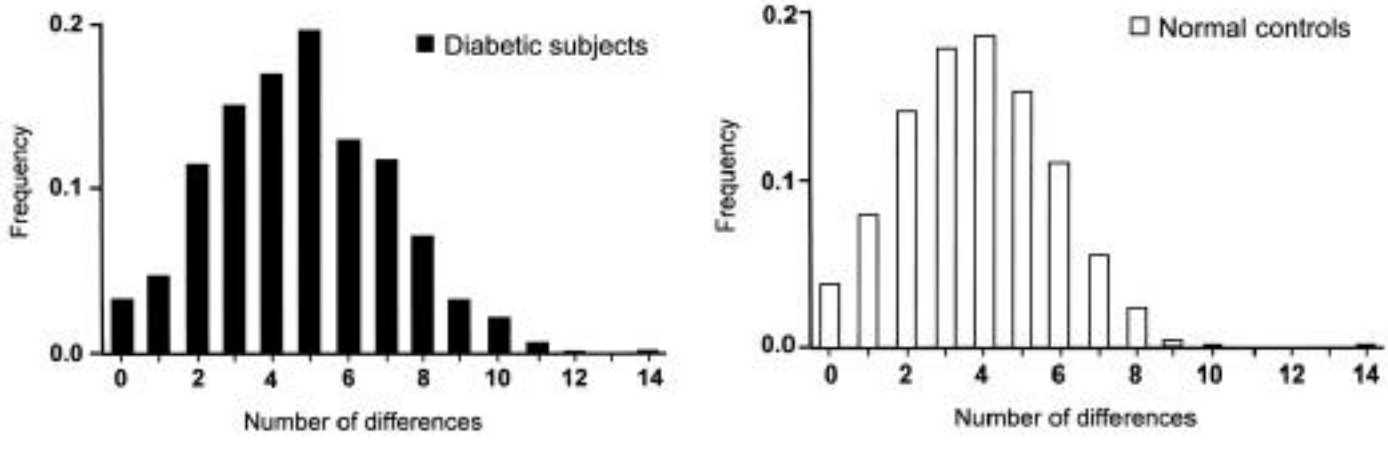

FIG. 2. Histograms showing the distributions of pairwise sequence differences in patients with diabetes and normal control subjects. Both distributions differ from Poisson ($P < 0.00001$) and from each other ($P < 0.0001$), indicating disparate nonlinear population expansions in the two groups.

