Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain

Karl W. Broman and James L. Weber

Marshfield Medical Research Foundation, Marshfield, WI

Summary

Using genotypes from nearly 8,000 short tandem-repeat polymorphisms typed in eight of the reference families from the Centre d'Étude du Polymorphisme Humain (CEPH), we identified numerous long chromosomal segments of marker homozygosity in many CEPH individuals. These segments are likely to represent autozygosity, the result of the mating of related individuals. Confidence that the complete segment is homozygous is gained only with markers of high density. The longest segment in the eight families spanned 77 cM and included 118 homozygous markers. All individuals in family 884 showed at least one segment of homozygosity: the father and mother were homozygous in 8 and 10 segments with an average length of 13 and 16 cM, respectively, and covering a total of 105 and 160 cM, respectively. The progeny in family 884 were homozygous over 5–16 segments with average length 11 cM. The progeny in family 102 were homozygous over 4–12 segments with average length 19 cM. Of the 100 individuals in the other six families, 1 had especially long homozygous segments, and 19 had short but significant homozygous segments. Our results indicate that long homozygous segments are common in humans and that these segments could have a substantial impact on gene mapping and health.

Introduction

The reference families from the Centre d'Étude du Polymorphisme Humain (CEPH) (White and Lalouel 1988; Dausset et al. 1990) were recruited in the effort to construct the first human genetic maps. Lymphoblastoid cell lines derived from the CEPH individuals have provided

a nearly limitless supply of DNA, making these families available for genotyping by investigators around the world. Many thousands of short tandem-repeat polymorphisms (STRPs) have been genotyped within a subset of eight of the CEPH families. These data provide a uniquely comprehensive view of the genomes of these individuals, which allows analyses that would not be possible on the basis of data from a more typical genome scan of 400 markers.

We recently constructed new genetic maps based on these families (Broman et al. 1998). As part of that work, we screened the data for apparent tight double-recombination events indicative of genotyping errors or mutations. In the process, we identified several long segments of noninformative markers in family 884, caused by long stretches of homozygous markers in the parents of that family.

These segments likely represent autozygosity. An individual is autozygous at a locus if he or she received two copies of a single ancestral allele at the locus—that is, if the two alleles are identical by descent (Hartl and Clark 1997). Autozygosity occurs when a couple shares a chromosomal segment identical by descent and both transmit the segment to one of their offspring.

Autozygosity may be detected at a single locus if genetic data are available on many individuals in an inbred pedigree. More generally, one needs to look at sets of linked markers: if an individual is homozygous at a large number of contiguous markers, it is likely that he or she is autozygous for the segment (i.e., that the two chromosomal segments have a common origin).

We developed a LOD score to characterize the significance of a segment of homozygosity, to distinguish autozygosity from chance homozygosity. We calculated this score for all possible subsets of contiguous autosomal markers in all 134 individuals from the eight CEPH families, in order to discover autozygous segments. All individuals in family 884 showed at least one segment of homozygosity, as did all of the progeny in family 102. Total lengths of homozygous segments in these individuals were 62–253 cM. Twenty percent of the individuals in the other six families also had significant homozygous segments. These unexpected findings

indicate that long homozygous segments are common in humans.

Material And Methods

Genetic Markers and Genotype Data

We used the data described by Broman et al. (1998); we summarize them again here. We considered eight of the CEPH families (102, 884, 1331, 1332, 1347, 1362, 1413, and 1416), excluding individuals 1332-09 and 1416-10, on whom few genotype data were available. These families contain a total of 134 individuals, including 92 progeny, 16 parents, and 26 grandparents. We considered genotypes for 7,746 autosomal STRPs, including 5,064 from Généthon (Dib et al. 1996), 1,334 from CHLC (Sheffield et al. 1995; Sunden et al. 1996), 849 from the Utah Marker Development Group (1995), 285 from the Center for Medical Genetics, Marshfield Medical Research Foundation, 35 telomeric markers (Rosenberg et al. 1997), and 179 miscellaneous markers. The genotypes for families 884, 1331, 1332, and 1362 were 96% complete. Families 102, 1347, 1413, and 1416 were not typed for the 849 Utah markers, but their genotypes were 94% complete for the other 6,897 markers. All the data are available at the Center for Medical Genetics, Marshfield Medical Research Foundation Web site.

