

AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm

Yupeng Wang^{1,2}, Xinyu Liu^{1,2}, Kelly Robbins¹ and Romdhane Rekaya^{*1,2,3}

Abstract

Background: Epistatic interactions of multiple single nucleotide polymorphisms (SNPs) are now believed to affect individual susceptibility to common diseases. The detection of such interactions, however, is a challenging task in large scale association studies. Ant colony optimization (ACO) algorithms have been shown to be useful in detecting epistatic interactions.

Findings: AntEpiSeeker, a new two-stage ant colony optimization algorithm, has been developed for detecting epistasis in a case-control design. Based on some practical epistatic models, AntEpiSeeker has performed very well.

Conclusions: AntEpiSeeker is a powerful and efficient tool for large-scale association studies and can be downloaded from <http://nce.ads.uga.edu/~romdhane/AntEpiSeeker/index.html>.

Background

Genetic association studies, which aim at detecting association between one or more genetic polymorphisms and a trait of interest such as a quantitative characteristic, discrete attribute or disease, have gained a lot of popularity in the past decade [1]. Although great progress in mapping genes responsible for Mendelian traits has been made, the genetic basis underlying many complex diseases remain unknown. It is widely accepted that these diseases may be caused by the joint effects of multiple genetic variations, which may show little effect individually but strong interactions. Such interactive effects of multiple genetic variations are often referred to as epistasis or epistatic interactions [2]. Recently, increasing numbers of studies have suggested the presence of epistatic interactions in complex diseases, e.g. breast cancer [3], type-2 diabetes [4] and atrial fibrillation [5].

A number of multi-locus approaches have been proposed to detect epistatic interactions, such as the combinatorial partitioning method (CPM) [6], restricted partitioning method (RPM) [7], the multifactor-dimensionality reduction (MDR) [3], the focused interaction

testing framework (FITF) [8] and the backward genotype-trait association (BGTA) [9]. Although these methods were tested and showed promising performance on small data sets, the computational burden prohibits their application on large scale datasets.

Typically, a large scale dataset for association studies may have several tens to hundreds of thousands of markers. For example, the genome-wide case-control data set for Age-related Macular Degeneration (AMD) contains more than 100 thousand SNPs genotyped on 96 cases and 50 controls [10]. An exhaustive search of two-locus interactions needs to evaluate at least 5.00×10^9 locus combinations, and this number increases to 1.67×10^{14} when three-locus interactions are considered. Although this process is computationally hard it could be enhanced by two recent approaches: the Bayesian epistasis association mapping (BEAM) [11] and SNPharvester [12], which were shown to be able to handle large scale datasets. However, more efficient and accurate methods are still desired.

The solution to this difficult search problem could be achieved using an optimization technique called ant colony optimization (ACO) algorithm. Ant colony algorithms, proposed first by Dorigio and Gambardella [13], are tools to solve difficult optimization problems such as the traveling salesman problem. ACO simulates how real

* Correspondence: rrekaya@uga.edu

¹ Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA

Full list of author information is available at the end of the article

ant colonies find the shortest route to a food source. Real ant colonies communicate through chemicals called pheromones, which are deposited along the path an ant travels. Ants that choose a shorter path will transverse the distance at a faster rate, resulting in more pheromones deposited along that path. Subsequent ants will then choose the path with more pheromones, thus creating a positive feedback. In ACO, artificial ants work as parallel units that communicate through a probability distribution function (PDF), which is updated by weights or pheromones. The change in pheromones is determined by some type of expert knowledge. As the PDF is updated, "paths" that perform better will be sampled at higher rates by subsequent artificial ants, and in turn deposit more pheromones. Thus, a positive feedback similar to real ant colonies is simulated.

Two recent studies showed the possibility of applying ant colony optimization to association studies [14,15]. However, the use of MDR for detecting epistatic interactions in these studies dramatically increased the computational burden. Besides, these studies did not test performance using the more practical epistatic models such as the ones proposed by Marchini et al. [16].

In this study, a new tool named AntEpiSeeker has been developed to search for epistatic interactions in large-scale association studies. The use of χ^2 values as score function to measure the association between an SNP set and the phenotype is computationally efficient. The two-stage design of ant colony optimization and the idea of searching bigger SNP sets harboring epistatic interactions enhance the power of ACO algorithms. AntEpiSeeker showed improved performance based on some practical epistatic models and large scale datasets.

