Metadata of the chapter that will be visualized online

Chapter Title	FastPCR Software for PCR, In Silico PCR, and Oligonucleotide Assembly and Analysis	
Copyright Year	2014	
Copyright Holder	Springer Science+Business Media, New York	
Corresponding Author	Family Name	**Kalendar**
	Particle	
	Given Name	**Ruslan**
	Suffix	
	Division	MTT/BI Plant Genomics Laboratory
	Organization	Institute of Biotechnology, University of Helsinki
	Address	65, 00014, Helsinki, Finland
	Organization	PrimerDigital Ltd
	Address	00790, Helsinki, Finland
Author	Family Name	**Lee**
	Particle	
	Given Name	**David**
	Suffix	
Author	Family Name	**Schulman**
	Particle	
	Given Name	**Alan H.**
	Suffix	
Abstract	This chapter introduces the software FastPCR as an integrated tools environment for PCR primer and probe design. It also predicts oligonucleotide properties based on experimental studies of PCR efficiency. The software provides comprehensive facilities for designing primers for most PCR applications and their combinations, including standard, multiplex, long-distance, inverse, real-time, group-specific, unique, and overlap extension PCR for multi-fragment assembly in cloning, as well as bisulphite modification assays. It includes a program to design oligonucleotide sets for long sequence assembly by the ligase chain reaction. The in silico PCR primer or probe search includes comprehensive analyses of individual primers and primer pairs. It calculates the melting temperature for standard and degenerate oligonucleotides including LNA and other modifications, provides analyses for a set of primers with prediction of oligonucleotide properties, dimer and G/C-quadruplex detection, and linguistic complexity, and provides a dilution and resuspension calculator. The program includes various bioinformatics tools for analysis of sequences with CG or AT skew, of CG content and purine–pyrimidine skew, and of linguistic sequence complexity. It also permits generation of	

random DNA sequence and analysis of restriction enzymes of all types. It finds or creates restriction enzyme recognition sites for coding sequences and supports the clustering of sequences. It generates consensus sequences and analyzes sequence conservation. It performs efficient and complete detection of various repeat types and displays them. FastPCR allows for sequence file batch processing, which is essential for automation. The FastPCR software is available for download at http://primerdigital.com/fastpcr.html and online version at http://primerdigital.com/tools/pcr.html.

Keywords (separated by “-”)	PCR primer design - Primer linguistic complexity - Sequence assembly - Software probe design - Ligase chain reaction - DNA primers

Chapter 18

FastPCR Software for PCR, In Silico PCR, and Oligonucleotide Assembly and Analysis

Ruslan Kalendar, David Lee, and Alan H. Schulman

Abstract

This chapter introduces the software FastPCR as an integrated tools environment for PCR primer and probe design. It also predicts oligonucleotide properties based on experimental studies of PCR efficiency. The software provides comprehensive facilities for designing primers for most PCR applications and their combinations, including standard, multiplex, long-distance, inverse, real-time, group-specific, unique, and overlap extension PCR for multi-fragment assembly in cloning, as well as bisulphite modification assays. It includes a program to design oligonucleotide sets for long sequence assembly by the ligase chain reaction. The in silico PCR primer or probe search includes comprehensive analyses of individual primers and primer pairs. It calculates the melting temperature for standard and degenerate oligonucleotides including LNA and other modifications, provides analyses for a set of primers with prediction of oligonucleotide properties, dimer and G/C-quadruplex detection, and linguistic complexity, and provides a dilution and resuspension calculator. The program includes various bioinformatics tools for analysis of sequences with CG or AT skew, of CG content and purine–pyrimidine skew, and of linguistic sequence complexity. It also permits generation of random DNA sequence and analysis of restriction enzymes of all types. It finds or creates restriction enzyme recognition sites for coding sequences and supports the clustering of sequences. It generates consensus sequences and analyzes sequence conservation. It performs efficient and complete detection of various repeat types and displays them. FastPCR allows for sequence file batch processing, which is essential for automation. The FastPCR software is available for download at http://primerdigital.com/fastpcr.html and online version at http://primerdigital.com/tools/pcr.html.

Key words PCR primer design, Primer linguistic complexity, Sequence assembly, Software probe design, Ligase chain reaction, DNA primers

Abbreviation

OE-PCR	Overlap extension PCR
PCR	Polymerase chain reaction
RT-PCR	Real-time PCR
SSR	Simple sequence repeat

Svein Valla and Rahmi Lale (eds.), *DNA Cloning and Assembly Methods*, Methods in Molecular Biology, vol. 1116, DOI 10.1007/978-1-62703-764-8_18, © Springer Science+Business Media New York 2014

1 Introduction

The polymerase chain reaction (PCR) is fundamental to molecular biology and is the most important practical molecular technique for the DNA research laboratory. However, the utility of the method is dependent on identifying unique primer sequences and designing PCR-efficient primers. Primer design is a critical step in all types of PCR methods to ensure specific and efficient amplification of a target sequence [1–7]. Even though there are currently many online and commercial bioinformatics tools, primer design for PCR is still not as convenient and practical as it might be for routine use. The adaptation of PCR for different applications has made it necessary to develop new criteria for PCR primer and probe design to cover uses such as RT-PCR, real-time PCR, group-specific and unique PCR, combinations of multiple primers in multiplex PCR, overlap extension PCR fort multi-fragments assembling cloning, and bisulphite modification assays. There is a need as well as for a program integrating design oligonucleotide sets for long sequence assembly by the ligase chain reaction (LCR), discovery of simple sequence repeats (SSRs) and their amplification as diagnostic markers, and for designing TaqMan, molecular beacon, and microarray oligonucleotides [6, 8, 9].

In developing FastPCR and Java web tools (Table 1), our aim was to create practical and easy-to-use software for routine manipulation and analysis of sequences for most PCR applications. The parameters adopted are based on our experimental data for efficient PCR and are translated into algorithms in order to design combinations of primer pairs for optimal amplification. This software, FastPCR, has been successfully used throughout the scientific community in a wide range of PCR and probe applications, and repeat searches and analysis. The first papers describing this software were published in 2001 and since 1999 this software has been applied in various projects, have been cited together over 500 times in scientific journals, patents, PhD theses, and over 1,000 downloads of the installation file per month. The program code is regularly updated.

2 Software, General Information

The FastPCR software (http://primerdigital.com/fastpcr.html) is written in Microsoft Visual Studio 6.0 and compiled to an executive file that, after installation, can be used with any version of Microsoft Windows. For Linux and Mac it requires "Wine" (http://www.winehq.org/) as a compatibility layer for running Windows programs. It is a completely free alternative implementation of the Windows API also for use with native Windows DLLs.

Table 1
Summary of the FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis

Features
PCR tool provides comprehensive facilities for
Design of primers for most PCR applications and their combinations, including standard, multiplex, long-distance, inverse, real-time, unique (specific primers for each from genetically related DNA sequences) or group-specific (universal primers for genetically related DNA sequences), linear-after-the-exponential (LATE)-PCR, bisulphite modification assays, polymerase extension PCR multi-fragment assembly cloning
Design of long oligonucleotides for microarray analyses and dual-labeled oligonucleotides for probes such as molecular beacons
Polymerase chain assembly (PCA) or oligo assembly—for automating the design of oligonucleotide sets for long sequence assembly by ligase chain reaction (LCR) and PCR
In silico (virtual) PCR or multiple primer or probe searches, or in silico PCR against whole genome(s) or a list of predictions by chromosome of probable PCR products, and search for potential mismatching locations of the specified primers or probes
Testing of individual primers, melting temperature calculation for standard and degenerate oligonucleotides including LNA and other modifications
Evaluation of PCR efficiency, linguistic complexity, dimer and G/C-quadruplex detection, dilution and resuspension calculator
Analysis of features of multiple primers simultaneously, including T_m, CG content, linguistic complexity, dimer formation; optimal T_a
Identification of simple sequence repeat (SSR) loci by analyzing the low-complexity regions of input sequences
Restriction digest analyses for Type I, II, and III restriction enzymes and homing endonucleases, finding or creating restriction enzyme recognition sites for coding sequences
Searches for similar sequences (or primers)
Translation of nucleotide (DNA/RNA) sequences to the corresponding peptide sequence in all six frames for standard and degenerate DNA and modifications (inosine, uridine)
Determination of CG:(G–C)/(G+C), AT:(A–T)/(A+T), SW:(S–W)/(S+W), MK:(M–K)/(M+K), purine–pyrimidine (R–Y)/(R+Y) skews, CG% content and the melting temperature, primer quality and linguistic sequence complexity profiles

The online FastPCR (jPCR) software (http://primerdigital.com/tools/) is written in Java with NetBeans IDE (Oracle) and requires the Java Runtime Environment (JRE) on a computer. It can be used with any operating system (64-bit OS preferred for large chromosome files).

3 The Interface

3.1 Inputs to FastPCR

The software contains the menus, the toolbars, and the ribbon and three text editors. The ribbon is designed to help the user quickly find the commands that are needed to complete a task. Commands are organized in logical groups, which are collected together under tabs (Fig. 1). Each tab relates to a type of activity, such as "PCR Primer Design," "in silico PCR," or "Primer Test."

Getting started with a basic project in FastPCR software is as easy as opening a new or existing file than using copy–paste or starting to type. There are three independent text editors on different tabs within the interface: "General sequence(s)," "Additional sequence(s) or pre-designed primers (probes) list," and "Result report." The first two text editors are necessary for loading sequences for analysis, the text editor "General sequence(s)" is designed for working with the project sequences, and the "Additional sequence(s) or pre-designed primers (probes) list" text editor is used for special and additional sequences such as pre-designed primers, multiple query sequences, or numbers for input. [AU2]

