IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 24, 2018, accepted November 26, 2018, date of publication November 30, 2018,
date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2884189

Evolutionary Computation-Based Memetic Algorithm Against Genetic Algorithm to Improve PCR-RFLP Assay Primers of SNP Genotyping

YU-HUEI CHENG[1], (Senior Member, IEEE), CHING-MING LAI[2], (Senior Member, IEEE), JIASHEN TEH[3], (Member, IEEE), AND CHE-NAN KUO[4]

[1]Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung 41349, Taiwan
[2]Department of Vehicle Engineering, National Taipei University of Technology, Taipei 10608, Taiwan
[3]School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Nibong Tebal 14300, Malaysia
[4]Department of Business Administration, CTBC Financial Management College, Tainan 709, Taiwan

Corresponding author: Che-Nan Kuo (fkikimo@hotmail.com)

This work was supported in part by the Ministry of Science and Technology (MOST) in Taiwan under Grant MOST107-2622-E-324-002-CC3, Grant MOST107-2221-E-324-020, and Grant MOST107-2821-C-324-001-ES, in part by the Chaoyang University of Technology in Taiwan under Grant CYUT107F0021213, in part by Chaoyang University of Technology (CYUT) and the Ministry of Education in Taiwan under Project: "The R&D and the cultivation of talent for Health-promotion products," in part by USM Research Universiti (RU) under Grant 304/PELECT/8014099, and in part by USM-UPE under External Grant 304/PELECT/6050385.

ABSTRACT A genetic algorithm (GA) combines the restriction enzyme mining core of single nucleotide polymorphism (SNP) restriction fragment length polymorphism (RFLP) to design polymerase chain reaction (PCR)-RFLP primer pairs for SNP-based genotyping with feasible estimated GA parameters. However, this GA method is easily trapped into local optima. An improved design of PCR-RFLP assay primers for SNP genotyping is needed. A memetic algorithm (MA) was used to design more robust primers for the PCR-RFLP assay to enable SNP genotyping. The novel restriction enzymes hunting (REHUNT) package was embedded into the MA method to provide available restriction enzymes. A formula to calculate more accurate thermodynamic primer melting temperatures was also introduced. Using the criteria of the GA method, *in silico* simulations for the MA method under different parameter settings were performed with the SNPs of SLC6A4, and results were compared. Appropriate MA parameter settings were superior in providing robust PCR-RFLP primers to achieve SNP genotyping compared with the GA method. Improvements included an accurate thermodynamic SantaLucia's formula for the calculation of melting temperature, use of the novel REHUNT for restriction enzymes mining, and selection of primers that better conformed to the primer constraints. The appropriate parameter settings for the proposed MA method were identified and carefully evaluated to design robust PCR-RFLP primers for SNP genotyping. Compared with the former GA method, the MA method is more feasible for PCR-RFLP SNP genotyping.

INDEX TERMS Genetic algorithm, memetic algorithm, polymerase chain reaction-restriction fragment length polymorphism, single nucleotide polymorphism genotyping.

I. INTRODUCTION

Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) is an established, inexpensive, and accurate laboratory technology to study the origins and variations in genetic mutations, as well as complex genetic diseases [1], [2]. In many genetic variations, single nucleotide polymorphisms (SNPs) are used as biological markers, and are the most common markers in humans. SNPs have been used for diverse applications, including forensics [3], personalized medicine [4], evolutionary studies [5], pharmacogenetic analysis [6], [7], preventive medicine [8], [9], and malignancy studies [10], [11]. To measure these genetic variations, SNP genotyping uses PCR-RFLP to determine the genotype included in the experimental sequence. Before performing PCR-RFLP it is absolutely necessary to identify the available restriction enzymes and practical primers.

There are several PCR-RFLP assay systems for SNP genotyping, which include V-MitoSNP [12], SNP-RFLPing [13]–[15], SNP Cutter [1], Methyl-Typing [16], and

Prim-SNPing [17]. These are useful for specific functions, but are limited in the ability to design PCR-RFLP assay primers to enable SNP genotyping [18]–[22]. In 2013, a valid method that implemented a genetic algorithm (GA) with the updated restriction enzymes mining core of SNP-RFLPing [13]–[15] was proposed as an effective means of designing PCR-RFLP assay primers for SNP genotyping. Pivotal parameters of the GA method have been measured with the goal of obtaining more feasible results. However, the GA method is fatally flawed because it is easily trapped into local optima. Thus, the design of PCR-RFLP assay primers for SNP genotyping needs to be improved.

A local search mechanism can effectively avoid the problem. A memetic algorithm (MA) that implies a local search mechanism was previously introduced to enable the design of PCR primers [23]. The method has been discussed and areas of refinement have been identified [24], [25], which inspired us to enhance and improve the method.

In this study, the previously reported novel restriction enzymes hunting (REHUNT) package [26] was applied. Many accessible primer evaluation criteria are available [20], [23], [27]–[30]. Furthermore, the calculation of a more accurate thermodynamic primer melting temperature [31] was considered for inclusion. Using the criteria of the GA method, the different parameter settings of the MA method were evaluated *in silico* and the results were compared with the GA method using SNPs of *SLC6A4*. Finally, the appropriate MA parameter settings were determined, which enables the feasible designing of PCR-RFLP assay primers for SNP genotyping.

II. MATERIALS AND METHODS

A. ABBREVIATIONS AND ACRONYMS DEFINITION OF THE PROBLEM OF PCR-RFLP ASSAY PRIMERS FOR SNP GENOTYPING

A target SNP is included in a DNA template sequence, T_D. The objective of SNP genotyping is to identify available restriction enzymes that can recognize the genotype of the target SNP, and to design feasible PCR-RFLP assay primers. The equation for the formulation of T_D is:

$$T_D = \{B_i | i \text{ is the index of DNA sequence, } \exists! B_i \in \text{target SNP}\} \quad (1)$$

where B_i represents the 'A,' 'T', 'C', and 'G' nucleotides or SNP. SNP is identified according to the SNP IUPAC code ('M', 'R', 'W', 'S', 'Y', 'K', 'V', 'H', 'D', 'B' or 'N') or the dNTP format ([dNTP1/dNTP2]). The symbol ∃! denotes existence and uniqueness. For the general design, this study only focused on true SNPs as described in dbSNP [32] of the National Center for Biotechnology Information as the target SNP. The presented method is not considered for deletion/insertion polymorphisms (DIPs) and multi-nucleotide polymorphisms (MNPs).

