Series: Engineering and Science Vol. 7 No. 1 (2021) E-ISSN: 2621-9794, P-ISSN: 2477-2097

DETEKSI GEN POTENSIAL HEWAN TERNAK UNTUK AUTENTIKASI PRODUK HALAL MENGGUNAKAN ANALISIS N-GRAM

Ruth Ema Febrita[1)] dan **Maghfirotul Amaniyah[2)]**

[1]Jurusan Teknik Informatika, Politeknik Negeri Banyuwangi, Jl. Raya Jember KM 13, Kawang, Labanasem, Banyuwangi, 68461
[2]Program Studi Teknik Pengolahan Hasil Ternak, Politeknik Negeri Banyuwangi, Jl. Raya Jember KM 13, Kawang, Labanasem, Banyuwangi, 68461
E-mail: ruthemafebrita@poliwangi.ac.id

Abstract

For a country with a majority Muslim population like Indonesia, the guarantee of the authenticity of halal food is a necessity to ensure consumer protection. One form of halal authentication testing on food is to use polymerase chain reaction (PCR) to obtain the DNA sequences. However, analyzing the whole sequences manually is time-consuming. Therefore, it is necessary detect the identifying markers found in the protein sequences of pigs and dogs that are not possessed by other animals using n-gram analysis. This study succeeded in finding two primer marker blocks that were only found in the protein structure of dogs and pigs, although the position of the markers could not be determined specifically due to different sequence lengths.

Keywords**:** *halal authentication, dog DNA, pig DNA, identifying marker, n-gram*

Abstrak

Jaminan akan keaslian makanan halal merupakan suatu kebutuhan untuk menjamin perlindungan terhadap konsumen, khususnya di negara dengan mayoritas penduduk beragama Islam seperti di Indonesia. Salah satu bentuk pengujian autentikasi halal pada makanan adalah dengan menggunakan *polymerase chain reaction* (PCR) yang digunakan untuk mendapatkan sekuens DNA. Melakukan analisis secara manual pada keseluruhan panjang sekuens akan memakan banyak waktu. Oleh karena itu perlu dilakukan deteksi marker penciri yang terdapat pada sekuens protein babi dan anjing yang tidak dimiliki oleh hewan lain menggunakan analisis n-gram. Penelitian ini berhasil menemukan dua blok marker yang hanya ditemukan pada struktur protein anjing dan babi, walaupun posisi marker tidak dapat ditentukan secara spesifik karena panjang sekuens yang berbeda-beda.

Kata Kunci: *autentikasi halal, dna anjing, dna babi, marker, n-gram*

PENDAHULUAN

Indonesia merupakan sebuah negara yang mayoritas penduduknya memeluk agama Islam. Dalam hal pangan, fakta tersebut sangat berdampak pada kebutuhan terhadap penjaminan keaslian makanan halal (autentikasi) yang menjadi hal yang sangat sensitif dan penting sebagai jaminan hukum dan jaminan kualitas, dan perlindungan konsumen. Sebuah makanan dikatakan halal apabila memenuhi beberapa kriteria, antara lain tidak mengandung zat yang berasal dari babi, tidak memabukkan, berasal dari hewan yang

Series: Engineering and Science Vol. 7 No. 1 (2021) E-ISSN: 2621-9794, P-ISSN: 2477-2097

tidak haram menurut syariat Islam, tidak termasuk dalam kategori najis, dan semua proses pengolahannya tidak digunakan untuk babi atau barang najis lainnya (Hidayatullah, 2020). Prosedur pengujian untuk autentikasi halal dapat menggunakan pendekatan kimiawi atau biologis, atau menggunakan metode berbasis lipid, protein, atau DNA (Premanandh & Salem, 2017; He & Yang, 2018).

