Research Article

Family-Specific Degenerate Primer Design: A Tool to Design Consensus Degenerated Oligonucleotides

Javier Alonso Iserte, 1 Betina Ines Stephan, 1 Sandra Elizabeth Goñi, 1 Cristina Silvia Borio, 1 Pablo Daniel Ghiringhelli, 2 and Mario Enrique Lozano 1

1 LIGBCM-Área Virosis Emergentes y Zoonóticas, Universidad Nacional de Quilmes, B1876BXD Buenos Aires, Argentina
2 LIGBCM-Área Virosis de Insectos, Universidad Nacional de Quilmes, B1876BXD Buenos Aires, Argentina

Correspondence should be addressed to Javier Alonso Iserte; jiserte@unq.edu.ar

Received 15 October 2012; Accepted 11 January 2013

Academic Editor: Goetz Laible

Copyright © 2013 Javier Alonso Iserte et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Designing degenerate PCR primers for templates of unknown nucleotide sequence may be a very difficult task. In this paper, we present a new method to design degenerate primers, implemented in family-specific degenerate primer design (FAS-DPD) computer software, for which the starting point is a multiple alignment of related amino acids or nucleotide sequences. To assess their efficiency, four different genome collections were used, covering a wide range of genomic lengths: Arenavirus (10 × 10^4 nucleotides), Baculovirus (0.9 × 10^5 to 1.8 × 10^5 bp), Lactobacillus sp. (1 × 10^6 to 2 × 10^6 bp), and Pseudomonas sp. (4 × 10^6 to 7 × 10^6 bp). In each case, FAS-DPD designed primers were tested computationally to measure specificity. Designed primers for Arenavirus and Baculovirus were tested experimentally. The method presented here is useful for designing degenerate primers on collections of related protein sequences, allowing detection of new family members.

1. Introduction

The polymerase chain reaction (PCR), one of the most important analytical tools of molecular biology, allows a highly sensitive detection and specific genotyping of environmental samples, specially important in the metagenomic era [1]. A large list of genome typing applications includes arbitrarily primed PCR (AP-PCR), random amplified primed DNAs (RAPDs), PCR restriction fragment length polymorphism (PCR-RFLP), and direct amplification of length polymorphism (DALP). All of these techniques require a high quality and purity of the specific target template, because any available DNA could be substrate for the amplification step. In view of this, genotyping procedures of large genomes or complex samples are more reliable if they are based on DNA amplification using specific oligonucleotides. Therefore, primer design is crucial for efficient and successful amplification.

Several primer design programs are available (e.g., OLIGO [6], OSP [7, 8], Primer Master [9], PRIDE [10], Primer3 [11], among others). Regardless of each computational working strategy, all of these use a set of common criteria (e.g., G/C content, melting temperature, etc.) to evaluate the quality of primer candidates in a specific target region selected by the user. Alternative programs are aimed at more specific purposes, such as selection of primers that bind to conserved genomic regions based on multiple sequence alignments [12, 13], primer design for selective amplification of protein-coding regions [14], oligonucleotide design for site-directed mutagenesis [15], and primer design for hybridization [16]. Usually, the design of truly specific primers requires the information of the complete nucleotide sequence. This is the starting point for most of the programs described in the literature. However, the need of designing specific primers is not always accompanied by the complete knowledge of the target genome sequence.

A primer, or more generally any DNA sequence, is called specific if it represents a unique sequence and is called degenerate if it represents a collection of unique sequences. For example, the amino acid sequence “YHP” could be
be detected. However, at the same time, it diminishes the proportion of primer specific for a given sequence. Therefore, we decided to be very strict in the search of conserved regions and minimize the amount of degeneracy incorporated at this end. If the input set of sequences is sufficiently large, it is highly probable that a region identified as conserved among all known sequences will likewise be conserved in any new member of the family.

2. Scoring and Primer Search Strategy

The method presented here can be used starting with DNA or protein sequence alignments (Figure I(a)). If the input was DNA, sequences were aligned to obtain one global degenerate DNA consensus. If the input was a protein alignment, each protein of the alignment is backtranslated into a degenerate DNA sequence. All the degenerate DNA sequences were combined in one global degenerate DNA consensus. This consensus sequence covers all the putative input sequences that could be the origin of each protein sequence (Figure I(b)). Also, the consensus sequence may code for amino acids that were not detected in the known sequences. This is inevitable given the kind of degeneracy of the genetic code.