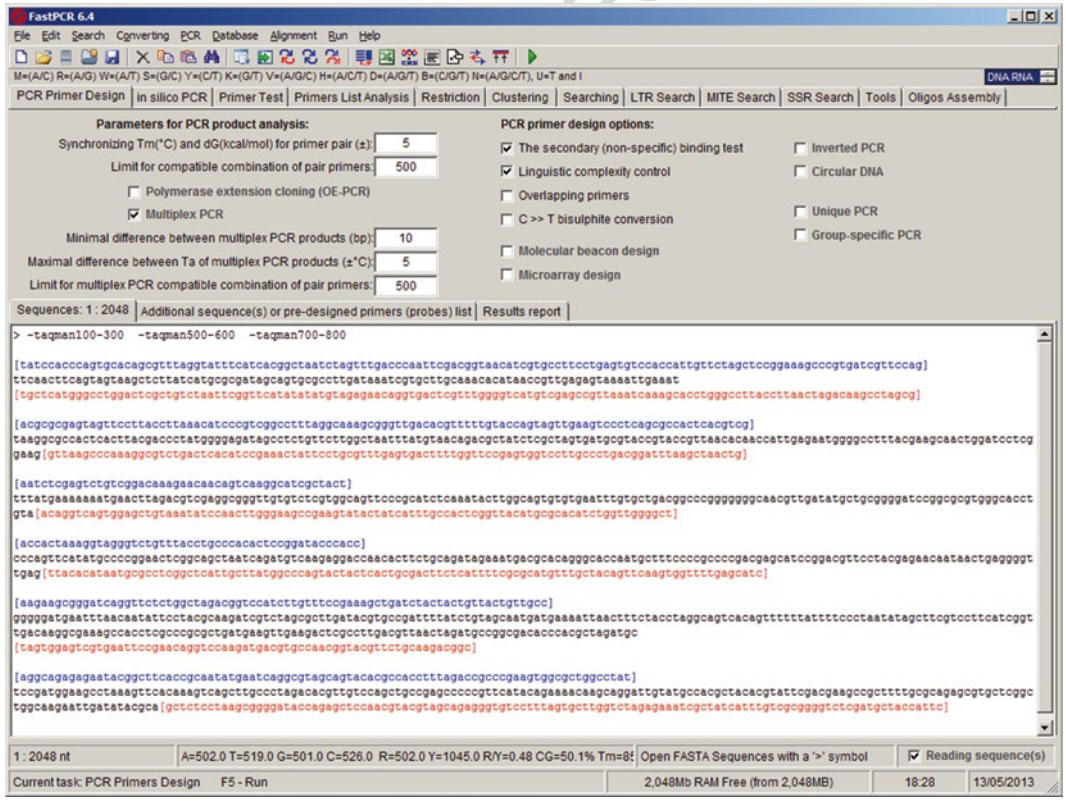


Fig. 1 The FastPCR sequence editor and user interface

3.2 Program Output

FastPCR automatically generates results at a third text editor, "Result report," in tabulated format for transferring to spreadsheet software from the clipboard using copy–paste. Output results are easy to save as Excel worksheet (.xls) or as Rich Text Format (.rtf) text files, compatible with MS Excel or Open Office. The separated output of the primer design is a list of primers, a set of primer pair sequences with their theoretical PCR products, and, for multiplex PCR, the result of the calculation of multiple-PCR primers for given target sequences. In addition, the output shows optimal annealing temperature for each primer pair, the size of PCR product, and complete information for each designed primer and for each multiplex PCR product set.

3.3 Sequence Entry

Sequence data files are prepared using a text editor (Notepad, WordPad, MS Word), and saved in ASCII as text/plain format (.txt) or in .rtf. The program either takes a single sequence or accepts multiple separate DNA sequences in FASTA, tabulated format (two columns from MS Excel sheet or MS Word table), EMBL, MEGA, GenBank, MSF, DIALIGN, simple alignment, or BLAST Queue web alignment formats. The template length is not limited. The FastPCR clipboard allows the user to copy and paste text or tables from MS Word documents or MS Excel worksheets or other programs and to paste them into another Office document. It is important that all target sequences are prepared in the same format. Users can type or import from file(s) into "General Sequence(s)" or "Additional sequence(s) or pre-designed primers (probes) list" editors.

FastPCR allows files to be opened in several ways: the original file can be opened as read-only for editing with text editors; files can be opened to memory without using text editors, which allows larger file(s), up to 200 Mb, to be analyzed; files within a folder can be selected and the files opened during task execution without the use of text editor program. Additionally, the program can open files within a selected folder in order to join all these files in a text editor. For example, this feature can be applied to convert all files from a selected folder into a single file of FASTA sequences. Alternatively this feature allows splitting FASTA sequences to individual files in a particular selected folder.

When a sequence file is opened, FastPCR displays the information about the opened sequence and its format. The information status bar shows the number of sequences, the total sequence length (in nucleotides), the nucleotide composition, and the purine, pyrimidine, CG percentage, and the melting temperature. Files can be saved in various formats including .rtf, .xls, or txt from the text editor in use.

3.4 FASTA Format Description

FastPCR normally expects to read sequence files in FASTA format [10]. FASTA format has the highest priority and is simple, because the raw sequence is merely preceded by a definition line. The definition line begins with a ">" sign and optionally followed

immediately by a name for the sequence with no length restriction. Many sequences can be listed in the file, the format indicating a new sequence at each ">" symbol found. It is important to press Enter at the end of each line after ">" to help FastPCR recognize the end and beginning of the sequence and the sequence name. It is important that the first line of each sequence starts with ">."

Degenerate DNA sequences are accepted as IUPAC code [11], an extended vocabulary of 11 letters, which allows the description of ambiguous DNA code. Each letter represents a combination of one or several nucleotides: M = (A/C), R = (A/G), W = (A/T), S = (G/C), Y = (C/T), K = (G/T), V = (A/G/C), H = (A/C/T), D = (A/G/T), B = (C/G/T), N = (A/G/C/T), U = T, and I (Inosine). The program accepts amino acid codes: A (Ala), C (Cys), D (Asp), E (Glu), F (Phe), G (Cly), H (His), I (Ile), K (Lys), L (Leu), M (Met), N (Asn), P (Pro), Q (Gln), R (Arg), S (Ser), T (Thr), U (Sec), V (Val), W (Trp), and Y (Tyr).

3.5 Alignment Format Description

There are many different programs that perform different types of alignment formats. Standardizing on a set of formats enables programs to be written that can read results from many different programs. In all alignment formats, gaps that have been introduced into the sequences to make them align are indicated by the "-" character. The exception to this rule is GCG/MSF format which uses "." as the gap character inside the sequences. The file may begin with as many lines of comment or description as required. The first mandatory line must contain the text "MSF," "Alignment as simple alignment format," "DIALIGN," or "MEGA" to be recognized as alignments from these programs. Following the first line are lines that start with the sequence name, which is separated from the aligned sequence residues by white space.

4 The PCR Primers or Probe Design Analysis Options

4.1 PCR Primer Design Generalities

Primer design is one of the key steps for successful PCR. For PCR applications, primers are usually 18–35 bases in length and should be designed such that they have complete sequence identity to the desired target fragment to be amplified. The parameters, either controllable by the user or selected automatically, are primer length (12–500 nt), melting temperature for short primers calculated by nearest neighbor thermodynamic parameters, theoretical primer PCR efficiency (quality at %) value, primer CG content, 3′ end terminal enforcement, preferable 3′ terminal nucleotide sequence composition in degenerated formulae, and added sequence tags at 5′ termini. The other main parameters used for primer selection are the general nucleotide structure of the primer such as linguistic complexity (nucleotide arrangement and composition), specificity,

t2.1 **Table 2**
t2.2 **Default primer design selection criteria**

t2.3 Criteria	Default	Ideal
t2.4 Length (nt)	20–24	>21
t2.5 T_m range (°C)[a]	52–68	60–68
t2.6 T_m[a] 12 bases at 3′ end	30–50	41–47
t2.7 CG (%)	45–65	50
t2.8 3′ end composition (5′-NNN-3′)[b]	SWW, SWS, SSW, WSS	SSA, SWS, WSS
t2.9 Sequence linguistic complexity (*LC*, %)[c]	>75	>90
t2.10 Sequence quality (*PQ*, %)	>70	>90

t2.11 [a]Nearest neighbor thermodynamic parameters according to SantaLucia [13]
t2.12 [b]Ambiguity codes
t2.13 [c]Sequence linguistic complexity measurement calculated using the alphabet-capacity *L*-gram method

the melting temperature of the whole primer, the melting temperature at the 3′ and 5′ termini, self-complementarity, and secondary (nonspecific) binding.

The software can dynamically optimize the best primer length for the entered parameters. All PCR primer (probe) design parameters are flexible and changeable according to the specifics of the analyzed SEQUENCES and tasks. Primer pairs are analyzed for cross-hybridization and specificity of both primers and, optionally, selected with similar melting temperatures. Primers with balanced melting temperatures (within 1–6 °C of each other) are desirable but not mandatory. The default primer design selection criteria are shown in Table 2. It is possible to use predesigned primers or probes or, alternatively, predesigned primers can act as references for the design of new primers. The program accepts a list of predesigned oligonucleotide sequences and checks the compatibility of each primer with a newly designed primer or probe.

[AU3]

4.2 Melting Temperature Calculation

The melting temperature (T_m) is defined as the temperature at which half the DNA strands are in the double-helical state and half are in the "random-coil" state. The T_m for short oligonucleotides with normal or degenerate (mixed) nucleotide combinations is calculated in the default setting using nearest neighbor thermodynamic parameters [12, 13]. The CG content of an oligonucleotide is the most important factor that influences the T_m value. The melting temperature for mixed bases is calculated by averaging nearest neighbor thermodynamic parameters—enthalpy and entropy values—at each mixed site; extinction coefficient is similarly predicted by averaging nearest neighbor values at mixed sites [2, 3]. Mismatched pairs can be taken into account since the parameters provide for DNA/DNA duplexes and dangling ends,

which are unmatched terminal nucleotides [14–16]. The melting temperature for primer (probe) self- or cross-dimers and for in silico PCR experiments with oligonucleotides having mismatches to the target is calculated using values for the thermodynamic parameters for a nucleic acid duplex.

The FastPCR allows the choice of other nearest neighbor thermodynamic parameters or simple non-thermodynamic T_m calculation formulae. For non-thermodynamic T_m calculation, we suggest using simple formulae; the Wallace–Ikatura rule [17] is often used as a rule of thumb when primer T_m is to be estimated at the bench. However, the formula was originally applied to the hybridization of probes in 1 M NaCl and is an estimate of the melting temperature for oligonucleotides shorter than 10 bases:

$$T_m(^\circ\mathrm{C}) = 2(\mathrm{L}+\mathrm{G}+\mathrm{C}),$$

for oligonucleotides longer than ten bases:

$$T_m(^\circ\mathrm{C}) = 64.9 + \frac{41([\mathrm{G}+\mathrm{C}]-16.4)}{L}$$

or the formula [18]:

$$T_m(^\circ\mathrm{C}) = 77.1 + 11.7\log_{10}[\mathrm{K}^+] + \frac{41[\mathrm{G}+\mathrm{C}]-528}{L}$$

where L is length of primer, [G+C] is the number of Gs and Cs, and [K^+] is salt molar concentration (default value is 50 mM). The two equations above assume that the stabilizing effects of cations are the same on all base pairs. The melting temperature of the PCR product is calculated using the formula [15]:

$$T_m(^\circ\mathrm{C}) = 81.5 + 16.6\log_{10}[\mathrm{K}^+] + \frac{41[\mathrm{G}+\mathrm{C}]-675}{L}$$

4.3 Linguistic Complexity of Sequence and Nucleotide-Skew Analysis

The sequence complexity calculation method can be used to search for conserved regions between the compared sequences in order to detect low-complexity regions including SSRs, imperfect direct or inverted repeats, polypurine and polypyrimidine triple-stranded DNA structures, and four-stranded structures (such as G/C-quadruplexes) [19]. Linguistic complexity measurements are performed using the alphabet-capacity L-gram method [20, 21] along the whole sequence length and calculated as the sum of the observed range (x_i), from 1- to L-size words in the sequence, divided by the sum of the expected (E) value for this sequence length. Linguistic complexity (LC) values for sequence length (s) are converted to percentages, in which 100 % means maximal "vocabulary richness" of a sequence:

$$LC(\%) = \frac{100 \times \sum_{i=1}^{L} x_i}{E},$$

253

where 254

$$E = \sum_{i=1}^{L} \begin{cases} s - i + 1, s < 4^i - 1 + i \\ 4^i, s \geq 4^i - 1 + i \end{cases},$$

255

$$L = \left[\log_4 \left(\frac{s}{3} \right) + 1 \right].$$

256

For example, the sequence 5′-ACACACACACACACAC, 16 nt ($L=3$), contains two nucleotides (A, C), but expected $E=4$ variants; two variants of dinucleotides (AC, CA), but expected $E=(16-1)$ variants; two variants of trinucleotides (ACA, CAC), and expected $E=(16-2)$ variants. The complexity value is $LC=100(2+2+2)/(4+16-1+16-2)=18.2\%$. 257 258 259 260 261 262

