MDPI

Article

Predicting DNA Motifs by Using Multi-Objective Hybrid Adaptive Biogeography-Based Optimization

Siling Feng [1,2] , Ziqiang Yang [1] and Mengxing Huang [1,2,]

[1] College of Information Science & Technology, Hainan University, No. 58 Renmin Avenue,
    Hai'kou 570228, China; fengsiling2008@163.com (S.F.); comcxziqiang@gmail.com (Z.Y.)
[2] State Key Laboratory of Marine Resource Utilization in the South China Sea, Hainan University,
    No. 58 Renmin Avenue, Hai'kou 570228, China
[] Correspondence: huangmx09@hainu.edu.cn; Tel.: +86-898-6618-7056

Abstract: The computational discovery of DNA motifs is one of the most important problems in molecular biology and computational biology, and it has not yet been resolved in an efficient manner. With previous research, we have solved the single-objective motif discovery problem (MDP) based on biogeography-based optimization (BBO) and gained excellent results. In this study, we apply multi-objective biogeography-based optimization algorithm to the multi-objective motif discovery problem, which refers to discovery of novel transcription factor binding sites in DNA sequences. For this, we propose an improved multi-objective hybridization of adaptive Biogeography-Based Optimization with differential evolution (DE) approach, namely MHABBO, to predict motifs from DNA sequences. In the MHABBO algorithm, the fitness function based on distribution information among the habitat individuals and the Pareto dominance relation are redefined. Based on the relationship between the cost of fitness function and average cost in each generation, the MHABBO algorithm adaptively changes the migration probability and mutation probability. Additionally, the mutation procedure that combines with the DE algorithm is modified. And the migration operators based on the number of iterations are improved to meet motif discovery requirements. Furthermore, the immigration and emigration rates based on a cosine curve are modified. It can therefore generate promising candidate solutions. Statistical comparisons with DEPT and MOGAMOD approaches on three commonly used datasets are provided, which demonstrate the validity and effectiveness of the MHABBO algorithm. Compared with some typical existing approaches, the MHABBO algorithm performs better in terms of the quality of the final solutions.

Keywords: multi-objective optimization; motif discovery; transcription factor binding site; hybrid adaptive biogeography-based optimization

1. Introduction

The motif discovery problem (MDP) in molecular biology is to find similar regions common to each sequence in a given set of DNA, RNA, or protein sequences [1]. It is an important problem for locating binding sites and finding conserved regions in unaligned sequences. From a computational point of view, finding motifs in many sequences is an NP-hard problem. Many methods have been applied to solve MDP and have achieved excellent results such as statistical methods, probabilistic methods etc. In recent years, with the development of evolutionary algorithms and their advantages, they have also been gradually applied to MDP.

Evolutionary computation (EC) is an optimization method based on the principles of biological evolution and is gaining more attention in recent years. EC has certain advantages for motif discovery [2]. Evolutionary algorithms (EA) carry out global search and have relatively low

sensitivity to initial conditions [3]. They are comparatively flexible in terms of how solutions are represented and evaluated, and do not require knowledge about the problem to which they are being applied. EC methods have been successfully applied to solve the motif discovery problem such as the genetic algorithm (GA) [4], bacterial foraging optimization algorithm integrating taboo search (TSBFO) [5], estimation of distribution algorithm with differential evolution (DE/EDA) [6], evolutionary multi-objective optimization (DEPT) [7], multi-objective artificial bee colony (MOABC) algorithm [8], Multi-objective genetic algorithm (MOGAMOD) [9], Non-dominated Sorting Genetic algorithm-III (NSGA-III) [10] and Multi-objective evolutionary algorithm based on decomposition (MOEA/D) [11], etc. The biogeography-based optimization (BBO) algorithm (Simon, 2008) [12] is a nature-inspired computational technique based on the mathematical models of biogeography. As a population-based stochastic algorithm, the BBO algorithm generates the next generation population by simulating the characteristics of the biological species migration. Because of information sharing in the migration process, the BBO algorithm has a better exploitation ability. The BBO algorithm is superior for solving the single-objective motif discovery problem [13–15], which has also been modified to solve multi-objective optimization problems (MOPs) [16–19]. However, in these papers, the BBO algorithm has still not been applied to solving the multi-objective motif discovery problem. In the literature [20], we have applied hybridization of adaptive biogeography-based optimization algorithm and differential evolution (DE) to multi-objective optimization problems (MOPs), and have achieved excellent performance on the convergence and the distribution of solutions.

The aim of this paper is based on our previous research, and is to apply the BBO algorithm to the multi-objective motif discovery problem. So far as we know, it is the first time that the multi-objective biogeography-based optimization has been applied to multi-objective MDP. In this paper, a new algorithm named MHABBO is presented and used to solve multi-objective MDP, and then presents a comparative study on twelve datasets with other different algorithms.

The motivation for proposing MHABBO for MDP in this research is threefold. First, based on the above literature review, there have been several successful applications based on multi-objective biogeography-based optimization (MBBO). Second, we have proposed a new MBBO algorithm and achieved excellent performance on multi-objective benchmark functions [20]. Finally, we will try to apply it to solve multi-objective MDP.

The key contributions of this paper are as follows. Firstly, we propose a new approach called MHABBO based on BBO algorithm to predict motifs. In the MHABBO algorithm, the migration in the BBO is implemented with the number of iterations to avoid the presence of a very stable local minimum. Secondly, motivated by the work described in References [21–24], the mutation is performed by integrating with DE to produce new feasible solutions. Simultaneously, the parameters of migration probability and mutation probability are adaptively changed. Furthermore, the immigration and emigration rates based on the cosine curve are modified. Finally, we apply MHABBO algorithm to the multi-objective motif discovery problem.

Compared with DEPT and MOGAMOD approaches on three commonly used datasets, the MHABBO algorithm performs better, or at least comparably, in terms of the quality of the final solutions. Statistical comparisons with some typical existing approaches demonstrate the validity and effectiveness of the MHABBO algorithm. Experimental results show that the obtained Pareto solutions can approximate to the Pareto optimal front and has good diversity and uniform distribution.

The paper is organized as follows. Section 2 describes the motif discovery problem. The MHABBO algorithm process for multi-objective motif discovery problem is proposed in Section 3. Section 4 shows the simulation and experiment results. Finally, a brief conclusion is illustrated in Section 5.

2. Motif Discovery Problem

In this paper, we also use the same objectives as those used in Reference [9] to find many long and strong motifs. The multi-objective motif discovery problem is converted into the following

three-objective optimization problem: Maximize similarity, Maximize motif length and Maximize support. These three objectives for MDP are defined as follows [7]:

(1) Support: Support indicates the level of the support of the candidate motifs to the consensus motif. The consensus motif is built by using the candidate motifs. The level of the support is measured by similarity rate of the candidate motif to the consensus motif. The similarity rate means the same number of the nucleosides between the candidate motif and the consensus motif. When the similarity rate is larger than 50%, the subsequent corresponding to candidate motif can be considered as a Support. For example, the consensus motif is assumed to be GACCTTTTGCAATCCTGG, the candidate motif of the sequence 1, i.e., GACCACTTGCAGTCTTAG, has 13 nucleotides identical to the consensus motif, and the consensus motif has 18 nucleotides, so its similarity rate is 13/18 = 72%.

(2) Motif Length: The motif length points to the number of the nucleotides of the consensus motif. In the example, the motif length is 18. According to real datasets used in this paper, the value of the motif length is limited to between 5 and 60.

(3) Similarity: the similarity objective function of motif is defined as the average of the dominance values of all position weight matrix columns. The similarity is calculated based on Equation (1). In which the *dv* in each column (dominant nucleotide) is the dominance value of the dominant nucleotide, it is calculated by Equation (2):

$$Similarity(Motif) = \frac{\sum_{i=1}^{l} dv(i)}{length} \tag{1}$$

$$dv(i) = \max_b\{f(b,i)\} \qquad i = 1,\ldots,l \tag{2}$$

where $f(b,i)$ is the score of nucleotide b in column i in the position weight matrix, and l is the motif length.

To better understand the similarity objective function, an example is used to illustrate it. Firstly, a position weight matrix (PWM) from the motif patterns found by MHABBO in every sequence is generated. Then, the percentage of occurrence of nucleotides at each motif position is calculated (see Table 1). The highest value of each matrix column is selected. The similarity is obtained by averaging all these dominance values. In this example, the similarity value is computed as 0.81 (81%) by using Equation (6):

(1 + 1 + 0.5 + 0.75 + 1 + 1 + 0.75 + 0.5 + 1 + 0.5 + 0.75 + 1 + 0.75 + 1 + 0.5 + 1 + 0.5 + 1)/18 = 0.81.