Pengujian berbasis DNA biasanya dapat digunakan untuk menganalisis kontaminan dalam bahan mentah maupun dalam makanan olahan. Untuk mendapatkan sekuens DNA biasanya menggunakan metode *polymerase chain reaction* (PCR). Sebuah sekuens yang dikodekan dapat memiliki panjang puluhan hingga ratusan asam amino yang dapat dikodekan dalam ratusan hingga ribuan karakter (huruf). Analisis n-gram dalam penelitian ini bertujuan untuk melakukan pendekatan berbeda dalam menganalisis kandungan yang terdapat pada suatu sekuens DNA/ gen yang pada umumnya memiliki panjang ratusan huruf dengan cara yang lebih cepat.

Penelitian ini bertujuan untuk mencari *marker* dalam susunan sekuens protein hewan yang diharapkan dapat menjadi penanda, baik secara struktur (sekuens) maupun posisi dengan menggunakan analisis n-gram. N-gram merupakan salah satu teknik dalam *machine learning* yang biasa digunakan dalam melakukan analisis berbasis teks. Hasil penelitian ini diharapkan dapat menjadi referensi dalam pengujian produk halal berbasis DNA.

Sebuah pendekatan analisis menggunakan kombinasi metode n-gram dan skip-gram yang dimodifikasi digunakan untuk mengekstraksi fitur pada sekuen protein (Islam, dkk., 2018). Dalam penelitian tersebut n-gram digunakan sebagai metode dalam melakukan ekstraksi fitur untuk memprediksi fungsionalitas dalam sebuah sekuen protein. Potongan sekuen yang diperoleh dengan menggunakan n-gram juga merepresentasikan motif dari sebuah sekuen protein. Sedangkan metode skip-gram digunakan untuk mengabaikan potongan sekuen tertentu dalam proses fitur ekstraksi. Pendekatan ini sangat berguna dalam membandingkan sebuah potongan mutasi yang memiliki panjang k karakter dengan sekuen protein.

Fitur k-skip-n-gram adaptif juga telah digunakan untuk melakukan ekstraksi informasi gen penyakit untuk mengidentifikasi gen penyakit Alzheimer (Xu dkk., 2019). Dalam penelitian tersebut n-gram yang digunakan adalah 2-gram. Sampel-sampel ekstraksi gen tersebut kemudian ditransformasikan ke dalam vektor dan

Series: Engineering and Science Vol. 7 No. 1 (2021) E-ISSN: 2621-9794, P-ISSN: 2477-2097

dianalisis menggunakan metode klasifikasi dengan random forest. K-skip-n-gram adaptif dapat digunakan untuk menemukan keterhubungan informasi baik pada residu yang bersinggungan maupun yang tidak bersinggungan.

N-gram juga digunakan untuk menganalisis sekuens DNA dari 15 keadaan kromatin dari Broad Histon Track dan mengaplikasikan pemodelan bahasa biologis ke dalam bentuk n-gram. Berdasarkan frekuensi kemunculan sekuens, ditemukan bahwa beberapa n-gram tertentu cenderung muncul di suatu keadaan kromatin, namun jarang muncul pada keadaan kromatin lainnya, sehingga disebut sebagai penciri dari suatu kromatin tersebut.

METODE PENELITIAN

Penelitian ini menerapkan teknik n-gram untuk mengekstraksi sekuens protein pada hewan menjadi beberapa potongan sekuens yang lebih kecil, yang kemudian akan diolah dan dibandingkan untuk dideteksi struktur sekuens yang menjadi penciri pada sekuens hewan yang “tidak halal”.

Dataset

Data sekuens protein yang digunakan dalam penelitian ini diambil dari *Universal Protein Resource* (UniProt) yang menyediakan *resource* berkaitan dengan sekuens protein dan anotasi (Bateman, dkk., 2021). Secara khusus bentuk sekuens protein yang digunakan adalah IGF2 (Insulin Like Growth Factor 2). Data yang digunakan dalam penelitian ini terdiri dari 34 sekuens protein hewan, dimana di dalamnya terdapat lima sekuens protein babi, lima sekuens protein anjing, dan sisanya sekuens protein hewan lainnya, seperti sapi, kuda, monyet, tikus, ayam, kambing, domba, dan ikan.

