Article

Not peer-reviewed version

FalseAmpHunter: A Bioinformatics Pipeline for Detecting and Characterizing False Amplicons in PCR

Muhammad Shoaib Akhtar *, Mian Sahib Zar, Abdul Rehman Haris, Samia Tahir

Posted Date: 27 April 2026

doi: 10.20944/preprints202604.1800.v1

Keywords: false amplicon; off-target amplification; paralogous sequences; polymerase chain reaction; primers

 doi:10.20944/preprints202604.1800.v1

Article

FalseAmpHunter: A Bioinformatics Pipeline for Detecting and Characterizing False Amplicons in PCR

Muhammad Shoaib Akhtar [1,*], Mian Sahib Zar [2], Abdul Rehman Haris [3] and Samia Tahir [4]

[1] Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

[2] School of Biology and Basic Medical Sciences, Soochow University, Suzhou, China

[3] Department of Microbiology and Molecular Genetics, Allama Iqbal Open University, Islamabad, Pakistan

[4] Department of Food and Nutritional Sciences, Faculty of Science and Technology, University of Central Punjab, Lahore, Pakistan

* Correspondence: xoaib@ymail.com

Abstract

Polymerase chain reaction (PCR) is a widely used molecular biology technique; however, it remains highly susceptible to non-specific primer binding, particularly in genomic regions with extensive sequence similarity. Such off-target amplification can generate false amplicons that are difficult to detect using conventional quality control methods and may lead to erroneous downstream interpretation. Here, we present FalseAmpHunter, a pipeline designed to detect, assemble, and characterize false amplicons from paired-end next-generation sequencing (NGS) data generated from PCR amplification. FalseAmpHunter reconstructs candidate amplicons, maps them genome-wide, evaluates primer-binding orientation, and distinguishes true target amplification from paralog-driven off-target products and sequencing artifacts. We validated FalseAmpHunter using a synthetic dataset derived from *in silico* PCR of paralogous olfactory receptor (OR) genes. The pipeline successfully identified both the intended target amplicon and false amplicons originating from paralogous loci, while excluding a random decoy control. FalseAmpHunter provides a systematic and transparent solution for investigating false PCR amplification events and is applicable to assay development, primer validation, and troubleshooting of targeted sequencing experiments. By transforming raw sequencing data into interpretable genomic evidence, it enhances confidence in PCR-based analyses, particularly in paralog-rich genomic contexts. The pipeline is accessible online at: https://github.com/xoaib4/FalseAmpHunter.

Keywords: false amplicon; off-target amplification; paralogous sequences; polymerase chain reaction; primers

1.. Introduction

PCR remains a foundational technique in molecular biology and widely used in basic biology, clinical diagnostics and forensics due to its sensitivity, speed, low input requirements and cost efficiency [1–7]. Despite its widespread use, PCR is inherently vulnerable to non-specific primer binding, which can lead to the amplification of unintended genomic regions [8,9]. Such false amplicons are particularly problematic when universal or degenerate primers are employed, or when assays are transferred across genomic contexts with high sequence similarity [10,11]. In many cases, these off-target amplification events remain undetected, especially when the amplicon size alone is used as a quality control metric [12]. Even if detected, it can be challenging to characterize false amplicons.

 doi:10.20944/preprints202604.1800.v1

Instead of yielding a single unintended product, off-target amplification usually generates a heterogeneous mixture of amplicons derived from multiple genomic loci, orientations, or paralogous regions [12]. Conventional quality control workflows, such as gel electrophoresis, melting curve analysis, or single-locus alignment, are designed to indicate off-target amplification but not to resolve it, leaving a critical gap between experimental observation and genomic interpretation [12–14]. While gel electrophoresis can indicate off-target amplification based on amplicon sizes, melting curve analysis indicates off-target amplification by variations in melting curve peaks. However, identifying a false amplicon is challenging in itself. Until a false amplicon is properly identified, a new more specific primer pair cannot be designed and deployed.