Methods

The generic ant colony optimization

The ACO has been proven to be a successful technique for numerous non-deterministic polynomial-time hard (NP-hard) combinatorial optimization problems such as the traveling salesman problem, the graph coloring problem, the frequency assignment problem, the quadratic assignment problem, feature selection for microarray classification and the vehicle routing problem [17-22]. ACO has the advantages of a positive feedback, and it lends itself to parallel computing, among other advantages.

As defined by Dorigio and Gambardella [13], ACO is comprised of parallel artificial ants that communicate through a probability density function (PDF) that is updated by weights or 'pheromone levels'. In this case, the ACO is an iterative procedure which stops at a pre-defined number of iterations and the weights are determined by the significance of the epistatic interaction of

the selected set of SNPs. The probability of selecting locus k at iteration i is defined as:

$$p_k(i) = \frac{(\tau_k(i))^\alpha \eta_k^\beta}{\sum_{j=1}^L (\tau_j(i))^\alpha \eta_j^\beta} \quad (1)$$

where $\tau_k(i)$ is the amount of pheromones for locus k at iteration i ; η_k^β is some form of prior information, which is set to 1 in this study as we treat each locus equally; α is the parameter determining the weight given to the pheromones deposited by ants. The ACO is initialized with all loci having an equal level of pheromone τ_0 . Using the PDF defined in equation (1), each artificial ant, m , will select an SNP set S_m of n loci from the whole set of genomic SNPs. The epistatic interaction for this SNP set is evaluated by the χ^2 test. The pheromone level of each locus k in S_m is then updated, based on the performance of S_m , as:

$$\tau_k(i+1) = (1-\rho)\tau_k(i) + \Delta\tau_k(i), \quad (2)$$

where ρ is a constant between 0 and 1 that represents the pheromone evaporation rate; $\Delta\tau_k(i)$ is the change in pheromone level for locus k at iteration i , which equals $0.1 \chi^2$ of S_m in this study, and is set to zero if locus $k \notin S_m$. This process is repeated for all artificial ants.

AntEpiSeeker algorithms

In an effort to increase the detection power of the generic ACO as outlined in the previous section, an advanced ACO algorithm called AntEpiSeeker, which employs a two-stage design of ACO, is proposed. The first stage of AntEpiSeeker searches SNP sets of sufficient size (larger than the number of SNPs in a given epistatic interaction) using the above ACO, which results in a pre-defined number of highly suspected SNP sets determined by χ^2 scores, and another SNP set of a pre-defined size, determined by pheromone levels. The second stage of AntEpiSeeker conducts exhaustive search of epistatic interactions within the highly suspected SNP sets, and within the reduced set of SNPs with top ranking pheromone levels. The use of highly suspected SNP sets (much smaller than the available SNPs in the data) enhances the power of detecting pure epistasis based on greatly reduced computational cost and the SNP set with top ranking pheromone levels is used to detect epistasis among the SNPs with big marginal effects. Additionally, we suggest two rounds of search: 1) using a relatively large size SNP set, which is sensitive to strong signals,

AntEpiSeeker algorithm

Input parameters

D: a dataset of *N* case and control samples genotyped at *L* loci
iEpiModel: number of SNPs in an epistatic interaction
Pvalue: statistical significance threshold
largesetsize: size of the large SNP sets
smallsetsize: size of the small SNP sets
iAntCount: number of ants
iItCountLarge, *iItCountSmall*: number of iterations for large or small SNP sets

τ_0 : initial pheromone level for each locus
 ρ : evaporation rate
 α : parameter determining the weight given to pheromone deposited by ants

for *setsize* in (*largesetsize*, *smallsetsize*)

 If *setsize*==*largesetsize*, *iItCount*=*iItCountLarge*; else, *iItCount*=*iItCountSmall*

 Initialize the pheromone level of each locus

i=0

 Iteration begins

 Each ant selects a SNP set with *setsize* loci sequentially according to equation (1)

 Calculate the χ^2 test for each SNP set and update the pheromone for each locus in the SNP set

 Record the SNP sets with the highest χ^2 test

 Clear ants

i=*i*+1

 Iteration ends if *i*==*iItCount*

Post-processing: conduct exhaustive search of epistatic interactions of a desired model within the recorded SNP sets with the highest χ^2 test and the loci with top ranking pheromone levels

End for

Optional: minimize false positives

Return epistatic interactions with p-values

Figure 1 Pseudocode for AntEpiSeeker.

and 2) using a relatively small size SNP set, which is sensitive to weak signals. The pseudocode for AntEpiSeeker is shown in Figure 1.