N-Gram

N-gram merupakan sebuah teknik probabilistik untuk memprediksi kata selanjutnya yang mungkin dari N-1 kata sebelumnya dengan cara memecah suatu string ke dalam bentuk yang lebih kecil sejumlah n kata atau n karakter/huruf. N pada n-gram dapat bernilai 1, 2, 3, dan seterusnya. Karena sebuah sekuens protein tidak berisi spasi di dalamnya, dengan demikian pemotongan sekuens protein dilakukan berdasarkan n karakter/huruf. Ilustrasi n-gram akan diberikan pada Tabel 1.

Series: Engineering and Science Vol. 7 No. 1 (2021) E-ISSN: 2621-9794, P-ISSN: 2477-2097

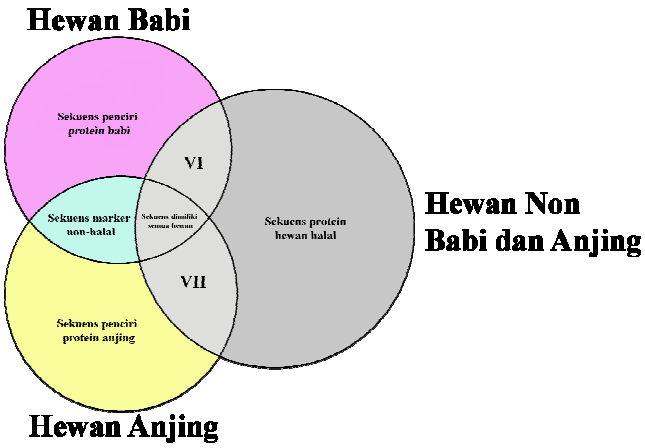
Tabel 1. Ilustrasi Proses Pemecahan Sekuens dengan n-gram

Sekuens	Subsekuens dengan n=5 (5-gram)
"MVSPDPQIIVVAPETELAS"	['MVSPD', 'VSPDP', 'SPDPQ', 'PDPQI', 'DPQII', 'PQIIV', 'QIIVV', 'IIVVA', 'IVVAP', 'VVAPE', 'VAPET', 'APETE', 'PETEL', 'ETELA', 'TELAS']

Dapat disaksikan pada Tabel 1 bahwa contoh sekuens protein terdiri dari 19 huruf yang kemudian dipecah menjadi sekuens yang lebih kecil yang terdiri dari masing-masing 5 huruf, sehingga menghasilkan 15 sekuens yang berbeda sesuai dengan urutan karakter pada sekuens asli.

Metode Analisis Data

Analisis data dilakukan dengan analisis himpunan, dimana hasil penelitian yang dicari adalah interseksi sekuens pada babi dan anjing yang tidak dimiliki oleh hewan lainnya. Sekuens anggota pada himpunan babi dan anjing dibandingkan dengan anggota pada himpunan sekuens hewan lainnya. Sekuens yang menjadi marker/penciri non-halal pada protein babi dan anjing merupakan sekuens yang tidak akan terdapat pada hewan lain. Ilustrasi metode analisis data akan dijelaskan pada Gambar 1.



Gambar 1. Ilustrasi teknik analisis data dengan teori himpunan.

Setelah analisis data berhasil dilakukan, sekuens penciri protein babi dan anjing yang terletak berurutan kemudian direkonstruksi ke dalam bentuk asli sekuens dan diberikan tanda pada posisi penciri. Rekonstruksi dilakukan dengan cara menyusun kembali urutan huruf protein pada kumpulan sekuens penciri, mulai dari huruf pertama

Series: Engineering and Science Vol. 7 No. 1 (2021) E-ISSN: 2621-9794, P-ISSN: 2477-2097

yang ditemukan hingga huruf terakhir pada potongan sekuens yang mengandung beberapa huruf pada potongan sekuens sebelumnya.

