Chapter 2

Introduction on Using the FastPCR Software and the Related Java Web Tools for PCR and Oligonucleotide Assembly and Analysis

Ruslan Kalendar, Timofey V. Tselykh, Bekbolat Khassenov, and Erlan M. Ramanculov

Abstract

This chapter introduces the FastPCR software as an integrated tool environment for PCR primer and probe design, which predicts properties of oligonucleotides based on experimental studies of the PCR efficiency. The software provides comprehensive facilities for designing primers for most PCR applications and their combinations. These include the standard PCR as well as the multiplex, long-distance, inverse, real-time, group-specific, unique, overlap extension PCR for multi-fragments assembling cloning and loop-mediated isothermal amplification (LAMP). It also contains a built-in program to design oligonucleotide sets both for long sequence assembly by ligase chain reaction and for design of amplicons that tile across a region(s) of interest. The software calculates the melting temperature for the standard and degenerate oligonucleotides including locked nucleic acid (LNA) and other modifications. It also provides analyses for a set of primers with the prediction of oligonucleotide properties, dimer and G/C-quadruplex detection, linguistic complexity as well as a primer dilution and resuspension calculator. The program consists of various bioinformatical tools for analysis of sequences with the GC or AT skew, CG% and GA% content, and the purine–pyrimidine skew. It also analyzes the linguistic sequence complexity and performs generation of random DNA sequence as well as restriction endonucleases analysis. The program allows to find or create restriction enzyme recognition sites for coding sequences and supports the clustering of sequences. It performs efficient and complete detection of various repeat types with visual display. The FastPCR software allows the sequence file batch processing that is essential for automation. The program is available for download at http://primerdigital.com/fastpcr.html, and its online version is located at http://primerdigital.com/tools/pcr.html.

Key words PCR primer design, Isothermal amplification of nucleic acids, Software probe design, DNA primers, DNA primers nucleic acid hybridization, Degenerate PCR, Tiling arrays, Primer linguistic complexity, Ligase chain reaction

1 Introduction

The polymerase chain reaction (PCR) is a fundamental tool for the molecular biology genetic analyses, and is the most important practical molecular technique for the research laboratory. At the

Lucília Domingues (ed.), *PCR: Methods and Protocols*, Methods in Molecular Biology, vol. 1620,
DOI 10.1007/978-1-4939-7060-5_2,

time, there is a variety of thermocycling and isothermal techniques used for amplification of nucleic acids. The thermocycling techniques use a temperature cycling to drive the repeated cycles of DNA synthesis to produce large amounts of new DNA being synthesized in proportion to the original amount of a template DNA. The template DNA strands are amplified by the repeated cycles of several temperature steps including (1) heat denature (denaturation of a double-strand DNA template on single strands), (2) annealing (annealing of primers on a single-strand DNA template), and (3) extension reaction (elongation of primers by the DNA polymerase) [1, 2]. The utility of the method is dependent on the identification of unique primer sequences and the PCR-efficient primers design.

The nucleic acid amplification techniques (NAAT) for detection of target sequences use a probe (quantitative PCR) or microarrays specifically modified to be detected during the course of an amplification reaction. For instance, the TaqMan and Molecular Beacons assays both use a reporter and a quencher dye attached to the probe. The TaqMan assay is an example of technique for the homogeneous nucleic acid detection of a target sequence that employs a modified probe. The TaqMan probes hybridize to the target sequence while it is being amplified. The enzyme responsible for amplification of the target sequence also degrades any hybridized probe in its path. Among the technologies developed for target detection and quantification, the most promising is probably the one involving the molecular beacons. The conventional molecular beacons are single-stranded oligonucleotide hybridization probes that form a stem-and-loop structure. The loop part contains the sequence being complementary to the target nucleic acid (either DNA or RNA). The stem part is formed due to hybridization of the complementary sequence of the 3′ end with the 5′ end. The ends of the molecular beacon are self-complementary and are not supposed to hybridize to a target sequence.

Another approach is the use of isothermal techniques for DNA amplification that do not rely on thermocycling to drive the amplification reaction. A number of such techniques have also been developed so far. The isothermal techniques utilize DNA polymerases with strand-displacement activity and are used as a nucleic acid amplification method that can obviate the need for the repeated temperature cycles. For example, to run a reaction in LAMP [3], the mixture of various reagents is held at a constant temperature (in the vicinity of 65 °C) and includes nucleic acid strands of the template, oligonucleotide primers, the strand displacement-type DNA synthetase, and nucleic acid monomers.

Primer design is a critical step in all types of PCR methods to ensure specific and efficient amplification of a target sequence [4–10]. The specificity of the oligonucleotides is one of the most important factors for good PCR, since optimal primers should hybridize only to the target sequence. Particularly it is essential when complex genomic DNA is used as the template. Amplification problems during reaction can arise due to primers annealing to repetitious sequences (retrotransposons, DNA transposons, or tandem repeats) [11]. 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). Primers complementary to repetitious DNA may produce many nonspecific bands in a single-primer amplification and compromise the performance of unique PCRs. However, the generation of inverted repeat sequences is widely exploited in the generic DNA fingerprinting methods. Often only one primer is used in these PCR reactions, the ends of the PCR products should consist of an inverted repeat complementary to the sequence of the primer.

Therefore, the use of primers is not limited to the PCR nucleic acid amplification but extends to a number of standard molecular biological methods. These considerations motivate the development of a new, high-throughput and stand-alone software that includes the PCR primers design capabilities.

The adaptation of the PCR method for different applications requires development of a new criteria for the PCR primer and probe design to cover approaches such as real-time PCR, group-specific and unique PCR, combinations of multiple primers in multiplex PCR. The criteria must also allow the possibility of the extension PCR for multi-fragments assembling cloning, bisulfite modification assays and of isothermal DNA amplification techniques, as well as a way to design oligonucleotide sets for long sequence assembly by ligase chain reaction, discovery of simple sequence repeats and their amplification as diagnostic markers, TaqMan, molecular beacon, and microarray oligonucleotides [9, 12, 13].