The LC tries to describe the "uniqueness" ("vocabulary richness") of a sequence and the likelihood of PCR success of each primer; this value varies from 100 for the best to 5 (e.g., poly(N)) for the worst primers. LC values of 80 and higher allow for the rapid choice of the best primer or probe sequences. 263 264 265 266 267

4.4 Primer Quality (Virtual PCR Efficiency) Determination

Our experimental data showed that the primer nucleotide composition and melting temperature of the 12 bases at the 3′ end of the primers are important factors for PCR efficiency. The melting temperature of the 12 bases at the 3′ terminus is calculated preferably by nearest neighbor thermodynamic parameters [13]. The composition of the sequence at the 3′ terminus is important; primers with two terminal C/G bases are recommended for increased PCR efficiency [22]. Nucleotide residues C and G form a strong pairing structure in the duplex DNA strands. Stability at the 3′ end in primer template complexes will improve the polymerization efficiency. 268 269 270 271 272 273 274 275 276 277 278

We specify an abstract parameter called primer quality (PQ) that can help to estimate the efficiency of primers for PCR. PQ is calculated by the consecutive summation of the points according to the parameters of total sequence and purine–pyrimidine sequence complexity and of the melting temperatures of the whole primer and of the terminal 12 bases at both the 3′ and 5′ ends. Self-complementarity, which gives rise to possible primer-dimer and hairpin structures, reduces the final value. The PQ tries to describe the likelihood of PCR success of each primer; this value varies from 100 for the best to 0 for the worst primers. To meet multiplexing demands, it is possible in the program to select the best primer with an optimal temperature range, allowing the design of qualified primers or probes for any target sequence with any CG 279 280 281 282 283 284 285 286 287 288 289 290 291

and repeat content. PQ values of 80 and higher allow for the rapid choice of the best PCR primer pair combinations. No adverse effects, due to the modification of the reaction buffer, chosen thermostable polymerases, or variations in annealing temperature, have been observed on the reproducibility of PCR amplification using primers with high PQ.

4.5 Hairpin (Loop) and Dimer Formation

Primer-dimers involving one or two sequences may occur in a PCR. The FastPCR tool eliminates intra- and inter-oligonucleotide reactions before generating a primer list and primer pair candidates. It is very important for PCR efficiency that the production of stable and inhibitory dimers is prevented, especially by avoiding complementarity in the 3′ ends of primers from whence the polymerase will extend. Stable primer-dimer formation is very effective at inhibiting PCR because the dimers formed are amplified efficiently and compete with the intended target.

Primer-dimer prediction is based on analysis of non-gapped local alignments and the stability of both the 3′ end and the central regions of the primers (Fig. 2). Primers will be rejected when they have the potential to form stable dimers, depending on the nucleotide composition and with at least five bases at the 3′ end or seven bases in the central region. Tools to calculate T_m for primer-dimers with mismatches for pure, mixed, or modified (inosine, uridine, or locked nucleic acid (LNA)) bases, using averaged nearest neighbor thermodynamic parameters, are provided for DNA/DNA duplexes [12–14, 23, 24].

In addition to Watson–Crick base-pairing, there is a variety of other hydrogen bonding configurations possible [19, 25–27], including G/C-quadruplexes and wobble base pairs, which the FastPCR software detects. The program provides for the detection of alternative hydrogen bonding during primer-dimer and in silico PCR primer binding site detection. The mismatch stability is examined in order of decreasing stability: **G-C**>**A-T**>G·G>G·T≥G·A>T·T≥A·A>T·C≥A·C≥C·C. Guanine is the most universal base, because it forms the strongest base pair and the strongest mismatches. However, cytosine is the most discriminating base, because it forms the strongest pair and the three weakest mismatches [23, 28]. Therefore, the software also looks for stable guanine mismatches: G·G, G·T and G·A.

G-rich (and C-rich) nucleic acid sequences can fold into four-stranded DNA structures that contain stacks of G-quartets [19]. These quadruplexes can be formed by the intermolecular association of two or four DNA molecules, dimerization of sequences that contain two G bases, or by the intermolecular folding of a single strand containing four blocks of guanines. These are easy to eliminate from primer design because of their low linguistic complexity; LC = 32 % for $(TTAGGG)_4$. The software predicts the presence of putative G- and C-quadruplexes in primer sequences.

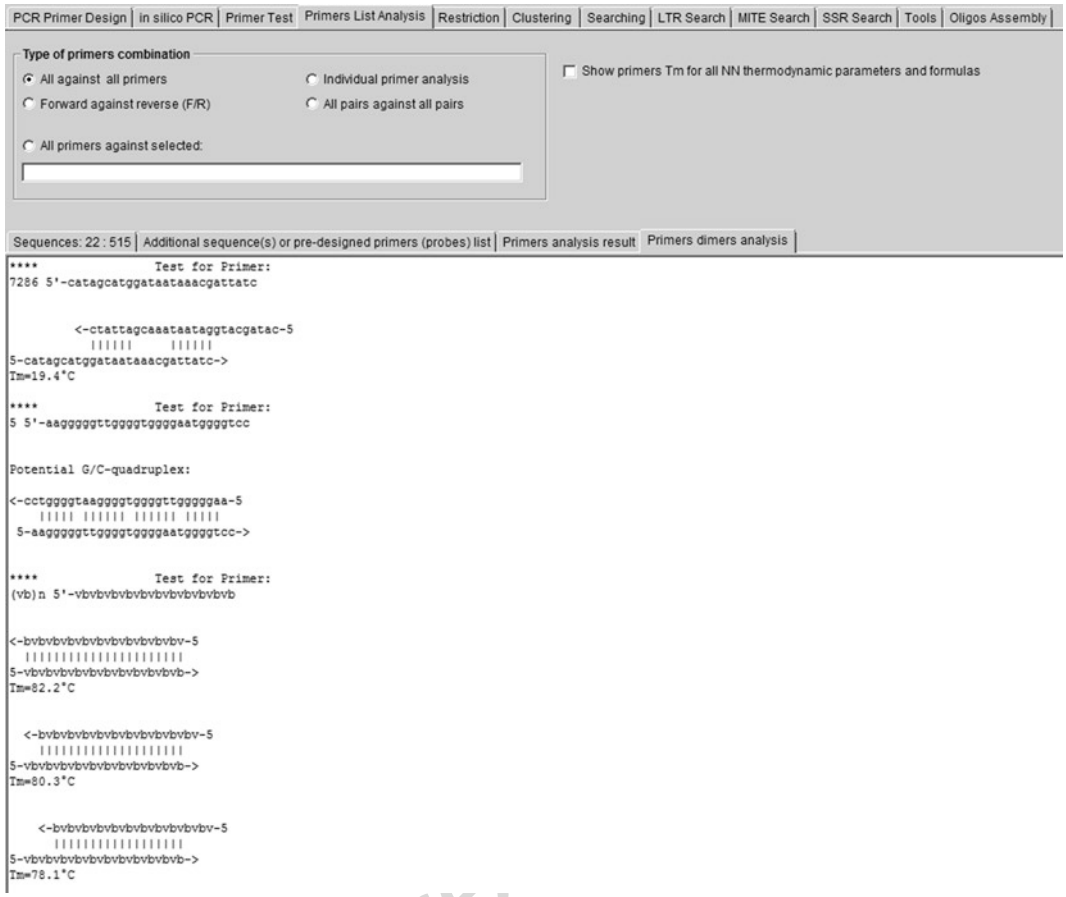


Fig. 2 One of the FastPCR duplex formation results

Intermolecular G-quadruplex-forming sequences are detected according to the formula $\ldots G_{m1}X_nG_{m2}\ldots$, where m is the number of G residues in each G-tract (m_1, $m_2 \geq 3$); the gap X_n ($n \leq 2$*minimal (m_1:m_2)) can be any combination of residues, including G [16]. The gap sequences (X_n) may have varying lengths, and a relatively stable quadruplex structure may still be formed with a loop more than seven bases long, but in general, increasing the length of the gap leads to a decrease in structure stability. It is also possible for one of the gaps to be zero length when there are long poly-G tracts of >6 bases.

4.6 Calculation of Optimal Annealing Temperature

The optimal annealing temperature (T_a) is the temperature, generally stated as a range, where efficiency of PCR amplification is maximal but nonspecific products minimal. The most important values for estimating the T_a are the primer quality, the T_m of the primers, and the length of PCR fragment. Primers with high T_m (>60 °C) can be used in PCRs with a wide T_a range compared to

primers with low T_m (<50 °C). The optimal annealing temperature for PCR is calculated directly as the value for the primer with the lowest T_m (T_m^{min}). However, PCR can work in temperatures up to 10° higher than the T_m of the primer so as to favor primer target duplex formation:

$T_a(^\circ C) = T_m^{min} + \ln L$, where L is the length of the PCR fragment.

4.7 The Secondary Nonspecific Binding Test; Alternative Amplification Products

The specificity of the oligonucleotides is one of the most important factors for good PCR; optimal primers should hybridize only to the target sequence, particularly when complex genomic DNA is used as the template. Amplification problems can arise due to primers annealing to repetitious sequences (retrotransposons, DNA transposons, or tandem repeats). Alternative product amplification can also occur when primers are complementary to inverted repeats and produce multiple bands. This is unlikely when primers have been designed using specific DNA sequences (unique PCR). However, the generation of inverted repeat sequences is exploited in two common generic DNA fingerprinting methods, RAPD and AP-PCR [29, 30]. Because only one primer is used in these PCRs, the ends of the products must be reverse complements and thus can form stem-loops.

The techniques of inter-repeats amplified polymorphism: inter-retrotransposon amplified polymorphism (IRAP), retrotransposon-microsatellite amplified polymorphism (REMAP), inter-MITE amplification [31, 32], and *Alu*-repeat polymorphism [33, 34] have exploited highly abundant, dispersed repeats as markers. However, primers complementary to repetitious DNA may produce many nonspecific bands in single-primer amplification and compromise the performance of unique PCRs. A homology search with the primer as the query sequence, for example, using BLASTn against all sequences in GenBank or EMBL-Bank, will determine whether the primer is likely to interact with dispersed repeats. Alternatively, one can create a small, local, specialized library of repeat sequences based on those in Repbase [35] or TREP [36] and use this for searches.

The mismatches at the 3′ end of the primers affect target amplification much more than mismatches at the 5′ end. A two-base mismatch at the 3′ end of the primer prevents amplification. A single-base mismatch at the 3′ end as well as several mismatches at the 5′ end of the primer permits amplification, albeit with reduced efficiency. However, the presence of multiple primer binding sites does not necessarily lead to alternative amplification products because, for successful amplification, the priming sites for both primers must be both located close to each other, in correct orientation, and sufficiently match the primer sequences.

By default, FastPCR performs a test for nonspecific binding by repeats masking and low-complexity regions detection and masking for each given sequence.

