DEVELOPMENT OF A WEBBASED APPLICATION TO DETECT PALINDROMES IN DNA SEQUENCES

Fatma Eltayeb¹, Muna Elbahir², Sahar Mohamed³, Mohamed Ahmed and⁴ Nazar Zaki⁵

¹, ², ³, ⁴Faculty of Mathematical Science, University of Khartoum, Sudan
⁵College of Information Technology, UAE University, Al Ain 17555, UAE
nzaki@uaeu.ac.ae

ABSTRACT
Detecting palindromes in DNA sequence is a central problem in computational biology. Identifying palindromes could help scientists advance the understanding of genomic instability. DNA sequences containing long adjacent inverted repeats (palindromes) are inherently unstable and are associated with many types of chromosomal rearrangements. In this paper, we present a simple web-base tool to assist biologist detecting palindromes in DNA sequence.

1. Introduction
A palindrome is a sequence of letters/words which reads the same in forward as well as backward directions. For example, the English word ‘madam’ is a palindrome [1]. DNA palindromes are words from the nucleotide base alphabet $A = \{A, C, G, T\}$ that are symmetrical in the sense that they read exactly the same as their complementary sequences in the reverse direction [2]. There are two types of DNA palindromes, Palindromes that occur on opposite strands of the same section of DNA helix and Inverted Repeats in which two different segments of the double helix read the same but in opposite directions [3]. DNA palindromes are crucial for gene regulation, DNA replication and initiation of gene amplification.

For the last few decades, vast of algorithms have been developed to aid in solving the problem through dynamic programming and heuristics techniques, however, very little attention has been made on producing a convenient and user-friendly system to be used by an average biologist. Therefore, in this work, we present a simple tool to assist biologist detecting palindromes in DNA sequence using a simple web-base application.

2. Online palindromes detection
The application is held in a dynamic website (www.bio-sdn.com) to be accessible 24 hours. For the development of the web site we used JSP technology, which is an exciting new technology that provides powerful and efficient creation of dynamic content [4]. JSP is a presentation layer technology that allows static Web content to be mixed with Java code [4]. MySql was used for creating database to hold user palindrome detection results, to be able to retrieve the saved results.

The algorithm for detecting the palindrome was written in the Java programming language, which presents the dynamic content. This dynamic content is Web pages that are generated by an application on the server.

To detect palindromes, the user should enter or upload a DNA sequence of interest in plain text format as input to the application as shown in Figure 1. A subsequence size should also be specified.
There are several steps in the algorithm of detecting palindromes, the algorithm will start by validating the input, which means ignoring numbers, spaces and all characters except the basic nucleotide acids (A, T, C, G) as shown in figure 2.

The next step is to convert all A’s to T’s, T’s to A’s, G’s to C’s, and C’s to G’s. The deliverable of this step is the complemented sequence. The DNA sequence will then be compared with the reverse of the complemented sequence. This process is illustrated by the following example. Suppose the DNA sequence “ATGACCAGGT CAT” is a palindrome of length 6 “ATGACCAGGT CAT”, then the complement for this sequence will be “TACTGGACCAGTA” and the reverse of the complemented sequence will be “ATGACCTGGTCAT”. By comparing the two sequences “TACTGGACCAGTA” and “ATGACCTGGTCAT” we can detect the Palindrome as: 1→5 ATGACC with 8→13 GGTCAT. This we call symmetrical Palindromes. Another example is shown in Figure 3.

The process of detecting Mirror Images in a DNA sequence is actually the same as mentioned above, except for the complementary step. An example is shown in Figure 4.
3. CONCLUSION

DNA sequences containing palindromes are associated with a high degree of genomic instability. The instability associated with palindromic sequences also creates difficulties in their molecular analysis. The main objective of this work is to develop a tool to assist biologist to detect any palindrome exists in a given DNA sequence. In this tool, the input sequence is reversed and complemented and then subsequences of the inputted sequence are compared with the resultant sequence using linear search under specific conditions. The matches found are then recorded as a palindrome. The tool has been developed using Java and Java Server Pages (JSP). In the future, the performance of this tool will be evaluated.