When developing the FastPCR software (Table 1), our aim was to create a practical and easy-to-use tool for the routine manipulation and analysis of sequences for most PCR applications. The adopted parameters are based on our experimental data for efficient PCR and are translated into the algorithms in order to design combinations of primer pairs for the optimal amplification.

This chapter describes the FastPCR software as a complete solution for the PCR primers design. In particular, we describe here the interface, configuration, and main capabilities of the

Table 1

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

Features
PCR tool provides comprehensive facilities for designing primers for most PCR applications and their combinations:
Standard, multiplex, long distance, inverse, real-time PCR (LUX and self-reporting), group-specific (universal primers for genetically related DNA sequences) or unique (specific primers for each from genetically related DNA sequences), overlap extension PCR (OE-PCR)—multi-fragments assembling cloning and loop-mediated isothermal amplification (LAMP); single primer PCR (design of PCR primers from close located inverted repeat), automatically detecting simple sequence repeat (SSR) loci and direct PCR primer design, amino acid sequence degenerate PCR, polymerase chain assembly (PCA), design amplicons that tile across a region(s) of interest
A long oligonucleotide can be designed for microarray analyses and dual-labeled oligonucleotides for probes such as molecular beacons
PCA or oligonucleotides assembly—created to automate the design oligonucleotide sets for long sequence assembly by PCR
In silico (virtual) PCR or multiple primer or probe searches, or in silico PCR against whole genome(s) or a list of chromosome prediction 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
PCR efficiency, linguistic complexity, dimer and G/C-quadruplex detection, dilution and resuspension calculator
Analysis of features of multiple primers simultaneously, including T_m, GC content, linguistic complexity, dimer formation; optimal T_a
Tool for identifying SSR loci by analyzing the low complexity regions of input sequences
Tool for restriction I–II–III types enzymes and homing endonucleases analysis, find or create restriction enzyme recognition sites for coding sequences
Tool for searching for similar sequences (or primers)
Translates nucleotide (DNA/RNA) sequences to the corresponding peptide sequence in all six frames for standard and degenerate DNA and modifications (inosine, uridine)
The program includes various bioinformatics tools for patterns analysis of sequences with GC:(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%, GA% content and purine–pyrimidine skew, the melting temperature, considers linguistic sequence complexity profiles

program. On specific examples, we show how to use the FastPCR software for the high-throughput PCR assay of large amounts of data. The chapter also provides the quick start guide for an immediate application of the FastPCR software for PCR primers design and genome analyses during planning of experiments as well as for prediction of the results.

2 Software

2.1 Supported Platforms and Dependencies

The online version of the FastPCR software (http://primerdigital.com/tools/pcr.html) is written in Java with NetBeans IDE (Oracle) and requires the Java Runtime Environment (JRE 8) on a computer. The program can be used with any operating system (64-bit OS preferred for large chromosome files). The stand-alone version of the FastPCR software (http://primerdigital.com/fastpcr.html) can be used with any version of Microsoft Windows.

2.2 Downloading and Installing

The online version of the FastPCR software requires the Java Runtime Environment (http://www.oracle.com/technetwork/java/javase/downloads/). The Oracle company strongly recommends that all Java users upgrade to the latest Java 8 release.

The online version FastPCR software users need to add the URL (http://primerdigital.com/) of this application to the Exception Site List (https://www.java.com/en/download/faq/exception_sitelist.xml), which is located under the Security tab of the Java Control Panel (http://www.java.com/en/download/help/appsecuritydialogs.xml). Adding this application URL to the list will allow it to run after presenting some security warnings. Existence of the application URL in the Exception list allows users to run Rich Internet Applications (RIAs) that would normally be blocked by security checks. The exception site list is managed in the Security tab of the Java Control Panel. The list is shown in the tab. To add, edit or remove a URL from the list, use the following:

- Click on the Edit Site List button.
- Click the Add in the Exception Site List window.
- Click in the empty field under Location field to enter the URL: http://primerdigital.com/.
- Click OK to save the URL that you entered. If you click Cancel, the URLs will not be saved.
- Click Continue in the Security Warning dialog.

In order to enhance security, the certificate revocation checking feature has been enabled by default (starting from Java 7). Before Java attempts to launch a signed application, the associated certificate will be validated to ensure that it has not been revoked by the issuing authority. This feature has been implemented using both Certificate Revocation Lists (CRLs) and Online Certificate Status Protocol (OCSP) mechanisms.

Optionally, users can download self-signed certificates file (http://primerdigital.com/j/primerdigital.cer) and import it to "Signer CA" (Certificate Authority) from the Java Control Panel.

Finally, users need to set "Security Level" to "High" under the Security tab of the Java Control Panel (as it is shown on the

picture: http://primerdigital.com/image/primerdigital_certificate_big.png).

Running and downloading online jPCR software from a desktop computer using the Java Web Start (JavaWS) command:

javaws http://primerdigital.com/j/pcr.jnlp

or

javaws http://primerdigital.com/j/pcr2.jnlp

Alternatively, users can run the software directly from the WEB site: http://primerdigital.com/tools/pcr.html.

2.3 Availability

The FastPCR software is available for download at http://primerdigital.com/fastpcr.html and online version at http://primerdigital.com/tools/pcr.html. Web tools are available here: http://primerdigital.com/tools/. The program manual, license agreement, and files for installation are available on the Internet at http://primerdigital.com/fastpcr/ and YouTube tutorial videos at http://www.youtube.com/user/primerdigital.

3 The Interface

3.1 Inputs to FastPCR

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

Getting started with a basic project in the FastPCR software is as easy as opening a new or existing file as well as a copy-paste or starting to type.

