DE GRUYTER OPEN

JAISCR, 2016, Vol. 6, No. 3, pp.203 – 212
10.1515/jaiscr-2016-0015

GENETIC ALGORITHM COMBINED WITH A LOCAL SEARCH METHOD FOR IDENTIFYING SUSCEPTIBILITY GENES

Cheng-Hong Yang[1], Sin-Hua Moi[1], Yu-Da Lin[1], and Li-Yeh Chuang[2]

[1]*Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan*
e-mail: chyang@cc.kuas.edu.tw, moi9009@gmail.com, e0955767257@yahoo.com.tw

[2]*Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan*
e-mail: chuang@isu.edu.tw

Abstract

Detecting genetic association models between single nucleotide polymorphisms (SNPs) in various disease-related genes can help to understand susceptibility to disease. Statistical tools have been widely used to detect significant genetic association models, according to their related statistical values, including odds ratio (*OR*), chi-square test (χ^2), *p*-value, *etc*. However, the high number of computations entailed in such operations may limit the capacity of such statistical tools to detect high-order genetic associations. In this study, we propose lsGA algorithm, a genetic algorithm based on local search method, to detect significant genetic association models amongst large numbers of SNP combinations. We used two disease models to simulate the large data sets considering the minor allele frequency (MAF), number of SNPs, and number of samples. The three-order epistasis models were evaluated by chi-square test (χ^2) to evaluate the significance (*P*-value $<$ 0.05). Analysis results showed that lsGA provided higher chi-square test values than that of GA. Simple linear regression indicated that lsGA provides a significant advantage over GA, providing the highest β values and significant *p*-value.

Keywords: Genetic algorithms, identifying susceptibility genes, local search algorithm

1 Introduction

Single nucleotide polymorphisms (SNPs) are important biomarkers in genomes [1], and gene expression may be influenced by the SNP alone or by interaction between SNPs [2]. Thus, improved understanding of associations between SNPs contributes to the analysis of diseases and cancers [3-5]. Genetic associations indicate that the effect of any single genetic variation (e.g., SNPs) will likely be dependent on other genetic variations (interaction between SNPs) [6]. Genetic association studies focus on which SNP combinations may be associated with high risk in genes related to diseases and cancers. Thus, epistasis identification can be regarded as a feature selection problem, and genetic association detection remains a challenge in bioinformatics [7].

Genetic associations can be identified by identifying significant differences between pathological (case) and normal (control) state. Many statistical methods have been proposed to identify significant genetic associations, such as PLINK [8] and BOOST [9]. However, these methods only identify two-order genetic associations. Identification of high order genetic associations is a NP-hard prob-

lem, especially for high-dimensional SNP combinations and large SNPs [10]. Traditional statistical methods, e.g., chi-square test (χ^2), are suitable for computationally intensive operations. Thus, evolutionary computations have been applied to improve statistical methods for identifying significant genetic associations. Particle swarm optimization (PSO) has been applied to identify significant genetic associations for facial emotion perception [11] and hypertension [12]. Genetic algorithm (GA) has been successfully used to identify significant genetic associations for chronic dialysis [13] and breast cancer [14]. These previous studies showed that the limitations imposed by large statistical evaluations can be overcome by evolutionary computation. Moreover, significant genetic associations with SNP combinations show that a marginal SNP may be excluded due to P-value > 0.05, but this SNP may be associated with disease when combined with other SNPs [13]. However, the search abilities of these methods are insufficiently robust for large numbers of SNPs.

In this study, we used local search to improve on GA to enhance population diversity. Local search can reduce the probability of the same vector being identified between two selected chromosomes to create a crossover operation. A high-dimensional data set was simulated using the biological parameters of SNPs. The results of the improved GA outperform those of the traditional GA.

2 Method

2.1 Problem definition

An SNP represents three types of genotypes, including 'AA' (homozygous reference genotype), 'Aa' (heterozygous genotype), and 'aa' (homozygous variant genotype). In this study, the genotype at SNP_i is defined as a set $G_i = \{1, 2, 3 | 1='AA', 2='Aa', 3='aa'\}$, where i is the i^{th} SNP in n SNPs (n=total number of SNPs) which is related to disease. Genetic association identification aims to select the m SNP ($m \geq 2$) by determining whether their combination has significant associations with disease or not. A genetic association can be regarded as a set $E = \{s_1, s_2, s_3, ..., s_m\}$, where $s_i = \{SNP_i, G_i\}$ and the problem space consists of the m-dimensional SNP selection. The objective function $f(E)(f : \delta \subseteq R^m R)$ is defined by chi-square test (χ^2) and the objective E^* is the set E with highest χ^2 value, i.e., $f(E^*) > f(E)$ for all $E \in \delta$, where δ is a non-empty large finite set serving as the problem space.