The data had previously been cleaned of genotypes causing apparent tight double-recombination events indicative of genotyping errors or mutations. The genetic maps and marker order were as determined by Broman et al. (1998). The total sex-averaged genetic length of the 22 autosomes was 3,488 cM, corresponding to approximately one marker per 0.5 cM. In the following, all reported genetic distances are sex averaged. Allele frequencies were estimated on the basis of the 28 founding individuals in the eight families. The average (SD) of the marker heterozygosities was .68 (.13). Cytogenetic locations of the markers, from Collins et al. (1996), were obtained from the Genetic Location Database at the University of Southampton.

Identification of Long Homozygous Segments

Our goal was to identify subsets of contiguous markers for which an individual showed an unusual proportion of homozygosity. For each individual, we looked at all possible subsets of contiguous markers. For each such subset, we calculated a LOD score comparing the hypotheses that the individual was or was not autozygous (i.e., homozygous by descent) at all markers in the subset. Large values of the statistic indicate segments in which the individual showed a significant amount of homozygosity.

In forming the LOD score, we assumed that the markers were in both linkage and Hardy-Weinberg equilib-

rium, and we used a simple model for genotyping errors and mutations, much like that used by Broman and Weber (1998) in the context of the inference of pairwise relationships. The LOD score has the following form:

$$\text{LOD}(j,k) = \sum_{i=j}^{k} \log_{10}\left[\frac{\Pr(g_i|\text{autozygous at } i)}{\Pr(g_i|\text{not autozygous at } i)}\right] ,$$

where g_i is the individual's observed genotype at marker i. The assumed genotype probabilities are displayed in table 1. In the model underlying the LOD score, it is assumed that errors and mutations occur with probability ε and that, when such events occur, genotypes are obtained according to their population frequencies. This model serves as a device for obtaining a statistic that facilitates the detection of homozygous segments while both taking account of the varying informativeness of the markers and allowing the presence of a small proportion of heterozygous markers within each segment; it was not intended to reflect the details of the error and mutation processes.

In calculating the LOD score, we assumed that the combined rate of genotyping errors and mutations was $\varepsilon = .001$. A smaller value of ε would give greater weight to heterozygous markers, resulting in shorter identified segments. If ε were made larger (smaller) by a factor of 5, the LOD score for a segment would be increased (decreased) by 0.7 for each heterozygous marker that it contained; for example, if a segment containing a single heterozygous marker had a LOD score of 7.9 when $\varepsilon = .001$, the LOD score would be ~7.2 and ~8.6 when $\varepsilon = .0002$ and .005, respectively.

The above-described LOD score was calculated for each possible subset of contiguous markers in each individual. When, for a particular individual, two subsets of markers with large LOD scores overlapped, we retained information only on that subset with the larger LOD score.

Simulations under Linkage Equilibrium

Even under complete linkage equilibrium, individuals may be homozygous for a number of contiguous mark-

Table 1

Probabilities of Observed Genotype at a Marker, Given the Autozygosity Status of the Surrounding Segment

OBSERVED GENOTYPE	PROBABILITY THAT SEGMENT IS		RATIO
	Autozygous	Not Autozygous	
AA	$(1-\varepsilon)p_A + \varepsilon p_A^2$	p_A^2	$(1-\varepsilon)/p_A + \varepsilon$
AB	$2\varepsilon p_A p_B$	$2p_A p_B$	ε

