



Article

ConsensusPrime—A Bioinformatic Pipeline for Efficient Consensus Primer Design—Detection of Various Resistance and Virulence Factors in MRSA—A Case Study

Maximilian Collatz^{1,2,*}, Martin Reinicke^{1,2}, Celia Diezel^{1,2}, Sascha D. Braun^{1,2}, Stefan Monecke^{1,2,3}, Annett Reissig^{1,2} and Ralf Ehricht^{1,2,4}

¹ Leibniz Institute of Photonic Technology (IPHT), 07745 Jena, Germany; ralf.ehricht@leibniz-ipht.de (R.E.)

² InfectoGnostics Research Campus, 07743 Jena, Germany

³ Institute for Medical Microbiology and Virology, Dresden University Hospital, 01307 Dresden, Germany

⁴ Institute of Physical Chemistry, Friedrich-Schiller University, 07743 Jena, Germany

* Correspondence: maximilian.collatz@leibniz-ipht.de

Abstract: Background: The effectiveness and reliability of diagnostic tests that detect DNA sequences largely hinge on the quality of the used primers and probes. This importance is especially evident when considering the specific sample being analyzed, as it affects the molecular background and potential for cross-reactivity, ultimately determining the test's performance. Methods: Predicting primers based on the consensus sequence of the target has multiple advantages, including high specificity, diagnostic reliability, broad applicability, and long-term validity. Automated curation of the input sequences ensures high-quality primers and probes. Results: Here, we present a use case for developing a set of consensus primers and probes to identify antibiotic resistance and virulence genes in *Staphylococcus (S.) aureus* using the ConsensusPrime pipeline. Extensive qPCR experiments with several *S. aureus* strains confirm the exceptional quality of the primers designed using the pipeline. Conclusions: By improving the quality of the input sequences and using the consensus sequence as a basis, the ConsensusPrime pipeline ensures high-quality primers and probes, which should be the basis of molecular assays.

Keywords: consensus primer; primer design; alignment filtering; *Staphylococcus aureus*; MRSA; qPCR; bioinformatics



Citation: Collatz, M.; Reinicke, M.; Diezel, C.; Braun, S.D.; Monecke, S.; Reissig, A.; Ehricht, R.

ConsensusPrime—A Bioinformatic Pipeline for Efficient Consensus Primer Design—Detection of Various Resistance and Virulence Factors in MRSA—A Case Study. *BioMedInformatics* **2024**, *4*, 1249–1261. <https://doi.org/10.3390/biomedinformatics4020068>

Academic Editors: Hans Binder and Themis Exarchos

Received: 8 March 2024

Revised: 30 April 2024

Accepted: 7 May 2024

Published: 10 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The efficiency and specificity of the widely used molecular biological method of polymerase chain reaction (PCR) for detecting target sequences depends on the quality of the primers designed for amplification. Consensus primers designed on a representative set of sequences ensure that the target can be efficiently amplified despite genetic variations in the target gene.

As researchers engage in applications ranging from gene expression studies to genetic diagnostics, the ability to design primers that capture consensus sequences becomes a linchpin for success. Consensus primer design mitigates the risk of nonspecific amplification and accommodates the genetic diversity inherent in various samples. This introduction aims to underscore the critical role of consensus primer design in shaping the reliability and accuracy of PCR, thus influencing the outcomes of diverse molecular biology applications with a heightened focus on versatility and adaptability.

Various applications, such as PrimerDesign-M [1], MPD [2], and Oli2Go [3], can predict primers and probes based on sequence alignments. However, none of these tools cure the input alignment and use the consensus sequence as a design basis. Instead, they generate degenerate primers that differ at the positions where the entered sequences differ. Consequently, they generate several suitable primers in case of sequence differences, leading to a larger number of primers and probes being designed and a higher probability

of cross-reactions. Since the number of primers that can be used in one assay is limited, an approach using only a single consensus primer is preferred.

Another problem is that Oli2Go [3] was only available via a web interface that is no longer accessible.

The ConsensusPrime [4] pipeline can streamline the automated curation of the input alignment, a feature no other available tool offers, and the so-predicted consensus primers strike a balance between a reduced number of primers and probes while maintaining sufficient performance. Also, the automatic visualization of the designed sequences in the report file provided by the ConsensusPrime pipeline serves as helpful first quality control.

Using consensus sequences to predict primers for clinical assays can be a valuable approach for several reasons. **Conserved Regions:** Consensus sequences are derived from aligning multiple sequences of the same gene or genomic region from different organisms. The regions that remain conserved across these sequences will likely be functionally essential for the gene's function. In the context of primer design, targeting these conserved regions increases the likelihood that the primers will successfully amplify the desired target across a wide range of organisms, including potentially unknown or diverse strains. **Broad Applicability:** Clinical assays often need to be sensitive and specific across a diverse range of samples, including various strains or variants of a pathogen. Consensus-based primers are more likely to work effectively in such situations because they target regions maintained throughout the organism's evolution. **Specificity:** Consensus sequences are generated by identifying the most common nucleotides at each position in the alignment. This process minimizes the chance of designing primers that might bind to nonspecific or unintended regions within the genome. Primers designed to target conserved regions are less likely to produce false-positive or false-negative results due to cross-reactivity or off-target amplification. **Diagnostic Reliability:** In clinical settings, accuracy and reliability are crucial. Consensus sequences help ensure that the designed primers are specific to the target of interest, minimizing the risk of false-positive or false-negative results caused by nonspecific binding to closely related sequences. **Reduced Variability:** In the case of rapidly mutating organisms, or more precisely, their genes, consensus sequences can reduce the impact of variability on primer efficiency. If one designs primers based on a single sequence, they may not work well if the target sequence is significantly mutated. Consensus sequences provide a more stable target. **Inter-laboratory Consistency:** Consensus-based primers can promote consistency between laboratories performing the same clinical assays. Since these primers are designed based on well-established sequence data, labs can expect similar results when using the same primers. **Long-term Validity:** Because the sequences are located in well-conserved areas, it is ensured that the primers do not lose their efficiency and relevance over time. **Ease of Design:** By selecting suitable consensus primers/probes, it can be avoided designing degenerate primers/probes which can simplify the experimental setup. Especially in assay design, the number of probes designed plays a role when the total number of possible probes is limited, for instance, due to color filters/channels and fluorescent dyes.

