

Seed optimization for i.i.d. similarities is no easier than optimal Golomb ruler design

Bin Ma^{a,*}, Hongyi Yao^b

^a David Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

^b Institute for Theoretical Computer Science, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Article history:

Received 24 October 2008
Received in revised form 2 July 2009
Accepted 15 July 2009
Available online 17 July 2009
Communicated by F.Y.L. Chin

Keywords:

Computational complexity
Spaced seeds
Golomb ruler

ABSTRACT

The spaced seed is a filtration method to efficiently identify the regions of interest in string similarity searches. It is important to find the optimal spaced seed that achieves the highest search sensitivity. For some simple distributions of the similarities, the seed optimization problem was proved to be *not* NP-hard. On the other hand, no polynomial time algorithm has been found despite the extensive researches in the literature. In this article we examine the hardness of the seed optimization problem by a polynomial time reduction from the optimal Golomb ruler design problem, which is a well-known difficult (but not NP-hard) problem in combinatorial design.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction and notations

1.1. Seed optimization

Similarity searches often utilize some types of filtrations to efficiently identify the similarity candidates before more accurate but time-consuming examination. Since some real targets can be incorrectly filtered out, the filtration trades the search sensitivity for speed. In DNA similarity searches, spaced seed was invented to achieve a better tradeoff [16].

A spaced seed x is represented by a binary string such as $111*1**1*1**11*111$. The positions with letter 1 are required matches, and the positions with letter * are “don't cares”. The length of the string is called the *length* of the seed, denoted by $l(x)$. The number of required matches is called the *weight* of the seed, denoted by $w(x)$. A similarity is *hit* by a seed x if there is a length- $l(x)$ segment of the similarity such that all the required matches specified

```
GAGTACTCAACACCAACATTAGTGGGCA
|| | | | | | | | | | | | | | | |
GAATACTCAACAGCAACACTAATGGGCA
111*1**1*1**11*111
```

Fig. 1. The seed $111*1**1*1**11*111$ hits the similarity region.

by x are satisfied by the segment. Fig. 1 shows an example. Given a probabilistic model of the similarities, the *sensitivity* of a seed is the probability that a randomly sampled similarity contains at least one hit.

For spaced seeds with the same weight, it is easy to see that their expected number of hits in random regions are approximately equal. However, surprisingly, their sensitivities can be very different [16]. It is therefore important to find the spaced seed that achieves the highest sensitivity under a given distribution of the similarities. This is called the *seed optimization problem*.

In a similarity, we use 1 to denote the matches and 0 to denote the mismatches. Then the similarity can be represented by a 0–1 string. The i.i.d. model assumes that each position of a similarity is independently 1 with probability p . Here p is called the *similarity level*. The (L, k) model assumes that the similarity has length L and there are exactly k uniformly random positions that are 1. The seed

* Corresponding author.

E-mail addresses: binma@uwaterloo.ca (B. Ma),
thy03@mails.tsinghua.edu.cn (H. Yao).

```

111*1**1*1**11*111      111*1**1*1**11*111      ...
111*1**1*1**11*111      111*1**1*1**11*111      ...
    
```

Fig. 2. No matter how the seed is “slided”, two overlapping 111*1**1*1**11*111 always give six or more extra required matches than one seed.

optimization problems for i.i.d. model and $\langle L, k \rangle$ model are called the *i.i.d. seed optimization* and the $\langle L, k \rangle$ *seed optimization*, respectively.

A spaced seed x can also be specified by the set of positions of the required matches. For example, the seed $x = 111*1**1*1**11*111$ can be denoted by its *set representation* $S(x) = \{0, 1, 2, 4, 7, 9, 12, 13, 15, 16, 17\}$. For a given set S and an integer i , we define $S + i = \{x + i \mid x \in S\}$.

1.2. Related work

Using several discontinuous positions as a filter for similarity search was proposed in some earlier literature such as [20,12,2]. Ma et al. [16] first studied the optimization of the discontinuous positions and coined the term “spaced seed”. Given a spaced seed, the computing of the exact sensitivity was proved to be NP-hard for i.i.d. model in [14,15] and for $\langle L, k \rangle$ model in [19]. Exponential time algorithms have been developed [10,13,3,1] and a polynomial time approximation scheme (PTAS) was given in [15] for computing the sensitivity of a given seed. However, the NP-hardness of computing seed sensitivity does not lead to the NP-hardness of seed optimization.