NOTE.—ε denotes the combined rate of genotyping errors and mutations.
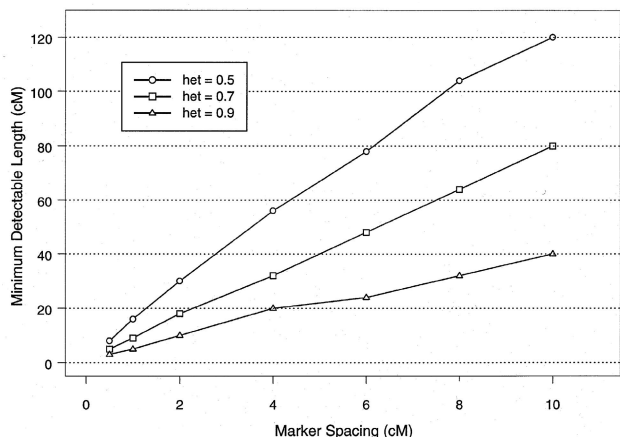
Figure 1 Minimum detectable length (in cM) of an autozygous region for a genome of length 3,488 cM, with equally spaced markers of constant heterozygosity.

ers somewhere in the genome, completely by chance. In order to estimate the distribution of the above-described LOD score under linkage equilibrium, we simulated the genetic data for $134 \times 20 = 2,680$ unrelated individuals having the same pattern of missing data as were seen in the 134 CEPH individuals whom we studied. We assumed Hardy-Weinberg and linkage equilibrium and used the allele frequencies estimated on the basis of the 28 founding individuals in the eight families. For each simulated genome, we calculated the maximum LOD score, among all possible subsets of contiguous markers, and the length of the subset with the maximum LOD score.

To determine the effects of marker density and marker informativeness on the minimum detectable length of an autozygous segment, we simulated genomes that were of length 3,488 cM and had markers that were equally spaced, had constant heterozygosity, and were in linkage equilibrium. To simplify the simulation, we assumed that the genome consisted of a single long chromosome. For each of various marker densities and heterozygosities, we simulated the genetic data for 10,000 genomes and calculated the 95th percentile of the maximum length

of contiguous homozygous markers. This percentile is the minimum detectable length of an autozygous segment when a 5% genomewide type I–error rate is used.

Results

The minimum detectable length of an autozygous segment, as a function of marker density, in the case of equally spaced markers with constant heterozygosity, is displayed in figure 1. For markers with .7 heterozygosity, an autozygous segment as short as 9 cM may be detected when the markers are 1 cM apart, whereas the segment must be ≥32 cM long if there are markers every 4 cM. At the density of a more typical genome scan (i.e., 10 cM), only very long autozygous segments can be detected.

The simulations under linkage equilibrium that made use of the genetic maps and allele frequencies estimated from the data indicated that, ~5% of the time, an individual will have a segment of at least nine contiguous homozygous markers, with a LOD score >4.67, somewhere in the genome. In the following, we present only those segments that had LOD scores >4.67.

All individuals in family 884 had at least one homozygous segment with a LOD score >4.67. The father, individual 884-01, had eight homozygous segments, which are summarized in table 2. LOD scores for these homozygous segments were as high as 27.85. One segment was ~2 cM long; the others were 10–23 cM long. Overall, these segments covered a total of 105 cM (3.0% of the autosomal genome). The mother, individual 884–02, had 10 homozygous segments, which are summarized in table 3. Five segments were <10 cM long; three were 30–40 cM long. Overall, these segments covered a total of 160 cM (4.6% of the autosomal genome).

Table 4 contains a summary of the homozygous segments in the other members of family 884. Three of the four grandparents in the family had multiple long homozygous segments; the other grandparent showed just one homozygous segment of 11 markers. The 12 progeny in the family showed 5–16 homozygous segments with average length 11.0 cM and covering, on average, 97 cM (2.8% of the autosomal genome). These homo-

Table 2

Homozygous Segments for Individual 884-01

Chromosome (Markers)	Cytogenetic Band(s)	Length (cM)	Proportion Homozygous	LOD Score
4 (D4S3021–D4S1597)	q28-q31	14.8	49/50	26.58
6 (D6S1017–D6S1623)	p21-p11	14.5	36/36	21.74
6 (D6S1631–D6S1580)	q16-q22	13.2	49/49	27.85
7 (D7S2486–D7S800)	q31-q32	12.8	38/38	20.18
11 (DllS1794–DllS4138)	p15	1.6	15/16	5.46
12 (D12S354–D12S386)	q24	23.4	36/38	17.08
19 (D19S605–D19S890)	q13	9.8	11/11	5.37
21 (D21S1891–D21S1897)	q22	15.0	20/20	12.78

