Open Access Journal of Information Systems (OAJIS)

Available at http://is.its.ac.id/pubs/oajis/index.php

The Third Information Systems International Conference

DNA QR Code Scanner for Identifying the Species Origin of Meat Products

Tigor Nauli *

Research Center for Informatics, Indonesian Institute of Sciences (LIPI), Kompleks LIPI Gd 20 Lt 3, Jalan Cisitu, Sangkuriang, Bandung 40135, Indonesia

Abstract

DNA barcoding is a molecular technology for species identification using a standard fragment of DNA sequence. Species identification of animal material in food samples is essential to authenticate the components of meat products from illegal substitutions. DNA barcoding make it possible to distinguish meat species of closely related animals. DNA barcoding is a molecular technology for species identification using a standard fragment of DNA sequence. Species identification of animal material in food samples is essential to authenticate the components of meat products from illegal substitutions. DNA barcoding make it possible to distinguish meat species of closely related animals. At the present, there is no practical application available for DNA barcoding because of the difficulty in retrieval of DNA sequences in a feasible data input. We attempt to eliminate the limitation by encoding the DNA sequence into a QR code. DNA extracted from meat products and had sequencing can be stored as DNA QR code. A DNA QR code scanner was developed to propose a practical application of DNA barcoding technology. This Android application is derived from ZXing barcode scanner and is equipped with DNA barcodes database of domestic animals. The DNA QR code of an unidentified specimen is compared with the reference DNA barcodes to find the matching species by local sequence alignment of Smith-Waterman. The application can help users to ascertain the meat content in food they purchase by scanning the DNA QR code label attached on the product.

Keywords: Android application, DNA barcoding; DNA QR code; meat species; scanner; sequence alignment; species identification

1. Introduction

DNA barcoding is a molecular technology for rapid identification of animals and plants in species level. It uses sequence comparisons of a standard fragment of DNA sequence ("DNA barcode") to distinguish species. DNA barcoding was first introduced for species identification of biological diversity

* Corresponding author. Tel.: +62-22-2504711; fax: +62-22-2504712.
E-mail address: tigor.nauli@lipi.go.id.

in taxonomic research [1]. At present, it has been shown valuable benefits to many applications such as ecological forensics, combating diseases vectors, protecting endangered species, food product regulation.

Species identification is essential for the detection and identification of animal material in food samples. Meat substitution with cheaper or less desirable or objectionable species has become our concerns for health, religious, aesthetic or legal reason. It is necessary to identify and/or authenticate the components of meat products from illegal substitutions [2]. DNA barcoding make it possible to distinguish meat species of closely related animals [3], or meat that occurs individually or in a complex mixture [4].

At the present, there is no practical application available for DNA barcoding because of the difficulty in information retrieval through direct scanning of DNA sequences [5]. The DNA sequences in plain format is very long string of characters which is also not feasible for data input. We attempt to eliminate the limitation by encoding the DNA sequence into a more compact form. Quick Response (QR) code has been identified as the best of the available barcode types for representing the DNA sequences.

In this study, a DNA QR code scanner was developed to propose a simple application of DNA barcoding technologies. This mobile application help users to recognize the species origin of meat product by scanning the DNA QR code label.

1.1. The DNA barcode

All organisms, including animals, plants, fungi, and bacteria, contain DNA within their cells. DNA comprises nucleotide “bases” adenine, cytosine, guanine, and thymine (A, C, G, T), which are arranged in very specific sequence, to encode the functional or structural proteins. DNA barcode refers to a short section of DNA from a standardized DNA region (“gene”) of the genome. That DNA sequence can be used to identify different species, in the same way a supermarket scanner uses the familiar black stripes of the UPC barcode to identify different groceries [6].

The DNA sequence successfully used in DNA barcoding in animals is the 5' end of the mitochondrial gene cytochrome oxidase 1 (*cox1* or *COI*). This sequence is recognized as universal barcode of the kingdom Animalia and is used to authenticate and trace animal species and breeds [7].