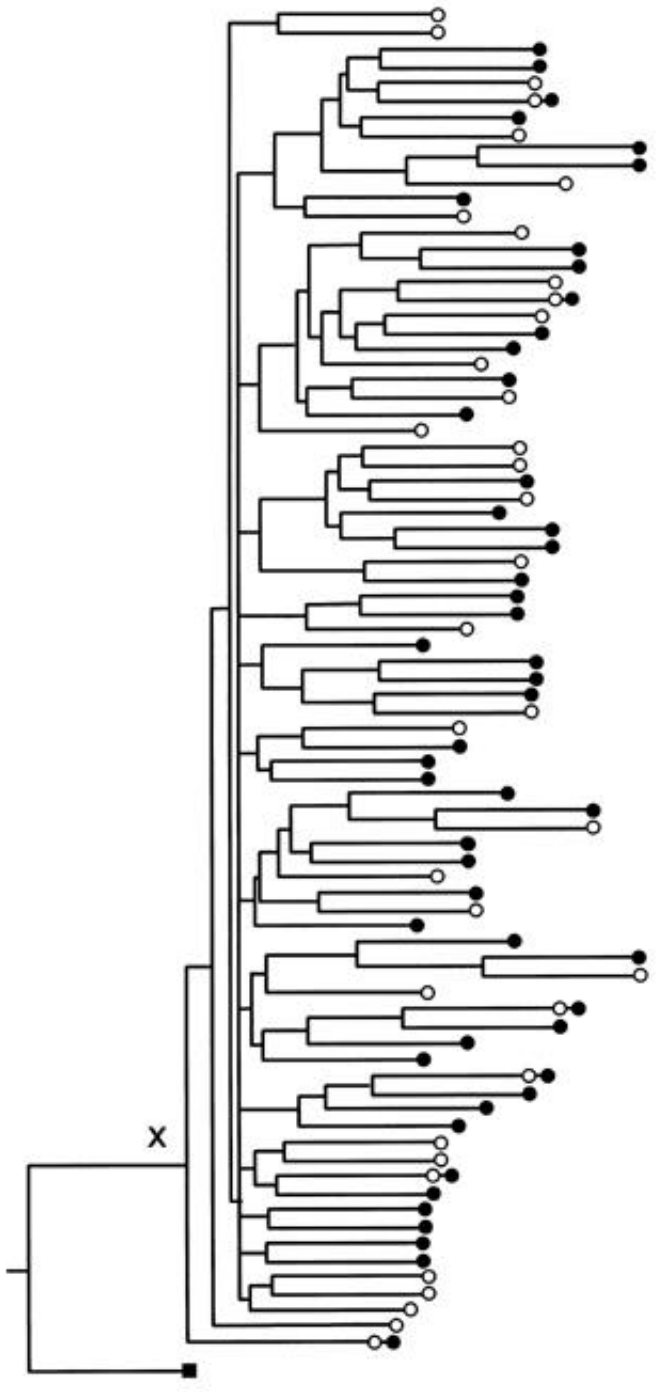

FIG. 3. Phylogenetic maximum parsimony tree derived from the 81 different mtDNA sequences obtained in this study. The symbols at the tips of the tree denote patients with diabetes (●) and control subjects (○). The twigs showing more than one symbol are those sequence types present in both study groups. Nodes marked with an X indicate that lineages to the right of the branching point were present in ≥50% of trees constructed from 100 times bootstrapped data. ■, the hypothetical ancestor that forms the root of the tree.

2 (non-insulin dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* 35:1060–1067, 1992

