Düzce University
Journal of Science & Technology

Research Article

A Bioinformatics-Based Approach for Designing Primer Sets in Determination of Meat Specificity

Nursel SÖYLEMEZ MİLLİ[a,*], İsmail Hakkı PARLAK[b], Ercan Selçuk ÜNLÜ [c], Mehmet MİLLİ [b], Ömer EREN [d]

[a]*Scientific, Industrial and Technological Application and Research Center (STARC),Bolu Abant Izzet Baysal University, Bolu, TURKEY*
[b] *Department of Computer Engineering, Bolu Abant Izzet Baysal University, Bolu, TURKEY*
[c]Department of Chemistry, *Bolu Abant Izzet Baysal University, Bolu, TURKEY*
[d] *Department of Food Engineering, Bolu Abant Izzet Baysal University, Bolu, TURKEY*
* *Corresponding author's e-mail address: nurselsoylemez@ibu.edu.tr*
DOI:10.29130/dubited.898519

ABSTRACT

Polymerase chain reaction (PCR) and its derivatives are one of the most widely used DNA-based methods in species determination studies in meat and meat products. Chromosomal or mitochondrial genes of the species can be targeted in PCR-based analyzes used in species detection studies. Many researchers are able to realize oligonucleotide differences between species through online alignment programs on mitochondrial DNA. Using chromosomal DNA would provide more concise results in quantification studies. However, determining the marker regions for genomic DNA is challenging due to the large size of the chromosomes. Bioinformatics approaches are available for selected applications. However, using those approaches requires intensive knowledge of computer science, molecular biology, and bioinformatics in addition to high computational power. In this study, a pipeline is presented that will provide a user-friendly approach to be adopted by facilities where contamination analyzes are routinely performed.

Keywords: *Sequence alignment, Bioinformatics, Biocomputing, Food quality*

Et Özgüllüğünün Belirlenmesinde Primer Setlerinin Tasarımına Yönelik Biyoinformatik Tabanlı Bir Yaklaşım

ÖZ

Polimeraz zincir reaksiyonu (PCR) ve türevleri, et ve et ürünlerinde tür belirleme çalışmalarında en yaygın kullanılan DNA bazlı yöntemlerden biridir. Tür tespit çalışmalarında kullanılan PCR tabanlı analizlerde türlerin kromozomal veya mitokondriyal genleri hedeflenebilir. Birçok araştırmacı, mitokondriyal DNA üzerindeki çevrimiçi hizalama programları aracılığıyla türler arasındaki oligonükleotid farklılıklarını gerçekleştirebilmektedir. Kromozomal DNA kullanmak, kantifikasyon çalışmalarında daha kısa sonuçlar sağlayacaktır. Bununla birlikte, genomik DNA için işaretleyici bölgelerin belirlenmesi, kromozomların büyüklüğünden dolayı zordur. Biyoinformatik yaklaşımlar, seçilmiş uygulamalar için mevcuttur. Ancak, bu yaklaşımları kullanmak, yüksek hesaplama gücüne ek olarak yoğun bilgisayar bilimi, moleküler biyoloji ve biyoinformatik bilgisi gerektirir. Bu çalışmada, kontaminasyon analizlerinin rutin olarak yapıldığı tesisler tarafından benimsenmesi için kullanıcı dostu bir yaklaşım sağlayacak bir kod akışı sunulmuştur.

Anahtar Kelimeler:*Sekans hizalama, Biyoinformatik, Biyo-hesaplama, Gıda kalitesi*

Received: 22/03/2021, Revised: 24/03/2021, Accepted: 26/04/2021

I. INTRODUCTION

Determining meat specificity is a serious problem, and verification of meat products is very important in the food industry [1]. Authentication of meat and meat products is essential for protecting public health, economic investment, and religious sanctity [2-4]. The integrity of food products is protected by national and international regulations that state that all ingredients must be labeled and all raw materials must be traceable [5] The basis of these regulations is the approaches applied to determining the source of meat and its limits. In general, proteomics and genomics-based approaches are among the most preferred approaches. However, electrophoretic [6], spectroscopic [7], chromatographic [8], [9], immunological [10], biosensors [11-13] and electronic nose [14,15] chemometric approaches such as are also being studied.

Especially, the PCR method, which is one of the genomic-based approaches, is more sensitive compared to other methods due to the stability of DNA in hard conditions. Therefore, DNA identification techniques have enormous potential for forensics, diagnostics, and food analysis[1]. Various DNA-based techniques have been proposed by researchers: sequence-specific PCR[16-18], qPCR [19-21], PCR-RFLP [22,23], PCR-RAPD [24], ddPCR[25-27], DNA Barcoding [28,29].

