Laboratory Note

Genetic Epistasis IX - Comparative Assessment of the Algorithms

LN-9-2014

Ricardo Pinho and Rui Camacho
FEUP
Rua Dr Roberto Frias, s/n,
4200-465 PORTO
Portugal
Fax: (+351) 22 508 1440
e-mail: ei09045@fe.up.pt
www : http://www.fe.up.pt/∼ei09045 rcamacho@fe.up.pt
www : http://www.fe.up.pt/∼rcamacho

May 2014

Abstract

This lab note contains the results obtained from the algorithms discussed in previous lab notes. All algorithms are compared by their characteristics and by their Power, scalability, and Type 1 Error Rates in epistatic detection, main effect detection, and full effect detection. From the results obtained, we can see that the algorithm BOOST has the highest Power in epistatic detection and main effect detection, but has a high error rate. Screen and Clean has a contant but high error rate overall, very low Power in epistatic detection and average Power in other models.SNPHarvester and SNPRuler have relatively low Power, but low error rates. TEAM has good Power, but high error rate. MBMDR has good Power and low Type 1 Error Rate, but very bad scalability. BEAM3 has high Power in main effect detection, but also high error rate. In terms of scalability, BOOST is the most scalable, with MBMDR being the least scalable.

1 Introduction

In this lab note, the epistasis detection algorithms used in earlier lab notes([PC14b] [PC14c] [PC14d] [PC14e] [PC14f] [PC14g] [PC14h]) will be compared, using the results from the data sets and measurements discussed in Lab Note LN-1-2014 [PC14a].
The algorithms used in this empirical study are BEAM 3.0 [Zha12]; BOOST [WYY+10a]; MBMDR [MVV11]; Screen and Clean [WDR+10]; SNPRuler [WYY+10b]; SNPHarvester [YHW+09]; and TEAM [ZHZW10]. Table 1 and Table 2 show the main characteristics of the search methods, scoring techniques, types of disease models detected, and the programming language of the tested algorithms [SZS+11].

Table 1: Similarities and differences between BEAM3, BOOST MBMDR, and Screen & Clean.

Features	BEAM 3	BOOST	MBMDR	Screen & Clean
Search	Stochastisc	Exhaustive	Exhaustive	Heuristic
Permutation Test	√	–	√	–
Chi-square Test	–*	√	–*	–*
Tree/Graph Structure	√	–	–	–
Bonferroni Correction	–	√	–	√
Interactive Effect	√	√	√	√
Main Effect	√	√	√	√
Full Effect	√	√	√	√
Programming Language	C++	C	R	R

*Although BEAM3 can evaluate interactive and full effects, the evaluation test is not comparable between methods. Only single SNPs are evaluated with χ^2 test. MBMDR and Screen & Clean results are comparable with other algorithms.

Features	SNPHarvester	SNPRuler	TEAM
Search	Stochastic	Heuristic	Exhaustive
Permutation Test	–	–	√
Chi-square Test	√	√	–
Tree Structure	–	√	√
Bonferroni Correction	√	√	–
Interactive Effect	√	√	√
Main Effect	√	–	–
Full Effect	√	–	–
Programming Language	Java	Java	C++

2 Comparative Assesmement

The measures used to assess the quality of each algorithm are: **Power**; **Scalability**; and **Type 1 Error Rate**.

2.1 Power

The Power of an algorithm is related to its ability to find the ground truth of the disease. In this case, the Power is evaluated by the number of data sets, out of 100, where the algorithm finds the ground truth and is measured as a percentage for each data set configuration. In each data set, the most significant interactions, i.e. $\alpha < 0.05$, are selected.

2.2 Scalability

Scalability is determined by 3 main factors: **execution time**, **cpu usage**, and **memory usage**. Execution time is measured in seconds, cpu usage is measured in precentage of processor usage by the algorithm, and memory usage is measured in Kilobytes of RAM memory used by the algorithm. All measures are averaged over the 100 data sets in each data set configuration.

2.3 Type 1 Error Rate

Similar to the Power, the Type 1 Error Rate is determined by the amount of false positives in the 100 data sets within the most significant interactions, i.e. $\alpha < 0.05$.

3 Experimental Procedure

As mentioned in Lab Note LN-1-2014 [PC14a], there are 270 different configurations of data sets, with different parameters: allele frequency (0.01,0.05,0.1,0.3, and 0.5); population size (500,1000, and 2000); odds ratio (1.1,1.5, and 2.0); prevalence (0.0001 and 0.02); and disease model (Epistasis, Main effect, and Epistasis + Main Effect).
To test the Power and Type 1 Error Rate of algorithms, the outputs of each algorithm is gathered for each data set configuration and the corresponding confusion matrix is created. The output of each algorithm is filtered, selecting only interactions with a statistical relevancy of 5%. From these confusion matrices, the number true positives and false positives of data sets within each configuration is obtained and used as comparison for Power and Type 1 Error Rate respectively. For scalability, the built-in shell command *time* was used to obtain al the scalability measures for all algorithms.

4 Results

To compare each criteria of the algorithms, Table 2, 3, and 4 were created to represent the Power and Type 1 Error Rate of each algorithm, by number of individuals and allele frequency for epistasis, main effect and full effect respectively. Table 5 shows the results of the scalability measures used to evaluate each algorithm.
For epistasis detection, we can see that, for data sets with 500 individuals, no algorithm has a Power above 26%. This shows ta big dificulty in detecting epistasis with few individuals. The algorithm with best Power for these data sets is BOOST, followed by TEAM and SNPRuler and SNPHarvester. In error rate however, the algorithm with the lowest values is SNPRuler, followed by TEAM, SNPHarvester and BOOST. For data sets with 1000 individuals, there is a big increase in Power in all algorithms, reaching a maximum of 91%. BOOST has the best Power in all allele frequencies, followed by TEAM, SNPRuler and SNPHarvester. SNPRuler is once again the algorithm with the lowest Type 1 Error Rate, followed by TEAM, BOOST and SNPHarvester. In 2000 individuals, BOOST has the best Power with a maximum of 100%, followed by TEAM and SNPHarvester, with SNPRuler being better than SNPHarvester for 0.5 minor allele frequency. The lowest error rate is achieved by SNPRuler. Each of the other algorithms has a high Type 1 Error Rate in at least 1 setting. Screen and Clean is clearly the worse algorithm due to its lack of Power and high Type 1 Error Rate across all data set sizes. The Power shows an increase with allele frequency in each

algorithm, reaching their maximum Power in 0.5 allele frequency. There is no clear correlation between error rate and allele frequency for any algorithm.