Minimizing false positives

AntEpiSeeker may report all detected epistatic interactions at a p-value threshold. In addition, AntEpiSeeker

incorporates a procedure for minimizing false positives, which can be described as:

1) The set of all detected epistatic interactions is denoted by EI_{all} and another null set, holding the epistatic interactions with minimized false positives, is denoted by EI_m .

2) Each reported epistatic interaction I_i in EI_{all} is attempted to be added into EI_m sequentially. If I_i does not have any locus overlapping with those of each epistatic interaction in EI_m , I_i is added to EI_m . Otherwise, assuming that the epistatic interaction J_j in EI_m has at least one locus overlapping with those of I_i , determine if the p-value of I_i is smaller than that of J_j . If so, J_j in EI_m is replaced by I_i . If not, I_i is not reported in EI_m .

The software

The AntEpiSeeker package was written in C++. Before compiling, the GNU Scientific Library (GSL) needs to be installed on the user's computer. A separate parameter file named "parameters.txt" specifies the parameters needed to run the program. The SNP data file should be comma-delimited, with the first row specifying the SNP names. All subsequent rows should contain SNP data for each sample. The SNP data should be coded by 0, 1 and 2. The last column indicates the sample status (0 indicates control and 1 indicates case). There are three output files. "AntEpiSeeker.log" records some intermediate results, "results_maximized.txt" reports all detected epistatic interactions, and the user-specified output file shows the epistatic interactions with minimized false positives. The user specified output file includes the locus name, χ^2 value and p-value. The software and its source code are available for download at <http://nce.ads.uga.edu/~romd-hane/AntEpiSeeker/index.html>.

Parameter Setting

The parameters needed to run AntEpiSeeker include $iAntCount$, $iItCountLarge$, $iItCountSmall$, α , $iTopModel$, $iTopLoci$, ρ , τ_0 , $largesetsize$, $smallsetsize$, $iEpiModel$, $pvalue$, $INPFILE$, $OUTFILE$. The parameter " $iEpiModel$ " specifies the number of SNPs in an epistatic interaction. The parameters " $largesetsize$ ", " $smallsetsize$ " must be greater than " $iEpiModel$ ". For a two-locus interaction model, we suggest $largesetsize = 6$, $smallsetsize = 3$, $iEpiModel = 2$; For a three-locus interaction model, we suggest $largesetsize = 6$, $smallsetsize = 4$, $iEpiModel = 3$. The parameters " $iItCountLarge$ ", " $iItCountSmall$ " should be chosen according to the number of SNPs genotyped in the data (Denoted by L). Typically, we suggest $iItCountSmall \geq 0.1 \times L$ and $iItCountLarge = 0.5 \times iItCountSmall$. $iAntCount$ may vary from 500 to 5,000, where larger $iAntCount$ should correspond to larger L . ρ should range from 0.01 to 0.1 for better performance, where smaller L should use larger ρ . The default parameters in the AntEpiSeeker package, used in our most simulation studies, were an optimal setting balanced between ρ and $iAntCount$, which should work well on medium size datasets ($2 \times 10^3 \leq L \leq 2 \times 10^4$).