3. Cook JT, Page RC, O'Rahilly S, Levy J, Holman R, Barrow B, Hattersley AT, Shaw AG, Wainscoat JS, Turner RC: Availability of type 2 diabetic families for detection of diabetes susceptibility genes. *Diabetes* 42:1536–1543, 1993
4. Mohan V, Sharpe PS, Aber VR, Mather HM, Kohner EM: Family histories in Asian and European non-insulin dependent diabetic patients. *Prac Diab* 3:254–257, 1986
5. Tattersall R, Fajans SS: A difference between the inheritance of classic juvenile-onset and maturity onset type diabetes of young people. *Diabetes* 24:44–53, 1975
6. Alcolado JC, Alcolado R: Importance of maternal history of non-insulin dependent diabetic patients. *BMJ* 302:1178–1180, 1991
7. Dorner G, Mohknike A: Further evidence for predominantly maternal transmission of maturity-onset type diabetes. *Endokrinologie* 68:121–124, 1976
8. Martin AO, Simpson JL, Ober C, Freinkel N: Frequency of diabetes mellitus in mothers of probands with gestational diabetes: possible maternal influence on the predisposition to gestational diabetes. *Am J Obstet Gynecol* 151:471–475, 1985
9. Thomas F, Balkau B, Vauzelle-Kervroedan F, Papoz L, The CODIAB-INSERM-ZENECA Group: Maternal effect and familial aggregation in NIDDM. *Diabetes* 43:63–67, 1984
10. Bagriacik N, Ipbuker A, Ilkora H, Gorpe U, Ersan E, Ozyazar M: Diabetes in Turkey: epidemiological characteristics 1959–1987 (Abstract). *Diabetes* 40 (Suppl. 1):1731, 1991
11. Young CA, Kumar S, Young MJ, Boulton AJM: Excess maternal history of diabetes in Caucasian and Afro-origin non-insulin dependent diabetic patients suggests dominant maternal factors in disease transmission. *Diabetes Res Clin Pract* 28:47–49, 1995
12. Simmons D, Galtand BA, Leakehe L, Fleming C: Frequency of diabetes in family members of probands with non-insulin dependent diabetes mellitus. *J Intern Med* 237:315–321, 1995
13. Gagg J, Sherratt EJ, Thomas AW, Peters J, Alcolado JC: Patients report maternal transmission of NIDDM but not IDDM (Abstract). *Diabet Med* (Suppl. 2): S30, 1995
14. Mitchell BD, Kammerer CM, Reinhart LJ, Stern MP, MacCluer JW: Is there an excess maternal transmission of NIDDM? *Diabetologia* 38:314–317, 1995
15. Alcolado R, Gagg J, Thomas AW, Sherratt EJ, Alcolado JC: Testing parents of NIDDM patients. *Diabetologia* 38:1482–1483, 1995
16. Cox NJ: Maternal component in NIDDM: how large an effect? *Diabetes* 43:166–168, 1994
17. Van den Ouweland JMW, Lemkes HHPJ, Ruitenbeek W, Sandkuijl LA, de Vijlder MF, Struyvenberg PA, Van de Kamp JJ, Maassen JA: Mutation in mitochondrial tRNAleu(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness. *Nat Genet* 1:368–371, 1992
18. Ballinger SW, Shoffner JM, Hedaya EV, Trounce I, Polak MA, Koontz DA, Wallace DC: Maternally transmitted diabetes and deafness associated with a 10.4 Kb mitochondrial DNA deletion. *Nat Genet* 1:11–15, 1992
19. Ballinger SW, Shoffner JM, Gebhart S, Koontz DA, Wallace DC: Mitochondrial diabetes revisited. *Nat Genet* 7:458–459, 1994
20. Wallace DC: Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci U S A* 91:8739–8746, 1994
21. Piercy R, Sullivan KM, Benson N, Gill, P: The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106:85–90, 1993
22. Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula, P, Peltonen L, Savontaus ML: The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am J Hum Genet* 58:1309–1322, 1996
23. Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC: Asian affinities and continental radiation of the four founding native American mtDNAs. *Am J Hum Genet* 53:563–590, 1993
24. Corte-Real HBSM, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC: Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Am J Hum Genet* 60:331–350, 1996
25. Cann RL, Stoneking M, Wilson AC: Mitochondrial DNA and human evolution. *Nature* 325:31–36, 1987
26. Slatkin M, Hudson RR: Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555–562, 1991
27. Neel JV: Diabetes mellitus: A "thrifty" genotype rendered detrimental by "progress"? *Am J Hum Genet* 14:352–362, 1962
28. Swinburn DA: The thrifty genotype hypothesis: how does it look after 30 years? *Diabet Med* 13:695–699, 1996
29. Strimmer K, Vonhaeseler A: Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol* 13:964–969, 1996
30. Hasegawa M, Iida Y, Yano T, Takaiwa F, Iwabuchi M: Phylogenetic relationships amongst eukaryotic kingdoms inferred from ribosomal RNA sequences. *J Mol Evol* 22:32–38, 1985

31. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon JC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG: Sequence and organisation of the human mitochondrial genome. *Nature* 290:457–465, 1981
32. Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Saontaus ML, Wallace DC: Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850, 1996
33. Rogers AR, Harpending H: Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9:552–569, 1992
34. Thomas AW, Edwards A, Sherratt EJ, Majid A, Gagg J, Alcolado JC: Molecular scanning of candidate mitochondrial tRNA genes in type 2 (non-insulin dependent) diabetes mellitus. *J Med Genet* 33:253–256, 1996
35. Tawata M, Ohtaka M, Iwase E, Ikegishi Y, Aida K, Onaya T: New mitochondrial DNA homoplasmic mutations associated with Japanese patients with type 2 diabetes. *Diabetes* 47:276–277, 1998