

Software

Open Access

Computing power of quantitative trait locus association mapping for haploid loci

Derek Gordon*¹ and Andrew R Zinn²

Address: ¹Department of Genetics, Rutgers University, 145 Bevier Road, Piscataway, NJ 08854, USA and ²McDermott Center for Human Growth and Development, University of Texas Southwestern Medical School, 5323 Harry Hines Blvd, Dallas, TX 75390-8591, USA

Email: Derek Gordon* - gordon@biology.rutgers.edu; Andrew R Zinn - Andrew.Zinn@UTSouthwestern.edu

* Corresponding author

Published: 23 August 2009

Received: 13 February 2009

BMC Bioinformatics 2009, **10**:261 doi:10.1186/1471-2105-10-261

Accepted: 23 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2105/10/261>

© 2009 Gordon and Zinn; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Statistical power calculations are a critical part of any study design for gene mapping. Most calculations assume that the locus of interest is biallelic. However, there are common situations in human genetics such as X-linked loci in males where the locus is haploid. The purpose of this work is to mathematically derive the biometric model for haploid loci, and to compute power for QTL mapping when the loci are haploid.

Results: We have derived the biometric model for power calculations for haploid loci and have developed software to perform these calculations. We have verified our calculations with independent mathematical methods.

Conclusion: Our results fill a need in power calculations for QTL mapping studies. Furthermore, failure to appropriately model haploid loci may cause underestimation of power.

Background

Statistical power calculations are a critical part of any study design for gene mapping. With regards to quantitative trait locus (QTL) mapping, it is typically assumed that there are two alleles at the QTL, and hence three genotypes (e.g., see Lynch and Walsh [1]). For each genotype, there is a corresponding mean quantitative value that is determined by the biometric model developed by Falconer and others [2]. To determine the mean of the heterozygote, we need a parameter known as the dominance parameter. Software to compute power for the biallelic situation has been made available in the Genetic Power Calculator, developed by Purcell et al. [3] However, for haploid organisms or loci, there is no such parameter because there are no heterozygotes. Thus, there is a different set of calculations necessary to compute power for association. There has been an extensive amount of literature written

on the analysis of haploid data (see, e.g., Wu et al. [4,5], Tulsieram et al. [6]) To our knowledge, there has been no power calculator developed for QTL mapping with haploid loci. This is an oversight, since there are several common situations in humans involving loci that are haploid or hemizygous [7,8].

The purpose of this work is to mathematically derive the biometric model for haploid loci, and to compute power for QTL mapping when the loci are haploid.

Implementation

The program for computing power is written in C++ and may be created on any computer that has a C++ compiler, for example those computers with a Unix or Linux operating system.

After making the executable, the user types the phrase "power-ttesthap" to implement the program. The user is queried for the sample size of the study. Next, the user is asked for the significance level at which simulation power is to be computed. When this value is entered, the threshold for the *t*-test is computed and reported to the screen. Following this, the user is asked about the locus-specific quantitative trait heritability *Q*. When this value is entered, the program determines the means of the allele 0 group and the allele 1 group, as well as the variance in each group. Finally, the user is asked to specify the number of replicates for which power is computed. We recommend at least 100,000 replicates. See Figure 1 for an example of how data are entered, and see additional file 1. Below, we place an example run.

Results and discussion

Calculations of simulation power as function of Locus specific QTL heritability and increaser allele frequency

In Table 1, we provide an abbreviation of all the variables used in the Results section. In Figure 2, we plot the simulation power as a function of the Locus specific QTL heritability and the increaser allele frequency, using the ranges of *Q* and *p* that we specified in the Methods. There are sev-

```
T-Test power calculator - Haploid data, version 1.0, May
2009

Written by Derek Gordon

Please specify number of subjects in study: 500
Please specify significance level for t test: 0.01

t test threshold: 2.58574
Genetic model parameters
Please specify locus specific QTL heritability: 0.05

Please specify allele frequency in Allele Group 1
(Increaser Allele): 0.45

Mean in Allele 0 Group: -0.24721
Mean in Allele 1 Group: 0.20226
Variance in each group: 0.95000
Please specify number of replicates in simulation: 100000
Power of t test for genetic model parameters at
0.0099999998 significance level: 0.9941600000
```

Figure 1
Example run with program T-Test power calculator – Haploid data.

