Published in final edited form as:
Genet Epidemiol. 2010 February ; 34(2): 146–150. doi:10.1002/gepi.20444.

Extent and Distribution of Linkage Disequilibrium in the Old Order Amish

Cristopher V. Van Hout[1,§], **Albert M. Levin**[1,§], **Evadnie Rampersaud**[2], **Haiqing Shen**[2], **Jeffrey R. O'Connell**[2], **Braxton D. Mitchell**[2], **Alan R. Shuldiner**[2,3], and **Julie A. Douglas**[1,*]

[1] Department of Human Genetics, University of Michigan School of Medicine, Ann Arbor, Michigan [2] Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland [3]Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland

Abstract

Knowledge of the extent and distribution of linkage disequilibrium (LD) is critical to the design and interpretation of gene mapping studies. Because the demographic history of each population varies and is often not accurately known, it is necessary to empirically evaluate LD on a population-specific basis. Here we present the first genome-wide survey of LD in the Old Order Amish (OOA) of Lancaster County Pennsylvania, a closed population derived from a modest number of founders. Specifically, we present a comparison of LD between OOA individuals and U.S. Utah participants in the International HapMap project (abbreviated CEU) using a high-density single nucleotide polymorphism (SNP) map. Overall, the allele (and haplotype) frequency distributions and LD profiles were remarkably similar between these two populations. For example, the median absolute allele frequency difference for autosomal SNPs was 0.05, with an inter-quartile range of 0.02 to 0.09, and for autosomal SNPs 10-20 kb apart with common alleles (minor allele frequency ≥ 0.05), the linkage disequilibrium measure r^2 was at least 0.8 for 15% and 14% of SNP pairs in the OOA and CEU, respectively. Moreover, tag SNPs selected from the HapMap CEU sample captured a substantial portion of the common variation in the OOA (~88%) at $r^2 \geq 0.8$. These results suggest that the OOA and CEU may share similar LD profiles for other common but untyped SNPs. Thus, in the context of the common variant-common disease hypothesis, genetic variants discovered in gene mapping studies in the OOA may generalize to other populations.

Keywords

single nucleotide polymorphism; population genetics; human genetics; founder population; linkage disequilibrium; haplotypes

Introduction

Many genetic studies of complex traits and diseases are being conducted in population isolates, including the Old Order Amish (OOA) of Lancaster County Pennsylvania [Ginns, et al. 1998; Hsueh, et al. 2000; Mitchell, et al. 2001; Streeten, et al. 2006; Post, et al. 2007; Douglas, et al. 2008; Mitchell, et al. 2008; Wang, et al. 2009]. Whether results from these

*Address correspondence to: Julie A. Douglas, Ph.D. Department of Human Genetics University of Michigan 1241 E. Catherine St. 5912 Buhl Building, SPC 5618 Ann Arbor, MI 48109-5618 Phone: 734-615-2616 Fax: 734-763-3784 jddoug@umich.edu.
§These authors contributed equally to this work.

studies will generalize to other populations is dependent (in part) on the similarity of allele frequencies and patterns of linkage disequilibrium between populations. To inform future genetic studies of the OOA and facilitate comparisons of findings with other populations, we conducted the first genome-wide survey of linkage disequilibrium in the OOA and compared our findings to the International HapMap project [Frazer, et al. 2007].

Most of the present-day OOA of Lancaster County are the descendants of approximately 200 individuals [Cross 1976] from central western Europe who immigrated to the United States in the early eighteenth century [McKusick, et al. 1964]. Although recent data indicate that the differences in LD between isolated and cosmopolitan populations for common alleles are modest [Bonnen, et al. 2006; Service, et al. 2006], the uncertain but unique demographic history of the OOA necessitates empirical evaluation of LD.

Subjects and Methods

OOA study subjects were recruited and genotyped (n=861) in the course of the Heredity and Phenotype Intervention (HAPI) Heart study [Mitchell, et al. 2008], which was designed to identify gene-environment interactions influencing cardiovascular traits. Because many closely related individuals were deliberately ascertained, we used a simulated annealing algorithm [Douglas and Sandefur 2008] to select a set of minimally related individuals (30 men and 30 women). The median [range] pair-wise kinship coefficient was 0.03 [0.01-0.04] for the set of 60 versus 0.03 [0.01-0.3] for the entire sample of 861. For comparison with the OOA, we also utilized 30 men and 30 women (or 60 unrelated parents) from a U.S. Utah population with northern and western European ancestry (abbreviated CEU) in the International HapMap project [Frazer, et al. 2007].