There are three independent text editors at different tabs: "General Sequence(s)," "Additional sequence(s) or pre-designed primers (probes) list," and "Result report."

The two first text editors are necessary for loading sequences for analysis, the "General Sequence(s)" text editor is designed for working with the project sequences; the "Additional sequence(s) or pre-designed primers (probes) list" text editor is applied for special and additional sequences, for example, for predesigned primers, multiple query sequences or for the numbers for input.

3.2 Program Output

The FastPCR software automatically generates results in the third text editor named as "Result report." It is performed in a tabulated format for transferring the results to Microsoft Excel sheet from a clipboard with a copy-paste method or to save them as .XLS or .RTF text file, compatible with both MS Excel and Open Office. Moreover, the program also produces results containing the list of primers, a set of primer pair sequences with their theoretical PCR

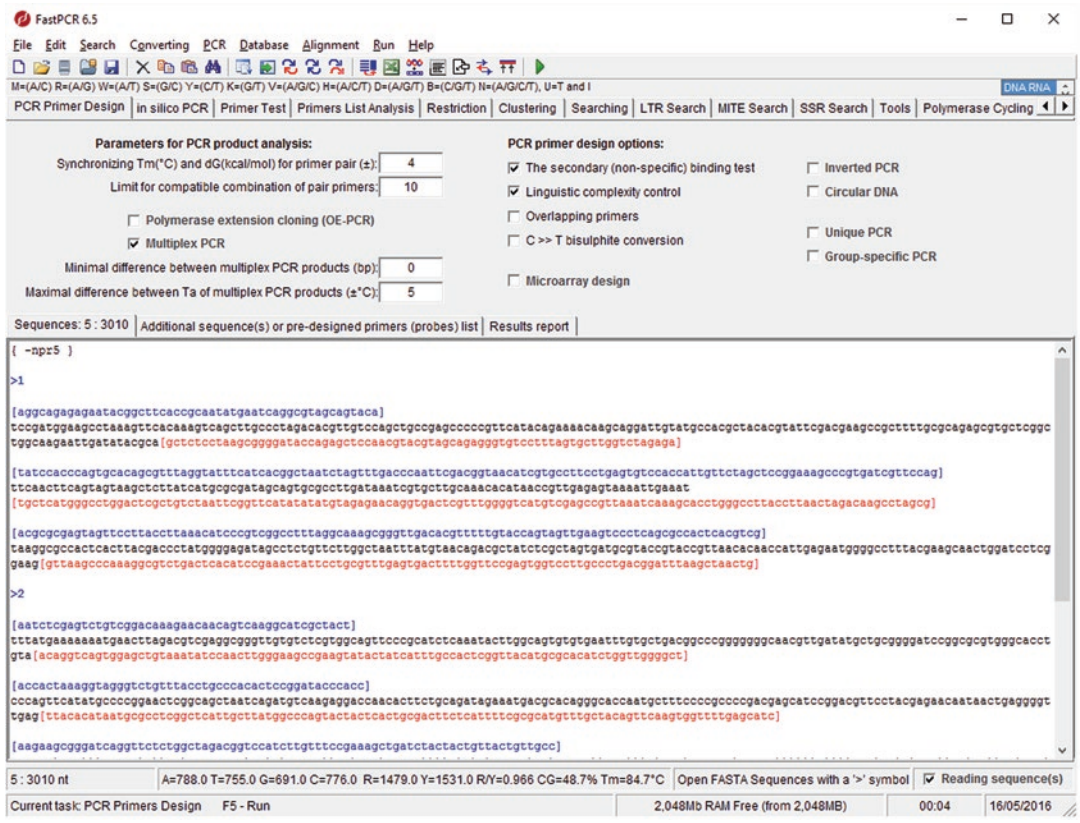

Fig. 1 The FastPCR sequence editor and user interface

products and for multiplex PCR. It also provides a user with the results of the calculation of multiple PCR primers for a given target sequences. In addition, the output shows the optimal annealing temperature for each primer pair as well as the size of PCR product and complete information for each primer designed and for the multiplex PCR product set.

3.3 Sequence Entry

Prepare your sequence data file using a text editor (Notepad, WordPad, Word), and save in a ASCII text/plain format or a Rich Text Format (.rtf). The program takes either a single sequence or accepts multiple separate DNA sequences in the FASTA, tabulated (two columns from either MS Excel sheet or Word table), EMBL, MEGA, GenBank and MSF, DIALIGN format or in the simple alignment and Blast Queue web alignments result formats. The template length is not limited.

The FastPCR software clipboard allows user to copy/paste operations with text or table from Microsoft Office documents or Excel worksheets (or other programs) and use them in another MS Office document. Importantly, the full target sequences must be

prepared in the same format. User can type or import data from file(s) into the "General Sequence(s)," "Additional sequence(s) or pre-designed primers (probes) list" editors. In the FastPCR software users have several options on how to open a file while starting the program. The user can open the original file as read-only in order to work with text editors, or open file to memory without opening to text editors, which is the better choice for large file(s). An alternative way to open files is by showing to the program the entered folder; the program will open each file while executing task without opening it to text editor. Additionally, users can open all files from a selected folder and program will combine all the files in a text editor. For example, this feature can be applied for converting all files from the selected folder into a single file presenting the list of FASTA sequences. As opposed to this feature, the program allows splitting of FASTA sequences into individual files in a selected folder. At the time, users can download file(s) from the NCBI Genebank server by accession number(s). The identifier may be a Genebank accession, accession.version or gi's (e.g., p01013, AAA68881.1, 129295) and a bar-separated NCBI sequence identifier (e.g., gi|129295). Spaces (or comma) between identifiers in the input will lead to downloading all sequences simultaneously to text-editor (spaces before or after the identifier are allowed).