Additionally the software allows this test to be performed against a reference sequence or sequences (e.g., a BAC or YAC) or one's own database. Primers that bind to more than one location on given sequences will be rejected. Even though the test for non-specific primer binding is performed as a default for all primers, the user may cancel the operation. Identification of secondary binding sites including mismatched hybridization is normally performed by considering the similarity of the primer to targets along the entire primer sequence. An implicit assumption is that stable hybridization of a primer with the template is a prerequisite for priming by DNA polymerase. FastPCR pays particular attention to the 3′ portion of the primers and calculates the similarity of 3′ end to the target (the length is chosen by user) to determine the stability of any potential interactions.

The secondary nonspecific primer binding test is based on repeat masking using a quick local alignment screen (which allows one mismatch within a hash index of 12-mers) between the reference and input sequences.

5 Methods

Once the input files are selected or sequences copied and pasted to the **General Sequence(s)** text editor, the FastPCR provides various execution features. Figure 3 provides an example for primer design from the user's perspective.

5.1 Execution of the Selected Task

The user selects the ribbon having the task needed. The program will only perform the selected task. The name of the selected executive task is shown on the status bar by "Press F5." The task is executed by using key **F5** or by clicking the arrow on the toolbar using the mouse. Once the executive task is completed, the result is shown in the **Result report** text editor (e.g., *see* Fig. 3).

5.2 PCR Primer Design Options

The "PCR Primer Design" Tab contains various execution options for commonly selected types of PCR and for the most important PCR parameters (Fig. 1). The option panel of "PCR Primers or Probe Design Options" is shown in Fig. 4. Once the user selects any attribute, the option attribute value field shows the default attributes value, which can then be modified. "PCR Primers or Probe Design Options" affects all sequences. PCR primer design options can be customized for each sequence using special commands at the header of the sequence (http://primerdigital.com/soft/pcr_help.html). Typically, it is not necessary to use these commands to manage typical PCR primer design and these are applied to advanced tasks. Default global parameters for primer design will be assigned by typing the help command "/?" in text editor:

Sequences: 3 : 844 | Additional sequence(s) or pre-designed primers (probes) list | PCR primers design result

PrimerID	Sequence(5'-3')	Length(bp)	Tm(°C)	dG(kcal/mol)	Tm_3'end(°C)	CG(%)	Linguistic_Complexity(%)	Primer_Quality(%)
example1								
1F1_1_18-39	tcgtattcaggcgtaccttctg	22	56.5	-27.3	41.2	50.0	88	88
1F2_1_42-62	gcctcgggctgcgttacttcg	21	64.0	-30.0	40.3	66.7	85	85
1F3_1_53-73	cgttacttcgttgcggatagg	21	56.6	-26.3	41.4	52.4	87	87
1F4_1_65-85	gcggataggactatctcggtg	21	57.3	-26.3	37.2	57.1	85	83
1F5_1_78-97	tctcggtgttttctgactgc	20	55.4	-25.0	36.3	50.0	76	73
1F6_1_99-118	tggcttccgcgagtcattgc	20	61.1	-27.9	43.4	60.0	92	92
1R1_1_348-368	atgtgggtttgccttacaagc	21	56.7	-26.1	40.2	47.6	92	90
1R2_1_264-286	tgttggttacttctcggtaaggt	23	55.3	-27.4	38.8	43.5	80	80
1R3_1_161-180	atgatgccctcatgtccgta	20	55.5	-25.1	37.7	50.0	82	82
1R4_1_136-157	tgttttatcgacaccttcgtcc	22	55.2	-26.6	42.8	45.5	85	80

Forward_PrimerID	Sequence(5'-3')	Tm(°C)	Primer_Quality(%)	Reverse_PrimerID	Sequence(5'-3')	Tm(°C)	Primer_Quality(%)	PCR_Fragment_Size(bp)	Topt(°C)
1F1_1_18-39	tcgtattcaggcgtaccttctg	56.5	88	1R1_1_348-368	atgtgggtttgccttacaagc	56.7	90	351	62.0
1F1_1_18-39	tcgtattcaggcgtaccttctg	56.5	88	1R2_1_264-286	tgttggttacttctcggtaaggt	55.3	80	269	61.0
1F1_1_18-39	tcgtattcaggcgtaccttctg	56.5	88	1R3_1_161-180	atgatgccctcatgtccgta	55.5	82	163	61.0
1F1_1_18-39	tcgtattcaggcgtaccttctg	56.5	88	1R4_1_136-157	tgttttatcgacaccttcgtcc	55.2	80	140	60.0
1F3_1_53-73	cgttacttcgttgcggatagg	56.6	87	1R1_1_348-368	atgtgggtttgccttacaagc	56.7	90	316	62.0
1F3_1_53-73	cgttacttcgttgcggatagg	56.6	87	1R2_1_264-286	tgttggttacttctcggtaaggt	55.3	80	234	61.0
1F3_1_53-73	cgttacttcgttgcggatagg	56.6	87	1R3_1_161-180	atgatgccctcatgtccgta	55.5	82	128	60.0
1F3_1_53-73	cgttacttcgttgcggatagg	56.6	87	1R4_1_136-157	tgttttatcgacaccttcgtcc	55.2	80	105	60.0
1F4_1_65-85	gcggataggactatctcggtg	57.3	83	1R1_1_348-368	atgtgggtttgccttacaagc	56.7	90	304	62.0
1F4_1_65-85	gcggataggactatctcggtg	57.3	83	1R2_1_264-286	tgttggttacttctcggtaaggt	55.3	80	222	61.0

example2									
2F1_1_9-28	gagcacataggaacgctcca	20	57.3	-25.8	44.5	55.0	92	92	
2F2_1_58-77	taatcgcggcacacggaggt	20	60.8	-28.1	47.6	60.0	84	80	
2F3_1_84-105	gtgctaacaaccacggtaagtc	22	56.8	-27.0	40.6	50.0	88	88	
2F4_1_96-115	acggtaagtctcacgtccaa	20	55.3	-25.2	40.3	50.0	82	80	
2R1_1_96-115	ttggacgtgagacttaccgt	20	55.3	-25.2	37.6	50.0	82	82	
2R2_1_84-105	gacttaccgtggttgttagcac	22	56.8	-27.0	38.3	50.0	88	86	
2R3_1_67-88	agcacatatttacctccgtgtg	22	55.2	-26.4	41.8	45.5	90	80	
2R4_1_56-75	ctccgtgtgccgcgattagg	20	61.7	-27.9	46.4	65.0	87	87	
2R5_1_16-35	cgagctctggagcgttccta	20	58.8	-26.6	42.0	60.0	84	84	
2R6_1_5-25	agcgttcctatgtgctccgca	21	60.7	-28.9	46.1	57.1	87	87	

Forward_PrimerID	Sequence(5'-3')	Tm(°C)	Primer_Quality(%)	Reverse_PrimerID	Sequence(5'-3')	Tm(°C)	Primer_Quality(%)	PCR_Fragment_Size(bp)	Topt(°C)
2F1_1_9-28	gagcacataggaacgctcca	57.3	92	2R1_1_96-115	ttggacgtgagacttaccgt	55.3	82	107	60.0

Fig. 3 Primer design result window

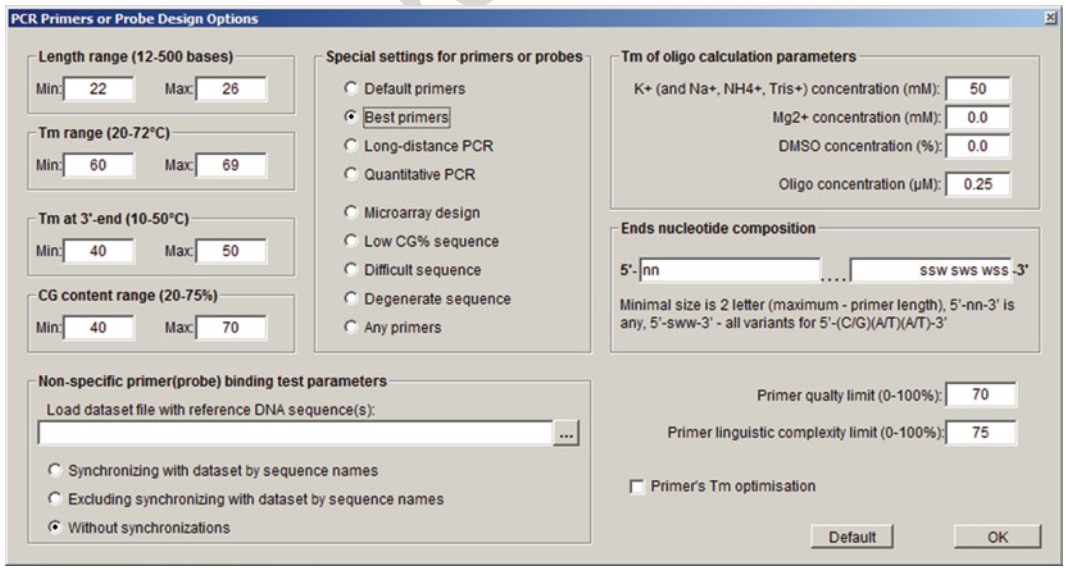


Fig. 4 The "PCR Primers or Probes Design Options" window

"{-ln20-23 -tm55-68 -3tm37-50 -cg41-70 -q70 -lc75 -npr400 -c5[NN] -c3[SWW SSW SWS WSS]}," where "-ln20-23" determines the range of primer length (20–23 bases), "-tm55-68" determines the range of primer T_m (55–68 °C), "-3tm37-50" determines the range of primer T_m at 3′ end (37–50 °C), "-cg41-70" determines the range of primer CG% contents (between 41 and 70 %), "-npr400" shows the maximum number of primers (400) designed to each target, "-c5[NN]" denotes a primer having no specific sequence pattern for 5′ ends, and "-c3[SWW SSW SWS WSS]" specifies primers that conform to particular patterns of ambiguity, such as that shown here for example.

5.3 Examples for Primer Selection Region

Users can specify, individually for each sequence, multiple locations for both forward and reverse primer placement with the commands: "-FpdN1-N2" for forward primers and "-RpdN1-N2" for reverse primers, where from N1 to N2 are bases from the selected regions; "-pdN1-N2" (see more at: http://primerdigital.com/soft/pcr_help.html). Alternatively, users can specify multiple locations for both forward and reverse primers using ["and"] inside each sequence: the software allows multiple and independent locations for both forward and reverse primers inside each of the sequences. Whilst PCR primer design will be performed independently for different targets, multiplex PCR primer design can be performed simultaneously with multiple amplicons within a single sequence as well as for different sequences, i.e., all possible combinations of ["and"] inside one or more sequences. By default, the software designs primers within the entire sequence length. Optionally, users can specify, individually for each sequence, multiple locations for both forward and reverse primers with the commands:

1. The same location for both forward and reverse primers will be designed in the central [nnnnnnnnnn] part ("[]" is used only once):
   .....[nnnnnn]....
2. Different locations for forward and reverse primers; forward primers will be chosen inside the "[1nnnnnn]" location and reverse primers inside "[2nnnnnn]" location (twice "[]"):
   .....[1nnnnnn]....[2nnnnnn].....
3. Primers must flank the central "]nnnnnn["; forward primers will be chosen from 1 to "A]" bases and reverse primers will be chosen from base "[C" to the end of sequence: .....A]nnnnnn[C.....
4. Forward and reverse primers have an overlapping part "[nnnnnn]"; forward primers will be chosen from "[A to n]" bases and reverse primers will be chosen from "[n base to C]":
   .....[A.....[nnnnnn].....C]....

