American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 124B:20–28 (2004)

Phenotypic Characterization and Genealogical Tracing in an Afrikaner Schizophrenia Database

Maria Karayiorgou,[1]* Marie Torrington,[2] Gonçalo R. Abecasis,[3] Herman Pretorius,[2] Brian Robertson,[4] Sean Kaliski,[4] Stephen Lay,[4] Christina Sobin,[1] Natalie Möller,[2] S. Laura Lundy,[1] Maude L. Blundell,[1] Joseph A. Gogos,[5] and J. Louw Roos[2]

[1]*Human Neurogenetics Laboratory, The Rockefeller University, New York, New York*
[2]*Department of Psychiatry and Weskoppies Hospital, University of Pretoria, Pretoria, Republic of South Africa*
[3]*Department of Biostatistics, University of Michigan, Ann Arbor, Michigan*
[4]*Department of Psychiatry and Valkenberg Hospital, University of Capetown, Capetown, Republic of South Africa*
[5]*Department of Physiology and Cellular Biophysics and Center for Neurobiology and Behavior, College of Physicians and Surgeons, Columbia University, New York, New York*

Founder populations hold tremendous promise for mapping genes for complex traits, as they offer less genetic and environmental heterogeneity and greater potential for genealogical research. Not all founder populations are equally valuable, however. The Afrikaner population meets several criteria that make it an ideal population for mapping complex traits, including founding by a small number of initial founders that likely allowed for a relatively restricted set of mutations and a large current population size that allows identification of a sufficient number of cases. Here, we examine the potential to conduct genealogical research in this population and present initial results indicating that accurate genealogical tracing for up to 17 generations is feasible. We also examine the clinical similarities of schizophrenia cases diagnosed in South Africa and those diagnosed in other, heterogeneous populations, specifically the US. We find that, with regard to basic sample descriptors and cardinal symptoms of disease, the two populations are equivalent. It is, therefore, likely that results from our genetic study of schizophrenia will be applicable to other populations. Based on the results presented here, the history and current size of the population, as well as our previous analysis addressing the extent of background linkage disequilibrium (LD) in the Afrikaners, we conclude that the Afrikaner population is likely an appropriate founder population to map genes for schizophrenia using both linkage and LD approaches.

KEY WORDS: founder population; schizophrenia genetics; South Africa; genealogical research; gene mapping

INTRODUCTION

The identification of susceptibility genes for common, complex disorders has been a task of unprecedented difficulty. A very high degree of genetic heterogeneity, as well as clinical heterogeneity in certain diseases, such as the psychiatric syndromes, are suspected to be at the core of the difficulty. There are certain expectations that founder populations may be useful to circumvent some of these difficulties by offering reduced genetic heterogeneity, more uniform environment, and the potential for genealogical research. Such populations have proven useful for mapping genes that underlie Mendelian disorders and multigenic disorders that segregate as Mendelian conditions in specific populations [Hastbacka et al., 1994; Kalaydjieva et al., 1996; Nystuen et al., 1996; Visapaa et al., 2002] and particular features of these populations should also facilitate the search for genes that influence susceptibility to complex traits.

Founder populations have originated from a relatively small number of individuals and have expanded over several generations in relative isolation. As a result, it may be reasonable to expect that a decreased number of

Grant sponsor: Essel Foundation; Grant sponsor: EJLB Foundation; Grant sponsor: National Institute of Mental Health (to MK); Grant number: R01 MH 61399-01.

*Correspondence to: Maria Karayiorgou, M.D., Associate Professor, Head, Human Neurogenetics Laboratory, The Rockefeller University, 1230 York Avenue, Box #313, New York, NY 10021. E-mail: karayim@mail.rockefeller.edu

Received 4 February 2003; Accepted 19 May 2003
DOI 10.1002/ajmg.b.20090

independent susceptibility loci are segregating in such a population. At any susceptibility locus, all affected individuals in the population may carry the same or, more likely, a limited set of alleles derived identical-by-descent (IBD) from a few common ancestors. The reduced locus and allelic heterogeneity in founder populations means that genetic mapping strategies may require smaller sample sizes for success. Furthermore, the relatively recent origin of the population ensures that the chromosomal region surrounding the disease allele is larger than in outbred populations, allowing relatively sparse genetic maps in these populations to be as informative as denser maps in outbred populations. As a result, founder populations are more amenable to linkage disequilibrium (LD) approaches, which may be more powerful than traditional linkage analysis for complex traits. It is important to note that the degree of resolution of mapping strategies based on LD between loci will vary between founder populations and depend on the number of generations since the founding. While initial gene localization may be more feasible in young founder populations because of the extensive LD, fine scale localization of a gene may be problematic for the same reason.

It is not known at this point which population history and structure is more advantageous for mapping complex traits. Current genetic linkage studies of schizophrenia include patient samples from founder populations of different ages (such as from Finland, Costa Rica, French Quebec, Palau, Kosrae) [Hovatta et al., 1999; Paunio et al., 2001; DeLisi et al., 2002; Devlin et al., 2002; Wijsman et al., 2003]. Here, we present the Afrikaner population of South Africa as an ideal genetic isolate for our gene mapping studies in schizophrenia. The Afrikaner population meets several criteria that make it an ideal population for mapping complex traits, including (a) founding by a small number of initial founders that likely allowed for a relatively restricted set of mutations to have been introduced in the population, (b) large current population size to allow identification of sufficient number of cases, and (c) strong infrastructure to allow genealogical research as well as collection of good quality clinical data.