Then, the degenerate consensus sequence was analyzed using an overlapping window-based strategy. The window length corresponds to the required oligonucleotide length, and each window corresponds to a putative primer. For each candidate primer a score is calculated. In the first place, for each position of a candidate primer a position score \( S_{pi} \) was calculated using (1):

\[
S_{pi} = 1 - \log_{10} \left( N_{D_i} \right),
\]

where \( N_{D_i} \) is the degeneracy value at the position \( i \) of the oligonucleotide \( (1 \leq i \leq n, \text{where } n \text{ is the length of the primer}) \). \( N_{D_i} \) is 1 for "A, C, G or T," 2 for "K, M, R, S, W or Y," 3 for "B, D, H or V," and 4 for "N." This expression takes a value of 1 for nondegenerate bases and decreases for more degenerated bases. On the other hand, it is known that in PCR reactions, the 3' end of the primer is more important than the 5' end. The region of the 3' end of the primer must be as little degenerated as possible. Therefore, a good annealing at this end is imperative in order to minimize unspecific amplifications. Considering this, the value of \( S_{pi} \) is multiplied by a weighting value \( W_{pi} \) defined by a straight line function that increases as it comes closer to the 3' end (2):

\[
W_{pi} = pA + \left( N_{y} - pA \right) \frac{i \times (N_{x} - pA)}{N_{x}},
\]

where \( i \) is the position from the 5' end along the oligonucleotide \( (1 \leq i \leq n, \text{where } n \text{ is the length of the primer}) \) and \( pA, N_{y}, \text{and } N_{x} \) are user adjustable parameters defining the straight line function. \( pA \) is the axis intersection and \( (N_{y} - pA)/N_{x} \) is the slope. Default values for \( pA, N_{y}, \text{and } N_{x} \) are 0, 1, and 1, respectively. Changing them will permit them to be more or less strict about including degenerations closer to the 3' end of the primer. Increasing \( pA \) or \( N_{x} \), or
Figure 1: Minimum degenerated sequence generation. (a) Diagram of the general strategy used. (b) Sample protein alignment showing an example for the steps of the strategy diagram. Each sequence is computationally backtranslated to hypothetical nucleic acid sequences. IUPAC codes were used to show ambiguous positions. These sequences are piled up in order to get the degenerated consensus sequence. Numbers below this indicate the degeneration value of each position.

decreasing $N_y$, results in lesser stringency on the designed primer. Finally, to obtain a scaled global score ($S_g$), the result of $W_{p_i} \times S_{p_i}$ is divided by the maximum possible score ($M_s$, (3)). Global normalized score ($S_g$) was calculated according to (4). In this way, $S_g$ value varies from 0 to 1. Maximum score is obtained when the value of the $S_{p_i}$ is 1 for each position. Therefore, $N_D$ must also be 1 too, and this only happened with nondegenerated primers:

$$M_s = n \times p_A + \frac{(n + 1) \times n \times (N_y - p_A)}{2 \times N_x},$$

$$S_g = \frac{\sum_{i=1}^n S_{p_i} \times W_{p_i}}{M_s},$$

3. Methods

3.1. Alignment and Sequence Comparison Tools. For global alignment of protein sequences, the program ClustalW 1.83 [24] was used with default parameters. Local alignments of proteins against genomes were made using stand-alone Blast 2.2.13 [25] with default parameters. Oligonucleotide match searches were made with specifically developed tools written in C language.

3.2. Sequence Data. Several sets of sequences were used in the tests of the program, for designing and comparison of the primer sequences against genomes. All sequences GenBank's accession numbers are presented in Table 1.

3.3. Filtering Primers. In addition to the scoring process, FAS-DPD can optionally filter the primers individually according to common criteria: melting point temperature (estimated using Santalucía's method [26]), G + C content, 5' versus 3' stability, presence of tandem repeats of the same base occurring at 3' end or any place in the sequence, presence of a degenerated position at the 3' end, and formation of homodimer structures. Also, primer pairs can be filtered according to amplification product size, melting point temperature compatibility, G + C content compatibility, and formation of heteroduplex structures.