HASIL DAN PEMBAHASAN

Pengujian N-Gram

Dalam penelitian ini telah dilakukan pengujian sejumlah nilai n pada n-gram. Pengujian nilai n dilakukan guna mengetahui batas nilai n terbaik yang memungkinkan untuk bisa membaca marker (sekuens protein penciri) yang dicari. Berdasarkan prosedur pengujian PCR, pada umumnya panjang primer yang digunakan berkisar antara 18-25 basa nukleotida (Yusuf, 2010). Berdasarkan referensi tersebut, maka nilai n=[18-25] digunakan untuk pencarian sekuens marker. Penggunaan nilai n yang lebih kecil juga digunakan untuk melihat kemungkinan adanya sekuens marker yang berukuran lebih kecil dari 18 karakter. Hasil pengujian nilai n akan disajikan pada Tabel 2.

Tabel 2
Pengujian Nilai N pada N-Gram

N	Σ sekuens unik semua hewan	Σ sekuens unik babi	Σ sekuens unik anjing	Σ sekuens babi di hewan lain	Σ sekuens anjing di hewan lain	Marker non-halal
8	1594	270	380	146	157	7
9	1649	270	383	137	149	7
10	1697	270	386	130	142	7
11	1741	270	389	123	134	6
12	1779	270	392	118	128	6
13	1816	270	395	113	122	7
14	1851	270	398	108	116	8
15	1880	270	401	104	112	8
16	1907	270	404	101	109	8
17	1932	270	407	99	106	9
18	1957	270	410	97	103	8
19	1979	270	413	95	99	7
20	2000	270	416	93	95	6
21	2021	270	419	91	91	5
22	2042	270	422	89	87	4
23	2063	270	425	87	83	3
24	2083	270	428	85	79	2
25	2103	270	431	83	75	1

Berdasarkan Tabel 2, ditemukan bahwa semakin besar nilai n akan menghasilkan jumlah sekuens unik yang semakin banyak, namun akan memperkecil jumlah sekuens marker. Apabila sekuens marker yang dihasilkan berada pada posisi yang berurutan, maka kumpulan sekuens tersebut dapat direkonstruksi ulang dalam bentuk urutan

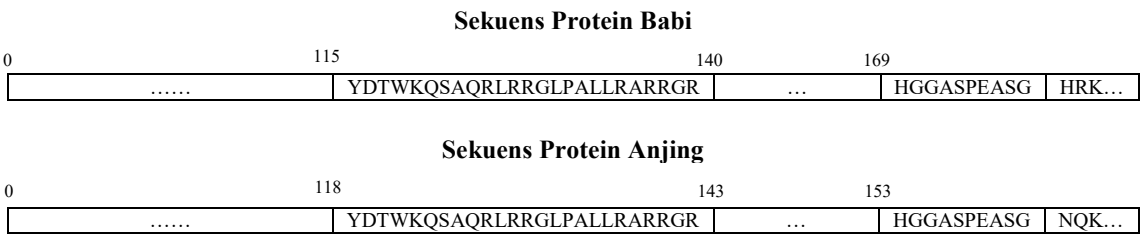
aslinya. Marker non-halal yang berhasil ditemukan selanjutnya akan dijelaskan lebih detail pada Tabel 3.

Marker Non-Halal

Tabel 3
Sekuens Marker Non-Halal

N	Marker non-halal	Posisi awal marker pada protein babi	Posisi awal marker pada protein anjing
8	YDTWKQSAQRLRR GASPEASG	[115-118], 147, 171 [170-173], 202, 226	[118-119], 172, 249, 273 154, [173-174], 227
9	YDTWKQSAQQRLRRG GGASPEASG	[115-118], 147, 171 [169-172], 201, 225	[118-119], 172, 249, 273 153, [172-173], 226
10	YDTWKQSAQRLRRGL HGGASPEASG	[115-118], 147, 171 [169-172], 201, 225	[118-119], 172, 249, 273 153, [172-173], 226
11	YDTWKQSAQRLRRGLP		
12	YDTWKQSAQRLRRGLPA		
13	YDTWKQSAQRLRRGLPALLR		
14	YDTWKQSAQRLRRGLPALLRA	[115-118], 147, 171	[118-119], 172, 249, 273
15	YDTWKQSAQRLRRGLPALLRAR		
16	YDTWKQSAQRLRRGLPALLRARR		
17-25	YDTWKQSAQRLRRGLPALLRARRGR		