Table 1. Position Weight Matrix for a Motif.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	0	1	0.25	0.25	0	0	0	0.25	0	0.25	0.75	1	0	0	0.25	0	0.25	0
C	0	0	0.5	0.75	0	0	0	0.25	0	0.5	0	0	0.25	1	0.5	0	0.25	0
T	0	0	0.25	0	1	1	0.75	0.5	0	0.25	0.25	0	0.75	0	0.25	1	0	0
G	1	0	0	0	0	0	0.25	0	1	0	0	0	0	0	0	0	0.5	1

3. MHABBO Algorithm

In this section, we describe the MHABBO algorithm for the motif discovery problem in detail. In the MHABBO algorithm, the migration in the BBO is implemented with the number of iterations, the mutation is performed by integrating with DE to produce new feasible solutions. Simultaneously the parameters about migration probability and mutation probability are adaptively changed. Meanwhile, the immigration rate and emigration rate based on a cosine curve are modified.

First, we describe the representation of the individuals in our algorithms. Because each individual contains the necessary information used to form a possible motif, an individual is represented as the motif length and the starting location s_i of the potential motif on all the sequences. Representation of an individual is shown as Table 2. This representation is the same as that used in [9].

Table 2. The representation of an individual.

Motif Length	Seq. 0	Seq. 1	Seq. 2	. . .	Seq. n
length	S0	S1	S2	...	Sn

3.1. Migration Operator for the MDP

The sharing of features between solutions is represented as immigration and emigration between the islands. The immigration rate λ and the emigration rate μ of each solution are used to probabilistically share features between solutions. Motivated by the work in [25], these parameters are modified based on the cosine curve. The immigration rate and emigration rate of each individual are changed respectively by Equations (3) and (4). In which NP is the size of population.

$$\lambda_i = \frac{1}{2}(\cos(\frac{i}{NP} \times \pi - \frac{\pi}{2}) + 1). \tag{3}$$

$$\mu_i = \frac{1}{2}(-\cos(\frac{i}{NP} \times \pi - \frac{\pi}{2}) + 1). \tag{4}$$

Motivated by blended migration operator in [26], in our algorithm, the coefficient of solution $\mathbf{H}_i$ is related to the number of iterations. The modified migration operator is based on the following considerations. First, blend combination operators have been widely used in other optimization algorithms. Second, good solutions will be not degraded due to migration. Besides, poor solutions can still accept a lot of new features from good solutions. The migration operator is designed to accelerate the speed of convergence based on the number of iterations. Modified migration is defined as:

$$\mathbf{H}_i(j) = \frac{t}{t_{\max}}\mathbf{H}_i(j) + (1 - \frac{t}{t_{\max}})\mathbf{H}_k(j) \quad . \tag{5}$$

where $\mathbf{H}_i$ is immigrating island, $\mathbf{H}_k$ is emigrating island, $\mathbf{H}_i(j)$ is the jth dimension of the ith solution, and t is the number of iterations, $t_{\max}$ is the maximum number of iteration. Equation (5) means a new solution after migration is comprised of two components: the migration of feature from itself and another solution. It accelerates the convergence speed of the algorithm. And modified migration operator is described as follows (Algorithm 1):

Algorithm 1: Migration for the MDP (MigrationDo($\mathbf{H}$, P_{modif}))

Input: Initial population $\mathbf{H}$ and migration probability

Output: The population $\mathbf{H}$ that have been optimized by migration

---

For i = 1 to NP // NP is the size of population

If rand < P_{modif}

Use λ_i to probabilistically decide whether to immigrate to $\mathbf{H}_i$

If $rand(0,1) < \lambda_i$ then

For $k = 1\ to\ NP$

Select the emigrating island $\mathbf{H}_k$ with probability $\infty\mu_i$

If $rand(0,1) < \mu_k$ then

For j = 1 to N_d // N_d is the dimension size

$\mathbf{H}_i(j) = \frac{t}{t_{\max}}\mathbf{H}_i(j) + (1 - \frac{t}{t_{\max}})\mathbf{H}_k(j)$

End for

End if

End for

End if

End if

End for

3.2. Mutation Operator for the MDP

Although the hybridization of the BBO with DE has achieved many good results [27–29], they incorporate DE into the migration procedure for single-objective optimization problems. In MHABBO algorithm, DE is incorporated into the mutation procedure for multi-objective optimization problems. The algorithm helps to find the non-dominated solutions. A mutated individual ($\mathbf{H}_i(j)$) is generated according to Equation (6)

$$\mathbf{H}_i(j) = \mathbf{H}_i(j) + c_1 \times (\mathbf{H}_{best}(j) - \mathbf{H}_i(j)) + c_1 \times (\mathbf{H}_{r1}(j) - \mathbf{H}_{r2}(j)) \tag{6}$$

where $\mathbf{H}_i(j)$ is selected for mutation, c_1 is the mutation scaling factor, usually its value is set as range between 0.1 and 0.15. $\mathbf{H}_{r1}(j)$, $\mathbf{H}_{r2}(j)$ is the randomly selected two solutions, $\mathbf{H}_{best}(j)$ is the best solution in this generation. In MHABBO algorithm, this mutation scheme tends to increase the diversity among the population. Modified mutation operator is described as follows (Algorithm 2):

Algorithm 2: Mutation for the MDP (Mutation Do(H, P_{muta}))
Input: The population **H** optimized by migration, mutation probability
Output: The population **H** that have been optimized by mutation

---

    For i = 1 to NP // NP is the size of population
Select mutating habitat $\mathbf{H}_i$ with probability $\propto P_{muta}$
If $\mathbf{H}_i$ is selected, then
For j = 1 to N_d // N_d is the dimension size
$\mathbf{H}_i(j) = \mathbf{H}_i(j) + c_1 \times (\mathbf{H}_{best}(j) - \mathbf{H}_i(j)) + c_1 \times (\mathbf{H}_{r1}(j) - \mathbf{H}_{r2}(j))$
End for
End if
End for

3.3. Adaptive BBO for MDP

Modification probability factors and mutation probability factors in the BBO algorithm are denoted as P_{modif} and P_{muta} respectively. The two factors with ranges between 0 and 1 are set by users. The settings of the parameters are related to the experience of the user, and they may be unfavorable for the selection of migration individual and mutation individual. In order to choose better migration individual and mutation individual, these parameters are changed dynamically with the fitness function.

In the adaptive BBO algorithm, the parameters of modification probability and mutation probability are altered according to Equations (7) and (8). In Equations (7) and (8), constant factor k_1, k_2, k_3 and k_4 which range between 0 and 1 are set by users. Usually k_1 = 0.4, k_2 = 0.95, k_3 = 0.1, k_4 = 0.25.

$$P_{modif} = \begin{cases} k_1 \times \frac{MaxCost - FitnessCost}{MaxCost - AvgCost} & FitnessCost \geq AvgCost \\ k_2 & FitnessCost < AvgCost \end{cases} \tag{7}$$

$$P_{muta} = \begin{cases} k_3 \times \frac{MaxCost - FitnessCost}{MaxCost - AvgCost} & FitnessCost \geq AvgCost \\ k_4 & FitnessCost < AvgCost \end{cases} \tag{8}$$

3.4. The Redefinition of the Fitness Function

In this paper, we propose the multi-objective MHABBO algorithm for multi-objective motif discovery problem. Generally speaking, solving multi-objective optimization problems is through Pareto non-dominated sorting and crowding distance sorting of different solutions. The fitness function is determined based on the Pareto dominance relation in [30]. However, only considering the Pareto dominance relation is not enough, if the distribution of solutions is also included, the definition of the fitness function will be more reasonable. Originating from SPEA2 algorithm [31], which measures

the Pareto dominance relationship and density relationship between different solutions as fitness function. So the fitness function is redefined based on density information and Pareto dominance relationship among the habitats. That is, we employ non-dominated sorting approach to determine the non-dominated rank of individuals. Specifically, the Pareto dominance relationship refers to the number of non-dominated solutions that dominate an individual. The density of each individual is calculated by the k nearest neighbor method. For any individual $\mathbf{H}_i = (H_{i1}, H_{i2} \ldots, H_{in})$ in the habitat population $\mathbf{H} = \{\mathbf{H}_1, \mathbf{H}_2 \ldots, \mathbf{H}_{NP}\}$, its fitness function is defined by Equation (9) as follows.