The software allows the selection of any number of independent PCR primer (or probe) designing tasks for each sequence using multiple combinations of "`[..]`" and `-FpdN1-N2`, `-RpdN1-N2`, or `-pdN1-N2` commands. Multiplex PCR can be carried out simultaneously within a single sequence with multiple tasks as well as for different sequences with multiple tasks or a combination of both. [AU4]

All possible combinations of "`[]`" (forward) with "`[]`" (reverse) within the sequence(s):

1. `[]`
2. `] [`
3. `[] []`
4. `[[]]`
5. `([] [])n or/and ([[]])n.`

5.4 User-Defined PCR Product Size

The PCR product size can be specified in a similar way, with the command: "`(N1-N2)`"; these values can be specified in the form of minimum and maximum values for the product size. For example, the "`(400-500)`" line defines that the product size ranges from 400 to 500 base pairs. In case a user wants to specify a fixed product size, the command should be a single number, for example, "`(500).`" FastPCR is flexible and allows PCR product sizes from 50 to 10,000 base pairs in length.

5.5 PCR Set-up Examples with Individual Commands

1. Where primers have already been designed, FastPCR can be used to predict the optimal annealing temperature and PCR product length for one or more predesigned primers (with the "`-npd`" command, which prohibits the primer design). For example, the command:

   "`-fpr[ggagagtagcttacctcgct cggtaaggttct-tcatgc] -npd`"

   will analyze the selected sequence between the two primers (5′ `ggagagtagcttacctcgct` and 5′ `cggtaaggttcttcatgc`).
2. Design primers with a difference in T_m of about 10°, e.g., for LATE-PCR:

   "`-Ftm50-55 -Rtm64-68 -pTMs10.`"
3. Design primers with a specific restriction enzyme site at the 3′ end ("`-z3eNameEnzyme`," "`-Fz3eNameEnzyme`," "`-Rz3eNameEnzyme`," where "`NameEnzyme`" is the name of the restriction enzyme: "`-z3eXceI`." The alternative command ("`-c3NN`") is also used for a special primer location. For example, "`-c3YCATGR`" is the same as "`-z3eXceI`": both commands will design primers with the recognition site for *Xce I* endonucleases `5'-YCATGR-3'` at the 3′ end of the primers. [AU5]
4. Additional bases can be added to primer ends using the commands "`-5eNN`" or "`-3eNN`," where 5 or 3 denotes the end at

which to add the extra bases and "NN" is a given sequence of one to more bases. For example, "`-F5eCGACG -R5eTTTTTT`," means adding the sequence "`CGACG`" to forward primers and "`TTTTTT`" to reverse primers, both at the 5′ ends.

5.6 Bisulphite-Modified DNA

The "C ≫ T bisulphite conversion" option allows the design of specific PCR primers for in silico bisulphite conversion for both strands. Only cytosine not followed by guanidine (CpG methylation) will be replaced by thymine:

```
5′aaCGaagtCC 3′          5′aaCGaagtTT 3′
  ||||||||||      ->       |||||| |
3′ttGCttCagg 5′          3′ttGCttTagg 5′
```

5.7 Uniqueness of Primers

Optionally, the user can synchronize the primer test for secondary, nonspecific binding with a dataset of sequence names. The program recognizes that a given sequence in the screening library dataset (from loading the dataset file) is the same as the sequence for which it is designing primers; this allows sequence selection to be made even if the selection matches the screening sequence perfectly. This allows the same dataset to be used for both primer design and screening without having to make a new screening database for each sequence. In other words, a dataset that contains sequences A, B, C, and D can be used both for choosing primers and for checking primer specificity. Alternatively, the user can input preexisting primers into a second "Additional sequence(s) or pre-designed primers (probes) list" text editor. These primers or probes will be checked for compatibility (inter-primer-dimer formation) with newly designed primers. The number of preexisting primers is not limited to one or two; it can be as many as the user needs.

5.8 PCR Primer Design

The PCR primer design algorithm generates a set of primers having a high likelihood of success in virtually any amplification protocol. All PCR primers designed by FastPCR can be used for PCR or sequencing experiments. The program is able to generate either long oligonucleotides or PCR primers for the amplification of gene-specific DNA fragments of user-defined length. FastPCR provides a flexible approach to designing primers for many applications and for both linear and circular sequences. It will check if either primers or probes have secondary binding sites in the input sequences that may give rise to additional PCR products. The selection of the optimal target region for the design of long oligonucleotides is performed in the same way as for PCR primers. The basic parameters in primer design are also used to measure the oligonucleotide quality and to evaluate the thermodynamic stability of the 3′ and 5′ terminal bases.

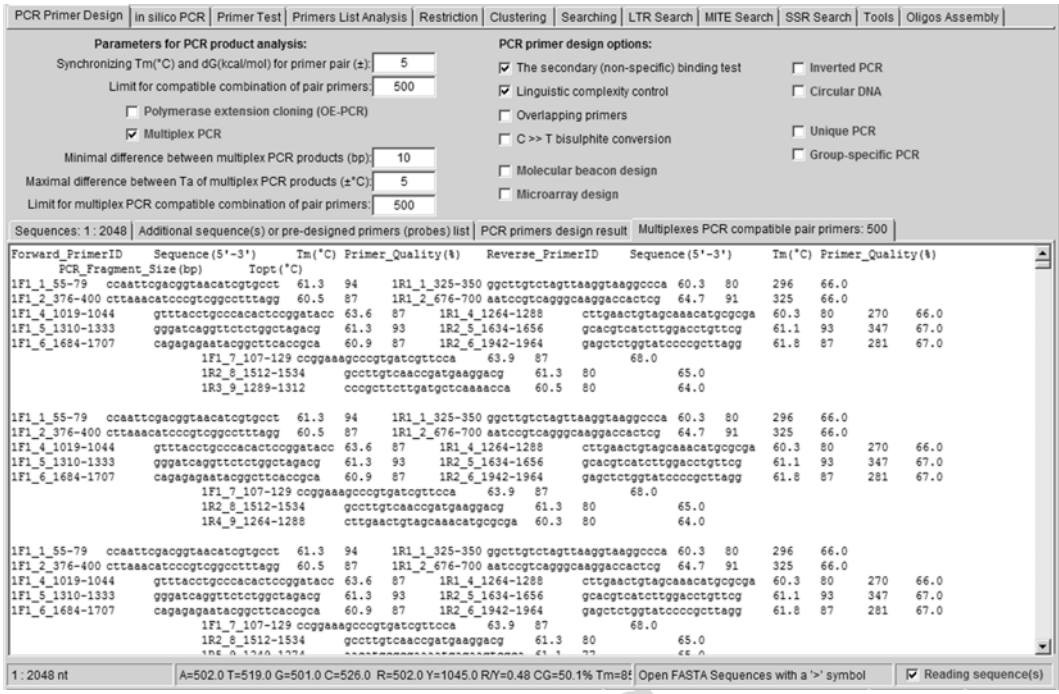


Fig. 5 An example of multiplex PCR result

Both the proposal of primer pairs and the selection of the best pairs are possible. The user can vary the product size or design primer pairs for the whole sequence without specifying parameters by using default or preset parameters. Preset parameters are specified for various situations: for example, for sequences with low CG content, for long-distance PCR, for degenerate sequences, or for manual input. A list of the best primer candidates and all compatible primer pairs that are optimal for PCR is generated. Users can specify, individually for each sequence, multiple locations for both forward and reverse primers within each sequence, whilst PCR design will be performed independently for different targets. Primers for multiplex PCRs can be designed from a single or from multiple targets (Fig. 5).

The program generates primer pairs (and probes) from the input sequences and shows the optimal annealing temperature for each primer pair and the sizes of PCR products together with information for each designed primer. Suggested primers and primer pairs are produced in tabulated format for MS Excel or Open Office (Table 3). The spreadsheets show the following properties: the automatically generated primer name, primer sequence, sequence location, direction, length, melting temperature, CG content (%), molecular weight, molar extinction coefficient, linguistic complexity (%), and *PQ*. For compatible primer pairs, the annealing temperature and PCR product size are also provided.

t3.1 **Table 3**
t3.2 **Program output of the primer design**

t3.3 Primer ID	Sequence (5′–3′)	Length (nt)	T_m (°C)	dG, kcal/mol	T_m 3′end (°C)	CG %	*LC*	*PQ*
t3.4 1F1_1_6-27	gggcggatcacttgaggtcagg	22	63.0	−29.7	40.0	63.6	88	87
t3.5 1F2_1_107-132	ggcaggagaatcacttcaacctggga	26	62.5	−33.6	40.2	53.8	89	89
t3.6 1F3_1_133-154	ggcggaggttgcagtgaaccga	22	64.4	−31.3	43.9	63.6	90	90
t3.7 1F4_1_153-175	gagaccgcgccactgcactccag	23	67.2	−33.7	42.8	69.6	76	76
t3.8 1F5_1_166-187	tgcactccagcctgggcgacag	22	66.3	−32.1	49.5	68.2	85	85
t3.9 1F6_1_176-197	cctgggcgacagcctgagactc	22	64.8	−30.8	41.4	68.2	80	80
t3.10 1F1_2_859-880	cttacggaggccgagatgggca	22	63.4	−30.6	46.9	63.6	88	87
t3.11 1F2_2_913-934	tggccaacatggtgaaaccctg	22	60.9	−28.9	41.6	54.5	82	80
t3.12 1F3_2_956-977	aattagctgggcatggtggcac	22	60.9	−29.1	47.8	54.5	80	73
t3.13 1F4_2_1013-1036	tggagctgaaccactgcactccag	24	63.2	−32.3	42.8	58.3	79	79
t3.14 1R1_1_412-434	acccagaagagcctgagtgggca	23	63.9	−31.7	46.7	60.9	83	83
t3.15 1R2_1_309-330	ccttgctcagctctggccatcc	22	63.6	−30.0	44.8	63.6	80	80
t3.16 1R3_1_299-320	ctctggccatcccagttcaagc	22	61.5	−28.9	42.2	59.1	88	80
t3.17 1R1_2_1206-1231	agatggggtttcaccatgttggccca	26	64.1	−34.8	44.6	53.8	80	80
t3.18 1R2_2_1166-1187	acctcaggtgatccacctgcct	22	61.7	−29.5	44.2	59.1	82	80
t3.19 1R3_2_1150-1174	cacctgcctcagcttcccaaagtgc	25	65.3	−34.2	40.6	60.0	84	84
t3.20 1R4_2_1132-1157	caaagtgcttggattacaggcgtgag	26	61.0	−33.0	42.4	50.0	93	93
t3.21 The separated output of a set of primer pair sequences with their theoretical PCR products								

(continued)

Table 3
(continued)

