Penggunaan Penyilangan Aritmatika dalam Algoritma Genetika untuk Desain Primer

Polymerase Chain Reaction

Dinda Eling K. Sasmito1, Zainudin Zukhri2, Wayan T. Artama3

1,2 Magister Teknik Informatikam Universitas Islam Indonesia

3 Fakultas Peternakan Hewan, Universitas Gadjah Mada

1 dindaelingsasmito@gmail.com

2 zainudin.zukhri@gmail.com

3 artama@ugm.ac.id

**Abstract.** Penelitian ini melanjutkan penelitian terdahulu mengenai Bioinformatika pada desain primer Polymerase Chain Reaction. Sebagian dari parameter ini dijadikan sebagai representasi kromosom pada Algoritma Genetika. Parameter utama yang digunakan sebagai representasi kromosom yaitu panjang primer forward, panjang primer reverse, indeks awal primer forward dan panjang produk. Penelitian ini mengembangkan penelitian terdahulu dalam mendesain primer PCR menggunakan algoritma genetika. Pengembangan yang dilakukan adalah pada proses penyilangan, menggunakan penyilangan aritmatika. Percobaan dilakukan terhadap populasi awal yang disilangkan menggunakan teknik penyilangan terdahulu dan penyilangan aritmatika yang diusulkan. Hasil penelitian menunjukkan bahwa penyilangan aritmatika memiliki nilai fitness yang lebih baik dibandingkan populasi awal dan teknik penyilangan terdahulu.

**Keywords:** desain primer, PCR, *polymerase chain reaction*, algoritma genetika, penyilangan aritmatika, *arithmetic* *crossover*

1. Pendahuluan
	1. *Polymerase Chain Reaction* (PCR)

Reaksi berantai polimerase (Polimerase Chain Reaction, PCR) adalah suatu metode enzimatis untuk melipatkgandakan secara eksponensial suatu sekuen nukleotida tertentu secara in vitro. Metode ini pertama kali dikembangkan pada tahun 1985 oleh Kary B. Mullis, seorang peneliti di perusahaan CETUS Corporation. Metode ini sekarang telah banyak digunakan untuk berbagai macam manipulasi dan analisis genetik 4.

Urutan sekuen DNA pada primer yang optimal dan konstentrasi primer yang tepat sangat penting untuk spesifisitas maksimal dan efisiensi PCR 5. Keberhasilan PCR sangat tergantung pada desain primer 5,6. Beberapa karakteristik primer menjadi parameter dalam desain primer seperti telah dipaparkan pada penelitian sebelumnya 1.

* 1. Penelitian Terdahulu

Dalam beberapa dekade terakhir, berbagai pendekatan untuk desain primer PCR telah diusulkan. Banyak program baik *online* maupun *offline* dikembangkan untuk desain primer ini. Pengembangan sistem GeneFisher pertama diusulkan oleh Meyer, berfungsi untuk mendesain primer dari urutan DNA yang tidak diketahui7. Pada 2004, sebuah penelitian mengusulkan desain primer menggunakan Algoritma Genetika2, yang kemudian dijadikan acuan untuk dikembangkan pada penelitian ini. Algoritma serupa juga digunakan kembali untuk permasalahan desain primer yang berbeda di tahun 20053.

Pada dasarnya, Algoritma Genetika merupakan metode pencarian yang didasarkan pada proses evolusi alamiah8. Algoritma genetika tidak hanya mengevaluasi solusi tunggal dari permasalahan, melainkan menyeleksi sekaligus sejumlah penyelesaian masalah yang mungkin. Menurut Michalewicz pada tahun 1998, keberhasilan penggunaan Algoritma Genetika sangat ditentukan oleh penentuan pernyataan masalah ke dalam bentuk titik-titik pencarian yang disebut dengan kromosom, serta pemilihan operator-operator yang digunakan8. Penelitian ini menggunakan teknik operator genetika yang berbeda dari penelitian terdahulu. Nilai fitness tiap kromosom yang dihasilkan akan dibandingkan, dengan harapan teknik operator genetika yang diusulkan mampu menghasilkan kromosom dengan nilai fitness yang lebih baik.

1. Materi dan Metode

Sebagai perbandingan yang valid, gen CYP1A dipilih karena telah digunakan dalam penelitian-penelitian sebelumnya9. Gen ini dikaitkan dengan peningkatan risiko kanker paru-paru10. Penelitian ini akan memodifikasi teknik penyilangan dalam Algoritma Genetika yang digunakan untuk desain primer PCR.

* 1. Representasi Kromosom

Pada penelitian Wu dkk2, sebuah sekuen DNA berupa kode asam basa nukleat didefinisikan sebagai , sedangkan basa komplemennya didefinisikan sebagai 2. Sebagai contoh, di bawah ini adalah sekuen DNA berupa kode asam basa nukleat dan basa komplemennya.