Table 3

Homozygous Segments for Individual 884-02

Chromosome (Markers)	Cytogenetic Band(s)	Length (cM)	Proportion Homozygous	LOD Score
3 (D3S1571–D3S1617)	q28	4.9	9/9	5.53
4 (GATA144E02–D4S189)	p11-q12	11.1	21/21	12.26
5 (D5S398–D5S401)	q11-q14	29.8	77/77	46.21
6 (D6S1711–D6S278)	q11-q22	35.3	109/113	48.12
8 (D8S506–D8S385)	q22-q23	8.0	28/30	12.35
9 (D9S1802–D9S250)	q33	6.5	18/18	9.53
12 (D12S103–D12S1680)	q13-q21	11.3	43/43	21.82
16 (D16S494–D16S3107)	q21-q22	8.8	26/26	17.23
16 (D18S450–GATA51E05)	q21-q22	40.3	84/84	49.79
22 (D22S1156–D22Sl179)	q13	3.9	21/21	15.81

zygous segments derived from shared haplotypes in the parents. The grandpaternal chromosome in the father matched the grandmaternal chromosome in the mother in 12 segments covering 130 cM (3.7% of the autosomal genome); the grandmaternal chromosome in the father matched the grandpaternal chromosome in the mother in 5 segments covering 82 cM (2.4% of the autosomal genome); the grandpaternal chromosomes in the father and mother were identical in 9 chromosomal segments covering 87 cM (2.5% of the autosomal genome); and the grandmaternal chromosomes in the father and mother were identical in 6 segments covering 65 cM (1.9% of the autosomal genome).

The 14 progeny in family 102 showed 4–12 homozygous segments with average length 18.5 cM and covering, on average, 155 cM (4.4% of the autosomal genome) (table 5). Neither parent in family 102 showed any significant homozygous segments; genotype data on the grandparents in this family were not available.

A 31-cM segment of chromosome 4, containing the 89 markers from D4S1604 to D4S3030, is of special interest with regard to family 102. All 14 progeny were homozygous for at least one portion of this segment. Child 102-16 was homozygous for 12 markers in this segment, spanning 7 cM; children 102-10 and 102-12 were homozygous for the entire 31-cM segment. The other 11 progeny were homozygous for a segment of intermediate size. An inspection of the inferred haplotypes in the two parents shows that they share a common haplotype on one pair of their chromosomes over this entire segment; on their other chromosome pair, they share a common haplotype for an 18-cM segment containing 44 markers. The homozygosity in 8 of the 14 progeny derives from the longer haplotype shared by the two parents; the homozygosity in the other 6 progeny derives from the other shared haplotype.

Of the 100 individuals in the other six families, 20 had at least one homozygous segment with LOD score >4.67. Individual 1416-14 (a grandparent) showed four homozygous segments, including three segments with

LOD scores >15, containing >25 markers and spanning 19–29 cM. Nine individuals were homozygous for segments <2 cM long, including one individual homozygous for two short but significant segments. Ten other individuals were homozygous for segments 3–7 cM long. One of these contained 27 markers; the others contained 6–11 markers. A complete list of significant homozygous segments for all individuals in the eight CEPH families is available with the electronic version of this article, at the *Journal*'s web site.

Many of the identified homozygous segments contained one or more heterozygous markers. These heterozygous markers could be due to errors in marker order, genotyping errors, mutations, or gene conversions. Of the 286 homozygous segments identified, 41 contained one heterozygous marker, and 8 contained two to four heterozygous markers. The other 237 seg-