POP	500 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BOOST	0%	4%	0%	7%	1%	7%	14%	6%	26%	4%
SnC	0%	15%	0%	15%	0%	14%	0%	19%	0%	18%
SNPH	0%	4%	0%	4%	0%	7%	4%	3%	2%	4%
SNPR	0%	0%	0%	0%	0%	0%	3%	0%	6%	0%
TEAM	0%	0%	0%	0%	0%	2%	6%	0%	8%	1%
POP	1000 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BOOST	0%	7%	0%	4%	41%	5%	66%	2%	91%	8%
SnC	0%	17%	0%	22%	0%	16%	0%	15%	0%	22%
SNPH	0%	4%	0%	13%	21%	9%	43%	9%	14%	3%
SNPR	0%	0%	0%	0%	10%	0%	35%	0%	71%	0%
TEAM	0%	2%	1%	4%	21%	5%	47%	1%	65%	0%
POP	2000 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BOOST	0%	2%	7%	2%	94%	21%	100%	6%	100%	8%
SnC	0%	18%	0%	20%	6%	21%	2%	16%	0%	14%
SNPH	0%	2%	18%	27%	85%	19%	70%	11%	33%	5%
SNPR	0%	0%	0%	1%	32%	8%	44%	0%	92%	0%
TEAM	0%	1%	43%	37%	92%	28%	92%	10%	95%	1%

Table 2: This table contains the results for epistasis detection. A comparison between the tested algorithms: BOOST, Screen and Clean, SNPHarvester, SNPRuler, and TEAM. The table is organized by population size (POP) and minor allele frequency (MAF), with an odds ratio of 2.0 and a prevalence of 0.02. For each allele frequency, there are two columns: the Power (P) obtained, and the Type 1 Error Rate (T1ER).

In main effect detection, for 500 individuals, the best algorithm is BEAM3, closely followed by BOOST, SNPHarvester and Screen and Clean far behind. The Type 1 Error Rate is lowest in MBMDR, and Screen and Clean, with BEAM3, SNPHarvester, and BOOST very close to each other, with very high error rates, BOOST having the highest error rate. for 1000 individu-

als, BOOST has better Power than BEAM3, followed by SNPHarvester, and Screen and Clean with MBMDR far behind. The Type 1 Error Rate is higher for BOOST, very closely followe by BEAM3, SNPHarvester, and Screen and Clean, with MBMDR having the lowest error rate. For data sets with 2000 individuals, BOOST and MBMDR have a better Power for data sets with allele frequency lower than 0.1, and BEAM3, BOOST and SNPHarvester equally good in allele frequencies higher than 0.1. The error rate is lower generally for MBMDR, followed by Scren and Clean.

Table 4 shows the full effect detection for BOOST, Screen and Clean, and SNPHarvester. BOOST and SNPHarvester have the highest Power detection for all allele frequencies but have a high Type 1 Error Rate. Screen and Clean has high Power for high allele frequencies, but 0 for configurations belo 0.3 and with a higher Type 1 Error Rate for configurations below 0.1. Screen and Clean has the lowest Type 1 Error Rate but also has the worst Power detection. BOOST has the best ratio of Power to Type 1 Error Rate.

Table 5 shows the running time, CPU usage and memory usage of all algorithms for scalability measure. Screen and Clean is the slowest recorded algorithm, followed by SNPHarvester, TEAM, BEAM3 and SNPRuler, with BOOST being the fastest algorithm. Screen and Clean also has the highest increase in running time, followed by SNPHarvester, TEAM, with BOOST, BEAM3, and SNPRuler far behind. SNPRuler is the algorithm with the highest CPU usage, having to resort to more than 1 core to finish each task. SNPHarvester, BOOST, BEAM3, and Screen and Clean are all close to 100%, with TEAM being the algorithm with the least required CPU usage. BEAM3, BOOST, and TEAM have an increase of CPU usage with data set size, TEAM being the algorithm with the highest increase. In memory usage, SNPRuler shows the highest usage of memory, closely followed by TEAM, Screen and Clean, SNPHarvester, BEAM3, and finally BOOST far behind.

POP	500 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BEAM3	0%	0%	0%	3%	0%	9%	100%	71%	100%	99%
BOOST	0%	1%	0%	1%	2%	12%	100%	78%	100%	97%
MBMDR	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%
SnC	0%	14%	0%	17%	0%	21%	20%	23%	54%	15%
SNPH	0%	1%	0%	5%	0%	11%	100%	78%	100%	99%
POP	1000 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BEAM3	0%	6%	0%	3%	32%	18%	100%	99%	100%	100%
BOOST	0%	7%	1%	3%	43%	23%	100%	99%	100%	100%
MBMDR	0%	0%	0%	2%	2%	0%	7%	0%	12%	0%
SnC	0%	14%	0%	21%	0%	23%	54%	28%	70%	30%
SNPH	0%	10%	0%	4%	38%	22%	100%	99%	100%	100%
POP	2000 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BEAM3	0%	1%	1%	17%	92%	67%	100%	100%	100%	100%
BOOST	0%	1%	14%	11%	97%	74%	100%	100%	100%	100%
MBMDR	0%	0%	14%	6%	54%	3%	71%	2%	85%	0%
SnC	0%	13%	0%	22%	39%	36%	58%	38%	62%	48%
SNPH	0%	1%	1%	24%	92%	79%	100%	100%	100%	100%

Table 3: This table contains the results for main effect detection. A comparison between the tested algorithms: BEAM3, BOOST, Screen and Clean, and SNPHarvester. The table is organized by population size (POP) and minor allele frequency (MAF), with an odds ratio of 2.0 and a prevalence of 0.02. For each allele frequency, there are two columns: the Power (P) obtained, and the Type 1 Error Rate (T1ER).

