

EVALUASI ALGORITMA GENETIKA BERBASIS MULTI-SEED UNTUK OPTIMASI PRIMER PCR

MULTI-SEED EVALUATION OF GENETIC ALGORITHM FOR PCR PRIMER OPTIMIZATION

¹Dinda Eling Kartikaning Sasmito, ²Fajar Suryani

¹Program Studi Informatika, Universitas Mercu Buana Yogyakarta, Indonesia

²Program Studi Teknik Informatika, Universitas Duta Bangsa Surakarta, Indonesia

dinda@mercubuana-yogya.ac.id, fajar_suryani@udb.ac.id

Received:
11 Juni 2026

Revised:
27 Juni 2026

Accepted:
31 July 2026

Published:
4 August 2026

ABSTRAK

Desain primer merupakan tahap penting dalam *Polymerase Chain Reaction* (PCR) karena kualitas primer memengaruhi spesifisitas, efisiensi amplifikasi, dan keberhasilan analisis molekuler. Penelitian ini bertujuan mengevaluasi *robustness*, konvergensi, dan efisiensi *pipeline* Algoritma Genetika (GA) untuk optimasi desain primer PCR melalui replikasi *multi-seed*. Kandidat primer direpresentasikan sebagai kromosom yang memuat indeks awal primer *forward*, panjang primer *forward*, panjang produk PCR, dan panjang primer *reverse*. Fungsi objektif berdasarkan akumulasi penalti dari pelanggaran parameter desain primer. *Objective optimum* didefinisikan sebagai kandidat tanpa pelanggaran terhadap kriteria internal yang digunakan. Dengan panjang sekuens target 2.608 bp, ruang pencarian kasar mencapai 153.758.619 kemungkinan kandidat sehingga pencarian *exhaustive* tidak efisien. Evaluasi dilakukan pada 20 replikasi dalam dua set *seed* menggunakan populasi 100 individu, maksimum 300 generasi, probabilitas *crossover* 0,8, dan probabilitas mutasi 0,1. Kinerja GA dianalisis melalui *best fitness*, *median fitness*, *diversity top-K* berbasis *normalized Hamming distance*, serta perbandingan dengan *random search baseline* pada *budget* evaluasi maksimum yang sama. Hasil menunjukkan GA mencapai *objective optimum* pada 19 dari 20 replikasi (95%). Seluruh replikasi pada *seed set* kedua mencapai *objective optimum* dengan rata-rata 580 evaluasi kandidat, sedangkan *random search* memerlukan rata-rata 10.980 evaluasi kandidat. Hasil ini menunjukkan bahwa GA memberikan pencarian yang lebih efisien, terarah, dan stabil terhadap variasi *seed*.

Kata Kunci : desain primer, PCR, algoritma genetika, *robustness*, konvergensi

ABSTRACT

Primer design is an important stage in the *Polymerase Chain Reaction* (PCR) because primer quality affects specificity, amplification efficiency, and the success of molecular analysis. This study evaluates the robustness, convergence, and efficiency of a Genetic Algorithm (GA) pipeline for PCR primer design optimization through multi-seed replication. Primer candidates were represented as chromosomes consisting of the forward primer start index, forward primer length, PCR product length, and reverse primer length. The objective function was constructed from accumulated penalties for violations of primer design parameters. The objective optimum was defined as a candidate without violations of the internal criteria used. With a target sequence length of 2,608 bp, the approximate search space reached 153,758,619 possible candidates, making exhaustive search inefficient. The evaluation involved 20 replications across two seed sets using a population size of 100 individuals, a maximum of 300 generations, a crossover probability of 0.8, and a mutation probability of 0.1. GA performance was analyzed using best fitness, median fitness, top-K diversity based on normalized Hamming distance, and comparison with a random search baseline under the same maximum evaluation budget. The results showed that GA achieved the objective optimum in 19 of 20 replications (95%). All replications in the second seed set reached the objective optimum within approximately 580 candidate evaluations, whereas random search required approximately 10,980 candidate evaluations on average. These findings indicate that GA provides more efficient, directed, and stable search performance across seed variations.

Keywords : primer design, PCR, genetic algorithm, robustness, convergence

PENDAHULUAN

Polymerase Chain Reaction (PCR) merupakan salah satu teknik penting dalam biologi molekuler yang digunakan untuk memperbanyak fragmen DNA tertentu secara *in vitro*. Keberhasilan PCR sangat dipengaruhi oleh kualitas primer yang digunakan. Primer yang kurang sesuai dapat menyebabkan amplifikasi non-spesifik, pembentukan *primer-dimer*, struktur *hairpin*, rendahnya efisiensi amplifikasi, atau bahkan kegagalan reaksi PCR. Oleh karena itu, proses desain primer perlu mempertimbangkan beberapa karakteristik penting, seperti panjang primer, *melting temperature* (T_m), selisih T_m antar primer, *GC content*, *GC clamp*, potensi *hairpin*, *self-dimer*, *cross-dimer*, *specificity*, serta panjang produk PCR [1].

Dalam tiga tahun terakhir, penelitian desain primer komputasional terus berkembang. Beberapa penelitian berfokus pada integrasi Primer3 dan pengembangan *pipeline* otomatis untuk desain primer, validasi spesifisitas, serta *in silico* PCR [2]–[8]. Penelitian lain mulai mengarah pada evaluasi kualitas primer berbasis *machine learning* dan pendekatan optimasi untuk meningkatkan efisiensi pemilihan kandidat primer [9]–[15]. Perkembangan tersebut menunjukkan bahwa desain primer tidak lagi hanya dipandang sebagai proses pemilihan berdasarkan aturan dasar, tetapi juga sebagai proses komputasional yang melibatkan pencarian kandidat, evaluasi kualitas, dan optimasi berdasarkan banyak kriteria.