$$
\begin{aligned}
F(\mathbf{H}_i) &= \sum_{\mathbf{H}_i \in \mathbf{H}, \mathbf{H}_j \prec \mathbf{H}_i} \frac{1}{NP} |\{k | \mathbf{H}_j, \mathbf{H}_k \in \mathbf{H} \wedge k \neq j \wedge k \neq i \wedge \mathbf{H}_j \prec \mathbf{H}_k\}| + D(i); \\
D(i) &= \frac{1}{\sigma_i^e + 2}, e = \left\lfloor \sqrt{NP + \overline{N}} \right\rfloor \\
i, j, k, e &\in \{1, 2, \ldots, NP\}
\end{aligned}
\tag{9}
$$

where $\mathbf{H}_i, \mathbf{H}_j, \mathbf{H}_k \in \mathbf{H}$ are habitats, σ_i^e is the distance between the habitat $\mathbf{H}_i$ and $\mathbf{H}_e$ in the objective space, the operator $|\cdot|$ is the cardinality of the set. e is integer value of the square root of the sum of population number NP and elitism number $\overline{N}$. According to Equation (14), the fitness function $F(\mathbf{H}_i)$, is the sum of $D(i)$ and the average of the sum of the number of dominated habitats in the total population. In which the number of dominated habitats means the total number of any other habitats whom every individual who dominates $\mathbf{H}_i$ can dominate in the population. Therefore, the lower the dominated degree of habitat $\mathbf{H}_i$ is, the smaller the fitness function of $\mathbf{H}_i$ is, when the fitness function of $\mathbf{H}_i$ is 0, it indicates that $\mathbf{H}_i$ is a non-dominated habitat.

3.5. Main Procedure of MHABBO for Multi-Objective Motif Discovery Problem

Firstly, the fitness function on the basis of density information and Pareto dominance relation is redefined, then the modified migration procedure and the mutation procedure are merged into the BBO. Furthermore, related parameters in the BBO such as modification probability and mutation probability, emigration rate and immigration rate, are altered. The procedure of the MHABBO is described in Algorithm 3.

Algorithm 3: The main pseudo-code of MHABBO algorithm for multi-objective MDP

Input: The Sequences S
Output: support, motif length, similarity and the non-dominated consensus motif instance and corresponding PWM.
1. Init(number of iterations, elitism parameter *keep*, migration probability P_{modif}, mutation probability P_{muta} etc.)
2. $\mathbf{P} \Leftarrow$ GenerateInitialRandomPopulation()
3.   EvaluateFitness($\mathbf{H}_i$) for each habitat $\mathbf{H}_i$ in $\mathbf{P}$ according to Equation (9).
4.   While the halting criterion is not satisfied do
5.   Elite $\Leftarrow \mathbf{P}$ (1:*keep*)
6.   Compute λ_i, μ_i for each habitat $\mathbf{H}_i$ according to Equations (3) and (4)
7. $\mathbf{P} \Leftarrow$ MigrationDo($\mathbf{P}, P_{modif}$)   Algorithm. (1)
8. $\mathbf{P} \Leftarrow$ MutationDo($\mathbf{P}, P_{muta}$)   Algorithm. (2)
9.   EvaluateFitness($\mathbf{H}_i$)
10.   SortPopulation($\mathbf{P}$)
11. $\mathbf{P} \Leftarrow$ ReplaceWorstbyElites ($\mathbf{P}$, Elite)
12. $\mathbf{P} \Leftarrow$ ClearDuplicates($\mathbf{P}$)
13.   [maximum cost, minimum cost, average cost]$\Leftarrow$EvaluateCostItems()
14.   [P_{modif},P_{muta}]$\Leftarrow$updateProbability()   Equations (7) and (8)
15.   End while

We generate the initial population for three different targets. That is, a solution is randomly generated when the length and support is different. Besides, we get the similarity between this

candidate motif and consensus motif. Each solution has these three different indicators, including the length, support and similarity. That is, each solution reflects multiple different objectives. The function GenerateInitialRandomPopulation() in Algorithm 3 is described as follows:

GenerateInitialRandomPopulation()

---

1. The length of candidate motif is randomly generated on the basis of the range of motif length
2. The beginning position of candidate motif in the sequence is randomly given according to the length of the sequence
3. Evaluate support

   3.1 Generate a candidate motif based on the beginning position and the length of motif for each sequence
   3.2 Generate PWM based on these candidate motifs
   3.3 Generate consensus motif according to PWM
   3.4 Compare candidate motif generated from each sequence with the consensus motif, if similarity rate is greater than 0.5, the number of support is added one.

4. Regenerate PWM according to support
5. Confirm the consensus motif $\mathbf{H}_i$ according to PWM
6. Repeat the above steps, generate a specified number of population

---

In the evaluation of the fitness function, the degrees of Pareto domination and distribution information between different solutions are reflected by fitness function. That is, the Pareto non-dominated sorting is equivalent to the ranking of the value of fitness function. The MHABBO measures the Pareto dominance relationship and density information between different solutions as fitness function. We employ non-dominated sorting approach to determine the non-dominated rank of individuals. The function EvaluateFitness($\mathbf{H}_i$) in Algorithm 3 is described as follows:

EvaluateFitness($\mathbf{H}_i$)

---

1. Confirm the Length, the Support of $\mathbf{H}_i$
2. Compute Similarity of $\mathbf{H}_i$ by Equation (7)
3. Count the number of the solution dominating $\mathbf{H}_i$ based on the three objective function (Length, Support and Similarity)
4. For each solution $\mathbf{H}_j$ dominating $\mathbf{H}_i$,
5. Count the number of the solution dominating $\mathbf{H}_j$
6. For each solution $\mathbf{H}_k$ dominating $\mathbf{H}_j$
7. Evaluate Fitness function for each habitat $\mathbf{H}_i$ according to Equation (14)

---

4. Simulation and Analysis

4.1. Simulation, Comparison and Discussion

4.1.1. Results Comparisons with Other Methods

In order to demonstrate the feasibility of the MHABBO algorithm for the MDP, MHABBO algorithm is compared with MOGAMOD and DEPT. Some experiments are carried out on a number of real sequence datasets which are selected from the TRANSFAC database [32]. Motif instances from different sequences of each dataset have already been tagged, so these datasets are used as a benchmark for the discovery of TBFSs [33]. The properties of datasets are shown in Table 3. Every real dataset corresponds to living beings in nature. More concretely, three of the datasets are from the fly (those beginning by dm), three from the human being (hm), three from the mouse (mus), and three from yeast (yst). Meanwhile, datasets with a different number of sequences and different sizes (nucleotides per

sequence) are selected to ensure that our algorithm works with several types of instances. For example, the yst04r sequence dataset contains 7 sequences of 1000 bps each. Motif instances from the yst04r sequence have 7 instances ranges from 5 to 25. The hm03r sequence dataset contains 10 sequences of 1500 bps each. Motif instances from the hm03r sequence have 15 instances ranges from 14 to 46. Using datasets from different species, the new algorithm can obtain the meaning motifs in all types of biological data. The times are also given in Table 3. The algorithm has been implemented by using the MATLAB R2014b programming language. All experiments were performed using windows10 OS, 64 bit processer, Inter(R) Core(TM) i5-6200U CPU (2.30 GHz) with 12 GB of RAM.

Table 3. The properties of the benchmark datasets.

Dataset	#Sequence	Length	#Instance	#Width of Motifs	Time (s)
Dm01g	4	1500	7	13–28	50
Dm04g	4	2000	9	10–26	51
Dm05g	5	2500	14	6–21	58
Hm03r	10	1500	15	14–46	42
Hm04m	13	2000	11	7–44	37
Hm16g	7	3000	7	9–54	38
Mus02r	9	1000	12	10–33	38
Mus07g	4	1500	4	15–33	53
Mus11m	12	500	15	6–27	42
Yst03m	8	500	18	6–24	44
Yst04r	7	1000	7	5–25	39
Yst08r	11	1000	14	12–49	39

Parameters used by the MHABBO algorithm are shown in Table 4. MHABBO algorithm has been run 5 times for each dataset with different random seeds. The top 20 results obtained by a non-dominated sort in the 5 runs are recorded. Due to the limit of the length of article, we only list the parameters used for yst04r dataset: DE mutation scheme is DE/rand/1/bin, the population size is 100, the maximal generation number is 100, number of variables in each individual is 8, the value of motif length is between 5 to 25, habitat modification probability is 0.75, mutation probability is 0.05, elitism parameter is 10, and scaling factor is 0.01, k_1 factor is 0.4; k_2 factor is 0.95; k_3 factor is 0.05; k_4 factor is 0.1. Other test problems have similar parameters to yst04r dataset. These parameters that are different from the yst04r dataset have a number of variables for each individual mutation probability and modification probability etc. We assume a motif instance is correctly discovered if the predicted binding site is within 3 bps away from the true binding site.