The conventional means of generating DNA sequence data to obtain a barcode for a species or a specimen are through PCR amplification using species-specific primers and Sanger sequencing of DNA barcode sequences from genomic DNA extracted from individual specimens [8].

1.2. Sequence Alignment

The barcode of an unidentified specimen can be compared with the reference barcodes to find the matching species. Searching database with a DNA sequence rely heavily on sequence comparison techniques. To identify regions of similarity between the DNA sequences, two sequences are aligned in an optimal arrangement. The optimal alignment of two sequences is chosen from the maximum score of matching pairs, mismatching pairs, and penalty score of the gaps.

The current algorithms for comparing biological sequences are based mostly on the technique of sequence alignment [9]. The algorithm includes hierarchical clustering (parsimony and neighbor joining), similarity methods, combines clustering/similarity methods, and diagnostic methods.

1.3. QR Code

Quick Response (QR) Code was the earliest 2D barcode. The codes carry meaningful information in the vertical direction as well as the horizontal. Therefore, the codes can carry up to several hundred times the amount of data carried by ordinary barcodes (which storing a maximum of 20 digit).

QR Code consists of black modules arranged in a square pattern on white background. Three big square marks on the three corners of the code are required for positioning the reader, while several smaller marks arranged in some places in the code are for aligning the pattern. Among 2D barcodes, QR Code has the largest capacity to carry 7,089 numeric, 4,296 alphanumeric characters, and 2,953 bytes of binary (8 bits) data. The QR Code has the best compression efficiency in encoding DNA barcode sequences among the other 2D Codes [5].

2. Materials and Methods

2.1. Data set of DNA sequences

The gen *COI* sequences of several animals were retrieved from GenBank nucleotide sequence database (http://ncbi.nlm.nih.gov). The DNA sequences were queried from the GenBank by keywords "barcode". Some of the DNA sequences were extracted from the complete genome as CDS for gene="COX1".

We considered with the natural meat species in the food, hence DNA barcode sequences of common domestic animals were chosen as reference barcodes. The DNA sequences with accession number JX426135 (617 bp), JN245997 (623 bp), JN245994 (808 bp), and JN632605 (781 bp) were chosen as test set ("the samples").

2.2. Encoding DNA sequence into DNA QR code

The open source QR Code library (in Java) was adapted for developing a program to encode the DNA sequences [10]. To generate a QR code, we first created an instance of the *QRCodeWriter*. Then called its *encode()* method with four settings: the text to encode ("the DNA sequence"), barcode type ("the QR code") and the desired width and height (in pixels) of the image produced. The *encode()* method does the job and returns a matrix of bytes. Next we used the *MatrixToImageWriter* class, and its *toBufferedImage()* method, to convert the matrix into a *BufferedImage* which is easily converted to a CF image object.

2.3. Developing a scanner for reading QR code

The QR code scanner was built as a separate component by extending the ZXing barcode scanner [10]. To grab the printed image of QR code, we first created *ScannerView* as the camera interface of the Android device [11]. This camera view overwrittes several methods of (ZXing) *BarcodeScannerView* class. The *PlanarYUVLuminanceSource* was positioned in the middle of the camera and acts as a viewfinder frame that scan any barcode inside. The picture of the barcode taken by camera snapshot will be converted into a binary image (bitmap) by *HibridBinarizer(picture)* method.

The camera will stop automatically once the focused picture gets captured. The *ResultHandler* is responsibles to start the camera again, if the previous snapshot was failed. Then, the bitmap will be converted into string of text by *decodeWithState(bitmap)* method of *MultiFormatReader()* class accordingly, depending on the thirteen types of *BarcodeFormat*.

2.4. Comparison of DNA sequences

To make comparative statements about DNA sequences, a sequence alignment is needed. The basic concept of selecting an optimal sequence alignment is simple. The two sequences are matched up in an arbitrary way. The quality of the match is scored. Then one sequence is moved with respect to the other and the match is scored again, until the best alignment is found. Sequence similarity refers to the occurrence of exactly the same nucleic acid in the same position in two aligned sequences.