For an arbitrary distribution of the similarities, seed optimization problem was proved to be NP-hard [13]. But the proof does not work for the simple i.i.d. and $\langle L, k \rangle$ models. In fact, when unarily encoded, the seed optimization problem for i.i.d. model defines a sparse language [14,15] and is therefore not NP-hard [18]. The same proof also applies to the $\langle L, k \rangle$ similarities. An exponential-time brute-forth algorithm was used in [16] to select the optimal seed. Many heuristic algorithms were also developed [23,9,4,13,21,24, 1]. However, the exact computational complexities of the seed optimization problem for i.i.d and $\langle L, k \rangle$ similarities remain unknown.

1.3. Golomb ruler

A w -mark Golomb ruler is a set of distinct nonnegative integers $0 = a_1 < a_2 < \dots < a_w$, called “marks”, such that $|a_i - a_j| \neq |a_k - a_l|$ for $\{i, j\} \neq \{k, l\}$ and $i \neq j$. The optimal Golomb ruler design problem seeks for a w -mark ruler with the least a_w [5].

It is relatively easy to construct a w -mark Golomb ruler with polynomial a_w . However, the finding of the optimal Golomb ruler is a well-known difficult problem. The largest known optimal Golomb ruler to date has $w = 24$, which was found in 1967 [22] and verified to be optimal with four years of distributed computation at distributed.net (<http://www.distributed.net>) in 2004. However, there is no mathematical proof on the hardness of Golomb ruler design either. In contrast, when unarily coded, the optimal Golomb ruler problem defines a sparse language and is therefore not NP-hard [18].

In this paper we reduce the optimal Golomb ruler design problem to seed optimization, and consequently

prove that seed optimization is at least as hard as optimal Golomb ruler design. Our results, together with the tremendous efforts on optimal Golomb ruler design, justify the exponential-time and heuristic algorithms for seed optimization.

2. I.I.D. seed optimization

It has been believed that the sensitivity increase of the spaced seed comes from the irregularities in the seed shape. When a spaced seed hits a similarity region, an extra hit right after the first hit requires many additional matching positions, as illustrated in Fig. 2. This makes the concurrent existence of more than one hits in the same similarity region a rare event; whereas for a consecutive seed, the second hit is relatively easy – only one additional required match is needed. As a result, while the total number of hits are similar, spaced seeds hit more similarity regions than a consecutive seed. If the set representation $S(x)$ of a seed x is a Golomb ruler, then $S(x) \cap (S(x) + i)$ has at most one element for any integer i , resulting in the minimum level of overlap between a seed and its sliding. Therefore, a Golomb ruler is likely to be the optimal spaced seed. In Theorems 1 and 2 we prove that this is true under certain conditions.

Theorem 1. Consider the i.i.d. seed optimization problem for seed length l , weight w , similarity length L , and similarity level p . Let $n = L - l + 1$ be the number of positions the seed can hit the similarity. Suppose $p \leq \frac{1}{n^3}$ and $n \geq 2l$. Then there is a w -mark Golomb ruler with $a_w = l - 1$ if and only if each optimal spaced seed is a Golomb ruler.

Proof. Suppose a length- l and weight- w seed is given by its set representation. When the context is clear, we also use S to refer to the seed. Define $\varphi(i) = |S \cap (S + i)|$. Define $\phi = \max_i \varphi(i)$.

Denote by $h(i_1, \dots, i_k)$ the probability of that the seed hits at every position of i_1, i_2, \dots, i_k . This event is equivalent to that all the positions in $\bigcup_{j=1}^k (S + i_j)$ are matches. Therefore, it is easy to verify that for any distinct numbers $i, j, k \in [0, n]$, we have $h(i) = p^w$, $h(i, j) = p^{2w - \varphi(j-i)} \leq p^{2w - \phi}$, and

$$h(i, j, k) \leq p^{2w - \phi + 1}. \tag{1}$$

According to Bonferroni Inequalities [8],

$$\Pr(S \text{ hits}) \geq \sum_{i=0}^{n-1} h(i) - \sum_{0 \leq i < j < n} h(i, j), \tag{2}$$

and

$$\begin{aligned} \Pr(S \text{ hits}) &\leq \sum_{i=0}^{n-1} h(i) - \sum_{0 \leq i < j < n} h(i, j) \\ &\quad + \sum_{0 \leq i < j < k < n} h(i, j, k). \end{aligned} \tag{3}$$