Given that off-target amplification typically yields a heterogeneous mixture of products derived from multiple genomic loci. Sanger sequencing of the bulk PCR product would produce overlapping, unreadable signals, making it unsuitable for resolving individual false amplicons [15]. In contrast, NGS sequences individual molecules independently, enabling the deconvolution of complex amplicon mixtures [16]. Although NGS library preparation and sequencing follow standard protocols, a dedicated bioinformatics workflow is required to identify false amplicons. Existing bioinformatics tools focus on read-level quality control (QC), de novo assembly, or reference-based alignment; however, they do not provide an integrated framework specifically designed to investigate false PCR amplification. Consequently, researchers are often forced to manually combine multiple tools and ad-hoc scripts to reconstruct false amplicons, assess primer-binding orientation, and identify genomic origins. Even if researchers can run tools manually, it is difficult for them to select the correct set of tools. Moreover, this process is time-consuming, error-prone, and difficult to reproduce, particularly for users without extensive bioinformatics expertise.

To address this gap, we developed FalseAmpHunter, an automated bioinformatics pipeline designed to identify, assemble, and characterize false amplicons from paired-end NGS read data generated from PCR products. FalseAmpHunter integrates read-level quality control, de novo assembly, reference-guided validation, and primer-centric interrogation into a single, reproducible workflow. The pipeline reconstructs candidate amplicons, maps them back to the genome, evaluates primer-binding orientation (sense vs. anti-sense), and reports coverage and alignment statistics to facilitate biological interpretation.

FalseAmpHunter is intended for applications in assay development, primer validation, and method troubleshooting, where understanding the genomic basis of off-target amplification is critical. By transforming raw sequencing data into interpretable genomic evidence, the pipeline provides a systematic approach to diagnosing off-target amplification events and improving primer design strategies. The pipeline is modular, transparent, and can be applied to both real and synthetic datasets, making it suitable for a wide range of experimental and computational settings. In this study, we utilized a synthetic sequencing dataset of highly similar paralogous genes to test the FalseAmpHunter pipeline.

2. Materials and Methods

2.1. Generation of False Amplicon

2.1.1. Targeted Sequences

Olfactory receptor (OR) gene family is the largest paralogous gene family in the human and mammalian genomes. This family has more than 800 OR intact and pseudogenes. We selected four intact OR genes from the same OR10G gene subfamily in the human genome. These genes included OR10G4, OR10G7, OR10G8 and OR10G9, and are paralogs with high sequence similarity [17]. The sequences of these paralogous genes are provided as supplementary file 1.

2.1.2. Primers

 doi:10.20944/preprints202604.1800.v1

We designed primers to amplify the OR10G4 gene, which has three other highly similar paralogs. OR genes are inherently difficult to amplify specifically due to high incidence of paralogs. The primers we designed for this amplification of OR10G4 are given in Table 1. The other three targeted genes, OR10G7, OR10G8, and OR10G9, were included to induce off-target amplification owing to their sequence similarity with OR10G4. The melting temperature (TM) of forward and reverse primers were 76.4°C and 55.1°C respectively. TMs were calculated assuming 50 mM salt and 50 nM annealing oligo concentration.

Table 1. Forward and reverse primer sequences of OR10G4 gene.

Orientation	Primer Sequence
Forward	5' ATGTCCAACGCCAGCCTCGTGACAGC 3'
Reverse	5' ACAGTGTAGAAAATGGCCAC 3'

2.1.3. PCR

After the primer sequences were designed, we used the *in silico* PCR tool of the University of California, Santa Cruz (UCSC; [18]) to amplify human DNA using the human reference genome assembly, hg38, under default conditions. This *in silico* PCR generated a heterogeneous mixture of amplicons.

2.2. Synthetic Sequencing of Amplicon

For the amplified product of UCSC's *in silico* PCR, we synthesized raw sequencing reads using the NGS simulator, ART (Q Version 2.5.8, released: June 6, 2016) for Illumina MiSeq version 3 [19]. We provided previously *in silico* amplified PCR product in a FASTA format and paired-end sequencing reads of 150 base pairs (bp) length for the total size of amplicon with standard deviation of 20 were generated at 2000x coverage. In addition, we also added primer sequence carryovers (5–30 bp) at the start of 5% sequencing reads and generated chimeric reads to make this synthetic dataset more representative of real false amplicon sequencing datasets using a custom script. As a negative control, we synthesized 1 KB long decoy dataset of random nucleotide bases and synthesized sequencing reads in the same manner as described above.