Salah satu pendekatan yang dapat digunakan untuk masalah optimasi desain primer adalah Algoritma Genetika atau *Genetic Algorithm* (GA). GA merupakan metode optimasi berbasis populasi yang terinspirasi dari proses evolusi biologis. Secara umum, GA bekerja melalui tahapan pembentukan populasi awal, evaluasi *fitness*, seleksi, *crossover*, mutasi, dan elitisme. Dalam konteks desain primer PCR, satu kandidat solusi dapat direpresentasikan sebagai kombinasi posisi primer, panjang primer, dan panjang produk PCR. Setiap kandidat kemudian dievaluasi berdasarkan fungsi objektif atau *fitness* yang disusun dari parameter desain primer. Karena ruang pencarian kandidat primer dapat sangat besar, GA dapat digunakan untuk mengeksplorasi solusi secara lebih terarah dibandingkan pencarian *exhaustive* [16].

Sebagai metode optimasi stokastik, kinerja GA dapat dipengaruhi oleh variasi populasi awal dan *seed* random. Oleh karena itu, evaluasi GA tidak hanya perlu mempertimbangkan kualitas solusi akhir, tetapi juga konsistensi hasil antar-replikasi, pola konvergensi, dan perubahan *diversity* populasi. *Robustness* menunjukkan kestabilan kinerja algoritma pada kondisi awal yang berbeda, sedangkan konvergensi menggambarkan kecenderungan populasi menuju solusi dengan nilai *objective* lebih rendah atau *fitness* lebih tinggi. *Diversity* berperan dalam menjaga eksplorasi ruang solusi dan mengurangi risiko *premature convergence* [16], [17].

Penelitian terdahulu telah menerapkan GA untuk desain primer PCR menggunakan representasi kromosom berbasis empat komponen, yaitu indeks awal primer *forward*, panjang primer *forward*, panjang produk PCR, dan panjang primer *reverse* [18]. Pendekatan optimasi primer berbasis GA juga telah dikembangkan untuk memilih kandidat primer dengan mempertimbangkan beberapa kriteria secara simultan [19]. Namun, penelitian tersebut belum secara khusus mengevaluasi *robustness* antar-*seed*, pola konvergensi, perubahan *diversity* kandidat, serta efisiensi pencarian dibandingkan *random search baseline*.

Berdasarkan celah tersebut, penelitian ini mengevaluasi *robustness*, konvergensi, dan efisiensi *pipeline* GA untuk optimasi desain primer PCR melalui replikasi *multi-seed*. Evaluasi dilakukan menggunakan dua set *seed*, *best-of-run*, tren *best fitness* dan *median fitness*, *diversity top-K* berbasis *normalized Hamming distance*, serta *random search baseline* dengan *budget* evaluasi maksimum yang sama. Dengan demikian, kontribusi utama penelitian ini bukan hanya menghasilkan kandidat primer dengan *fitness* tinggi, tetapi juga menilai stabilitas, efisiensi, dan keterarahan proses pencarian GA terhadap variasi *seed*.

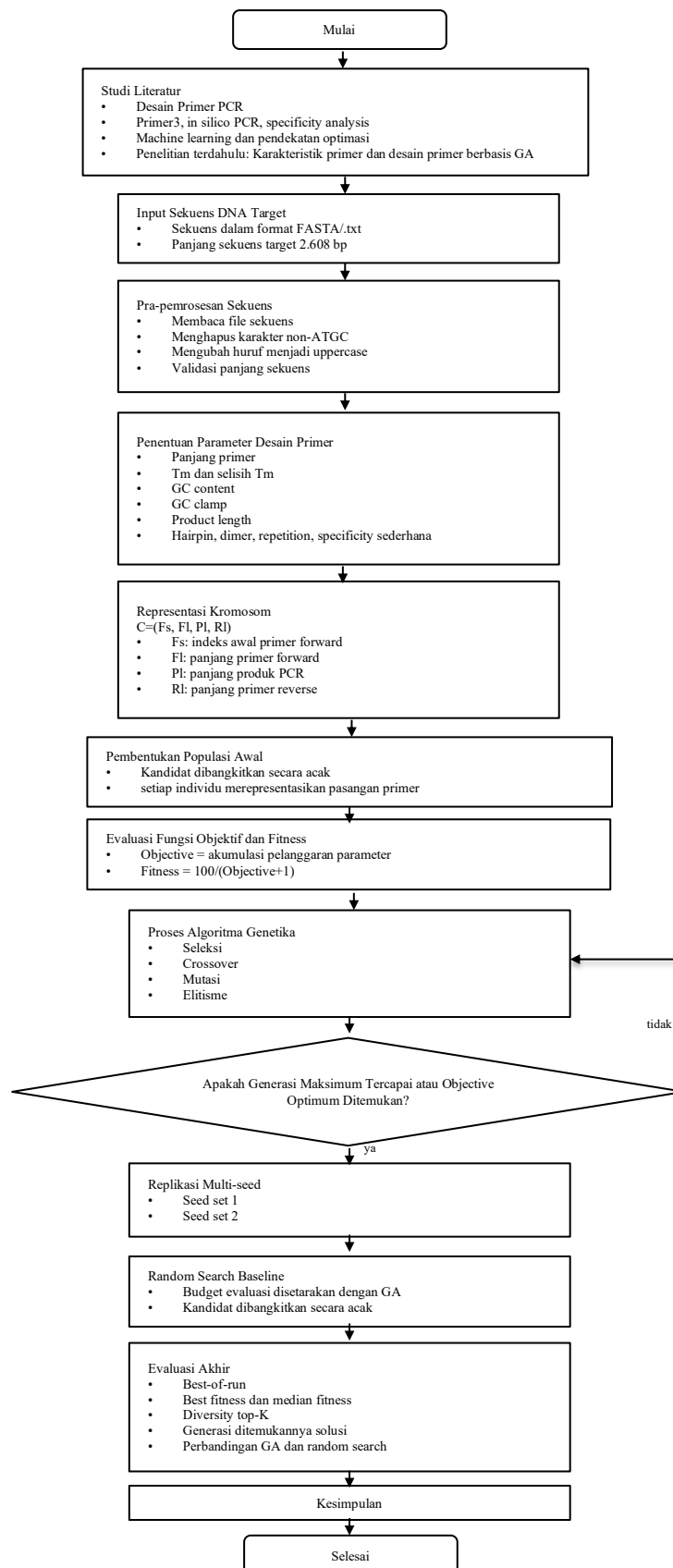
Rumusan masalah penelitian ini adalah: (1) apakah *pipeline* GA mampu menghasilkan kandidat primer dengan *objective optimum* secara konsisten pada *seed* berbeda; (2) bagaimana stabilitas hasil optimasi pada dua set *seed*; (3) bagaimana pola konvergensi berdasarkan *best fitness* dan *median fitness*; (4) bagaimana perubahan *diversity* kandidat *top-K*; dan (5) bagaimana performa GA dibandingkan *random search baseline* dengan *budget* evaluasi maksimum yang sama.