POP	500 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BOOST	0%	10%	0%	4%	1%	15%	100%	100%	100%	100%
SnC	0%	18%	0%	15%	0%	19%	30%	19%	49%	37%
SNPH	0%	2%	0%	8%	0%	9%	100%	100%	100%	100%
POP	1000 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BOOST	0%	11%	2%	16%	42%	38%	100%	100%	100%	100%
SnC	0%	14%	0%	21%	0%	28%	58%	35%	73%	45%
SNPH	0%	4%	0%	8%	32%	27%	100%	100%	100%	100%
POP	2000 individuals									
MAF	0.01		0.05		0.1		0.3		0.5	
	P	T1ER	P	T1ER	P	T1ER	P	T1ER	P	T1ER
BOOST	0%	7%	15%	17%	98%	81%	100%	100%	100%	100%
SnC	0%	14%	0%	21%	0%	33%	40%	68%	91%	84%
SNPH	0%	1%	0%	20%	95%	79%	100%	100%	100%	100%

Table 4: This table contains the results for full effect detection. A comparison between the tested algorithms: BOOST, Screen and Clean, and SNPHarvester. The table is organized by population size (POP) and minor allele frequency (MAF), with an odds ratio of 2.0 and a prevalence of 0.02. For each allele frequency, there are two columns: the Power (P) obtained, and the Type 1 Error Rate (T1ER).

	Running Time (s)			CPU Usage(%)			Memory Usage (MB)		
	500	1000	2000	500	1000	2000	500	1000	2000
BEAM3	4.9	7	8	87.8	96.3	95.5	4	4.3	5.8
BOOST	0.16	0.22	0.34	95.7	98.79	97.87	0.98	1	1.2
MBMDR*	–	–	–	–	–	–	–	–	–
SnC	8.05	18.65	34.65	75.7	98.99	77.25	129.8	137.2	152.5
SNPHarvester	9.29	25.89	33	102.1	86.5	101.6	68.35	71.3	76.86
SNPRuler	2.7	3.09	4.1	130.2	141.9	156.28	312.7	316	320.2
TEAM	3.28	5.28	9.81	66.99	69.71	74.75	162.7	176	228.1

Table 5: Scalability test containing the average running time, CPU usage, and memory usage by data set population size. *MBMDR does not contain scalability results because these were obtained from different computers with different hardware settings from all other results. The data sets have a minor allele frequency is 0.5, 2.0 odds ratio, 0.02 prevalence.

5 Results Discussion

The results obtained from all the different algorithms show interesting qualities among them. BOOST is clearly the algorithm with the highest Power, but has high Type 1 Error Rate. SNPRuler has low Type 1 Error Rate, but not very high Power and only works for epistasis detection. Screen and Clean is ineffective in most settings, but has a relatively low Type 1 Error Rate and high Power for main effect and full detection in data sets with high allele frequency. BEAM3 only works for main effect detection but has high Power with slightly lower error rate than BOOST. SNPHarvester has low Power, but also low Type 1 Error Rate in all model types. MBMDR has good Power for certain configurations and a very low Type 1 Error Rate, however it has a very high running time for each data set. TEAM has good Power, with slightly high Type I Error Rate in certain configurations.
BOOST is the most scalable algorithm, followed by SNPRuler and BEAM3. This is specially important for large data sets and their ability to work in an ensemble approach. In epistasis detection, considering the Power, Screen and Clean and SNPHarvester show the worse potential. For main effect, The Power is lowest for Scren and Clean and MBMDR. For full effect, Screen and Clean is once again the weakest algorithm.
With this information, the best algorithms for each scenario can be used together to maximize Power and lower Type 1 Error Rate.

References

[MVV11] Jestinah M Mahachie John, Francois Van Lishout, and Kristel Van Steen. Model-Based Multifactor Dimensionality Reduction to detect epistasis for quantitative traits in the presence of error-free and noisy data. *Eur J Hum Genet*, 19(6):696–703, June 2011.

[PC14a] Ricardo Pinho and Rui Camacho. Genetic Epistasis I - Materials and methods. 2014.

[PC14b] Ricardo Pinho and Rui Camacho. Genetic Epistasis II - Assessing Algorithm BEAM 3.0. 2014.

[PC14c] Ricardo Pinho and Rui Camacho. Genetic Epistasis III - Assessing Algorithm BOOST. 2014.

[PC14d] Ricardo Pinho and Rui Camacho. Genetic Epistasis IV - Assessing Algorithm Screen and Clean. 2014.

[PC14e] Ricardo Pinho and Rui Camacho. Genetic Epistasis V - Assessing Algorithm SNPRuler. 2014.

[PC14f] Ricardo Pinho and Rui Camacho. Genetic Epistasis VI - Assessing Algorithm SNPHarvester. 2014.

[PC14g] Ricardo Pinho and Rui Camacho. Genetic Epistasis VII - Assessing Algorithm TEAM. 2014.

[PC14h] Ricardo Pinho and Rui Camacho. Genetic Epistasis VIII - Assessing Algorithm MBMDR. 2014.

[SZS+11] Junliang Shang, Junying Zhang, Yan Sun, Dan Liu, Daojun Ye, and Yaling Yin. Performance analysis of novel methods for detecting epistasis, 2011.

[WDR+10] Jing Wu, Bernie Devlin, Steven Ringquist, Massimo Trucco, and Kathryn Roeder. Screen and clean: a tool for identifying interactions in genome-wide association studies. *Genetic epidemiology*, 34:275–285, 2010.

[WYY+10a] Xiang Wan, Can Yang, Qiang Yang, Hong Xue, Xiaodan Fan, Nelson L S Tang, and Weichuan Yu. BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. *American journal of human genetics*, 87:325–340, 2010.

[WYY+10b] Xiang Wan, Can Yang, Qiang Yang, Hong Xue, Nelson L S Tang, and Weichuan Yu. Predictive rule inference for epistatic interaction detection in genome-wide association studies. *Bioinformatics (Oxford, England)*, 26:30–37, 2010.