The Afrikaners were founded by approximately 2,000 initial settlers who came from The Netherlands and other parts of Northern Europe in consecutive waves, starting at 1,652. The first immigrants settled in the Cape. They later spread inland, founding geographically isolated communities. Besides geography, other factors, such as culture, language differences, and religious practices, contributed to the isolation of the Afrikaner. Most Afrikaners were members of the Dutch Reformed Church. They lived in large families scattered over the country, so that consanguinity was common, especially in early generations. Population growth was almost entirely through reproduction and immigration subsequent to the founding was minimal. In more recent years, some admixture has occurred. The current size of the Afrikaner population is approximately 3 million and their age of ancestry is on average 12.5 generations old.

The demographic history of this population is reflected in the unusually high frequency of certain rare Mendelian disorders (as much as five to ten times higher than in most other population groups), such as variegate porphyria (VP), Huntington's chorea, familial hypercholesterolemia (FH), Fanconi anemia, myotonic dystrophy [Hayden et al., 1980a,b; Brink et al., 1987; Rosendorff et al., 1987; Goldman et al., 1996; Warnich et al., 1996; Groenewald et al., 1998]. Moreover, unusually low allelic diversity at the associated disease loci has been observed in the Afrikaner. This is an expected result of genetic drift in a population of this size. Indeed, studies on FH identified only three founder mutations responsible for more than 95% of FH cases in the Afrikaner population [Leitersdorf et al., 1989]. These Afrikaner mutations were also identified in FH patients in The Netherlands [Defesche et al., 1993] and one of them was shown to have originated there and been introduced to South Africa by a single individual of Dutch origin [Defesche et al., 1996]. Finally, unusually large extent (8–11 cM) of conserved haplotypes around rare disease genes has also been described in the Afrikaners [Pronk et al., 1995; Roby et al., 1999].

In a previous study, supporting the appropriateness of the Afrikaner population for mapping complex traits, we have shown that the level of background LD in the Afrikaner sample is detectable at 6 cM [Hall et al., 2002], whereas it decays rapidly beyond 3 cM distances in the other founder and outbred populations examined, including the Finns, previously proven a valuable population for monogenic disease mapping. Here, we address two additional issues regarding this population, namely, the feasibility of detailed genealogical research and also the clinical similarity of schizophrenia diagnosed in this population and in more cosmopolitan populations.

We show that it is feasible in this population to trace each proband to the 16th century ancestors, an approach that promises to be very powerful when assessing alleles or chromosomal segments shared by descent. Initial results from our genealogical research are presented. Finally, we address the clinical comparability between Afrikaner schizophrenic patients and US schizophrenic patients and conclude that the cardinal symptoms of the disorder are stable across the two populations. This is important because while there is general agreement that using founder populations to map genes for complex traits may be meritorious, concern exists about the generalization of findings to the population at large, given the uncertainty about the accuracy of the diagnosis in disorders diagnosed exclusively as clinical syndromes, as all psychiatric disorders are at present.

METHODOLOGY

Sample Collection

Since 1997, in collaboration with clinicians from two of the major psychiatric hospitals in South Africa (Weskoppies Hospital in Pretoria and Valkenberg Hospital in Capetown), we ascertained and evaluated probands with a history of psychotic illness. Approval for the protocol was obtained from the Rockefeller IRB, as well as from the IRB committees at each site. All participants gave written informed consent.

At each site, an Afrikaans-speaking mental health worker (recruiter) identified individuals potentially ap-

propriate for enrollment in the study through patient referral from hospital staff, chart reviews in local hospitals, presentations to local support groups, and advertisements in local newspapers and magazines. After identifying a potential proband, the recruiters complete a brief psychiatric screen, explain the study, ascertain the patient's interest and motivation, determine the Afrikaner lineage and availability of both biological parents, obtain informed consent from the patient and parents, and record detailed pedigree information. A blood sample is then obtained (from proband and parents) and a clinical interview with the proband is scheduled. For probands who are severely affected or actively psychotic, an informant interview with a knowledgeable first-degree relative (usually a parent) is conducted, and a brief, unstructured interview is administered to the proband.

Our current sample includes 319 probands of Afrikaner descent (251 from Pretoria and 68 from Capetown), who meet diagnostic criteria for schizophrenia or schizoaffective disorder according to the American Psychiatric Association [1994]. Of the 319 probands, 71 are members of families with two or more affected individuals, while the rest are members of triads (affected individual and both biological parents).

Clinical Evaluation

The Diagnostic Instrument for Genetic Studies (DIGS) [Nurnberger et al., 1994], after it was translated and back translated into Afrikaans, was used for all diagnostic evaluations. Two senior psychiatrists (HP and JLR) from the Weskoppies Hospital in Pretoria and a senior Psychologist (SL) from the Valkenberg Hospital in Capetown, with a minimum of 10 years clinical experience, conducted all clinical interviews after they were trained in the use of the DIGS and the research application of DSM-IV. Reliability interviews with each clinical interviewer were conducted at the beginning of the study and the clinical validity of hospital chart diagnoses was also assessed. In order to ensure continued interviewer reliability and to reduce interviewer 'drift,' additional reliability interviews were conducted at regular intervals (once a year). In addition, each site videotaped two interviews per interviewer each year for the purposes of continuing quality control. Reliability of final diagnoses was found to be excellent at both sites. All clinical interviewers were fluent in both English and Afrikaans. The diagnostic interviews were conducted in Afrikaans. Following each interview, diagnosticians completed DSM-IV checklists for each category in which positive symptoms were identified in the DIGS; a narrative chronologic summary of prodromal traits, symptom onset, and functional impairment; and an English-language data summary form of selected DIGS items.