Because of Eq. (1), when $p \leq \frac{1}{n^3}$,

$$\sum_{0 \leq i < j < k < n} h(i, j, k) \leq \binom{n}{3} p^{2w-\phi+1} < p^{2w-\phi} \times \frac{1}{2}. \quad (4)$$

If $\phi = 1$, Eq. (2) becomes

$$\Pr(S \text{ hits}) \geq np^w - p^{2w-1} \times \frac{n^2}{2}. \quad (5)$$

If $\phi \geq 2$, because there is at least one pair of i and j such that $\varphi(j-i) = \phi \geq 2$, as well as Eq. (4), Eq. (3) becomes

$$\begin{aligned} \Pr(S \text{ hits}) &\leq np^w - p^{2w-\phi} + p^{2w-\phi} \times \frac{1}{2} \\ &= np^w - \frac{1}{2} \times p^{2w-\phi} \\ &< np^w - p^{2w-1} \times \frac{n^2}{2}. \end{aligned} \quad (6)$$

When there is a Golomb ruler of length l with w markers, the seed defined by the ruler has $\phi = 1$ and the hit probability is lower bounded by Eq. (5). Because $\phi \geq 2$ implies Eq. (6), the optimal seed must be such that $\phi = 1$. It is easy to verify that when $n \geq 2l$, $\phi = 1$ implies that the seed is a Golomb ruler. \square

Corollary 1. *The i.i.d. seed optimization problem is at least as hard as optimal Golomb ruler design.*

Proof. Theorem 1 says that the finding of w -mark Golomb ruler with length l can be reduced to the seed optimization problem. Then the optimal Golomb ruler problem for a given weight w can be solved by trying different length l in polynomial steps. \square

One problem of Theorem 1 is that the upper bound of p is $O(n^{-3})$, which is very small and not practical. We relax this upper bound in Theorem 2.

Theorem 2. *Let l, w, L, p, n be as defined in Theorem 1. Suppose $p \leq \frac{1}{4l} \cdot \frac{n-1}{n}$ and $2l \leq n \leq (2\sqrt{l})^{w-1}$. Then there is a w -mark Golomb ruler with $a_w = l - 1$ if and only if each optimal spaced seed is a Golomb ruler.*

Proof. We only need to prove the “only if”. Suppose there is a w -mark Golomb ruler with $a_w = l - 1$, denoted as S^* . Denote the optimal spaced seed as S . We prove by contradiction that S is also a Golomb ruler.

Define $\varphi^*(i) = |S^* \cap (S^* + i)|$. Because S^* is a Golomb ruler, $\varphi^*(i) \leq 1$. Define $\varphi(i) = |S \cap (S + i)|$ and $\phi = \max_i \varphi(i)$. If S is not a Golomb ruler, then $\phi > 1$. Let $h^*(i_1, \dots, i_k) = \Pr(S^* \text{ hits at } i_1, \dots, i_k)$ and $h(i_1, \dots, i_k) = \Pr(S \text{ hits at } i_1, \dots, i_k)$.

Because both S and S^* have weight w , $h(i) = h^*(i) = p^w$. In addition, if $j - i \geq l$, then $h(i, j) = h^*(i, j) = p^{2w}$. Thus, by replacing S by S^* and h by h^* in Eq. (2) and then subtracting Eq. (3) from Eq. (2), we get the following:

$$\begin{aligned} &\Pr(S^* \text{ hits}) - \Pr(S \text{ hits}) \\ &\geq \sum_{0 \leq i < j < \min(i+l, n)} h(i, j) - \sum_{0 \leq i < j < \min(i+l, n)} h^*(i, j) \\ &\quad - \sum_{0 \leq i < j < k < n} h(i, j, k) \\ &\geq (n-l)p^{2w-\phi} - nlp^{2w-1} - \sum_{0 \leq i < j < k < n} h(i, j, k). \end{aligned} \quad (7)$$

Here the last inequality is due to the following two facts: (1) There is at least one d such that $\varphi(d) = \phi$. Therefore $h(i, i+d) = p^{2w-\phi}$ for at least $n-l$ different i . (2) $h^*(i, j) \leq p^{2w-1}$.