Sebuah individu pada algoritma genetika dilambangkan sebagai sepasang primer, yang disajikan sebagai vektor, dan ditulis sebagai Persamaan (1)

 (1)

dimana:

Fs : Indeks awal primer forward

Fl : Panjang primer forward

Pl : Panjang produk yang dihasilkan diantara primer forward dan primer reverse

Rl : Panjang primer reverse

Pada penelitian ini, nilai panjang primer forward (Fl) dan panjang primer reverse (Rl) yang diperbolehkan adalah pada rentang 15 sampai dengan 35. Indeks awal forward primer (Fs) merupakan nilai random antara 0 dan selisih panjang template DNA dikurangi panjang produk maksimum, yang ditetapkan pada penelitian ini sebagai 300 basa. Panjang produk yang dihasilkan (Pl) merupakan nilai random antara panjang produk minimum (pada penelitian ini ditetapkan sebagai 150 basa) dan panjang produk maksimum (pada penelitian ini ditetapkan sebagai 300 basa). Sebagai contoh, sebuah individu yang dinotasikan dengan berarti memiliki indeks awal primer forward 140 dan panjang primer forward 20, sehingga menghasilkan primer forward ATGTCGGCCACGGAGTTTCT dari gen CYP1A. Individu tersebut memiliki panjang produk 200 dan panjang primer reverse 21, artinya indeks awal primer reverse berada pada indeks 319, dan diakhiri pada indeks 340 dan merupakan komplementer dari basa pada sekuen DNA nya, sehingga menghasilkan primer reverse ACGTCCCCATACTGCTGGCTC dari gen CYP1A. Dengan demikian, solusi yang diperoleh menggunakan representasi kromosom satu vektor berupa sepasang primer untuk sebuah sekuen DNA. Ilustrasi representasi kromosom pada contoh di atas dapat dilihat melalui Gambar 1.



**Gambar 1.** Ilustrasi Representasi Kromosom pada gen CYP1A

* 1. Evaluasi Kromosom

Penelitian terdahulu telah mengemukakan parameter yang digunakan untuk menilai primer yang dihasilkan1,2. Penelitian ini menambahkan parameter selisih panjang primer, *Missmatch, Repetition*. Dari parameter tersebut, diberikan batasan untuk menentukan nilai fitnessnya sebagai berikut.

1. Panjang Primer:

Panjang primer optimum untuk amplifikasi DNA dalam beberapa penelitian berada pada rentang 18-22 basa nukleotida6, rentang 18-26 basa nukleotida2, dan rentang 16-28 basa nukleotida9. Nilai panjang primer pada penelitian ini berada pada rentang 15 hingga 35 basa, dengan nilai optimum mengikuti penelitian Yang dkk (2008) yakni 16-28 basa nukleotida. Fungsi panjang primer yang digunakan untuk menentukan nilai panjang primer pada fungsi objektif dapat dilihat pada Persamaan (2).

 (2)

1. Perulangan: *Repetation*

Perulangan yang tidak boleh terjadi pada primer adalah perulangan lebih dari tiga basa berurutan sama (contoh: AGCGGGGGATG memiliki 5 basa berurutan G). Perulangan lain yang tidak boleh terjadi adalah perulangan 2 basa yang terjadi sebanyak lebih dari 4 kali (contoh: ATATATAT, masih dapat ditoleransi). Fungsi repetation yang digunakan untuk memberi nilai pada fungsi objektif dapat dilihat pada Persamaan (3).

(3)

1. Selisih panjang primer: *Lendiff*

Selisih antara panjang forward primer dan reverse primer sebaiknya tidak lebih dari 3 basa 2,9. Fungsi selisih panjang primer yang digunakan pada fungsi objektif dapat dilihat pada Persamaan (4).

 (4)

1. GCcontent

Fungsi GC Content yang digunakan untuk dimasukkan dalam fungsi objektif seperti pada Persamaan (5).

 (5)

1. GCclamp

Fungsi GC Clamp yang digunakan untuk dimasukkan dalam fungsi objektif seperti pada Persamaan (6).

 (6)

1. Suhu leleh: *Tm*

Fungsi suhu leleh yang digunakan untuk dimasukkan dalam fungsi objektif seperti pada Persamaan (7).

 (7)

1. Selisih suhu leleh: *Tmdiff*

Selisih antara suhu leleh forward primer dan reverse primer sebaiknya tidak lebih dari 5°C 2,9. Fungsi selisih suhu leleh yang digunakan pada fungsi objektif dapat dilihat pada Persamaan (8).