Multiple sequence alignments (MSA) form the basis for determining the consensus sequence, but aligning sequences is challenging due to their divergence, variations in length, composition, and evolutionary history. Aligning numerous sequences requires significant computational resources and efficient algorithms [5]. Different alignment algorithms optimize accuracy and speed based on sequence characteristics and available resources [6]. Interpreting MSAs involves integrating computational predictions with experimental data to identify conserved regions, functional motifs, and evolutionary relationships. Aligning sequences with insertions or deletions requires the addition of gaps. However, balancing the need to align similar regions with the risk of introducing gaps in unrelated areas is crucial. As gaps are only a theoretical concept to align divergent sequences, they are removed from the consensus sequence before the primer design process.

Since consensus primers are based on the combined information content of several sequences, the quality of these input sequences is crucial. In order to keep the design with sequences of varying quality from databases or BLAST searches as simple as possible and

still have no loss of quality, the pipeline already starts with the automatic curation of the starting sequences. In order to keep this process as transparent as possible, the results of each filter step are output in the form of a separate alignment for manual visualization, and the quantitative results are recorded in the final report file. The initial data processed this way forms the basis for the subsequent consensus primer design. This combination of qualitative enhancement of the initial sequences and the automatic design of ideal consensus primers makes the ConsensusPrime Pipeline a valuable and unique bioinformatics tool.

A decisive criterion for the functionality and quality of the primers and probes designed with the pipeline is the quality of the sequence data used in the design. The sequences used for the input alignment must represent the target sequences. Otherwise, the predicted primers may not precisely and reliably recognize the expected target sequences.

Since genomic DNA changes through constant mutation, predicting primers and probes in the least affected regions is essential to improve the long-term utilization of the designed sequences. The ever-increasing number of sequences available for design is another crucial reason for automating the selection of ideal regions. In this way, additional sequences can be reliably and quickly added to an existing design in the future to illustrate their influence on the need for potential new primers and probes.

Primers and Probes designed using the previously described principles can be used to identify pathogens and their resistances in a molecular biological assay. One very relevant and widespread pathogen is *Staphylococcus (S.) aureus*. Due to its multidrug resistance and virulence, methicillin-resistant *S. aureus* (MRSA) presents a formidable global challenge in healthcare and community settings. *S. aureus*, a bacterium commonly colonizing the skin and nasal passages of humans and animals, has evolved mechanisms to resist the effects of β -lactam antibiotics, including methicillin and other penicillin derivatives, by producing β -lactam lysing enzymes and the expression of penicillin-binding protein 2a (PBP 2a) that resists the antibiotics [7]. The emergence and spread of MRSA have significantly increased the difficulty of treating staphylococcal infections, leading to elevated morbidity, mortality, and healthcare costs.

Accurate detection of *S. aureus* and resistance classification is essential in clinical diagnostics. Based on these classifications, antibiotics can be targeted to reduce side effects and prevent further resistance. One method for detecting and predicting this resistance is qPCR (quantitative polymerase chain reaction), making detecting various target genes possible. In this paper, seven genes were selected and divided into one species marker, four virulence genes, and three resistance genes.

In order to be able to reliably identify *S. aureus*, *eno* was selected as a species marker. *eno* encodes an enolase/phosphopyruvate hydratase/laminin-binding protein and is part of the mRNA degradosome holoenzyme-like complex [8,9].

The *mecC* gene is a recently discovered homolog to the beta-lactam resistance gene *mecA* [10,11]. It is located on the staphylococcal cassette chromosome *mec* (SCC*mec*) XI and originates from zoonotic *S. aureus*. It has a sequence similarity of only 69% to *mecA*. Thus, it might cause diagnostic problems as *mecA* primers and probes, as well as antibodies for the *mecA* gene product, might fail to detect *mecC* or, respectively, its gene product.

Fusidic acid resistance protein C, encoded by the *fusC* gene, provides resistance against fusidic acid, a commonly used topical antibiotic [12–15]. *fusC* is often found in conjunction with *mecA* on SCC elements, further complicating the treatment of MRSA infections [16–18].

The *lukF-PV* gene encodes a component of the bicomponent leukocidin PVL (Pantone Valentine Leukocidin), which is associated with necrotizing pneumonia and chronic/recurrent skin and soft tissue infections in humans [19]. Similarly, *lukF-P83*, a component of another phage-borne bi-component leukocidin, is linked to bovine mastitis, and it is present in *S. aureus* isolates from various animal hosts [20–22].

Lastly, the genes *sea* and *seb* (formerly known as *entA* and *entB*) encode Enterotoxins A, respectively B, superantigenic toxins implicated in (non-menstrual) toxic shock syndrome and food intoxication [23–26]. These toxins induce a massive immune response by unspecific binding to MHC receptors, leading to T-cell activation and subsequent clinical manifestations [27].

Understanding the genetic determinants associated with MRSA virulence and resistance is essential for implementing effective strategies to combat and prevent the spread of MRSA infections in healthcare facilities and communities.