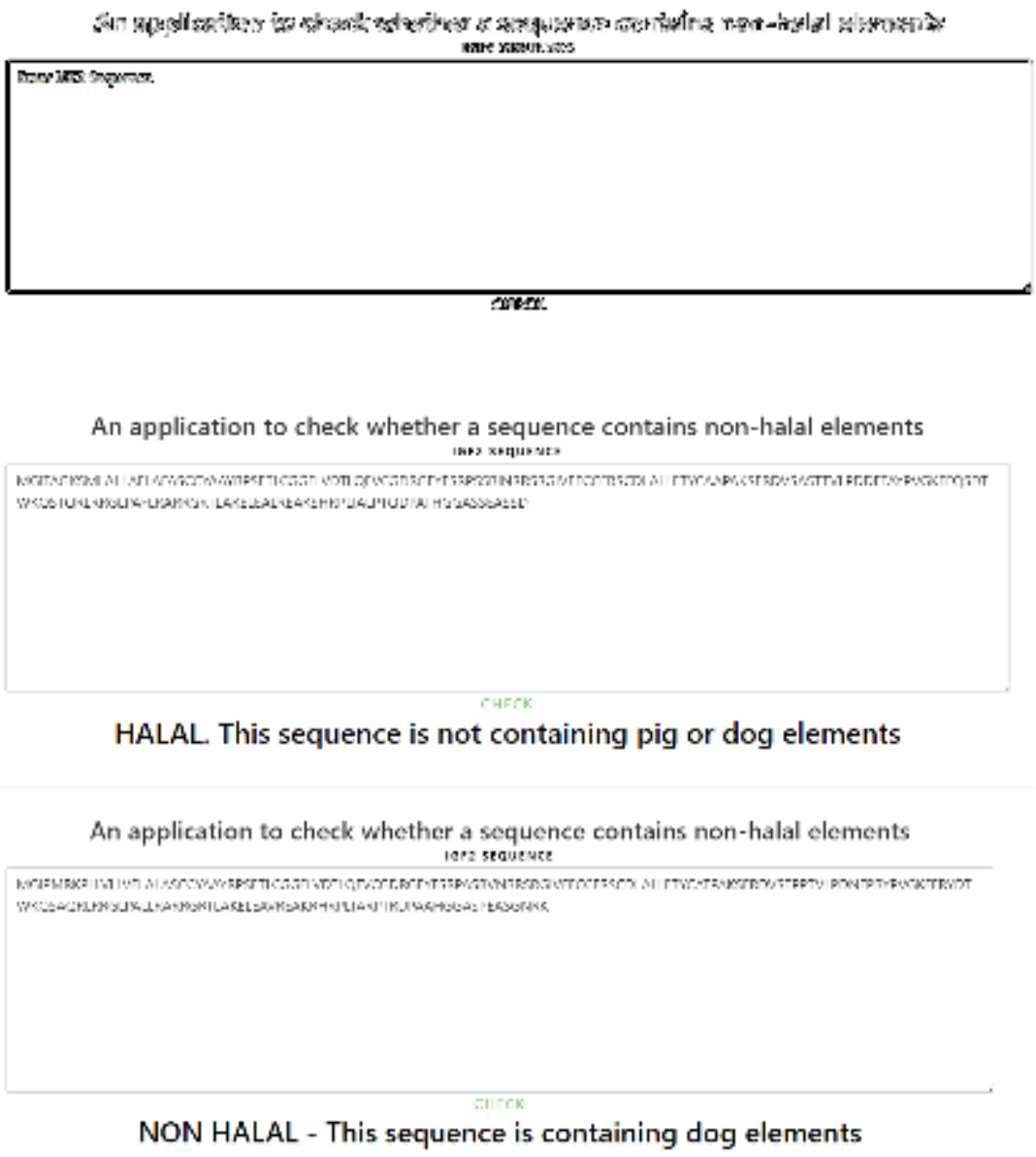
Tabel 3 menunjukkan bahwa terdapat dua blok marker non-halal yang ditemukan pada sekuens protein anjing dan babi. Blok marker pertama 'YDTW....RGR' ditemukan sejak n bernilai paling kecil hingga mengalami titik jenuh pada nilai n=17. Akan tetapi blok marker kedua HGGASPEASG mulai ditemukan saat nilai n kecil hingga maksimum n=10. Hal ini menandakan bahwa karakter sebelum dan setelah HGGASPEASG bukan merupakan penciri pada protein anjing dan babi. Posisi marker pada keseluruhan sekuens berbeda-beda tergantung pada panjang sekuens protein. Akan tetapi posisi blok marker YDTWKQSAQRLRRGLPALLRARRGR selalu muncul lebih dahulu daripada blok marker HGGASPEASG. Ilustrasi letak marker non-halal akan ditampilkan pada Gambar 2.