Genealogical Research

The psychiatric nurse-study recruiter at each site collected initial pedigree information from all probands who met study criteria and were enrolled in the study. The genealogical query includes questions about place and date of birth, place, and date of death for each great grandparent, as well as name and order of birth for all children in each generation. The proband and/or parents supplied the information. Typically, informants were able to provide pedigree information for three to four generations [up to great grandparents]. A systematic search was then conducted to identify the generations prior to the great grandparents. This search involved reviewing records in the State Archives, Church Archives, the Master's office, the Deeds office, the office of the Registrar of Births, Marriages, and Deaths, museums. The painstaking archival research continued until the pedigree was dated to the early 1800s. From that point, the Genealogies of Old South African Families [De Villiers, 1981], which record in detail the history of all families up until the initial 2,000 founders, are used to bring the pedigree back to its original founders. The genealogical information is then set out in the form of a family tree, or pedigree chart, and entered into a secure searchable File Maker Pro-Computerized database, where each person is assigned a unique numerical ID. This genealogical ID is different than the clinical ID and the DNA ID. Before tracing of an individual begins, the database is searched (by last name, first name, or both, and by year of birth) to confirm that the individual has not already been traced.

Predicting the Extent of Shared Haplotypes

We used simulation to predict the extent of the original founder haplotype that might surround a disease mutation carried by affected individuals in our sample. Carrying out parametric gene dropping simulations [Ploughman and Boehnke, 1989] that include a particular disease model, is impractical in a large pedigree with many loops such as the Afrikaner pedigree described here. Instead, we used a simple non-parametric procedure to study the fraction of a risk haplotype introduced by a single founder chromosome that might be transmitted to each affected individual.

In each simulation, we first labeled a single founder chromosome as the disease chromosome. We then iterated through the pedigree, examining each individual before any of his descendants. Each individual was then assigned either two normal chromosomes (if neither parent carried the disease chromosome) or one normal and one disease chromosome (if at least one parent carried a disease chromosome). Whenever a disease chromosome was transmitted, we sampled the location of flanking recombination events from an exponential distribution and tracked the erosion of the ancestral haplotype. Finally, we recorded the extent of the original ancestral haplotype carried by each affected individual in the pedigree.

This simulation represents a simplification of the processes by which disease alleles segregate through a pedigree and shared haplotypes are generated. Two simplifications are particularly important: (1) it assumes that all affected individuals share a particular founder allele and (2) it does not account for shared haplotype stretches, other than the original founder

chromosome. In the presence of genetic heterogeneity and phenocopies it is unlikely that all affected individuals will share a particular mutation, and subsets of affected individuals likely to share a particular founder mutation (because they are linked to the same locus or exhibit similar phenotypes) must be identified. Nevertheless, our simulation provides an accurate estimate of the length of the original founder chromosome that will flank the disease allele in each affected carrier. Since we do not account for sharing of haplotypes other than the original founder chromosome, this estimate serves as a lower bound for the extent of haplotype sharing among carriers of a shared disease allele.

RESULTS AND DISCUSSION

Clinical Comparison of the Afrikaner Schizophrenia Sample to a US Schizophrenia Sample

In the analyses reported here, we characterized the demographic and clinical features of a subset of our Afrikaner patients (N = 260) and assessed their clinical comparability to the manifestations of schizophrenia in North America by comparing them to our current US schizophrenia sample (N = 238). In both samples, only one affected individual per family (proband) was included. Affected family members were excluded from these analyses, as were probands who were unavailable for interview and thus diagnosed solely on the basis of medical records and/or family report. All US patients have completed in full the same DIGS interview administered to South African patients [Sobin et al., 2001]. Our results are reported in Table I.

Eleven demographic and clinical diagnostic characteristics were indistinguishable across both populations, including mean age at time of interview, gender distribution, percent of left-handed patients, percent of patients with a family history of schizophrenia, or other psychiatric diagnosis, percent of patients with co-morbid substance abuse or dependence, and the reported rates of delusions, auditory hallucinations, visual hallucinations, sensory hallucinations, disordered behavior, and alogia/flat affect. Thus with regard to basic sample descriptors and cardinal symptoms of disease, the two populations are equivalent.

Three demographic and four clinical variables were different between the Afrikaner and the US samples: (a) Afrikaner patients completed fewer years of education (12 vs. 14 years, $P \leq 0.01$); (b) fewer Afrikaner patients were single (i.e., never married) (74 vs. 83%, $P \leq 0.01$); (c) more Afrikaner patients were living with parents or care-taking family members (living dependently) (80 vs. 59%, $P \leq 0.01$); (d) mean onset age of DSM-IV schizophrenia or schizoaffective disorder was reported to be approximately 3 years later in the Afrikaner sample (23 vs. 20 years, $P \leq 0.01$); (e) rates of reported schizoaffective disorder were lower in the Afrikaner sample (26 vs. 46%, $P \leq 0.01$); (f) suicide attempts were reported to occur less frequently among Afrikaner patients with schizophrenia or schizoaffective disorder (28 vs. 52%, $P \leq 0.01$); and finally (g) rates of reported thought disorder were higher in the Afrikaner sample (68 vs. 50%, $P \leq 0.01$). The differences in level of education, marital status, and dependent living arrangements likely reflect cultural variation. For example, the rate of high school completion among the Afrikaner population overall is