Table 4. Parameters used by MHABBO algorithm.

MHABBO	DEPT	MOGAMOD
Population Size: 100	Population Size: 200	Population Size: 200
Migration Probability: 0.75	Crossover Probability: 0.25	Crossover: SPX with probability 0.6
Mutation Probability: 0.05	Mutation Factor: 0.03	Mutation Factor: 0.5
Maxgen: 100	Selection Scheme:	Parents choose: Binary Tournament
Elitism parameter: 10	Rand/1/Binomial	New Generation Selection: Elitist
Scaling factor c_1: 0.01		
k_1: 0.4, k_2: 0.95, k_3: 0.05; k_4: 0.1		

The performance of MHABBO for the MDP has been compared with other different methods such as MOGAMOD and DEPT. Comparisons of the motif predicted by MHABBO and corresponding Support, Similarity and Motif Length with other methods for yst04r are shown in Table 5. Comparisons of three objectives and the motif predicted by MHABBO with other methods for yst08r are shown in Table 6. Comparisons of three objectives and the motif predicted by MHABBO with different methods

for hm03r are shown in Table 7. The motifs predicted with "*" in the Table 6 indicate that the motif predicted is consistent with the known motif instance.

Table 5. Comparisons of the predicted motif with different methods for yst04r.

Method	Support	Length	Similarity	Predicted Motif
AlignACE [34]	N/A	10	N/A	C ATTCCA
MEME [35]	N/A	11	N/A	C ATTCCCC
Weeder [36]	N/A	10	N/A	TTTTCT CA
MOGAMOD [9]	5	14	0.84	C A CTTCCACTAA
	6	14	0.77	C ATTCCTCTAT
DEPT	5	22	0.854	TAAATCTTTTACTTTTTTTTCT
	6	19	0.842	CTAATTCATTCTTTTTCAA
	7	15	0.847	TTTCT CAAACACA
MHABBO	6	5	0.85	AAATC*
	2	19	0.82	GAGCAAGAAGCCAATGAAA
	2	10	0.8	TAACCAAGAA*
	3	5	0.93	TTTCT

Table 6. Comparisons of the predicted motif with different methods for yst08r.

Method	Support	Length	Similarity	Predicted Motif
AlignACE	N/A	11	N/A	CACCCA ACAC
	N/A	12	N/A	T ATT CACT A
MEME	N/A	11	N/A	CACCCA ACAC
Weeder	N/A	10	N/A	ACACCCA AC
MOGAMOD	7	15	0.84	C ACT T CCT
	8	14	0.83	CCA AAAAA C
	8	13	0.85	ACACCCA ACATC
DEPT	7	20	0.84	TCAATTTTTTTTTTCTATTC
	8	19	0.83	TTATTTTTTTCTCTTTC
	8	15	0.85	CCATATTTCTTCTA
MHABBO	2	40	0.74	CACTACAATTGCTTTGAGTGGTGTATTCTCAGTCGCCAAG
	3	16	0.75	GGTGTATGTCCTAATA*
	3	34	0.68	AACCAGACAAAC*AAAAGAAAAAAAAAAATTAAAAG
	2	31	0.81	AGAACAAAAAAAAAAAAAAAAAAAAAAAAAAA

Table 7. Comparisons of the predicted motif with different methods for hm03r.

Method	Support	Length	Similarity	Predicted Motif
AlignACE	N/A	13	N/A	T T ATAAAAAA
MEME	N/A	20	N/A	A T TA ATAAAA AAAAAC
Weeder	N/A	10	N/A	T ATCACT
MOGAMOD	7	22	0.74	TATCATCCCT CCTA ACACAA
	7	18	0.82	T ACTCT TCCCTA TCT
	10	11	0.74	TTTTTTCACCA
	10	10	0.79	CCCA CTTA
	10	9	0.81	A T TCC
DEPT	7	22	0.78	A CTTA T CCT ACACA A A
	9	12	0.83	A TCTCA T CC
	10	9	0.85	T A ACTCA
MHABBO	2	29	0.85	ATCATAGGACCTCCCTTGCTTCCCAATGG
	2	25	0.76	CCTTTTATTGTTCTATT*
	2	13	0.85	AATTAGGAGACAA*
	3	36	0.68	AACAACAAAAGATAAAAAGTCAAATGAATGAACTCA

From these tables above, MHABBO achieves better results than MOGAMOD, while MHABBO achieves solutions similar to solutions obtained by DEPT and several motif instances predicted by MHABBO are very similar with the known motif instances. So we conclude that the MHABBO algorithm can predict meaningful motifs, therefore it is a promising method for multi-objective motif discovery. As the length of the predicted motifs becomes longer, the similarity does not obviously decrease. From Tables 5 and 6, we observe that there are some motifs only one of algorithms can predict. The reason is that different search strategies explore different search spaces. Hence, MHABBO is chosen for motif discovery. Additionally, the MHABBO algorithm can not only predict some motifs acquired by other known methods but also find novel motifs. However, the accuracy of the predicted motifs is not high enough. The reason is that the performance of MHABBO is influenced by randomly selecting an SIV during the process of migration and mutation between the islands. Another reason is that the definition of fitness function just considers some factors, so it may lead them away from accuracy.

4.1.2. The Consensus Motifs Obtained by MHABBO Algorithm

Sequence logos are a graphical representation of an amino acid or nucleic acid multiple sequence alignment developed by Tom Schneiderand Mike Stephens [37]. A sequence logo provides a richer and more precise description of, for example, a binding site, than a consensus sequence. WebLogo is a web based application designed to make the generation of sequence logos as easy and painless as possible, so the consensus motifs predicted by our algorithm on different datasets are expressed by WebLogo in Table 8.