3.4. PCR Amplification. The PCR conditions used in all experiments follow a common protocol. The reaction mix contained 1X Taq DNA polymerase buffer (Productos Biológicos, Argentina), 0.2 mM dNTPs, 0.5 μM of each primer, 20 μM template, and different concentration of MgCl2 and dimethyl sulfoxide (DMSO) in different reactions. The MgCl2 was used from 2 mM to 3 mM, and DMSO was used from 0% (v/v) to 5% (v/v). The reactions were performed in a total volume of 10 μL, and the thermal profile consisted of an initial denaturation step of 94°C for 2 min, followed by 35 cycles of denaturation/annealing/extension steps. The denaturation step was at 92°C for 10 seconds, the temperature of the annealing step was not the same in all experiments, varying from 45°C to 60°C, and the time was always 15 seconds (see Figure 4). The extension step was at 72°C; the time of this step was 15 seconds. In all cases, one of the primers is specific for the template, while the other primer was designed by the method described in this work. The last step was a final extension of 5 minutes at 72°C. For Junin Virus, the template used was a plasmid containing a copy of cDNA of JUNV S genomic segment. For Baculovirus, the template was a plasmid containing a fragment of Anticarsia gemmatalis.
| Acc. number | Sequence description | Accum. number | Sequence description |
|-------------|---------------------|--------------|---------------------|
| **Arenaviruses** | **Accession numbers and brief description** | | |
| AY129248.1 | Machupov. v. st. Carvallo | U41071.1 | Sabia v. |
| AF485260.1 | Machupov. v. st. Carvallo | EU260463.1 | Chapare v. st. 810419 |
| AY924206.1 | Machupov. v. st. MARU-216606 | AY081210.1 | Allpahuayo v. CLHP-2098 |
| AY924202.1 | Machupov. v. st. Chicava | AY012686.1 | Allpahuayo v. from Peru |
| AY624355.1 | Machupov. v. st. Chicava | AY012687.1 | Allpahuayo v. st. CLHP-2472 |
| AY924205.1 | Machupov. v. st. MARU-216606 | AF485262.1 | Pirital v. |
| AY619645.1 | Machupov. v. st. Mallele | AF277659.1 | Pichinde v. |
| AY924203.1 | Machupov. v. st. 9430084 | M16735.1 | Pichinde v. |
| AY924208.1 | Machupov. v. st. MARU-249121 | AF485261.1 | Parana v. st. 12056 |
| AY924204.1 | Machupov. v. st. 200002427 | AF512829.1 | Parana v. st. 10256 |
| AY924207.1 | Machupov. v. st. MARU-222688 | AF512831.1 | Flexal v. st. BeAn 293022 |
| AY571959.1 | Machupov. v. st. 9350537 | AF485257.1 | Flexal v. st. Pinheiro |
| AY746535.1 | Junin v. st. Candid-1 | AF512831.1 | Flexal v. st. BeAn 293022 |
| AY385023.2 | Junin v. st. XJ13 | AF512830.1 | Latino v. st. MARU 10924 |
| AY619641.1 | Junin v. st. Rumero | AF485259.1 | Latino v. st. Maru 10924 |
| DI0072.2 | Junin v. st. MC2 | U34248.1 | Oliveros v. |
| M20304.1 | Tacaribe v. | AY847350.1 | LCM v. st. Armstrong 53b |
| AF485256.1 | Amapari v. st. BeAn 70563 | M20869.1 | LCM v. st. Armstrong 53b |
| AF512834.1 | Amapari v. st. BeAn 70563 | EU136038.1 | Dandenong v. is. 0710-2678 |
| AF512832.1 | Cupixi v. st. BeAn 11930 | DQ328874.1 | Mopeia v. st. Mozambique |
| AY129247.1 | Guanarito v. st. INH-95551 | DQ328877.1 | Ippy v. st. Dak-An-B-188-d |
| AF485258.1 | Guanarito v. st. INH-95551 | X52400.1 | Nigeria Lassa v. |
| AY947548.1 | Guanarito v. st. CVH-960101 | AY628206.1 | Lassa v. st. Weller |
| AY924392.1 | Bear Canyon v. st. AV 98470029 | AY628201.1 | Lassa v. st. Macenta |
| AY924391.1 | Bear Canyon v. st. AV 980070039 | AY628205.1 | Lassa v. st. ZI48 |
| AF512833.1 | Bear canyon v. st. A0060209 | J04324.1 | Lassa v. st. Josiah |
| DQ865244.1 | Catarina v. st. AV A0400135 | AY772168.1 | Mopeia Lassa reasortant 29 |
| DQ865245.1 | Catarina v. st. AV A0400212 | AY628203.1 | Lassa v. st. Josiah |
| EU123328.1 | Skinner Tank v. st. AV D1000090 | AF181853.1 | Lassa v. st. LP |
| EU123331.1 | North American arenav. st. AV 96010024 | AY628207.1 | Lassa v. st. Pinneo |
| EU123330.1 | North American arenav. st. AV 96010051 | AY628208.1 | Lassa v. st. Acar-3080 |
| AF288063.1 | Whitewater Arroyo v. st. 9310135, | AF181854.1 | Lassa v. st. 803213 |
| AF485264.1 | Whitewater Arroyo v. st. 9310141 | AF485261.1 | Mobala v. st. ACAR-3080-MRC5-P2 |
| EU123329.1 | North American arenav. st. AV D1240007 | M33879.1 | Mopeia v. st. AN-21366 |
| AF485263.1 | Tamiami v. st. CDC W-10777 | AY772170.1 | Mopeia v. st. AN-20410 |
| AF512828.1 | Tamiami v. st. W 10777 | AY628207.1 | Mopeia v. st. 20410 |
| **Baculoviruses** | **Accession numbers and brief description** | | |
| AP066270.1 | Adoxophyes homnai nucleopolyhedrovirus DNA | X77048.1 | Cryptophlebia leucotreta granulosis |
| AF547984.1 | Adoxophyes orana granulovirus | X79569.1 | Cryptophlebia leucotreta granulosis |
| NC_005839.2 | Agrotis segetum granulovirus | NC_002816.1 | Cydia pomonella granulovirus |
| L22858.1 | Autographa cannabifera nucleopolyhedrovirus clone C6 | NC_003083.1 | Epiphyas postvittana NPV |
| L33801.1 | Bombbyx mori nuclear polyhedrosis virus isolate T3 | NC_002654.2 | Helicoverpa armigera |
| NC_005137.2 | Choristoneura fumiferana DEF MNPV | AF081810.1 | Lymna antra dispar |
| NC_004778.3 | Choristoneura fumiferana MNPV | NC_003259.1 | Mamestra configurata NPV-A |
| AY864330.1 | Chrysodeixis chalcites NPV | U75930.2 | Orgyia pseudotsugata MNPV |
| AY456389.1 | Chrysodeixis chalcites NPV | AF499596.1 | Phthorimaea opercula granulovirus |
| AY456390.1 | Chrysodeixis chalcites NPV | NC_002593.1 | Plutella xylostella granulovirus |
| AY545786.1 | Chrysodeixis chalcites NPV | NC_004323.1 | Rachiplasia ou MNPV |
| AY545787.1 | Chrysodeixis chalcites NPV | NC_002169.1 | Spodoptera exigua MNPV |
| AY229987.1 | Cryptophlebia leucotreta granulovirus | NC_003012.1 | Spodoptera littoralis NPV |
Table 1: Continued.