TABLE I. Comparison of Demographic and Clinical Characteristics of Afrikaner and US Probands

	Afrikaners (N = 260)		US (N = 238)	
	Mean	SD	Mean	SD
Demographics				
Interview age	32.63	10.52	33.63	8.93
Years of education	11.87	2.06	13.60	2.18*
Female	31.54%	(82)	38.66%	(92)
Left-handed	9.77%	(21)	9.40%	(22)
Single (never married)	73.54%	(189)	83.41%	(191)*
Living dependently	80.16%	(206)	58.95%	(135)*
Family history/schizophrenia	33.33%	(86)	31.03%	(54)
Family history/other psychiatric diagnosis	68.99%	(178)	62.64%	(109)
Syndrome features				
Age DSM-IV schizophrenia onset	23.19	6.78	20.01	5.31*
Schizoaffective	25.58%	(66)	45.99%	(109)*
Comorbid subst. abuse/dependence	33.33%	(85)	36.13%	(86)
Attempted suicide	27.76%	(68)	52.34%	(123)*
Symptom history				
Delusions	97.26%	(248)	97.40%	(225)
Auditory hallucinations	79.17%	(190)	73.93%	(173)
Visual hallucinations	38.89%	(84)	46.36%	(102)
Sensory hallucinations	27.32%	(59)	35.71%	(80)
Disordered behavior	69.03%	(156)	63.32%	(145)
Thought disorder	67.87%	(150)	50.00%	(115)*
Alogia/flat affect	63.85%	(136)	68.12%	(156)

All parenthetical values indicate the value of 'N.'
*Fisher's exact test $P \leq 0.01$ or t-test $P \leq 0.01$.

lower than that of the US. Also, the increased percentages of patients living at home as adults may grossly reflect the greater proportion of nuclear intact families in South Africa.

On the other hand, differences in clinical variables may reflect a higher allelic homogeneity in this population, that predisposes to a subtype of the disease with fewer mood symptoms, higher rates of thought disorder and somewhat later onset. For example, recent analysis of families from the Irish Study of High-Density Schizophrenia Families (a more heterogeneous dataset) indicated that affected individuals from families with evidence for linkage to chromosome 8p had significantly more affective deterioration, poorer outcome, more thought disorder, and fewer depressive symptoms than affected individuals from the other families in the study [Kendler et al., 2000]. However, we cannot exclude currently the possibility that the differences found in clinical variables may be attributable to sample differences. Since diagnosticians for both the US and Afrikaner samples were trained and monitored identically by the same supervising clinician (CS), and used identical clinical instruments, it is unlikely that methodologic ascertainment bias accounts for the differences. It is possible that the populations themselves may differ. For example, the later onset age may be a secondary effect of the closer-knit family structure in the Afrikaner population. Unlike the US sample, Afrikaner probands tend to stay in the family home longer, eliminating the stress associated with leaving one's family of origin, and perhaps temporarily delaying the onset of symptoms. Similarly, the lower rates of schizoaffective disorder in the Afrikaner sample may reflect greater current severity of the patient population. The US sample was comprised almost entirely of patients who, at the time of their diagnostic interviews, were stabilized on medication, while the Afrikaner population included many more severely affected patients who were either chronically hospitalized or whose medication was only partially effective. A less well-stabilized sample of patients may be less able to report the details of past episodes that would provide the clinical substantiation for a diagnosis of schizoaffective disorder. Finally, the decreased rates of suicide may be the result of the closer supervised care provided to Afrikaner patients living with family members or in long-term institutionalized/hospital settings, an option that still exists in South Africa.

Genealogical Tracing in the Afrikaner Dataset

Genealogical approaches are best practiced in the context of a population that has been isolated over many centuries with little migration and where extensive genealogy is available and accessible to allow for tracking of the flow of genetic information. Information about genealogical relationships may decrease the locus heterogeneity for the disease and will significantly improve our ability to assess the evidence for increased allele sharing among affected individuals. Through detailed archival research (see Materials and Methods), we are able to trace genealogical relations back to the original founders. At present, we have traced 98 of our probands who meet diagnostic criteria for schizophrenia or schizoaffective disorder (~31% of the current Afrikaner sample). Remarkably, of the 98 traced affecteds, 87 descend from a single couple (Fig. 1).

In Table II, we list the maximum number of affected descendants for any one individual in each generation. It is of interest to note that we only observe individuals with a large number of affected descendants ten generations ago from present, indicating relatively distant connections between the individual pedigrees. Importantly, this does not pose a problem as the early portion of the genealogy is extremely well-documented in multiple Archival books. Currently, the largest connected portion of the pedigree includes 8,935 individuals in 17 generations. Of the 2,000 original settlers there are 644 represented in the master pedigree (making the fact that the majority of our probands link back to one settler and his wife even more impressive). There are 1,414 inbred individuals. The average coefficient of inbreeding for the pedigree is 0.00389. Among inbred individuals this rises to 0.0246. The maximum inbreeding coefficient for an individual is 0.15625. There are >2,500 loops within the pedigree, the majority of which are marriage loops. The majority of the consanguineous relationships occur in the early generations. To evaluate the completeness of the genealogy, we calculated the number of individuals and the proportion of founders in each generation. As seen in Table III, the genealogy seems complete, since we have identified the parents for the vast majority of individuals in generations 1–11 from present. One notable exception is the 3rd generation from present, where the fraction of individuals with no known ancestors is 44.4%, a dramatic increase compared to earlier and later generations. This reflects the history of the population, which in the period between 1900 and 1950 (coinciding with 3rd generation from present) experienced a dramatic movement following the discovery of minerals. The ensuing economic growth brought about a breakdown of tradition resulting to changes of spelling of surnames from Afrikaans to more anglicized versions, as well as to changes in the naming system of children, which up until then followed a very strict and predictable pattern, allowing inferences about lineage. These changes pose serious difficulties in the tracing of people in the 3rd generation from present.