Table 4

Homozygous Segments for Progeny and Grandparents in Family 884

Individual	No. of Segments	Length (cM)	No. of Markers	Total (Average) Length
Progeny:				
3	9	2.8–39.1	14–93	111.5 (12.4)
4	5	2.8–25.8	15–77	61.7 (12.3)
5	5	11.7–15.7	14–58	65.7 (13.1)
6	9	3.2–17.9	9–64	93.1 (10.3)
7	9	3.8–17.9	13–39	84.0 (9.3)
8	16	1.7–36.2	9–87	195.8 (12.2)
9	8	4.2–19.5	9–71	77.6 (9.7)
10	6	5.9–12.9	13–35	57.0 (9.5)
11	10	2.1–36.2	15–90	124.8 (12.5)
12	7	1.7–16.5	7–32	66.4 (9.5)
13	9	4.2–30.0	9–64	118.0 (13.1)
14	13	2.7–15.9	14–45	109.7 (8.4)
Grandparents:				
15	4	4.8–16.0	6–45	35.5 (8.9)
16	1	6.5	11	6.5 (6.5)
17	7	1.6–17.2	8–45	57.4 (8.2)
18	9	1.7–29.3	10–51	116.2 (12.9)

Table 5

Homozygous Segments for Progeny in Family 102

Individual	No. of Segments	Length (cM)	No. of Markers	Total (Average) Length
3	10	4.1–23.1	11–69	157.9 (15.8)
4	6	8.2–28.2	9–70	105.9 (17.7)
5	10	7.5–38.0	9–61	198.4 (19.8)
6	6	3.3–27.0	9–55	98.6 (16.4)
7	4	8.7–74.1	27–118	123.5 (30.9)
8	12	5.1–38.4	8–68	200.7 (16.7)
9	12	2.5–35.7	16–95	253.1 (21.0)
10	9	8.2–38.6	10–87	200.7 (22.3)
11	9	3.9–38.4	7–60	158.3 (17.6)
12	5	7.6–31.4	19–76	106.9 (21.4)
13	11	5.2–27.2	7–77	175.1 (15.9)
14	7	1.7–31.8	10–71	128.7 (18.4)
15	12	3.2–32.9	8–69	183.5 (15.3)
16	4	9.0–37.5	17–69	77.7 (19.4)

ments contained no heterozygous markers. Among the eight segments with two or more heterozygous markers, there were three cases in which two heterozygous markers were adjacent to each other, separated by 1 cM. (These occurred in a segment on chromosome 13 in progeny from family 102.) In all other cases, the heterozygous markers were relatively isolated from each other.

The lengths of the homozygous segments that we identified, as well as the proportion of the genome that they cover, are likely to be underestimates. The markers were ordered on the basis of the 184 meioses in these families, and so the order is generally correct only to ~1–3 cM. At the ends of a homozygous segment, the presence of a few heterozygous markers, caused by mutation or incorrect marker order, may cause an early truncation of the identified segment. When one assumes that $\varepsilon = .001$, a heterozygous genotype subtracts 3.0 from the LOD score, whereas homozygous genotypes for alleles with frequencies .1 and .25 add 1.0 and 0.6, respectively, to the LOD score. Thus, for a homozygous segment to be extended past a heterozygous genotype, one must see three to five additional homozygous genotypes. Of the 18 homozygous segments described in tables 2 and 3, 2 contained a single heterozygous genotype, 2 contained two heterozygous genotypes, and 1 contained four heterozygous genotypes; in all of these cases, there were at least five homozygous genotypes between the heterozygous marker and the end of the identified segment.

It is likely that a number of autozygous segments were missed, since the segments need to be of a sufficient size, containing a sufficient number of markers, to allow detection (see fig. 1). Although the marker data that we considered were of unusually high density, marker density varied considerably across the genome, with some regions containing 15 markers in <1 cM and others containing gaps as long as 8 cM with no markers. Marker heterozygosity varied considerably as well. Although the average heterozygosity was .68, 43% of the markers had heterozygosities ≤.60 or ≥.80.

Discussion

Autozygosity occurs when two copies of an ancestral haplotype come together in an individual. This may be the result of the mating of close relatives or of linkage disequilibrium in a population, but these are just the two extremes in a continuum of ancestral sharing. It is customary to distinguish between linkage disequilibrium and inbreeding: homozygous segments that are the result of linkage disequilibrium would not ordinarily be called autozygous. However, insofar as linkage disequilibrium is the result of the persistence of ancestral haplotypes in a finite population, homozygosity as the result of linkage disequilibrium is indeed the result of the mating of (very distantly) related individuals. Linkage disequilibrium is a local phenomenon and thus would cause only short homozygous segments. Long homozygous segments, such as that on chromosome 18 in individual 884-02, which is 40 cM in length and contains 84 markers, cannot be explained by linkage disequilibrium.