| Acc. number | Sequence description                  | Acc. number | Sequence description                  |
|-------------|---------------------------------------|-------------|---------------------------------------|
| AY096241.1  | Cryptophlebia leucoatra granulovirus   | NC_007383.1 | Trichoplusia ni SNPV                  |
| AY096242.1  | Cryptophlebia leucoatra granulovirus   |             |                                       |

**Pseudomonas** sp. sequences

| Acc. number | Sequence description                  | Acc. number | Sequence description                  |
|-------------|---------------------------------------|-------------|---------------------------------------|
| NC_007492.2 | *Pseudomonas fluorescens* Pf0-1       | NC_004578.1 | *Pseudomonas syringae*                 |
| NC_005773.3 | *Pseudomonas syringae*                | NC_002947.3 | *Pseudomonas putida*                  |
| NC_004129.6 | *Pseudomonas fluorescens*             | NC_002516.2 | *Pseudomonas aeruginosa*               |
| NC_007005.1 | *Pseudomonas syringae*                |             |                                       |

**Lactobacillus** sp. sequences

| Acc. number | Sequence description                  | Acc. number | Sequence description                  |
|-------------|---------------------------------------|-------------|---------------------------------------|
| NC_005362.1 | *Lactobacillus johnsonii*             | NC_002662.1 | *Lactococcus lactis* subsp.           |
| NC_007576.1 | *Lactobacillus sakei* subsp.          | NC_004567.1 | *Lactobacillus plantarum*             |

Figure 2: Primer distribution along one ORF. A collection of the best scoring primers for the nucleoprotein of Arenavirus, comprised of 50 primers for the genomic sequence and 50 for the antigenomic sequence, were represented in the corresponding alignment position. The height of each point indicates the cumulative number of the selected primers corresponding at this position. The alignment was made with 71 arenavirus N protein sequences.