We expect that a substantial fraction of the affected individuals in our pedigree may share risk alleles introduced by one or a few founders. We used simulation to estimate the length of the haplotype surrounding each of these alleles. Two individuals in our current extended pedigree have a total of 87 affected descendants, using strict diagnostic criteria (schizophrenia or schizoaffective disorder, depressed type), or 121 descendants, using the broadest classification criteria (any psychiatric disorder). We generated 1,000 pedigrees where a risk allele introduced by one of these individuals was transmitted to all affected descendants. In each case, we recorded the proportion of the original founder haplotype carried by each affected descendant. On average, there are 12.5 (range 10–17) intervening generations between this founder couple and their affected descendants so that

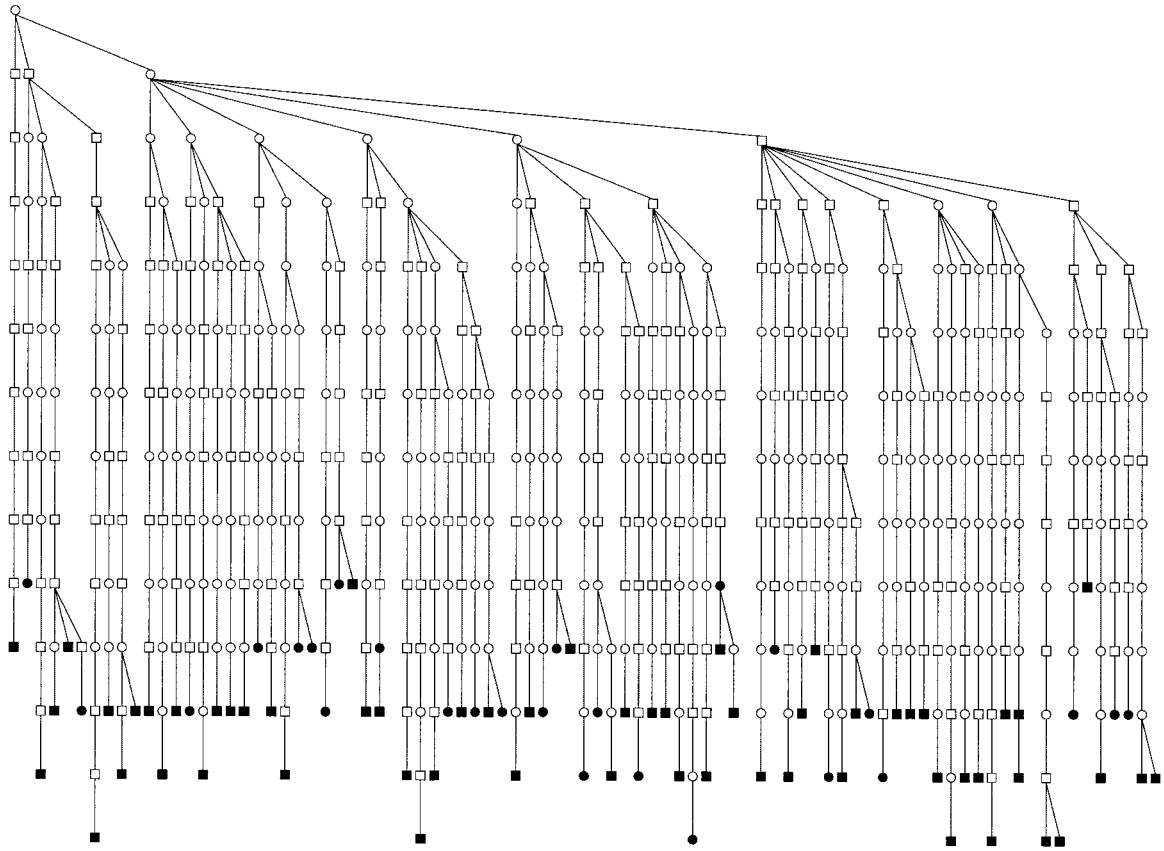

Fig. 1. Eighty seven of the affected individuals in our sample descend from a single couple, who lived 16 generations from present. Typically, each affected individual is connected to this couple through multiple lines of descent. The figure summarizes the shortest path (smallest number of meioses) between this couple and each affected descendant in our sample. On average, there are four alternative lines of descent connecting this founder couple and each affected individual. When multiple alternative lines of the same length were present, a single line was selected at random. Filled squares (males) and circles (females) represent affected individuals in our sample.