The application implemented the local pairwise sequence alignment of Smith-Waterman algorithm [12] with Gotoh's improvement to compare two DNA sequences. The Gotoh's algorithm uses affine gap penalties [13], where initiation of gaps is more expensive than extension of an existing gap.

This "in-exact matching" algorithm uses the dynamic programming method to optimize the number of base pairs with minimum gap scoring. The optimal alignment of two sequences is chosen from the maximum score of matching pairs, mismatching pairs, and penalty score of the gaps.

We extend the functionality of Java libraries of JAligner to perform the sequence alignment [14]. The parameters are set prior the alignment and the matrix score (for match and mismatch) was built previously.

2.5. The Application

The "DNA QR Code" scanner was built under Android 21 platform on Linux operating system (IGOS Nusantara X variant, kernel 4.0.5). We used the Android Studio 1.3 for fastened prototyping of the application [15]. All deployment tests to the application have been verified in the Android emulator (AVD) and in the real handset through DDMS perspective.

We implemented four *activity* classes, one for each feature of the application:

- *MainActivity* -This activity serves as the default activity to launch. Its layout has two display for scanned DNA QR codes and has three buttons, each corresponding to a feature of the application. The *onClick()* handlers for each button trigger cause the associated activity to launch.
- *ScannerActivity* -This activity handles the camera interface for capturing the DNA QR Codes and decoding back to DNA sequences.
- *AlignmentActivity* -This activity provides the comparison of two DNA sequences through the sequence alignment and a calculation to the sequence similarity.

QueryActivity -This activity does the main feature of the application. It takes the DNA sequence query; comparing it with reference DNA barcodes, which were stored in SQLite database, in pairwise basis for similarity; and reports the corresponding species name which has the highest similarity score.

3. Results and Discussion

3.1. DNA QR code of samples

Four DNA QR codes of the samples appear in more dense pixels compare with the common QR code for the URL, as seen in Fig. 1. More data we put into QR code, the more rows and columns of modules (the little black squares) will be introduced. Therefore, the DNA QR codes of the samples look darker than the common QR codes.

The minimum width of a printed QR code image depends on the size of an individual module when viewed by the camera (from a distance at some resolution). We found that the minimum width of a printed DNA QR code is 2.8 cm. Smaller than that minimum width, the scanner cannot read the DNA QR code correctly.

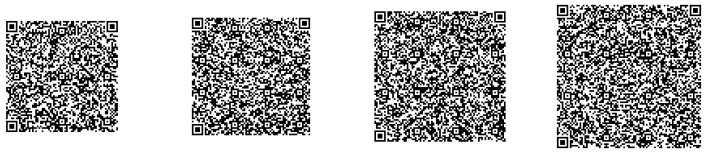

Fig. 1. DNA QR codes of (a) JX426135; (b) JN245997; (c) JN245994; (d) JN632605.

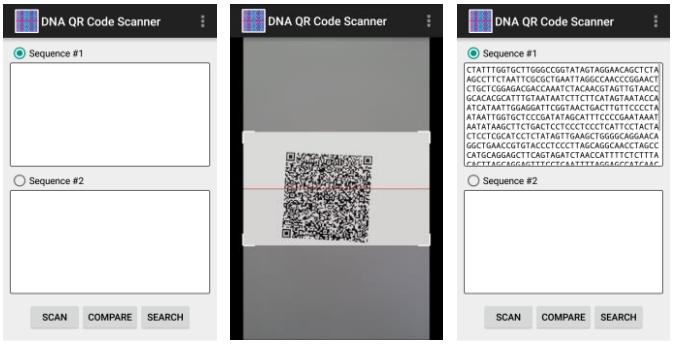

Fig. 2. (a) First screen after the application was started; (b) DNA QR code was captured by device's camera; (c) Previous screen after DNA QR code was decoded back to DNA sequence.