MNPV p74 gene. Sensitivity of the PCR assay was determined by dilution of cloned fragments from Junin virus [27] and Baculovirus template.

4. Results

4.1. Distribution of Generated Primers. The distribution of the resulting primers along the input sequence was analyzed. For this, the best one hundred primers obtained from a protein alignment were selected. For each position in the alignment, the number of the selected primers that correspond to this position was recorded (Figure 2). The test was repeated for different protein alignments.

The selected primers were located around a few hot spots in the alignment. This behavior indicates that there are generally few regions in a sequence alignment useful for degenerate primer design. Many primers found by the program are almost identical, shifting one or two bases between them, and located for most cases in a 30–40 base run. Similar results were obtained with all proteins tested.

4.2. Intragenomic Specificity and Score Analysis. Because it is possible that the best primers are not the less degenerated substrings in the collection of candidates, their specificity was tested. Also, it was necessary to get a more precise understanding of the score assigned by FAS-DPD in terms of specificity. To achieve this, the primers were compared with the complete genome sequences used to design them, looking for unspecific perfect matches.

For this task, a wide range of genome sizes was covered. Four collections of complete genome sequences were used: Arenavirus (genome in $10^4$ bases order), Baculovirus (genome in $10^5$ bases order), *Lactobacillus* (genome in $10^6$ bases order), and *Pseudomonas* (genome in $10^6$ bases order). For each set, a randomly selected genome was used as reference. Each annotated ORF of this genome was used to search related ORFs in the other genomes of the collection using the local Blast tool. The expected value of Blast was used to decide when two ORFs were related. When an ORF of the reference genome had a related one in all other genomes, all of them were aligned with ClustalW and used in further analysis.

Each resulting alignment was used as input for FAS-DPD to search primers. For each genome polarity the best fifty nonoverlapping primers were selected. This selection was made to avoid concentration of overrepresented, hot-spot-derived, high score primers. This allowed us to find a balanced set of primers, with high and low scores.

In order to find the relationship between the score calculated for each primer and its specificity, all the primers were compared with all the oligonucleotides of the same size derived from each genome, searching for perfect matches (Figure 3). The results were similar for the four systems despite their differences in genome size.

There is an inverse correlation between primer score and the number of unspecific perfect matches. But this correlation is not linear. The quantity of unspecific perfect matches of primers with a minimal score of 0.85 and their target genome was generally zero. The number of unspecific perfect matches grew enormously with lower primer scores.
Figure 3: Specificity of primers. Primers designed for all ORFs shared among each model organism used were compared against the complete set of genomes for perfect matches with oligonucleotides of the same length. Each point represents the number of perfect matches (in log₁₀ scale) of a primer in relation to its score. The length of the primers was 20 nucleotides. (a) Arenavirus genomes: 71 for S (small) RNA, 24 for L (large) RNA. (b) 22 Baculovirus genomes. (c) 5 Lactobacillus sp. genomes. (d) 7 Pseudomonas sp. genomes. (e) A set of primers for Lactobacillus sp. with scores between 0.85 and 0.90 were tested for nonperfect matches that could anneal unspecifically in PCR. Each bar represents the number of matches against the complete set of Lactobacillus genomes. The number below the bar indicates how many bases are shared.
4.3. Experimental Challenge. In addition to theoretic tests to determine the usefulness of FAS-DPD designed primers, experimental challenges were performed using Arenavirus and Baculovirus as models. The assay consisted in performing PCRs using a pair of primers, including a degenerated FAS-DPD designed primer and a standard nondegenerated primer (this allowed testing individually each designed primer), optimizing the reaction conditions and measuring its sensitivity.

For arenavirus, the primers were designed using sequences of 71 different GenBank records for the nucleoprotein (N protein) and the glycoprotein precursor (GPC protein). From the lists of the highest scored primers, three were randomly selected and synthesized for experimental evaluation, one for GPC (GR1058: RCNWHRTTNYCAARCYTT, score: 0.8596) and two for N (N527: GGNRYNSWNCCRAAYTGRTT, score: 0.8494; N918: NANRTTYTCRTANGGRTTNC, score: 0.8437) (Figure 4(a)).