Results

Power and computational time evaluation on a simulated data set

The performance of AntEpiSeeker was evaluated by comparison with two recent methods, BEAM [11] and SNPHarvester [12], as well as the generic ACO algorithm, using simulated data generated by the simulation program provided in the BEAM package. Note that the generic ACO algorithm does not select SNP sets of bigger size, and thus the parameters " $largesetsize$, $smallsetsize$, $iItCountLarge$, $iItCountSmall$ " were not needed for the generic ACO algorithm. The simulation study was conducted following the procedure and parameter settings of many previous studies [11,12,16,23,24]. For each combination of parameter settings, 50 datasets containing 4,000 samples (2,000 cases and 2,000 controls) and 2000 SNPs were simulated. The detection power was calculated as the ratio of the number of successful identifications to the number of datasets at the significance level 0.01 after Bonferroni correction. Data was simulated following three genetic models: 1) additive model, 2) epistatic interactions with multiplicative effects and 3) epistatic interactions with threshold effects, as defined by Marchini et al. [16]. Other parameters for data simulation were the effective size λ (a measure of marginal effects as defined by Marchini et al. [16]), linkage disequilibrium between SNPs measured by r^2 and minor allele frequencies (MAFs). λ was set to 0.3 for Model 1 and 0.2 for Models 2 and 3. For r^2 , two values (0.7 and 1.0) were used for each model. For MAF, three values (0.1, 0.2, and 0.5) were considered. The parameters for BEAM were set as default. The parameter settings for SNPHarvester were: $1 \leq k \leq 5$ and $paths = 50$, as suggested by its original simulation study. The parameter settings for AntEpiSeeker were: $largesetsize = 6$, $smallsetsize = 3$, $iItCountLarge = 150$, $iItCountSmall = 300$, $iEpiModel = 2$, $iAntCount = 1000$, $\alpha = 1$, $\rho = 0.05$ and $\tau_0 = 100$ (also available in the software package of AntEpiSeeker). The parameters of the generic ACO algorithm were set as $iAntCount = 1000$, $\alpha = 1$, $\rho = 0.05$, $\tau_0 = 100$, $iItCount$ (number of iterations) = 900, $iEpiModel = 2$. The comparison of detection power for AntEpiSeeker, BEAM, SNPHarvester and the generic ACO is presented in Figure 2. The results show that AntEpiSeeker outperforms BEAM in all parameter settings and is superior to SNPHarvester and the generic ACO in most parameter settings.

In addition, AntEpiSeeker is computationally efficient. In the above simulation study, the average running time of AntEpiSeeker, SNPHarvester and BEAM were 27, 54 and 133 minutes respectively, using a Linux system based on Dual Core AMD Opteron(tm) Processor 275.