2.2 Genetic algorithm

Genetic algorithm (GA) was proposed by Holland [15] and has been applied to research in artificial intelligence, such as gene expression in biology problems. Thus, GA has been applied to the problems of classification [16] and primer design [17]. In GA, a chromosome is represented as an available solution in the search space, i.e., a genetic association set E. Each chromosome is evaluated by the objective function, and the good chromosomes have a higher probability to precede the evolutionary operation. Furthermore, bad chromosomes will be eliminated from the population, leaving the promising elements in the good chromosomes for the next generation. The evolutionary strategy in GA includes six operations: (1) chromosome initialization, (2) population estimation, (3) selection operation, (4) crossover operation, (5) mutation operation, and (6) replacement operation. Algorithm 1 shows the GA process.

2.3 Genetic algorithm based on the local search algorithm

The local search algorithm searches the k-exchange neighborhood to improve the chromosome from the current solution by exchanging at most k elements [18]. Various studies have successfully applied the local search algorithm to improve evolutionary algorithms, such as multi-objective flexible job-shop scheduling problem [19], multi-modal optimization [20] and best-offspring hybrid genetic algorithm [21]. This study use the local search algorithm to enhance the population diversity after the mutation operation in GA process (lsGA). Algorithm 2 shows the lsGA pseudo-code. The detailed operations are explained in the following sections.

Chromosomal representation

The chromosomes are defined by the definition of genetic association and are shown below:

$$C_l = \{SNP_l, G_l\},$$

where SNP_l is a set included the selected m SNPs, where l is the l^{th} chromosome in the population, in

which each SNP cannot be selected repeatedly. $G_l$is a set including the genotypes which correspond to SNP_l. Let $C_l = \{10, 17, 1, 2\}$, which indicates that the i^{th} chromosome consists of the 'AA' genotype of SNP_{10} and 'Aa' genotype of SNP_{17}, in which the number of the SNP is its order in the dataset.

Objective function

In the GA process, the objective function is used to estimate the values of the chromosomes, referred to as fitness values. The chi-square test (χ^2) aims to identify the significant epistasis. The objective function can be written as:

$$F(C_l) = \frac{(a+b+c+d)(a\times d - b\times c)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (1)$$

where a, b, c, and d are respectively the four cells in the contingency table (see Table 1) The a is the total number of matched C_l in the cases, b is the total number of matched C_l in the controls, c is the total number of unmatched C_l in the cases, and d is the total number of unmatched C_l in the control. In this study, a high objective function indicates a better chromosome.

Selection operation

In GA, genetic operations requires two parents (P_1 and P_2) to produce two children (P'_1 and P'_2), and the parents are selected by the selection operation. We used rank-based tournament selection which ranks the chromosomes according to their fitness values and selects the two top chromosomes as the parents.

Crossover operation

The crossover operation performed a one-point crossover that randomly generated the D binary strings (D is the dimension of the parent). The first string indicates that the elements of two parents P_1 and P_2 need to be exchanged, while remaining strings are unchanged. Let binary strings = $\{1, 0, 0, 1\}$, $P_1 = \{1, 4, 2, 1\}$, and $P_2 = \{2, 4, 1, 3\}$, the two offsprings P'_1 and P'_2 are $\{2, 4, 2, 3\}$ and $\{1, 4, 1, 1\}$, respectively.

Mutation operation

The mutation operation performed the binary string mutation in which each bit in the binary string randomly generated a probability. If the probability is smaller than mutation threshold, this point in the offspring randomly generates a possible element. If the binary string remains unchanged after the mutation operation, this operation is repeatedly performed until a single bit is mutated.

Local search algorithm

The local search algorithm was used to find the better solution in the offspring neighborhood, and it could enhance the population diversity, especially when the production of offsprings is similar in the population. Algorithm 3 shows the pseudo-code of the local search algorithm. P' indicates the offsprings and C' is the neighboring offspring. d is the increased distance value between P' and C'. If the fitness value of C' is better than the fitness value of P', then C' replaces P'.

Replacement operation

The replacement operation aims to keep the good chromosomes for genetic operations in the following generation. The two producing offspring are added into the population and the least two chromosomes with low fitness values are deleted from the population.

Algorithm 1 – GA pseudo-code

```
01: begin
02: Initial population
03: while (generation ≠ termination)
04: Evaluate population
05: Selection
06: Crossover
07: Mutation
08: Replacement
09: Output best chromosome
10: end
```

Algorithm 2 – lsGA pseudo-code

```
01: begin
02: Initial population
03: while (generation ≠ termination)
04: Evaluate population
05: Selection
06: Crossover
07: Mutation
08: Local search
09: Replacement
10: Output best chromosome
11: end
```

Algorithm 3 – Local search algorithm pseudo-code

```
01: begin
02: for (i=1; i < total number of offsprings; i++)
03: Copy i^th offspring P'_i into C'_i;
04: for (j=1; j < the dimension of C'_i; j++)
05: d_j=l_j × Rand [0:1]; l_j={ l_j ∈N: l_j <max(C_j)}
06: C'_ij = C'_ij + d_j
07: If fitness (P'_i) > fitness (C'_ij)
08: Replace P'_i by C'_ij
09: end
```