Amplification reactions were performed using each of these primers together with the Arena primer CGCACCGGGAGTCAGTTT (score: 0.8596) and two for N (N527: GGNRYNSWNCCRAAYTGRTT, score: 0.8494; N918: NANRTTYTCRTANGGRTTNC, score: 0.8437) (Figure 4(a)).

4.4. Increment of Degeneration of FAS-DPD Designed Primers in relation to Minimum Degenerated Substring. The aim of FAS-DPD is to design universal degenerated primers that are not necessarily the less degenerated sequences of the collection of candidates. In order to know how much degeneration FAS-DPD designed primers acquire, another test was performed. Given an alignment of homologous ORFs, the degeneration was calculated for the highest scoring primer selected with FAS-DPD and for the minimum degenerated substring of the same length. Then, the ratio of these two values was obtained. The comparison was made with the complete set of ORF alignments used before (Arenavirus, Baculovirus, Pseudomonas, and Lactobacillus) (Figure 5). In more than 90% of the cases the increase of degeneration value is at most fourfold (e.g., changing “...A...” to “...N...” or “...A...A...” to “...R...W...”). Therefore, these primers...
lar biology teaches us that in the real world the specificity of the set of primers designed with FAS-DPD could be used to design generalized degenerate primers for detection of known or unknown members of gene families or organism families, including different types of pathogens. Also, this tool would allow a more efficient search for enzymes and other proteins with commercial or biotechnological importance, making for a faster and cheaper research process.

5. Discussion

In this work we presented a new algorithm, implemented in the FAS-DPD software, as an alternative strategy to solving DP problems. FAS-DPD was designed to use multiple alignments of proteins or nucleic acids as input data and constructs a consensus degenerate sequence from that, which is then used to design the putative primers.

The experimental background knowledge from molecular biology teaches us that in the real world the 3' ends of primers are key determinants of a successful amplification. FAS-DPD takes into account this property and incorporates special considerations in the global score calculation becoming more strict for the 3' end than for the 5' end.

The specificity of the set of primers designed with FAS-DPD was computationally tested with several collections of whole genomes, ranging from $10^4$ bp to $10^6$ bp. The restriction to higher lengths was due to the lack of whole genome collections for genus of bigger sizes with several individuals. In all genome collections assayed the results showed the same behavior; there is a relationship between the score value and the number of unspecific perfect matches. This analysis allows us to suggest a cut-off score (0.85) for primers that could be more successful.

PCRs were successfully performed on arenaviral and baculoviral models. For arenavirus, the designed GPC or N primers were used with the universal Arena primer [30]. For Baculovirus, the designed p74 primer was used with a specific p74 primer [28]. Each reaction was tested in different conditions in order to optimize its yield.

FAS-DPD software is licensed under GNU General Public License Version 3 and is available at http://www.github.com/javieriserte/fas-dpd.

In general, the results suggest that FAS-DPD could be used to design generalized degenerate primers for detection of known or unknown members of gene families or organism families, including different types of pathogens. Also, this tool would allow a more efficient search for enzymes and other proteins with commercial or biotechnological importance, making for a faster and cheaper research process.