2.3. False Amp Hunter Tool

The FalseAmpHunter pipeline we designed to identify false amplicons included six steps including sequencing reads QC, assembly of QC-passed reads, alignment of sequencing reads to assembled sequence, BLAST assembled sequence against human reference genome (hg38 genome assembly in current paper), finding primer hits in BLAST hits, and generation of coverage report for aligned sequencing reads (Supplementary Figure 1). The pipeline by itself is given as supplementary file 2 and online accessible at https://github.com/xoaib4/FalseAmpHunter. A description of all six steps is provided here to help with conceptual understanding of how this tool works.

1. The first step of our tool is QC of raw sequencing reads to avoid any results at later steps created by noise. We used the PRINSEQ tool for this purpose to remove any sequencing reads with lengths less than 20 bp, any reads with bases on the left or right with base quality scores less than 20, or any reads with mean base quality (BQ) scores less than 20 [20].
2. Subsequently, we generated QC reports using the fastQC tool (Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which are also one of the output logs of our FalseAmpHunter tool. fastQC is an established QC reporting tool in sequencing reads and produces a QC report in HTML.
3. Then using QC-passed paired-end reads, a *de novo* sequence was assembled using AbySS [21,22] given a k-mer 81 and bloom filter 6G. The k-mer size of 81 was chosen after testing several k-

 doi:10.20944/preprints202604.1800.v1

mer values, including 70, 81, and 96. The N50 and most contiguous assembly were produced with a k-mer size of 81 and were included in all downstream analyses.

4. The scaffolded assembled sequence of AbySS was then indexed using *samtools* [23], and QC-passed paired-end reads were mapped to this sequence using the Burrows-Wheeler Aligner (BWA) under default conditions [24]. This step ensures that the assembled sequence is representative of QC-passed reads and that both match each other.

5. Post-mapping *samtools coverage* was used to determine the efficiency of mapping on the assembled sequence. This coverage represented the best-assembled sequences, as the mapped reads were highest when best-matched with the assembled sequence. A mapping quality score (MPQ) of 20 or higher was used to shortlist the sequences. The MPQ is a Phred scale score, and an MPQ of 20 indicates a 99% probability of mapping with confidence to the correct genomic region.

6. The assembled sequence was then aligned with the human reference genome hg38 assembly database to identify similar sequences using the National Center for Biotechnology Information (NCBI) BLAST tool [25]. For NCBI BLAST, we used all default configurations. Among the similar sequences, primer sequences were searched to confirm that the amplified product was a false amplicon of the same primer. Our default sequence requires only one primer sequence to search; however, it can be searched again with a different primer sequence as many times as needed.

3. Results

3.1. False Amplicon Generation

To generate false amplicons, we selected OR genes, which are inherently difficult to amplify because of the presence of highly similar paralogs. We selected OR10G4, OR10G7, OR10G8, and OR10G9 to demonstrate off-target amplification and validate our pipeline, which can identify false amplicons. We used a set of forward and reverse primers to amplify OR10G4. The forward primer was 100% aligned with the OR10G4 gene, 88% with OR10G7, and 92% with OR10G8 and OR10G9, as shown in Figure 1A. The reverse primer was 100% aligned with the OR10G4 gene, 94% with OR10G7, 95% with OR10G8, and 100% with OR10G9, as shown in Figure 1B. We also presented sequence logos for both forward and reverse primers after alignment with similarity hits in the four OR genes, as shown in Figure 1.