3 Result and Discussion

3.1 Data set

In the performance comparison, two epistasis models, ZZ model [22, 23] and XOR model [24], were selected to test all methods. The XOR model is the nonlinear epistasis, and high risk of disease is dependent on inheriting a heterozygous genotype from one locus or a heterozygous genotype from another locus, but not all loci. In the ZZ model, high risk of disease is dependent upon inheriting exactly two high-risk alleles from two loci. MAFs of disease-associated SNPs were set at 0.1 and 0.2, and MAFs of unassociated SNPs were set from [0.05, 0.5]. Total numbers of SNPs were 50 and 100, and total numbers of samples were 400 (cases = 200 and controls = 200) and 1000 (cases = 500 and controls = 500). GAMETES was used to generate the SNP dataset using the above parameters [25]. Each parameter combination generated 100 data sets in each disease model. The objective is to identify the significant genetic association models.

3.2 Parameter settings

In this study, all methods used the same parameters and the same initial population to test statistical ability to identify genetic associations. The exchange probability for the one-point selection operation is 1.0 and the exchange probability for the one-point mutation operation is 0.1. The population size is 50 and the total number of generations is 1000.

3.3 Evaluation of identified genetic association models in 12 XOR models and 12 ZZ models

In this study, each initial population between GA and lsGA is the same and the random seed in the program is also the same. Figures 1 and 2 show the results of three-order genetic association models in GA and lsGA. The symbols, upper side and lower side, in each point indicate the mean ± standard deviation (SD), and each point saves all fitness values of the population for every 50 generations over 100 data sets.

In Figure 1, the results for all generations showed that the mean best fitness values from lsGA outperform those of the traditional GA in 12 XOR models. The difference of chi-square test (χ^2) values (fitness values) between GA and lsGA is very large, indicating that lsGA outperforms GA in identifying the most significant genetic association model, and the increased χ^2 values indicate that the p-value is decreased, i.e., p-value $<<$ 0.05. Both total number of SNPs and samples can influence the χ^2 values, in which the χ^2 values of large samples are higher than small samples because the a and d in Table 1 increases when the total number of samples are increased. The χ^2 values of SNP = 500 are lower than other XOR models with SNP = 50 and 100. This clearly shows that a high total number of SNPs can increase the degree of difficulty in processing the evolutionary algorithm. However, lsGA can enhance the χ^2 values, especially in XOR model with SNP = 500 and sample = 1000. This shows that the local search algorithm facilitates the finding of better solutions.

Table 1. Contingency table of a chromosome

	Case	Control	Total
C_l	a	b	a+b
C_l	c	d	c+d
Total	a+c	b+d	a+b+c+d

C_l indicates unmatching C_l

In Figure 2, results for all generations show that the mean of best fitness values from lsGA outperform those of GA in 12 ZZ models, and the high χ^2 values indicate that significant genetic association models are identified by GA and lsGA. However, lsGA identified more significant genetic asso-

ciation models than GA did. In the ZZ model, the χ^2 values slowly increase over generations due to the high risk of disease in ZZ model being dependent upon inheriting the two high risk alleles from two loci, resulting in only three possible genotype combinations, in which the combination with only one homozygous reference genotype may have high risk in the three-order genetic association models. Therefore, the population may easily be trapped in a local optima. However, lsGA can continually enhance the χ^2 values, especially in ZZ models with SNP = 500. This shows that the local search algorithm can avoid the population being trapped in a local optima.

Table 2 shows the average results of a simple linear regression for best fitness values and the average amount of fitness values for the population using GA and lsGA methods in the XOR and ZZ models. The positive β values indicate that the lsGA is superior to GA, and the high value indicates the greater improvement. The p-value ($P>$t) is used to determine whether lsGA significantly improves on GA in XOR and ZZ models. The lsGA shows a significant advantage as compared to the GA, providing the highest β values and significant p-value.

3.4 Comparison of GA and lsGA for population

Figures 3 and 4 show the mean sum of fitness values of the population in the form of a $\log_{10}$ value over the number of generation in GA and lsGA. The symbols, upper side and lower side, are the mean $\pm$ standard deviation (SD). Each point is the mean sum of the fitness values in the population over 100 data sets.

The distribution curves in Figures 3 and 4 are respectively similar with those in Figure 1 and 2, indicating the values of chromosomes are improved by the genetic operations of GA. In addition, the improvement trend in GA is relatively slow, while that in lsGA is more obvious. This indicates that the local search algorithm can provide better offspring to advance the population for finding better epistasis models in the XOR and ZZ models.

3.5 Effectiveness comparison of GA and lsGA

The effectiveness of the proposed lsGA is shown by computer simulations on genetic association models consisting of 12 XOR models and 12 ZZ models. The results clearly showed that lsGA can effectively escape from the local optima. Thus, the more significant genetic association models could be identified by lsGA, and these genetic association models with high risk included several SNPs which can help improve understanding of the associations between genes and disease. Several local search algorithms have been proposed to improve evolutionary algorithms in various problems, including multi-objective optimization [26], location-routing problem [27], and so on. Therefore, our proposed lsGA may be able to solve other problems. Furthermore, these local search algorithms may be more effectively in improving the search ability of GA for identifying better genetic association models.