Firstly, a pair of sub-sequences of corresponding constraints from T_D and one restriction enzyme that can recognize the genotype of the target SNP must be identified. The pair of sub-sequences, designated the forward primer (P_f) and reverse primer (P_r) are presented as in equation (2) and (3), respectively.

$$P_f = \{B_i | \forall B_i \in \{\text{'A', 'T', 'C', 'G'}\}, F_s \le i \le F_e, i \text{ is the index of } T_D\} \quad (2)$$

$$P_r = \{\overline{B_i} | \forall B_i \in \{\text{'A', 'T', 'C', 'G'}\}, R_s \le i \le R_e, i \text{ is the index of } T_D\} \quad (3)$$

where F_s and F_e indicate the start index and the end index of P_f in T_D, respectively; R_s and R_e indicate the start index and the end index of P_r in T_D, respectively; and $\{\overline{B_i}\}$ is the anti-sense sequence of B_i. For example, for the sequence CATCGAATCTGCGTCTTATGCC, the complementary sequence is GTAGCTTAGACGCAGAATACGG, based on the established A-T and C-G nucleotide complementary pairings, while the anti-sense sequence is the reverse of the complementary sequence, i.e., GGCATAAGACGCAGATTCGATG.

B. OTHER RECOMMENDATIONS PROPOSED ENVIRONARY COMPUTATION METHOD FOR DESIGN OF PCR-RFLP ASSAY PRIMERS

In the proposed evolutionary computation methods, a chromosome encoding vector P_v (termed "individual" in MA and GA) involves four elements—F_s, F_l, P_l, and R_l (as shown in Figure 1)—to derive feasible PCR-RFLP assay primers.

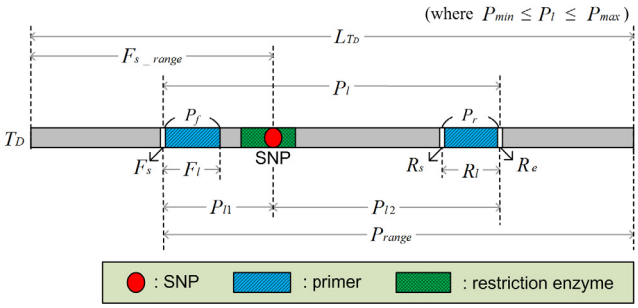

FIGURE 1. Illustration of parameters for PCR-RFLP primer design.

The chromosome encoding vector is expressed by:

$$P_v = (F_s, F_l, P_l, R_l, E) \quad (4)$$

The start position, R_s, of the reverse primer can be calculated by the chromosome encoding vector P_v:

$$R_s = F_s + P_l - R_l \quad (5)$$

From the known five variables of F_s, F_l, P_l, R_s, and R_l, the forward primer and the reverse primer can be immediately derived. Therefore, P_v is the main vector applied to perform the evolutionary computations in designing the PCR-RFLP assay primers.

The proposed method involves seven separate steps. In order, the steps are: 1) mining for available restriction

enzymes from the restriction enzyme database (REBASE), 2) evaluation of availability for the design of PCR-RFLP assay primers, 3) generation of a random initial population, 4) evaluation of fitness, 5) local search, 6) judging termination criteria, and 7) evolutionary operations. These steps are detailed subsequently and outlined in the flowchart shown in Figure 2.

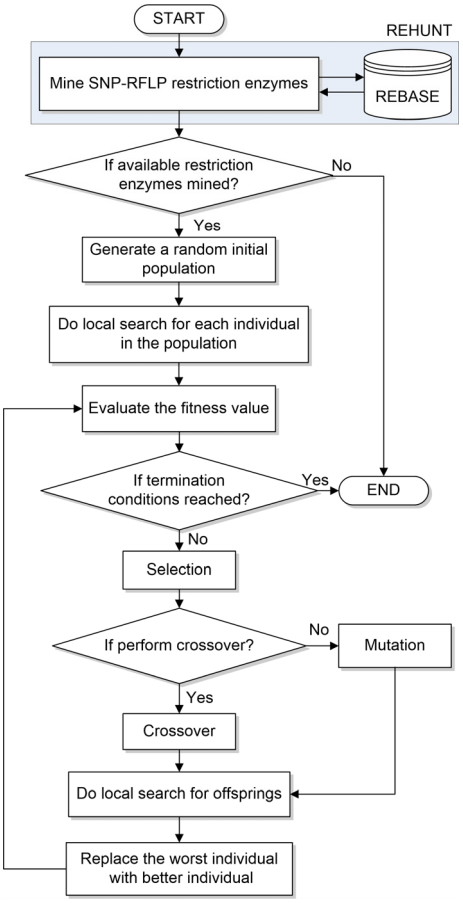

FIGURE 2. Work flowchart for the PCR-RFLP primer design using the MA method.

1) MINING FOR AVAILABLE RESTRICTION ENZYMES FROM REBASE

It is essential to provide the available restriction enzymes to recognize the genotype of the target SNP. This is achieved using REBASE [33]. The database is updated periodically. The proposed method also used the novel REHUNT [26], which provides reliable and effective means to mine available restriction enzymes from REBASE.

2) EVALUATION OF AVAILABILITY FOR THE DESIGN OF PCR-RFLP ASSAY PRIMERS

All restriction enzymes in REBASE [33] are evaluated in turn to discover the available restriction enzymes. If no available restriction enzymes are discovered, the design of the PCR-RFLP assay primers will be insignificant. Therefore, the proposed method terminates directly when no available restriction enzymes are discovered. Otherwise, the following steps are performed in sequence.

3) GENERATION OF RANDOM INITIAL POPULATION

The initial population is the incipient solution provided in the proposed evolutionary computation methods. A fixed number of unduplicated individuals (P_v) are generated randomly as an initial population. In P_v, F_s is generated randomly between 1 and ($L_{TD} - P_{min} + 1$), in which the range is F_{s_range} as shown in Figure 1. F_l is generated randomly between the minimum and the maximum primer lengths. In this study, the minimum and maximum primer lengths were set to 16 and 28 nt, respectively. The method also randomly generates P_l between P_{min} and P_{max} (i.e., the range, P_{range}, as shown in Figure 1) to limit the PCR product length. In this study, P_{min} was set to 100 bps and P_{max} is set to 300 bps. Eventually, R_l is generated randomly, as for F_l.