eral conclusions we may draw from studying this graph. The first is that, as *Q* increases, the power increases for any value of *p*. The second conclusion is that, if *Q* is held constant, the power differs minimally over the range of *p*. In fact, if we consider the set of consecutive differences ((power for *Q* = 0.05 and *p* = 0.25) – (power for *Q* = 0.05 and *p* = 0.1)), ((power for *Q* = 0.05 and *p* = 0.5) – (power for *Q* = 0.05 and *p* = 0.25)),..., ((power for *Q* = 0.1 and *p* = 0.9) – (power for *Q* = 0.1 and *p* = 0.75)), the maximum difference is 0.012, and the average difference is -0.00025. The maximum difference occurs for the difference ((power for *Q* = 0.07 and *p* = 0.25) – (power for *Q* = 0.07 and *p* = 0.1)). We note that the power was calculated using the executable file that is compiled from the material in Additional File 2.

By contrast, if we consider the set of consecutive differences ((power for *Q* = 0.06 and *p* = 0.1) – (power for *Q* = 0.05 and *p* = 0.1)), ((power for *Q* = 0.07 and *p* = 0.1) – (power for *Q* = 0.06 and *p* = 0.1)),..., ((power for *Q* = 0.1 and *p* = 0.9) – (power for *Q* = 0.09 and *p* = 0.9)), the maximum difference is 0.182 and the average difference is 0.085. The maximum difference occurs for ((power for *Q* = 0.06 and *p* = 0.75) – (power for *Q* = 0.05 and *p* = 0.75)).

Another result from this graph is that power of at least 80% may be achieved when the Locus specific QTL heritability *Q* is at least 0.07, suggesting that genes for QTLs with reasonable variance may be mapped even when the significance level is stringent.

Real data example – EFHC2 gene SNP rs7055196 typed in 45, X Turner Syndrome subjects

Zinn et al. genotyped 97 45, X Turner Syndrome subjects at the SNP rs7055196. They also phenotyped the subjects for Facial Affect Fear Recognition score and performed a *t*-test to determine whether there was a significant difference in the scores based on the allele at the rs7055196 locus [9]. They found no significant difference among the two groups (49.8 +/- 26.5 for those with A allele versus 46.2 +/-26.1 for those with G allele; *p*-value = 0.67; see Figure 1 of Zinn et al.).

Because there are 11 subjects with the G allele and 86 subjects with the A allele, we estimate the increaser allele frequency to be 0.87. The sample size we consider is 97. We specify a range of locus-specific QTL heritabilities from 0.005 to 0.15 in increments of 0.005. Since we are only performing a single test, we consider a significance level of 0.05. We present our results in Figure 3, where each simulation data point is created using 100,000 replicates. In that figure, we see that power increases as the heritability goes up (similar to Figure 2). Also, a heritability of 0.04 gives a simulation power of approximately 0.50. These

Table 1: List of Abbreviations

Variable	Description
N	Sample size
$\hat{\alpha}$	Significance level
T	Threshold for t test (uses N and α)
Q	Variance of quantitative trait locus (QTL)
P	Frequency of QTL increaser allele
μ_0	Mean QTL value in group "0" (those with non-increaser allele)
μ_1	Mean QTL value in group "1" (those with increaser allele)
R	Number of replicates

results suggest that the heritability is lower rather than higher.

Conclusion

To our knowledge, this is the first work to determine the power of QTL mapping for haploid loci or organisms. We compute power by simulation, and our results with exact analytic power suggest that the power by simulation is highly accurate. The most common application of these results is SNP association studies of loci on the X or Y chromosomes in male subjects. Other situations where haploidy arises in humans include chromosome disorders such as 45,X Turner syndrome [7] and segmental aneuploidies such as Cri-du-Chat (5p-) syndrome [8]. Last, the power to map QTLs in haplo/diploid organisms such as *Saccharomyces cerevisiae* can be increased by studying organisms in the haploid state [10].

Natural extensions of this work are situations such as gene-gene interactions (epistasis of different kinds), gene-environment interactions, and multi-trait mapping. This is the subject of future research. The purpose of this work is to provide researchers with the simplest non-trivial example of power calculations for haploid data.