To prove the theorem, it suffices to show that when $\phi \geq 2$, formula (7) is greater than zero, which is a contradiction to the optimality of S . Clearly, when p is small, nlp^{2w-1} can be bounded by a fraction of $(n-l)p^{2w-\phi}$ in formula (7). We need to examine the third term $\sum_{0 \leq i < j < k < n} h(i, j, k)$ more carefully and bound it by a fraction of $(n-l)p^{2w-\phi}$.

The set of indexes $I = \{(i, j, k) : 0 \leq i < j < k < n\}$ can be divided into two sets $I_1 = \{(i, j, k) \in I : j < i+l \text{ and } k < j+l\}$ and $I_2 = I \setminus I_1$. Clearly I_1 corresponds to the situations where the seed at j overlaps both the seed at i and the seed at k ; and I_2 corresponds to the situations where at least one seed does not overlap the other two seeds. For any $(i, j, k) \in I_2$, $|(S+i) \cup (S+j) \cup (S+k)| \geq 3w - \phi$. Hence

$$\sum_{(i, j, k) \in I_2} h(i, j, k) \leq \frac{n^3}{6} \times p^{3w-\phi}. \quad (8)$$

When p is small, this can also be bounded by a fraction of $(n-l)p^{2w-\phi}$.

Again, I_1 can be divided into two sets

$$J_1 = \{(i, j, k) \in I_1 : |(S+k) \setminus ((S+i) \cup (S+j))| = 1\}$$

and $J_2 = I_1 \setminus J_1$. That is, providing that there are hits at i and j , J_1 contains the indexes where the seed at k requires only one additional match in the similarity region, and J_2 contains the indexes where the seed at k requires at least two additional matches. Therefore, for any $(i, j, k) \in J_2$, $|(S+i) \cup (S+j) \cup (S+k)| \geq 2w - \phi + 2$. Hence

$$\sum_{(i, j, k) \in J_2} h(i, j, k) \leq nl^2 \times p^{2w-\phi+2}. \quad (9)$$

When p is small, this can be bounded by a fraction of $(n-l)p^{2w-\phi}$ again. The rest of the proof is to bound $\sum_{(i, j, k) \in J_1} h(i, j, k)$. For $(i, j, k) \in J_1$, we consider the possibilities of k for fixed i and j . As shown in Fig. 3, because the last letter 1 in the seed at k already contributed an additional match, the second last letter 1 in the seed at k must coincide with the last letter 1 of either the seed at i or the seed at j . Otherwise it contributes another additional match (keep in mind that the seeds at i, j and k are the same seed), contradicting the definition of J_1 . Thus, once i is fixed, there are l choices of j , and then there are at most two choices of k . As a result, $|J_1| \leq 2nl$. Because $h(i, j, k) \leq p^{2w-\phi+1}$ for $(i, j, k) \in J_1$,

$$\sum_{(i, j, k) \in J_1} h(i, j, k) \leq p^{2w-\phi+1} \times 2nl. \quad (10)$$

$$\begin{array}{l}
 i: \quad \dots 1***1 \qquad \dots 1***1 \\
 j: \quad \dots 1***1 \qquad \dots 1***1 \\
 k: \quad \dots 1***1 \qquad \dots 1***1 \\
 \end{array}
 \tag{A} \qquad \tag{B}$$

Fig. 3. The two possible choices of k for fixed i and j in J_1 .

Combining Eqs. (7), (8), (9) and (10),

$$\begin{aligned}
 & \Pr(S^* \text{ hits}) - \Pr(S \text{ hits}) \\
 & \geq (n-l)p^{2w-\phi} - nlp^{2w-1} \\
 & \quad - \sum_{(i,j,k) \in I_2 \cup J_2 \cup J_1} h(i,j,k) \\
 & \geq (n-l)p^{2w-\phi} - nlp^{2w-1} - \frac{n^3}{6} \cdot p^{3w-\phi} \\
 & \quad - nl^2 \cdot p^{2w-\phi+2} - 2nl \cdot p^{2w-\phi+1} \\
 & = p^{2w-\phi+1} \left((n-l) \cdot p^{-1} - nl \cdot p^{\phi-2} \right. \\
 & \quad \left. - \frac{n^3}{6} \cdot p^{w-1} - nl^2 \cdot p - 2nl \right). \tag{11}
 \end{aligned}$$