4) EVALUATION OF FITNESS

The minimized fitness value is used to determine the designed result between good and bad in the proposed method. An experienced fitness function is designed consistent with the universal primer constraints for evaluating the fitness value of each individual in the population. More details for the fitness function have been previously provided [19], [22]. In GA method, the melting temperature, T_{m}, was calculated by the Wallace's formula [34] (Eq. 6), which is simple and considered suitable for approximate nearest-neighbor thermodynamic calculations [35]. However, many researchers consider the Wallace's formula to be inaccurate and do not support its use in primer design. Therefore, the nearest-neighbor thermodynamic calculation is applied to improve the evaluation of T_{m}. The improved T_{m} is calculated by SantaLucia's formula (Eq. 7) [31].

$$Tm_{wallace}(P) = (\#G + \#C) \times 4 + (\#A + \#T) \times 2 \quad (6)$$

$$Tm_{santalucia}(P) = \frac{\Delta H^{\circ}(predicted) \times 1000}{(\Delta S^{\circ}(\text{salt\_correction}) + R \times \ln(C_T/4))} - 273.15 \quad (7)$$

where $\Delta H^{\circ}(predicted)$ is the enthalpy; $\Delta S^{\circ}(salt\_correction)$ is the entropy correction; R is the gas constant (1.987 cal/Kmol) and C_T is the DNA concentration.

The $\Delta H^{\circ}(predicted)$ is calculated by

$$\begin{aligned}\Delta H^{\circ}(predicted) &= \Delta H^{\circ}(AA) + \Delta H^{\circ}(AC) + \Delta H^{\circ}(AG) + \Delta H^{\circ}(AT) \\ &+ \Delta H^{\circ}(CA) + \Delta H^{\circ}(CC) + \Delta H^{\circ}(CG) + \Delta H^{\circ}(CT) \\ &+ \Delta H^{\circ}(GA) + \Delta H^{\circ}(GC) + \Delta H^{\circ}(GG) + \Delta H^{\circ}(GT) \\ &+ \Delta H^{\circ}(TA) + \Delta H^{\circ}(TC) + \Delta H^{\circ}(TG) + \Delta H^{\circ}(TT) \\ &+ \Delta H^{\circ}(init.w/term.\,GC) + \Delta H^{\circ}(init.w/term.\,AT) \quad (8)\end{aligned}$$

The $\Delta S^{\circ}(predicted)$ is calculated by

$$\begin{aligned}\Delta S^{\circ}(predicted) &= \Delta S^{\circ}(AA) + \Delta S^{\circ}(AC) + \Delta S^{\circ}(AG) + \Delta S^{\circ}(AT) \\ &+ \Delta S^{\circ}(CA) + \Delta S^{\circ}(CC) + \Delta S^{\circ}(CG) + \Delta S^{\circ}(CT) \\ &+ \Delta S^{\circ}(GA) + \Delta S^{\circ}(GC) + \Delta S^{\circ}(GG) + \Delta S^{\circ}(GT) \\ &+ \Delta S^{\circ}(TA) + \Delta S^{\circ}(TC) + \Delta S^{\circ}(TG) + \Delta S^{\circ}(TT) \\ &+ \Delta S^{\circ}(init.w/term.\ GC) + \Delta S^{\circ}(init.w/term.\ AT) \quad (9)\end{aligned}$$

5) LOCAL SEARCH

A local search mechanism exploits a preset limited area based on current individuals to provide other outstanding individuals. Local optima can be obtained by the local search operation to improve the experiences of individuals in the population. The proposed method firstly performs the local search operation in an initial iteration to provide an improved initial population. Furthermore, the local search operation is also applied to the new offspring after crossover or mutation operation to further improve the offspring. Iteration-by-iteration, the global optimum is eventually achieved. The pseudo-code of the local search operation in the proposed MA method for the design of PCR-RFLP assay primers is determined in the following sequence of operations:

```
1 Begin;
2   Select an incremental value d = a*Rand();
3   For a given individual i ∈ Population: calculate fitness
    (i);
4   For j = 1 to the number of variables of individual i;
5     value(j) = value(j) + d;
6   If the fitness of the individual does not be improved
    then
7     value(j) = value(j)-d;
8   else if the fitness of the individual is improved then
9     retain value(j);
10  Next j;
11 End;
```

In the above pseudo-code, d is an incremental value that is used to assist an individual in exploiting neighboring individuals. Constant a is a preset value set to limit the range of local search. To appropriately set the constant a, the variables of F_l and R_l are selected to determine the constant a based on a maximum permitted range for a chromosome vector, P_v. Therefore, the constant a is determined by the difference of the maximum primer length and the minimum primer length (here a was obtained to 12 bps).

6) JUDGING TERMINATION CRITERIA

The proposed method implements two termination criteria: in first, the fitness value of zero is discovered for one individual; in second, a preset iteration is completed. Continuing iterations will increase the computation time and waste the computation resources. In addition, the preset iteration settings are sensitive for the result of primer design, and hence must be set carefully.

7) EVOLUTIONARY OPERATIONS

Both MA and a GA have evolutionary operations, including selection, crossover, mutation, and replacement. Two individuals are selected randomly from the population during the selection operation. In particular, alternative of crossover and mutation are used in the proposed method. As long as a performed probability meets the preset crossover rate, the two selected individuals will apply the uniform crossover operation to generate two new offspring. When a performed probability meets the preset mutation rate, one of the selected two individuals will be selected randomly and apply one point mutation to generate a new offspring. More details for the operations are provided elsewhere [19].

III. RESULTS AND DISCUSSION

Many small- and medium-sized laboratories utilize the easy-to-use PCR-RFLP assay for SNP genotyping. However, the design of the assay primers limits its use to only a few methods and systems. An improved method for providing PCR-RFLP assay primers is proposed in this study to facilitate PCR-RFLP assay for SNP genotyping. The method was assessed by using different parameter settings, including population size, crossover rate, and mutation rate. Better parameter settings were selected for PCR-RFLP primer design in a data set with 288 SNPs in *SLC6A4* and the results were compared with the GA method. Simultaneously, the SantaLucia's thermodynamic formula was used to measure the method.