 (8)

1. *Hairpin*

Fungsi hairpin yang digunakan untuk dimasukkan dalam fungsi objektif seperti pada Persamaan (9).

 (9)

1. *Dimer*

Fungsi dimer yang digunakan untuk dimasukkan dalam fungsi objektif seperti pada Persamaan (10).

 (10)

1. *Specivicity*

Fungsi specificity yang digunakan untuk dimasukkan dalam fungsi objektif seperti pada Persamaan (11).

 (11)

1. *Missmatch*

Fungsi missmatch yang digunakan untuk dimasukkan dalam fungsi objektif seperti pada Persamaan (12).

 (12)

Bobot setiap parameter primer pada fungsi objektif didasarkan pada tingkat kepentingan masing-masing menurut pengguna sistem yang akan dirancang dinamis pada sistem yang akan dibangun. Nilai default bobot pada fungsi objektif di atas didasarkan pada hasil diskusi dengan pakar. Bobot 1, 3, 10 dan 50 masing-masing menunjukkan tingkat kepentingan dengan 50 sebagai batasan desain terpenting, diikuti dengan bobot 10, 3, kemudian 1 sebagai bobot terkecil. Nilai objektif ditentukan menggunakan Persamaan (13).

 (13)

Fungsi fitness yang digunakan harus dipetakan dari fungsi objektifnya. Cara yang digunakan pada penelitian ini adalah dengan memetakan setiap penambahan nilai fungsi objektif sebagai penurunan nilai fungsi fitness pada Persamaan (14).

 (14)

* 1. Penyilangan

Proses penyilangan pada penelitian Wu dkk menggunakan metode penyilangan *n* titik2. Untuk individu yang terpilih disilangkan, dibangkitkan nilai random *R* dimana *0<R<16*. Nilai R kemudian dikonversi ke biner untuk menentukan titik mana yang terkena penyilangan. Sebagai contoh, nilai random *R*=11(10) dikonversi ke biner menjadi 1011(2). Berdasarkan biner yang dihasilkan, maka titik penyilangan adalah titik pertama, ke-tiga dan ke-empat. Nilai *offspring* yang dihasilkan adalah penyilangan nilai induk dari masing-masing titik tersebut.

Pada penelitian ini, proses penyilangan menggunakan metode penyilangan aritmatika setelah terlebih dahulu terpilih titik yang disilangkan, menggunakan teknik yang sama seperti penelitian sebelumnya. Pada penyilangan aritmatika, nilai *offspring* yang dihasilkan tidak langsung ditukarkan dari induk1 dan induk2, melainkan terlebih dahulu dibangkitkan bilangan random *r* dimana 0<*r<*1. Nilai *offspring* yang dihasilkan akan mengikuti Persamaan (15) dan Persamaan (16).

 (15) (16)

Contoh ilustrasi penyilangan kedua teknik dapat dilihat pada gambar 2 berikut.



**Gambar 2.** Proses penyilangan, (a) Teknik terdahulu, (b) Teknik yang diusulkan

1. Hasil dan Diskusi

Penelitian ini menitikberatkan pada modifikasi teknik penyilangan dari penelitian terdahulu2. Percobaan dilakukan pada sebuah populasi genetika berupa solusi desain primer dengan ukuran populasi 20 individu, dengan vektor sepasang primer dengan nilai random berdasarkan Persamaan (1). Untuk melihat proses penyilangan menggunakan teknik penyilangan pada penelitian terdahulu dan teknik penyilangan aritmatika, disajikan 4 sampel data pada Tabel 1. Grafik perbandingan populasi awal sebelum penyilangan, setelah penyilangan menggunakan teknik terdahulu dan setelah penyilangan menggunakan penyilangan aritmatika dapat dilihat pada Gambar 3.

Pada Gambar 3 terlihat bahwa penyilangan menggunakan teknik pada penelitian terdahulu lebih banyak menghasilkan *offspring* dengan nilai fitness yang lebih rendah dibandingkan induknya (populasi awal), sementara menggunakan penyilangan aritmatika lebih banyak menghasilkan *offspring* yang memiliki nilai fitness lebih tinggi.