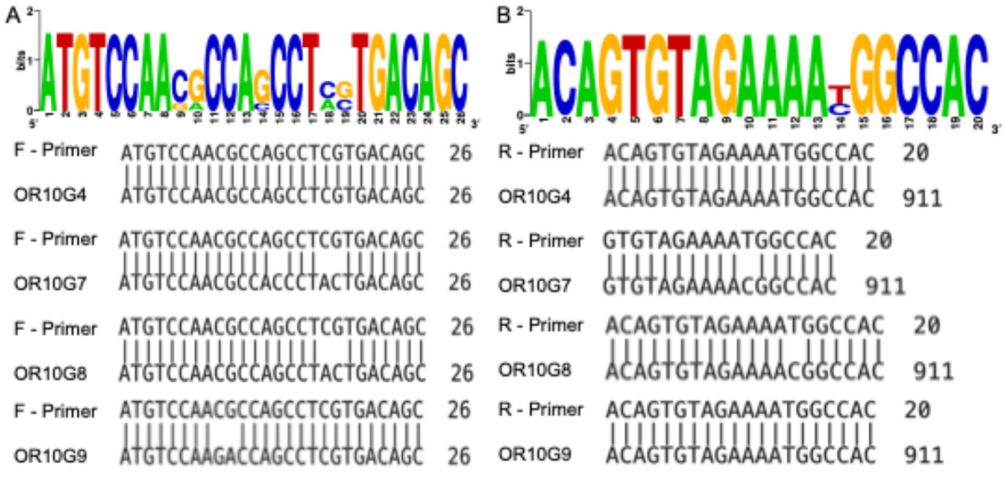

Figure 1. Sequence logos and alignments of designed primers. A: Sequence logo and alignment of designed forward primer with targeted gene. B: Sequence logo and alignment of designed reverse primer with targeted gene.

 doi:10.20944/preprints202604.1800.v1

Using this set of forward and reverse primers, we conducted an *in silico* PCR at the UCSC genome browser under default conditions. This PCR produced two different amplicons, as shown in Figure 2. These amplicons were similar to two sequences on chromosome 11. The coordinates of these sequences on chromosome 11 were 124015575-124016510 and 124023013-124023842. When we examined the annotations for these sequences, they were identified as OR10G4 and OR10G9. Thus, we successfully produced one false amplicon (OR10G9) in addition to one targeted (OR10G4). The remaining two ORs, OR10G7 and OR10G8, were not amplified by this PCR.

Figure 2. Output of UCSC *in silico* PCR. Two amplicons with their sequences and hg38 genomic coordinates are shown.

3.2. Sequence Reads Synthesis

The best and ultimate practice to identify a false amplicon is to sequence the entire amplicon. To mimic these steps in the discovery of false amplicons, we used the ART tool to synthesize paired-end raw NGS reads in the FASTQ format for both amplicons. In addition, we used a random decoy sequence of 1kb as a negative control. We also synthesized paired-end raw FASTQ files for this decoy sequence. For the amplicon FASTQ, we also contaminated these files with primer sequences and artificially designed chimeric sequences to simulate real sequencing artifacts commonly observed in PCR amplicon datasets. Each amplicon-FASTQ file was 4.3 MB in volume.

3.3. Application of FalseAmpHunter to find false amplicon

Once we had synthetic FASTQ reads for our amplicon, we ran FalseAmpHunter to identify a false amplicon. FalseAmpHunter ran successfully and assembled 1181 contigs, which were scaffolded into 336 scaffolds. The high number of contigs and scaffolds relative to the two-amplicon

 doi:10.20944/preprints202604.1800.v1

input mixture can be attributed to two factors: the intentionally induced synthetic noise, including primer carryovers and chimeric reads, which fragmented the assembly, and the inherent challenge posed to assemblers by highly similar paralogous sequences, where near-identical regions across OR10G4 and OR10G9 introduced ambiguity during contig extension and scaffolding. There were four output files in total, including assembled scaffolds in the FASTA format, BLAST results of scaffolded FASTA in the human reference genome hg38 in a text file, primer similarity in BLAST results in a text file, and coverage of each scaffold in a text file. We began by examining the outputs from the coverage results and identified scaffolds with the highest coverage, depth, and mapping quality score. We identified results with an MPQ of 20 or higher; three scaffolds met these criteria and are shown in Table 2. All three scaffolds were 100% covered, mapped by reads of mean BQ over 32, and had a mean MPQ between 39.1 and 54.4.

Table 2. Coverage statistics of top three scaffolds.

Scaffold	Start Position	End Position	Number of mapped reads	Covered bases	Coverage	Mean depth	Mean BQ score	Mean MPQ score
1	1	391	3043	391	100	1107.28	32.6	54.4
2	1	350	2777	350	100	1117.29	32.8	53.9
13	1	204	1292	204	100	878.451	32.8	39.1