A. DATA SET AND COMPUTATIONAL ENVIRONMENT

The same 288 SNP *SLC6A4* data set used in the GA method was used for the comparison of the MA and GA methods. DIPs and MNPs were excluded from the 288 SNPs. SNP-Flankplus [36] retrieved a total length of 500 bp flanking sequences for every SNP to evaluate the proposed method; these DNA templates have been described [19]. As done in this previous study, the method was presently also run on the same computational environment with an Intel(R) Core 2 CPU of 1.86 GHz and 1 GB RAM with the Microsoft Windows XP SP3 and JAVA 6.0 platforms.

B. PARAMETER SETTINGS

The method set common primer constraints [23], [28] that included primer length between 16 and 28 nt, GC% between 40 and 60%, primer T_m between 45 and 62°C, primer T_m difference within 5°C, PCR product length larger than 100 bp, and PCR product length ratio 1:2:3. As with the GA, the number of iterations, population size, crossover rate, and mutation rate were set as 1000, 50, 0.6, and 0.001, respectively. We first applies the parameter settings based on the DeJong and Spears measure [37] to evaluate the proposed method, followed by different population sizes, crossover rates, and mutation rates to evaluate the proposed method. The evaluations provided better parameter settings.

TABLE 1. Computer simulation results for 251 *SLC6A4* SNPs based on DeJong and Spears parameters.

Methods (T_m formula)	Constraints									
	Primer length difference	GC%	GC clamp	T_m	T_m difference	PCR product length	Dimer	Hairpin	Specificity	Average fitness
MA (Wallace)	99.20%	83.27%	97.01%	97.61%	97.21%	100.00%	100.00%	99.40%	100.00%	1.17
MA (SantaLucia)	100.00%	89.78%	99.72%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0.05
GA (Wallace)	97.61%	78.69%	65.94%	97.61%	94.82%	99.87%	100.00%	99.80%	100.00%	4.74

TABLE 2. Computer simulation results in different population sizes for 251 SNPS of *SLC6A4*.

Population size	Constraints									
	Primer length difference	GC%	GC clamp	T_m	T_m difference	PCR product length	Dimer	Hairpin	Specificity	Average fitness
100	100.00%	85.66%	98.80%	98.61%	99.60%	100.00%	100.00%	100.00%	100.00%	0.79
200	99.60%	90.84%	99.60%	99.00%	99.60%	100.00%	100.00%	100.00%	100.00%	0.56
300	99.60%	87.65%	100.00%	98.61%	99.60%	100.00%	100.00%	99.80%	100.00%	0.51
400	99.20%	91.63%	99.80%	98.61%	99.60%	100.00%	100.00%	100.00%	100.00%	0.50
500	100.00%	91.24%	100.00%	98.80%	100.00%	100.00%	100.00%	100.00%	100.00%	0.46
600	100.00%	91.43%	99.60%	98.80%	100.00%	100.00%	100.00%	100.00%	100.00%	0.44
700	100.00%	89.64%	99.80%	98.80%	100.00%	100.00%	100.00%	100.00%	100.00%	0.45
800	100.00%	93.03%	100.00%	98.80%	100.00%	100.00%	100.00%	100.00%	100.00%	0.43
900	100.00%	91.83%	99.60%	98.80%	100.00%	100.00%	100.00%	100.00%	100.00%	0.46
1000	99.20%	92.23%	97.61%	99.20%	99.20%	99.87%	100.00%	100.00%	100.00%	0.43
mean±	99.76%±	90.52%±	99.48%±	98.80%±	99.76%±	99.99%±	100.00%±	99.98%±	100.00%±	0.54±
SD	0.34%	4.54%	1.49%	0.38%	0.28%	0.13%	0.00%	0.13%	0.00%	0.14

Highlighted rows show better average fitness

C. IN SILICO SIMULATION RESULTS

In the 288 *SLC6A4*SNPs, 251 SNPs with available restriction enzymes were identified from the REBASE version 906 updated on 29 May 2009, and 251 feasible primer pairs were designed. When the REBASE version 402 updated on 30 January 2014 was used, 181 SNPs with available restriction enzymes were identified, and 181 feasible primer pairs were designed. The results of the GA method are provided in a supplemental file that can be downloaded from https://sites.google.com/site/yhcheng1981/downloads/ma-ga-pcr-rflp-assay-primers.

1) RESULTS OF THE DEJONG AND SPEARS PARAMETER SETTINGS IN THE WALLACE'S FORMULA

The computer simulation results based on the DeJong and Spears parameter settings are shown in Table 1.

When the method used the same Wallace's formula to calculate T_m, the accuracy of the MA method was 1.59% higher than the GA method in terms of primer length difference. The accuracy of the MA method was 4.58% and 31.07% higher than the GA method in terms of GC% and GC clamp, respectively. The accuracy of the MA and GA methods were equivalent in terms of T_m. The accuracy of the MA method was 2.39% higher than the GA method in terms of T_m difference. The accuracy of the MA method in achieving perfect PCR product length was 0.13% higher than the GA method. The MA and GA methods were 100% in terms of dimmers. The accuracy of the MA method was 0.4% lower than that the GA method in terms of hairpin. Finally, the MA and GA methods displayed 100% specificity. These evaluations indicated the superiority of the MA method compared to the GA method. The MA and GA methods had and an average fitness of 1.17 and 4.74, respectively, for 251 SNPs. Therefore, the design of PCR-RFLP assay primers using the MA method was more suitable for designing primers that conformed to the primer constraints than the GA method based on the DeJong and Spears parameter settings in the Wallace's formula.

2) RESULTS OF DIFFERENT POPULATION SIZES

The population size provides the permitted solutions for PCR-RFLP primers. To estimate the influences of the population size, the population size was set to 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000. These results are shown in Table 2.

A better average fitness value of 0.43 was evident in the population sizes of 800 and 1000. All primer constraints in the mean and standard deviation (SD) were satisfactory (means >90%, SDs <0.4%, excluding the GC clamp, which was 1.49%), except GC% (mean, 90.52%; SD, 4.54%). Furthermore, the average fitness seemed to improve gradually with increased population size. The population size of 1000 displayed a better average fitness value of 0.43, which was better than that of GA (0.82) (Please see the supplemental file).