Genotyping and QC Methods

DNA was extracted from whole blood by standard methods as described previously [Mitchell, et al. 2008]. The Affymetrix GeneChip® Human Mapping 500K Array Set was used for the comparison of LD patterns in both the OOA and CEU samples. Genotype calls were made using a Bayesian Robust Linear Model with Mahalanobis (BRLMM) distance classifier [Affymetrix 2006]. Genotype data for the CEU sample and corresponding annotation for the platform, including chromosome and genomic positions for all SNPs on the array, were obtained from the Affymetrix website (www.affymetrix.com).

Individuals with >5% missing genotypes, and/or for men, >1% heterozygous genotypes on the X chromosome, were excluded. A subset of autosomal SNPs (2,068), which were selected to have high information content (minor allele frequency (MAF) ≥0.3), low pair-wise LD (maximum r^2 of 0.44), and coverage across all autosomes (average intermarker spacing of 1.3 cM) in the OOA, were used to infer relationships using the maximum likelihood method implemented in Relpair [Epstein, et al. 2000]. We excluded individuals who had an inferred relationship that differed from the pedigree relationship with a likelihood ratio greater than 10^6. Based on these combined criteria, a total of 24 individuals (out of 861) were excluded from further analysis.

SNPs were required to satisfy the following quality control criteria in both samples: (1) ≤ 5% uncalled genotypes; (2) ≤5 and ≤1 Mendelian inconsistencies in OOA and CEU samples, respectively, using pedigree diagnostics as implemented in PedCheck [O'Connell and Weeks 1998]; and (3) Hardy Weinberg Equilibrium (HWE) p-value≥10^{-6} by Fisher's exact test [Wigginton, et al. 2005] as implemented in Haploview [Barrett, et al. 2005]. To assess genotyping accuracy, we used duplicate genotype data for 61 of the 861 OOA subjects for whom data from the Affymetrix Genome-Wide Human SNP Array 6.0 (overlap of 482,235 SNPs with Affymetrix GeneChip® Human Mapping 500K Array Set) were also

available. Only SNPs with <2 duplicate inconsistencies were retained for analysis. Of the 500,447 genotypes that mapped to a single location in the human genome, 82,404 failed at least one QC measure in at least one sample. Those SNPs were removed, leaving a total of 409,071 autosomal [Table 1] and 8,972 X chromosome [Table 1 in the Appendix] SNPs. For the SNPs that passed our quality control criteria, the genotype consistency rate among 61 duplicate pairs was 99.4%.

Statistical Analyses

Fisher's exact test was used to compare allele frequency distributions between the OOA and CEU. For common SNPs (MAF≥0.05) on the same chromosome and within 10 Mb of each other, we used the Expectation-Maximization (EM) algorithm to obtain maximum likelihood estimates of two-SNP haplotype frequencies and measured pair-wise LD by the r^2 and D'statistics [Lewontin 1964]. Based on common SNPs, we also identified haplotype blocks in the CEU using an extension of the 4-gamete rule [Wang, et al. 2002] and estimated haplotype frequencies in both the CEU and OOA using the EM algorithm with a partition-ligation method [Qin, et al. 2002] for blocks with >10 SNPs as implemented in Haploview [Barrett, et al. 2005]. For each sample, we then calculated and compared the effective number of haplotypes in each block, i.e., $(\Sigma p_i^2)^{-1}$, where p_i is the frequency of the i^{th} haplotype in the block. As a measure of redundancy, we identified the number of SNPs (or proxies) that were in strong LD with each SNP at various thresholds of r^2 in each sample. To evaluate the extent to which SNPs selected to tag variation in the CEU capture common variation in the OOA, we selected common tag SNPs in the CEU using the greedy algorithm [Carlson, et al. 2004] implemented in Haploview [Barrett, et al. 2005] such that every unselected SNP had an $r^2 \geq 0.8$ with one or more selected SNPs. We then calculated r^2 between the tag SNPs and the remaining 'non-tagged' but typed SNPs in the OOA. Unless specified otherwise, all analyses were carried out using a combination of in-house R, Perl, and C programs.

Results

For the 418,043 SNPs that passed QC, mean heterozygosity was 0.26 and 0.27 for the autosomes in the OOA and CEU, respectively, and 0.23 and 0.24 for the X chromosome. The slightly lower heterozygosity in the OOA reflects the larger number of monomorphic SNPs in the OOA relative to the CEU, e.g., 68,869 versus 57,669 for the autosomes [Table 1]. Among all SNPs that were polymorphic in at least one sample, the median absolute allele frequency difference was 0.05 for the autosomes and 0.07 for the X chromosome. At p-value$<10^{-6}$, OOA and CEU allele frequencies were significantly different for 799 autosomal and 137 X chromosome SNPs.