When a sequence file is open, the FastPCR software displays the information about the opened sequences and the sequences format. The information status bar shows the amount of sequences, total sequences length (in nucleotides), nucleotide compositions, purine, pyrimidine, and CG% contents.

When users save a file from the current text editor, they must choose the file format to save the file, e.g., Rich Text Format (.rtf), MS Excel worksheet (.xls), or text/plain format (.txt).

3.4 FASTA Format Description

The FastPCR software normally is expected to read files with sequences in FASTA format (http://blast.ncbi.nlm.nih.gov/blastcgihelp.shtml). The FASTA format have a highest priority and is simple as the raw sequence proceeded by definition line. The definition line begins with a ">" sign that can be optionally followed by a sequence name of any length and amount of words with no space in between. There can be many sequences listed in the same file. The format requires that a new sequence always starts with a new ">" symbol. It is important to press "Enter" key at the end of each definition line to help the FastPCR software recognize the end and beginning of sequence and the sequence name. Make sure the first line starts with a ">" and, optionally, a header description.

Degenerate DNA sequences are accepted as IUPAC code that is an extended vocabulary of 11 letters that allows the description of ambiguous DNA code [14]. 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-accepted 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), Y(Tyr).

3.5 Alignment Format Description

There are many different programs, which produce many different types of alignment formats. The use of a standard set of formats enables creation of programs that can read the results originating from many different sources. 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 alignment file may begin with as many lines of comment or description as required. The first mandatory line that is recognized as part of the MSF file contains the text MSF, or contains the text Alignment as simple alignment format, or contains the texts DIALIGN or MEGA recognized as alignments from these programs. There then follows lines with each sequence line starting 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 primer design parameters, controllable either by the user or automatically, include the primer length (12–500 nt), melting temperature for short primers calculated by nearest neighbor thermodynamic parameters, the 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 represented by the general nucleotide structure of the primer such as linguistic complexity (nucleotide arrangement and composition), primer specificity, the melting temperature of the whole primer and the melting temperature at the 3′ and 5′ termini as well as the self-complementarity and secondary (nonspecific) binding.

The software can dynamically optimize the best primer length for entered parameters. All PCR primer (probe) design parameters are flexible and changeable according to the features of the

Table 2

Default primer design selection criteria

Criteria	Default	Ideal
length (nt)	18–22	>21
T_m range (°C)[a]	49–60	60–68
T_m[a] 12 bases at 3′ end	30–50	40–45
GC (%)	40–70	45–55
3′-end composition (5′-nnn-3′)	swh ssw wsh	ssa, sws, wss
Sequence linguistic complexity (LC, %)[b]	≥70	>90
Sequence Quality (PQ, %)	≥75	>80

[a]Nearest neighbor thermodynamic parameters SantaLucia [16]
[b]Sequence linguistic complexity measurement was performed using the alphabet-capacity *l*-gram method

analyzed sequence and research task. Primer pairs are analyzed for cross-hybridization, specificity of both primers and, optionally, are 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.

4.2 Melting Temperature (T_m) Calculation

The 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 are calculated in the default setting using nearest neighbor thermodynamic parameters [15, 16]. The T_m is calculated using a formula based on nearest neighbor thermodynamic theory with unified d*S*, d*H* and d*G* parameters:

$$T_m\left(^{\circ}C\right) = \frac{dH}{dS + R\ln\left(\frac{c}{f}\right) + 0.368\left(L-1\right)\ln\left(\left[K^+\right]\right)} - 273.15,$$

where d*H* is enthalpy for helix formation, d*S* is entropy for helix formation, *R* is molar gas constant (1.987 cal/K mol), *c* is the nucleic acid molar concentration (250 pM), [K+] is salt molar concentration (default value is 50 mM). The $f = 4$ when the two strands are different and $f = 1$ when self-hybridization takes place.

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 [5, 6]. Mismatched pairs can be taken into account since the parameters provide for DNA–DNA duplexes and the dangling ends, the unmatched terminal nucleotides [17–19]. The melting temperature for primer (probe) self- or cross-dimers and for in silico PCR experiments with oligonucleotides with mismatches to the target sequence is calculated using values for the thermodynamic parameters for a nucleic acid duplex.

The FastPCR software allows the choice of other nearest neighbor thermodynamic parameters or simple nonthermodynamic T_m calculation formulae. For nonthermodynamic T_m calculation, we suggest using simple formulae; the Wallace–Ikatura rule [20] 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 shorter 15 bases:

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

Alternative and more advanced nonthermodynamic formulae:

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

or formulae [21]:

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

where L is the length of primer and $[G + C]$ is the number of G's and C's, $[K+]$ 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. Alternatively, the melting temperature of the PCR product may be calculated using the formulae [18]:

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

4.3 Linguistic Complexity of Sequence and Nucleotide-Skew Analysis

The sequence analysis complexity calculation method can be used to search for conserved regions between compared sequences for the detection of low-complexity regions including simple sequence repeats, imperfect direct or inverted repeats, polypurine and polypyrimidine triple-stranded DNA structures, and four-stranded structures (such as G/C-quadruplexes).

Linguistic complexity measurements are performed using the alphabet-capacity *L*-gram method [22, 23] along the whole sequence length and calculated as the sum of the observed range (*xi*) 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, 100% being the highest level:

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

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

$$L = \lceil \log_4 \left(\frac{s}{3} \right) + 1 \rceil$$

For example, the sequence 5′-*GTGTGTGTGTGTGTGC*, 17 nt ($L = 3$), contains three nucleotides (G, T, and C), but expected $E = 4$ variants; three variants of two-nucleotides (GT, TG, and GC), but expected $E = (16 - 1)$ variants; three variants of three-nucleotides (GTG, TGT, and TGC), and expected $E = (16 - 2)$ variants. The complexity value is LC = 100 × (3 + 3 + 3) / (4 + 16 − 1 + 16 − 2) = 27.3%.