4. References


DEVELOPMENT OF A WEBBASED APPLICATION TO DETECT PALINDROMES IN DNA SEQUENCES

Fatma Eltayeb¹, Muna Elbahir², Sahar Mohamed³, Mohamed Ahmed and⁴ Nazar Zaki⁵
¹,²,³,⁴ Faculty of Mathematical Science, University of Khartoum, Sudan
⁵ College of Information Technology, UAE University, Al Ain 17555, UAE
nzaki@uaeu.ac.ae
ABSTRACT
Detecting palindromes in DNA sequence is a central problem in computational biology. Identifying palindromes could help scientists advance the understanding of genomic instability. DNA sequences containing long adjacent inverted repeats (palindromes) are inherently unstable and are associated with many types of chromosomal rearrangements. In this paper, we present a simple web-base tool to assist biologist detecting palindromes in DNA sequence.

1. Introduction

A palindrome is a sequence of letters/words which reads the same in forward as well as backward directions. For example, the English word ‘madam’ is a palindrome [1]. DNA palindromes are words from the nucleotide base alphabet A = {A, C, G, T} that are symmetrical in the sense that they read exactly the same as their complementary sequences in the reverse direction [2]. There are two types of DNA palindromes, Palindromes that occur on opposite strands of the same section of DNA helix and Inverted Repeats in which two different segments of the double helix read the same but in opposite directions [3]. DNA palindromes are crucial for gene regulation, DNA replication and initiation of gene amplification.

For the last few decades, vast of algorithms have been developed to aid in solving the problem through dynamic programming and heuristics techniques, however, very little attention has been made on producing a convenient and user-friendly system to be used by an average biologist. Therefore, in this work, we present a simple tool to assist biologist detecting palindromes in DNA sequence using a simple web-base application.

2. Online palindromes detection

The application is held in a dynamic website (www.bio-sdn.com) to be accessible 24 hours. For the development of the web site we used JSP technology, which is an exciting new technology that provides powerful and efficient creation of dynamic content [4]. JSP is a presentation layer technology that allows static Web content to be mixed with Java code [4]. MySql was used for creating database to hold user palindrome detection results, to be able to retrieve the saved results.

The algorithm for detecting the palindrome was written in the Java programming language, which presents the dynamic content. This dynamic content is Web pages that are generated by an application on the server.

To detect palindromes, the user should enter or upload a DNA sequence of interest in plain text format as input to the application as shown in Figure 1. A subsequence size should also be specified.

Figure 1. Input the DNA sequence.

There are several steps in the algorithm of detecting palindromes, the algorithm will start by validating the input, which means ignoring numbers, spaces and all characters except the basic nucleotide acids (A, T, C, G) as shown in figure 2.
The next step is to convert all A’s to T’s, T’s to A’s, G’s to C’s, and C’s to G’s. The deliverable of this step is the complemented sequence. The DNA sequence will then be compared with the reverse of the complemented sequence. This process is illustrated by the following example. Suppose the DNA sequence “ATGACCAGGTCAT” is a palindrome of length 6 “ATGACCAGGTCAT”, then the complement for this sequence will be “TACTGGACCAGTA” and the reverse of the complemented sequence will be “ATGACCTGGTCAT”. By comparing the two sequences “TACTGGACCAGTA” and “ATGACCTGGTCAT” we can detect the Palindrome as: 1→5 ATGACC with 8→13 GGTCAT. This we call symmetrical Palindromes. Another example is shown in Figure 3.

The process of detecting Mirror Images in a DNA sequence is actually the same as mentioned above, except for the complementary step. An example is shown in Figure 4.

The final output shown in figure 5 is the list of all possible palindromes in the DNA sequence of interest.
3. CONCLUSION

DNA sequences containing palindromes are associated with a high degree of genomic instability. The instability associated with palindromic sequences also creates difficulties in their molecular analysis. The main objective of this work is to develop a tool to assist biologist to detect any palindrome exists in a given DNA sequence. In this tool, the input sequence is reversed and complemented and then subsequences of the inputted sequence are compared with the resultant sequence using linear search under specific conditions. The matches found are then recorded as a palindrome. The tool has been developed using Java and Java Server Pages (JSP). In the future, the performance of this tool will be evaluated.