References

[1] K. Nelson, Metagenomics as a Tool to Study Biodiversity, ASM Press, Washington, DC, USA, 2008.
[2] J. Welsh and M. McClelland, “Fingerprinting genomes using PCR with arbitrary primers,” Nucleic Acids Research, vol. 18, no. 24, pp. 7213–7218, 1990.
[3] J. G. K. Williams, A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey, “DNA polymorphisms amplified by arbitrary primers are useful as genetic markers,” Nucleic Acids Research, vol. 18, no. 22, pp. 6531–6535, 1990.
[4] W. C. Nichols, S. E. Lyons, J. S. Harrison, R. L. Cody, and D. Ginsburg, “Severe von Willebrand disease due to a defect at the level of von Willebrand factor mRNA expression: detection by exonic PCR-restriction fragment length polymorphism analysis,” Proceedings of the National Academy of Sciences of the United States of America, vol. 88, no. 9, pp. 3857–3861, 1991.
[5] E. Desmarais, I. Lanneluc, and J. Laguel, “Direct amplification of length polymorphisms (DALP), or how to get and characterize new genetic markers in many species,” Nucleic Acids Research, vol. 26, no. 6, pp. 1458–1465, 1998.
[6] W. Rychlik and R. E. Rhoads, “A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA,” Nucleic Acids Research, vol. 17, no. 21, pp. 8543–8551, 1989.
[7] L. Hillier and P. Green, “OSP: a computer program for choosing PCR and DNA sequencing primers,” PCR Methods and Applications, vol. 1, no. 2, pp. 124–128, 1991.
[8] P. Li, K. C. Kupfer, C. J. Davies, D. Burbree, G. A. Evans, and H. R. Garner, “PRIMO: a primer design program that applies base quality statistics for automated large-scale DNA sequencing,” Genomics, vol. 40, no. 3, pp. 476–485, 1997.
[9] V. Proutski and E. C. Holmes, “Primer Master: a new program for the design and analysis of PCR primers,” Computer Applications in the Biosciences, vol. 12, no. 3, pp. 253–255, 1996.
S. Haas, M. Vingron, A. Poustka, and S. Wiemann, “Primer design for large scale sequencing,” *Nucleic Acids Research*, vol. 26, no. 12, pp. 3006–3012, 1998.

S. Rozen and H. Skaltsky, “Primer3 on the WWW for general users and for biologist programmers,” *Methods in molecular biology*, vol. 132, pp. 365–386, 2000.

A. Gibbs, J. Armstrong, A. M. Mackenzie, and G. F. Weiller, “The GPIPRIME package: computer programs for identifying the best regions of aligned genes to target in nucleic acid hybridisation-based diagnostic tests, and their use with plant viruses,” *Journal of Virological Methods*, vol. 74, no. 1, pp. 67–76, 1998.

M. D. Gadberry, S. T. Malcomber, A. N. Doust, and E. A. Kellogg, “Primeclade: a flexible tool to find conserved PCR primers across multiple species,” *Bioinformatics*, vol. 21, no. 7, pp. 1263–1264, 2005.

C. E. López-Nieto and S. K. Nigam, “Selective amplification of protein-coding regions of large sets of genes using statistically designed primer sets,” *Nature Biotechnology*, vol. 14, no. 7, pp. 857–861, 1996.

A. Turchin and J. F. Lawler, “The primer generator: a program that facilitates the selection of oligonucleotides for site-directed mutagenesis,” *BioTechniques*, vol. 26, no. 4, pp. 672–676, 1999.

D. Hyndman, A. Cooper, S. Pruzinsky, D. Coad, and M. Mitsuhashi, “Software to determine optimal oligonucleotide sequences based on hybridization simulation data,” *BioTechniques*, vol. 20, no. 6, pp. 1090–1097, 1996.

C. Linhart and R. Shamir, “Degenerate primer design: theoretical analysis and the HYDEN program,” *Methods in Molecular Biology*, vol. 402, pp. 221–244, 2007.

C. Linhart and R. Shamir, “The degenerate primer design problem,” *Bioinformatics*, vol. 18, supplement 1, pp. S172–S180, 2002.

X. Wei, D. N. Kuhn, and G. Narasimhan, “Degenerate primer design via clustering,” *IEEE Computer Society Bioinformatics Conference*, vol. 2, pp. 75–83, 2003.

T. M. Rose, J. G. Henikoff, and S. Henikoff, “CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design,” *Nucleic Acids Research*, vol. 31, no. 13, pp. 3763–3766, 2003.

T. M. Rose, “CODEHOP-mediated PCR: a powerful technique for the identification and characterization of viral genomes,” *Virology Journal*, vol. 2, article 20, 2005.

R. Boyce, P. Chilana, and T. M. Rose, “iCODEHOP: a new interactive program for designing COnsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences,” *Nucleic Acids Research*, vol. 37, no. 2, pp. W222–W228, 2009.

S. Balla and S. Rajasekaran, “An efficient algorithm for minimum degeneracy primer selection,” *IEEE Transactions on Nanobioscience*, vol. 6, no. 1, pp. 12–17, 2007.

J. D. Thompson, D. G. Higgins, and T. J. Gibson, “CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice,” *Nucleic Acids Research*, vol. 22, no. 22, pp. 4673–4680, 1994.

F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

J. SantaLucia, “A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 4, pp. 1460–1465, 1998.