Primer ID	Sequence (5′–3′)	Length (nt)	T_m (°C)	dG, kcal/mol	T_m 3′end (°C)	CG %	*LC*	*PQ*	
Forward primer ID	Sequence (5′–3′)	T_m (°C)	PQ	Reverse primer ID	Sequence (5′–3′)	T_m (°C)	PQ	PCR Fragment Size (bp)	T^{opt} (°C)
1 F1_1_8-27	gcggatcacttgaggtcagg	58.9	87	1R3_1_301-320	ctctggccatcccagttcaa	57.6	80	313	63
1 F1_1_8-27	gcggatcacttgaggtcagg	58.9	87	1R4_1_291-310	cccagttcaagccatcccct	60.2	76	303	65
1 F2_1_49-68	caacgtggagctaggtatgg	56.0	80	1R1_1_409-429	gaagagcctgagtgggcacaa	60.0	85	381	62
1 F2_1_49-68	caacgtggagctaggtatgg	56.0	80	1R3_1_301-320	ctctggccatcccagttcaa	57.6	80	272	62
1 F2_1_49-68	caacgtggagctaggtatgg	56.0	80	1R4_1_291-310	cccagttcaagccatcccct	60.2	76	262	62
1 F3_1_109-130	caggagaatcacttcaacctgg	56.4	87	1R1_1_409-429	gaagagcctgagtgggcacaa	60.0	85	321	62
1 F3_1_109-130	caggagaatcacttcaacctgg	56.4	87	1R3_1_301-320	ctctggccatcccagttcaa	57.6	80	212	62
1 F3_1_109-130	caggagaatcacttcaacctgg	56.4	87	1R4_1_291-310	cccagttcaagccatcccct	60.2	76	202	62
1 F4_1_134-154	gcggaggttgcagtgaaccga	62.6	90	1R1_1_409-429	gaagagcctgagtgggcacaa	60.0	85	296	66
1 F4_1_134-154	gcggaggttgcagtgaaccga	62.6	90	1R2_1_311-333	tgtccttgctcagctctggccat	62.3	83	200	68
1 F4_1_134-154	gcggaggttgcagtgaaccga	62.6	90	1R3_1_301-320	ctctggccatcccagttcaa	57.6	80	187	63
1 F4_1_134-154	gcggaggttgcagtgaaccga	62.6	90	1R4_1_291-310	cccagttcaagccatcccct	60.2	76	177	65
1 F5_1_153-172	gagaccgcgccactgcactc	64.3	73	1R2_1_311-333	tgtccttgctcagctctggccat	62.3	83	181	68
1 F5_1_153-172	gagaccgcgccactgcactc	64.3	73	1R4_1_291-310	cccagttcaagccatcccct	60.2	76	158	65

5.9 Multiplex PCR Primer Design

Multiplex PCR is an approach commonly used to amplify several DNA target regions in a single reaction. The simultaneous amplification of many targets reduces the number of reactions that needs to be performed; multiplex PCR thus increases throughput efficiency. The design of multiplex PCR assays can be difficult because it involves extensive computational analyses of primer pairs for interactions. The multiplex PCR algorithm is based on the fast non-recursion method, with the software performing checks on product size and primers' thermodynamics parameters (enthalpy—**dH** and Gibb's free energy—**dG**) compatibility and cross-dimer formation for all primers. To achieve uniform amplification of the targets, the primers must be designed to bind with equal efficiencies to their targets. FastPCR can quickly design a set of multiplex PCR primers for all input sequences and/or multiplex targets within each sequence. PCR conditions may need to be adjusted, for example, by increasing or decreasing the annealing temperature so that all products are amplified equally efficiently.

To achieve uniform amplification, most existing multiplex primer design packages use primer melting temperature. In practical terms, the design of almost identical T_as and T_ms is very important. The melting temperatures of the PCR products are also important because these are related to annealing temperature values. The T_m of a PCR product directly depends on its CG content and its length; short products are more efficiently amplified at low PCR annealing temperatures (100 bp, 50–55 °C) than are long products (>3,000 bp, 65–72 °C). For most multiplex PCRs, there is usually a small variation (up to 5 °C) between the optimal T_as of all primer pairs. The annealing temperature must be optimal in order to maximize the likelihood of amplifying the target genomic sequences whilst minimizing the risk of nonspecific amplification. Further improvements can be achieved by selecting the optimal set of primers that maximize the range of common T_ms. Once prompted, FastPCR calculates multiplex PCR primer pairs for given target sequences. The speed of calculation depends on the numbers of target sequences and primer pairs involved.

An alternative way to design compatible multiplex PCR primer pairs is to use predesigned primers as references for the design of new primers. The user can select input options for the PCR products such as the minimum product size differences between the amplicons. Primer design conditions can be set individually for each given sequence or all primers can be designed using the same values; selected settings have higher priority for PCR primer or probe design than the general settings. The results include primers for individual sequences, compatible primers, product sizes, and annealing temperatures. Because clear differentiation of the products is dependent on using compatible primer pairs in the single reactions, the program recovers all potential variants of primer combinations for analyses of the chosen DNA regions and provides,

in tabular form, their compatibility including information on primer-dimers, cross-hybridization, product size overlaps, and similar alternative primer pairs based on T_m. The user may choose those alternative compatible primer pair combinations that provide the desired product sizes. Using the program, researchers can select predesigned primer pairs from a target for their desired types of PCRs by changing the filtering conditions as mentioned above. For example, a conventional multiplex PCR requires differently sized (at least by 10 bp) amplicons for a set of target genes, so the value for the minimum size difference between PCR products can be selected.

In addition to avoid amplifying different amplicons of the same size, multiplex PCR must also minimize the generation of primer-dimers and secondary products, which becomes more difficult with increasing numbers of primers in a reaction. To avoid the problem of nonspecific amplification, FastPCR selects primer pairs that give the most likelihood of producing only the amplicons of the target sequences by choosing sequences which avoid repeats or other motifs. The program allows the user to design not only compatible pairs of primers but also compatible single primers for different targets or sequences. The input data can be either a single sequence with a minimum two internal tasks or many sequences with or without internal tasks. Most of the parameters on the interface are self-explanatory. Optionally, the user is asked to provide the sequence and select oligonucleotide designing parameters.

On the **PCR Primer Design** tab, the user clicks on the **Multiplex PCR** option. The user then selects the limit for the number of primer pairs (the default is 100), the minimal size difference between multiplex PCR products (the default 10 bp), and the maximal difference between the T_as of the PCR products (the default is ±5 °C). After specifying inputs and primer design options, the user can execute the primer design task. Once the design of the primer set is completed, the result will appear in two **Result** text editors: **PCR primer design result** and **Multiplex PCR compatible pair primers**. Figure 8 shows the access to the PCR primer design output. The result text editor **PCR primer design result** will display the individual PCR primer design data, including the primer list and the compatible primer pairs for all the sequences and their internal tasks. The second **Multiplex PCR compatible pair primers** text editor collects final search results that are presented as a list of the sets of the compatible primer pairs for multiplex PCR.

5.10 Group-Specific PCR Primers

Group-specific amplification, also called family-specific and sequence-specific amplification, is an important tool for comparative studies of related genes, sequences, and genomes that can be applied to studies of evolution, especially for gene families and for cloning new related sequences. Specific targets such as homologous genes

or members of a transposable element family can be amplified to uncover DNA polymorphisms associated with these sequences or other genetic investigations. The overall strategy of designing group-specific PCR primers uses a hash index of 12-mers to identify common regions in the target sequences, followed by standard PCR primer design for the current sequence, and then testing of complementarity of these primers to the other sequences. FastPCR performs sequences multiple alignment or does accept alignment sequences input, giving it the flexibility to use a different strategy for primer design. If required, it can design degenerate PCR primers to amplify the polymorphic region of all related sequences.

The FastPCR package designs large sets of universal primer pairs for each given sequence, identifies conserved regions, and generates suitable primers for all given targets. The steps of the algorithm are performed automatically and the user can influence the settings for the primer design options. The quality of primer design is dependent on sequence relationships, genetic similarity, and suitability of the consensus sequence for the design of good primers. The software is able to generate group-specific primers for each set of sequences independently, which are suitable for all sequences. Primer alignment parameters for group-specific PCR primers are similar to those used for in silico PCR. The user chooses **Group-specific PCR** on the **PCR Primer Design** tab. After specifying inputs and PCR primer design options, the user can execute the PCR primer design task. The program takes multiple separate DNA sequences either in FASTA or at alignment formats.

[AU6]

Once the primer set design is complete, the result will appear in **Result** text editor, as the **PCR primer design result**. Figure 6 shows the access to PCR primer design output (Table 3). The result text editor **PCR primer design result** displays the individual group-specific PCR primer design data, including the primer list and compatible primer pairs for all the sequences and their internal tasks where suitable primers are found. In the case where an alignment has been input, the result text editor displays only one group-specific PCR primer design set, including degenerate primers, in the primer list as well compatible primer pairs for all the sequences.

5.11 Simple Sequence Repeat Locus Search and PCR Primer Design

SSRs or microsatellites are short tandem repeats of one or more bases. Microsatellites are ubiquitously distributed throughout eukaryotic genomes, often highly polymorphic in length, and thereby an important class of markers for population genetic studies. Our approach to locating SSRs is to analyze low-complexity regions in DNA by using linguistic sequence complexity. This method allows the detection of perfect and imperfect SSRs with a single, up to 10-base, repeat motif. Each entry sequence is processed for identification of SSRs and the SSR flanking regions are used to design compatible forward and reverse primers for their amplification by PCR.

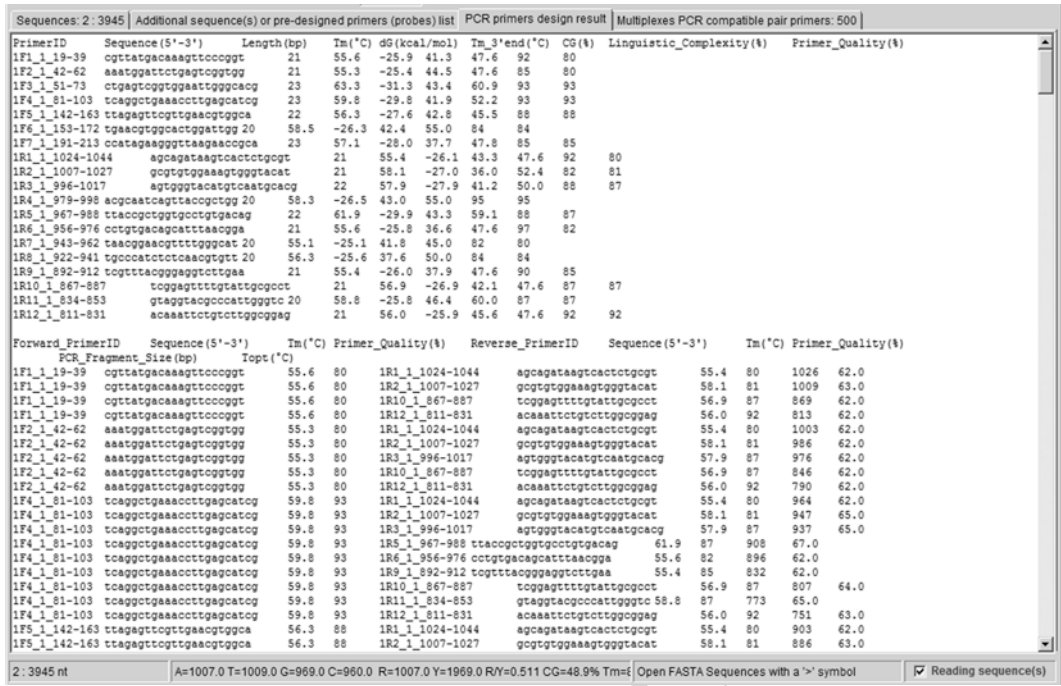


Fig. 6 An example of group-specific PCR result

FastPCR identifies all SSRs within each entry sequence and designs compatible PCR primer pairs for each SSR locus. The default PCR primer design parameters are that the primers must be within 100 bases from either side of the identified SSR. Often the sequences available around SSR loci are not suitable for designing good primers; the user can increase or decrease the distance from either side to find more efficient and compatible primer pairs. The capabilities of FastPCR make it a complete bioinformatics tool for the use of microsatellites as markers, from discovery through to primer design. For example, the user can specify PCR primer design to SSR loci within 200 bp around an SSR, with the command: "`-ssr/200.`" The software finds all SSR sites and then will design PCR primers and compatible primer pairs independently for each SSR locus.