3.2. Scanning the DNA QR code

When the application was started, it show two blank rectangles in the middle of device screen and three captioned buttons place at the bottom. By touching the [SCAN] button, the screen switches to the camera interface. User can place the DNA QR code inside the viewfinder. The scanner captured the DNA QR code and decoding back into DNA sequence. One of the rectangles is now fill with the decoded DNA sequence, as in Fig. 2. The selected radio button on the top of each rectangle indicates which rectangle is currently focused. If the scanner failed to decode the DNA QR code, the rectangle contains some random numbers only. In this case, the user should repeat the scanning again.

3.3. Comparison of two DNA sequences

A sequence alignment is a way of arranging the sequences of DNA to identify regions of similarity. Two DNA sequences are exactly the same when the similarity score is 100%. These identical DNA sequences came from the same species. For this study, we chose the arbitrary 95% similarity for identical species [16].

The application will conduct a sequence alignment when [COMPARE] button is touched. The detailed picture of the alignment between JX426135 (Seq. #1) and JN632605 (Seq. #2) are showed in Fig. 3. DNA sequence JX426135 (Seq. #1) from 1 to 617 aligned with DNA sequence JN632605 (Seq #2) from 61 to 677. About 5.51% of their nucleotides are difference. This indicates that the two DNA sequences are belongs to different species.

3.4. Searching the DNA sequence of species

Query DNA sequence in reference DNA QR code is the main feature of the application. It involves the pairwise sequence alignment between query sequence and each of barcode sequence in reference database. While repeating the pairwise alignment, we keep the larger similarity score than the previous one. The conclusion was taken from the pair DNA sequences with the largest similarity score.

Searching the DNA sequence will start when [SEARCH] button is touched. The application takes the sequence in one of the rectangles as the query DNA and reports the result in a new display. Fig. 4 exhibits the query result for each of samples.

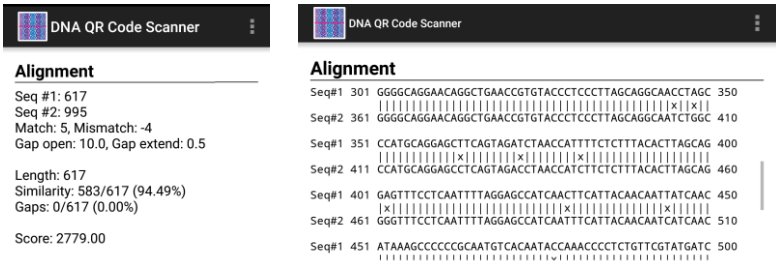

Fig. 3. A sequence alignment of DNA sequence JX426135 (Seq. 1) and DNA sequence JN632605 (Seq. 2).

Sample JX426135 is exactly the DNA of meat cow (cattle). Its DNA sequence (617 bp) matched with cow's barcode. Sample JN245997 was identified as the DNA of meat pig with 98.56% similarity. And sample JN245994 was identified as the DNA of meat goat with 99.63% similarity.

However, the application failed to recognize the JN632605 as the DNA of banteng (*Bos javanicus*). The highest similarity score between JN632605 and reference barcodes is 94.21%. This failure was mainly caused by missing the species barcode in the reference database. The closest barcode is the DNA sequence of cow (*Bos taurus*).

The successful querying the DNA sequence depends on the local sequence similarity. DNA scoring matrices are usually implemented as match/mismatch score. The similarity uses the scoring matrix which scores a match as +5 and a mismatch as -4. By the default, the FASTA program [17] also uses +5/-4. While the NCBI nucleotide BLAST web site (megablast) [18] uses match/mismatch scores of +1/−3 with target sequences that are 99% identical. We implemented the scoring matrix that similar with FASTA, because more sensitive DNA scoring parameters (BLAST) are effective for longer DNA evolutionary distances, e.g. mouse-human [19]. The DNA sequence of our barcodes are considered short.

3.5. Extension of DNA barcodes

The important part of the application is the scanner and the reference barcodes. The scanner is responsible for decoding DNA QR codes. The reference barcodes and the sequence alignment determine the successful species identification. The species identification of sample JN632605 has better result if the reference barcodes database comprises a complete list of species barcodes.