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 base 3′ terminus is calculated preferably by nearest-neighbor thermodynamic parameters [16]. The composition of the sequence at the 3′ terminus is important; primers with two terminal C/G bases are recommended for increased PCR efficiency [24]. 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.

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 following parameters: total sequence and purine–pyrimidine sequence complexity, the melting temperature of the whole primer and of the terminal 3′ and 5′ 12 bases. Self-complementarity, which gives rise to possible dimer and hairpin structures, reduces the final value. 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 and repeat content. PQ values of 80 and higher allow for the rapid choice of the best PCR primer pair combination. 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 reaction. 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 are avoided, especially avoiding complementarity in the 3′-ends of primers from where the polymerase will extend. Stable primer dimer formation is very effective at inhibiting PCR since the dimers formed are amplified efficiently and compete with the intended target.

Primer dimer prediction is based on analysis of non-gap local alignment and the stability of both the 3′ end and the central part of the primers. Primers will be rejected when they have the potential to form stable dimers depending on nucleotides composition and with at least five bases at the 3′ end or seven bases at the central part. Tools calculate T_m for primer dimers with mismatches for pure, mixed, or modified (inosine, uridine, or locked nucleic acid) bases using averaged nearest neighbor thermodynamic parameters provided for DNA–DNA duplexes [15–17, 25, 26].

Besides Watson–Crick base pairing, there is a variety of other hydrogen bonding possible configurations [27–30] such as G/C-quadruplexes or wobble base pairs that the FastPCR software detects.

The program includes the detection of the alternative hydrogen bonding to Watson–Crick base pairing at the primer-dimers and in silico PCR primer binding site detection. The mismatches stability follows the trend 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, since it forms the strongest base pair and the strongest mismatches. On the other hand, "C" is the most discriminating base, since it forms the strongest pair and the three weakest mismatches [25, 31].

Therefore, the tools are also looking at 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 [30]. 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. Intermolecular G-quadruplex-forming sequences are detected according to the formula …$G_{m1}X_nG_{m2}$…, where m is the number of G residues in each G-tract (m_1, $m_2 \geq 3$); the gap Xn ($n \leq 2 \times$ minimal (m_1:m_2)) can be any combination of residues, including G [19]. The gap sequences (Xn) 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 range of temperatures where efficiency of PCR amplification is maximal without nonspecific products. The most important values for estimating the T_a is the primer quality, the T_m of the primers and the length of PCR fragment. Primers with high T_m's (>60 °C) can be used in PCRs with a wide T_a range compared to primers with low T_m's (<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 °C higher than the T_m of the primer to favor primer target duplex formation:

$$T_a\left(^{\circ}\mathrm{C}\right) = T_m^{min} + \ln L\,,$$

where L is length of PCR fragment.

4.7 The Secondary Nonspecific Binding Test; Alternative Amplification

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—Random amplified polymorphic DNA (RAPD) and Arbitrarily Primed (AP)-PCR [32, 33]. Because only one primer is used in these PCR reactions, the ends of the products must be reverse complements and thus can form stem-loops.

The techniques of inter-retrotransposon amplification polymorphism (IRAP), retrotransposon-microsatellite amplification polymorphisms (REMAP), inter-MITE amplification [34, 35], and *Alu*-repeat polymorphism [36, 37] have exploited these 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 of the primer 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 [38] or TREP (http://wheat.pw.usda.gov/ITMI/Repeats/).

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 as well as several mismatches at the 5′ end of the primer allow amplification, with reduced efficiency of the amplification.

On the other hand, the presence of multiple primer binding sites does not necessarily lead to an alternative amplification, because, for amplification, both primers have to be located close to each other. The close location of the primers at correct orientation to each other and efficient binding of each DNA target determines the probability of alternative amplification.

By default, the FastPCR performs a nonspecific binding test for each given sequence. Additionally, the software allows this test to be performed against a reference sequence or sequences (e.g., BAC, YAC) or one's own database. Primers that bind to more than one location on current sequences will be rejected. Even though the nonspecific primer binding test 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. The software pays particular attention to the 3′ end portion of the primer and calculates the similarity of 3′ end of the primer to target (the length is chosen by the user) to determine the stability of the 3′-terminus. The secondary nonspecific primer binding test is based on a quick, non-gapped local alignment (that allows one mismatch within a hash index of 9-mers) screening between the reference and input sequence.

The software offers flexible specificity stringency options. User can specify the number of mismatches that primers must have to unintended targets at 3′ end region where these mismatches must be present. The default specificity settings are that at least one mismatch in the last five bases at the 3′ end of primer.

5 Methods

Once the input files are selected or sequence copy-paste to **General Sequence(s)** text editor, the software provides various execution features. Figure 2 provides a primer design window from user perspective.