5.12 Oligonucleotide Design for In Vitro Long Sequence Synthesis

The application to make long synthetic DNA molecules relies on the in vitro assembly of a set of short oligonucleotides, either for the LCR [37] or for assembly PCR [38]. These oligonucleotides should be adjacent on the same strand and overlap the complementary oligonucleotides from the second strand. There are two major parameters for designing oligonucleotides for gene synthesis for LCR or assembly PCR. First, the oligonucleotides

should share similar T_m values. Second, a given oligonucleotide sequence should be unique to avoid multiple nonspecific binding that may lead to incorrect assembly. The software must dynamically choose the length of the oligonucleotides to ensure both the specificity and the uniform T_m. The algorithm of FastPCR is able to design oligonucleotides for long sequences containing repeats and to minimize their potential nonspecific hybridization during 3′ end extension in PCR. For long sequence assembly, oligonucleotide design starts from the 5′ end of a given sequence; the oligonucleotide length is dynamically changed until a unique 3′ end has been found and the T_m of the oligonucleotide has reached the T_m threshold. All oligonucleotides are designed without gaps between them. The other strand is used for the design of the overlapping oligonucleotides using the same algorithm as above but with the T_m of the overlapping regions reaching the T_m −15 °C threshold. The composition of the sequence at the 3′ terminus is important because stability at the 3′ end of the double-stranded complexes will improve the specificity of extension by the polymerase. To reduce nonspecific polymerase extension (and ligation), the algorithm chooses only unique sequences for the 3′ terminus. Minimally, the last two nucleotides at the 3′ terminus must not be complementary to any nonspecific targets. Other complementary regions are less important for assembling multiple fragments by PCR and ligation.

The input data can comprise either a single or many sequences. Most of the parameters on the interface are self-explanatory. The user is asked to provide the sequence and select oligonucleotide designing parameters. The user clicks on **Oligo options** on the **Oligos Assembly** tab, and chooses the minimal oligonucleotide length and T_m threshold, which by default are 40 nt and 60 °C, respectively. The interface allows changing T_m calculation parameters. The search process runs after pressing **F5** or from menu bar or toolbox. The research result is presented as a list of oligonucleotides for both strands. On each strand, all oligonucleotides are adjacent with no gap between neighboring primers. An oligonucleotide will overlap two oligonucleotides from the complementary strand. The algorithm pays attention to avoid nonspecific oligonucleotide hybridization to repeated regions. Where it is not possible to design primers outside of repeated sequences, it is likewise difficult to find short specific oligonucleotides. The solution to this problem is to divide the sequence into short segments, design a set of oligonucleotides for each segment independently, and then combine all these segments in the second PCR for final amplification.

5.13 Polymerase Extension PCR for Fragment Assembly

Sequence-independent cloning, including ligation-independent cloning, requires generation of complementary single-stranded overhangs in both the vector and insertion fragments. Similarly,

multiple fragments can be joined or concatenated in an ordered manner using overlapping primers in PCR. Annealing of the complementary regions between different targets in the primer overlaps allows the polymerase to synthesize a contiguous fragment containing the target sequences during thermal cycling, a process called "overlap extension PCR" (OE-PCR) (Chapter 8) [39]. The efficiency depends on the T_m and on the length and uniqueness of the overlap. To achieve this, the program designs compatible forward and reverse primers at the ends of each fragment, and then extends the 5′ end of primers using sequences from the primers of the fragment that will be adjacent in the final product. The input sequence can be made of either a single or many sequences. The user needs to pay special attention to the preparation of the given sequences for assembly.

Users can specify the locations for both forward and reverse primers design using "[]" to bracket the region. The bracketed sequences will be used by the program for designing the overlapping primers. The program selects the overlapping area so that the primers from overlapping fragments are similar in size with optimal annealing temperatures. The program adds the required bases so that the T_m of the overlap is similar to, or higher than, the T_m of the initial primers. Primers are tested for dimer formation within the appropriate primer pairs. The user chooses **Polymerase extension cloning** (**OE-PCR**) on the **PCR Primer Design** tab and selects the limit for multiple-PCR-compatible combinations of pair primers (default is 100). After specifying sequence inputs and PCR primer design options, the user can execute the search task. Once the design of the primer sets is complete, the result will appear in two text editors: **PCR primer design result** and **PCR fragments assembling compatible pair primers**. The text editor **PCR primer design result** window displays the individual PCR primer design data, including the primer list and the compatible primer pairs for all sequences whose primers are found. The **PCR fragments assembling compatible pair primers** text editor collects the final search result and presents it as a list of sets of compatible primer pairs for individual fragment amplification and assembly. Figure 7 shows a sample result visualization window.

5.14 In Silico PCR

Modelling the hybridization of primers to targeted annealing sites is the only way to predict PCR products [7, 24, 40–44]. The last 10–12 bases at the 3′ end of primers are important for binding stability; single mismatches can reduce PCR efficiency, the effect increasing with proximity to the 3′ terminus. FastPCR allows simultaneous testing of single primers or a set of primers designed for multiplex PCR. It performs a fast, gapless alignment to test the complementarity of the primers to the target sequences. For in silico PCR, a quick alignment to detect primer locations on the reference sequence is performed by analyses of both strands using

this figure will be printed in b/w

Fig. 7 An example of polymerase extension PCR for fragment assembly result

a hash index of 7- to 12-mers (allowing up to one mismatch) and by calculating the local similarity for the whole primer. The parameters can be altered to allow different degrees of mismatches at the 3′ end of the primers. The parameters for quick alignment may be set: the minimum is 0–5 mismatches (default 2 mismatches) at 3′ end of primer. The program can also handle degenerate primers or probes, including those with 5′ or 3′ tail sequences. It includes the detection of non-Watson–Crick base-pairing in in silico PCR, e.g., the stable guanine mismatches G·G, G·T, and G·A. Probable PCR products can be found for both linear and circular templates in both standard and inverse PCR, as well as in multiplex PCR and using bisulphite-treated DNA. This in silico tool is useful for quickly analyzing primers or probes against target

sequences, for determining primer location, orientation, and efficiency of binding, and for calculating primer T_m and annealing temperature in PCR.

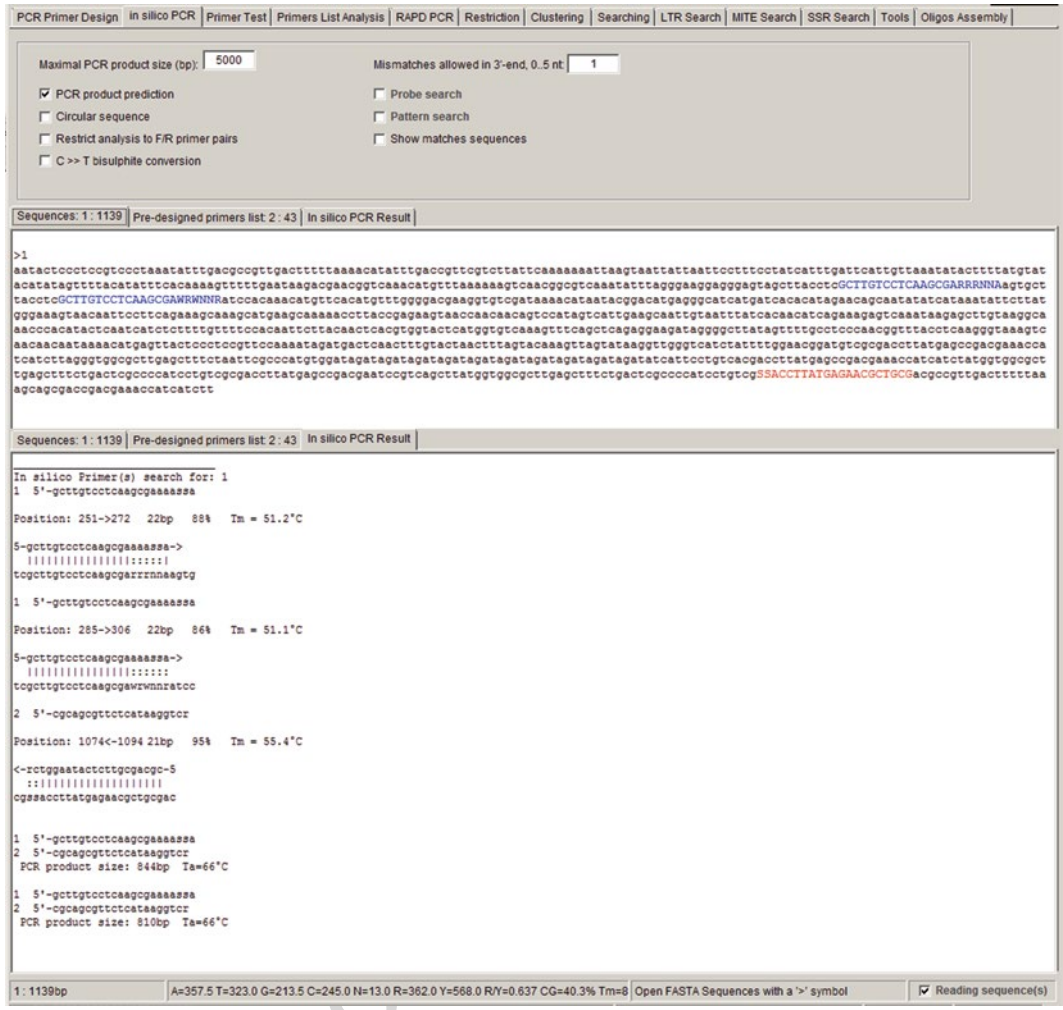
The user must input a preexisting primer list into a second **Additional sequence(s) or pre-designed primers (probes) list** text editor. The number of preexisting primers is not limited; it can be as many as the user needs. The target sequences can be entered either as multiple separate DNA sequences or by opening files from the selected folders. For in silico PCR against whole genome(s) or a list of chromosomes, the user must specify the directory containing the input. The program will be consistent: it will look at each file to find the position of the primers. The user can execute the search task with **F5** on the in silico **PCR** tab or can specify search options including stringency and PCR product detection settings. For the stringency options, users can specify the number of mismatches that the primers are allowed at 3′ terminus. The default specificity settings allow a maximum two mismatches within the 3′ end region of the primers. These mismatches within the 3′ end of the primers should not be located close to each other. Once the primer set design is complete, the results will appear in the text editors In silico **PCR Result**.

In silico **PCR Result** text editor reports the specificity of the primers (locations, including target position, similarity, and T_m), a summary of primer pairs in relation to the PCR template, and detailed information on each primer pair, including its length and T_a. It will show the target-specific primers that have been found. The actual targets will be listed along with detailed alignments between primers and targets (Fig. 8).