A notion of a “species” is difficult to formally define. There still debate about what constitutes a species. Definitions of species tend to fall into the morphological and the biological species concepts [20]. Cattle (*Bos taurus*) and banteng (*Bos javanicus*) have same morphology (appearance) as cow, but they are genetically different species. To extend the reference DNA barcodes, we should add several barcodes of species, for example, banteng (*Bos javanicus*) in cow group; domestic pig (*sus scrofa domesticus*) in pig group; and so on.

It is not necessary to list all species of the animals, because we are only interested in the species origin of meat products. Hence, we can limit the list of reference DNA barcodes to the barcodes of “edible” meat species. With the complete reference DNA barcodes, we assume that the “unknown species” are now belongs to meat of uncommon or objectionable species.

To protect consumer rights, the food authority can adopt the DNA barcoding technique for species authentication. The legislation should therefore impose DNA QR code as a genetic label of food products declaring the species origin inside. Then, by using “DNA QR code scanner”, customer can check the meat content of food products they purchase easily.

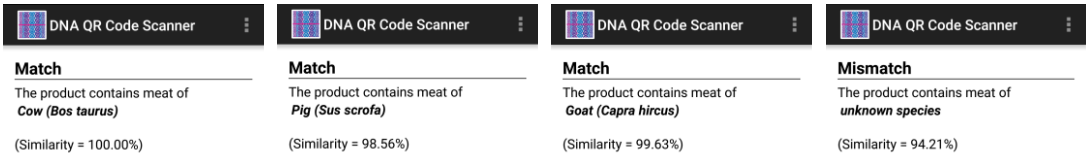

Fig. 4. Query results for (a) JX426135; (b) JN245997; (c) JN245994; (d) JN632605.

The scanner application was targeted for meat species identification because it includes a small set of barcodes which can easily fit in device memory, and enable working offline. With small modification to the system, the application can also be applied for species identification in general. We should switch the link from barcode database in device memory to the public barcode database, such as in GenBank (http://ncbi.nlm.nih.gov) or iBOL (http://boldsystems.org). However, the scanner is not longer a simple application, since it requires online access to a large barcode database.

4. Conclusion

The DNA QR code scanner provides tool for identifying the species origin of meat products by direct scanning printed DNA QR code. The DNA QR code of an unidentified specimen is compared with the reference DNA barcodes stored in device memory to find the matching species by local sequence alignment of Smith-Waterman. The reference DNA barcodes comprises with the barcodes of common meat species.

Acknowledgments

The author gratefully acknowledges the discussions and suggestions offered by Dr. Zalinar Udin.

References

[1] Ratnasingham S, Hebert PD. BOLD: the barcode of life data system (www.barcodinglife.org). *Mol Ecol Notes* 2007;**7**:355–64.
[2] Pascal G, Mahe S. Identity, traceability, acceptability and substantial equivalence of food. *Cell Mol Biol* 2001;**47**:1329–42.
[3] Spychaj A, Mozdziak PE, Pospiech E. PCR methods in meat species identification as a tool for the verification of regional and traditional meat products. *Acta Sci. Pol., Technol. Aliment*. 2009;**8**(2),5-20.
[4] Tillmar AO, Barbara Dell'Amico B, Welander J, Holmlund G. A universal method for species identification of mammals utilizing next generation sequencing for the analysis of DNA mixtures. *PloS ONE* 2013;**8**(12):e8376.
[5] Liu C, Shi L, Xu X, Li H, Xing H, Liang D, et al. DNA barcode goes two-dimensions: DNA QR Code Web Server. *PLoS ONE* 2012;**7**(5):e35146.
[6] Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc R Soc Lond B*, 2003;**270**:313-21.
[7] Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Mol Ecol Resources* 2014;**14**,892–901.
[8] Kress WJ, Erickson DL. DNA barcodes: methods and protocols. New York: Humana Press; 2012.
[9] Zhang Y, Chen W. A new measure for similarity searching in DNA sequences. *MATCH Commun Math Comput Chem* 2011;**65**:477-88.
[10] ZXing ("zebra crossing"): Open-source, multi-format 1D/2D barcode image processing library implemented in Java, with ports to other languages., (https://github.com/zxing/zxing) (December 12, 2014)