2. Materials and Methods

In this paper, we have used the ConsensusPrime pipeline to design a set of primers and probes to detect the clinically relevant bacterium *S. aureus* and several of its relevant resistance and virulence genes. They were analyzed in extensive qPCR experiments and tested with different strains to prove the high quality of the designed primers and probes. Algorithmic details of the ConsensusPrime pipeline can be found in the original publication (<https://doi.org/10.3390/biomedinformatics2040041>, accessed on 11 January 2024).

2.1. Function of the ConsensusPrime Pipeline

The ConsensusPrime pipeline uses multiple homologous sequences to predict primers and probes in the most conserved regions. Therefore, the pipeline combines various quality filters for the input sequences with automatic primer prediction using Primer3 in the conserved regions. The pipeline filters the input sequences by a similarity threshold to their consensus sequence and removes duplicate sequences. Based on this filtered alignment, the most homolog regions are identified and used for the primer and probe design.

2.2. Experimental Evaluation Setup

The quality of the predicted primers and probes was evaluated by a qPCR dilution series for each primer/probe pair with 14 different *Staphylococcus aureus* strains and one *Staphylococcus epidermidis* strain.

2.2.1. Bacterial Strains

In order to assess the efficacy of the predicted primer and probe sets, we used 15 strains from our in-house strain collection. Several criteria were considered when selecting the strains. These included the availability of the bacteria as a culture, the representative coverage of the target genes, i.e., the presence of the corresponding resistance and virulence genes, and, in some cases, the availability of previously published genomes. If the strains had not already been published with their genomes, the genomes of the strains were sequenced and published on GenBank. These strains were cultivated on Columbia blood agar (Becton Dickinson, Heidelberg, Germany) for about 12 h at 37 °C. All strains for which the genomic sequence was not already available in databases were sequenced using ONT next-generation whole genome sequencing, see Table 1. The expected qPCR results were derived based on the genomic sequences and used as ground truth for comparison against our experimental results, see Table 2.

Table 1. Reference strains and their genomic sequences were used to evaluate the performance of the predicted qPCR primer and probes.

Organism	Strain ID	Accession Number
<i>Staphylococcus aureus</i>	Jena-IPHT-124322	CP155060
<i>Staphylococcus aureus</i>	MW2	BA000033.2
<i>Staphylococcus aureus</i>	NRS158	SRS3408576
<i>Staphylococcus aureus</i>	Jena-IPHT-124622	CP155059
<i>Staphylococcus aureus</i>	Jena-IPHT-97437	CP155061
<i>Staphylococcus aureus</i>	Jena-IPHT-124982	CP155058
<i>Staphylococcus aureus</i>	Jena-IPHT-124984	CP155057
<i>Staphylococcus aureus</i>	Jena-IPHT-94881	CP155063
<i>Staphylococcus aureus</i>	Jena-IPHT-95377	CP155062
<i>Staphylococcus aureus</i>	M10/0061	FR823292.1
<i>Staphylococcus aureus</i>	N315	BA000018.3
<i>Staphylococcus aureus</i>	NCTC8325	CP000253.1
<i>Staphylococcus aureus</i>	TCH1516	CP000730.1
<i>Staphylococcus aureus</i>	MSSA476	BX571857.1
<i>Staphylococcus epidermidis</i>	RP62A	NC_002976.3

2.2.2. Genome Sequencing

DNA extraction for Nanopore MinION sequencing (Oxford Nanopore Technology, Oxford, UK) was performed using the Nucleospin Microbial DNA Kit by Macherey Nagel (MN, Düren, Germany). Initially, all strains were cultured from cryo-cultures (Microbank; Thermo Fisher Scientific, Waltham, MA, USA) on blood agar plates at 37 °C overnight. Subsequently, one entire inoculation loop per strain was washed with 500 µL 1× PBS (pH 7.4), centrifuged, and resuspended in 100 µL buffer BE. Following the manufacturer's instructions, with two minor adaptations: (1) Samples underwent lysis using a BeadBug™ microtube homogenizer (Benchmark Scientific, Sayreville, USA) for 12 min (for Gram-positive bacteria) or 4 min (for Gram-negative bacteria) at full speed. (2) Before binding the DNA onto Nucleospin microbial DNA columns, proteinase K was inactivated by incubating the samples at 70 °C for 5 min. After cooling down, 4 µL of RNase (100 mg/mL; Sigma Aldrich, Steinheim, Germany) was added, and the samples were incubated at 37 °C for 5 min. Finally, DNA was eluted twice with 75 µL of nuclease-free water (Carl Roth, Karlsruhe, Germany).

The genome sequencing of the 15 strains was performed on one MinION flow cell (12 strains using Barcoding for multiplexing) and three different Flongle flow cells. Library preparations were carried out using the 1D genomic DNA ligation kit (SQK-LSK 109) and the native barcoding expansion kits. In brief, size selection and DNA clean-up were performed using Agencourt AMPure XP beads (Beckman Coulter GmbH, Krefeld, Germany) at a ratio of 1:1 (*v:v*) before library preparation. Repair of potential nicks in DNA and DNA ends was conducted in a combined step using NEB Next FFPE DNA Repair Mix and the NEB Next Ultra II End repair/dA-tailing module (New England Biolabs, Ipswich, MA, USA) with the incubation time tripled. Before adapter ligation, barcodes were ligated to the dA-prepared ends of the DNA, followed by a second AMPure clean-up. A subsequent third AMPure bead purification was carried out before the ligation of sequencing adapters onto the prepared ends. The initial quality check of the flow cell indicated a minimum of 1200 active pores at the start of sequencing. For the Flongles, the average number of active pores ranged between 80 and 150. Genomic DNA samples used for loading comprised around 40 to 60 ng per strain (measured by Qubit 4 Fluorometer; Thermo Fisher Scientific). The sequencing was conducted for 72 h for the MinION flow cell and 12 h for each Flongle using the MinKNOW software version 22.05.5.