3.6 Runtime comparison of GA and lsGA

The computational running time of lsGA was similar to that of GA. The local search algorithm evaluates the D-dimensional vectors in the two offsprings after the mutation operation. The computational complexity of GA can be represented as big-O(NM), where N is the total number of generations and M is the total population size. lsGA is big-O($N(M+2D)$), where D is the chromosome dimension. The D is very small in the problem of identifying genetic associations, e.g., a 3-order genetic association only uses a 6-dimensional vector. However, lsGA is superior to GA in terms of finding better genetic association models with higher χ^2 values.

4 Conclusions

The local search algorithm is used to improve the GA (named as lsGA) to detect genetic associations amongst disease-related genes. Two disease models are used to evaluate the ability of lsGA to detect significant genetic association models regarding the marks of SNPs located in susceptibility genes. Our results show that lsGA can detect more significant models than GA, and continued to effectively enhance the χ^2 values for finding better mod-

Table 2. Comparison of GA and lsGA in the mean of best fitness values and in mean sum of fitness values of population in XOR and ZZ model by simple linear regression

	MAF = 0.1, Sample size = 400		MAF = 0.1, Sample size = 1000		MAF = 0.2, Sample size = 400		MAF = 0.2, Sample size = 1000	
	β	P>t	β	P>t	β	P>t	β	P>t
Mean of best fitness values								
XOR model								
SNPs = 50	0.23	1.10E-98	0.21	4.58E-81	0.39	4.12E-154	0.16	4.94E-24
SNPs = 100	0.25	5.49E-122	0.21	6.04E-82	0.41	7.81E-171	0.33	7.84E-106
SNPs = 500	0.33	2.83E-219	0.30	8.80E-170	0.37	3.41E-137	0.34	6.28E-115
ZZ model								
SNPs = 50	0.29	1.79E-158	0.34	2.04E-223	0.35	5.57E-124	0.30	1.40E-123
SNPs = 100	0.26	6.21E-126	0.35	3.00E-245	0.27	1.51E-69	0.34	1.22E-84
SNPs = 500	0.13	4.06E-31	0.11	9.65E-24	0.19	1.69E-34	0.06	5.18E-61
Mean sum of fitness values of population								
XOR model								
SNPs = 50	0.26	9.20E-134	0.26	2.32E-125	0.37	1.62E-138	0.24	9.64E-58
SNPs = 100	0.28	8.96E-153	0.22	3.48E-95	0.39	2.15E-149	0.35	5.20E-120
SNPs = 500	0.34	1.68E-232	0.32	1.08E-199	0.37	1.14E-134	0.34	8.40E-116
ZZ model								
SNPs = 50	0.32	2.69E-197	0.36	8.52E-261	0.35	8.05E-91	0.36	4.73E-131
SNPs = 100	0.27	1.40E-138	0.36	5.06E-260	0.29	9.41E-117	0.36	2.83E-129
SNPs = 500	0.18	4.93E-59	0.10	2.47E-18	0.25	2.76E-05	0.08	5.02E-08