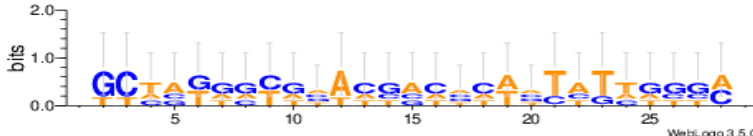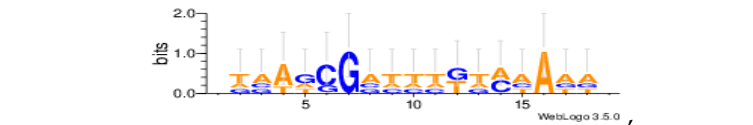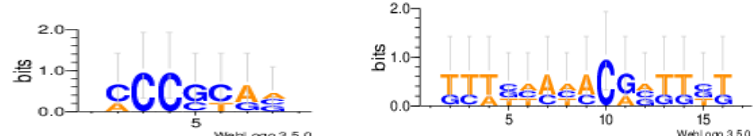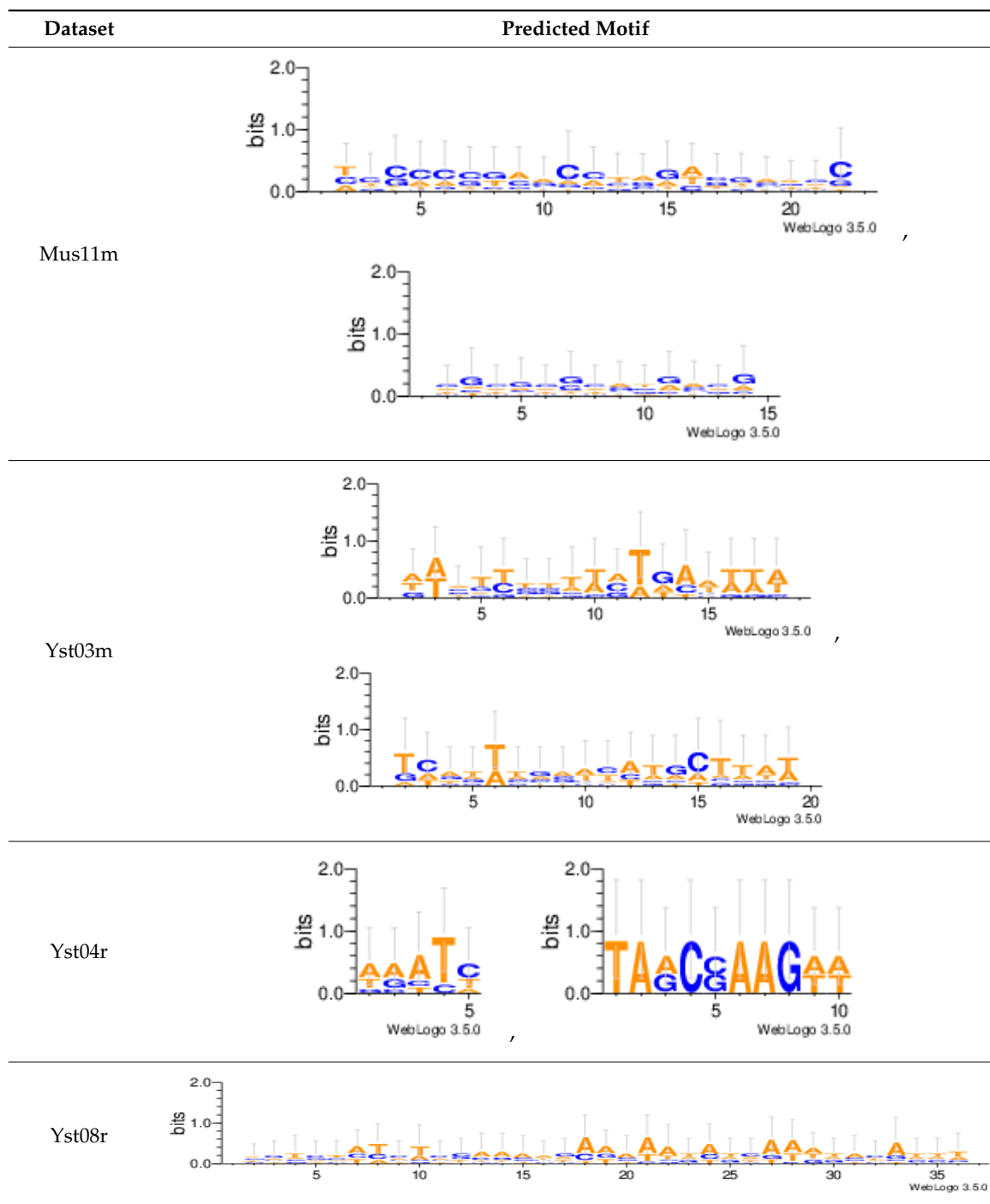**Table 8.** The Consensus motif predicted by MHABBO.

Dataset	Predicted Motif
Dm01g	
Dm04g	
Dm05g	

Table 8. *Cont*.

Dataset	Predicted Motif
Hm03r	
Hm04m	
Hm16g	
Mus02r	
Mus07g	

Table 8. *Cont.*

Dataset	Predicted Motif
Mus11m	
Yst03m	
Yst04r	
Yst08r	

4.1.3. Representation of the Pareto Fronts Obtained by MHABBO Algorithm

In order to have a visual perspective on the results, we show the graphs corresponding to the solutions obtained by MHABBO for each dataset (see Figure 1). The graphs show the Pareto front points (blue points) that are obtained by running the algorithm configured with the optimal parameters. The motif length is represented in the X-axis, the similarity in the Y-axis, and the support in the Z-axis. Furthermore, we show the projection of each point at the planes XY (purple points), XZ (red points) and ZY (yellow points).
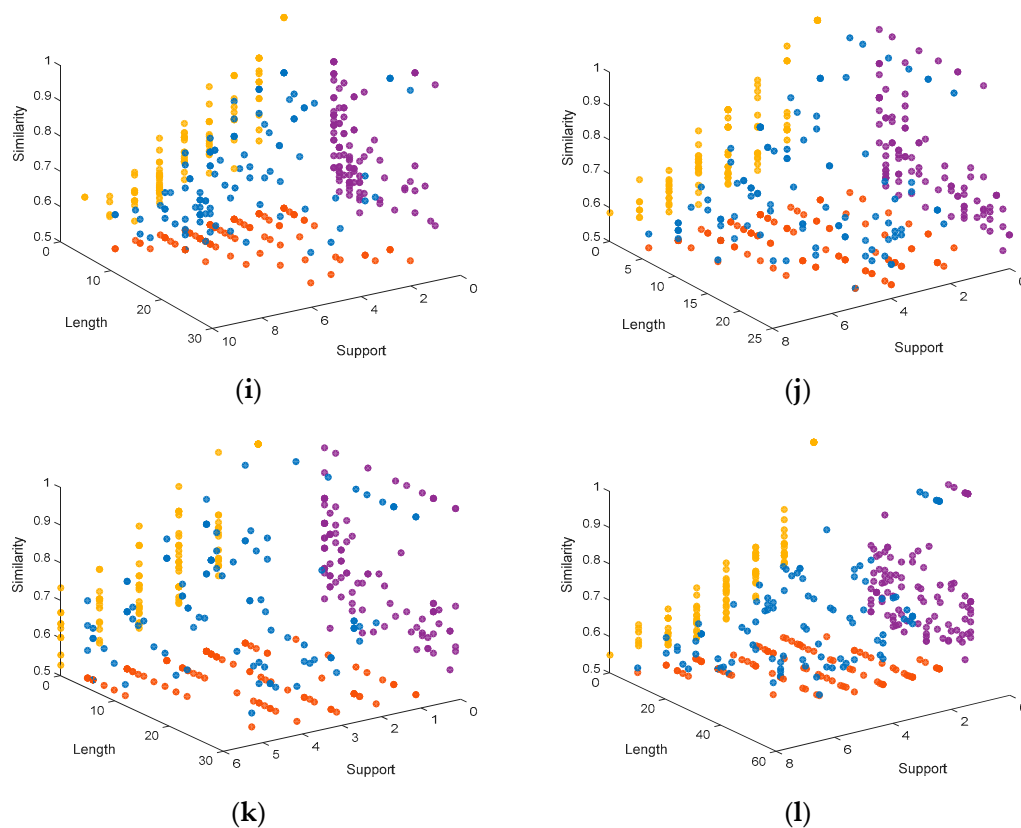
(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

(**f**)

(**g**)