The percentage of SNP pairs within 10 Mb of each other and between which strong LD was observed was remarkably similar between the OOA and CEU for the autosomes [Table 2] and the X chromosome [Table 2 in the Appendix]. For example, for autosomal SNPs at an inter-marker distance of <10 kb, no evidence of recombination (D'=1) was observed for 79% and 75% of SNP pairs, perfect LD ($r^2=1$) was observed for 20% and 19% of SNP pairs, and useful LD ($r^2 \geq 0.8$) was observed for 30% and 29% of SNP pairs in the OOA and CEU, respectively. Based on the CEU sample, we identified 58,097 autosomal haplotype blocks, with a median of 3 SNPs per block and an inter-quartile range of [3, 4]. Among all autosomal blocks, the median effective number of haplotypes (n_e) was 2.43 and 2.47 in the OOA and CEU, respectively, and the median of the differences in n_e (CEU minus OOA) per block was 0.04, with an inter-quartile range of −0.2 to 0.3, suggesting modestly greater haplotype diversity in the CEU. Results based on haplotype blocks defined in the OOA did not qualitatively differ from those based on blocks defined in the CEU (data not shown).

Of common autosomal SNPs, 72% and 64% had at least one proxy at $r^2 \geq 0.8$ and 55% and 44% had at least one perfect proxy ($r^2=1$) in the OOA and CEU, respectively, indicating that fewer independent SNPs are required to represent variation in the OOA relative to the CEU. At $r^2 \geq 0.8$, 170,979 of 310,704 common SNPs in the CEU were selected as tag SNPs and captured ~88% of the 'non-tagged' SNPs in OOA, suggesting that SNPs selected to tag common variation in the CEU capture much of the same variation in the OOA. SNPs not captured by the CEU tag SNPs tended to be of lower minor allele frequency (data not shown). Results for the X chromosome were qualitatively similar.

Discussion

In general, we found a high degree of similarity in allele frequencies and LD patterns in the OOA and CEU samples. Allele frequencies were not significantly different between the OOA and CEU for >99% of SNPs. Based on common SNPs, which comprised 74% and 66% of autosomal SNPs in the OOA and CEU, respectively, the distribution and extent of LD were remarkably similar between these two samples. These data are consistent with previous theoretical predictions [Kruglyak 1999; Pritchard and Przeworski 2001] and recent empirical data [Bonnen, et al. 2006; Service, et al. 2006; Navarro, et al. 2009; Thompson, et al. 2009], all of which point to modest differences in LD between isolated and cosmopolitan populations for common alleles. The situation for rare alleles, however, is likely to be different as has been demonstrated in applications of LD mapping for monogenic diseases and traits.

Demographic and historical information indicate that the OOA were founded relatively recently (~10 to 15 generations ago) by a modest number of individuals (several hundred) and then expanded rapidly to a current census population size exceeding 30,000 [Lancaster County Amish 2002]. Though the precise demographic details are unknown, it is apparent that the number of founders and rate of growth were sufficient and that the subsequent isolation of the OOA was too short for genetic drift and/or recombination to have meaningfully altered the common allele or haplotype frequency spectrum. Our recent study of variation on the Y chromosome supports these observations in that much of the diversity observed in non-isolated populations of similar ancestry is present in the OOA [Pollin, et al. 2008]. It appears that inbreeding due to the finite population size of the OOA was also insufficient to meaningfully alter the allele frequency distribution or extent of LD. Based on the 60 OOA individuals included in our analyses, the average inbreeding coefficient F [Wright 1922] was 0.026 (range of 0.0003 to 0.046), which is too weak to generate substantial differences in LD relative to a non-isolated population [Hill and Robertson 1968].

Owing to similar allele frequencies and LD patterns in the OOA and CEU, CEU-derived tag SNPs performed well in capturing common variation in the OOA, consistent with previous studies in other samples of European ancestry, including those from isolated populations [Willer, et al. 2006; Service, et al. 2007]. These results suggest that the OOA and CEU samples may also share similar LD profiles for other common but untyped SNPs. Thus, findings from gene mapping studies in the OOA may generalize to other populations in the context of the common variant-common disease hypothesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the Amish Research Clinic Staff, our Amish liaisons, and the Amish community, whose extraordinary support and cooperation made this study possible. We also thank Drs. Alejandro Schaffer and Richa Agarwala at the NIH/NCBI for providing the pedigree information and the Center for Inherited Disease Research (CIDR), NIH for providing duplicate genotypes from the Affymetrix Genome-Wide Human SNP Array 6.0. This study was supported in part by NIH grants U01 HL72515 and R01 CA122844.