The guppy basecaller (version 4.5.2 (flow cell) or 6.0.1 (flongles), Oxford Nanopore Technologies) was employed to translate MinION raw reads (FAST5) into high-quality tagged sequence reads with 4000 reads per FASTQ-file. The barcode trimming option (model version: dna_r9.4.1_450bps_sup.cfg, and dna_r10.4.1_e8.2_400bps_sup) was utilized. The flye software (version 2.8.3) was utilized to assemble each strain's quality tagged sequence reads into a complete, circular contig. The assemblies were polished in two stages. Firstly, four iterative rounds of racon (v1.4.21) were conducted with parameters including match 8, mismatch 6, gap 8, and window lengths of 500. Subsequently, medaka (version 1.4.3) was employed on the last racon-polished assembly using the models r941_min_sup_g507 and r10.4.1_e82_400bps_sup_g615. Finally, Abricate (v1.0.0) was utilized to screen the resulting corrected assembly for resistance and virulence genes. To access the sequenced data, see Table 1. Quality check of the sequencing data was performed by NanoPlot and NanoComp to assess sequencing quality and errors. Genome data were validated by checking with IDEEL. In this step, all open reading frames from sequencing data were compared against the UniProt TREMBL database to identify sequence variances.

2.2.3. Genomic DNA Dilution

We created a 10-fold dilution series of genomic DNA for each reference strain to validate and measure the experiments. By calculating genomic equivalents (GE) using the genome size of sequenced specimens of the bacterial pathogen species, we achieved a relative quantification of all marker genes situated on the chromosome, encompassing species markers. The genes associated with resistance in different reference strains were

situated either on the chromosome or plasmid. The calculation offered a semi-quantitative assessment of the copy number for resistance marker genes encoded on plasmids, relying on the genome copy number.

2.2.4. qPCR Analysis

To analyze the predicted primers' quality, qPCR assays were conducted in 20 μ L volumes. The qPCR mix comprised 1 \times PCR buffer B1V10 (BLINK AG, Jena, Germany), 3 mM MgCl₂, 200 nM each of primer and TaqMan probe, 4 U HotStart-Taq polymerase (Biotechrabbit, Berlin, Germany), 1 mg/mL BSA (NEB, Ipswich, MA, USA), 200 nM dNTPs/dUTP (Biotechrabbit, Berlin, Germany), and 0.2 U Uracil-DNA glycosylase (Biotechrabbit, Berlin, Germany). Amplification was carried out in a QuantStudio5 qPCR cycler (Thermo Fisher Scientific Inc., Waltham, MA, USA) with an initial denaturation at 95 °C for 5 min followed by 40 cycles of 20 s denaturation at 95 °C and annealing and elongation at 61 °C for 30 s.

For validation of the monoplex qPCR assays, a calibration curve was established for each marker using a 10-fold dilution series from 10⁶ down to 10¹ GE/ μ L, with a 2 μ L template volume of genomic DNA from reference strains. The efficiency of qPCR was determined based on the slope of the calibration curve.

3. Results

In a comparison between in-silico prediction and in-vitro experiments, we could show that the prediction corresponds perfectly with the results generated in the laboratory, proving the excellent quality of the primers and probes predicted with the ConsensusPrime pipeline. The detailed results of the comparison for all predictions against the measured qPCR results are listed in Table 2.

3.1. Primer and Probe Design

Using the ConsensusPrime pipeline, we designed a set of seven primers/probes and tested them against 15 selected reference strains of *S. aureus/epidermidis*. Dilution series were prepared for each primer/probe pair to validate their efficiency. All primer/probe pairs showed high efficiency in all dilutions, displaying the capability of the pipeline to predict well-working primers/probes from multiple homolog sequences.

3.1.1. Target Genes

Staphylococcus aureus will be specifically identified in this study, and a small selection of virulence and resistance genes will be detected. For this purpose, one species marker (*eno*), four virulence markers (*sea*, *seb*, *lukF-PV* (human), *lukF-P83* (bovine)), and two resistance genes (*mecC*, *fusC*) were selected for which suitable consensus primers and probes will be designed in the following. In order to make this selection, extensive literature and database searches are usually necessary, which is why this process cannot be automated very well. For each target gene selected, a reference sequence must be defined. In order to design suitable consensus primers for the target gene, homologous sequences of the gene are required. These can be found either in suitable homology databases or by searching for the homologous sequences with the help of the reference sequence and a BLAST search. In both cases, the sequences found should first be checked for their organism of origin and whether this is relevant for the consensus primer. A distinction must be made between sequences that are not relevant, i.e., that can be ignored, and those that the primer must not recognize under any circumstances. As a practical example, a PCR for the clinically highly relevant staphylococcal virulence factor *lukF-PV* must not recognize the common gene *lukF*, which is present in (nearly) all isolates of that species. Those that must not be recognized can be included as a consensus sequence in the final alignment. In this way, it is possible to check how similar the designed primers are to the unwanted sequences and whether this leads to unspecific results. These homologous sequences are then the starting point for the ConsensusPrime pipeline.

3.1.2. Primer Design Parameters

Another prerequisite for the design of good primers is the selection of suitable design parameters. These include the T_m values, GC content, the primer/probe sequences' length, and the desired product length. For all information on possible primer design parameters, please refer to the Primer3 manual [28].

These parameters are stored in the primer3 parameter text file and transferred to the pipeline. All details can be found in the Supplementary Materials 1, 2 and 3. In this way, the user can either provide individual parameters for each primer design or use one primer3 parameter file for all designs of an assay. The primer3 parameter file contained the parameters about the desired T_m values, the size of the primer/probe, and the product of the sequences to be chosen.

