ELECTRONIC LETTER

Potential for expanded power in linkage studies using the ALLEGRO and MERLIN software programs

E Rampersaud, W K Scott, E R Hauser, M C Speer

J Med Genet 2005;42:e68 (<http://www.jmedgenet.com/cgi/content/full/42/12/e68>). doi: 10.1136/jmg.2005.032029

Multipoint linkage analysis in complex diseases requires the use of fast algorithms that can handle many markers and a large number of moderately sized pedigrees with unknown mode of inheritance. This need has led to the development of several competitive software programs. We recently completed a genomic screen of neural tube defects using GENEHUNTER-PLUS and the more recent ALLEGRO. The ALLEGRO software was found to offer expanded power for linkage studies, particularly for childhood onset diseases like neural tube defects, though the results must be treated with caution.

Multipoint linkage analysis in complex diseases necessitates the use of fast algorithms that can handle many markers and a large number of moderately sized pedigrees with unknown mode of inheritance. This need has spawned the development of several competitive software programs that employ the same allele sharing method based on the Lander-Green algorithm,^{1,2} but differ in the computational approaches used. The most recent programs, ALLEGRO³ and MERLIN,⁴ improve on the speed of the original GENEHUNTER and GENEHUNTER-PLUS (GH+) programs and allow for larger sized pedigrees.

METHODS

We recently completed a genomic screen of neural tube defects (NTDs)⁵ where our analyses were performed using ALLEGRO and GH+. We found some variation in the overall results from the two analytic packages, which led to an investigation of the pedigree specific results. Of interest to us was our finding that ALLEGRO not only calculates identity by descent sharing between genotyped affected individuals but also calculates a full likelihood using affected individuals who were not genotyped, even those without genotyped descendants available for inferring missing linkage phase. In our analysis of the NTD dataset, this meant that pedigrees having only one genotyped affected individual were informative when we used ALLEGRO, but this was not the case with GH+.

To illustrate this better, we present a simple example using pedigree structure from our dataset. Figure 1 shows two pedigrees, each containing affected individuals who were not genotyped. There are two affected individuals in family 8007; individual 12 is an affected distant cousin who was not genotyped. Family 8776 has four affected individuals and individual 102 was not genotyped; for the purpose of our illustration, only a subset of individuals from this family is shown. We calculated affecteds-only multipoint parametric LOD scores for chromosome 7 markers from our screen (9 cM sex-average genetic map) using GH+ and ALLEGRO for comparison (table 1). We compare non-parametric results from GH+ version 1.2, ALLEGRO version 1.2c, and MERLIN version 0.10.2. We report parametric LOD scores under a

dominant genetic model, and non-parametric LOD scores based on a linear model and using the S_{pairs} statistic. For simplicity, we examine LOD scores at marker d7s2201 (3.1 cM).

RESULTS

ALLEGRO and GH+ give discrepant parametric linkage results for pedigree 8007 (GH+ 0.0, ALLEGRO 0.79) and pedigree 8776 (GH+ 1.75, ALLEGRO 2.01). However, the non-parametric LOD scores are closer for family 8776 (GH+ 0.72, ALLEGRO 0.75, MERLIN 0.75). For family 8007, only ALLEGRO (LOD 0.88) and MERLIN (LOD 0.88) appear to use the full linkage information to calculate non-parametric LODs. Changing the affection status of individual 102 in family 8776 from affected to unknown, results in the exclusion of that person from the analysis and produces a parametric LOD score from ALLEGRO of 1.77 at 3.1 cM, a value that is closer to the score calculated by GH+ (1.75). In all our calculations, pedigree size is not a factor since the number of bits calculated for each pedigree is below the maximum number ($n = 17$) allowed by GH+, ALLEGRO, and MERLIN. Hence, no "trimming" of less informative individuals in these pedigrees was performed.

The parametric LOD scores showed the largest difference between the programs. Hence, we also performed parametric linkage analysis using VITESSE version 1.0,⁶ which implements the Elston-Stewart algorithm;⁷ we used marker map *d7s3056_3.1cM_d7s2201_6.4cM_d7s641*. As expected, the results from VITESSE (LOD 0.79) are the same as from ALLEGRO (LOD 0.79) for family 8007 and for family 8776 (ALLEGRO 2.01, VITESSE 2.01).