The length of an autozygous segment reflects its age: haplotypes are broken up by recombination at meiosis, and so a short autozygous region is likely to be of distant origin. (Haplotypes can also be brought together by recombination; if such an event were to occur twice independently, the result could be an autozygous segment composed of portions derived from two distinct ancestors.) The proportion of an individual's genome that is autozygous, the expected value of which is the inbreeding coefficient (Hartl and Clark 1997), is a measure of the degree of relationship between his or her parents.

Family 102 was of Venezuelan origin. The long segments of homozygosity in the progeny of this family, covering on average 4.4% of the autosomal genome, indicates that the parents were relatively closely related. Six of the 14 progeny had homozygous segments covering >5% of their genome. Although neither parent showed any inbreeding, the parental haplotypes corresponding to the homozygous segments on chromosome 4 imply that the homozygosity in this family is the result of relatedness between at least two pairs of the four grandparents. For example, the grandparents could have been composed of two sets of first cousins, making the parents in the family double–second cousins. If that were true, then one would expect the progeny in the family to be autozygous at an average of 3.1% of the autosomal genome.

Family 884 is from the Old Order Amish in Pennsylvania. The progeny in this family were autozygous at a smaller proportion of the genome (2.8%), and the

autozygous segments were somewhat smaller, than those in family 102. The parents and grandparents in this family, however, also showed a large proportion of autozygosity. The grandparental origins of the haplotypes shared between the two parents indicates that each of the four grandparents was related to each of the others. The genealogy of this family (Egeland 1972) indicates that the grandparents have no ancestors in common in the three prior generations. For example, if one looks at the five generations preceding father 884-01, he is indicated to be the child of third cousins: one of his father's great-grandmothers is the sister of one of his mother's great-grandfathers. The parents of mother 884-02 show a similar relationship. The child of third cousins would be expected to be autozygous at, on average, 0.4% of the genome. Individuals 884-01 and 884-02, however, were autozygous at 3.0% and 4.6% of their genomes, respectively. Similarly, this part of the genealogy, on its own, would imply an average of 0.7% autozygosity in the progeny, one-fourth the observed average amount of autozygosity in the 12 progeny. Thus the homozygosity in family 884 cannot be explained by the inbreeding in four generations preceding the family's grandparents. Prior generations likely showed much more inbreeding, so that the observed autozygosity was the result of drawing from a limited pool of DNA five or more generations back in time.

The presence of long homozygous segments in the parents of family 884 make this family a less than ideal choice as a reference for genetic maps. Future analyses of the CEPH data should treat this family with special care. For example, Bugge et al. (1998) used the CEPH genetic data to calculate crossover counts for each autosome. Family 884 should have been excluded from those counts for chromosomes for which a parent was homozygous—and, therefore, not informative—for a large portion of the chromosome, since several double crossovers were likely missed.

Of the 100 individuals in the other six CEPH families, all from Utah, 1 individual, grandparent 1416-14, had one homozygous segment 6 cM long and three segments 19–25 cM long, the four segments together covering 2.2% of the autosomal genome, indicating a close relationship between her parents; 19 of the other individuals in these families had quite small homozygous segments. These six families clearly do not display the sort of inbreeding observed in families 102 and 884, but the results cannot reasonably be ascribed to chance homozygosity; rather, we conclude that the families were subject to a small amount of inbreeding, at least in the past, and that the markers studied show considerable linkage disequilibrium, a modern trace of older relationships in the population from which these families were drawn.

The long homozygous segments provide opportunity for study of mutation. Heterozygous markers within these homozygous regions are likely to be the result of a mutation event in one of the generations between the last common ancestor and the autozygous individual. An example is marker GATA44 within the long homozygous region on chromosome 6 in 884-02 (genotype 152,148). These heterozygous markers will provide valuable information about mutation rate—and, when combined with data on the relevant haplotype or allele frequencies, about the nature of the nucleotide change.