We identified these sequences from FASTA file and went to second file in results with BLAST results of scaffolded FASTA in human reference genome hg38. The first scaffold, which was 391 bp long, was aligned to the chromosome 11 sequence from 124023652 to 124024042 without any gaps (Figure 3A). This region is part of the OR10G9 gene. The second scaffold was 350 bp long and was also aligned with the chromosome 11 sequence from 124016217 to 124016566 without any gaps (Figure 3B). This region is part of OR10G4, for which primers were targeted. The third scaffold was 204 bp long and was aligned with the chromosome 11 sequence from 124022939 to 124023142 without any gaps (Figure 3C). This region is also part of the OR10G9 gene. Both assembled OR10G9 sequences were distinct parts of the same gene. Thus, based on these coverage and assembled sequence results, we were able to identify sequences that were falsely amplified. Although the assembled scaffold lengths (391 bp, 350 bp, and 204 bp) were shorter than the expected full-length PCR products, OR10G4 (830 bp, coordinates 124015575–124016404) and OR10G9 (830 bp, coordinates 124023013–124023842), the pipeline nonetheless successfully identified both the target and false amplicons. The truncation of assembled scaffolds likely reflects the high sequence similarity between paralogous OR10G loci, which introduces ambiguity at assembly boundaries and causes assemblers to terminate contigs prematurely rather than risk mis-joining near-identical sequences.

 doi:10.20944/preprints202604.1800.v1

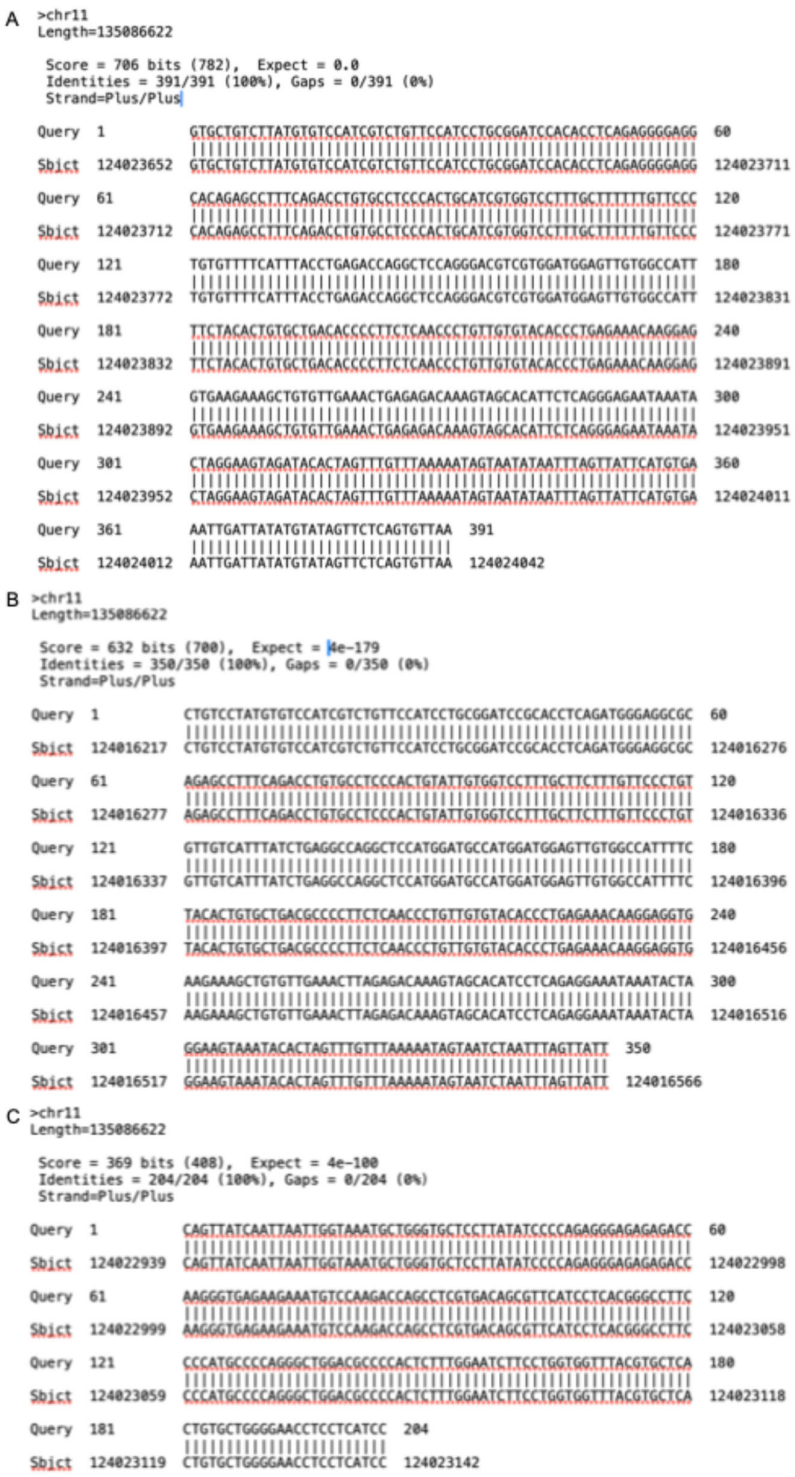

Figure 3. Alignment of three scaffolds with hg38 genome assembly. Each scaffold aligned with a different genomic region. Alignment shows BLAST statistics, any gaps and lengths of each scaffold.