3.1.3. Alignment Filter Thresholds

The last parameters passed via the command line at the start of the pipeline concern the filtering of the input sequences. These parameters determine the threshold values for selecting the homologous regions for the following primer/probe design. The pipeline uses the default values if these parameters are not explicitly specified. The default value for the `--consensussimilarity` is 0.8, which ensures that sequences less than 80% consistent with the consensus sequence of all entered sequences are removed for the following steps. The default value for the `--consensusthreshold` is 0.95, meaning only regions with a consistently high homology of at least 95% and above are used for primer/probe design.

3.2. Example Case *eno*

Based on the *eno* gene sequences as an example, the primer/sample prediction process with the ConsensusPrime pipeline is described below. The command used to generate the following example looks as follows.

```
consensus_prime.py --infile eno/eno.fas --primer3 eno/primer3_parameters.txt --outdir eno
--consensusthreshold 1.0
```

First, the pipeline begins to filter the sequences entered, in this case, 304 sequences from the in-house database. Alternatively, the user can use the reference sequence for a BLAST [29] search and then use the sequences found as input. It is essential to check the matches for meaningfulness in order to remove unwanted matches early to avoid their unwanted influence on the primer/probe prediction. In the first filtering step, duplicates are removed to reduce the influence of over-represented sequences, resulting in 46 unique sequences. The sequences that differ too much from the rest are removed in the next step. The sequence removed in this case was only partial and was therefore excluded.

The remaining 45 sequences form the basis for the consensus sequence on which the primers and probes are finally predicted. The final report file lists the number of processed sequences after each filter step; see Table 3. For each position of the consensus sequence, the consensus score is now calculated, reflecting the homogeneity. A consensus score of 0.95 means that 95% of the nucleotides in the alignment at this position are identical to the consensus sequence. The pipeline creates alignments for each filter step for complete traceability and manual control. In case of unexpected results, these can be checked to find possible causes quickly.

Due to the excellent homology in our input alignment, the parameter for the `--con-sensus` threshold was set to 1.0, which results in only regions with perfect homology being used for primer/probe prediction. This parameter often needs to be adjusted individually based on the data entered. If the pipeline cannot predict primers/probes, this parameter is one of the best ways to influence the prediction.

Table 3. Overview of the filtering steps performed by the ConsensusPrime Pipeline. Shown are the number of sequences after each filtering step. The original fasta file contained 304 Sequences. After the first filtering step, 46 unique sequences remained. One of these sequences was removed for being too different from the rest. In this case, the sequence removed was partial. The remaining 45 sequences were the sequences used for the consensus primer design.

Fasta/Alignment	Number of Sequences
Input sequences	304
Unique sequences	46
Unique similar sequences	45

The final results are visualized in a multiple-sequence alignment and presented in the HTML report file. See Figure 1. In addition to the alignment, detailed information is provided for each predicted set of primers/probes. This information includes the parameters defined in the primer3 parameter file, such as Tm values, GC content, or information about melting properties for self-bonding and hairpin structure formation. Based on the alignment combined with the detailed information, the best-desired primers can be selected manually or based on the penalty scores calculated by primer3. In the case of *eno*, we chose the second primer/probe pair because the reverse primer starts and ends with G/C, which ensures good binding properties, see Table 4. Note that the numbering of the primers by primer3 starts with 0. The melting values for possible self-bonding are far below the PCR temperatures and should, therefore, not hurt the efficiency of the primers/probes.



Figure 1. Shown is a section of the visualization of the final alignment containing. The alignment structure is always in the order of filtered sequences, consensus sequence, sequence parts used for the primer prediction, and predicted primer/probe pairs. This alignment enables the user to check the alignment of the filtered input sequences and the resulting consensus regions considered for the primer prediction. It also gives an excellent overview of the predicted Primers to choose the best pair if multiple predictions have been made.

Table 4. Shown is the listing of predicted primers/probes from the *eno* example as listed in the report file. Note that the sequence named “PRIMER_RIGHT_1_SEQUENCE_REV” matches the sequence in the alignment. For a functional set of forward and reverse primers, the second reverse complementary sequence of the reverse primer (5'-3') must be ordered.

PRIMER_LEFT_1_SEQUENCE_FWD (5'-3')	TGCCAATTATTACAGATGTTTACGC
PRIMER_RIGHT_1_SEQUENCE_REV	CCATCAGGTGCTTCAACTGG
PRIMER_RIGHT_1_SEQUENCE_REVCOMP (5'-3')	CCAGTTGAAGCACCTGATGG
PRIMER_INTERNAL_1_SEQUENCE	CGCGAAGTCTTAGACTCTCGTGGTAACCCAAGTGT

The prediction process described for *eno* using the ConsensusPrime pipeline was repeated for the other five target genes. The predicted qPCR primer/probe pairs were then synthesized and tested with in-house reference strains where relevant resistances/virulences are known by microarray and WGS (whole genome sequencing).

In order to use the primers and probes under identical conditions, they were predicted using the same primer3_parameters.txt file. This means that the target Tm values and optimal sizes were identical for each primer and probe design. All parameters used can be found in the Supplementary Materials under “Content of the primer3 parameter file”.

The predicted primer and probe sequences are listed in Table 5. In order to ensure a consistently high quality of the primers and probe pairs, small modifications were made to one primer and two probes after initial tests. The idea behind this was that primers and probes with guanine or cytosine at the 5' or 3' end bind more easily to the target sequence than primers and probes with adenosine or thymine. Different binding efficiencies are caused by the differing number of hydrogen bonds formed between the bases. Three hydrogen bonds between G and C, and two hydrogen bonds between A and T.