METODE

Penelitian ini merupakan penelitian komputasional eksperimental yang bertujuan mengevaluasi *robustness* dan konvergensi *pipeline* GA untuk optimasi desain primer PCR. Secara umum, penelitian dilakukan melalui beberapa tahap, yaitu studi literatur, penyiapan sekuens DNA target, pra-pemrosesan sekuens, penentuan parameter desain primer, penyusunan fungsi objektif dan *fitness*, implementasi GA, replikasi *multi-seed*, *random search baseline*, serta evaluasi hasil berdasarkan konvergensi dan *diversity*.

Studi literatur dilakukan terhadap penelitian desain primer komputasional terbaru, termasuk penelitian yang menggunakan integrasi Primer3, *in silico* PCR, analisis spesifisitas, *machine learning*, dan pendekatan optimasi [2]–[15]. Selain itu, penelitian ini juga mengacu pada penelitian sebelumnya mengenai karakteristik primer PCR [1] dan desain primer menggunakan GA [18]. Karakteristik primer dari penelitian sebelumnya digunakan sebagai dasar penyusunan parameter evaluasi, sedangkan pendekatan representasi

kromosom empat komponen digunakan sebagai dasar pengembangan pipeline GA. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

Dataset Sekuens Target dan Ruang Pencarian

Penelitian ini menggunakan satu sekuens DNA target sebagai template desain primer. Sekuens disimpan dalam format teks sederhana atau FASTA, kemudian dibersihkan sehingga hanya menyisakan karakter A, T, G, dan C. Panjang sekuens target yang digunakan adalah 2.608 bp. Penelitian ini berfokus pada evaluasi pipeline optimasi, sehingga validasi biologis spesifik terhadap organisme atau gen tertentu belum menjadi fokus utama.

Ruang pencarian kandidat primer dihitung berdasarkan representasi kromosom yang digunakan. Setiap individu dibentuk dari kombinasi empat komponen, yaitu indeks awal primer *forward* (Fs), panjang primer *forward* (Fl), panjang produk PCR (Pl), dan panjang primer *reverse* (Rl). Pada implementasi ini, Fs memiliki 2.309 kemungkinan posisi, Fl memiliki 21 kemungkinan nilai, Pl memiliki 151 kemungkinan nilai, dan Rl memiliki 21 kemungkinan nilai. Dengan demikian, jumlah kandidat yang mungkin terbentuk secara kasar adalah sebagai berikut:

Jumlah individu = jumlah Fs x jumlah Fl x jumlah Pl x jumlah Rl = $2.309 \times 21 \times 151 \times 21 = 153.758.619$ individu.

Jumlah tersebut menunjukkan bahwa pencarian *exhaustive* terhadap seluruh kemungkinan individu tidak efisien. Oleh karena itu, pendekatan optimasi seperti GA dan *baseline random search* digunakan untuk mengevaluasi strategi pencarian kandidat primer.

Representasi Kromosom

Representasi kromosom mengikuti pendekatan penelitian desain primer PCR berbasis GA yang telah dilakukan sebelumnya [1], [18]. Setiap individu direpresentasikan sebagai vektor empat komponen:

$$C = (Fs, Fl, Pl, Rl)$$

dengan *Fs* adalah indeks awal primer *forward* pada sekuens target, *Fl* adalah panjang primer *forward*, *Pl* adalah panjang produk PCR, dan *Rl* adalah panjang primer *reverse*. Primer *forward* diperoleh dari potongan sekuens target mulai posisi *Fs* sepanjang *Fl*. Posisi primer *reverse* dihitung dari posisi awal primer *forward* dan panjang produk PCR, kemudian primer *reverse* diperoleh dari *reverse-complement* potongan sekuens pada posisi tersebut. Dengan representasi ini, satu kromosom menghasilkan satu pasangan primer *forward* dan *reverse* serta satu ukuran produk PCR. Selain itu, representasi ini memungkinkan operator genetika seperti *crossover* dan mutasi bekerja langsung pada komponen numerik kromosom.

Parameter Desain Primer

Parameter desain primer yang digunakan mengacu pada karakteristik primer yang umum digunakan dalam penelitian desain primer PCR [1]. Parameter tersebut digunakan untuk menilai apakah suatu kandidat primer memenuhi kriteria desain primer yang diharapkan, dan dapat dilihat pada Tabel 1.

Tabel 1. Parameter Desain Primer

Parameter	Nilai
Panjang primer minimum	18 nt
Panjang primer maksimum	26 nt
Selisih panjang primer maksimum	3 nt
Tm minimum	52°C
Tm maksimum	62°C
Selisih Tm maksimum	5°C
GC content minimum	40%
GC content maksimum	60%
Panjang produk minimum	150 bp
Panjang produk maksimum	300 bp
GC clamp	Ya
Batas repetition	4
Skor hairpin maksimum	4
Skor self-dimer maksimum	8
Skor cross-dimer maksimum	8
Specificity sederhana	Ya

Pada tahap ini, penghitungan beberapa karakteristik seperti hairpin dan dimer dilakukan menggunakan pendekatan heuristik sederhana. Dengan demikian, hasil evaluasi primer dalam penelitian ini diposisikan sebagai penilaian komputasional awal berdasarkan fungsi fitness internal, bukan sebagai validasi termodinamika final. Validasi lebih lanjut dapat dilakukan menggunakan perangkat seperti Primer3, Primer-BLAST atau *in silico* PCR sebagaimana digunakan dalam penelitian desain primer komputasional [2]-[8].