**Tabel 1**. Proses penyilangan menggunakan teknik terdahulu dan penyilangan aritmatika

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *individu ke-* | *Fs* | *Fl* | *Pl* | *Rl* | *Fitness* | *Nilai random r* |
| Data sampel populasi awal |
| 1 | 507 | 33 | 289 | 22 | 1.754385965 | - |
| 2 | 991 | 30 | 259 | 34 | 1.470588235 | - |
| 3 | 1510 | 26 | 218 | 18 | 2.325581395 | - |
| 4 | 1057 | 25 | 205 | 23 | 2.127659574 | - |
| Data sampel penyilangan menggunakan teknik pada penelitian terdahulu |
| 1 | 991 | 33 | 259 | 34 | 1.515151515 | - |
| 2 | 507 | 30 | 289 | 22 | 1.886792453 | - |
| 3 | 1510 | 25 | 218 | 23 | 2.5 | - |
| 4 | 1057 | 26 | 205 | 18 | 3.03030303 | - |
| Data sampel penyilangan menggunakan teknik penyilangan aritmatika |
| 1 | 846 | 33 | 268 | 30 | 1.538461538 | *r* = 0.3 |
| 2 | 652 | 30 | 280 | 26 | 2 |
| 3 | 1510 | 25 | 218 | 22 | 2 | *r* = 0.2 |
| 4 | 1057 | 26 | 205 | 19 | 3.333333333 |



**Gambar 3.** Perbandingan nilai fitness

Rata-rata nilai fitness untuk populasi awal, hasil penyilangan menggunakan teknik penyilangan terdahulu dan hasil penyilangan menggunakan teknik penyilangan aritmatika pada penelitian ini secara terurut adalah 1.8701, 1.940, dan 2.1382. Nilai ini membuktikan bahwa nilai fitness yang dihasilkan menggunakan teknik penyilangan aritmatika memiliki rata-rata nilai fitness yang lebih baik dari populasi awal dan teknik penyilangan terdahulu. Harapannya, menggunakan teknik penyilangan aritmatika, populasi yang dihasilkan akan memiliki nilai fitness yang lebih tinggi pada setiap generasinya.

1. Kesimpulan

Algoritma genetika dapat digunakan untuk desain primer pada PCR menggunakan representasi kromosom berupa vector yang menyimpan nilai parameter-parameter primer berupa indeks awal primer forward, panjang prmer forward, panjang produk dan panjang primer reverse. Pengujian membuktikan bahwa operator penyilangan aritmetika yang diusulkan dalam penelitian ini mampu menghasilkan nilai fitness yang lebih baik dibandingkan populasi awal dan teknik penyilangan pada penelitian terdahulu. Selanjutnya, penelitian ini dapat dilanjutkan dengan pengembangan sistem secara keseluruhan untuk desain primer menggunakan Algoritma Genetika yang termasuk proses evaluasi, seleksi, penyilangan dan mutasi. Penelitian lain yang dapat dilakukan terkait parameter genetika (ukuran populasi, maksimum generasi, probabilitas penyilangan dan probabilitas mutasi) yang tepat juga dapat dilakukan untuk memperoleh hasil yang optimum, serta perbedaan fungsi objektif yang digunakan serta pemberian nilai bobot pada fungsi objektif tersebut.

Referensi

1. Sasmito DEK, Kurniawan R, Muhimmah I. Karakteristik Primer pada Polymerase Chain Reaction(PCR) untuk Sekuensing DNA: Mini Review. *Semin Nas Inform Medis 2014*. 2014:93-102. http://snimed.fit.uii.ac.id/.

2. Wu J-S, Lee C, Wu C-C, Shiue Y-L. Primer design using genetic algorithm. *Bioinformatics*. 2004;20(11):1710-1717. doi:10.1093/bioinformatics/bth147.

3. Lin F-M, Huang H-D, Huang H-Y, Horng J-T. Primer design for multiplex PCR using a genetic algorithm. *Proc 2005 Conf Genet Evol Comput - GECCO ’05*. 2005:475. doi:10.1145/1068009.1068087.

4. Yuwono T. *Teori Dan Aplikasi Polymerase Chain Reaction*. Andi Publisher; 2007.

5. Abd-elsalam KA. Minireview Bioinformatic tools and guideline for PCR primer design. 2003;2(May):91-95.

6. Burpo FJ. A critical review of PCR primer design algorithms and cross- hybridization case study. 2001:1-12.

7. Amoozegar M, Rezvannejad E. Primer design using gravitational search algorithm. *2014 Iran Conf Intell Syst*. February 2014:1-6. doi:10.1109/IranianCIS.2014.6802608.

8. Zukhri Z. *Algoritma Genetika - Metode Komputasi Evolusioner Untuk Menyelesaikan Masalah Optimasi*. Penerbit Andi; 2014.

9. Yang C-H, Cheng Y-H, Chang H-W, Chuang L-Y. Primer design with specific PCR product size using Memetic algorithm. *2008 IEEE Conf Soft Comput Ind Appl*. June 2008:332-337. doi:10.1109/SMCIA.2008.5045985.

10. National Center for Biotechnology Information. Homo sapiens cytochrome P450 family 1 subfamily A member 1 (CYP1A1). http://www.ncbi.nlm.nih.gov/nuccore/189339226?report=fasta. Accessed January 13, 2016.