When FalseAmpHunter was applied to the negative control decoy dataset, the de novo assembly step produced no output, as the random nucleotide sequence lacked the biological structure and k-mer redundancy necessary for successful assembly. This early termination without amplicon

 doi:10.20944/preprints202604.1800.v1

identification confirms that the pipeline does not generate false-positive calls from non-specific or random sequence inputs. AbySS produced empty output for the decoy dataset, likely because the k-mer size of 81 was sufficiently large; random sequences lacked the k-mer connectivity required to build assembly graphs, effectively filtering non-biological inputs at the assembly stage.

4. Discussion

PCR-based amplification remains a foundational technique in molecular biology and genomics [2–6,26,27]; however, its reliability is fundamentally constrained by primer specificity, particularly in genomic contexts characterized by extensive sequence similarity. Multigene families, such as olfactory receptors (ORs) and other paralog-rich loci, pose a persistent challenge, in which primers designed for a single target gene may inadvertently bind homologous regions and generate false amplicons. These off-target amplification events are difficult to detect using conventional quality control approaches, which rely solely on size or melting temperatures. In this study, we present FalseAmpHunter, a bioinformatics pipeline specifically designed to detect and distinguish true target amplicons from off-target amplification products arising from paralogous primer binding and/or PCR artifacts.

A key strength of FalseAmpHunter is its post-sequencing, data-driven validation strategy. Conventional strategies to address off-target amplification in PCR rely on either pre-PCR design or post-PCR observations, including *in silico* primer design metrics, pre-amplification specificity predictions, electrophoresis-based size prediction, and melting temperatures [12–14,18]. These strategies are observational and do not resolve the underlying problem. Rather than relying solely on observations, FalseAmpHunter evaluates the empirical sequencing output of false amplicons. By integrating read QC, de novo assembly, genome-wide alignment, and explicit primer matching, the pipeline reconstructs the full context in which an amplicon arises. This enables the identification of false amplicons that would otherwise remain unidentified. Importantly, the framework captures cases in which false amplicons display high coverage and apparent consistency, mimicking true target amplification, thereby posing a substantial risk for downstream misinterpretation.

The application of FalseAmpHunter to paralogous OR loci highlights the magnitude of this problem in sensory gene families. OR genes are among the most duplicated and sequence-conserved gene families in the human genome [17], and often share extensive homology across coding and flanking regions. As demonstrated in the workflow, a single primer pair designed for OR10G4 showed similarity to multiple paralogous loci, including OR10G7, OR10G8, and OR10G9, and co-amplified OR10G4 and OR10G9. Standard amplicon sequencing pipelines may incorrectly assign such reads to the intended target, particularly when relying on local alignments or gene-centric reference mapping. FalseAmpHunter resolves this ambiguity by enforcing genome-wide alignment of assembled scaffolds and explicitly verifying primer–scaffold concordance, thereby distinguishing genuine locus-specific amplification from paralog-driven artifacts.

Another important contribution of this work is the explicit handling of primer carryover and chimeric reads, which are common but often underappreciated sources of error in amplicon sequencing datasets. Primer-derived sequences can persist in sequencing libraries and generate misleading alignment signals, especially when present in low-complexity or repetitive regions [28]. By incorporating primer identification as a terminal step in the pipeline, FalseAmpHunter flags scaffolds and alignments that are dominated by primer matches rather than true genomic sequences, providing an additional layer of validation that is absent from most amplicon analysis workflows. Notably, the negative control decoy dataset did not assemble or produce any results, consistent with the experimental design.

From a broader methodological perspective, FalseAmpHunter addresses a critical gap between primer design theory and sequencing reality. Although numerous tools exist to design primers with minimal predicted off-target binding [29], these predictions are inherently limited by reference completeness, mismatch tolerance assumptions, and thermodynamic simplifications. Moreover, these tools cannot account for stochastic amplification dynamics or sequencing-induced artifacts.

 doi:10.20944/preprints202604.1800.v1

FalseAmpHunter complements these approaches by offering a retrospective validation framework that can be applied to existing datasets, including legacy studies in which primer specificity concerns may not have been fully appreciated at the time of data generation.