Fungsi Objektif dan Fitness

Fungsi objektif dibangun sebagai akumulasi penalti terhadap pelanggaran parameter desain primer. Jika kandidat melanggar batas atau kondisi yang ditentukan, maka kandidat tersebut memperoleh penalti. Semakin besar penalti yang diterima, semakin besar nilai objektif dan semakin rendah kualitas kandidat.

Kondisi yang diberi penalti meliputi panjang primer di luar batas, selisih panjang primer yang terlalu besar, T_m di luar rentang, selisih T_m terlalu tinggi, GC content di luar rentang, tidak terdapat GC *clamp*, *repetition* berlebih, potensi *hairpin*, potensi *self-dimer*, potensi *cross-dimer*, dan kemunculan primer lebih dari satu kali pada *template*.

Nilai fitness dihitung menggunakan rumus $\text{fitness} = 100 / (\text{objective} + 1)$. Dengan formulasi tersebut, kandidat dengan *objective* lebih rendah akan memiliki fitness lebih tinggi. Kandidat ideal memiliki *objective* 0 dan fitness 100. Pada penelitian ini, kondisi tersebut disebut sebagai *objective optimum*, yaitu kondisi ketika kandidat primer tidak memiliki pelanggaran terhadap parameter desain primer yang digunakan.

Algoritma Genetika dan Replikasi Multi-Seed

GA dijalankan melalui tahapan inisialisasi populasi, evaluasi fitness, seleksi, crossover, mutasi, dan elitisme. Populasi awal dibentuk secara acak berdasarkan batas nilai kromosom yang valid. Setiap individu dalam populasi dievaluasi menggunakan fungsi objektif dan fitness yang telah ditentukan. Individu dengan nilai fitness lebih tinggi memiliki peluang lebih besar untuk dipertahankan atau dipilih sebagai induk.

Seleksi digunakan untuk memilih individu yang akan menjadi induk. Crossover digunakan untuk menghasilkan offspring dari dua induk dengan menggabungkan sebagian komponen kromosom. Mutasi digunakan untuk mengubah sebagian nilai kromosom, misalnya posisi primer, panjang primer, atau panjang produk PCR. Elitisme digunakan untuk mempertahankan individu terbaik agar tidak hilang pada generasi berikutnya.

Pada implementasi ini, seleksi dilakukan menggunakan *roulette-wheel selection* berdasarkan nilai fitness, sehingga individu dengan fitness lebih tinggi memiliki peluang lebih besar untuk dipilih sebagai induk. Crossover diterapkan menggunakan *arithmetic crossover* pada komponen kromosom yang terpilih secara acak. Setiap kromosom direpresentasikan oleh empat gen, yaitu Fs, Fl, Pl, dan Rl. Mutasi dilakukan pada setiap gen dengan probabilitas mutasi yang telah ditentukan melalui pembangkitan kembali nilai gen dalam batas ruang pencarian internal yang valid. Elitisme diterapkan dengan mempertahankan kandidat terbaik global pada generasi berikutnya melalui penggantian kandidat dengan fitness terendah apabila diperlukan.

Karena GA merupakan metode stokastik, hasil optimasi dapat dipengaruhi oleh *seed* random. Oleh karena itu, penelitian ini menggunakan replikasi *multi-seed* untuk mengevaluasi kestabilan hasil. Dua set *seed* digunakan dalam penelitian ini. *Seed set* pertama digunakan untuk mengamati stabilitas awal *pipeline*. Sedangkan *seed set* kedua digunakan untuk menguji konsistensi hasil pada konfigurasi *seed* berbeda tanpa mengubah parameter GA maupun parameter desain primer. Pada setiap *seed*, data yang dicatat meliputi kandidat primer terbaik, nilai *objective*, nilai *fitness*, generasi saat *objective optimum* ditemukan, tren *best fitness* per generasi, median *fitness* per generasi, serta *diversity* kandidat *top-K*.

Tabel 2. Parameter Algoritma Genetika

Parameter	Nilai
Ukuran populasi	100
Jumlah generasi maksimum	300
Probabilitas crossover	0,8
Probabilitas mutasi	0,1
Jumlah replikasi	20 (dua set seed)
Seed set 1	835, 852, 772, 503, 833, 820, 297, 863, 186, 490
Seed set 2	101, 202, 303, 404, 505, 606, 707, 808, 909, 1001
Top-K untuk diversity	10

Random Search Baseline

Random search baseline digunakan sebagai pembandingan sederhana terhadap GA. Pada metode ini, kandidat primer dibangkitkan secara acak tanpa seleksi, *crossover*, mutasi, maupun elitisme. Setiap kandidat dievaluasi menggunakan fungsi objektif dan fitness yang sama dengan GA, kemudian kandidat terbaik dipilih sebagai hasil *random search*.

Budget random search disamakan dengan *budget* evaluasi maksimum GA, yaitu $\text{population_size} \times \text{max_generation}$. Dengan ukuran populasi 100 dan jumlah generasi maksimum 300, *random search* mengevaluasi 30.000 kandidat acak per *seed*. Penggunaan *budget* yang sama dimaksudkan agar perbandingan dilakukan pada jumlah evaluasi maksimum yang setara. *Random search* digunakan untuk

menilai apakah kandidat *objective optimum* dapat ditemukan hanya melalui sampling acak dalam jumlah besar, serta untuk melihat apakah GA memberikan keuntungan dari sisi efisiensi dan keterarahan pencarian.

Evaluasi Konvergensi dan Diversity

Konvergensi dianalisis berdasarkan perubahan *best fitness* dan *median fitness* per generasi. *Best fitness* menunjukkan performa kandidat terbaik pada setiap generasi, sedangkan *median fitness* menggambarkan kualitas populasi secara umum. *Median fitness* digunakan karena *best fitness* saja dapat memberi gambaran yang terlalu optimistis terhadap performa populasi.