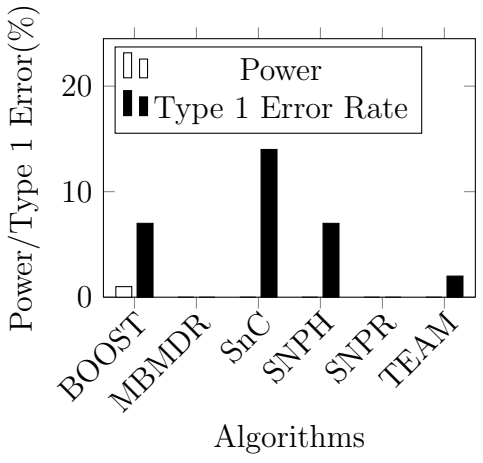
[YHW+09] Can Yang, Zengyou He, Xiang Wan, Qiang Yang, Hong Xue, and Weichuan Yu. SNPHarvester: a filtering-based approach for detecting epistatic interactions in genome-wide association studies. *Bioinformatics (Oxford, England)*, 25:504–511, 2009.

[Zha12] Yu Zhang. A novel bayesian graphical model for genome-wide multi-SNP association mapping. *Genetic Epidemiology*, 36:36–47, 2012.

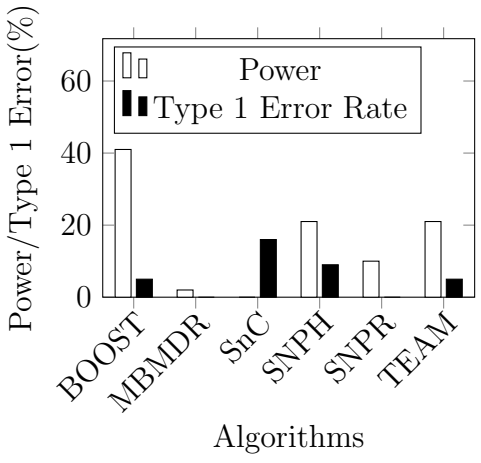
[ZHZW10] Xiang Zhang, Shunping Huang, Fei Zou, and Wei Wang. TEAM: efficient two-locus epistasis tests in human genome-wide association study. *Bioinformatics (Oxford, England)*, 26:i217–i227, 2010.

A Bar Graphs

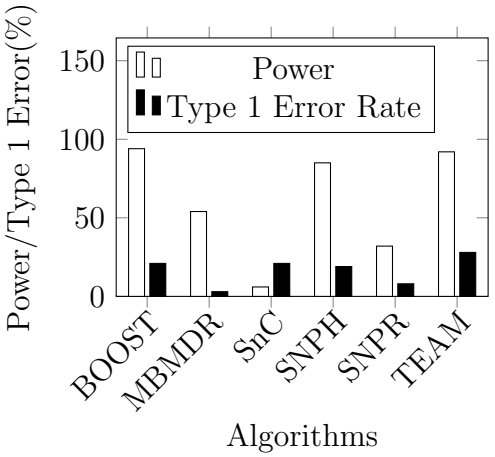
A.1 Population size



(a) 500 individuals.

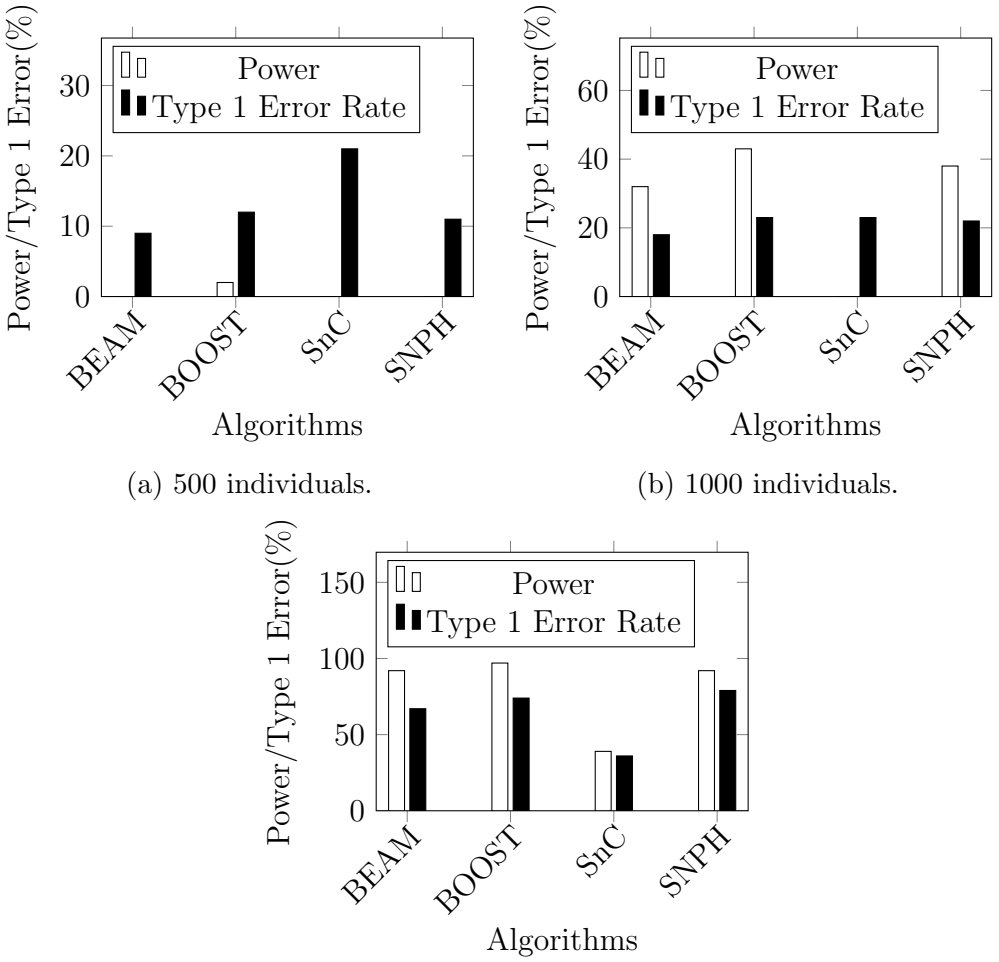


(b) 1000 individuals.



(c) 2000 individuals.

Figure 1: These results correspond to epistasis detection by population size, with a 0.1 minor allele frequency, 2.0 odds ratio, and 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 500 individuals (a), 1000 individuals (b), and 2000 individuals (c).