Table 5. The table shows a complete list of predicted primers and probes. All probes are modified with a 5' CY5 and a 3' NFQ-MGB quencher. The Tm values are calculated using the Santa Lucia formula. Manual modifications have been made to the predicted sequences for different reasons: ¹ Three nucleotides have been added to the 5' end, and one was removed from the 3' end to increase the Tm-value to fit the target specifications better. It was impossible to predict this sequence with Primer3 because Primer3 is limited to a probe length of 36 nucleotides, and the designed probe is 37nt long. ² Two nucleotides have been added to the 5' end. Three nucleotides have been removed from the 3' end of the primer to ensure better binding qualities. ³ One nucleotide was removed from the 5' and the 3' ends to increase the binding qualities of the probe.

Name	Sequence 5'-3'	Tm-Value	Product Size
eno_0_fwd	TGCCAATTATTACAGATGTTTACGC	55.8	130
eno_1_rev	CCAGTTGAAGCACCTGATGG	56.15	130
eno_0_probe	CGCGAAGTCTTAGACTCTCGTGGTAACCCAAGTGT	68.54	130
mecC_4_fwd	TTTGCCCGCATTGCATTAGC	57.8	178
mecC_4_rev	CTAGTATCTCGCCTTGGCCA	56.26	178
mecC_4_probe	TGCAAGATTTGGGAATCGGTGAAAATATCCCG	64.59	178
fusC_0_fwd	CGGACTTTATTACATCGATTGACG	55.46	281
fusC_0_rev	TGAAATTCGCCATATATACCTTCG	54.93	281
fusC_0_probe	CCAAGATTTTGAAATACCTTCATCAAGTCAACTGG	62.12	281
entA_0_fwd	CCACCCGCACATTGATAACC	56.6	300
entA_0_rev	TGGTAGCGAGAAAAGCGAAG	55.66	300
entA_0_probe	TGCCATAAGCTGTTCCCTGCAATTCAGACT	65.34	300
lukF-PV_P83_0_fwd	TGCCCATATTAGCACGTGGT	56.73	191
lukF-PV_P83_0_rev	TGATGTGTGTGTTGCTCTCT	54.39	191
lukF-PV_P83_0_probe ¹	GGATCGGTATGAAAATTTTGGAAACAACCTTGCACTGG	65.5	191
lukF-PV_2_fwd ²	CAATTGCATTGCTTTTGCTATCC	55.18	111
lukF-PV_6_rev	TGATGTTGCAGTTGTTTTGTACA	54.97	111
lukF_PV_2_probe ³	GATGCAGCTCAACATATCACACCTGTAAGTGAG	64.17	111
entB_0_fwd	ACGTAGATGTGTTGGAGCTAA	54.76	266
entB_0_rev	CACCAAATAGTGACGAGTTAGGT	55.47	266
entB_0_probe	TGTATGGTGGTGTAACTGAGCATAATGGAAACCA	64.65	266

3.3. Primer and Probe Evaluation

The functionality of the primers and probes was compared to the expected result based on the genome sequences of each reference strain used. For this purpose, the presence of the corresponding target genes was checked. If the gene was present/not present, the expected value (Exp.) was defined as POS/NEG. This expected value was then compared with the qPCR results, see Table 2. The detailed results of the dilution series can be found in the Supplementary Materials.

The comparison between expected and measured qPCR results in Table 2 shows a perfect match, corresponding to an accuracy of 100%, as all expected positive and all expected negative results could be perfectly replicated by qPCR with the designed consensus primers, demonstrating the potential of the primers and probes designed with the ConsensusPrime pipeline.

4. Discussion

4.1. Pitfalls and Challenges in Consensus Primer Design

If the pipeline is not able to predict primers and probes, there may be several reasons for this. One common cause is that the `--consensusthreshold` is too restrictive if the alignment is not sufficiently homogeneous, meaning that the areas passed to primer3 for the prediction of the primers/probes are insufficient. This problem can be recognized by the consensus regions already output in the terminal or the `consensusregions_alignment.fna` file. Other possible causes may be the target parameters for Tm values or GC content set in the `primer3_parameter.txt` file. If this is the case, the pipeline displays detailed information about filtering the processed primers/probes directly in the terminal or in the `.html` report file. Primer3 details why potential primers/probes were excluded from the selection, including the unsuitable length of the primer/probe sequence or the target product and Tm values outside the target temperatures. If this is the case, the design parameters must be adjusted, or primers/probes must be searched for in less homologous regions by adjusting the `--consensusthreshold` parameter. If no adjustments lead to appropriate primers/probes, a different target gene may need to be selected.

4.2. Consensus Primer vs. Degenerate Primers

Even if the initial hurdle in the form of multiple sequences and the adjustment of additional parameters is slightly higher than for the design of primers and probes on just one sequence, this study impressively shows how high quality and versatile the primers and probes designed based on multiple sequences are. Another aspect of the sequences designed in homologous regions is that the sequences designed in this way will also work well in the future because they have target areas that have a lower mutation rate than inhomogeneous areas. The ConsensusPrime Pipeline tries to keep this entry barrier as minimal and straightforward as possible by offering helpful user filter functions to improve the quality of the multiple sequences entered effectively. Furthermore, the pipeline allows the prediction of primers and probes under identical target parameters to ensure their use in more extensive assays and multiplex applications.

The symbiosis of alignment filters, primer/probe design, documentation, and visualization are missing from the tools already available, such as Genefisher [30]. In addition, the ConsensusPrimer pipeline is available as a download and can be used locally, which avoids the problem of web services such as CODEHOP-PCR [31] that are no longer maintained or available. The source code can also be viewed and modified by anyone. Another widely used method to design primers/probes for multiple sequence alignments is degenerate primers. However, degenerate primers and probes are just multiple designed sequences to cover possible mismatches and variable sequences. However, this leads to lower specificity and increases the number of probes that may be limited in assays.