The implications of this study extend beyond the OR research. Off-target amplification can affect diverse applications [30], including clinical variant validation, pathogen detection, copy number estimation, and targeted resequencing panels. In clinical and diagnostic contexts, false-positive amplification of paralogous genes can lead to incorrect variant calls or gene misassignment, with potential downstream consequences for interpretation and decision-making. By providing a systematic approach to identify and flag such events, FalseAmpHunter enhances confidence in PCR-based sequencing assays and supports more robust downstream analyses.

Several limitations should be acknowledged. FalseAmpHunter is currently optimized for short-read amplicon sequencing data and relies on accurate genome references for alignment-based disambiguation. The biggest challenge in the current dataset was that the assembled false amplicon sequences were shorter than the actual amplicons. This difference was representative of assembly challenges in highly similar paralogous sequences [17]. Long-read sequencing (LRS)-based approaches can address this challenge when applied. However, the LRS has higher error rates, and a modified pipeline based on the same concept will be required. Highly fragmented or incomplete reference genomes may reduce its effectiveness, particularly in non-model organisms. Additionally, while the pipeline is computationally tractable for typical amplicon datasets, scaling to extremely large multiplexed panels may require further optimization. Future extensions could incorporate long-read sequencing data, thermodynamic primer-binding models, or probabilistic scoring schemes to further refine amplicon classification.

5. Conclusions

FalseAmpHunter provides a principled and practical solution to a pervasive but under-addressed problem in PCR-based genomics. By explicitly modeling the consequences of paralogous primer binding and sequencing artifacts, it enables reliable discrimination between true and false amplicons. This framework strengthens confidence in targeted sequencing studies and encourages more rigorous validation practices in paralog-rich genomic regions. As targeted sequencing continues to play a central role in both research and clinical settings, tools such as FalseAmpHunter will be increasingly important for ensuring data accuracy and biological interpretability.

References

1. Joshi, M., Deshpande, J. Polymerase chain reaction: methods, principles and application. International Journal of Biomedical Research. 2010, 2, 81-97.
2. Chudhary, S.A., Imtiaz, S., Iqbal, N. Laboratory detection of novel corona virus 2019 using polymerase chain reaction. The International Journal of Frontier Sciences. 2020, 4, 101-.
3. Shehzad, H., Sajjad, O. Detection of single nucleotide polymorphism rs2013162 of IRF6 gene in patient with cleft lip and palate. The International Journal of Frontier Sciences. 2019, 3, 28-40.
4. Pervaiz, A., Akhtar, M.S., Mahmood, S., Sajjad, O., Khaliq, S., Berger, M.R. Molecular basis of cell cycle arrest induced by erufosine in metastatic breast cancer cells. Cancer Research. 2018, 78, 4307-.
5. Akhtar, M.S., Ali, S.S. Erufosine alters the genes associated with G2/M Phase of cell cycle in cancers: Molecular evidence from gene expression analysis. Biomedical Current Insights. 2024, 1.
6. Ali, S.S., Akhtar, M.S. Antitumor potential of plant protein (Riproximin) against breast and colorectal cancer: Facts from functional and molecular investigations. Biomedical Current Insights. 2025, 2, 21-6.
7. Zar, M.S., Bibi, A., Akhtar, M.S. Unravelling time after death: A comprehensive review of multi-omics approaches in postmortem interval estimation. Forensic Science International. 2026, 379, 112779. https://doi.org/10.1016/j.forsciint.2025.112779
8. Borah, P. Primer designing for PCR. Science Vision. 2011, 11, 134-6.
9. Apte, A., Daniel, S. PCR primer design. Cold Spring Harbor Protocols. 2009, 2009, pdb. ip65.