TABLE II. Maximum Number of Affected Descendants for a Single Individual in Each Generation

Generations from present[a]	Maximum affected descendants (using strict diagnostic criteria)
1	2
2	3
3	3
4	3
5	3
6	4
7	8
8	8
9	16
10	26
11	39
12	52
13	63
14	78
15	83
16	87

[a]Our genealogical database includes overlapping generations. For the purpose of this table individuals in generation 0 are those with no descendants in the pedigree. For other individuals, "generations from present" is defined as the maximum number of meioses required for connecting each individual to a descendant in generation zero.

TABLE III. Completeness of the Sample

Generations from present	People[a]	Founders[b]	Fraction (%)[c]
0	268	0	0.0
1	287	19	6.6
2	471	109	23.1
3	662	294	44.4
4	680	154	22.6
5	904	138	15.3
6	1,131	112	9.9
7	1,251	139	11.1
8	1,098	137	12.5
9	834	149	17.9
10	583	168	28.8
11	374	173	46.3
12	202	134	66.3
13	92	53	57.6
14	60	42	70.0
15	32	29	90.6
16	6	6	100.0

[a]Total number of individuals in each generation from present (people).
[b]The number of individuals who have no other known ancestors in the pedigree (founders).
[c]The proportion of all individuals who are founders (fraction).

disease alleles in each affected descendant should be flanked by an ~8 cM (100 cM/12.5) stretch of founder chromosome [Boehnke, 1994]. The results of Boehnke [1994] assumed independent lines of descent between the original founder and each affected descendant. In contrast, the affected individuals in our sample share a significant proportion of their genealogy and a single recombination event can affect the extent of the original haplotype present in a large number of individuals, generating great variability between different replicates of our pedigree (see Fig. 2a,b). Despite this, our simulations are in good agreement with the results of Boehnke [1994] and show that, on average, disease alleles are flanked by 8.8 cM (±8.3 cM, standard deviation) of founder chromosome, although this value is inflated by a small number of descendants carrying very long stretches of the original chromosome (>40 cM).

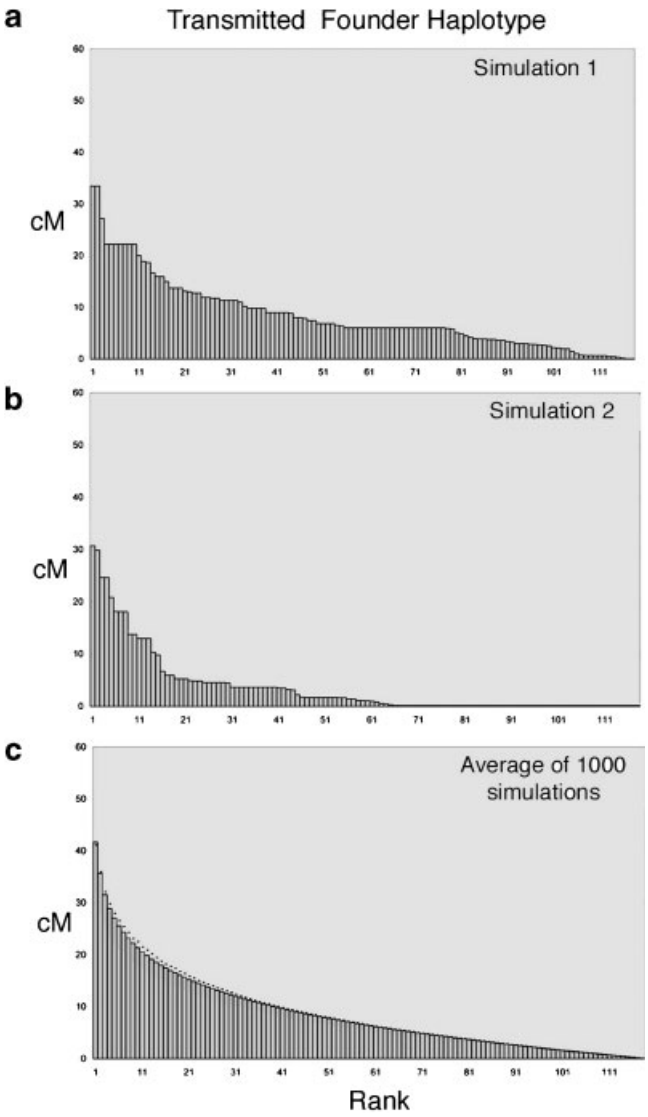

Fig. 2. The histogram illustrates the extent of the original founder chromosome flanking the shared disease allele in each affected individual. **Panels** (**a**) and (**b**) are the results of two representative replicates. Note how there are a number of discrete steps in the extent of the shared haplotype in each simulation. Each of these steps results from truncation of the ancestral haplotype by a single recombination event. **Panel** (**c**) summarizes the results of our 1,000 simulations. Filled bars represent the average extent of the original founder haplotype. Squares and lines represent the median and inter-quartile range.

On average (Fig. 2c), we found that the disease allele was flanked by ~6.4 cM or more of the original founder chromosome in half (60 of 121) of all affected individuals and by ~2.2 cM or more of the same chromosome in 80% of all affected individuals (96 of 121). We note that the shared genealogy between these individuals means that a small number of recombination events in key meioses can dramatically reduce the proportion of the founder haplotype transmitted to affected individuals (e.g., Fig. 2b). Nevertheless, in 95% of replicates the disease allele was flanked by >2.9 cM or more of the original founder chromosome in half of all affected carriers. In summary, our results suggest that shared disease alleles are likely to be flanked by >2 cM of the original founder chromosome in this pedigree. We found similar results for narrower phenotype definitions and for a small number of other founders with large numbers of affected descendants. These results suggest that approaches designed to search for haplotypes shared among affected individuals are feasible in this population using maps of microsatellite markers at 1–2 cM spacing.