References

Affymetrix. BRLMM: an Improved Genotype Calling Method for the GeneChip Human Mapping 500K Array Set. 2006. http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21(2):263–5. [PubMed: 15297300]

Bonnen PE, Pe'er I, Plenge RM, Salit J, Lowe JK, Shapero MH, Lifton RP, Breslow JL, Daly MJ, Reich DE. Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. Nat Genet. 2006; 38(2):214–7. others. [PubMed: 16429162]

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet. 2004; 74(1):106–20. [PubMed: 14681826]

Cross HE. Population studies and the Old Order Amish. Nature. 1976; 262(5563):17–20. [PubMed: 934323]

Douglas JA, Roy-Gagnon MH, Zhou C, Mitchell BD, Shuldiner AR, Chan HP, Helvie MA. Mammographic breast density--evidence for genetic correlations with established breast cancer risk factors. Cancer Epidemiol Biomarkers Prev. 2008; 17(12):3509–16. [PubMed: 19029399]

Douglas JA, Sandefur CI. PedMine--a simulated annealing algorithm to identify maximally unrelated individuals in population isolates. Bioinformatics. 2008; 24(8):1106–8. [PubMed: 18321883]

Epstein MP, Duren WL, Boehnke M. Improved inference of relationship for pairs of individuals. Am J Hum Genet. 2000; 67(5):1219–31. [PubMed: 11032786]

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449(7164):851–61. others. [PubMed: 17943122]

Ginns EI, Jean P, Philibert RA, Galdzicka M, Damschroder-Williams P, Thiel B, Long RT, Ingraham LJ, Dalwaldi H, Murray MA. 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 U S A. 1998; 95(26):15531–6. others. [PubMed: 9861003]

Hill WG, Robertson A. Linkage Disequilibrium in Finite Populations. Theoretical and Applied Genetics. 1968; 38:226–231.

Hsueh WC, Mitchell BD, Aburomia R, Pollin T, Sakul H, Gelder Ehm M, Michelsen BK, Wagner MJ, Jean PL, Knowler WC. Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. Diabetes Care. 2000; 23(5):595–601. others. [PubMed: 10834415]

Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet. 1999; 22(2):139–44. [PubMed: 10369254]

Lancaster County Amish. Church Directory of the Lancaster County Amish: The Diary. Gordonville, PA: 2002.

Lewontin RC. The Interaction of Selection and Linkage. II. Optimum Models. Genetics. 1964; 50:757–82. [PubMed: 14221879]

McKusick VA, Hostetler JA, Egeland JA. Genetic Studies of the Amish, Background and Potentialities. Bull Johns Hopkins Hosp. 1964; 115:203–22. [PubMed: 14209042]

Mitchell BD, Hsueh WC, King TM, Pollin TI, Sorkin J, Agarwala R, Schaffer AA, Shuldiner AR. Heritability of life span in the Old Order Amish. Am J Med Genet. 2001; 102(4):346–52. [PubMed: 11503162]

Mitchell BD, McArdle PF, Shen H, Rampersaud E, Pollin TI, Bielak LF, Jaquish C, Douglas JA, Roy-Gagnon MH, Sack P. The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. Am Heart J. 2008; 155(5):823–8. others. [PubMed: 18440328]

Navarro P, Vitart V, Hayward C, Tenesa A, Zgaga L, Juricic D, Polasek O, Hastie ND, Rudan I, Campbell H. Genetic comparison of a Croatian isolate and CEPH European Founders. Genetic Epidemiology. 2009 others. In this issue.

O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet. 1998; 63(1):259–66. [PubMed: 9634505]

Pollin TI, McBride DJ, Agarwala R, Schaffer AA, Shuldiner AR, Mitchell BD, O'Connell JR. Investigations of the Y chromosome, male founder structure and YSTR mutation rates in the Old Order Amish. Hum Hered. 2008; 65(2):91–104. [PubMed: 17898540]

Post W, Bielak LF, Ryan KA, Cheng YC, Shen H, Rumberger JA, Sheedy PF 2nd, Shuldiner AR, Peyser PA, Mitchell BD. Determinants of coronary artery and aortic calcification in the Old Order Amish. Circulation. 2007; 115(6):717–24. [PubMed: 17261661]

Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. Am J Hum Genet. 2001; 69(1):1–14. [PubMed: 11410837]