Diversity kandidat dihitung menggunakan rata-rata *normalized Hamming distance* antar kandidat *top-K*. Setiap kandidat direpresentasikan sebagai gabungan primer *forward* dan *reverse*. Jika panjang kandidat berbeda, *string* diselaraskan menggunakan *padding* sebelum jarak dihitung. Nilai *diversity* yang rendah menunjukkan bahwa kandidat *top-K* semakin mirip. Penurunan *diversity* yang disertai peningkatan *fitness* dapat dibaca sebagai konvergensi produktif, sedangkan penurunan *diversity* yang disertai stagnasi *fitness* dapat mengindikasikan *premature convergence* [16,17].

Indikator evaluasi yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3. Dengan evaluasi tersebut, penelitian ini tidak hanya menilai kualitas kandidat primer akhir, tetapi juga menilai perilaku proses optimasi GA, termasuk stabilitas antar *seed*, efisiensi pencarian, pola perbaikan populasi, dan kecenderungan konvergensi.

Tabel 3. Indikator Evaluasi

Aspek Evaluasi	Indikator
<i>Robustness</i>	Jumlah <i>seed</i> yang mencapai <i>objective optimum</i> , <i>mean fitness</i> , <i>median fitness</i>
Konvergensi	<i>Best fitness</i> dan <i>median fitness</i> per generasi
Efisiensi	Generasi atau jumlah evaluasi saat <i>objective optimum</i> ditemukan
<i>Diversity</i>	Rata-rata <i>normalized Hamming distance</i> kandidat <i>top-K</i>
<i>Baseline</i>	Perbandingan GA dengan <i>random search</i> pada <i>budget</i> evaluasi yang sama

HASIL DAN PEMBAHASAN

Evaluasi Dua Set Seed

Evaluasi dilakukan pada dua set *seed*. *Seed set* pertama digunakan untuk mengamati stabilitas awal *pipeline*, sedangkan *seed set* kedua digunakan untuk menguji konsistensi hasil pada konfigurasi *seed* berbeda. Pada setiap *seed*, GA dijalankan dengan parameter yang sama, yaitu ukuran populasi 100, jumlah generasi maksimum 300, probabilitas *crossover* 0,8, dan probabilitas mutasi 0,1. Ringkasan hasil ditunjukkan pada Tabel 4.

Tabel 4. Ringkasan Hasil GA pada Dua Set Seed

Eksperimen	Jumlah <i>seed</i>	<i>Objective optimum</i>	<i>Mean objective</i>	<i>Mean fitness</i>	<i>Median fitness</i>	<i>Mean generasi objective optimum</i> ditemukan
<i>Seed set 1</i>	10	9/10	0,10	95,00	100,00	12,00
<i>Seed set 2</i>	10	10/10	0,00	100,00	100,00	4,80
Gabungan	20	19/20	0,05	97,50	100,00	8,40

Pada *seed set* pertama, GA mencapai *objective optimum* pada 9 dari 10 *seed*. Satu *seed* menghasilkan *objective* 1 dan *fitness* 50, yang menunjukkan bahwa kandidat terbaik masih memiliki satu pelanggaran terhadap parameter *Tm*. Hasil tersebut mengindikasikan bahwa proses pencarian pada *seed* tersebut belum sepenuhnya mencapai kondisi optimum dalam batas generasi yang digunakan. Sebaliknya, pada *seed set* kedua, seluruh *seed* mencapai *objective optimum* dengan *fitness* 100. Secara gabungan, GA mencapai *objective optimum* pada 19 dari 20 replikasi (95%).

Hasil tersebut menunjukkan bahwa *pipeline* GA relatif *robust* terhadap variasi *seed*. *Seed set* kedua juga menunjukkan bahwa ketika konfigurasi *seed* diubah tanpa mengubah parameter algoritma, *pipeline* tetap mampu menghasilkan kandidat primer dengan *objective optimum* pada seluruh replikasi. Hal ini memperkuat indikasi bahwa *pipeline* tidak hanya berhasil pada satu kombinasi *seed* tertentu.

Karakteristik Hasil pada Seed Set Kedua

Seed set kedua digunakan sebagai uji konsistensi hasil pada konfigurasi *seed* berbeda. Seluruh *seed* pada *set* ini menghasilkan *objective optimum* dan *fitness* 100. Kandidat primer terbaik memiliki panjang produk antara 171 bp hingga 280 bp, dengan rata-rata 234,4 bp. *Tm* primer *forward* berada pada rentang 56-62°C, sedangkan *Tm* primer *reverse* berada pada rentang 54-60°C. *GC content* primer *forward* berada pada rentang 40,000-55,556%, sedangkan *GC content* primer *reverse* berada pada rentang 47,368-57,895%.

Seluruh kandidat pada *seed set* kedua memenuhi fungsi *fitness* yang ditentukan. Rata-rata generasi ditemukannya solusi adalah 4,8 dengan median 5. Dengan ukuran populasi 100, jumlah evaluasi kandidat hingga *objective optimum* ditemukan dapat diperkirakan sekitar 580 kandidat. Hal ini menunjukkan bahwa

GA mampu menemukan kandidat valid dengan jumlah evaluasi yang relatif lebih kecil dibandingkan *budget* maksimum 30.000 kandidat per seed.