3) RESULTS OF DIFFERENT CROSSOVER RATES

The crossover operation helps individuals to exchange information to attain a more feasible solution than before

TABLE 3. Computer simulation results in different crossover rates for 251 SNPS of *SLC6A4*.

Crossover rate	Constraints									
	Primer length difference	GC%	GC clamp	T_m	T_m difference	PCR product length	Dimer	Hairpin	Specificity	Average fitness
0.1	98.41%	80.28%	94.62%	97.21%	98.01%	100.00%	100.00%	99.80%	99.80%	1.64
0.2	99.60%	80.48%	96.22%	97.01%	99.60%	100.00%	100.00%	99.80%	100.00%	1.6
0.3	98.41%	79.88%	97.41%	98.41%	98.80%	99.73%	100.00%	100.00%	99.80%	1.14
0.4	98.01%	81.67%	96.81%	97.61%	99.20%	100.00%	100.00%	100.00%	100.00%	1.45
0.5	99.20%	83.67%	96.22%	98.21%	98.41%	99.73%	100.00%	100.00%	100.00%	1.16
0.6	99.20%	83.27%	97.01%	97.61%	97.21%	100.00%	100.00%	99.40%	100.00%	1.17
0.7	99.60%	84.46%	97.21%	98.80%	98.80%	99.87%	100.00%	100.00%	100.00%	0.95
0.8	98.41%	82.07%	98.61%	98.41%	99.20%	100.00%	100.00%	99.80%	100.00%	0.93
0.9	99.60%	82.07%	98.21%	98.80%	99.20%	99.73%	100.00%	100.00%	100.00%	0.9
1.0	99.60%	83.07%	99.00%	98.61%	99.60%	100.00%	100.00%	100.00%	100.00%	0.81
mean±	99.00%±	82.09%±	97.13%±	98.07%±	98.8%±	99.91%±	100.00%±	99.88%±	99.96%±	1.18±
SD	0.63%	2.98%	2.00%	1.24%	0.74%	0.39%	0.00%	0.40%	0.13%	0.26

Highlighted rows show better average fitness

TABLE 4. Computer simulation results in different mutation rates for 251 SNPS of *SLC6A4*.

Mutation rate	Constraints									
	Primer length difference	GC%	GC clamp	T_m	T_m difference	PCR product length	Dimer	Hairpin	Specificity	Average fitness
0.01	97.21%	82.07%	95.22%	95.42%	93.23%	99.07%	100.00%	98.61%	100.00%	0.93
0.02	97.21%	80.48%	93.63%	94.42%	91.24%	98.27%	100.00%	99.00%	99.60%	0.98
0.03	97.21%	80.28%	91.04%	93.63%	87.25%	96.81%	100.00%	97.41%	99.80%	0.74
0.04	94.82%	83.86%	91.24%	92.63%	83.27%	97.34%	100.00%	98.41%	100.00%	0.75
0.05	92.43%	79.88%	87.85%	88.65%	74.90%	96.28%	100.00%	97.81%	99.60%	0.61
0.06	92.43%	78.69%	84.66%	87.05%	72.91%	95.88%	100.00%	95.62%	99.60%	0.69
0.07	93.63%	80.48%	84.06%	88.05%	73.31%	95.48%	100.00%	97.01%	99.80%	0.68
0.08	94.82%	79.68%	82.47%	86.85%	70.92%	94.82%	100.00%	95.82%	99.60%	0.67
0.09	89.24%	75.30%	79.48%	83.27%	65.74%	93.89%	100.00%	96.22%	99.20%	0.54
0.10	90.84%	77.49%	80.08%	84.86%	66.14%	94.02%	100.00%	97.21%	99.80%	0.58
mean±	93.98%±	79.82%±	86.97%±	89.48%±	77.80%±	96.19%±	100.00%±	97.31%±	99.70%±	0.72±
SD	2.79%	4.70%	11.26%	8.49%	10.11%	5.18%	0.00%	2.35%	0.47%	0.14

Highlighted rows show better average fitness

exchange. Therefore, estimating the influences of crossover rates is useful for the design of PCR-RFLP assay primers. The results of the crossover rates of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 in this study are shown in Table 3.

The crossover rate of 1.0 yielded a better average fitness of 0.81. All primer constraints in the mean and SD produced satisfactory results, except GC% (means >97%, SDs <0.8%, excluding the GC clamp and T_m, which were 2.00% and 1.24%, respectively). Furthermore, the average fitness improved gradually with increased crossover rate. When the crossover rate was set to the highest value of 1.0, the average fitness value with the best result was also obtained. The crossover rate of 1.0 produced a better average fitness value of 0.81 for the MA method, which was obviously better than that of the 3.49 for the GA method (Please see the supplemental file).

4) RESULTS OF DIFFERENT MUTATION RATES

Mutation is used to generate a diverse solution to avoid solutions falling into the local optimum. The mutation rate was set to 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, and 0.10 to perform the MA method to design PCR-RFLP assay primers. The results are shown in Table 4.

The mutation rate of 0.09 yielded a better average fitness value of 0.54. All primer constraints in the mean displayed satisfactory results (means >86%), except for GC% and T_m difference (both <80%). However, the primer constraints in SD were not ideal, and only dimer and specificity were preferable at <2%. The mutation rate of 0.09 in the MA method produced an average fitness value of 0.54 that was obviously better than that of the GA method, which, with the mutation rate set to 0.02, displayed a better average fitness value of 4.37 (Please see the supplemental file).

5) RESULTS OF THE BETTER PARAMETER SETTINGS

After the evaluations for the population sizes, crossover rates, and mutation rates, the better parameter settings among these parameters are used to evaluate the design of PCR-RFLP assay primers. The results are shown in Table 5.

In the table, the population sizes were set to 800 and 1000, crossover rate was set to 1.0, and mutation rate was set to 0.09. The average fitness was 0.42 and 0.45 for the population size of 800 and 1000, respectively. The results for average fitness are similar to those in Table 2 when the population size was set to 800 and 1000. Therefore, the population sizes may

TABLE 5. Computer simulation results based on better parameter settings and crossover rate set to 1.0, and mutation rate set to 0.09 for the 251 SNPs of *SLC6A4*.