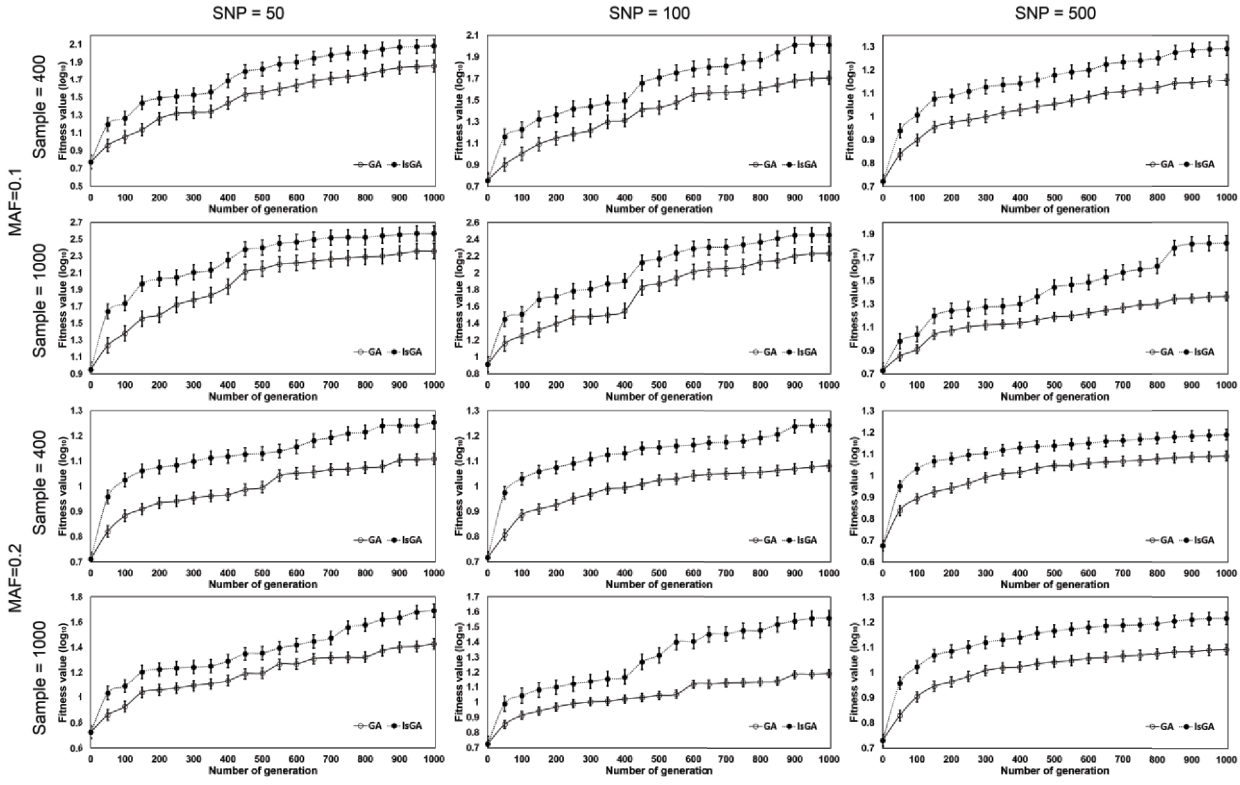

Figure 1. Mean best fitness values in the form of a log10 value over the number of generations for GA and lsGA in the 12 XOR models. The error bar is evaluated by the standard deviation in each point.

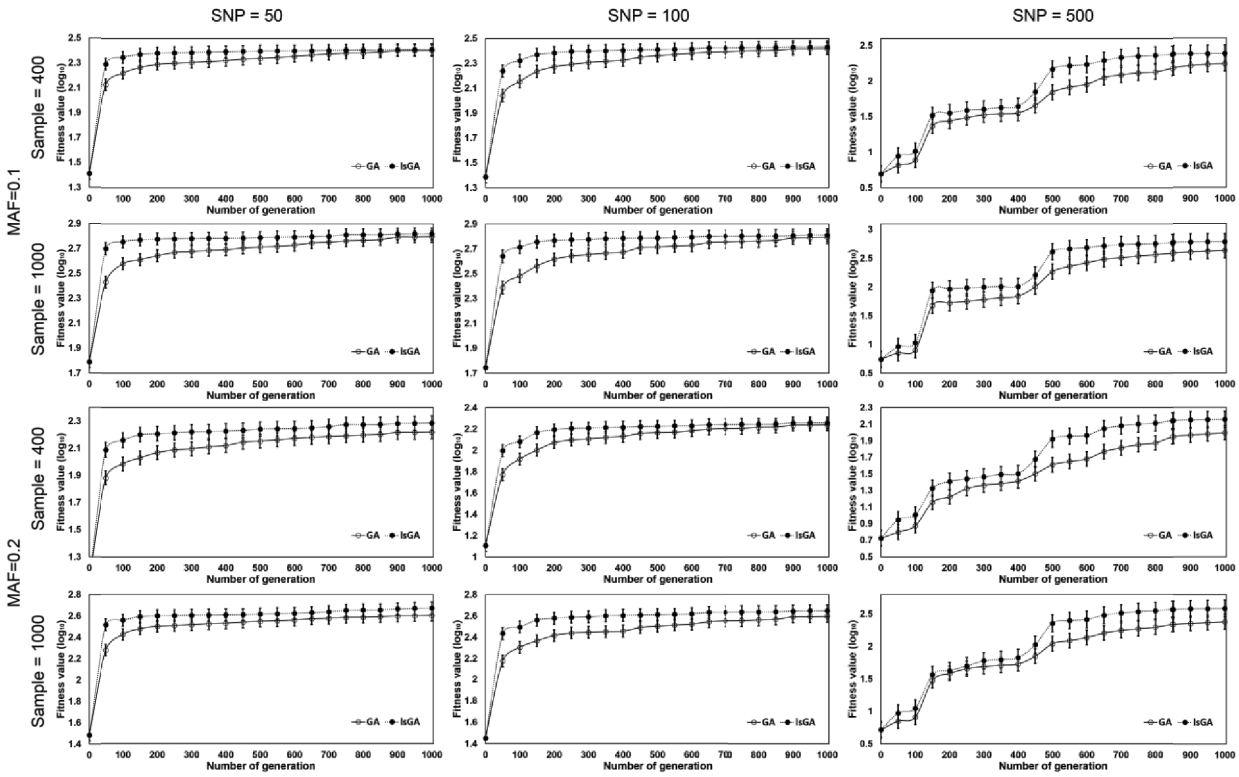

Figure 2. Mean best fitness values in the form of a log10 value over the number of generations for GA and lsGA in the 12 ZZ models. The error bar is evaluated by the standard deviation in each point.