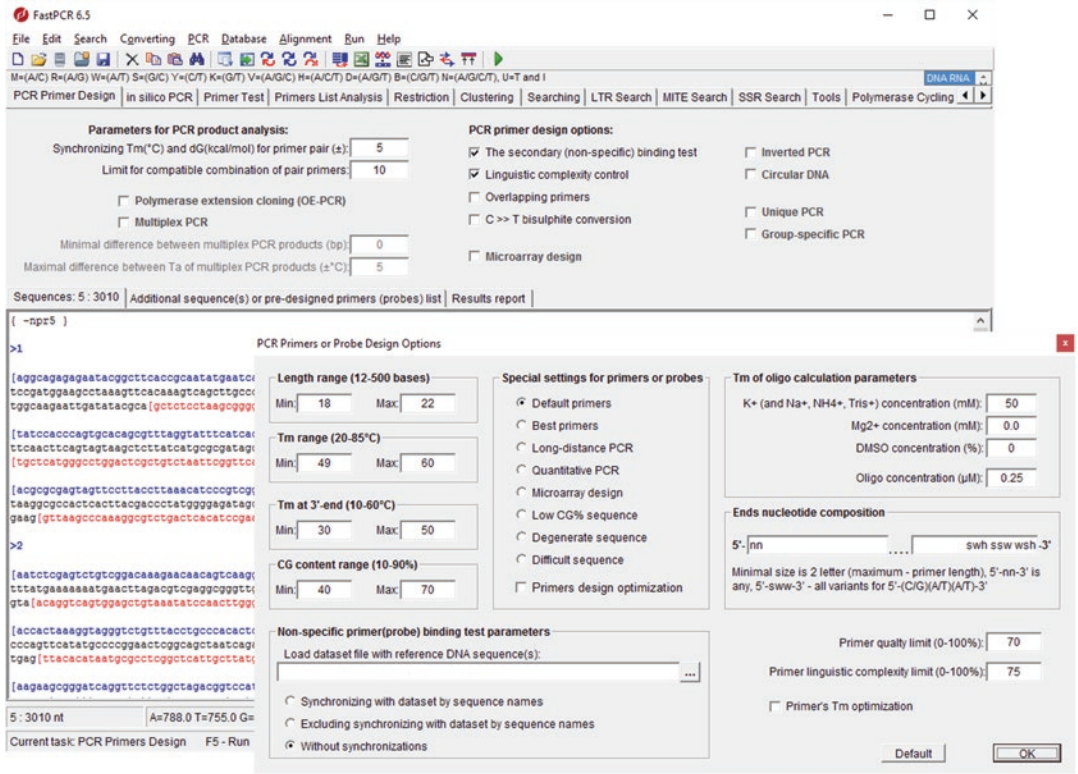

Fig. 2 The "PCR Primers or Probes Design Options" window

5.1 Execution of the Selected Task

User selects the ribbon with the task needed. The program will only perform the selected task. Depending on thetask selected, the program will show on the status bar the name of the executive task (Fig. 3) by "**Press F5**." To execute the current task, user can either press the **F5** key or click the arrow on the toolbar with the mouse.

Once the executive task is complete, the result is shown in the **Result report** text editor. Figure 4 shows a sample result visualization window.

5.2 PCR Primer Design Options

The "**PCR Primer Design**" Tab contains various execution options to easily select the type of PCR and most important PCR parameters. Figure 2 shows "**PCR Primers or Probe Design Options**" panel. Once 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 to all sequences. For individual PCR primer design options for each sequence, user can type special commands at the header of sequence (http://primerdigital.com/soft/pcr_help.html). Typically, the user does not need to use commands to manage PCR primer design, all these commands use optionally and only for advanced tasks. User can type at text editor this help command:

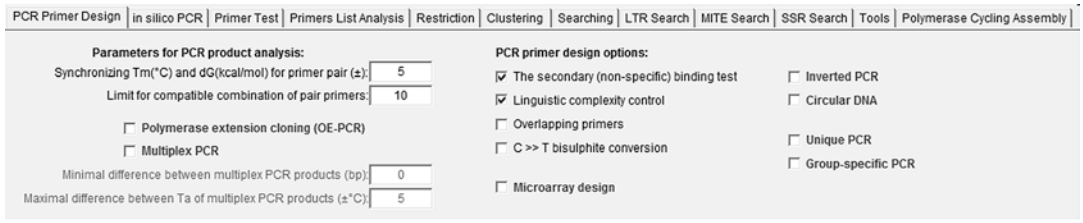

Fig. 3 The ribbon with the tasks. The program will only perform selected task

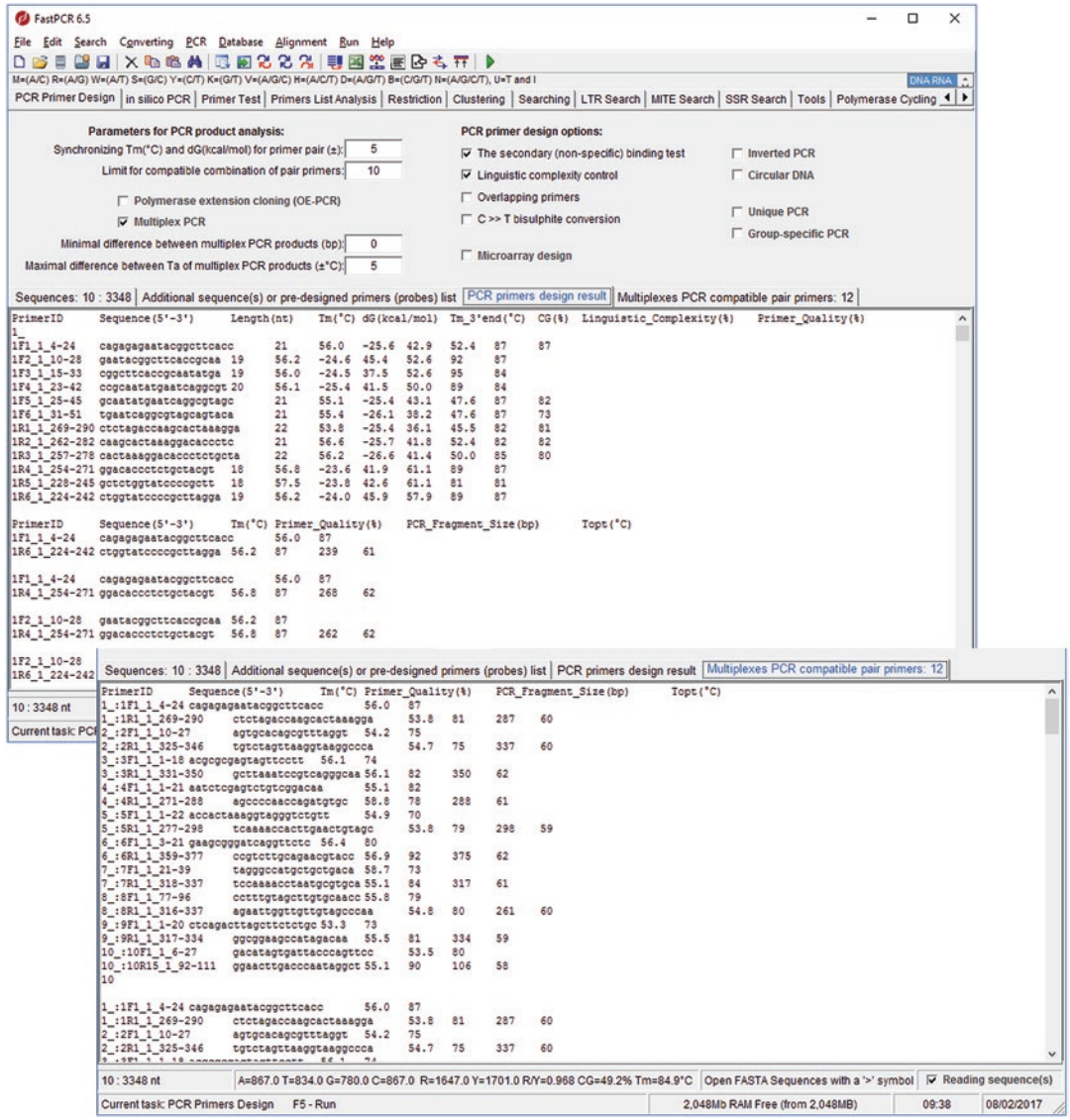

Fig. 4 An example of **Result report** text editor

/? and software replaces it with default global parameters for primer design:

{ -ln20-24 -tm51-60 -3tm30-50 -cg40-70 -q70 -lc75 -npr200 -c5[nn] -c3[swh ssw wsh] },

where

-ln20-24 determines the range of primer length (20–24 bases);

-tm51-60 determines the range of primer Tm (51–60 °C);

-3tm30-50 determines the range of primer Tm at 3′ end (30–50 °C);

-cg40-70 determines the range of primer CG% contents (between 40% and 70%);

-npr200 shows the limit for maximal primers amount designed to each target (200);

-c5[nn] denotes a primer having no specific sequence pattern for 5′ ends;

-c3[swh ssw **wsh]** makes specification for primers' 3′ ends with all these pattern with three bases per pattern.

Commands can be customized for each sequence, and global. Global commands are determined through "**PCR Options**" under menu "**PCR**." In addition, before all sequences in the text-editor (**General Sequence(s)**), as header can specify general parameters for all between **{...}**, as in the previous example, for help command: **/?** .

5.3 Examples for Primer Selection Regions

Users can specify, individually for each sequence, multiple locations for both forward and reverse primer design 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 or with the command **-pdN1-N2** (see more at: http://primerdigital.com/soft/pcr_help.html). Alternatively, users can specify the multiple locations for both forward and reverse primers design using **[** and **]** inside each sequence: the software allows multiple and independent locations of both forward and reverse primer design inside each of the sequences, whilst PCR design will be performed independently for different targets. Multiplex PCRs can be performed simultaneously within a single sequence with multiple amplicons as well as for different sequences, or combinations of both, i.e., all possible combinations of **[** and **]** inside the sequence(s). By default, software is designing primers within the entire sequence length.

Optionally, users can specify, individually for each sequence, multiple locations for both forward and reverse primer design with the commands:

1. The same location for both Forward and Reverse primers will be designed in the central **[nnnnnnnnn]** part ("**[ ]**" is Used only once):
   .....[nnnnnnnnnn].....

2. Different locations for Forward and Reverse primers; Forward primers will be chosen inside **[1nnnnnn]** location and Reverse primers inside **[2nnnnnn]** location ("**[ ]**" is used twice):
   .....[1nnnnnn]....[2nnnnnn].....
3. Designed primers must flank the central **]nnnnnn[**; Forward primers will be chosen from 1 to " **A]**" bases and Reverse primers will be chosen from "**[C**" base to the end of sequence:
   .....A]nnnnnn[C.....
4. Design primers with overlapping part **[nnnnnn]** for Forward and Reverse primers; 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 selecting any amount of independent PCR primers (probe) designing tasks for each sequences 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 or multiple tasks or both cases together.

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 user can specify the PCR product size in a similar way, with the command: (**N1**-**N2**); these values can be specified in the form of minimum and maximum value for the product size. For example, the (400–500) line defining the PCR product size ranges from 400 to 500 bases. In case a user wants to specify a fixed product size, the command should be a single number, for example (500). As default, the program allows PCR product sizes ranging between 30 and 10,000 bases.

5.5 PCR Set-Up Examples with Individual Commands

1. Prediction of optimal annealing temperature and PCR fragment(s) length for one or more predesigned primers (with -**npd** command, which prohibits the primer design). For a sequence to which these two primers (5′**ggagagtagcttacctcgct**, 5′**cggtaaggttcttcatgc**) have been designed:
   > **–fpr[ggagagtagcttacctcgct cggtaaggttcttcatgc] -npd**
2. Design of forward and reverse primers with a difference in T_m of about 10 °C for LATE-PCR:
   > **-Ftm50-55 -Rtm64-68 -pTMs10**

3. Design of primers with a specific restriction enzyme site at their 3′ end:

   -z3eNameEnzyme,**-Fz3eNameEnzyme**,**-Rz3eNameEnzyme**,

   where **NameEnzyme** is the name of the restriction enzyme: **-z3e**XceI.

   The alternative command (**-c3NN**) is used for special primer location. For example **-c3YCATGR** is the same as **-z3eXceI**, as the result: newly designed primers will locate on restriction site for *Xce I* endonucleases and 3′ end of primers will contain this sequence: **5′-(C/T)CATG(A/G)-3′**.