Gambar 2. Ilustrasi letak marker non-halal

Series: Engineering and Science Vol. 7 No. 1 (2021) E-ISSN: 2621-9794, P-ISSN: 2477-2097

Hasil temuan marker non-halal di atas menjadi dasar pembuatan aplikasi berbasis web untuk menentukan apakah sebuah sekuens protein IGF2 mengandung elemen halal atau tidak. Gambar 3 akan menunjukkan tampilan aplikasi web pendeteksi kandungan halal dan haram. Aplikasi tersebut dapat diakses pada https://halal-authentication-detector.netlify.app/ .



Gambar 3. Simulasi Aplikasi Pendeteksi Autentikasi Halal

SIMPULAN

Analisis n-gram berhasil diterapkan dalam mendeteksi marker penciri kandungan non-halal pada struktur protein (IGF2) pada hewan. Terdapat dua blok marker penciri yang ditemukan dalam struktur protein hewan babi dan anjing, yaitu YDTWKQSAQRLRRGLPALLRARRGR dan HGGASPEASG. Untuk penelitian lebih lanjut diharapkan dapat menerapkan analisis n-gram untuk mencari marker non-halal pada struktur biologi lainnya selain sekuens IGF2 yang dapat digunakan juga untuk proses autentikasi makanan halal dan non-halal.

DAFTAR PUSTAKA

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., … Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

He, Z., & Yang, H. (2018). Colourimetric detection of swine-specific DNA for halal authentication using gold nanoparticles. *Food Control*, *88*, 9–14. https://doi.org/10.1016/j.foodcont.2018.01.001

Hidayatullah, M. S. (2020). Sertifikasi dan Labelisasi Halal Pada Makanan dalam Perspektif Hukum Islam (Perspektif Ayat Ahkam). *YUDISIA : Jurnal Pemikiran Hukum Dan Hukum Islam*, *11*(2), 251. https://doi.org/10.21043/yudisia.v11i2.8620

Islam, S. M. A., Heil, B. J., Kearney, C. M., & Baker, E. J. (2018). Protein classification using modified n-grams and skip-grams. *Bioinformatics*, *34*(9), 1481–1487. https://doi.org/10.1093/bioinformatics/btx823

Premanandh, J., & Salem, S. Bin. (2017). Progress and challenges associated with halal authentication of consumer packaged goods. *Journal of The Science of Food and Agriculture*, *97*, 4672–4678. https://doi.org/https://doi.org/10.1002/jsfa.8481

Xu, L., Liang, G., Liao, C., Chen, G. Den, & Chang, C. C. (2019). K-SkIP-N-Gram-RF: A random forest based method for Alzheimer's disease protein identification. *Frontiers in Genetics*, *10*(FEB), 1–7. https://doi.org/10.3389/fgene.2019.00033

Yusuf, Z. K. (2010). Polymerase Chain Reaction (PCR). *Saintek*, *5*(6). https://doi.org/10.1016/B978-0-12-801238-3.08997-2