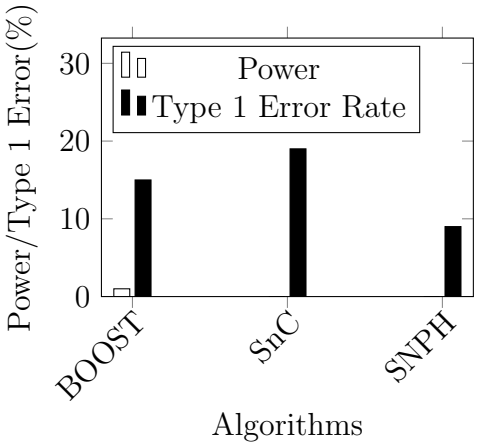


(a) 500 individuals.

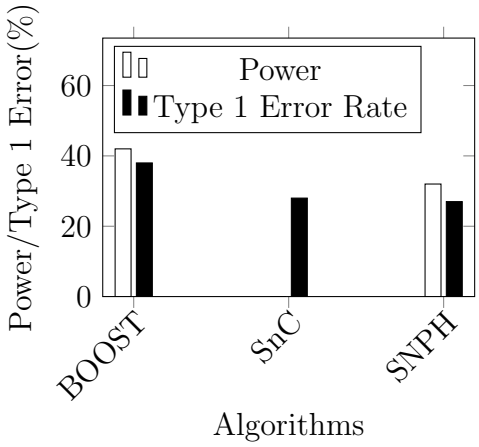
(b) 1000 individuals.

(c) 2000 individuals.

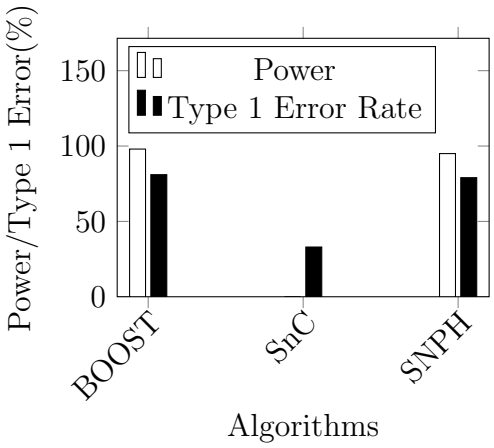
Figure 2: These results correspond to main effect detection by population size, with a 0.1 minor allele frequency, 2.0 odds ratio, and 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 500 individuals (a), 1000 individuals (b), and 2000 individuals (c).



(a) 500 individuals.



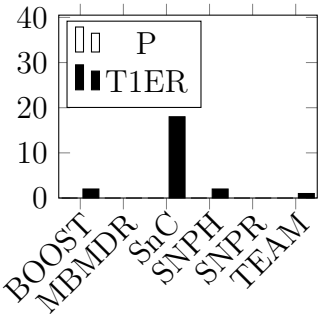
(b) 1000 individuals.



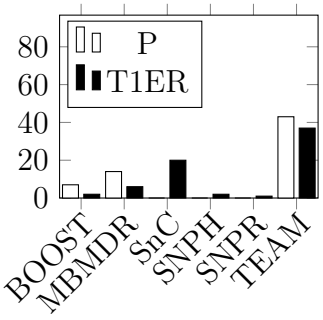
(c) 2000 individuals.

Figure 3: These results correspond to full effect detection by population size, with a 0.1 minor allele frequency, 2.0 odds ratio, and 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 500 individuals (a), 1000 individuals (b), and 2000 individuals (c).

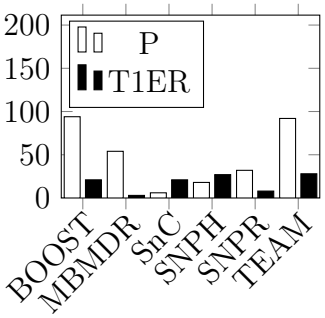
A.2 Frequency



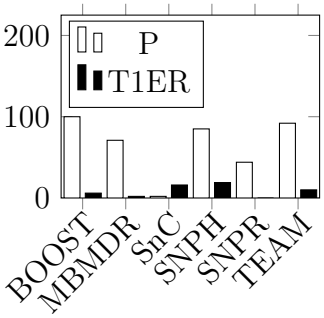
(a) 0.01 allele frequency.



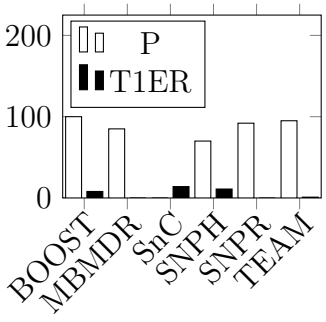
(b) 0.05 allele frequency.



(c) 0.1 allele frequency.

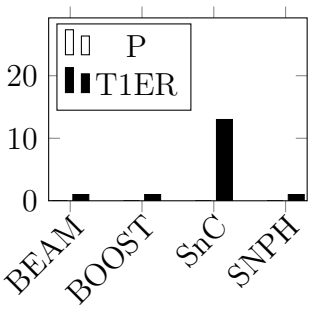


(d) 0.3 allele frequency.

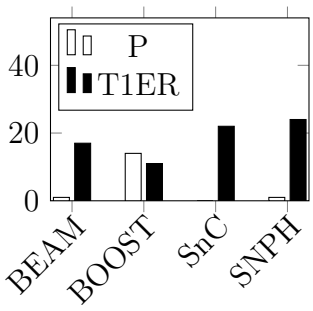


(e) 0.5 allele frequency.

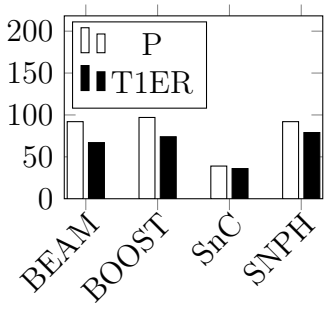
Figure 4: These results correspond to epistasis detection by minor allele frequency, with 2000 individuals, 2.0 odds ratio, and 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 0.01 (a), 0.05 (b), 0.1 (c), 0.3 (d), and 0.5 (e) allele frequencies.