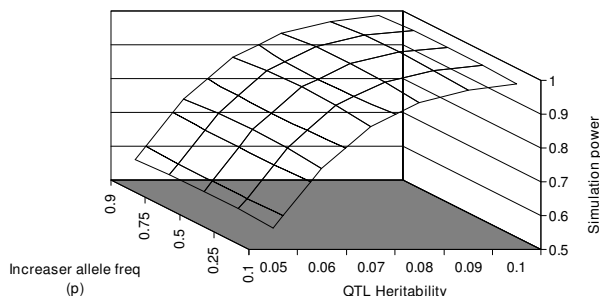


Figure 2 Simulation power results. In this figure, we present simulation power results as a function of the increaser allele p and the Locus specific QTL heritability Q . The total sample size is 500 subjects and the significance level is 1×10^{-6} .

Availability and requirements

Project name: t-test power calculator for haploid data

Project home page: <http://compgen.rutgers.edu/ttesthap/>

Operating system(s): Unix Solaris, Linux

Programming language: C++

Other requirements: None.

License: None.

Any restrictions to use by non-academics: None.

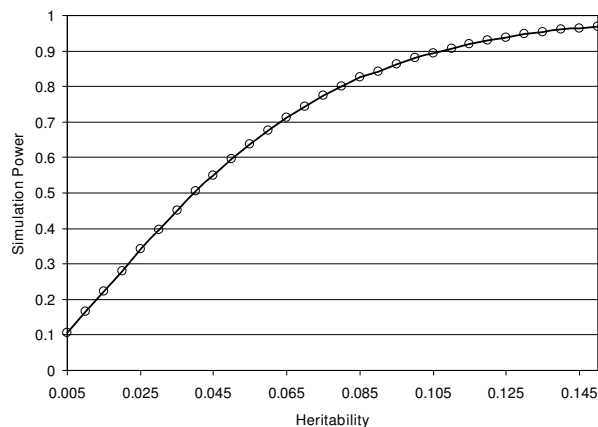


Figure 3 Simulation power results as a function of locus-specific QTL heritability for EFHC2 gene SNP rs7055196 example. In this figure, we present simulation power results as a function of the Locus specific QTL heritability Q . The total sample size is 97 subjects, the increaser allele frequency is 0.87 and the significance level is 0.05.

Authors' contributions

DG and ARZ conceived of the study and wrote the manuscript. DG created the software program. Both authors read and approved the final manuscript.

Additional material

Additional file 1

Methods for haploid power. This file provides the description of the method used for calculating power.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-10-261-S1.doc>]

Additional file 2

Files for compiling power-ttesthap (executable). This file containing the information needed to produce the executable that computes haploid power for a fixed sample size.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-10-261-S2.tar>]

Acknowledgements

The authors gratefully acknowledge grant NIH NS35554 to ARZ. Also, the authors gratefully acknowledge the help of Dr. Tara Matisse, who created the home page for their software.

References

1. Lynch M, Walsh B: **Genetics and Analysis of Quantitative Traits**. Sunderland: Sinauer Associates, Inc; 1998.
2. Falconer DS, Mackay TFC: **Introduction to Quantitative Genetics**. Fourth edition. San Francisco: Benjamin Cummings; 1996.
3. Purcell S, Cherny SS, Sham PC: **Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits**. *Bioinformatics* 2003, **19(1)**:149-150.
4. Wu RL: **Mapping quantitative trait loci by genotyping haploid tissues**. *Genetics* 1999, **152(4)**:1741-1752.
5. Wu RL, O'Malley DM, McKeand SE: **Understanding the genetic architecture of a quantitative trait in gymnosperms by genotyping haploid megagametophytes**. *Theor Appl Genet* 1999, **99(6)**:1031-1038.
6. Tulsieram LK, Glaubitz JC, Kiss G, Carlson JE: **Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes**. *Biotechnology (N Y)* 1992, **10(6)**:686-690.
7. Ford CE, Jones KW, Polani PE, De Almeida JC, Briggs JH: **A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome)**. *Lancet* 1959, **1(7075)**:711-713.
8. Lejeune J, Lafourcade J, Berger R, Vialatte J, Boeswillwald M, Sringue P, Turpin R: **[3 Cases of Partial Deletion of the Short Arm of a 5 Chromosome.]**. *C R Hebd Seances Acad Sci* 1963, **257**:3098-3102.
9. Faul F, Erdfelder E, Lang AG, Buchner A: **G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences**. *Behav Res Methods* 2007, **39(2)**:175-191.
10. Zinn AR, Kushner H, Ross JL: **EFHC2 SNP rs7055196 is not associated with fear recognition in 45,X Turner syndrome**. *Am J Med Genet B Neuropsychiatr Genet* 2008, **147B(4)**:507-509.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