4. References


DEVELOPMENT OF A WEBBASED APPLICATION TO DETECT PALINDROMES IN DNA SEQUENCES

Fatma Eltayeb¹, Muna Elbahir², Sahar Mohamed³, Mohamed Ahmed and⁴ Nazar Zaki⁵
¹, ², ³, ⁴Faculty of Mathematical Science, University of Khartoum, Sudan
⁵College of Information Technology, UAE University, Al Ain 17555, UAE
nzaki@uaeu.ac.ae

ABSTRACT

Detecting palindromes in DNA sequence is a central problem in computational biology. Identifying palindromes could help scientists advance the understanding of genomic instability. DNA sequences containing long adjacent inverted repeats (palindromes) are inherently unstable and are associated with many types of chromosomal rearrangements. In this paper, we present a simple web-base tool to assist biologist detecting palindromes in DNA sequence.

1. Introduction

A palindrome is a sequence of letters/words which reads the same in forward as well as backward directions. For example, the English word ‘madam’ is a palindrome [1]. DNA palindromes are words from the nucleotide base
alphabet A = \{A, C, G, T\} that are symmetrical in the sense that they read exactly the same as their complementary sequences in the reverse direction [2]. There are two types of DNA palindromes, Palindromes that occur on opposite strands of the same section of DNA helix and Inverted Repeats in which two different segments of the double helix read the same but in opposite directions [3]. DNA palindromes are crucial for gene regulation, DNA replication and initiation of gene amplification.

For the last few decades, vast of algorithms have been developed to aid in solving the problem through dynamic programming and heuristics techniques, however, very little attention has been made on producing a convenient and user-friendly system to be used by an average biologist. Therefore, in this work, we present a simple tool to assist biologist detecting palindromes in DNA sequence using a simple web-base application.

2. Online palindromes detection

The application is held in a dynamic website (www.bio-sdn.com) to be accessible 24 hours. For the development of the web site we used JSP technology, which is an exciting new technology that provides powerful and efficient creation of dynamic content [4]. JSP is a presentation layer technology that allows static Web content to be mixed with Java code [4]. MySql was used for creating database to hold user palindrome detection results, to be able to retrieve the saved results.

The algorithm for detecting the palindrome was written in the Java programming language, which presents the dynamic content. This dynamic content is Web pages that are generated by an application on the server.

To detect palindromes, the user should enter or upload a DNA sequence of interest in plain text format as input to the application as shown in Figure 1. A subsequence size should also be specified.

![Figure 1. Input the DNA sequence.](image1)

There are several steps in the algorithm of detecting palindromes, the algorithm will start by validating the input, which means ignoring numbers, spaces and all characters except the basic nucleotide acids (A, T, C, G) as shown in figure 2.

![Figure 2. Input the DNA sequence.](image2)

Figure 2. Input the DNA sequence.

The next step is to convert all A’s to T’s, T’s to A’s, G’s to C’s, and C’s to G’s. The deliverable of this step is the complemented sequence. The DNA sequence will then be compared with the reverse of the complemented sequence. This process is illustrated by the following example. Suppose the DNA sequence “ATGCCAGGTCAT” is a palindrome of length 6 “ATGCCAGGTCAT”, then the complement for this sequence will be “TACTGGACCAGTA” and the reverse of the complemented sequence will be “ATGACCTGGTCAT”. By comparing the two sequences “TACTGGACCAGTA” and “ATGACCTGGTCAT” we can detect the Palindrome as: 1 5 ATGACC with 8 13 GGTCAT. This we call symmetrical Palindromes. Another example is shown in Figure 3.

![Figure 3. Symmetrical Palindrome.](image3)

The process of detecting Mirror Images in a DNA sequence is actually the same as mentioned above, except for the complementary step. An example is shown in Figure 4.

![Figure 4. Mirror Image.](image4)

The final output shown in figure 5 is the list of all possible palindromes in the DNA sequence of interest.
3. CONCLUSION

DNA sequences containing palindromes are associated with a high degree of genomic instability. The instability associated with palindromic sequences also creates difficulties in their molecular analysis. The main objective of this work is to develop a tool to assist biologist to detect any palindrome exists in a given DNA sequence. In this tool, the input sequence is reversed and complemented and then subsequences of the inputted sequence are compared with the resultant sequence using linear search under specific conditions. The matches found are then recorded as a palindrome. The tool has been developed using Java and Java Server Pages (JSP). In the future, the performance of this tool will be evaluated.

4. References