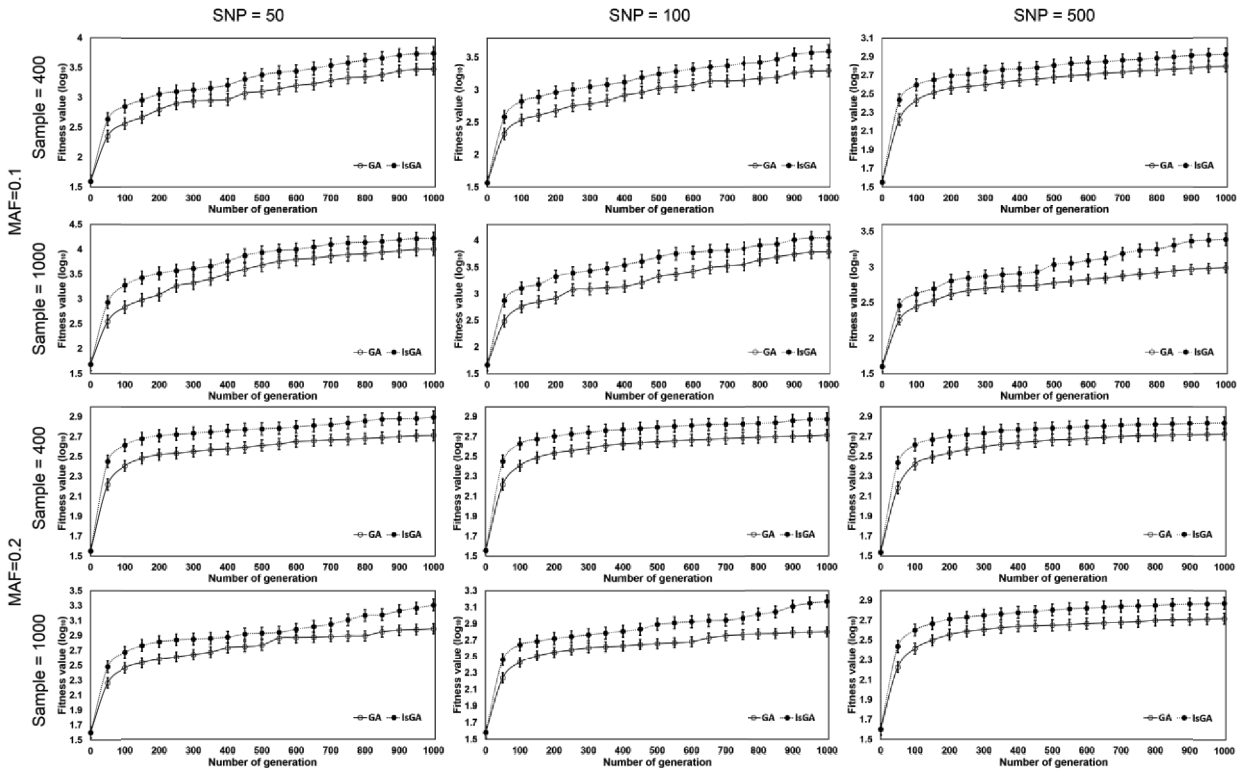

Figure 3. Mean sum fitness values of population in the form of a $\log_{10}$ value over the number of generations for GA and lsGA in the 12 XOR models. The error bar is evaluated by the standard deviation in each point.