Tabel 5. Best-of-run GA pada Seed Set Kedua

Seed	Forward primer	Reverse primer	Prod uk	Tm F	Tm R	GC F	GC R	Obj	Fit	Gen
101	TGAAGAAC TCTCTGGAA GC	GGAAGTGT GTCCTAGA CC	228	56	56	47, 37	55, 56	0	100	3
202	GCGACTTTA ACCCCTACA GG	TAGGGATG AAGTCAGC TGG	171	62	58	55	52, 63	0	100	5
303	ATCAACCAT GACCAGAA GC	AGCCAGGA AGAGAAAG ACC	177	56	58	47, 368	52, 63	0	100	7
404	GCTTTTACA TCCCCAAG GG	GTCATGTCC ACCTTCAC G	280	58	56	52, 632	55, 56	0	100	7
505	ATCAACCAT GACCAGAA GC	GAATTCCA CCCGTTGC AGC	201	56	60	47, 368	57, 89	0	100	5
606	TGTCATCTG TGCCATTTG C	AGGCTGTC TGTGATGTC CC	267	56	60	47, 368	57, 89	0	100	9
707	TCAAGGAG CACTACAA AACC	TGGATCTTT CTCTGTACC C	235	58	56	45	47, 37	0	100	5
808	TCTTCCTGG CTATCCTGC	CAGATAGC AAAAGTGC AGC	268	56	56	55, 556	47, 367	0	100	3
909	TGGAACCTT CCCTGATCC	CATGTGCC CAATCAGA GGC	254	56	60	55, 556	57, 89	0	100	0
1001	ATGTATTTG GTGATGAA CCC	TGGTCATG GTTGATCTG C	263	56	54	40	50	0	100	4

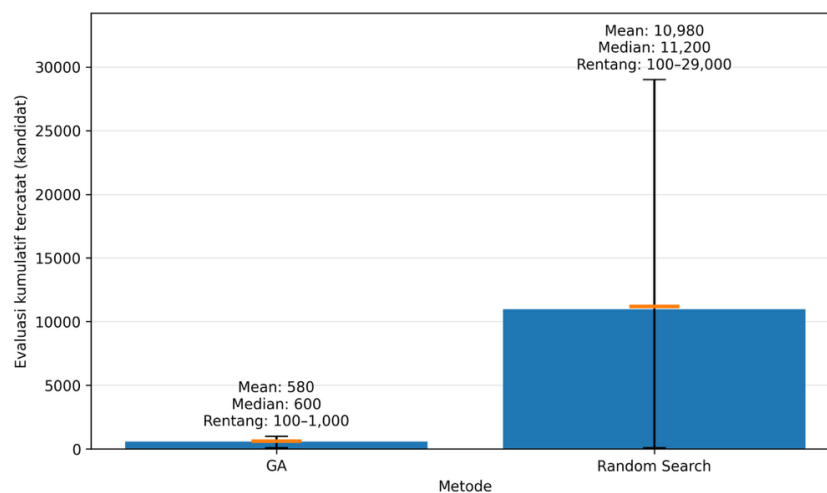
Perbandingan GA dengan *Random Search Baseline*

Untuk memastikan perbandingan dilakukan pada kondisi acak yang setara, GA dan *random search* dijalankan menggunakan *seed set* kedua yang sama, yaitu 101, 202, 303, 404, 505, 606, 707, 808, 909, dan 1001. Kedua metode menggunakan fungsi *objective*, parameter desain primer, serta *budget* maksimum yang sama. *Random search* tidak menggunakan mekanisme seleksi, *crossover*, mutasi, maupun elitisme. Ringkasan hasil perbandingan ditunjukkan pada Tabel 6.

Tabel 6. Perbandingan GA dan *Random Search* pada Seed Set Kedua

Metode	Jumlah seed	Objective optimum	Mean objective	Mean fitness	Mean evaluasi tercatat	Median evaluasi	SD evaluasi	Rentang evaluasi
GA	10	10/10	0,00	100,00	580	600	252,98	100–1.000
Random search	10	10/10	0,00	100,00	10.980	11.200	9.951,42	100–29.000

Tabel 6 menunjukkan bahwa GA dan *random search* sama-sama mencapai *objective optimum* pada seluruh 10 *seed*. Oleh karena itu, kualitas *objective* akhir tidak membedakan kedua metode pada konfigurasi sekuens target dan fungsi *objective* yang digunakan. Perbedaan utama terlihat pada jumlah evaluasi hingga *objective optimum* pertama kali tercatat. GA mencapai *objective optimum* pada rata-rata 580 evaluasi kandidat, dengan median 600 dan rentang 100–1.000 evaluasi. Sebaliknya, *random search* memerlukan rata-rata 10.980 evaluasi kandidat, dengan median 11.200 dan rentang 100–29.000 evaluasi. Dengan demikian, GA mencapai kondisi optimum dengan jumlah evaluasi yang secara substansial lebih rendah dibandingkan *random search*.



Gambar 2. Perbandingan evaluasi GA dan *random search*

Tren Konvergensi

Tren *best fitness* dan *median fitness* digunakan untuk mengamati perubahan kualitas kandidat sepanjang generasi. Untuk menjaga konsistensi agregasi lintas generasi, nilai metrik dari *seed* yang telah mencapai *objective optimum* dan berhenti lebih awal dibawa tetap hingga generasi terakhir pengamatan. Dengan demikian, nilai pada setiap generasi dihitung berdasarkan seluruh 10 *seed* pada *seed set* kedua. Ringkasan tren konvergensi ditunjukkan pada Tabel 7.

Tabel 7. Konvergensi Seed Set Kedua dengan Agregasi Carry-Forward

Generasi	Jumlah seed dalam agregasi	Mean best fitness	Mean median fitness	Mean best objective	Mean diversity top-K
0	10	19,022	1,639	11,700	0,791
1	10	19,857	1,752	9,800	0,696
2	10	24,160	2,092	5,800	0,620
3	10	42,143	2,476	3,600	0,601
5	10	76,429	3,622	1,200	0,574
9	10	100,000	4,409	0,000	0,594

Tabel 7 menunjukkan bahwa *mean best fitness* meningkat dari 19,022 pada generasi ke-0 menjadi 100,000 pada generasi ke-9. Sebaliknya, *mean best objective* menurun dari 11,700 menjadi 0,000. *Mean median fitness* juga meningkat dari 1,639 menjadi 4,409, meskipun peningkatannya lebih lambat dibandingkan *mean best fitness*. Pola tersebut menunjukkan bahwa kandidat terbaik dapat diperbaiki lebih cepat daripada kualitas populasi secara keseluruhan. Sementara itu, *mean diversity top-K* menurun dari 0,791 menjadi 0,594, yang menunjukkan bahwa kandidat terbaik semakin terkonsentrasi pada wilayah solusi tertentu selama proses evolusi.