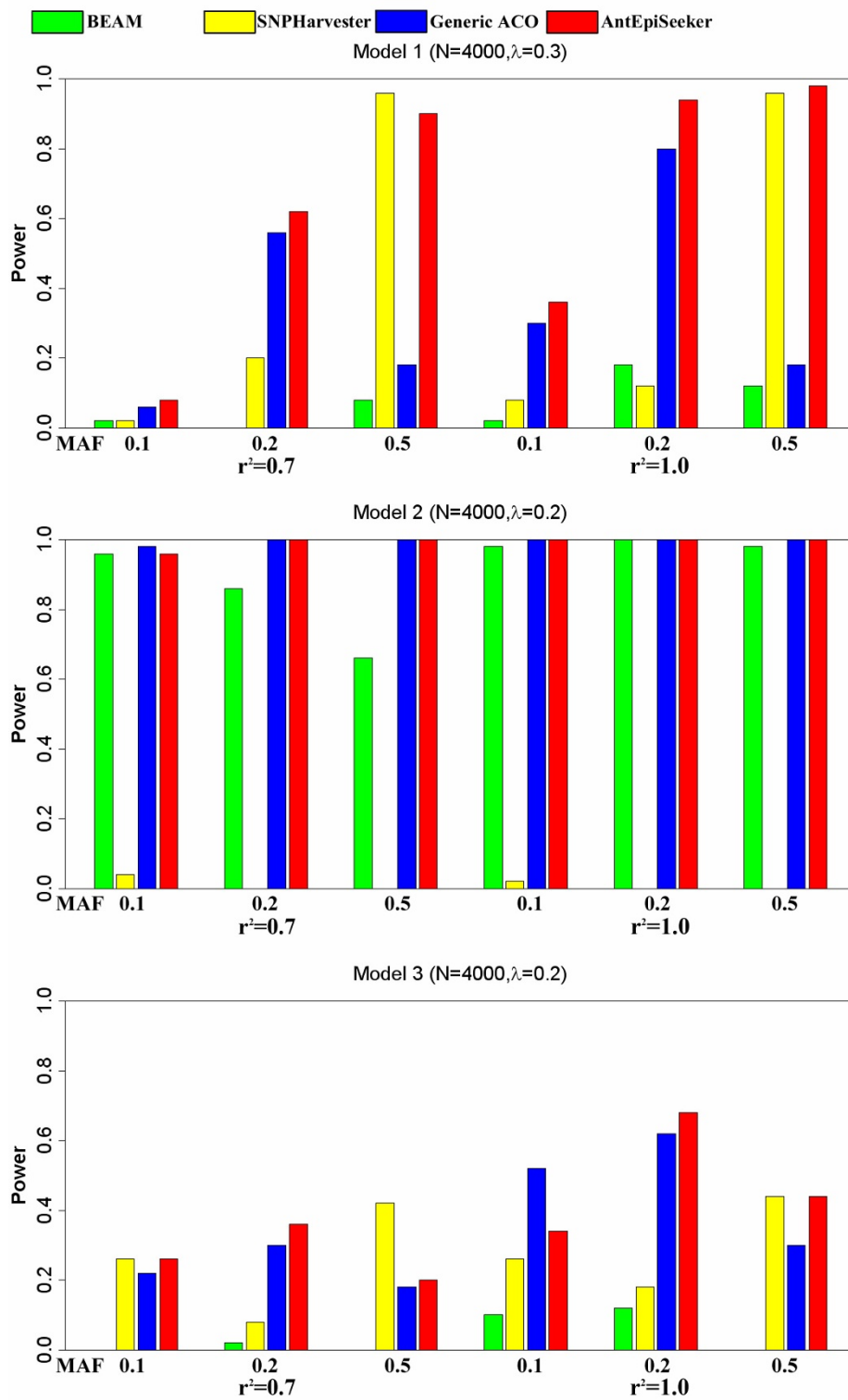


Figure 2 Power comparison between AntEpiSeeker, SNPHarvester, BEAM and Generic ACO. The absence of bars indicates zero power.

Table 1: False positive rate of different methods on a null simulation.

P value threshold	False positive rate				
	Exhaustive search	BEAM	SNPHarvester	AntEpiSeeker	
				Before minimizing false positives	After minimizing false positives
10^{-5}	5.5×10^{-6}	No positives reported	1.4×10^{-2}	3.5×10^{-6}	3.0×10^{-6}
10^{-4}	5.3×10^{-5}	No positives reported	1.6×10^{-2}	3.0×10^{-5}	1.1×10^{-5}
10^{-3}	6.9×10^{-4}	No positives reported	2.0×10^{-2}	2.9×10^{-4}	3.7×10^{-5}
10^{-2}	8.4×10^{-3}	No positives reported	2.4×10^{-2}	2.0×10^{-3}	6.6×10^{-5}

False positive rate evaluation on a null simulation

To approximate the false positive rates of AntEpiSeeker, a dataset without any genetic effects was simulated. The dataset contained 2,000 SNPs and 4,000 samples (2000 cases and 2000 controls), whose SNPs were generated independently with MAF uniformly distributed in [0.1, 0.5]. The parameters for running programs were the same as used in the first experiment. At different p-values, the false positive rates of the exhaustive search, BEAM, SNPHarvester and AntEpiSeeker are shown in Table 1. BEAM did not report any false positive, while the false positive rate of SNPHarvester is much higher than exhaustive search. AntEpiSeeker has a false positive rate that is comparable to but no larger than the desired significance level.

Evaluation of AntEpiSeeker on a simulated large scale dataset

To further test the performance of AntEpiSeeker, a real data based simulation study was carried out. The dataset consisted of the SNP genotypes on human chromosome 1 from 912 individuals (11 populations) of the International HapMap project (Phase 3) [25]. Loci with missing genotypes or $MAF < 0.1$ were removed, resulting in 73,355 SNP markers for analysis. Because it has no case/control status

attached, the data was randomly and equally divided between cases and controls (456 individuals in each group). Additionally, 132 epistatic interactions following the above mentioned 3 models were added to the data with randomly selected causative loci. The p-value threshold was set at 0.0001. The parameters for running programs were the same as used in the first experiment, for a fair comparison. The available software BEAM was not able to handle this dataset. The performances of SNPHarvester, generic ACO and AntEpiSeeker were compared in terms of true positive rates and false discovery rates, as summarized in Table 2. The results suggest that AntEpiSeeker significantly outperforms other methods on this large scale dataset.

Results on WTCCC RA data

AntEpiSeeker was used to perform a large-scale association study on the rheumatoid arthritis(RA) data from the Wellcome Trust Case Control Consortium (WTCCC) [26], which consisted of 332,831 SNP markers and 3,503 individuals (1,504 controls and 1,999 cases). Chromosomes were scanned separately first and then jointly. The parameters for AntEpiSeeker were adjusted according to the general rule presented in the section of parameter setting. Different methods were also compared based on this

Table 2: Performance comparison of different methods on a simulated large-scale dataset.

Methods	True positive rate	False discovery rate
SNPHarvester	26.5%	98.6%
Generic ACO	0	100%
AntEpiSeeker	66.7%	97.1%
AntEpiSeeker with minimized false positives	52.3%	18.8%

Table 3: Some epistatic interactions identified by AntEpiSeeker on WTCCC RA data.