When $\phi \geq 2$, $p \leq \frac{1}{4l} \cdot \frac{n-1}{n}$ and $n \leq (2\sqrt{l})^{w-1}$, the following inequalities hold: $(n-l) \cdot p^{-1} \geq 4nl$, $nl \cdot p^{\phi-2} \leq nl$, $n^2 \cdot p^{w-1} \leq (4lp)^{w-1} \leq (\frac{n-1}{n})^{w-1} \leq 1$, and $nl^2 \cdot p \leq l \cdot \frac{n-1}{4} \leq \frac{1}{4}nl$. Thus, it is easy to verify that formula (11) has a positive value. Hence the theorem is proved. \square

3. Discussion

Obviously, the main factor for the upper bound of p in Theorem 2 is $O(\frac{1}{l})$. With a more sophisticated analysis, it is also possible to bound p by $O(\frac{1}{w})$. The analysis is omitted here.

A natural question is to ask whether the upper bounds on p in Theorems 1 and 2 can be removed. The answer is no. The following is a counterexample. For $w = 5$, $n = 150$, $p = 0.999$, the optimal 5-mark Golomb ruler is $\{0, 2, 7, 10, 11\}$ [6]. The corresponding spaced seed has sensitivity $1 - 4.3376 \times 10^{-116}$. On the other hand, the spaced seed $\{0, 3, 4, 6, 11\}$, which is not a Golomb ruler, has a better sensitivity $1 - 3.3674 \times 10^{-117}$. One may argue that this may be due to that the region is not sufficiently long, and the boundary effects cause this to happen. Fig. 4 excludes this possibility. In the figure the no-hit probability curve of the non-Golomb ruler seed $\{0, 3, 4, 6, 11\}$ goes down faster than the curve of the Golomb ruler seed $\{0, 2, 7, 10, 11\}$, indicating the non-Golomb ruler seed is asymptotically better.

For the $\langle L, k \rangle$ seed optimization problem, similar results still hold. That is, for certain values of k , if there is a Golomb ruler, then each optimal spaced seed is a Golomb ruler. On the other hand, there is a counterexample showing the condition on k cannot be removed. These results can be found in the conference version of this paper [17].

There are also researches going on to hit all the $\langle L, k \rangle$ similarities with the maximum weighted seed [7] or with the minimum number of multiple spaced seeds [11]. It is

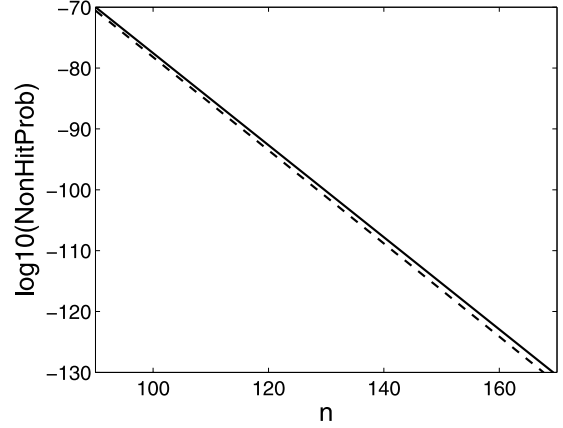


Fig. 4. The curves of $\log_{10}(\Pr(\text{There is no hit}))$ with respect to the region length when the similarity level $p = 0.999$. The upper curve is for the Golomb ruler seed $\{0, 2, 7, 10, 11\}$, and the lower curve is for the seed $\{0, 3, 4, 6, 11\}$.

noteworthy that despite the title of [7], the paper only gave algorithms to find the *almost* optimal spaced seed. Therefore the results in [7] do not contradict the results in this paper.

Acknowledgements

The work was supported in part by China High-tech R&D Program 2008AA02Z313, China NSF 60553001, National Basic Research Program of China 2007CB807900, 2007CB807901, NSERC and Canada Research Chair. B.M.'s work was partially done when he visited Prof. Andrew Yao at ITCS at Tsinghua University, and when he was an associate professor at the University of Western Ontario. B.M. thanks Dr. Ming Li and Dr. John Tromp for commenting on an earlier version of this manuscript. The authors thank the useful discussion with the attendees of a seminar course at ITCS at Tsinghua University. In particular, Yifei Zhang pointed out the Golomb ruler design problem during the course. We thank two anonymous reviewers' important suggestions on improving the manuscript.