6 Primer Analyses

Individual and sets of primers are evaluated using FastPCR or the online software. They calculate primer T_ms using default or other formulae for features of the primers including normal and degenerate nucleotide combinations, CG content, extinction coefficient, unit conversion (nmol per OD), mass (μg per OD), molecular weight, and linguistic complexity and consider primer PCR efficiency. Users can select either DNA or RNA primers (online: PrimerAnalyser, http://primerdigital.com/tools/PrimerAnalyser.html) with normal or degenerate oligonucleotides or modifications with various labels (for example, inosine, uridine, or fluorescent dyes). Tools allow the choice of other nearest neighbor thermodynamic parameters or non-thermodynamic T_m calculation formulae.

For LNA modifications the four symbols, dA = E, dC = F, dG = J, and dT = L, are used. Both programs perform analyses on-



this figure will be printed in b/w

Fig. 8 An example of in silico PCR result

type, allowing users to see the results immediately on screen. They can also calculate the volume of solvent required to attain a specific concentration from the known mass (mg), OD, or moles of oligonucleotide.

All primers are analyzed for intra- and inter-primer interactions regarding formation of dimers. Primer(s) can efficiently hybridize using the 5′ end or the middle of the oligonucleotides. Even though such interactions are not efficiently extended by DNA polymerase, their formation reduces the effective primer concentration available for binding to the targets and their presence can strongly inhibit PCR because double-stranded DNA at high concentrations is a strong inhibitor of DNA polymerase (Fig. 9).

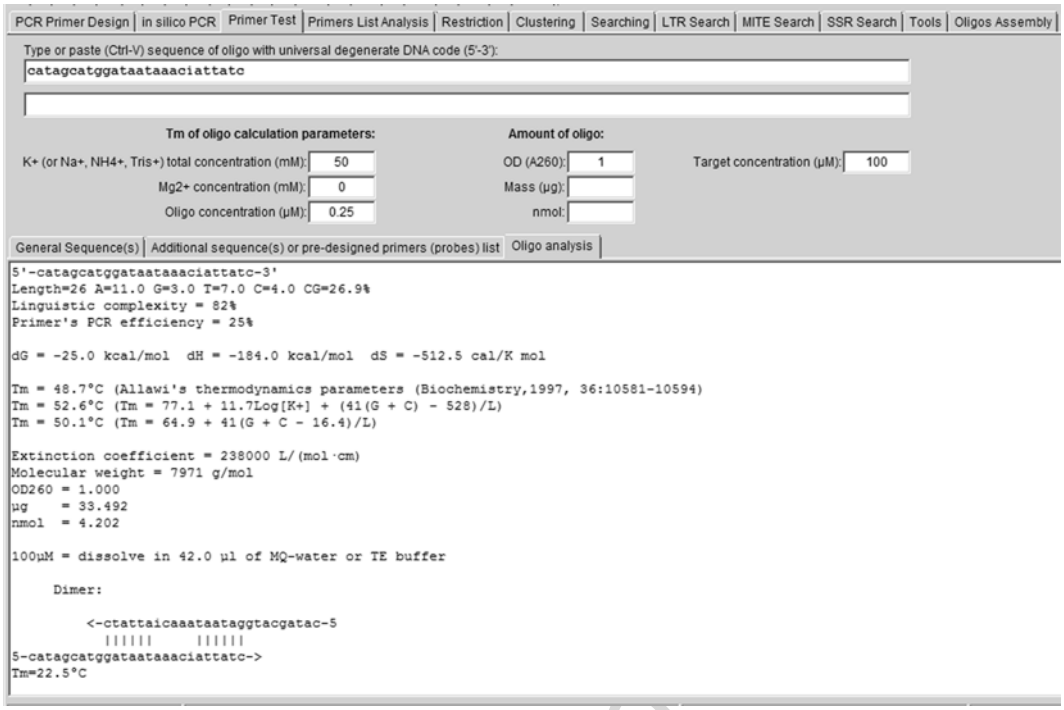


Fig. 9 Example result of the oligonucleotide analysis

7 Availability

The FastPCR software is available for download at http://primerdigital.com/fastpcr.html; the online version is available at http://primerdigital.com/tools/pcr.html. The program manual, licence agreement, and installation files can be found at http://primerdigital.com/fastpcr/. YouTube tutorial videos have been placed at http://www.youtube.com/user/primerdigital. Web tools are accessible at http://primerdigital.com/tools/.

Acknowledgments

Web tools are available free to academic institutions, provided that they are used for noncommercial research and education only. They may not be reproduced or distributed for commercial use. This work was partially supported by the companies PrimerDigital Ltd. and Oligomer Ltd. and by the Academy of Finland, Project 134079.

References

1. Untergasser A et al (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115. doi:10.1093/nar/gks596
2. Kalendar R, Lee D, Schulman AH (2009) FastPCR software for PCR primer and probe design and repeat search. Genes, Genomes and Genomics 3:1–14
3. Kalendar R, Lee D, Schulman AH (2011) Java web tools for PCR, *in silico* PCR, and oligonucleotide assembly and analysis. Genomics 98:137–144
4. Marshall OJ (2004) PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics 20:2471–2472
5. Owczarzy R et al (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res 36:W163–9. doi:10.1093/nar/gkn198
6. Bekaert M, Teeling EC (2008) UniPrime: a workflow-based platform for improved universal primer design. Nucleic Acids Res 36:e56. doi:10.1093/nar/gkn191
7. Ye J et al (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134. doi:10.1186/1471-2105-13-134
8. Giegerich R, Meyer F, Schleiermacher C (1996) GeneFisher—software support for the detection of postulated genes. Proc Int Conf Intell Syst Mol Biol 4:68–77
9. Gadberry MD et al (2005) Primaclade—a flexible tool to find conserved PCR primers across multiple species. Bioinformatics 21:1263–1264. doi:10.1093/bioinformatics/bti134
10. National Center for Biotechnology Information, National Library of Medicine, Building 38A, Bethesda, MD, 20894. http://blast.ncbi.nlm.nih.gov/blastcgihelp.shtml
11. Nomenclature for incompletely specified bases in nucleic acid sequences (1984) http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html
12. Allawi HT, SantaLucia J Jr (1997) Thermodynamics and NMR of internal G.T mismatches in DNA. Biochemistry 36:10581–10594. doi:10.1021/bi962590c
13. SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA 95:1460–1465
14. Le Novere N (2001) MELTING, computing the melting temperature of nucleic acid duplex. Bioinformatics 17:1226–1227. doi:10.1093/bioinformatics/17.12.1226
15. Bolton ET, McCarthy BJ (1962) A general method for the isolation of RNA complementary to DNA. Proc Natl Acad Sci USA 48:1390–1397
16. Guedin A et al (2010) How long is too long? Effects of loop size on G-quadruplex stability. Nucleic Acids Res 38:7858–68. doi:10.1093/nar/gkq639
17. Wallace RB et al (1979) Hybridization of synthetic oligodeoxyribonucleotides to ΦX 174 DNA: the effect of single base pair mismatch. Nucleic Acids Res 6:3543–57. doi:10.1093/nar/6.11.3543
18. von Ahsen N, Wittwer CT, Schutz E (2001) Oligonucleotide melting temperatures under PCR conditions: nearest-neighbor corrections for Mg^{2+}, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. Clin Chem 47:1956–1961
19. Kypr J et al (2009) Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res 37:1713–25. doi:10.1093/nar/gkp026
20. Gabrielian A, Bolshoy A (1999) Sequence complexity and DNA curvature. Comput Chem 23:263–74. doi:10.1016/S0097-8485(99)00007-8
21. Orlov YL, Potapov VN (2004) Complexity: an internet resource for analysis of DNA sequence complexity. Nucleic Acids Res 32:W628–33. doi:10.1093/nar/gkh466
22. Gilson MK et al (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 72:1047–69. doi:10.1016/S0006-3495(97)78756-3
23. Peyret N et al (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. Biochemistry 38:3468–3477. doi:10.1021/bi9825091
24. Watkins NE Jr, SantaLucia J Jr (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. Nucleic Acids Res 33:6258–67. doi:10.1093/nar/gki918
25. Sen D, Gilbert W (1992) Guanine quartet structures. Methods Enzymol 211:191–199
26. Il'icheva IA, Florent'ev VL (1992) Four-stranded complexes of oligonucleotides-quadruplexes. Mol Biol (Mosk) 26:512–531
27. Shing HP (1994) The non-B-DNA structure of d(CA/TG)n does not differ from that of Z-DNA. Proc Natl Acad Sci USA 91:9549–9553

28. SantaLucia J Jr, Hicks D (2004) The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct 33:415–40. doi:10.1146/annurev.biophys.32.110601.141800
29. Williams JGK et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. Nucleic Acids Res 18:6531–5. doi:10.1093/nar/18.22.6531
30. Welsh J, Mcclelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–8. doi:10.1093/nar/18.24.7213
31. Kalendar R, Schulman A (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nat Protoc 1:2478–84. doi:10.1038/nprot.2006.377
32. Chang RY, O'Donoughue LS, Bureau TE (2001) Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. Theor Appl Genet 102:773–781
33. Nelson DL et al (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci USA 86: 6686–6690
34. Sinnett D et al (1990) Alumorphs—human DNA polymorphisms detected by polymerase chain reaction using Alu-specific primers. Genomics 7:331–334
35. Jurka J et al (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genom Res 110:462–7. doi:10.1159/000084979
36. TREP, the Triticeae Repeat Sequence Database (2008) http://wheat.pw.usda.gov/ITMI/Repeats/
37. Landegren U et al (1988) A ligase-mediated gene detection technique. Science 241:1077–1080
38. Higasa K, Hayashi K (2002) Ordered catenation of sequence-tagged sites and multiplexed SNP genotyping by sequencing. Nucleic Acids Res 30:E11
39. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4:e6441. doi:10.1371/journal.pone.0006441
40. Cao YF et al (2005) Information theory-based algorithm for *in silico* prediction of PCR products with whole genomic sequences as templates. BMC Bioinformatics 6:190. doi:10.1186/1471-2105-6-190
41. Rubin E, Levy AA (1996) A mathematical model and a computerized simulation of PCR using complex templates. Nucleic Acids Res 24:3538–45. doi:10.1093/nar/24.18.3538
42. Lexa M, Valle G (2003) PRIMEX: rapid identification of oligonucleotide matches in whole genomes. Bioinformatics 19:2486–2488
43. Nishigaki K et al (2000) Whole genome sequence-enabled prediction of sequences performed for random PCR products of *Escherichia coli*. Nucleic Acids Res 28:1879–1884
44. Rotmistrovsky K, Jang W, Schuler GD (2004) A web server for performing electronic PCR. Nucleic Acids Res 32:W108–12. doi:10.1093/nar/gkh450

Author Queries

Chapter No.: 18 0002067381

Queries	Details Required	Author's Response
AU1	Please check if the edit made in the sentence "It calculates the melting … resuspension calculator." is appropriate.	
AU2	Please check if the edit made in the sentence "Getting started with a … to type." is appropriate.	
AU3	Please check if the edit made in the sentence "Primer pairs are analyzed … melting temperatures." is appropriate.	
AU4	Please check if the edit made in the sentence "Multiplex PCR can be … of both." is appropriate.	
AU5	Please provide closing parenthesis in sentence "Design primers with a specific restriction …"	
AU6	Please check if the edit made in the sentence "The program takes multiple … alignment formats." is appropriate.	