Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

In silico analysis of methylation of the selected genes using computer programs based on various analytical techniques

Magdalena Gryzińska [a], *Andrzej Jakubczak* [a], *Robert Stryjecki* [b,*], *Grażyna Jeżewska-Witkowska* [a]

[a] *Department of Biological Basis of Animal Production, Faculty of Biology and Animal Breeding, University of Life Sciences in Lublin, Poland*

[b] *Department of Zoology, Ecology and Wildlife Management, University of Life Sciences in Lublin, Poland*

ARTICLE INFO

Article history:
Received 5 March 2014
Received in revised form
4 September 2014
Accepted 29 September 2014
Available online 12 October 2014

Keywords:
Bioinformatics
DNA methylation
Epigenetics
In silico
MSP
BSP

ABSTRACT

Selected genes were analyzed *in silico* in three species: red fox (*Vulpes vulpes*), raccoon dog (*Nyctereutes procyonoides*), and dog (*Canis lupus familiaris*). This type of analysis exemplifies current and potential research on gene expression. Four nucleotide sequences, of the genes *IGF1*, *MYO15A*, *PAX3* and *MC1R*, were obtained from the NCBI online database. The analyses focused on the presence of CpG islands and two analytical techniques, BSP and MSP. The results from three computer programs, CpG Island Searcher®, BiSearch® and MethPrimer®, were discussed in detail. The applications were compared in terms of their functionality and usefulness.

1. Introduction

The need to process and quickly transmit the vast amounts of data used in genetics has led to the constant evolution of software, operating systems and equipment for their analysis. Currently data analysis can be performed in fractions of a second, which is crucial in areas such as medicine, genetics, *etc.* Specialized programs for analysing and ordering data have found a permanent place in hospitals, laboratories, and pharmaceutical companies. Most current genetic research is conducted in two stages – computer modeling and laboratory testing. *In silico* analysis reduces the time necessary to obtain results and substantially lowers the cost of research. This type

* *Corresponding author at*: Department of Zoology, Ecology and Wildlife Management, University of Life Sciences in Lublin, ul. Akademicka 13, 20-950 Lublin, Poland.
E-mail address: robstry@wp.pl (R. Stryjecki).

http://dx.doi.org/10.1016/j.bbe.2014.09.002

of analysis has led to the development of bioinformatics, whose tools have contributed to such breakthrough discoveries as the sequencing of the human genome.

There are few publications in the literature describing the use of bioinformatics in the field of epigenetics. Key papers dealing with this topic include a study by Jörg [1] which gives step-by-step descriptions of many analyses of epigenome modification and one by Li and Dahiya on primer design [2]. The use of bioinformatics tools in epigenetics is continually growing. Familiarity with the advantages, disadvantages and limitations of programs for analysing and ordering genetic data is of vital importance for further work. Some of them are available for installation, but the use of online software is increasing. These programs enable easier and faster access to a wide variety of databases. The available software enables simulation of epigenetic testing. There are many techniques that make it possible to examine the course and significance of methylation. The available programs use different algorithms, which leads to discrepancies in the results. However, work with vast amounts of data without advanced biotechnological tools would be impossible [3].

There are many programs designed for epigenetic research. The selection is wide and the choice depends on the profile of the research. The main features that should be taken into account in selecting software include the type of input file or data record, computational algorithms based on a given analytical technique, the possibility of setting parameters for the simulation, and the effectiveness of the program, shown in a clear, easy-to-read output file. The programs presented in this paper are available free of charge and enable analysis of DNA methylation based on different analytical techniques.

Computer programs used to design MSP and BSP primers are based on a search for CpG islands, which are located at the ends of the 5′ promoter regions of genes, which are essential for cell function. Lack of methylation in promoter regions can be a precondition for controlled active gene transcription. Methylation of CpG islands may lead to a lack of or impaired synthesis of these gene products [4]. Current criteria for CpG islands are based on arbitrary threshold parameters that have little biological justification and do not take into account the diversity of CpG islands. An epigenome prediction pipeline links the DNA sequence of CpG islands, DNA methylation, histone modifications and chromatin accessibility, which are correlated with open chromatin structure [5].

1.1. *A review of epigenetic programs*

The generally accepted search algorithm for CpG islands used in all of the programs was proposed in 1987 by Gardiner-Garden and Frommer [6]. In this algorithm a CpG island is defined as a 200-bp stretch of DNA with 50% C + G content and a value for observed CpG/expected CpG of over 0.6. Each of the programs allows individual settings for DNA length in base pairs (defining a CpG island), percentage content of C + G and the value for observed CpG/expected CpG.

CpG Island Searcher® is a simple program based on the Java™ programming language. It has several features, the most important of which include setting the desired percentage of GC dinucleotides and the desired the size of CpG islands. The main advantage of the program is that there is no restriction on the amount of input sequence data [7]. The algorithm CpG Island Searcher® used stringent criteria (usually referred to as Takai-Jones algorithm [8]): length ≥ 500 bp, GC content ≥ 55%, and ObsCpG/ExpCpG ≥ 0.65 [9].

BiSearch® is an online program written in Java™. Its most important functions are designing primers for classic PCR, BSP and MSP; determining primer melting temperature; and determining the possibility of amplification of undesired products based on entire genomes. Many simulation parameters can be edited and the program is frequently updated, which is an indication of its substantive value and of the continual development of its script. BiSearch software is composed of two basic algorithms (can be used stepwise or separately) – a primer design algorithm and a search with the selected primers through genomic sequences to find potential non-specific PCR products [10].

MethPrimer® is an online program written in Java™. Based on Primer3®, it designs primers for MSP and BSP. Certain modifications of the program make it possible to obtain primers for MS-SnuPE and COBRA as well. Primer designs and results are returned *via* a web browser in text and graphic view [11]. The algorithm MethPrimer features CpG island prediction with a sliding window and uses the detected CpG islands as a starting point for the primer project [2].

The aim of the study was to design and compare *in silico* MS-PCR (methylation-specific PCR) primers for four genes, *IGF1*, *MYO15A*, *PAX3* and *MC1R*, using selected computer programs, and then evaluate the usefulness of these programs. The research is justified by the fact that there have been few studies on this subject.

1.2. *A review of analytical techniques*

Current methods allow for the analysis of genome-wide methylation [12,13] and methylation of individual genes (mainly MSP and BSP).

Methylation-specific PCR (MSP) is the most common technique used to examine the methylation of cytosine in DNA. It differs from standard PCR in that two reactions are carried out in parallel, *i.e.* with primers specific for methylated and for unmethylated sequences. MSP is always preceded by a reaction with sodium bisulfate ($NaHSO^{\bullet}_3$), by which unmethylated cytosine is converted to uracil but methylated cytosine remains unchanged. It becomes possible to distinguish methylated and unmethylated DNA sequences using appropriately selected primers [14].

Sodium bisulfite PCR (BSP) is also a technique determining DNA cytosone methylation. In the first step there is a reaction with sodium IV bisulfate ($NaHSO_3$), as in MSP, followed by PCR using primers that are specific for modified DNA but do not contain any CpG sites within their sequence. The resulting PCR product is subjected to sequencing and the result obtained is compared with the unmodified DNA product [15].

2. Materials and methods

The material for the study consisted of files containing nucleotide sequences in FASTA format from three species – red fox *Vulpes vulpes* (Linnaeus, 1758), raccoon dog *Nyctereutes*