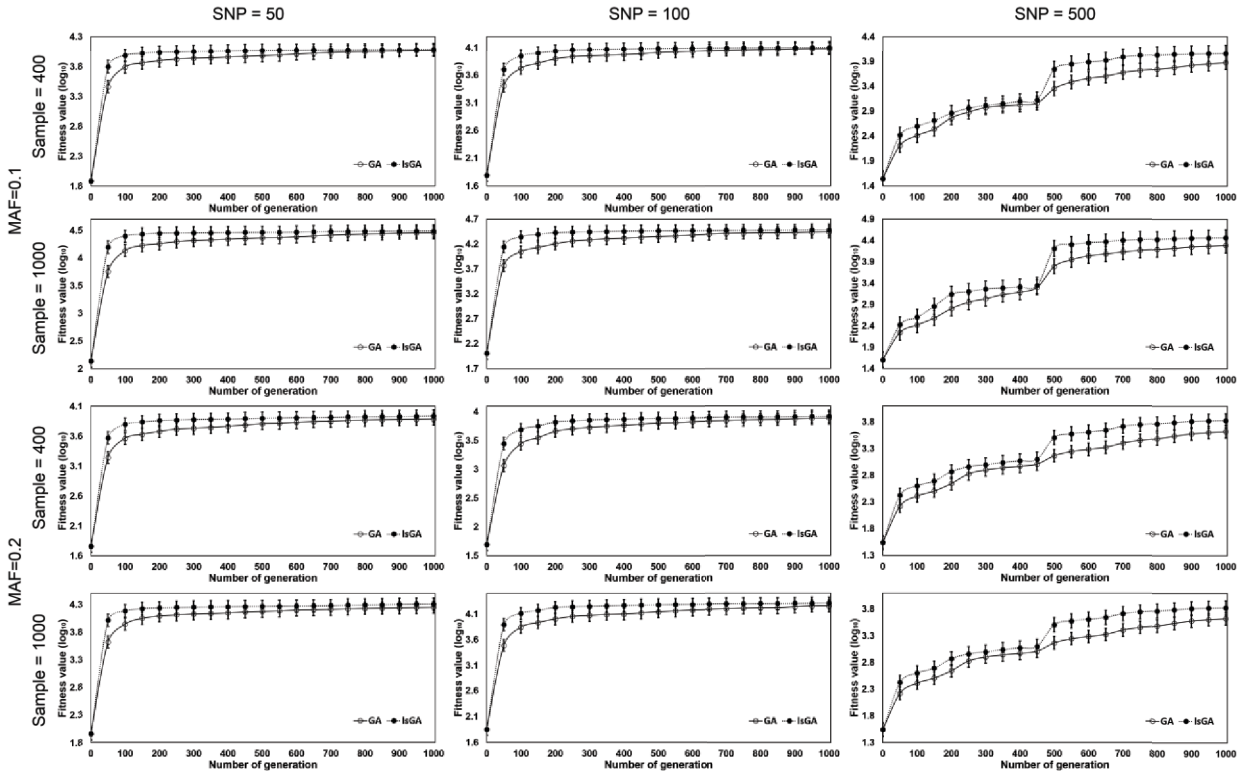

Figure 4. Mean sum fitness values of population in the form of a $\log_{10}$ value over the number of generations for GA and lsGA in the 12 ZZ models. The error bar is evaluated by the standard deviation in each point.

els, indicating that lsGA can be applied to identify complex genetic association models in large data sets.

Acknowledgment

This study was partly supported by the National Science Council of Taiwan for Grant NSC 102-2221-E-151-024 -MY3, 102-2622-E-151-003-CC3, and 102-2221-E-214 -039.

References

[1] A. D Roses, A. M. Saunders, Y. Huang, J. Strum, K. H. Weisgraber, and R. W. Mahley, Complex disease-associated pharmacogenetics: drug efficacy, drug safety, and confirmation of a pathogenetic hypothesis (Alzheimer's disease), Pharmacogenomics Journal, vol. 7, pp. 10-28, Feb 2007.

[2] R. Sanjuan and M. R. Nebot, A Network Model for the Correlation between Epistasis and Genomic Complexity, PLos One, vol. 3, pp. e2663, Jul 2008.

[3] C. H. Yang, L. Y. Chuang, Y. H. Cheng, Y. D. Lin, C. L. Wang, C. H. Wen, et al., Single nucleotide polymorphism barcoding to evaluate oral cancer risk using odds ratio-based genetic algorithms, Kaohsiung Journal of Medical Sciences, vol. 28, pp. 362-368, Jul 2012.

[4] J. B. Chen, Y. H. Yang, W. C. Lee, C. W. Liou, T. K. Lin, Y. H. Chung, et al., Sequence-based polymorphisms in the mitochondrial D-Loop and potential SNP predictors for chronic dialysis, PLoS One, vol. 7, pp. e41125, Jul 2012.

[5] C. Y. Yen, S. Y. Liu, C. H. Chen, H. F. Tseng, L. Y. Chuang, C. H. Yang, et al., Combinational polymorphisms of four DNA repair genes XRCC1, XRCC2, XRCC3, and XRCC4 and their association with oral cancer in Taiwan, Journal of Oral Pathology & Medicine, vol. 37, pp. 271-277, May 2008.

[6] J. H. Moore, A global view of epistasis, Nature Genetics, vol. 37, pp. 13-14, Jan 2005.

[7] J. H. Moore, F. W. Asselbergs, and S. M. Williams, Bioinformatics challenges for genome-wide association studies, Bioinformatics, vol. 26, pp. 445-455, Feb 15 2010.

[8] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, et al., PLINK: A tool set for whole-genome association and population-based linkage analyses, American Journal of Human Genetics, vol. 81, pp. 559-575, Sep 2007.

[9] X. A. Wan, C. Yang, Q. A. Yang, H. Xue, X. D. Fan, N. L. S. Tang, et al., BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies, American Journal of Human Genetics, vol. 87, pp. 325-340, Sep 2010.

[10] C. S. Greene, B. C. White, and J. H. Moore, Ant colony optimization for genome-wide genetic analysis, in Ant Colony Optimization and Swarm Intelligence, ed: Springer, pp. 37-47, 2008.

[11] L. Y. Chuang, H. Y. Lane, Y. D. Lin, M. T. Lin, C. H. Yang, and H. W. Chang, Identification of SNP barcode biomarkers for genes associated with facial emotion perception using particle swarm optimization algorithm, Annals of General Psychiatry, vol. 13, pp. 15, May 2014.