CONCLUSIONS

The first set of results we present here is a clinical comparison between schizophrenia cases identified and diagnosed in the Afrikaner population and a sample of schizophrenia cases identified and diagnosed in the US. This comparison suggests that our genetic findings in the Afrikaner study will be applicable to other populations and schizophrenia in general. The diagnostic evaluations at the two sites were carried out by independent groups of clinicians specially trained in administering the DIGS. Eleven demographic and clinical diagnostic characteristics were stable across both populations and, with regard to basic sample descriptors and cardinal symptoms of disease, the two populations are equivalent. Three demographic and four clinical variables were different. It is not known whether these differences reflect genetic or cultural features of the Afrikaner sample.

The second set of results summarizes the genealogical research we have been able to perform for a large (and increasing) number of the schizophrenia families recruited to date. The consistency of our results suggests that it is feasible to trace genealogical relations with accuracy in the Afrikaner population. In addition, our genealogical data supports the hypothesis that a substantial portion of the schizophrenia cases we have identified will share the same or a small number of susceptibility alleles, since 87 affecteds traced so far link back to one founder couple. Establishment of detailed genealogical relationships between affecteds will allow us to cluster together patients who are more likely to share the genetic basis of the disease and search for segments of DNA shared IBD on a genome-wide level.

In principle, we expect to undertake this search in two stages. In a first stage, a map of widely spaced microsatellite markers (10–20 cM) should identify regions of the genome shared between close relatives, such as siblings. Statistical methods for evaluating linkage findings in small pedigrees are now mature and well-developed [e.g., Abecasis et al., 2002]. In a second stage, a search for shared founder haplotypes in genomic regions previously linked to schizophrenia, should allow us to hone in on disease alleles. Since disease alleles are likely to be flanked by >2 cM of the original founder haplotype, a moderate panel of microsatellite markers (at 1–2 cM spacing) should allow us to identify any founder haplotypes that are shared between more distant relatives. Statistical methods for comprehensively evaluating evidence for haplotype sharing are less well-developed and evaluating any evidence for a shared ancestral haplotype among affected individuals is likely to require an empirical standard, such as the extent of shared haplotypes in parental chromosomes not transmitted to affected individuals.

The number of meioses, patients, and segregating disease alleles in a pedigree are related to mapping power and resolution. In appropriate populations, the genealogical approach may represent an effective way of calibrating these quantities and a very powerful method for dissection of complex traits provided that appropriate statistical tests designed to harness this power are developed.

ACKNOWLEDGMENTS

The authors thank all individuals who have participated in this research. They also gratefully acknowledge nursing sisters Ria van Wyk, Cheryl Botha, and Glenda Jansen for assistance with family identification and recruitment, as well as Christina Seubert and Samantha Brenner for assistance with genealogical research and data entering.

REFERENCES

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101.

American Psychiatric Association. 1994. Diagnostic and statistical manual. 4th edition. Washington, DC: American Psychiatric Association.

Boehnke M. 1994. Limits of resolution of genetic linkage studies, implications for the positional cloning of human disease genes. Am J Hum Genet 55:379–390.

Brink PA, Steyn LT, Coetzee GA, Van der Westhuyzen DR. 1987. Familial hypercholesterolemia in South African Afrikaners. PvuII and StuI DNA polymorphisms in the LDL-receptor gene consistent with a predominating founder gene effect. Hum Genet 77:32–35.

De Villiers CC. 1981. Genealogies of old South African families. Completely revised edition augmented and rewritten by Pama C. Rotterdam: A.A. Balkema.

Defesche JC, van Diermen DE, Lansberg PJ, Lamping RJ, Reymer PW, Hayden MR, Kastelein JJ. 1993. South African founder mutations in the low-density lipoprotein receptor gene causing familiar hypercholeserolemia in the Dutch population. Hum Genet 92:567–570.

Defesche JC, Van Diermen DE, Hayden MR, Kastelein JP. 1996. Origin and migration of an Afrikaner founder mutation FHAfrikaner-2. V408M. causing familial hypercholesterolemia. Gene Geogr 10:1–10.

DeLisi LE, Mesen A, Rodriguez C, Bertheau A, LaPrade B, Llach M, Riondet S, Razi K, Relja M, Byerley W, Sherrington R. 2002. Genome-wide scan for linkage to schizophrenia in a Spanish-origin cohort from Costa Rica. Am J Med Genet 114:497–508.

Devlin B, Bacanu SA, Roeder K, Reimherr F, Wender P, Galke B, Novasad D, Chu A, Cveno KT, Tiobek S, Otto C, Byerley W. 2002. Genome-wide multipoint linkage analyses of multiplex schizophrenia pedigrees from the oceanic nation of Palau. Mol Psychiatry 7:689–694.

Goldman A, Krause A, Ramsay M, Jenkins T. 1996. Founder effect and the prevalence of myotonic dystrophy in South Africans, Molecular studies. Am J Hum Genet 59:445–452.

Groenewald JZ, Liebenberg J, Groenewald IM, Warnich L. 1998. Linkage disequilibrium analysis in a recently founded population, evaluation of the variegate porphyria founder in South African Afrikaners. Am J Hum Genet 62:1254–1258.