Qin ZS, Niu T, Liu JS. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. Am J Hum Genet. 2002; 71(5):1242–7. [PubMed: 12452179]

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat Genet. 2006; 38(5):556–60. others. [PubMed: 16582909]

Service S, Sabatti C, Freimer N. Tag SNPs chosen from HapMap perform well in several population isolates. Genet Epidemiol. 2007; 31(3):189–94. [PubMed: 17323370]

Streeten EA, McBride DJ, Pollin TI, Ryan K, Shapiro J, Ott S, Mitchell BD, Shuldiner AR, O'Connell JR. Quantitative trait loci for BMD identified by autosome-wide linkage scan to chromosomes 7q and 21q in men from the Amish Family Osteoporosis Study. J Bone Miner Res. 2006; 21(9):1433–42. [PubMed: 16939402]

Thompson EE, Sun Y, Nicolae D, Ober C. Shades of gray: A comparison of linkage disequilibrium between Hutterites and Europeans. Genetic Epidemiology. 2009 In this issue.

Wang N, Akey JM, Zhang K, Chakraborty R, Jin L. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. Am J Hum Genet. 2002; 71(5):1227–34. [PubMed: 12384857]

Wang Y, O'Connell JR, McArdle PF, Wade JB, Dorff SE, Shah SJ, Shi X, Pan L, Rampersaud E, Shen H. Whole-genome association study identifies STK39 as a hypertension susceptibility gene. PNAS. 2009; 106(1):6. others. [PubMed: 19118201]

Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet. 2005; 76(5):887–93. [PubMed: 15789306]

Willer CJ, Scott LJ, Bonnycastle LL, Jackson AU, Chines P, Pruim R, Bark CW, Tsai YY, Pugh EW, Doheny KF. Tag SNP selection for Finnish individuals based on the CEPH Utah HapMap database. Genet Epidemiol. 2006; 30(2):180–90. others. [PubMed: 16374835]

Wright S. Coefficients of Inbreeding and Relationship. American Naturalist. 1922; 56:330–338.

Table 1

Summary of autosomal SNPs

	OOA	CEU	Overlap
Total genotyped	489,922	489,922	489,922
>1 duplicate inconsistency[1]	51,459	NA	NA
>5% missing data[2]	50,085	16,896	8,973
Mendelian inconsistencies[2,3]	3,188	1,168	202
$p<10^{-6}$ for HWE test[4]	379	217	116
Passed QC filter[5]	415,440	472,851	409,071
Passed QC in both OOA and CEU			
Monomorphic[4]	68,869	57,669	52,467
Polymorphic[4]			
MAF≥0.05	297,605	310,704	287,476
MAF≥0.10	256,614	267,149	240,375
MAF≥0.20	182,941	189,133	161,062

OOA = Old Order Amish

CEU = U.S. Utah residents from HapMap

MAF = Minor Allele Frequency

Note: SNPs that failed a QC measure in either sample were excluded from further analysis, and SNPs with MAF≥0.05 passing QC in both samples (n=287,476) were used for LD analysis.

[1] Based on the 61 OOA individuals who were also genotyped on the Affymetrix 6.0 array; SNPs with more than one duplicated genotype discrepancy were excluded.

[2] Based on 837 OOA and 90 CEU individuals (30 trios).

[3] SNPs with >5 and >1 Mendelian inconsistencies in OOA and CEU, respectively.

[4] Based on 60 unrelated individuals (30 men and 30 women) from each sample.

[5] SNPs may fail QC in more than one way, so rows do not sum to the subtotal passing QC.

Table 2

Percentage of autosomal SNP pairs[1] showing no evidence of recombination (D′=1), perfect LD (r^2=1), or where useful LD is observed (r^2≥0.8)

Inter-SNP distance (kb)	D′=1		r^2=1		r^2≥0.8	
	OOA	CEU	OOA	CEU	OOA	CEU
≤10	79	75	20	19	30	29
10-20	60	53	9	7	15	14
20-50	43	34	4	3	9	7
50-100	28	20	1	1	3	2
100-200	20	11	0	0	1	1
200-500	14	7	0	0	0	0
500-1,000	12	6	0	0	0	0
1,000-2,000	11	5	0	0	0	0
2,000-5,000	10	5	0	0	0	0
5,000-10,000	8	5	0	0	0	0

OOA = Old Order Amish (n=60)

CEU = U.S. Utah residents from HapMap (n=60)

[1]Restricted to SNPs with minor allele frequency ≥0.05 in both samples (n=287,476).