Chromosomal or mitochondrial genes can be targeted for PCR-based applications in species detection [30]. Using mitochondrial DNA for analysis provides a low limit of detection but cannot be used for quantitation. Due to the varying number of mitochondrial DNA (mtDNA) between cells, single-copy chromosomal DNA was generally preferred as the target gene to ensure the reproducibility of quantitative PCR measurements [26,31]. It has been suggested that the use of mtDNA cannot be recommended in full quantification studies since it varies at least 5 times between different tissues (fat vs muscle) compared to chromosomal DNA [26].

Using chromosomal DNA would provide more concise results in quantification studies. However, determining the marker regions for genomic DNA is challenging due to the large size of the chromosomes. Bioinformatics approaches are available for selected applications. However, using those approaches requires intensive knowledge of computer science, molecular biology, and bioinformatics in addition to high computational power. In this study, a pipeline is presented that will provide a user-friendly approach to be adopted by facilities where contamination analyzes are routinely performed.

II. MATERIALS AND METHODS

To obtain the results, the alignment software offered by LAST and the proposed program for the appropriate primer design in chromosomal DNA alignment files for the mentioned species was run on Linux (Ubuntu) at a personal computer. The LAST algorithm finds similar regions between genome sequences and aligns them accordingly. It is an algorithm designed to compare vertebrate genomes or large genome sequences such as chromosomal DNA. The installation and update of the LAST alignment software are available on the official site [32].

Alignment files, produced by LAST, may take up very large spaces and may become unfavorable for researchers to search for useful information manually. Basic text editors are not suitable to handle the files on large scales. In order to locate and search divergent sub-sequence pairs in alignment files, we created an open-source Python project [33]. The project consists of two files: differ.py and srch.py. differ.py finds and stores all divergent sub-sequences of aligned pairs with respect to given parameters. Mandatory parameters for differ.py are input file (-i), output file prefix (-o), and minimum value of divergent sequence length (-md). Two optional parameters are similarity tolerance (-ms) and early stop limit (-sa). An example command to run differ.py can be given as:

python differ.py -i*input_file* -o *output_prefix* -md *N* -ms*T* -sa*E*

The above command reads input_file, finds all divergent sub-sequence pairs of minimum length N with tolerance T, and stores the results in output_prefix_K_J.txt files, each of which are containing at most 10MB of data, where K ≥ N and J ≥ 1. Also, the program stops execution when E matches satisfying requirements given by former parameters. When the analysis is completed, differ.py outputs the length of the largest divergent sub-sequence to the terminal window. An example of the output of the proposed program is provided in Figure 1.

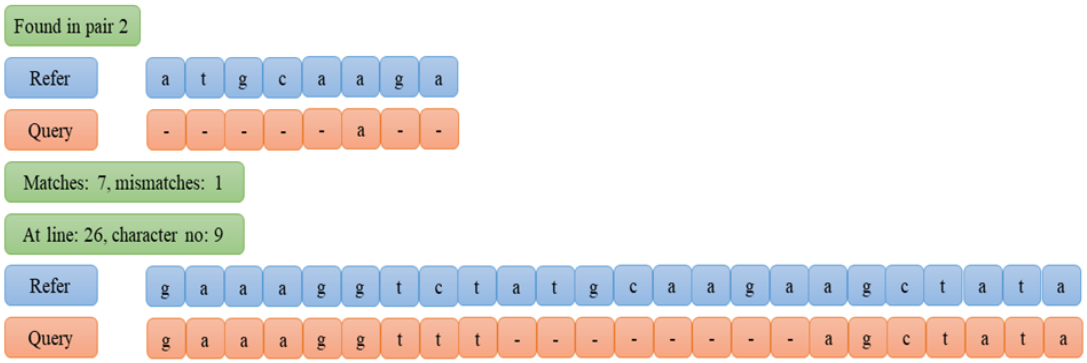

Figure 1. *An example of pairs in output*

The above statement can be interpreted as: at second pair in source alignment file, there is a 7 character long divergent sub-sequence with 1 erroneous character within, starting from 9th character at line 26. Reference sub-sequence is atgcaaga and query sub-sequence is -----a--. Also, the original reference and query sequences are given at the two bottom-most lines.