Population size	Constraints									
	Primer length difference	GC%	GC clamp	T_m	T_m difference	PCR product length	Dimer	Hairpin	Specificity	Average fitness
800	99.60%	91.04%	98.41%	98.01%	97.61%	99.73%	100.00%	99.60%	100.00%	0.42
1000	99.20%	90.24%	97.61%	97.41%	97.21%	99.47%	100.00%	100.00%	99.80%	0.45

TABLE 6. Computer simulation results based on better parameter settings and crossover rate set to 1.0, and mutation rate set to 0.09 for the 251 SNPs of *SLC6A4*.

Characteristics	Methods	
	MA	GA
Melting temperature formula	Wallace's formula / thermodynamic SantaLucia's formula	Wallace's formula
REBASE version	REBASE version 402 (updated on 30 January 2014)	REBASE version 906 (updated on 29 May 2009)
Primer constraints satisfied	Most designed primers conformed to the preset primer constraints, and the results were superior to those of the GA method.	Most designed primers conformed to the preset primer constraints, but the results were worse than those of the MA method.
Best average fitness	0.42	0.49
Worst average fitness	1.64	5.92

be the main parameter influencing the design of PCR-RFLP assay primers.

6) RESULTS OF THE DEJONG AND SPEARS PARAMETER SETTINGS IN THE SANTALUCIA THERMODYNAMIC FORMULA

When the method used the SantaLucia's thermodynamic formula to calculate T_m, all results were better than those obtained using the Wallace's formula for the MA method and the GA method (see Table 1). Most of the designed primers displayed 100% conformance to the preset primer constraints, excepting GC% (89.78%) and GC clamp (99.72%). Accordingly, the design of PCR-RFLP assay primers using the MA method with the SantaLucia's thermodynamic formula was more suitable for designing primers that conformed to the primer constraints.

D. COMPARISON OF THE METHOD WITH THE FORMER GA

The former GA method is the first evolutionary computation method for PCR-RFLP primer design for SNP genotyping. Its abilities have been proven [19]. In this study, the MA method was applied to improve the PCR-RFLP primer design for SNP genotyping. The characteristics of the GA and MA methods were compared. The results are summarized in Table 4.

The best and worst average fitness values were computed based on the evaluation in the data set for 288 SNPs of *SLC6A4*. In the melting temperature formula calculation, the MA method retained the Wallace's formula and provided an accurate SantaLucia's formula. However, the GA method only used the Wallace's formula. In restriction enzymes mining, the MA method uses the REBASE version 402 updated on 30 January 2014, but the GA method used the older REBASE version 906 updated on 29 May 2009. In the primer constraints evaluation, the primers designed using the MA method conformed to the preset primer constraint criteria better than those prepared using the GA method. Finally, the MA method obtained the best average fitness value and the worst average fitness value of 0.42 and 1.64, respectively, while the GA method obtained the best average fitness value and the worst average fitness value of 0.49 and 5.92, respectively (Please see the supplemental file).

Furthermore, the time complexity analysis of the GA and MA methods were also analyzed. In the GA primer design method, the iteration number and the population size are the important factors that primarily affects the time complexity. Since the crossover and mutation operations in the GA primer design method are computed by constant time, we can ignore them in the time complexity analysis. To consider n is the iteration number and p is the population size, we can get the worst case complexity $T(n) = n \times p$. Therefore, the time complexity of the GA primer design is $O(n \times p)$.

In the MA primer design method, in addition to the iteration number and the population size, the local search mechanism is another extra important factor that must be considered affecting the time complexity. The crossover and mutation operations in the MA primer design method are also can be ignored in the time complexity analysis due to their constant time. To consider n is the iteration number, p is the population size and v is the length of chromosome encoding vector, we can get the worst case complexity $T(n) = n \times p + n \times v = n \times (p + v)$. The computed time of $n \times v$ is for the firstly performs the local search operation in an initial iteration to provide an improved initial population. The time for the local search applied to the new offspring after crossover or mutation operation can be ignored due to only two new offsprings with the constant time. Therefore, the time complexity of the MA primer design is $O(n \times (p + v))$.

Finally, the new parts proposed compared to the former GA method is listed below to show the innovation:

(1) The accurate thermodynamic SantaLucia's formula for calculation of melting temperature is employed in the MA method.
(2) The novel reliable and open source REHUNT is integrated into the MA method for efficiently restriction enzymes mining.

(3) The distinctive local search mechanism implemented in the MA method assists the selection of primers that better conformed to the primer constraints.
(4) More robust PCR-RFLP primers for SNP genotyping are designed by the MA method than the former GA method.

IV. CONCLUSION

Providing effective and available PCR-RFLP primers is extremely important for small-scale basic research studies of complex genetic diseases associated with SNPs. The MA method is considered capable of designing more feasible PCR-RFLP assay primers to achieve SNP genotyping. The present improvements to the method included an accurate thermodynamic SantaLucia's formula for melting temperature calculation, use of the novel REHUNT package for mining of available restriction enzymes, and the design of primers that better conformed to the preset primer constraints. The appropriate parameter settings for the MA method have been carefully evaluated for the design of PCR-RFLP assay primers for SNP genotyping. In contrast to the GA method, the proposed MA method is more feasible for achieving PCR-RFLP SNP genotyping.