[11] Maguluru D. Android library projects that provides easy to use and extensible Barcode Scanner views based on ZXing and Zbar., (https://github.com/dm77/barcodescanner) (December 27, 2014)
[12] Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol* 1981;**147**(1):195-7.
[13] Gotoh O. An improved algorithm for matching biological sequences. *J Mol Biol* 1982;**162**(3):705-8.
[14] Moustafa A, Jaligner: Open-source Java implementation of the Needleman–Wunsch and Smith-Waterman algorithms for biological pairwise sequence alignment with the affine gap penalty model., (https://github.com/ahmedmoustafa/JAligner) (January 20, 2015)
[15] Google. Android Studio: the official IDE for Android application development, based on IntelliJ IDEA., (http://developer.android.com/ tools/studio/index.html) (November 21, 2014)
[16] Akashi, H. Within- and between-species DNA sequence variation and the 'footprint' of natural selection. *Gene* 1999; **238**:39-51.
[17] Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. *Meth Enzymol* 1990;**183**:63-98,
[18] Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. *Bioinformatics* 2008;**15**:1757-64
[19] Pearson WR. Selecting the Right Similarity-Scoring Matrix. *Curr Protoc Bioinformatics* 2013; **43**:3.5.1-3.5.9.
[20] Schuh RT, Brower AVZ. Biological Systematics: Principles and Applications. Ithaca: Cornell University Press; 2009.

Appendix A.

A.1. List of DNA barcodes of Some Domestic Animals as reference database

No.	Common Name	Scientific Name	GenBank Accession	Length (bp)
1	Buffalo	*Bubalus bubalis*	AF547270	901
2	Cat	*Felis catus*	JF443237	596
3	Chicken	*Gallus gallus*	JF498862	694
4	Cow	*Bos taurus*	HQ860420	658
5	Crab			
	Circular crab	*Atelecyclus rotundatus*	JQ305993	625
	Sand crab	*Portunus pelagicus*	EF661948	573
6	Dog	*Canis lupus familiaris*	JF443206	558
7	Donkey	*Equus asinus*	KC694097	627
8	Duck	*Anas platyrhyncha*	GU571722	648
9	Fish			
	Longtail tuna	*Thunnus tongkol*	JN644307	623
	Skipjack tuna	*Katsuwonus pelamis*	HQ167708	515
	Milk fish	*Chanos chanos*	JN242679	652
	Snakehead murrel	*Channa striata*	HM117206	582
	Sardine	*Sardina pilchardus*	KJ205157	628
	Narrow-barred mackerel	*Scomberomorus commerson*	HQ167712	515
	Indian anchovy	*Stolephorus indicus*	JX676139	649
10	Goat	*Capra hircus*	NC005044	901*
11	Horse	*Equus caballus*	HM102300	616
12	Mouse	*Mus musculus*	JF459217	657
13	Pig	*Sus scrofa*	NC000845	901*
14	Pigeon	*Columbidae livia*	GU571831	648
15	Prawn			
	Giant freshwater prawn	*Macrobrachium rosenbergii*	AB235295	608
	Black tiger shrimp	*Penaeus monodon*	KC409381	656
	American lobster	*Homarus americanus*	DQ889104	654
16	Rabbit	*Oryctolagus cuniculus*	HM102307	599
17	Rat	*Rattus rattus*	JF445260	558
18	Sheep	*Ovis aries*	JF443355	657
19	Squid			
	European squid	*Loligo vulgaris*	KC311395	654
	Common cuttlefish	*Sepia officinalis*	KC789480	654
	Common octopus	*Octopus vulgaris*	KC789315	654
20	Turkey	*Meleagris gallopavo*	HM102303	639

*complete CDS of gene *COI*

**DNA sequences of barcodes are not included