 doi:10.20944/preprints202604.1800.v1

10. Wen, D., Zhang, C. Universal Multiplex PCR: a novel method of simultaneous amplification of multiple DNA fragments. Plant Methods. 2012, 8, 32. 10.1186/1746-4811-8-32
11. Barghouthi, S.A. A Universal Method for the Identification of Bacteria Based on General PCR Primers. Indian Journal of Microbiology. 2011, 51, 430-44. 10.1007/s12088-011-0122-5
12. Ruiz-Villalba, A., van Pelt-Verkuil, E., Gunst, Q.D., Ruijter, J.M., van den Hoff, M.J.B. Amplification of nonspecific products in quantitative polymerase chain reactions (qPCR). Biomolecular Detection and Quantification. 2017, 14, 7-18. https://doi.org/10.1016/j.bdq.2017.10.001
13. Pan, Y.-B., Burner, D., Ehrlich, K., Grisham, M., Wei, Q. Analysis of primer-derived, nonspecific amplification products in RAPD-PCR. BioTechniques. 1997, 22, 1071-7.
14. Ruijter, J.M., Ruiz-Villalba, A., van den Hoff, A.J.J., Gunst, Q.D., Wittwer, C.T., van den Hoff, M.J.B. Removal of artifact bias from qPCR results using DNA melting curve analysis. The FASEB Journal. 2019, 33, 14542-55. https://doi.org/10.1096/fj.201901604R
15. Crossley, B.M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M.L., et al. Guidelines for Sanger sequencing and molecular assay monitoring. Journal of Veterinary Diagnostic Investigation. 2020, 32, 767-75. 10.1177/1040638720905833
16. ten Bosch, J.R., Grody, W.W. Keeping Up With the Next Generation: Massively Parallel Sequencing in Clinical Diagnostics. The Journal of Molecular Diagnostics. 2008, 10, 484-92. https://doi.org/10.2353/jmoldx.2008.080027
17. Akhtar, M.S., Ashino, R., Oota, H., Ishida, H., Niimura, Y., Touhara, K., et al. Genetic variation of olfactory receptor gene family in a Japanese population. Anthropological Science. 2022, 130, 93-106. 10.1537/ase.211024
18. Perez, G., Barber, G.P., Benet-Pages, A., Casper, J., Clawson, H., Diekhans, M., et al. The UCSC Genome Browser database: 2025 update. Nucleic Acids Res. 2025, 53, D1243-d9. 10.1093/nar/gkae974
19. Huang, W., Li, L., Myers, J.R., Marth, G.T. ART: a next-generation sequencing read simulator. Bioinformatics. 2011, 28, 593-4. 10.1093/bioinformatics/btr708
20. Schmieder, R., Edwards, R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011, 27, 863-4. 10.1093/bioinformatics/btr026
21. Jackman, S.D., Vandervalk, B.P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S.A., et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Research. 2017, 27, 768-77. 10.1101/gr.214346.116
22. Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., Birol, I. ABySS: a parallel assembler for short read sequence data. Genome Research. 2009, 19, 1117-23. 10.1101/gr.089532.108
23. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009, 25, 2078-9. 10.1093/bioinformatics/btp352
24. Li, H., Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009, 25, 1754-60. 10.1093/bioinformatics/btp324
25. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L. NCBI BLAST: a better web interface. Nucleic Acids Research. 2008, 36, W5-9. 10.1093/nar/gkn201
26. Ehnert, S., Linnemann, C., Braun, B., Botsch, J., Leibiger, K., Hemmann, P., et al. One-Step ARMS-PCR for the Detection of SNPs—Using the Example of the PADI4 Gene. Methods and Protocols. 2019, 2, 63.
27. Akhtar, M.S., Chan, A., Liu, H., Coffey, L.L., Gong, Q., editors. Rapid and Widespread Interferon Stimulated Response in Olfactory Sensory Neurons Upon SARS-CoV-2 infection. CHEMICAL SENSES; 2023: OXFORD UNIV PRESS GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND.
28. Nhu, H.N., Dylan, S., Kabir, P., Peter, K. Parsing ecological signal from noise in next generation amplicon sequencing. The New Phytologist. 2015, 205, 1389-93.
29. Guo, J., Starr, D., Guo, H. Classification and review of free PCR primer design software. Bioinformatics. 2020, 36, 5263-8. 10.1093/bioinformatics/btaa910
30. Borst, A., Box, A.T.A., Fluit, A.C. False-Positive Results and Contamination in Nucleic Acid Amplification Assays: Suggestions for a Prevent and Destroy Strategy. European Journal of Clinical Microbiology and Infectious Diseases. 2004, 23, 289-99. 10.1007/s10096-004-1100-1

 doi:10.20944/preprints202604.1800.v1