References

- [1] J. Buhler, U. Keich, Y. Sun, Designing seeds for similarity search in genomic DNA, in: Proc. of the 7th International Conference on Computational Biology (RECOMB), 2003, pp. 67–75.
- [2] S. Burkhardt, J. Kärkkäinen, Better filtering with gapped q-grams, in: Proc. 12th Annual Symposium Combinatorial Pattern Matching (CPM'01), 2001, pp. 73–85.
- [3] K. Choi, L. Zhang, Sensitive analysis and efficient method for identifying optimal spaced seeds, J. Comput. System Sci. 68 (2004) 22–40.
- [4] K.P. Choi, F. Zeng, L. Zhang, Good spaced seeds for homology search, Bioinformatics 20 (2004) 1053–1059.
- [5] C.J. Colbourn, J.H. Dinitz (Eds.), CRC Handbook of Combinatorial Designs, CRC Press, Boca Raton, FL, 1996.
- [6] A. Dollas, W. Rankin, D. McCracken, A new algorithm for Golomb ruler derivation and proof of the 19 mark ruler, IEEE Trans. Inform. Theory 44 (1) (1998) 379–382.
- [7] M. Farach-Colton, G. Landau, S. Sahinalp, D. Tsur, Optimal spaced seeds for faster approximate string matching, J. Comput. System Sci. 73 (7) (2007) 1035–1044.

- [8] J. Galambos, I. Simonelli, *Bonferroni-Type Inequalities with Applications*, Springer-Verlag, New York, 1996.
- [9] L. Ilie, S. Ilie, Fast computation of good multiple spaced seeds, in: *Proc. of 7th Workshop on Algorithms in Bioinformatics, 2007*, pp. 346–358.
- [10] U. Keich, M. Li, B. Ma, J. Tromp, On spaced seeds for similarity search, *Discrete Appl. Math.* 3 (2004) 253–263.
- [11] G. Kucherov, L. Noé, M.A. Roytberg, Multiseed lossless filtration, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2 (1) (2005) 51–61.
- [12] O. Lehtinen, E. Sutinen, J. Tarhio, Experiments on block indexing, in: *Proc. of the 3rd South American Workshop on String Processing, 1996*, pp. 183–193.
- [13] M. Li, B. Ma, D. Kisman, J. Tromp, PatternHunter II: Highly sensitive and fast homology search, *J. Bioinform. Comput. Biol.* 2 (3) (2004) 417–440.
- [14] M. Li, B. Ma, L. Zhang, Superiority and complexity of the spaced seeds, in: *Proc. of the 17th ACM–SIAM Symposium on Discrete Algorithms (SODA), 2006*, pp. 22–26.
- [15] B. Ma, M. Li, On the complexity of spaced seeds, *J. Comput. System Sci.* 73 (7) (2007) 1024–1034.
- [16] B. Ma, J. Tromp, M. Li, PatternHunter: faster and more sensitive homology search, *Bioinformatics* 18 (3) (2002) 440–445.
- [17] B. Ma, H. Yao, Seed optimization is no easier than optimal Golomb ruler design, in: *Proc. The 6th Asia Pacific Bioinformatics Conference, 2008*, pp. 133–144.
- [18] S. Mahaney, Sparse complete sets for NP: Solution of a conjecture of Berman and Hartmanis, *J. Comput. System Sci.* 25 (1982) 130–143.
- [19] F. Nicolas, E. Rivals, Hardness of optimal spaced seed design, in: *Proc. 16th Annual Symposium Combinatorial Pattern Matching (CPM'05), 2005*, pp. 144–155.
- [20] P. Pevzner, M. Waterman, Multiple filtration and approximate pattern matching, *Algorithmica* 13 (1995) 135–154.
- [21] F. Preparata, L. Zhang, K. Choi, Quick, practical selection of effective seeds for homology search, *J. Comput. Biol.* 12 (2005) 137–152.
- [22] J. Robinson, A. Bernstein, A class of binary recurrent codes with limited error propagation, *IEEE Trans. Inform. Theory* 13 (1967) 106–113.
- [23] J. Xu, D. Brown, M. Li, B. Ma, Optimizing multiple spaced seeds for homology search, in: *Proc. of the 15th Symposium on Combinatorial Pattern Matching (CPM), in: LNCS, vol. 3109, 2004*, pp. 47–58.
- [24] I.H. Yang, S.H. Wang, H.H. Chen, P.H. Huang, K.M. Chao, Efficient methods for generating optimal single and multiple spaced seeds, in: *Proc. of IEEE 4th Symp. on Bioinformatics and Bioengineering, 2004*, pp. 411–418.