Diversity Kandidat Top-K

Diversity *top-K* digunakan untuk melihat apakah kandidat terbaik tetap beragam atau semakin terkonsentrasi pada pola sekuens tertentu. Pada *seed set* pertama, diversity *top-K* menurun dari 0,787 pada generasi 0 menjadi 0,246 pada generasi 10. Pada *seed set* kedua, diversity *top-K* menurun dari 0,791 pada generasi ke-0 menjadi 0,594 pada generasi ke-9 berdasarkan agregasi carry-forward.

Penurunan *diversity* menunjukkan bahwa kandidat *top-K* semakin mirip seiring proses optimasi. Pada *seed set* kedua, penurunan tersebut dapat dibaca sebagai konvergensi produktif karena seluruh *seed* mencapai *objective optimum*. Pada *seed set* pertama, terdapat satu *seed* yang hanya mencapai *objective* 1. Kondisi ini mengindikasikan bahwa konvergensi cepat belum selalu menjamin ditemukannya *objective optimum* pada semua kondisi stokastik. Oleh karena itu, mekanisme eksplorasi seperti pengaturan *mutation rate*, *restart*, atau strategi mempertahankan *diversity* dapat dipertimbangkan pada pengembangan berikutnya.

Efisiensi dan *Robustness* Algoritma Genetika

Hasil penelitian menunjukkan bahwa *pipeline* GA mampu menghasilkan kandidat primer dengan kualitas tinggi pada sebagian besar replikasi. Secara gabungan, GA mencapai *objective optimum* pada 19 dari 20 *seed* (95%). Hal ini menunjukkan bahwa *pipeline* relatif stabil terhadap variasi *seed*, meskipun masih terdapat satu kasus pada *seed set* pertama yang berhenti pada *objective* 1.

Perbandingan dengan *random search baseline* menunjukkan bahwa kandidat *objective optimum* juga dapat ditemukan melalui *sampling* acak apabila budget evaluasi cukup besar. GA tidak selalu lebih unggul dari *random search* dalam fitness akhir. Kontribusi utama GA adalah pada efisiensi dan keterarahan pencarian. Pada *seed set* kedua, GA menemukan *objective optimum* pada rata-rata generasi ke-4,8 atau sekitar 580 evaluasi kandidat, sedangkan *random search* membutuhkan rata-rata sekitar 10.980 evaluasi kandidat.

Dari sisi konvergensi, GA menunjukkan pola perbaikan melalui peningkatan *best fitness* dan penurunan *best objective* pada generasi awal. *Median fitness* meningkat lebih lambat dibandingkan *best fitness*, yang menunjukkan bahwa kandidat terbaik ditemukan lebih cepat daripada perbaikan kualitas populasi secara keseluruhan. Sementara itu, penurunan *diversity top-K* menunjukkan bahwa kandidat terbaik semakin mirip selama proses evolusi. Pada *seed set* kedua, penurunan *diversity* disertai keberhasilan seluruh *seed* mencapai *objective optimum*, sehingga dapat dibaca sebagai konvergensi produktif. Namun, pada *seed set* pertama masih terdapat satu *seed* yang belum mencapai *objective optimum*, sehingga penurunan *diversity* tetap perlu diwaspadai sebagai kemungkinan awal *premature convergence* [16,17].

Penelitian terdahulu telah menunjukkan bahwa algoritma genetika dapat digunakan untuk mencari kandidat primer melalui representasi kromosom dan operator evolusioner [18], sedangkan GADPO menunjukkan pengembangan optimasi primer berbasis algoritma genetika dengan mempertimbangkan beberapa kriteria secara simultan [19]. Penelitian ini tidak membandingkan nilai fitness secara langsung dengan penelitian tersebut karena target sekuens, fungsi *objective*, dan batas parameter yang digunakan berbeda. Kontribusi penelitian ini terletak pada evaluasi perilaku *pipeline* secara eksperimental melalui dua *seed set*, analisis konvergensi dan *diversity top-K*, serta perbandingan efisiensi dengan *random search* pada *seed* yang sama.

Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan, yaitu evaluasi yang masih menggunakan satu sekuens target sehingga hasil belum dapat digeneralisasi untuk seluruh jenis template DNA. Pengujian pada beberapa sekuens target dengan panjang dan karakteristik berbeda diperlukan pada penelitian berikutnya.

Selain itu, fungsi fitness masih berbasis penalti heuristik. Evaluasi hairpin dan dimer belum menggunakan pendekatan termodinamika yang rinci, begitu pula dengan *specificity* yang belum menggunakan pemeriksaan terhadap basis data eksternal, misalnya Primer-BLAST. Oleh karena itu, hasil penelitian ini masih memerlukan validasi biologis eksperimental.

KESIMPULAN

Penelitian ini mengevaluasi *robustness*, konvergensi, dan efisiensi *pipeline* Algoritma Genetika (GA) untuk optimasi desain primer PCR melalui replikasi *multi-seed*. Kandidat primer direpresentasikan dalam kromosom empat komponen dan dievaluasi menggunakan fungsi *objective* berbasis penalti parameter desain primer. Pada ruang pencarian kasar sebesar 153.758.619 kandidat, GA mencapai *objective optimum* pada 19 dari 20 replikasi (95%). Seluruh *seed* pada *seed set* kedua mencapai *objective optimum* dengan rata-rata 580 evaluasi kandidat, sedangkan *random search* memerlukan rata-rata 10.980 evaluasi kandidat pada *seed* yang sama. Hasil ini menunjukkan bahwa GA memberikan pencarian yang lebih efisien, terarah, dan relatif stabil terhadap variasi *seed*. Analisis konvergensi menunjukkan perbaikan kualitas kandidat pada generasi awal, sedangkan *diversity top-K* menunjukkan kandidat terbaik semakin terkonsentrasi selama proses evolusi. Pipeline ini dapat digunakan sebagai tahap penyaringan kandidat primer awal sebelum evaluasi lebih lanjut menggunakan Primer3, Primer-BLAST, in silico PCR, dan validasi eksperimental. Penelitian berikutnya dapat menguji lebih banyak sekuens target serta mengintegrasikan evaluasi termodinamika dan spesifisitas yang lebih rinci.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Informatika, Universitas Mercu Buana Yogyakarta, atas dukungan pendanaan hibah penelitian dalam pelaksanaan penelitian dan penyusunan artikel ini.