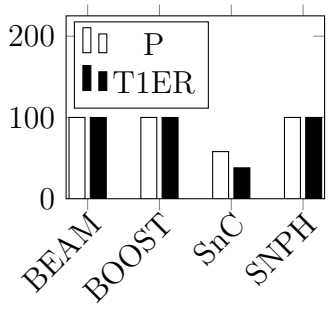
(a) 0.01 allele frequency.



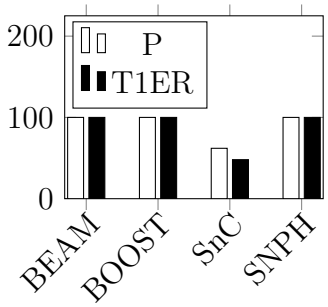
(b) 0.05 allele frequency.



(c) 0.1 allele frequency.

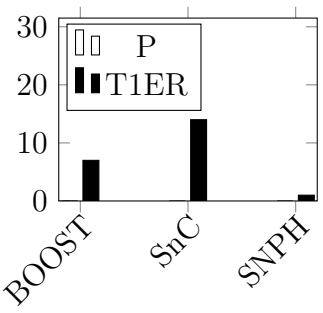


(d) 0.3 allele frequency.

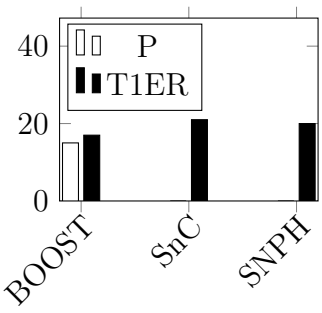


(e) 0.5 allele frequency.

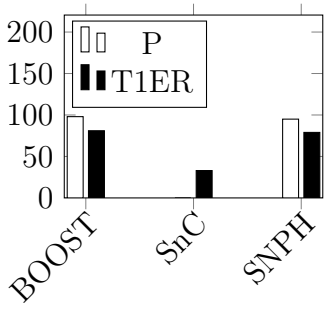
Figure 5: These results correspond to main effect detection by minor allele frequency, with 2000 individuals, 2.0 odds ratio, and 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 0.01 (a), 0.05 (b), 0.1 (c), 0.3 (d), and 0.5 (e) allele frequencies.



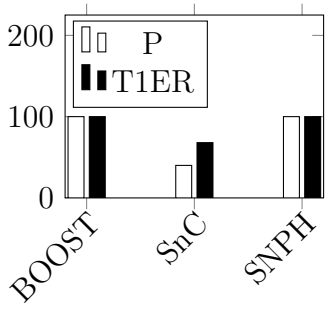
(a) 0.01 allele frequency.



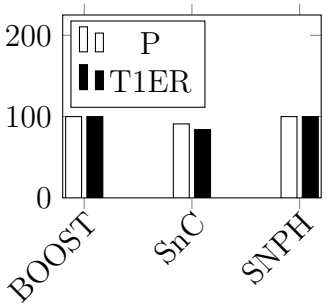
(b) 0.05 allele frequency.



(c) 0.1 allele frequency.



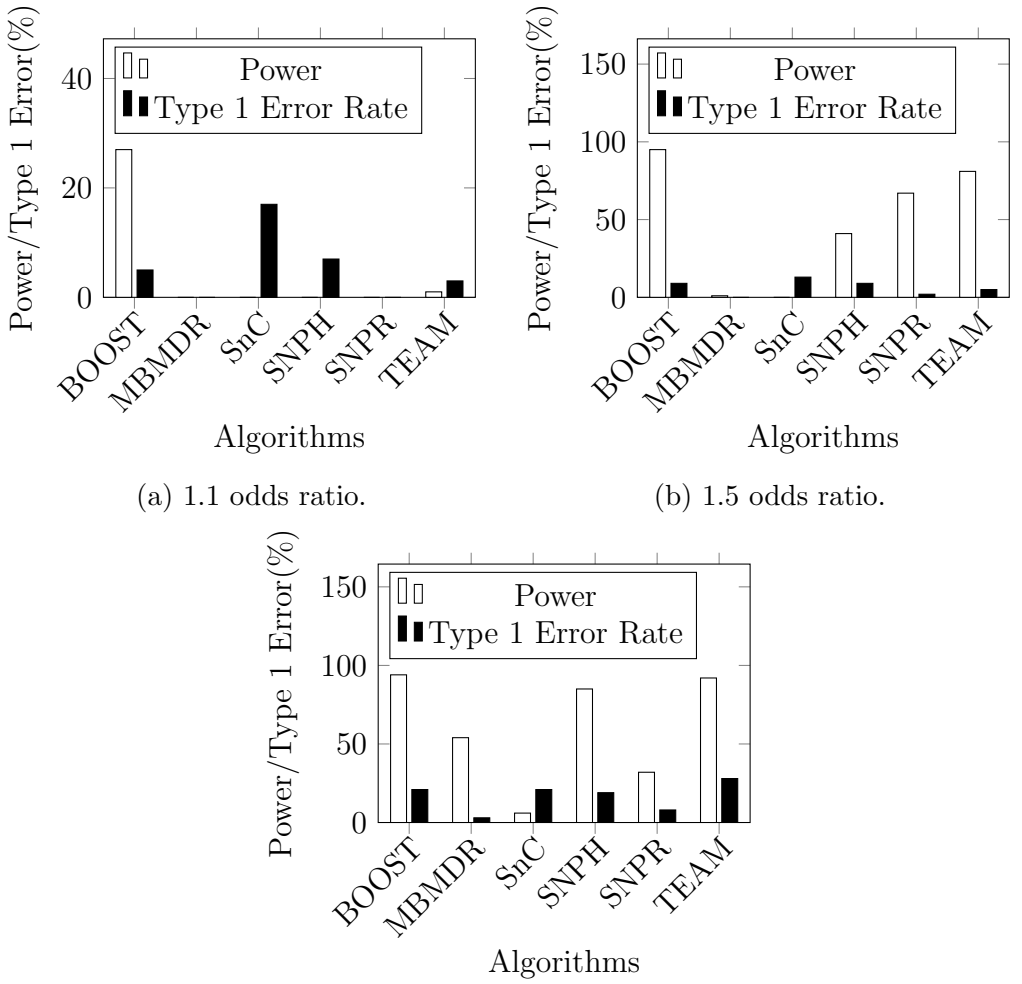
(d) 0.3 allele frequency.