In addition to the higher initial effort of collecting homologous sequences, consensus primers have a disadvantage. If sequences differ at a position relevant to primer and probe design, the designed primers and probes will not perfectly fit all target sequences.

If this is necessary and more primers and probes are not a problem, degenerate primers and probes are the better alternative, as designed by DPPrimer [32]. These degenerate probes and primers differ at the relevant position and fit the target sequences perfectly. Choosing the proper method for the target application is left to the user. Detecting the sequences generated in this way would be a possible improvement to the current pipeline. The ability to predict degenerate primers and probes under certain circumstances would also be a valuable expansion for the pipeline.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/biomedinformatics4020068/s1>. The supplementary file contains the following information: 1. The content of the primer3 parameter file used for the primer and probe design. 2. Information about the efficiency of each primer and probe set. 3. The detailed results of the dilution series for each primer and probe set against every strain.

Author Contributions: M.C. predicted the primer and probes using the pipeline and wrote the manuscript. M.R. helped edit the paper and analyzed the sequencing results. C.D. sequenced the genomes and evaluated them. S.D.B. analyzed the sequencing results and helped edit the paper. S.M. helped with the conceptual development, editing of the paper, and selection of target genes. A.R. performed and analyzed the qPCRs. R.E. helped with the conceptual development and editing of the paper. All authors revised and corrected the various drafts and approved the final version of the manuscript.

Funding: ADA (13GW0456C: BMBF): Adaptable Decentralized Diagnostics for Veterinary and Human Medicine. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The ConsensusPrime source code is accessible and available on github (<https://github.com/mcollatz/ConsensusPrime>, accessed on 11 January 2024). Further details about the pipeline can be found in the associated publication at: <https://www.mdpi.com/2673-7426/2/4/41>, accessed on 11 January 2024. All genome sequences of the newly sequenced strains have been published to GenBank.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yoon, H.; Leitner, T. PrimerDesign-M: A multiple-alignment based multiple-primer design tool for walking across variable genomes. *Bioinformatics* **2015**, *31*, 1472–1474. [[CrossRef](#)]
2. Wingo, T.S.; Kotlar, A.; Cutler, D.J. MPD: Multiplex primer design for next-generation targeted sequencing. *BMC Bioinform.* **2017**, *18*, 14. [[CrossRef](#)] [[PubMed](#)]
3. Hendling, M.; Pabinger, S.; Peters, K.; Wolff, N.; Conzemius, R.; Barišić, I. Oli2go: An automated multiplex oligonucleotide design tool. *Nucleic Acids Res.* **2018**, *46*, W252–W256. [[CrossRef](#)] [[PubMed](#)]
4. Collatz, M.; Braun, S.D.; Monecke, S.; Ehricht, R. ConsensusPrime—A Bioinformatic Pipeline for Ideal Consensus Primer Design. *BioMedInformatics* **2022**, *2*, 637–642. [[CrossRef](#)]
5. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)] [[PubMed](#)]
6. Zielezinski, A.; Vinga, S.; Almeida, J.; Karlowski, W.M. Alignment-free sequence comparison: Benefits, applications, and tools. *Genome Biol.* **2017**, *18*, 186. [[CrossRef](#)] [[PubMed](#)]
7. Fuda, C.; Fisher, J.; Mobashery, S. β -Lactam resistance in *Staphylococcus aureus*: The adaptive resistance of a plastic genome. *Cell. Mol. Life Sci.* **2005**, *62*, 2617–2633. [[CrossRef](#)] [[PubMed](#)]
8. Tristan, A.; Ying, L.; Bes, M.; Etienne, J.; Vandenesch, F.; Lina, G. Use of multiplex PCR to identify *Staphylococcus aureus* adhesins involved in human hematogenous infections. *J. Clin. Microbiol.* **2003**, *41*, 4465–4467. [[CrossRef](#)]
9. Roux, C.M.; DeMuth, J.P.; Dunman, P.M. Characterization of components of the *Staphylococcus aureus* mRNA degradosome holoenzyme-like complex. *J. Bacteriol.* **2011**, *193*, 5520–5526. [[CrossRef](#)]
10. Shore, A.C.; Deasy, E.C.; Slickers, P.; Brennan, G.; O’Connell, B.; Monecke, S.; Ehricht, R.; Coleman, D.C. Detection of staphylococcal cassette chromosome mec type XI carrying highly divergent *mecA*, *mecI*, *mecR1*, *blaZ*, and *ccr* genes in human clinical isolates of clonal complex 130 methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **2011**, *55*, 3765–3773. [[CrossRef](#)]