[12] S. J. Wu, L. Y. Chuang, Y. D. Lin, W. H. Ho, F. T. Chiang, C. H. Yang, et al., Particle swarm optimization algorithm for analyzing SNP-SNP interaction of renin-angiotensin system genes against hypertension, Molecular Biology Reports, vol. 40, pp. 4227-4233, Jul 2013.

[13] J. B. Chen, L. Y. Chuang, Y. D. Lin, C. W. Liou, T. K. Lin, W. C. Lee, et al., Genetic algorithm-generated SNP barcodes of the mitochondrial D-loop for chronic dialysis susceptibility, Mitochondrial DNA, vol. 25, pp. 231-237, Jun 2014.

[14] W. C. Chang, Y. Y. Fang, H. W. Chang, L. Y. Chuang, Y. D. Lin, M. F. Hou, et al., "Identifying association model for single-nucleotide polymorphisms of ORAI1 gene for breast cancer," Cancer Cell International, vol. 14, pp. 29, Mar 2014.

[15] J. H. Holland, Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence: U Michigan Press, 1975.

[16] L. P. Li, C. R. Weinberg, T. A. Darden, and L. G. Pedersen, Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method, Bioinformatics, vol. 17, pp. 1131-1142, Dec 2001.

[17] C. H. Yang, Y. H. Cheng, L. Y. Chuang, and H. W. Chang, Confronting two-pair primer design for enzyme-free SNP genotyping based on a genetic algorithm, BMC Bioinformatics, vol. 11, pp. 509, Oct 2010.

[18] E. H. Aarts and J. K. Lenstra, Local search in combinatorial optimization: Princeton University Press, 2003.

[19] G. Moslehi and M. Mahnam, A Pareto approach to multi-objective flexible job-shop scheduling problem using particle swarm optimization and local

search, International Journal of Production Economics, vol. 129, pp. 14-22, Jan 2011.

[20] B. Y. Qu, J. J. Liang, and P. N. Suganthan, Niching particle swarm optimization with local search for multi-modal optimization, Information Sciences, vol. 197, pp. 131-143, Aug 2012.

[21] Wan W, Birch JB: An Improved Hybrid Genetic Algorithm with a New Local Search Procedure. Journal of Applied Mathematics 2013, vol. 2013, Article ID 103591, Aug 2013.

[22] J. H. Moore, L. W. Hahn, M. D. Ritchie, T. A. Thornton, and B. C. White, Application of genetic algorithms to the discovery of complex models for simulation studies in human genetics, in Proceedings of the Genetic and Evolutionary Computation Conference/GECCO. Genetic and Evolutionary Computation Conference, 2002, pp. 1150.

[23] W. N. Frankel and N. J. Schork, Who's afraid of epistasis?, Nature genetics, vol. 14, pp. 371-373, 1996.

[24] J. H. Moore, L. W. Hahn, M. D. Ritchie, T. A. Thornton, and B. C. White, Routine discovery of complex genetic models using genetic algorithms, Applied Soft Computing, vol. 4, pp. 79-86, Feb 2004.

[25] R. J. Urbanowicz, J. Kiralis, N. A. Sinnott-Armstrong, T. Heberling, J. M. Fisher, and J. H. Moore, GAMETES: a fast, direct algorithm for generating pure, strict, epistatic models with random architectures, Biodata Mining, vol. 5, pp. 16, Oct 2012.

[26] A. Mousa, M. El-Shorbagy, and W. Abd-El-Wahed, Local search based hybrid particle swarm optimization algorithm for multiobjective optimization, Swarm and Evolutionary Computation, vol. 3, pp. 1-14, Apr 2012.

[27] H. Derbel, B. Jarboui, S. Hanafi, and H. Chabchoub, Genetic algorithm with iterated local search for solving a location-routing problem, Expert Systems with Applications, vol. 39, pp. 2865-2871, Feb 2012.

Cheng-Hong Yang is a professor of the Department of Electronic Engineering at National Kaohsiung University of Applied Sciences, Taiwan. He received his M.S. and Ph.D. degrees in computer engineering from North Dakota State University in 1988 and 1992, respectively. His main areas of research are evolutionary computation, bioinformatics, and assistive tool implementation.

Sin-Hua Moi received her M.S. degree from the Department of Oral Hygiene, Kaohsiung Medical University, Taiwan, in 2014. She is currently working towards the Ph.D. degree in the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan. She received diploma with specialization in biomedicine statistical analysis, database management and analysis. Her main areas of research are bioinformatics and biostatistics.

Yu-Da Lin received his M.S. and Ph.D. degree in the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan, in 2009 and 2015, respectively. He has rich experiences in computer programming, database design and management, and systems programming and design. His main areas of research are bioinformatics and computational biology.

Li-Yeh Chuang is a professor and director of the Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering at I-Shou University, Kaohsiung, Taiwan. She received her M.S. degree from the Department of Chemistry at the University of North Carolina in 1989 and her Ph.D. degree from the Department of Biochemistry at North Dakota State University in 1994. Her main areas of research are bioinformatics, biochemistry and genetic engineering.