(**h**)
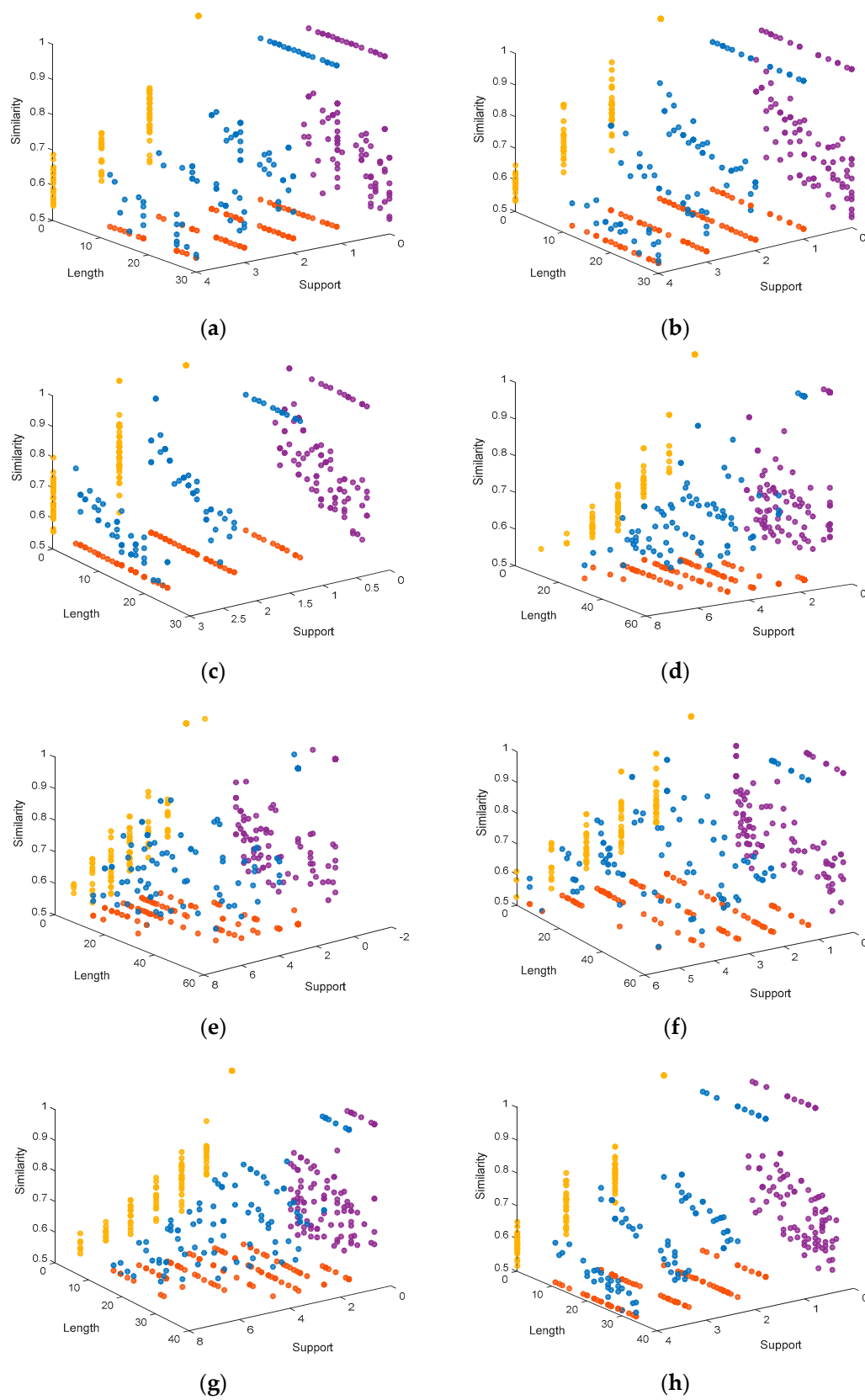
Figure 1. *Cont.*

Figure 1. Representation of the Pareto fronts obtained by MHABBO: (**a**) the distribution of solutions on the DM01g dataset; (**b**) the distribution of solutions on DM04g; (**c**) the distribution of solutions on DM05g; (**d**) the distribution of solutions on HM03r; (**e**) the distribution of solutions on HM04r; (**f**) the distribution of solutions on HM16g; (**g**) the distribution of solutions on MUS02r; (**h**) the distribution of solutions on MUS07g; (**i**) the distribution of solutions on MUS11m; (**j**) the distribution of solutions on YST03m; (**k**) the distribution of solutions on YST04r; (**l**) the distribution of solutions on YST08r.

The Pareto fronts obtained by MHABBO are shown in Figure 1, which shows that MHABBO on these datasets achieves better distribution of solutions. For example, there are 7 motif instances with a length range between 9 and 54 in the hm16r dataset. The length value of most parts of these motif instances is about 20. It can be seen from Figure 1 that the length of most obtained solutions is about 20. So the distribution of the obtained solutions is consistent with the distribution of the standard solution. As the length of the predicted motifs becomes longer, the similarity does not obviously decrease and as the support of the predicted motifs becomes larger, the similarity does not obviously decrease. The results demonstrate that the proposed MHABBO algorithm is competitive on the quantity and the distribution of final solutions. The results also present the distribution of the solutions and the convergence to Pareto-optimal front. It indicates that our approach performed well on multi-objective MDP.

4.2. Metrics to Assess Performance

Performance metrics play an important role in returning a scalar quantity, which reflects the quality of solutions. For each tool T and each data set D, we now have the set of known binding sites and the set of predicted binding sites. The correctness of T on D can be assessed both at the nucleotide level and at the motif level. There are many metrics that can be used to measure the quality of MDP [30], for example, nucleotide-level sensitivity (nSn), nucleotide-level positive predictive value (nPPV), the nucleotide-level correlation coefficient (nCC), nucleotide-level performance coefficient

(*n*PC), the motif-level correlation coefficient (*m*CC) and the motif-level F-score[35] etc. The following metrics are used in this paper: the *n*PC and F-score.

4.2.1. The Nucleotide-Level Performance Coefficient (*n*PC)

To measure the prediction accuracy of methods with respect to motif location, we have used the nucleotide-level performance coefficient (*n*PC). It was also adopted by Tompa et al. to evaluate binding site predictions in their single motif discovery benchmark study. The *n*PC is defined as follows:

$$n\text{PC} = \frac{n\text{TP}}{n\text{TP} + n\text{FN} + n\text{FP}} \tag{10}$$

Here, *n*TP is the number of nucleotide positions in both known sites and predicted sites, while *n*FN is the number of nucleotide positions in known sites but not in predicted sites, *n*FP is the number of nucleotide positions not in known sites but in predicted sites.

In order to furtherly measure the efficiency of this algorithm, the *n*PC value obtained by the MHABBO algorithm on different test functions are compared with other 14 algorithms. The results obtained by the fifteen different algorithms used to the different test problems are given in Figure 2.
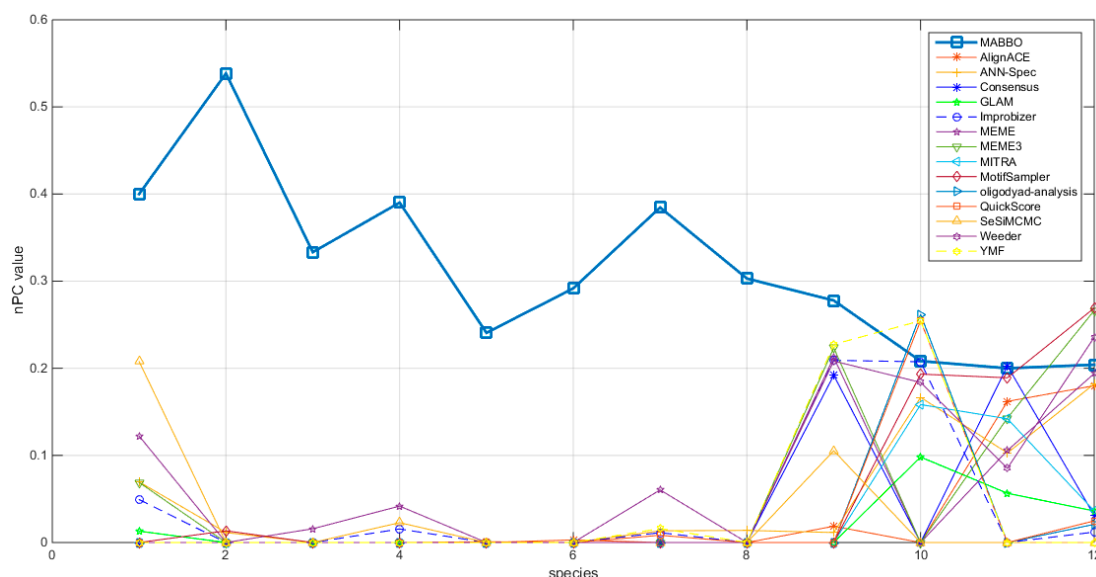

Figure 2. Comparisons of nPC value obtained by MHABBO with other algorithms on different test problems (Dm01g = 1, Dm04g = 2, Dm05g = 3, Hm03r = 4, Hm04m = 5, Hm16g = 6, Mus02r = 7, Mus07g = 8, Mus11m = 9, Yst03m = 10, Yst04r = 11 and Yst08r = 12).

It can be seen from Figure 2 that the *n*PCs obtained by the MHABBO algorithm on datasets from 1 to 9 are significantly better than the other fourteen algorithms. Except that the *n*PCs obtained by the MHABBO algorithm on datasets (Yst) are worse than several algorithms. This algorithm shows better performance on higher organisms than simpler organisms. It can be concluded that the performance of the algorithm is not obviously decreased with the increase of the dimension of the problem.

4.2.2. F-Score

To assess the performance of our algorithm at the motif-level, Precision, Recall and F-score are adopted on the basis of Equation (10) [38], where the operator $|\cdot|$ is the cardinality of the set. The

candidate motif instances obtained by MHABBO need to be verified by biological experiments. We hope to have a high Precision and a high Recall. The F-score is a tradeoff between Precision and Recall.

$$Precision = \frac{|correct\ motif|}{|motif\ found|}, \quad Recall = \frac{|correct\ motif|}{|true\ motif\ |},$$
$$F-score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{11}$$

Average results (precisions (P), recalls (R) and F-scores (F)) obtained by MHABBO on the twelve datasets is shown in Table 9. The comparisons of MHABBO with other methods [33] on the three datasets are given in Table 10.

Table 9. Average results (precisions (P), recalls (R) and F-scores (F)) of MHABBO on the twelve datasets.