The second program, srch.py, perform a search for finding similar sub-sequences between output files generated by differ.py and a given FNA file. Parameters for srch.py are input file prefix (-ipx), the minimum length of similar sub-sequences (-min), the maximum length of similar sub-sequences (-max), which sequence in output files to look at (-ord), allowed percentage of indel symbols (-rid), tolerance of non-similarity (-tol) and target FNA file to perform the search (-target). An example command for running srch.py can be shown as:

python srch.py -ipxinput_prefix -min A -max B -ord R -rid I -tol T -target F

The above command performs a similarity search between all files with name input_prefix_X and F. Any sub-sequence, satisfying the parametrized requirements will be saved to output files named as FNA_R_Y.txt, where Y ≥ A and Y equals to the length of matching sub-sequence. The resulting output files are formatted as following:

> ATATA

Line:15, Src: results_5_1.txt: 44
Line:24, Src: results_5_1.txt: 44

The above statement can be interpreted as: results_5_1.txt file contains the sequence ATATA at its 44th line. This sequence is found at lines 15 and 24 of the given FNA file.

III. RESULTS

The proposed algorithm was implemented on a computer equipped with a Core I7 4720HQ 2.6 Ghz processor, 16 Gb DDR3 Memory, AMD® R9 M265X graphics card. 16GB of physical memory on the computer was not enough during the alignment processes used. This situation was solved with the SWAP operation command provided by the Ubuntu system during the installation and an area of 100 Gb from the HDD was recognized as virtual RAM to the operating system. Swap operation is a partition on the hard drive reserved by the operating system. When the size of the data exceeds the maximum RAM capacity, this part is used as RAM and thus the operations can continue.Performance indicators of proposed algorithms are shown Table 1.

Table 1.*Performance indicators of proposed algorithms*

Process	**Operating System**	**Device Features**				**Species**		**Time**
		CPU	**RAM**	**HDD**	**GPU**	**First**	**Second**	
Alignment (LAST)	Linux (Ubuntu)	Core I7 4720HQ 2.6 Ghz	16 GB DDR3+ 100GB SWAP	1 TB	AMD® R9 M265X	Pig (File Size:2.5 GB)	Cattle (File Size: 2.7 GB)	122h. 16 min. 3 sec.
First program (differ.py)	Linux (Ubuntu)	Core I7 4720HQ 2.6 Ghz	16 GB DDR3+ 100GB SWAP	1 TB	AMD® R9 M265X	Pig (File Size:2.5 GB)	Cattle (File Size: 2.7 GB)	153 h. 4 min. 26 sec.
Second program (srch.py)	Linux (Ubuntu)	Core I7 4720HQ 2.6 Ghz	16 GB DDR3+ 100GB SWAP	1 TB	AMD® R9 M265X	Pig (File Size:2.5 GB)		91 h. 56 min. 45 sec.
Second program (srch.py)	Linux (Ubuntu)	Core I7 4720HQ 2.6 Ghz	16 GB DDR3+ 100GB SWAP	1 TB	AMD® R9 M265X	Cattle (File Size:2.7 GB)		85 h. 4 min. 51 sec

Note:No otheroperationwasperformed on thecomputerduringtheseprocesses.

Abbreviations: CPU: Computing ProcessingUnit – RAM:ReadOnly Memory – HDD: Hard Disk Drive – GPU: Graphics ProcessingUnit – HQ: High Quality – Ghz: Gigahertz.

In the second program, srch.py, it took 91 hours, 56 minutes, and 45 seconds to search the outputs of differ.py, the first program, in the Pig.FNA, on the personal laptop with the above features. The same situation took 85 hours, 4 minutes, and 51 seconds for the Cattle.FNA file.

IV. DISCUSSION & CONCLUSION

Most of the primers used in species identification and classification studies have been designed to target genes on mitochondrial DNA. On the other hand, primers designed based on chromosomal DNA sequences will be more useful than mitochondrial DNA, especially in comparing genomes close to each other, such as breeds of the same breed (two different bovine genomes). However, processing chromosomal DNA information is challenging to carry out necessary analysis using user-friendly online tools due to the large size of the sequence data.Another drawback is that using stand-alone-tools requires extensive knowledge and practice to understand the executable implementations of the tools.

In this study, we encoded a tool to represent species-specific chromosomal DNA regions belonging to pig and bovine species by aligning DNA sequences with each other. The Output file of this tool is

created by parsing the different oligonucleotide sequences between the species separately considering user-determined INDEL frame lengths.

In summary, we introduced a tool written in Python that can easily design primers for researchers who want to identify races close to each other. The developed tool along with its implementation documents is available for the academic community.

Optimization of software development will continue in order to increase the performance of the developed tool and to produce output in a more reasonable time frame. An output data of this program will be tested on meat samples to assess the efficiency of the primer sets for detection of contaminations in meat samples for closely related species.