REFERENCES

[1] R. Zhang *et al.*, "SNP cutter: A comprehensive tool for SNP PCR-RFLP assay design," (in English), *Nucleic Acids Res.*, vol. 33, pp. W489–492, Jul. 2005.
[2] M. Ota, H. Fukushima, J. K. Kulski, and H. Inoko, "Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism," (in English), *Nat. Protocol*, vol. 2, no. 11, pp. 2857–2864, 2007.
[3] A.-M. Divne and M. Allen, "A DNA microarray system for forensic SNP analysis," *Forensic Sci. Int.*, vol. 154, nos. 2–3, pp. 111–121, 2005.
[4] W. Sadée, "Pharmacogenomics: Harbinger for the era of personalized medicine?" *Mol. Intervent.*, vol. 5, no. 3, pp. 140–143, 2005.
[5] J. G. Hacia *et al.*, "Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays," (in English), *Nat. Genet.*, vol. 22, no. 2, pp. 164–167, Jun. 1999.
[6] V. Mooser, D. M. Waterworth, T. Isenhour, and L. Middleton, "Cardiovascular pharmacogenetics in the SNP era," (in English), *J. Thrombosis Haemostasis*, vol. 1, no. 7, pp. 1398–1402, Jul. 2003.
[7] C.-H. Yang, Y.-H. Cheng, L.-Y. Chuang, and H.-W. Chang, "Drug-SNPing: An integrated drug-based, protein interaction-based tagSNP-based pharmacogenomics platform for SNP genotyping," *Bioinformatics*, vol. 29, no. 6, pp. 758–764, 2013.
[8] C. R. Cantor, "The use of genetic SNPs as new diagnostic markers in preventive medicine," *Ann. New York Acad. Sci.*, vol. 1055, no. 1, pp. 48–57, 2005.
[9] T. Bernig and S. J. Chanock, "Challenges of SNP genotyping and genetic variation: Its future role in diagnosis and treatment of cancer," *Expert Rev. Mol. Diagnostics*, vol. 6, no. 3, pp. 319–331, 2006.
[10] H. C. Erichsen and S. J. Chanock, "SNPs in cancer research and treatment," *Brit. J. Cancer*, vol. 90, no. 4, pp. 747–751, 2004.
[11] L. J. Engle, C. L. Simpson, and J. E. Landers, "Using high-throughput SNP technologies to study cancer," *Oncogene*, vol. 25, no. 11, pp. 1594–1601, 2006.
[12] L. Y. Chuang *et al.*, "V-MitoSNP: Visualization of human mitochondrial SNPs," (in English), *BMC Bioinf.*, vol. 7, p. 379, Aug. 2006.
[13] H.-W. Chang, C.-H. Yang, P.-L. Chang, Y.-H. Cheng, and L.-Y. Chuang, "SNP-RFLPing: Restriction enzyme mining for SNPs in genomes," *BMC Genomics*, vol. 7, p. 30, Dec. 2006.
[14] L. Y. Chuang *et al.*, "Restriction enzyme mining for SNPs in genomes," (in English), *Anticancer Res.*, vol. 28, no. 4A, pp. 2001–2007, Jul./Aug. 2008.
[15] H.-W. Chang, Y.-H. Cheng, L.-Y. Chuang, and C.-H. Yang, "SNP-RFLPing 2: An updated and integrated PCR-RFLP tool for SNP genotyping," (in English), *BMC Bioinf.*, vol. 11, p. 173, Apr. 2010.
[16] C.-H. Yang, L.-Y. Chuang, Y.-H. Cheng, D.-L. Gu, C.-H. Chen, and H.-W. Chang, "Methyl-typing: An improved and visualized COBRA software for epigenomic studies," *FEBS Lett.*, vol. 584, no. 4, pp. 739–744, 2010.
[17] H.-W. Chang *et al.*, "Prim-SNPing: A primer designer for cost-effective SNP genotyping," (in English), *Biotechniques*, vol. 46, no. 6, pp. 421–431, May 2009.
[18] Y.-H. Cheng, L.-Y. Chuang, and C.-H. Yang, "Single nucleotide polymorphisms polymerase chain reaction-restriction fragment length polymorphism primer design using genetic algorithm with fixed and adaptive stopping criteria," *J. Adv. Math. Appl.*, vol. 2, no. 2, pp. 172–181, 2013.
[19] L.-Y. Chuang, Y.-H. Cheng, C.-H. Yang, and C.-H. Yang, "Associate PCR-RFLP assay design with SNPs based on genetic algorithm in appropriate parameters estimation," *IEEE Trans. Nanobiosci.*, vol. 12, no. 2, pp. 119–127, Jun. 2013.
[20] Y.-H. Cheng, "A novel teaching-learning-based optimization for improved mutagenic primer design in mismatch PCR-RFLP SNP genotyping," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 1, pp. 86–98, Jan./Feb. 2016.
[21] Y.-H. Cheng, C.-N. Kuo, and C.-M. Lai, "An improved evolutionary method with test in different crossover rates for PCR-RFLP SNP genotyping primer design," *Int. J. Mining, Metall., Mech. Eng.*, vol. 4, no. 1, pp. 25–29, 2016.
[22] Y.-H. Cheng, C.-N. Kuo, and C.-M. Lai, "Effective natural PCR-RFLP primer design for SNP genotyping using teaching-learning-based optimization with elite strategy," *IEEE Trans. Nanobiosci.*, vol. 15, no. 7, pp. 657–665, Oct. 2016.
[23] L.-Y. Chuang, Y.-H. Cheng, and C.-H. Yang, "Specific primer design for the polymerase chain reaction," *Biotechnol. Lett.*, vol. 35, no. 10, pp. 1541–1549, 2013.
[24] Y.-H. Cheng, L.-Y. Chuang, and C.-H. Yang, "Estimation of the influence for different population sizes in MA-based natural SNP-RFLP primer design," in *Proc. Conf. Technol. Appl. Artif. Intell. (TAAI)*, 2012, pp. 68–73.
[25] Y.-H. Cheng, L.-Y. Chuang, and C.-H. Yang, "Natural SNP-RFLP primer design using memetic algorithm," in *Proc. 29th Workshop Combinat. Math. Comput. Theory*, 2012, pp. 124–129.
[26] Y.-H. Cheng, J.-J. Liaw, and C.-N. Kuo, "REHUNT: A reliable and open source package for restriction enzyme hunting," *BMC Bioinf.*, vol. 19, no. 1, p. 178, 2018.
[27] C.-H. Yang, Y.-H. Cheng, L.-Y. Chuang, and H.-W. Chang, "Specific PCR product primer design using memetic algorithm," (in English), *Biotechnol Prog*, vol. 25, no. 3, pp. 745–753, May/Jun. 2009.
[28] L.-Y. Chuang, Y.-H. Cheng, and C.-H. Yang, "URPD: A specific product primer design tool," (in English), *BMC Res. Notes*, vol. 5, no. 1, p. 306, Jun. 2012.
[29] Y.-H. Cheng, "Computational intelligence-based polymerase chain reaction primer selection based on a novel teaching-learning-based optimisation," *IET Nanobiotechnol.*, vol. 8, no. 4, pp. 238–246, 2014.
[30] Y.-H. Cheng, "Estimation of teaching-learning-based optimization primer design using regression analysis for different melting temperature calculations," *IEEE Trans. Nanobiosci.*, vol. 14, no. 1, pp. 3–12, Jan. 2015.
[31] J. SantaLucia, Jr., "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," (in English), *Proc. Nat. Acad. Sci. USA*, vol. 95, no. 4, pp. 1460–1465, Feb. 1998.
[32] S. T. Sherry *et al.*, "dbSNP: The NCBI database of genetic variation," *Nucleic Acids Res.*, vol. 29, no. 1, pp. 308–311, Jan. 2001.
[33] R. J. Roberts, T. Vincze, J. Posfai, and D. Macelis, "REBASE—A database for DNA restriction and modification: Enzymes, genes and genomes," (in English), *Nucleic Acids Res.*, vol. 38, pp. D234–D236, Jan. 2010.
[34] R. B. Wallace, J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura, "Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: The effect of single base pair mismatch," (in English), *Nucleic Acids Res.*, vol. 6, no. 11, pp. 3543–3557, Aug. 1979.
[35] V. R. Prezioso, "General notes on primer design in PCR," BioSyst. Lab., Brinkmann Instrum. Inc., Westbury, NY, USA, Tech. Rep., 2006.
[36] C.-H. Yang, Y.-H. Cheng, L.-Y. Chuang, and H.-W. Chang, "SNP-Flankplus: SNP ID-centric retrieval for SNP flanking sequences," (in English), *Bioinformation*, vol. 3, no. 4, pp. 147–149, 2008.