Algorithms		Dm 01g	Dm 04g	Dm 05g	Hm 03r	Hm 04m	Hm 16g	Mus 02r	Mus 07g	Mus 11m	Yst 03m	Yst 04r	Yst 08r
MHABBO	P	3/100	2/20	4/20	6/20	8/20	5/10	8/20	2/10	8/20	8/20	4/10	5/20
	R	3/7	2/9	4/14	6/15	8/10	5/7	8/12	2/4	8/15	8/18	4/7	5/14
	F	0.06	0.14	0.24	0.34	0.53	0.59	0.5	0.29	0.46	0.42	0.47	0.29

Table 10. Comparisons of MHABBO with other methods on the three datasets: average results (precisions (P), recalls (R) and F-scores (F)).

Algorithms for MDP		Dataset			Algorithms for MDP		Dataset		
		Hm03	Mu02	Yst08			Hm03	Mu02	Yst08
YMF [39]	P	0/25	1/12	0/11	AlignACE [34]	P	0/14	0/0	9/41
	R	0/15	1/12	0/14		R	0/15	0/12	9/14
	F	0	0.08	0		F	0	0	0.33
SeSiMCMC [40]	P	1/10	0/9	0/21	MEME [35]	P	1/12	2/14	6/11
	R	1/15	0/12	0/14		R	1/15	2/12	6/14
	F	0.08	0	0		F	0.074	0.154	0.48
QuickScore [41]	P	0/22	1/22	3/56	MOTIFSAMPLE [42]	P	0/21	1/18	7/9
	R	0/15	1/12	3/14		R	0/15	1/12	7/14
	F	0	0.06	0.08		F	0	0.07	0.61
MITRA [43]	P	0/10	0/9	1/12	ANN-SPEC [44]	P	0/13	1/32	7/26
	R	0/15	0/12	1/14		R	0/15	1/12	7/14
	F	0	0	0.08		F	0	0.05	0.35
Improbizer [45]	P	1/20	0/18	1/22	MEME3 [35]	P	0/7	0/0	9/17
	R	1/15	0/12	1/14		R	0/15	0/12	9/14
	F	0.06	0	0.06		F	0	0	0.58
MHABBO	P	6/20	8/20	5/20	ABBO/DE/GEN [15]	P	5/30	5/30	8/30
	R	6/15	8/12	5/14		R	5/15	5/12	8/14
	F	0.34	0.5	0.29		F	0.22	0.24	0.36

Table 10 shows the average results of these algorithms in 5 runs. According to the F-score, MHABBO on hm03r and mus02r dataset is the best algorithm of all twelve algorithms, and it is better than ABBO/DE/GEN for single-objective motif discovery problems on hm03r and mus02r dataset. However, it is worse on yst08r than MEME, MEME3, ABBO/DE/GEN and MOTIFSAMPLE. This algorithm shows better performance on higher organisms than simpler organisms. The experiments demonstrate the validity of the proposed MHABBO algorithm for multi-objective motif discovery problems.

Assessing performance of the MHABBO algorithm at the nucleotide level and at the motif level, similar results have been obtained. That is to say, the more dimensions of the problem there are, the performance of the MHABBO does not worsen. This algorithm can obtain a more significant motif. It also shows that the algorithm on the convergence has better performance on higher organisms than simpler organisms.

5. Conclusions and Future Research

Since multi-objective, biogeography-based optimization has not been applied to the multi-objective motif discovery problem, we propose a hybrid multi-objective optimization algorithm named MHABBO to solve three-objective motif discovery problem on the basis of our previous research work. Compared with the existing methods, the proposed algorithm has the following advantages. Firstly, the redefinition of fitness evaluation based on MOEA can simplify the multi-objective optimization problem and use the Pareto dominance relationship to preserve population diversity. Secondly, modifying migration operations can speed up the convergence of the algorithm, and the mutation is performed by integrating with DE to produce new feasible solutions. In such a way, population diversity can be maintained. Finally, the robustness of the algorithm is enhanced by adaptively changed parameters related to the BBO algorithm.

Statistical comparisons with some typical existing approaches on several commonly used datasets are provided. The main work has been done in this paper as follows. Firstly, the motif instances obtained by the MHABBO algorithm on three commonly used datasets are compared with five other algorithms. Secondly, according to the PWMs corresponding to the obtained motif instances on twelve commonly used datasets, the logos of the motif instances are acquired using the online WebLogo software. Thirdly, the Pareto fronts of obtained motif instances on twelve commonly used datasets are drawn according to three-objective of the motif discovery problem. Finally, based on the NPC and F-score methods, the new algorithm is compared with other classical algorithms.

The experiments have indicated that the MHABBO algorithm outperforms other algorithms on the hm03r and mus02r datasets. From the Pareto fronts obtained by MHABBO, the results demonstrate that the proposed MHABBO algorithm is competitive on the convergence to Pareto-optimal front and the distribution of final solutions. It also shows that the algorithm on the convergence performs better on higher organisms than simpler organisms. It demonstrates the validity and effectiveness of the proposed MHABBO algorithm used to predict motifs from DNA sequences.

In this paper, we mainly discuss the multi-objective motif discovery problem. In the future, we will continue to improve the multi-objective BBO algorithm. We will try to combine NSGA-III or MOEA/D with the BBO algorithm for motif discovery problem. Additionally, in our earlier work, we discussed the portfolio optimization problem in second-order stochastic dominance constraint based on the BBO algorithm [46], and we will try to apply the multi-objective BBO algorithm to the multi-objective portfolio optimization problem [47,48] in the future.

Author Contributions: Siling Feng contributed the new processing method, and conceived and designed the experiments; Siling Feng and Ziqiang Yang performed the experiments; Siling Feng and Ziqiang Yang analyzed the data; Mengxing Huang contributed analysis tools; and Siling Feng wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patrik, D. What are DNA sequence motifs. *Nat. Biotechnol.* **2006**, *24*, 423–425.
2. Lones, M.A.; Yo, Y.; Tyrrell, A.M. The Evolutionary Computation Approach to Motif Discovery in Biological Sequences. In Proceedings of the 7th Annual Workshop on Genetic and Evolutionary Computation (GECCO'052005), Washington, DC, USA, 25–29 June 2005.
3. Lou, Y.; Li, J.; Jin, L.; Li, G. A CoEvolutionary Algorithm Based on Elitism and Gravitational Evolution Strategies. *J. Comput. Inf. Syst.* **2012**, *7*, 2741–2750.
4. Che, D.; Song, Y.; Rashedd, K. MDGA: Motif discovery using a genetic algorithm. In Proceedings of the 2005 Conference on Genetic and Evolutionary Computation (GECCO 2005), Washington, DC, USA, 25–29 June 2005; pp. 447–452.

5. Shao, L.; Chen, Y. Bacterial Foraging Optimization Algorithm Integrating Tabu Search for Motif Discovery. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2009), Washington, DC, USA, 1–4 November 2009; pp. 415–418.

6. Shao, L.; Chen, Y.; Abraham, A. Motif Discovery using Evolutionary Algorithms. In Proceedings of the International Conference of Soft Computing and Pattern Recognition (SOCPAR 2009), Malacca, Malaysia, 4–7 December 2009; pp. 420–425.

7. Gonzalez-Álvarez, D.L.; Vega-Rodriguez, M.A.; Gomez-Pulido, J.A.; Sanchez-Pérez, J.M. Predicting DNA Motifs by Using Evolutionary Multiobjective Optimization. *IEEE Trans. Syst. Man Cybern.* **2012**, *42*, 913–925. [CrossRef]

8. González-Álvarez, D.L.; Vega-Rodríguez, M.A.; Pulido, J.A.G.; Sánchez-Pérez, J.M. Finding Motifs in DNA Sequences Applying a Multiobjective Artificial Bee Colony (MOABC) Algorithm. In Proceedings of the 9th European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (EvoBIO), Torino, Italy, 27–29 April 2011; pp. 89–100.

9. Kaya, M. MOGAMOD: Multi-objective genetic algorithm for motif discovery. *Int. J. Expert Syst. Appl.* **2009**, *36*, 1039–1047. [CrossRef]

10. Deb, K.; Jain, H. An evolutionary many-objective optimization algorithm using reference-point-based non-dominated sorting approach. Part I: Solving problems with box constraints. *IEEE Trans. Evol. Comput.* **2014**, *18*, 577–601. [CrossRef]

11. Zhang, Q.; Li, H. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. Evol. Comput.* **2014**, *11*, 712–731. [CrossRef]

12. Simon, D. Biogeography-based optimization. *IEEE Trans. Evol. Comput.* **2008**, *12*, 702–771. [CrossRef]

13. Feng, S.L.; Zhu, Q.X.; Gong, X.J.; Zhong, S. Biogeography-Based Optimization for Motif Discovery Problem. *J. Comput. Inf. Syst.* **2013**, *9*, 6999–7010.