Epistatic interactions	Location	Related genes	P value
(rs6457617, rs41443144)	6p21.32-6p25.1	MHC, RP3	<10 ⁻¹⁶
(rs743777, rs9627642)	22q12.3-22q13.31	NR, NR	<10 ⁻¹⁶
(rs2837960, rs41492246)	21q22.3-21q22.12	NR, NR	<10 ⁻¹⁶
(rs6920220, rs17165379)	6q23.3-6p25.3	TNFAIP3, NR	<10 ⁻¹⁶
(rs6457620, rs41454544)	6p21.32-6q21	HLA-DRB1, OSTM1	<10 ⁻¹⁶
(rs3890745, rs41348151)	1p36.32-1p21.3	MME11, NR	<10 ⁻¹⁶
(rs4810485, rs2748666)	20q13.12-20q11.23	CD40, NR	<10 ⁻¹⁶

real dataset. The available software BEAM was not able to handle such a big dataset. Compared with SNPHarvester, AntEpiSeeker identified more SNP markers, which were previously identified as having remarkable marginal effects [26-28], as being involved in epistatic interactions. We summarized these epistatic interactions in Table 3. Other epistatic interactions suggested by AntEpiSeeker were posted on our project web site at <http://nce.ads.uga.edu/~romdhane/AntEpiSeeker/index.html>. It took about 5 days for AntEpiSeeker to handle the WTCCC RA data, based on Dual Core AMD Opteron(tm) Processor 275, while SNPHarvester took about 2 weeks to handle it, a similar time to the one reported in [12].

Conclusion

In this paper, we proposed a novel tool (AntEpiSeeker) for the discovery of epistatic interactions in large scale case-control studies. AntEpiSeeker was assessed through comparison with two recent approaches on both simulated and real datasets. AntEpiSeeker, which adopts a two-stage optimization procedure, is a modified algorithm derived from the generic ACO. To demonstrate the advantages of the two-stage optimization, we also compared the performance of AntEpiSeeker with that of the generic ACO. AntEpiSeeker is a continuous research project and may be upgraded in the future.

Availability and requirements

Project name: AntEpiSeeker

Project home page: <http://nce.ads.uga.edu/~romdhane/AntEpiSeeker/index.html>

Operating system(s): Windows, Linux

Programming language: C++

Other requirements: GNU Scientific Library (GSL) is needed for recompile

License: None for usage

Any restrictions to use by non-academics: None

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YW and KR conceived the project. YW developed the software. YW and XL analyzed the data. YW and RR wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This study was supported in part by resources and technical expertise from the University of Georgia Research Computing Center, a partnership between the Office of the Vice President for Research and the Office of the Chief Information Officer.

Author Details

¹Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA, ²Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA and ³Department of Statistics, University of Georgia, Athens, GA 30602, USA

Received: 1 October 2009 Accepted: 28 April 2010

Published: 28 April 2010

References

1. Cordell HJ, Clayton DG: **Genetic association studies.** *Lancet* 2005, **366**:1121-1131.
2. Cordell HJ: **Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans.** *Hum Mol Genet* 2002, **11**:2463-2468.
3. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH: **Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer.** *Am J Hum Genet* 2001, **69**:138-147.
4. Cho YM, Ritchie MD, Moore JH, Park JY, Lee KU, Shin HD, Lee HK, Park KS: **Multifactor-dimensionality reduction shows a two-locus interaction associated with type 2 diabetes mellitus.** *Diabetologia* 2004, **47**:549-554.
5. Tsai CT, Lai LP, Lin JL, Chiang FT, Hwang JJ, Ritchie MD, Moore JH, Hsu KL, Tseng CD, Liau CS, Tseng YZ: **Renin-angiotensin system gene polymorphisms and atrial fibrillation.** *Circulation* 2004, **109**:1640-1646.
6. Nelson MR, Kardina SL, Ferrell RE, Sing CF: **A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation.** *Genome Res* 2001, **11**:458-470.
7. Culverhouse R, Klein T, Shannon W: **Detecting epistatic interactions contributing to quantitative traits.** *Genet Epidemiol* 2004, **27**:141-152.
8. Millstein J, Conti DV, Gilliland FD, Gauderman WJ: **A testing framework for identifying susceptibility genes in the presence of epistasis.** *Am J Hum Genet* 2006, **78**:15-27.
9. Zeng T, Wang H, Lo SH: **Backward genotype-trait association (BGTA) - based dissection of complex traits in case-control design.** *Hum Hered* 2006, **62**:196-212.
10. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, et al.: **Complement factor H polymorphism in age-related macular degeneration.** *Science* 2005, **308**:385-389.
11. Zhang Y, Liu JS: **Bayesian inference of epistatic interactions in case-control studies.** *Nat Genet* 2007, **39**:1167-1173.