ACKNOWLEDGMENT: This study was supported by Bolu Abant İzzet University Scientific Research Projects (Project no: 2017.09.04.1123). In addition the authors would like to thank the Scientific, Industrial and Technological Application and Research Center (STARC) of Bolu Abant İzzet Baysal University for utilization of laboratories.

V. REFERENCES

[1] Q. Zia, M. Alawami, N. F. K. Mokhtar, R. M. H. R. Nhari and I. Hanish, "Current analyticalmethods for porcine identification in meat and meat products," *Food Chem.*, vol. 324, no. April 2019, pp. 126664, 2020.

[2] M. A. M. Hossain, S. M. K. Uddin, S. Sultana, S.Q. Bonny, M. F. Khan and Z. Z. Chowdhury, "Heptaplex polymerase chain reaction assay for the simultaneousdetection of beef, buffalo, chicken, cat, dog, pork, and fish in raw and heat-treated food products," *J. Agric. Food Chem.*, vol. 67, no. 29, pp. 8268–8278, 2019.

[3] A. Lopez-Oceja, C. Nuñez, M. Baeta, D. Gamarra, and M. M. de Pancorbo, "Speciesidentification in meat products: A new screening method based on high resolution melting analysis ofcyt b gene," *Food Chem.*, vol. 237, pp. 701–706, 2017.

[4] M. E. Ali, M.A. Razzak, S. B. A. Hamid, M. M. Rahman, M. Al Amin, N. R. A. Rashid, and Asign, "Multiplex PCR assay for the detection of five meat species forbidden inIslamic foods," *Food Chem.*, vol. 177, pp. 214–224, 2015.

[5] M. E. Ali, U. Hashim, S. Mustafa, Y. B. Che Man, Th S. Dhani, M. Kashif, M. K. Uddin, and S. B. A. Hamid, "Analysis of pork adulteration in commercial meatballs targeting porcinespecific mitochondrial cytochrome b gene by TaqMan probe real-time polymerase chain reaction," *Meat Sci.*, vol. 91, no. 4, pp. 454–459, 2012.

[6] R. Grujić and D. Savanović, "Analysis of myofibrillar and sarcoplasmic proteins in pork meatby capillary gel electrophoresis," *Foods Raw Mater.*, vol. 6, no. 2, pp. 421–428, 2018.

[7] M. Montowska and E. Pospiech, "Differences in two-dimensional gel electrophoresis patternsof skeletal muscle myosin light chain isoforms between Bos taurus, Sus scrofa and selected poultryspecies," *J. Sci. Food Agric.*, vol. 91, no. 13, pp. 2449–2456, 2011.

[8] M. Alikord, H. Momtaz, J. Keramat, M. R. Kadivar, and A. Homayouni, "Species identification and animal authentication in meat products : a review," *J. Food Meas. Charact.*, vol. 12, no. 1, pp. 145–155, 2018.

[9] J. M. N. Marikkar, M. E. Mirghani, and I. Jaswir, “Application of chromatographic andinfra-red spectroscopic techniques for detection of adulteration in food lipids: a review,” *J. Food Chem. Nanotechnol.*, vol. 2, no. 1, pp. 32–41, 2016.

[10] J. Mandli, I. EL Fatimi, N. Seddaoui and A. Amine, “Enzyme immuno assay (ELISA/immuno sensor) for a sensitive detection of pork adulteration in meat,” *Food Chem.*, vol. 255, no. January, pp. 380–389, 2018.

[11] P. K. Singh, G. Jairath, S. S. Ahlawat, A. Pathera and P. Singh, “Biosensor: an emerging safety tool for meat industry,” *J. Food Sci. Technol.*, vol. 53, no. 4, pp. 1759–1765, 2016.

[12] S. Roy, I. A. Rahman, J. H. Santos, and M. U. Ahmed, “Meat species identification usingDNA-redox electrostatic interactions and non-specific adsorption on graphene biochips,” *FoodControl*, vol. 61, pp. 70–78, 2016.

[13] S. Roy, N. F. Mohd-Naim, M. Safavieh and M. U. Ahmed, “Colorimetric nucleic aciddetection on paper microchip using loop mediated isothermal amplification and crystal violetdye,” *ACS Sensors*, vol. 2, no. 11, pp. 1713–1720, 2017.

[14] X. Tian, J. Wang, and S. Cui, “Analysis of pork adulteration in minced mutton using electronic nose of metal oxide sensors,” *J. Food Eng.*, vol. 119, no. 4, pp. 744–749, 2013.