(e) 0.5 allele frequency.

Figure 6: These results correspond to full effect detection by minor allele frequency, with 2000 individuals, 2.0 odds ratio, and 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 0.01 (a), 0.05 (b), 0.1 (c), 0.3 (d), and 0.5 (e) allele frequencies.

A.3 Odds Ratio

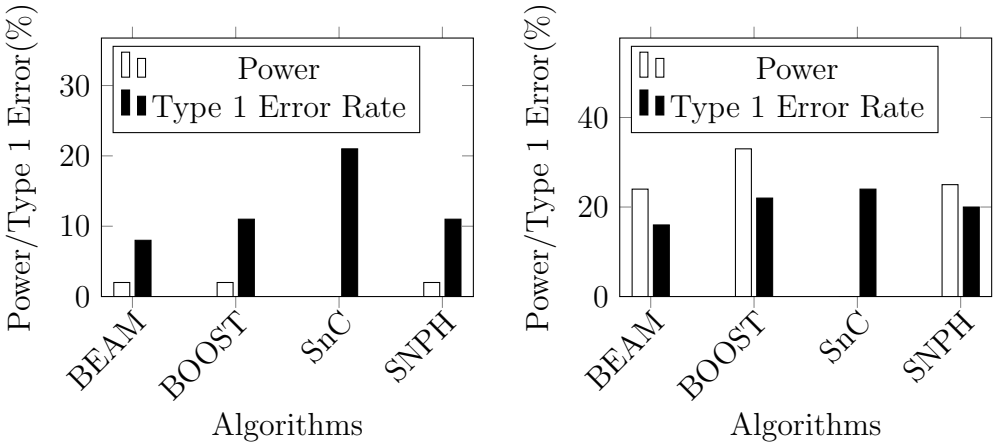


(a) 1.1 odds ratio.

(b) 1.5 odds ratio.

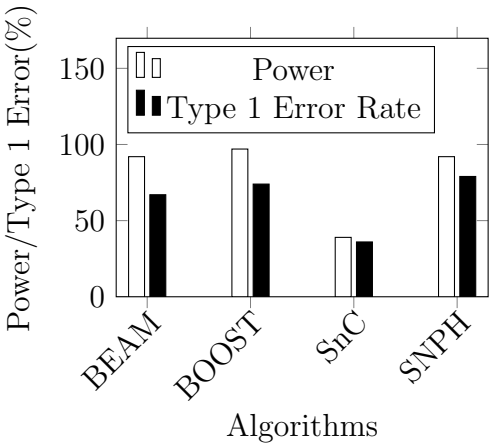
(c) 2.0 odds ratio.

Figure 7: These results correspond to epistatic detection by odds ratio, with a minor allele frequency of 0.1, 2000 individuals, and a 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 1.1 (a), 1.5 (b), and 2.0 (c) odds ratio.



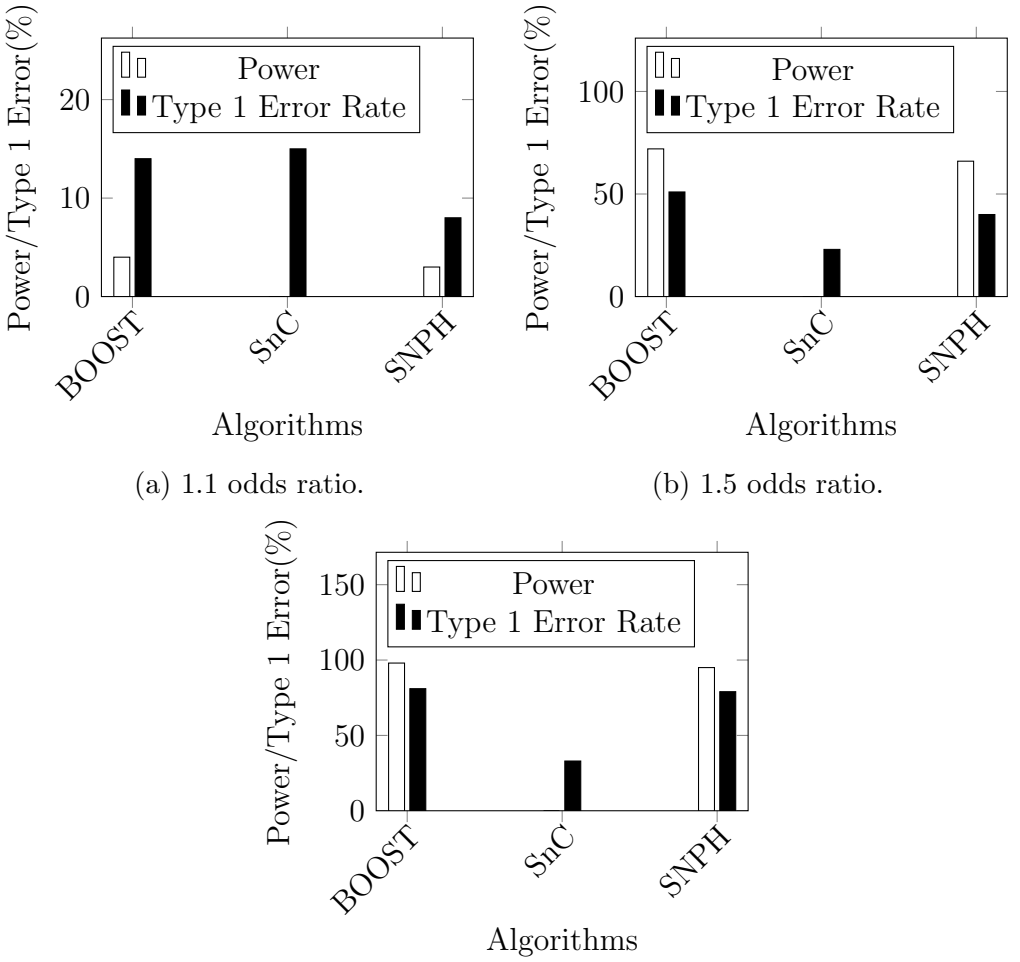
(a) 1.1 odds ratio.