14. Feng, S.L.; Zhu, Q.X.; Gong, X.J.; Zhong, S. Hybridizing Biogeography-Based Optimization with Differential Evolution for Motif Discovery Problem. *ICIC Express Lett.* **2013**, *7*, 3343–3348.

15. Feng, S.L.; Zhu, Q.X.; Zhong, S.; Gong, X.J. Hybridizing Adaptive Biogeography-Based Optimization with Differential Evolution for Motif Discovery Problem. *Sens. Transducers* **2014**, *162*, 233–237.

16. Chutima, P.; Wong, N. A Pareto biogeography-based optimisation for multi objective two-sided assembly line sequencing problems with a learning effect. *Comput. Ind. Eng.* **2014**, *69*, 89–104. [CrossRef]

17. E Silva, M.D.A.C.; Coelho, L.D.S.; Lebensztajn, L. Multi objective biogeography-based optimization based on predator-prey approach. *IEEE Trans. Magn.* **2012**, *48*, 951–954. [CrossRef]

18. Ma, H.; Su, S.; Simon, D.; Fei, M. Ensemble multi-objective biogeography-based optimization with application to automated warehouse scheduling. *Eng. Appl. Artif. Intell.* **2015**, *44*, 79–90. [CrossRef]

19. Goudos, S.K.; Plets, D.; Liu, N.; Martens, L.; Joseph, W. A multi-objective approach to indoor wireless heterogeneous networks planning based on biogeography-based optimization. *Comput. Netw.* **2015**, *91*, 564–576. [CrossRef]

20. Feng, S.; Yang, Z.; Huang, M. Hybridizing Adaptive Biogeography-Based Optimization with Differential Evolution for Multi-Objective Optimization Problems. *Information* **2017**, *8*, 83. [CrossRef]

21. Jadon, S.S.; Tiwari, R.; Sharma, H.; Bansal, J.C. Hybrid Artificial Bee Colony algorithm with Differential Evolution. *Appl. Soft Comput.* **2017**, *58*, 11–24. [CrossRef]

22. Loris, V.; Marco, S. Coupling response surface and differential evolution for parameter identification problems. *Comput.-Aided Civil Infrastruct. Eng.* **2015**, *30*, 376–393. [CrossRef]

23. Trivedi, A.; Srinivasan, D.; Biswas, S.; Reindl, T. Hybridizing genetic algorithm with differential evolution for solving the unit commitment scheduling problem. *Swarm Evol. Comput.* **2015**, *23*, 50–64. [CrossRef]

24. Wang, X.; Xu, Z. Multi-objective optimization algorithm based on biogeography with chaos. *Int. J. Hybrid Inf. Technol.* **2014**, *7*, 225–234. [CrossRef]

25. Ma, H.; Simon, D. Analysis of migration models of biogeography-based optimization using Markov theory. *Eng. Appl. Artif. Intell.* **2011**, *24*, 1052–1060. [CrossRef]

26. Ma, H.; Simon, D. Blended biogeography-based optimization for constrained optimization. *Eng. Appl. Artif. Intell.* **2011**, *24*, 517–525. [CrossRef]

27. Boussaïd, I.; Chatterjee, A.; Siarry, P.; Ahmed-Nacer, M. Two-stage update biogeography-based optimization using differential evolution algorithm (DBBO). *Comput. Oper. Res.* **2011**, *38*, 1188–1198. [CrossRef]

28. Cai, Z. A Novel Hybrid Biogeography-based with Differential Mutation. In Proceedings of the International Conference on Electronic & Mechanical Engineering and Information Technology, Harbin, China, 12–14 August 2011; pp. 2710–2714.

29. Gong, W.; Cai, Z.; Ling, C.X. DE/BBO: A hybrid differential evolution with biogeography-based optimization for global numerical optimization. *Soft Comput.* **2010**, *15*, 645–665. [CrossRef]

30. Bi, X.; Wang, J.; Li, B. Multi-objective optimization based on hybrid biogeography-based optimization. *Syst. Eng. Electron.* **2014**, *36*, 179–186.

31. Zitzler, E.; Laumanns, M.; Thiele, L. SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization. In Proceedings of the Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems, Athens, Greece, 19–21 September 2001; pp. 95–100.

32. Wingender, E.; Dietze, P.; Karas, H.; Knüppel, R. TRANSFAC: A database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* **1996**, *24*, 238–241. [CrossRef] [PubMed]

33. Tompa, M.; Li, N.; Bailey, T.L.; Church, G.M.; De Moor, B.; Eskin, E.; Favorov, A.V.; Frith, M.C.; Fu, Y.; Kent, W.J.; et al. Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.* **2005**, *23*, 137–144. [CrossRef] [PubMed]

34. Hughes, J.D.; Estep, P.W.; Tavazoie, S.; Church, G.M. Computational identification of cis-regulatory elements associated with functionally coherent groups of genes in Saccharomyces cerevisiae. *J. Mol. Biol.* **2000**, *296*, 1205–1214. [CrossRef] [PubMed]

35. Bailey, T.L.; Elkan, C. The value of prior knowledge in discovering motifs with MEME. In Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology, Cambridge, UK, 16–19 July 1995; AAAI Press: Menlo Park, CA, USA, 1995; pp. 21–29.

36. Pavesi, G.; Mereghetti, P.; Mauri, G.; Pesole, G. Weeder Web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.* **2004**, *32*, W199–W203. [CrossRef] [PubMed]

37. WebLogo 3. Available online: http://weblogo.threeplusone.com/create.cgi (accessed on 19 September 2017).

38. Li, G.; Chan, T.M.; Leung, K.S.; Lee, K.H. A Cluster Refinement Algorithm for Motif Discovery. *IEEE Trans. Comput. Biol. Bioinform.* **2010**, *7*, 1–15. [CrossRef] [PubMed]

39. Sinha, S.; Tompa, M. YMF: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Res.* **2003**, *31*, 3586–3588. [CrossRef] [PubMed]

40. Favorov, A.V.; Gelfand, M.S.; Gerasimova, A.V.; Mironov, A.A.; Makeev, V.J. Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length and its validation on the ArcA binding sites. In Proceedings of the Fourth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS 2004), Novosibirsk, Russia, 25–30 July 2004.

41. Régnier, M.; Denise, A. Rare events and conditional events on random strings. Discrete Math. *Theor. Comput. Sci.* **2004**, *6*, 191–214.

42. Thijs, G.; Lescot, M.; Marchal, K.; Rombauts, S.; De Moor, B.; Rouze, P.; Moreau, Y. A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. *Bioinformatics* **2001**, *17*, 1113–1122. [CrossRef] [PubMed]

43. Eskin, E.; Pevzner, P. Finding composite regulatory patterns in DNA sequences. *Bioinformatics* **2002**, *18* (Suppl. 1), S354–S363. [CrossRef] [PubMed]

44. Workman, C.T.; Stormo, G.D. ANN-Spec: A method for discovering transcription factor binding sites with improved specificity. In *Pacific Symposium on Biocomputing*; Altman, R., Dunker, A.K., Hunter, L., Klein, T.E., Eds.; Stanford University: Stanford, CA, USA, 2000; pp. 467–478.

45. Ao, W.; Gaudet, J.; Kent, W.J.; Muttumu, S.; Mango, S.E. Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. *Science* **2004**, *305*, 1743–1746. [CrossRef] [PubMed]

46. Ye, T.; Yang, Z.; Feng, S. Biogeography-Based Optimization of the Portfolio Optimization Problem with Second Order Stochastic Dominance Constraints. *Algorithms* **2017**, *10*, 100. [CrossRef]

47.　Sawik, B. Survey of multi-objective portfolio optimization by linear and mixed integer programming. In *Applications of Management Science*; Lawrence, K.D., Kleinman, G., Eds.; Emerald Group Publishing Limited: Bingley, UK, 2013; Volume 16, pp. 55–79.

48.　Sawik, B. A Review of Multi-Criteria Portfolio Optimization by Mathematical Programming. In *Recent Advances in Computational Finance*; Dash, G.H., Thomaidis, N., Eds.; Nova Science Publishers: New York, NY, USA, 2013; pp. 149–172.