11. García-Álvarez, L.; Holden, M.T.; Lindsay, H.; Webb, C.R.; Brown, D.F.; Curran, M.D.; Walpole, E.; Brooks, K.; Pickard, D.J.; Teale, C.; et al. Methicillin-resistant *Staphylococcus aureus* with a novel mecA homologue in human and bovine populations in the UK and Denmark: A descriptive study. *Lancet Infect. Dis.* **2011**, *11*, 595–603. [[CrossRef](#)] [[PubMed](#)]
12. O’neill, A.; McLaws, F.; Kahlmeter, G.; Henriksen, A.; Chopra, I. Genetic basis of resistance to fusidic acid in staphylococci. *Antimicrob. Agents Chemother.* **2007**, *51*, 1737–1740. [[CrossRef](#)] [[PubMed](#)]
13. Chen, H.-J.; Hung, W.-C.; Tseng, S.-P.; Tsai, J.-C.; Hsueh, P.-R.; Teng, L.-J. Fusidic acid resistance determinants in *Staphylococcus aureus* clinical isolates. *Antimicrob. Agents Chemother.* **2010**, *54*, 4985–4991. [[CrossRef](#)]
14. Chen, C.-M.; Huang, M.; Chen, H.-F.; Ke, S.-C.; Li, C.-R.; Wang, J.-H.; Wu, L.-T. Fusidic acid resistance among clinical isolates of methicillin-resistant *Staphylococcus aureus* in a Taiwanese hospital. *BMC Microbiol.* **2011**, *11*, 98. [[CrossRef](#)] [[PubMed](#)]
15. Farrell, D.J.; Castanheira, M.; Chopra, I. Characterization of global patterns and the genetics of fusidic acid resistance. *Clin. Infect. Dis.* **2011**, *52* (Suppl. S7), S487–S492. [[CrossRef](#)]
16. Ito, T.; Katayama, Y.; Asada, K.; Mori, N.; Tsutsumimoto, K.; Tiensasitorn, C.; Hiramatsu, K. Structural comparison of three types of staphylococcal cassette chromosome mec integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **2001**, *45*, 1323–1336. [[CrossRef](#)]
17. Oliveira, D.C.; Wu, S.W.; de Lencastre, H. Genetic organization of the downstream region of the mecA element in methicillin-resistant *Staphylococcus aureus* isolates carrying different polymorphisms of this region. *Antimicrob. Agents Chemother.* **2000**, *44*, 1906–1910. [[CrossRef](#)]
18. Elements IWGotCoSCC. Classification of staphylococcal cassette chromosome mec (SCCmec): Guidelines for reporting novel SCCmec elements. *Antimicrob. Agents Chemother.* **2009**, *53*, 4961. [[CrossRef](#)]
19. Spaan, A.N.; Henry, T.; Van Rooijen, W.J.; Perret, M.; Badiou, C.; Aerts, P.C.; Kemmink, J.; de Haas, C.J.; van Kessel, K.P.; Vandenesch, F.; et al. The staphylococcal toxin Pantone-Valentine Leukocidin targets human C5a receptors. *Cell Host Microbe* **2013**, *13*, 584–594. [[CrossRef](#)] [[PubMed](#)]
20. Zou, D.; Kaneko, J.; Narita, S.; Kamio, Y. Prophage, ϕ PV83-pro, carrying Pantone-Valentine leukocidin genes, on the *Staphylococcus aureus* P83 chromosome: Comparative analysis of the genome structures of ϕ PV83-pro, ϕ PVL, ϕ 11, and other phages. *Biosci. Biotechnol. Biochem.* **2000**, *64*, 2631–2643. [[CrossRef](#)]
21. Yamada, T.; Tochimarui, N.; Nakasuji, S.; Hata, E.; Kobayashi, H.; Eguchi, M.; Kaneko, J.; Kamio, Y.; Kaidoh, T.; Takeuchi, S. Leukotoxin family genes in *Staphylococcus aureus* isolated from domestic animals and prevalence of lukM–lukF-PV genes by bacteriophages in bovine isolates. *Vet. Microbiol.* **2005**, *110*, 97–103. [[CrossRef](#)] [[PubMed](#)]
22. Kaneko, J.; Muramoto, K.; Kamio, Y. Gene of LukF-PV-like component of Pantone-Valentine leukocidin in *Staphylococcus aureus* P83 is linked with lukM. *Biosci. Biotechnol. Biochem.* **1997**, *61*, 541–544. [[CrossRef](#)]
23. Betley, M.J.; Mekalanos, J.J. Nucleotide sequence of the type A staphylococcal enterotoxin gene. *J. Bacteriol.* **1988**, *170*, 34–41. [[CrossRef](#)] [[PubMed](#)]
24. Borst, D.W.; Betley, M.J. Phage-associated differences in staphylococcal enterotoxin A gene (sea) expression correlate with sea allele class. *Infect. Immun.* **1994**, *62*, 113–118. [[CrossRef](#)]
25. Kohler, P.L.; Greenwood, S.D.; Nookala, S.; Kotb, M.; Kranz, D.M.; Schlievert, P.M. *Staphylococcus aureus* isolates encode variant staphylococcal enterotoxin B proteins that are diverse in superantigenicity and lethality. *PLoS ONE* **2012**, *7*, e41157. [[CrossRef](#)]
26. Johler, S.; Sihto, H.-M.; Macori, G.; Stephan, R. Sequence variability in staphylococcal enterotoxin genes seb, sec, and sed. *Toxins* **2016**, *8*, 169. [[CrossRef](#)] [[PubMed](#)]
27. Abril, A.; Gonzalez-Villa, T.; Barros-Velázquez, J.; Cañas, B.; Sánchez-Pérez, A.; Calo-Mata, P.; Carrera, M. *Staphylococcus aureus* exotoxins and their detection in the dairy industry and mastitis. *Toxins* **2020**, *12*, 537. [[CrossRef](#)]
28. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)]
29. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
30. Giegerich, R.; Meyer, F.; Schleiermacher, C. (Eds.) *GeneFisher—Software Support for the Detection of Postulated Genes*; ISMB: San Diego, CA, USA, 1996.
31. Rose, T.M.; Henikoff, J.G.; Henikoff, S. CODEHOP (COnsensus-DEgenerate hybrid oligonucleotide primer) PCR primer design. *Nucleic Acids Res.* **2003**, *31*, 3763–3766. [[CrossRef](#)]
32. Gahoi, S.; Arya, L.; Anil, R. DPPrimer—A Degenerate PCR Primer Design Tool. *Bioinformatics* **2013**, *9*, 937. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.