[37] K. A. De Jong and W. M. Spears, *An Analysis of the Interacting Roles of Population Size and Crossover in Genetic Algorithms*. Berlin, Germany: Springer, 1990, pp. 38–47.

YU-HUEI CHENG (M'12–SM'17) received the M.S. and Ph.D. degrees from the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan, in 2006 and 2010, respectively. He crosses many professional fields, including biological and medical engineering, electrical and electronic engineering, and information engineering. He is currently an Associate Professor with the Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung, Taiwan. He has rich experiences in evolutionary computation, optimal design, computer programming, database design and management, and systems programming and design. His research interests include artificial intelligence, automatic control, bioinformatics, biomedical engineering, computational intelligence, embedded systems, electric and hybrid vehicles, Internet of Things, machine learning, mobile medical, power electronics, renewable energy, and robotics. He is also a Senior Member of the Universal Association of Computer and Electronics Engineers and a member of the IAENG (International Association of Engineers) and the Taiwanese Association for Artificial Intelligence (TAAI). He has authored and co-authored over 100 technical papers. A few of his papers have been published in IEEE sponsored leading journals (e.g., the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON RELIABILITY, the IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, the IEEE TRANSACTIONS ON NANOBIOSCIENCE, and the IEEE ACCESS) and conference proceedings (e.g., BIBE, FUZZ, iCAST, ICCI*CC, PEDS, TENCON, VPPC, and so on). In addition, he has also been actively participating in various activities by serving as a Publication Chair of IEEE iCAST 2017, a Technical Program Committee Members of ICAT2E2017, ICSEVEN 2017, and ICSEVEN 2018, a Publicity Chair of IEEE iCAST 2018, and a session chairs/co-chairs of several international conferences, and has reviewed numerous papers for high visibility journals (e.g., *Methods in Ecology and Evolution*, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computational and Structural Biotechnology Journal*, the *International Journal of Hydrogen Energy*, the IEEE ACCESS, *Clinical Therapeutics*, the IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS, *Energies*, *Applied Mathematical Modelling*, the *International Journal of Fuzzy Systems*, *Soft Computing*, the IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, and so on). In terms of academic activities, he is currently the Associate Editor of the IEEE ACCESS.

CHING-MING LAI (S'06–M'10–SM'17) received the B.S. degree (Hons.) in aeronautical engineering from the National Huwei University of Science and Technology, Yunlin, Taiwan, in 2004, the M.S. degree in electrical engineering from National Central University, Chungli, Taiwan, in 2006, and the Ph.D. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2010. From 2009 to 2012, he has served as a Senior R&D Engineer with the Power SBG, Lite-ON Technology Corporation, Taipei, Taiwan, where he worked on the high-efficiency and high-power-density ac/dc power supply. In 2012, he established UPE-Power Technology, Company, Ltd., Taichung, Taiwan, the company is developing switching power supplies and power converters for renewable energy resources. Since 2016, he has been with the Graduate Institute of Mechanical and Electrical Engineering, National Taipei University of Technology (Taipei Tech), Taiwan. He is currently an Associate Professor with the Department of Vehicle Engineering, Taipei Tech. His research interests include electric vehicles, power electronics, and high-efficiency energy power conditioning systems. He is a Life Member of the Taiwan Power Electronics Association and the Society of Automotive Engineers Taipei and a member of the IEEE Power Electronics, the IEEE Industry Applications, and the IEEE Industrial Electronics Societies. He is the fellow of IET. He was a recipient of the 2008 Young Author's Award for Practical Application from the Society of Instrument and Control Engineers, Japan, and the Best Paper Award at the 2013 IEEE International Conference on Power Electronics and Drive Systems. He is also an Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEM II, the IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS, and *IEICE Electronics Express*.

JIASHEN TEH (M'17) received the B.Eng. degree in electrical and electronics engineering from Universiti Tenaga Nasional, Selangor, Malaysia, in 2010, and the Ph.D. degree in electrical and electronics engineering from The University of Manchester, Manchester, U.K., in 2016. He is currently a Senior Lecturer at Universiti Sains Malaysia, Penang, Malaysia. His research interests include power system reliability analyses, renewable energy sources, and smart grid technologies.

CHE-NAN KUO was born in Tainan, Taiwan. He received the B.S. degree from the Department of Computer Science, Tunghai University, Taichung, Taiwan, in 2002, and the M.S. and Ph.D. degrees from the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, in 2004 and 2009, respectively. He is currently an Assistant Professor with the Department of Business Administration, CTBC Financial Management College, Tainan. He has many excellent research papers about folded hypercubes published on famous journals, such as *Theoretical Computer Science*, *Discrete Applied Mathematics*, *Information Sciences*, and *Computers and Mathematics with Applications*. His current research interests include interconnection networks, discrete mathematics, computation theory, graph theory, and algorithm analysis.