Hall D, Wijsman EM, Roos JL, Gogos JA, Karayiorgou M. 2002. Extended intermarker linkage disequilibrium in the Afrikaners. Genome Res 12:956–961.

Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, Coloma A, Lovett M, Buckler A, Kaitila I, Lander ES. 1994. The diastrophic dysplasia gene encodes a novel sulfate transporter, positional cloning by fine structure linkage disequilibrium mapping. Cell 78:1073–1087.

Hayden MR, MacGregor JM, Beighton PH. 1980a. The prevalence of Huntington's chorea in South Africa. S Afr Med J 58:193–196.

Hayden MR, Hopkins HC, Macrea M, Beighton PH. 1980b. The origin of Huntington's chorea in the Afrikaner population of South Africa. S Afr Med J 58:197–200.

Hovatta I, Varilo T, Suvisaari J, Terwilliger JD, Ollikainen V, Arajarvi R, Juvonen H, Kokko-Sahin ML, Vaisanen L, Mannila H, Lonnqvist J, Peltonen L. 1999. A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. Am J Med Genet 65:1114–1124.

Kalaydjieva L, Hallmayer J, Chandler D, Savov A, Nikolova A, Angelicheva D, King RHH, Ishpekova B, Honeyman K, Calafell F, Shmarov A, Petrova J, Turnev I, Hristova A, Moskov M, Stancheva S, Petkova I, Bittles AH, Georgieva V, Middleton L, Thomas PK. 1996. Gene mapping in Gypsies identifies a novel demyelinating neuropathy on chromosome 8q24. Nat Genet 14:214–217.

Kendler KS, Myers JM, O'Neill FA, Martin R, Murphy B, MacLean CJ, Walsh D, Straub RE. 2000. Clinical features of schizophrenia and linkage to chromosomes 5q, 6p, 8p, and 10p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry 157: 402–408.

Leitersdorf E, Van der Westhuyzen DR, Coetzee GA, Hobbs HH. 1989. Two common low density lipoprotein receptor gene mutations cause familial hypercholesterolemia in Afrikaners. J Clin Invest 84:954–961.

Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T, NIMH Genetics Initiative. 1994. Diagnostic interview for genetic studies, rationale, unique features, and training. Arch Gen Psychiatry 51:949.

Nystuen A, Benke PJ, Merren J, Stone EM, Sheffield VC. 1996. A cerebellar ataxia locus identified by DNA pooling to search for linkage disequilibrium in an isolated population from the Cayman Islands. Hum Mol Genet 5:525–531.

Paunio T, Ekelund J, Varilo T, Parker A, Hovatta I, Turunen JA, Rinard K, Foti A, Terwilliger JD, Juvonen H, Suvisaari J, Arajarvi R, Suokas J, Partonen T, Lonnqvist J, Meyer J, Peltonen L. 2001. Genome-wide scan in a nationwide study sample of schizophrenia families in Finland reveals susceptibility loci on chromosomes 2q and 5q. Hum Mol Genet 10:3037–3048.

Ploughman LM, Boehnke M. 1989. Estimating the power of a proposed linkage study for a complex genetic trait. Am J Hum Genet 44: 543–551.

Pronk JC, Gibson RA, Savoia A, Wijker M, Morgan NV, Melchionda S, Ford D, Temtamy S, Ortega JJ, Jansen S, Havenga C, Cohn RJ, de Ravel TJ, Roberts I, Westerveld A, Easton DF, Joenje H, Mathew CG, Arwert F. 1995. Localisation of the Fanconi anaemia complementation group A gene to chromosome 16q24.3. Nat Genet 11:338–340.

Roby P, Eyre S, Worthington J, Ramesar R, Cilliers H, Beighton P, Grant M, Wallis G. 1999. Autosomal dominant 'Beukes' premature degenerative osteoarthropathy of the hip joint maps to an 11-cM region on chromosome 4q35. Am J Hum Genet 64:904–908.

Rosendorff J, Bernstein R, Macdougall L, Jenkins T. 1987. Fanconi anaemia, another disease of high prevalence in the Afrikaans population of South Africa. Am J Med Genet 27:793–797.

Sobin C, Blundell ML, Conry A, Weiller F, Gavigan C, Haiman C, Karayiorgou M. 2001. Early, non-psychotic deviant behavior in schizophrenia: A possible endophenotypic marker for genetic studies. Psychiatry Res 101:101–113.

Visapaa I, Fellman V, Vesa J, Dasvarma A, Hutton JL, Kumar V, Payne GS, Makarow M, Van Coster R, Taylor RW, Turnbull DM, Suomalainen A, Peltonen L. 2002. Gracile syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. Am J Hum Genet 71:863–876.

Warnich L, Kotze MJ, Groenewald IM, Groenewald JZ, van Brakel MG, van Heerden CJ, de Villiers JN, van de Ven WJ, Schoenmakers EF, Taketani S, Retief AE. 1996. Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South Africa. Hum Mol Genet 5:981–984.

Wijsman EM, Rosenthal EA, Hall D, Blundell ML, Sobin C, Heath SC, Williams R, Brownstein MJ, Gogos JA, Karayiorgou M. 2003. Genome-wide scan in a large complex pedigree with predominantly male schizophrenics from the island of Kosrae: Evidence for linkage to chromosome 2q. Mol Psychiatry 8:695–705.