DAFTAR PUSTAKA

- Al-Mahrami, N., Al Yazidi, S., Alrashdi, H., Al-Jahwari, Z., Al-Mahrizi, M., Yousif, A. A., Al-Aamri, H., Al-Rahbi, B., & Al Jahdhami, N. (2026). PrimerAST: A predictive machine learning tool for primer design and quality assessment. *Scientific Reports*, 16, 14980. <https://doi.org/10.1038/s41598-026-38238-8>
- Filomena, E., Picardi, E., Pesole, G., & D'Erchia, A. M. (2025). IsoPrimer: A pipeline for designing isoform-aware primer pairs for comprehensive gene expression quantification. *Bioinformatics Advances*, 5(1), vbaf171. <https://doi.org/10.1093/bioadv/vbaf171>
- Fulghum, B., Tanker, S. H., & White, R. A., III. (2024). DeGenPrime provides robust primer design and optimization unlocking the biosphere. *Bioinformatics Advances*, 4(1), vbae044. <https://doi.org/10.1093/bioadv/vbae044>
- Guo, J., Grojean, J., & Guo, H. (2025). ARMSprimer3: An open-source primer design Python program for amplification refractory mutation system PCR (ARMS-PCR). *Journal of Pathology Informatics*, 17, 100442. <https://doi.org/10.1016/j.jpi.2025.100442>
- He, W., Zhuo, Y., Wang, C., Huang, Y., Zang, X., Yang, C., Deng, H., Zhou, Y., Liu, J., Zhang, P., Xue, X., & Zhang, L. (2025). Genome-wide primer scan (GPS): A Python package for a flexible, reliable and large-scale primer design toolkit. *Frontiers of Computer Science*, 19, 192906. <https://doi.org/10.1007/s11704-024-40392-z>
- Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: Past, present, and future. *Multimedia Tools and Applications*, 80(5), 8091–8126. <https://doi.org/10.1007/s11042-020-10139-6>
- Kuzdraliński, A. (2025). PrimeSpecPCR: Python toolkit for species-specific DNA primer design and specificity testing. *SoftwareX*, 31, 102249. <https://doi.org/10.1016/j.softx.2025.102249>
- Molloy, T. G., Banerjee, A., Balayan, A., Robbins, J. K., & Engler, A. J. (2026). Primer PICKR: Literature-mined scoring platform for robust RT-qPCR primers. *Nature Communications*. <https://doi.org/10.1038/s41467-026-73648-2>
- Monfort-Lanzas, P., Rusu, E. C., Parrakova, L., Karg, C. A., & Hackl, H. (2024). ExonSurfer: A web-tool to design primers at exon-exon junctions. *BMC Genomics*, 25, 594. <https://doi.org/10.1186/s12864-024-10456-2>
- Nagi, S. C., Ashraf, F., Miles, A., & Donnelly, M. J. (2024). AnoPrimer: Primer design in malaria vectors informed by range-wide genomic variation. *Wellcome Open Research*, 9, 255. <https://doi.org/10.12688/wellcomeopenres.20998.1>
- Pitsch, J. W., Wirth, S. A., Costantino, N. T., Mejia, J., Doss, R. M., Warren, A. V. A., Ustanik, J., Yang, X., & Breuss, M. W. (2025). CREPE (CREate Primers and Evaluate): A computational tool for large-scale primer design and specificity analysis. *Genes*, 16(9), 1062. <https://doi.org/10.3390/genes16091062>
- Rodríguez-Bejarano, F. M., Vega-Rodríguez, M. A., & Santander-Jiménez, S. (2024). GADPO: Genetic algorithm based on dominance for primer optimization. *Expert Systems with Applications*, 238, 122206. <https://doi.org/10.1016/j.eswa.2023.122206>
- Sasmito, D. E. K., Kurniawan, R., & Muhimmah, I. (2014). Karakteristik primer pada polymerase chain reaction (PCR) untuk sekuensing DNA: Mini review. *Seminar Nasional Informatika Medis (SNIMed)*, 93–102.
- Sasmito, D. E. K., Zuhri, Z., & Artama, W. T. (2016). Penggunaan penyilangan aritmatika dalam algoritma genetika untuk desain primer polymerase chain reaction. *Seminar Nasional Informatika Medis (SNIMed) VII*, 1–8.
- Soliman, M. A., Azab, M. S., Hussein, H. A., Roushdy, M. M., & Abu el-Naga, M. N. (2024). FBPP: Software to design PCR primers and probes for nucleic acid base detection of foodborne pathogens. *Scientific Reports*, 14, 1229. <https://doi.org/10.1038/s41598-024-51372-5>
- Sudholt, D. (2020). The benefits of population diversity in evolutionary algorithms: A survey of rigorous runtime analyses. In B. Doerr & F. Neumann (Eds.), *Theory of evolutionary computation: Recent developments in discrete optimization* (pp. 359–404). Springer. https://doi.org/10.1007/978-3-030-29414-4_8
- Vázquez-González, L., Regueira-Iglesias, A., Balsa-Castro, A., Vila-Blanco, C., Tomás, I., & Carreira, M. J. (2024). PrimerEvalPy: A tool for *in-silico* evaluation of primers for targeting the microbiome. *BMC Bioinformatics*, 25(1), 157. <https://doi.org/10.1186/s12859-024-05805-7>
- Wirth, J. S., Katz, L. S., Williams, G. M., & Chen, J. C. (2024). primerForge: A Python program for identifying primer pairs capable of distinguishing groups of genomes from each other. *Journal of Open Source Software*, 9(101), 6850. <https://doi.org/10.21105/joss.06850>
- Zhang, J., Yang, J., & Wen, C. (2025). EasyKASP: A simple and fast tool for KASP primer design. *BMC Bioinformatics*, 26, 292. <https://doi.org/10.1186/s12859-025-06322-x>