The presence of long homozygous regions has at least two implications for gene mapping in genetically isolated populations. On the positive side, it may be possible to extend the widely used homozygosity (autozygosity) mapping for rare recessive disorders (Smith 1953; Lander and Botstein 1987), to include genetically more complex disorders. The presence of two copies of a relatively common allele often increases disease risk compared with the presence of a single copy (e.g., see Farrer et al. 1997; Rosendaal 1997). By searching for shared homozygous regions among individuals from an isolated population with extreme phenotypes, it may be possible to map responsible genes. Furthermore, a modification of the LOD score described here could be used to identify haplotypes shared by distantly related diseased individuals. In this case, one would search for segments of many contiguous markers for which the pair shared at least one allele identical by state.

On the negative side, if a large sibship contributes substantially to the power of a gene-mapping study and one (or both) of the parents is homozygous at a major gene locus, then evidence in support of linkage may be reduced below the detection threshold. In this respect, it is interesting to note that indications of a gene, on chromosome 18, predisposing to bipolar affective disorder have been reported in some studies (Berrettini et al. 1994; Stine et al. 1995; Freimer et al. 1996; McMahon et al. 1997)—but not in the Old Order Amish (Pauls et al. 1995; Polymeropoulos and Schaffer 1996; Ginns et al. 1996, 1998)—and that the mother of Amish family 884 (a family with bipolar affective disorder) is homozygous for a long segment of chromosome 18.

Although it is impossible to accurately extrapolate from data on only eight families, we suspect that long homozygous segments are common in many human populations. The Old Order Amish are a clearly recognized genetic isolate, but neither the Venezuelan kindred, 102, nor the remaining six, Utah Mormon, kindreds derived from any obvious genetic isolates. Furthermore, as described above, the grandparents in Amish family 884 were not particularly closely related, and many human populations are probably as isolated genetically as the Amish. The total lengths of autozygous segments within individuals from a specific population may provide a useful measure of the isolation of that population.

Clinical implications of the long homozygous seg-

ments are substantial. First-cousin matings will produce, on average, autozygosity of 6%. Fitness and health of the progeny of first-cousins pairs are well known to be reduced (Morton 1978; Jaber et al. 1992; Stoll et al. 1994; Vogel and Motulsky 1997; Ober et al. 1999). Yet, our results indicate that many individuals who are not the offspring of obviously consanguineous matings have degrees of autozygosity near or even exceeding that of the offspring of first-cousin matings. Autozygosity may play a much larger role in morbidity than previously has been suspected.

Analogous to a more powerful telescope that makes possible study of new (and often more ancient) astronomic phenomena, the high-density whole-genome polymorphism scan opens to examination a whole new genetics realm. The results presented in this report could not have been obtained with low-density scans. Current technology does not permit 0.5-cM-density scans to be easily completed today, but many developments are underway that should eventually open this new realm to routine exploration (e.g., see Wang et al. 1998).

It is interesting to speculate what would be observed, in terms of autozygosity, with an even-higher-density polymorphism scan; undoubtedly, at least some shorter—and, hence, older—homozygous regions would be detected. The many reported examples of strong linkage disequilibrium in humans virtually guarantee that these short autozygous segments exist (although their number and length are unknown). Mutation, especially for STRPs, will obscure detection of these more ancient segments. For efficient detection, it may become necessary to turn to diallelic polymorphisms with lower mutation rates. Homologous human chromosome pairs may even prove to be a patchwork of alternating short segments that are either largely homozygous or largely heterozygous for low-mutation-rate polymorphisms.