[15] X. Tian, J. Wang, Z. Ma, M. Li, Z. Wei, and J. M. Díaz-Cruz, “Combination of an E-Nose andan E-Tongue for adulteration detection of minced mutton mixed with pork,” *J. Food Qual.*, vol. 2019, 2019.

[16] E. Novianty, L. R. Kartikasari, J. H. Lee, and M. Cahyadi, “Identification of pork contamination in meatball using genetic marker mitochondrial DNA cytochrome b gene by duplex PCR,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 193, no. 1, 2017.

[17] Z. Dai, J. Qiao, S. Yang and S. Hu, “Species authentication of common meat based on PCR analysis of the mitochondrial COI gene,” *Appl Biochem Biotechnol.*, no. 461, pp. 1770–1780, 2015.

[18] A. Doosti and P. G. Dehkordi, “Molecular assay to fraud identification of meat products,” *JFood Sci Technol.*, vol. 51, no. January, pp. 148–152, 2014.

[19] A. Di Pinto, M. Bottaro, E. Bonerba, G. Bozzo, E. Ceci, and P. Marchetti, “Occurrence ofmislabeling in meat products using DNA-based assay,” *J Food Sci Technol.*, vol. 52, no. April, pp. 2479–2484, 2015.

[20] R. Köppel, A. Ganeshan, S. Weber, K. Pietsch, C. Graf, R. Hochegger, K. Griffiths, and S. Burkhardt, “Duplex digital PCR for the determination of meat proportions of sausagescontaining meat from chicken, turkey, horse, cow, pig and sheep,” *Eur. Food Res. Technol.*, vol. 245,

[21] J. Ha, S. Kim, J. Lee, S. Lee and H. Lee, “Identification of pork adulteration in processed meat products using the developedmitochondrial DNA-based primers,” *Korean J. Food Sci. Anim. Resour.*, vol. 37, no. 3, pp. 464–468,2017.

[22] F. Guan, Y. Jin, J. Zhao, A. Xu and Y. Luo, “A PCR Method That Can Be Further Developedinto PCR-RFLP Assay for Eight Animal Species Identification,” *J. Anal. Methods Chem.*, vol. 2018, 2018.

[23] B. G. Mane and C. S. K. Hpkv, “PCR-RFLP assay for identification of species origin of meatand meat products 1,” vol. 2, no. 2, pp. 31–36, 2014.

[24] M. Huang, Y. Horng, H. Huang, Y. Sin and M. Chen, “RAPD fingerprinting for the speciesidentification of animals,” *Asian-Aust. J. Anim. Sci.*, vol. 16, pp. 1406–1410, 2003.

[25] M. Baker, “Digital PCR hits its stride,” *Nat. Methods*, vol. 9, no. 6, pp. 541–544, 2012.

[26] C. Floren, I. Wiedemann, B. Brenig, E. Schütz and J. Beck, “Species identification and quantification in meat and meat products using droplet digital PCR (ddPCR),” *Food Chem.*, vol. 173,pp. 1054–1058, 2015.

[27] H. R. Shehata, J. Li, S. Chen, H. Redda, S. Cheng, N. Tabujera, H. Li, K. Warriner, and R. Hanner, “Droplet digital polymerase chain reaction (ddPCR) assays integrated withan internal control for quantification of bovine, porcine, chicken and turkey species in food and feed,” *Plos One*, vol. 12, no. 8, 2017.

[28] R. Köppel, F. Zimmerli and A. Breitenmoser, “Heptaplex real-time PCR for the identification and quanti W cation of DNA from beef , pork , chicken , turkey , horse meat , sheep (mutton)and goat,” *Eur. Food Res. Technol.*, pp. 125–133, 2009.

[29] G. Barcaccia, M. Lucchin and M. Cassandro, “DNA barcoding as a molecular tool to trackdown mislabeling and food piracy,” *Diversity*, vol. 8, no. 1, 2016.

[30] K. Nakyinsige, Y. B. C. Man and A. Q. Sazili, “Halal authenticity issues in meat and meatproducts,” *Meat Sci.*, vol. 91, no. 3, pp. 207–214, 2012.

[31] N. Z. Ballin, F. K. Vogensen and A. H. Karlsson, “Species determination - Can we detect andquantify meat adulteration?,” *Meat Sci.*, vol. 83, no. 2, pp. 165–174, 2009.

[32] LAST, “No Title,” *Genome-Scale Sequence Comparison*. (2020, September 22). [Online]. Available:http://last.cbrc.jp/doc/last.html.

[33] SBPD, “SBPD,” *Software Based Primer Design*. (2020, October 06) [Online]. Available: https://github.com/ihpar/FnaSrch.