DISCUSSION

The ALLEGRO software offers expanded power for linkage studies, particularly for childhood onset diseases like neural tube defects, in which many affected individuals die soon after birth or in early childhood and are unsampled. However, the results must then be interpreted with caution. In such pedigrees, we cannot confirm that the affected unsampled child did inherit a disease associated haplotype present in a sampled affected individual; instead, we can only confirm that the disease associated haplotype is or is not present in the unsampled affected individual's parent and could (or could not) have been transmitted. Misspecification of the unsampled individual's genotype could in this way result in a reduction of power to detect linkage. However, for rare diseases such as NTD, this is unlikely to be the case. MERLIN appears to handle ungenotyped individuals in the calculation of non-parametric LOD scores under the linear model in a similar manner to ALLEGRO. Additionally, MERLIN can now be used to calculate parametric LOD scores and these results are similar to the LOD scores from ALLEGRO.

Abbreviations: GH+, GENEHUNTER-PLUS; NTD, neural tube defect

Table 1 Comparison of multipoint parametric LOD scores calculated by GH+, ALLEGRO, VITESSE, and MERLIN								
Family	Location (cM)	Parametric LOD score*				Non-parametric LOD score†		
		GH+	ALLEGRO	MERLIN	VITESSE	GH+	ALLEGRO	MERLIN
8007	3.1	0.00	0.79	0.79	0.79	0.00	0.88	0.88
8776	3.1	1.75	2.01	2.01	2.01	0.72	0.75	0.75

*Assuming an autosomal dominant genetic model; allele frequency 0.001; †based on the s_{Pairs} statistic, using a linear model.

Figure 1 Two pedigrees each containing affected individuals (individual 12 in family 8007, individual 102 in family 8776) who have not been genotyped. The plus symbol (+) indicates that an individual has been genotyped.

ACKNOWLEDGEMENTS

We thank Michael Frigge (DeCode Genetics, Inc., Iceland) for helpful correspondence related to this publication.

Authors' affiliations

E Rampersaud, W K Scott, E R Hauser, M C Speer, Center for Human Genetics, Duke University Medical Center, Durham, NC, USA

This work was supported by a Ruth L Kirschstein National Research Service Awards predoctoral fellowship (NS046249) and by grants from the National Institutes of Health (NS39818, HD39948, ES11375, ES011961).

Competing interests: none declared

Correspondence to: Dr Marcy C Speer, Duke University Medical Center, Box 3445, Durham, NC 27710, USA; marcy@chg.duhs.duke.edu

Revised version received 19 April 2005
Accepted for publication 24 April 2005

REFERENCES

- 1 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996;**58**:1347–63.
- 2 Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997;**61**:1179–88.
- 3 Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000;**25**:12–3.
- 4 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;**30**:97–101.
- 5 Rampersaud E, Bassuk AG, Enterline DS, Timothy M, Siegel DG, Melvin EC, Aben J, Allen J, Aylsworth A, Brei T, Bodhurtha J, Buran C, Floyd LE, Hammock P, Iskandar B, Ito J, Kessler JA, Lasarsky N, Mack P, Mackey J, McLone D, Meeropol E, Mehlretter L, Mitchell LE, Oakes WJ, Nye JS, Powell C, Sawin K, Stevenson R, Walker M, West SG, Worley G, Gilbert JR, Speer MC. Whole genome-wide linkage screen for neural tube defects reveals regions of interest on chromosomes 7 and 10. *J Med Genet* 2005;**12**:940–6.
- 6 O'Connell JR, Weeks DE. The VITESSE algorithm for rapid exact multilocus linkage analysis data via genotype set-recording and fuzzy inheritance. *Nat Genet* 1995;**11**:402–8.
- 7 Elston RC, Stewart J. A general model for genetic analysis of pedigree data. *Hum Hered* 1971;**21**:523–42.