12. Yang C, He Z, Wan X, Yang Q, Xue H, Yu W: **SNPHarvester: a filtering-based approach for detecting epistatic interactions in genome-wide association studies.** *Bioinformatics* 2009, **25**:504-511.
13. Dorigo M, Gambardella LM: **Ant colonies for the travelling salesman problem.** *Biosystems* 1997, **43**:73-81.
14. Greene CS, White BC, Moore JH: **Ant colony optimization for genome-wide genetic analysis.** *Lect Notes Comput Sci* 2008, **5217/2008**:37-47.
15. Greene CS, Gilmore JM, Kiralis J, Andrews PC, Moore JH: **Optimal use of expert knowledge in ant colony optimization for the analysis of epistasis in human disease.** *Lect Notes Comput Sci* 2009, **5483/2009**:92-103.
16. Marchini J, Donnelly P, Cardon LR: **Genome-wide strategies for detecting multiple loci that influence complex diseases.** *Nat Genet* 2005, **37**:413-417.
17. Talbi EG, Roux O, Fonlupt C, Robillard D: **Parallel Ant Colonies for the quadratic assignment problem.** *Future Generation Computer System* 2001, **17**:441-449.
18. Maniezzo V, Carbonaro A: **An ANTS heuristic for the frequency assignment problem.** *Future Generation Computer Systems* 2000, **16**:927-935.
19. Dorigo M, Di Caro G: **New Ideas in Optimization.** In *New Ideas in Optimization* Edited by: Corne D, Dorigo M, Glover F. McGraw-Hill; 1999.
20. Dorigo M, Di Caro G, Gambardella LM: **Ant Algorithms for Discrete Optimization.** *Artificial Life* 1999, **5**:137-172.
21. Robbins KR, Zhang W, Bertrand JK, Rekaya R: **The ant colony algorithm for feature selection in high-dimension gene expression data for disease classification.** *Math Med Bio* 2007, **24**:413-26.
22. Dorigo M, Stützle T: *Ant Colony Optimization* The MIT Press; 2004.
23. Jiang R, Tang W, Wu X, Fu W: **A random forest approach to the detection of epistatic interactions in case-control studies.** *BMC Bioinformatics* 2009, **10**(Suppl 1):S65.
24. Wan X, Yang C, Yang Q, Xue H, Tang NL, Yu W: **MegaSNPHunter: a learning approach to detect disease predisposition SNPs and high level interactions in genome wide association study.** *BMC Bioinformatics* 2009, **10**:13.
25. The International HapMap Consortium: **The International HapMap Project.** *Nature* 2003, **426**:789-796.
26. Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447**:661-678.
27. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Cobyln J, Weinblatt ME, Gabriel SB, Purcell S, Klareskog L, Gregersen PK, Shadick NA, Daly MJ, Altshuler D: **Two independent alleles at 6q23 associated with risk of rheumatoid arthritis.** *Nat Genet* 2007, **39**:1477-1482.
28. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt NP, Gianniny L, Korman BD, Padyukov L, Kurreeman FA, Chang M, Catanese JJ, Ding B, Wong S, Helm-van Mil AH van der, Neale BM, Cobyln J, Cui J, Tak PP, Wolbink GJ, Crusius JB, Horst-Bruinsma IE van der, Criswell LA, Amos CI, Seldin MF, Kastner DL, Ardlie KG, Alfredsson L, Costenbader KH, Altshuler D, Huizinga TW, Shadick NA, Weinblatt ME, de Vries N, Worthington J, Seielstad M, Toes RE, Karlson EW, Begovich AB, Klareskog L, Gregersen PK, Daly MJ, Plenge RM: **Common variants at CD40 and other loci confer risk of rheumatoid arthritis.** *Nat Genet* 2008, **40**:1216-1223.

doi: 10.1186/1756-0500-3-117

Cite this article as: Wang et al., AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm *BMC Research Notes* 2010, **3**:117

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