Acknowledgments

Electronic-Database Information

URLs for data in this article are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation, http://www.marshmed.org/genetics

Genetic Location Database, http://cedar.genetics.soton.ac.uk/public_html/ldb.html

References

Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JI, Gershon ES (1994) Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. Proc Natl Acad Sci USA 91: 5918–5921

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: Individual and sex-specific variation in recombination. Am J Hum Genet 63:861–869

Broman KW, Weber JL (1998) Estimation of pairwise relationships in the presence of genotyping errors. Am J Hum Genet 63:1563–1564

Bugge M, Collins A, Petersen MB, Fisher J, Brandt C, Hertz JM, Tranebjærg L, et al (1998) Non-disjunction of chromosome 18. Hum Mol Genet 7:661–669

Collins A, Frezal J, Teague J, Morton NE (1996) A metric map of humans: 23,500 loci in 850 bands. Proc Natl Acad Sci USA 93:14771–14775

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R (1990) Centre d'Étude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. Genomics 6:575–577

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Egeland JA (ed) (1972) Descendents of Christian Fisher and other Amish-Mennonite pioneer families. Johns Hopkins Hospital, Baltimore

Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, et al (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. JAMA 278: 1349–1356

Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P, Service SK, et al (1996) Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nat Genet 12: 436–441

Ginns EI, Ott J, Egeland JA, Allen CR, Fann CSJ, Pauls DL, Weissenbach J, et al (1996) A genome-wide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. Nat Genet 12:431–435

Ginns EI, St Jean P, Philibert RA, Galdzicka M, Damschroder-Williams P, Thiel B, Long RT, et al (1998) A genome-wide search for chromosomal loci linked to mental health wellness in relatives at high risk for bipolar affective disorder among the Old Order Amish. Proc Natl Acad Sci USA 95: 15531–15536

Hartl DL, Clark AG (1997) Principles of population genetics, 3d ed. Sinauer Associates, Sunderland, MA

Jaber L, Merlob P, Bu X, Rotter JI, Shohat M (1992) Marked parental consanguinity as a cause for increased major malformations in an Israeli Arab community. Am J Med Genet 44:1–6

Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567–1570

McMahon FJ, Hopkins PJ, Xu J, McInnis MG, Shaw S, Cardon L, Simpson SG, et al (1997) Linkage of bipolar affective disorder to chromosome 18 markers in a new pedigree series. Am J Hum Genet 61:1397–1404

Morton NE (1978) Effect of inbreeding on IQ and mental retardation. Proc Natl Acad Sci USA 75:3906–3908

Ober C, Hyslop T, Hauck WW (1999) Inbreeding effects on fertility in humans: evidence for reproductive compensation. Am J Hum Genet 64:225–231

Pauls DL, Ott J, Paul SM, Allen CR, Fann CSJ, Carulli JP, Falls KM, et al (1995) Linkage analysis of chromosome 18 markers do not identify a major susceptibility locus for bipolar affective disorder in the Old Order Amish. Am J Hum Genet 57:636–643

Polymeropoulos MH, Schaffer AA (1996) Scanning the genome with 1772 microsatellite markers in search of a bipolar disorder susceptibility gene. Mol Psychiatry 1:404–407

Rosenberg M, Hui L, Ma J, Nusbaum HC, Clark K, Robertson L, Dziadzio L, et al (1997) Characterization of short tandem repeats from thirty-one human telomeres. Genome Res 7:917–923

Rosendaal FR (1997) Risk factors for venous thrombosis: prevalence, risk, and interaction. Semin Hematol 34:171–187

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, et al (1995) A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. Hum Mol Genet 4:1837–1844

Smith CAB (1953) Detection of linkage in human genetics. J R Stat Soc B 15:153–192

Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C, Clark CD, et al (1995) Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. Am J Hum Genet 57:1384–1394

Stoll C, Alembik Y, Dott B, Feingold J (1994) Parental consanguinity as a cause of increased incidence of birth defects in a study of 131,760 consecutive births. Am J Med Genet 49:114–117

Sunden SLF, Businga T, Beck J, McClain A, Gastier JM, Pulido JC, Yandava CN, et al (1996) Chromosomal assignment of 2900 tri- and tetranucleotide repeat markers using NIGMS somatic cell hybrid panel #2. Genomics 32:15–20

Utah Marker Development Group (1995) A collection of ordered tetranucleotide-repeat markers from the human genome. Am J Hum Genet 57:619–628

Vogel F, Motulsky AG (1997) Human genetics: problems and approaches, 3d ed. Springer-Verlag, Berlin

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, et al (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082

White R, Lalouel JM (1988) Sets of linked genetic markers for human chromosomes. Annu Rev Genet 22:259–279