(b) 1.5 odds ratio.



(c) 2.0 odds ratio.

Figure 8: These results correspond to main effect detection by odds ratio, with a minor allele frequency of 0.1, 2000 individuals, and a 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 1.1 (a), 1.5 (b), and 2.0 (c) odds ratio.



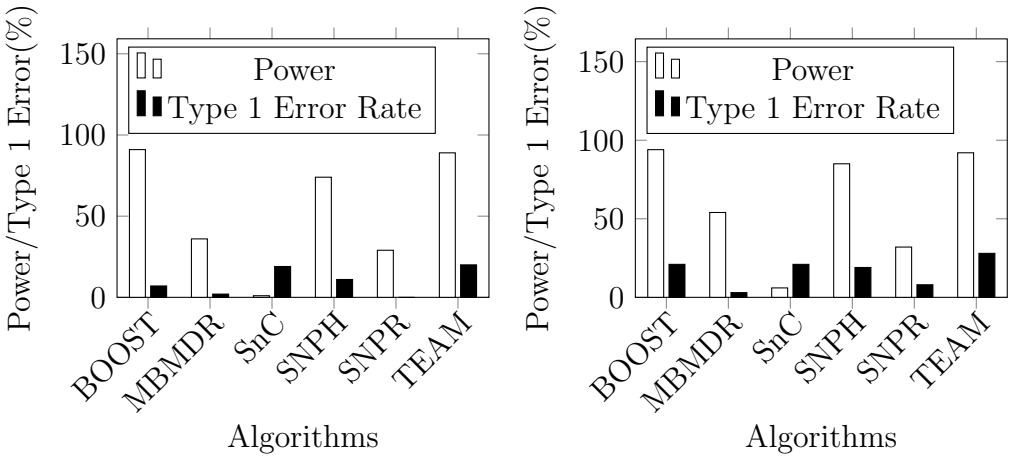
(a) 1.1 odds ratio.

(b) 1.5 odds ratio.

(c) 2.0 odds ratio.

Figure 9: These results correspond to full effect detection by odds ratio, with a minor allele frequency of 0.1, 2000 individuals, and a 0.02 prevalence. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 1.1 (a), 1.5 (b), and 2.0 (c) odds ratio.

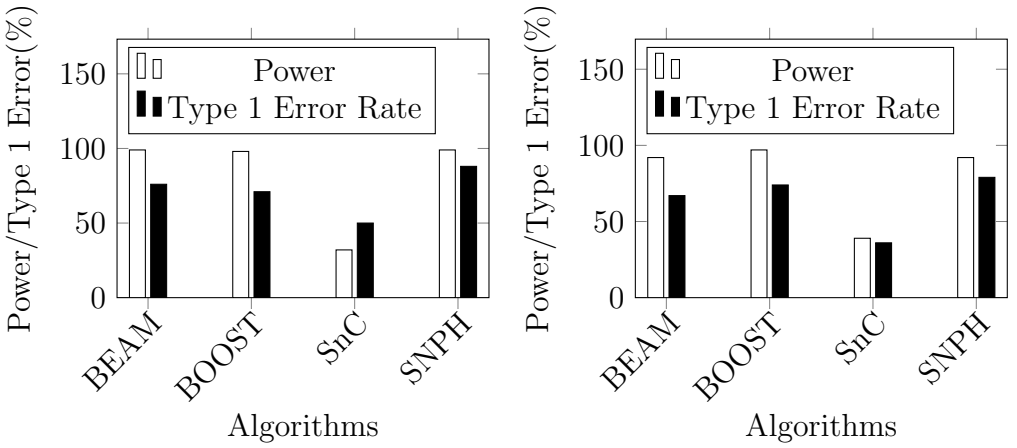
A.4 Prevalence



(a) 0.0001 prevalence.

(b) 0.02 prevalence.

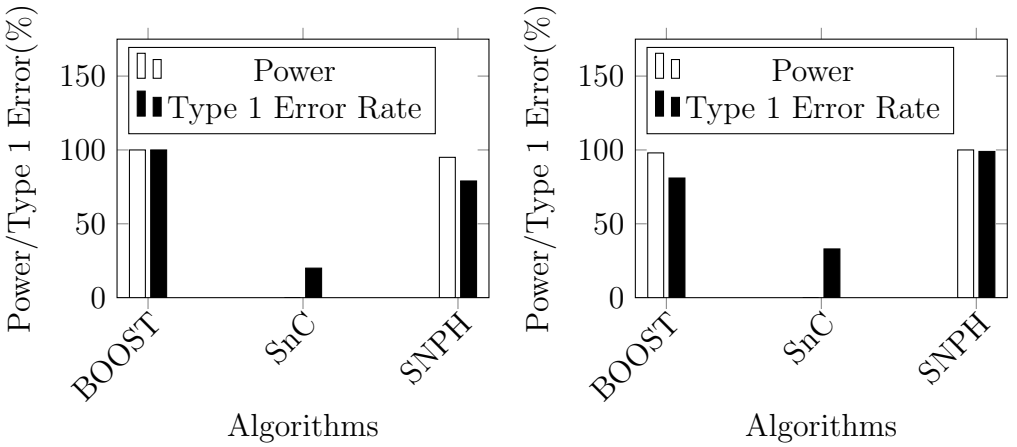
Figure 10: These results correspond to epistasis detection by prevalence, with a minor allele frequency of 0.1, 2000 individuals, and a 2.0 odds ratio. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 0.0001 (a), and 0.02 (b) prevalence.



(a) 0.0001 prevalence.

(b) 0.02 prevalence.

Figure 11: These results correspond to main effect detection by prevalence, with a minor allele frequency of 0.1, 2000 individuals, and a 2.0 odds ratio. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 0.0001 (a), and 0.02 (b) prevalence.



(a) 0.0001 prevalence.

(b) 0.02 prevalence.

Figure 12: These results correspond to full effect detection by prevalence, with a minor allele frequency of 0.1, 2000 individuals, and a 2.0 odds ratio. The results of the Power and Type 1 Error Rate of BOOST, MBMDR, Screen and Clean, SNPHarvester, SNPRuler and TEAM. Each subfigure contains the values for all algorithms in data sets with 0.0001 (a), and 0.02 (b) prevalence.