4. Addition of non-template DNA sequences to primer's ends. Add a sequence to 5′ end with command **-5eNN** or **-3eNN**, where **NN** is some sequence from one to more bases, for example:

   -F5eCGACG -R5eTTTTTT, adding sequence **CGACG** to Forward primers and sequence **TTTTTT** to Reverse primers at 5′ ends.

5. LAMP primer design examples:

   > -LAMP -LN17-24 -Tm52-62

   where **-LAMP** assay design (two nested primer pairs, four primers based on six nested targets) without using Loop Primers (http://loopamp.eiken.co.jp/e/lamp/loop.html), with primer lengths from 17 to 14 nt (**–LN17-24**), with the melting temperature from 52 to 62 °C (**–Tm52-62**). By default, the distance between 5′ end of F2 and B2 is considered 120–200 bp and the distance between F2 and F3 as well as B2 and B3 is 0–20 bp. The distance for loop forming regions (5′ of F2 to 3′ of F1, 5′ of B2 to 3′ of B1) is 0–40 bp. For LAMP assay design including Loop primer is necessary to apply the command:

   > -LAMP2 -LN17-24 -Tm52-62,

   where **-LAMP2**, will indicate the program for the development of LAMP assay design (two nested primer pairs, six primers based on eight nested targets) using Loop Primers (Fig. 5).

6. Design of tiling arrays: design both overlapping and nonoverlapping PCR primer pairs to generate either distinct or overlapping amplicons. Many unique tiling array platforms using both PCR products and short oligonucleotide probes have been created for a variety of applications including whole-genome arrays and specifically targeted arrays encompassing certain classes of genomic regions. FastPCR can automate the design of PCR primers for tiling arrays using both PCR products and short oligonucleotides, and avoiding repetitive regions. The user can specify the lengths of overlapping PCR primer pairs with the command: **-TILLING[N1N2]**; these values can be specified in the form of minimum and maximum value for the overlapping lengths. For example, the target DNA sequence using the **"Overlapping Amplicon"** option:

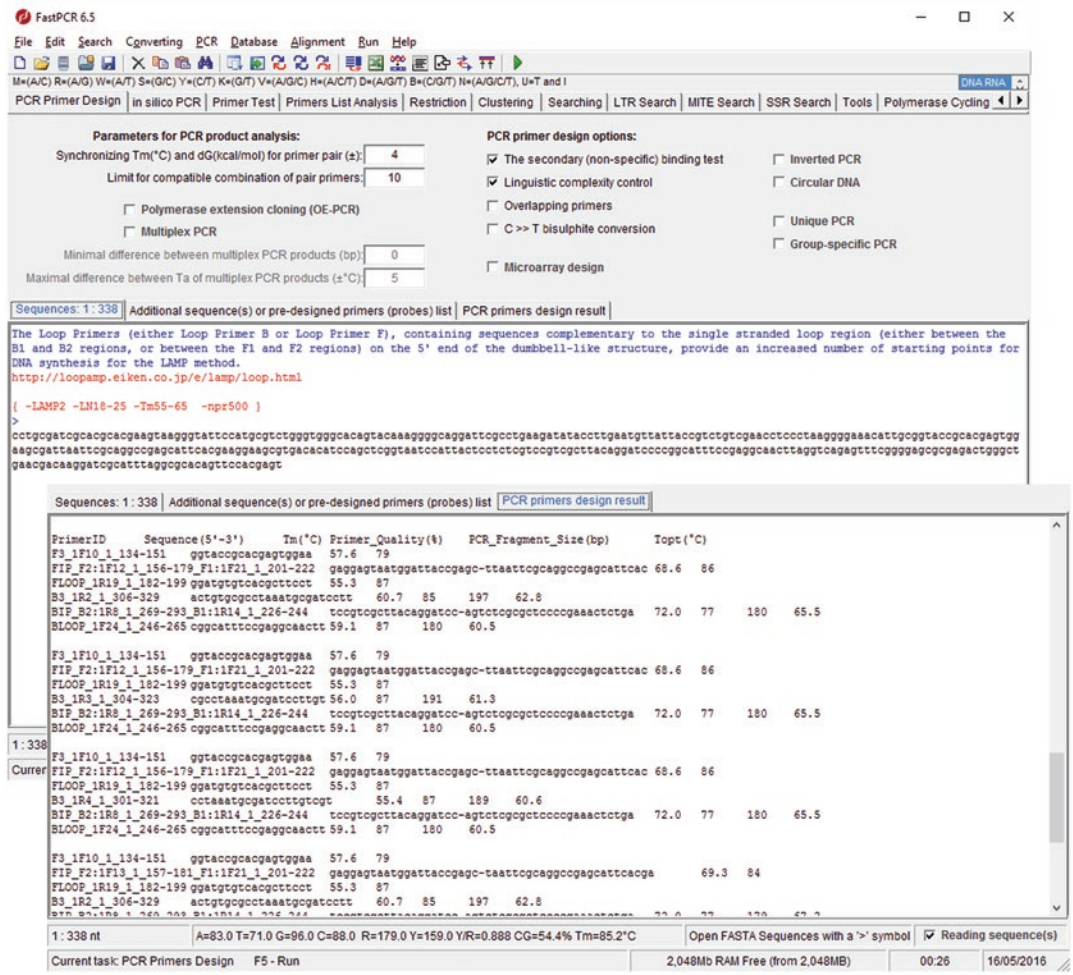

Fig. 5 An example of LAMP assay design using Loop Primers result

> **(110–250) -TILLING[-100-10]**,

where both **N1** and **N2** negative values allow overlapping between amplicons from 10 to 100 bases, command **(110–250)** defines the amplicons' size range from 110 to 250 bases.

The other option available for PCR primer pairs design is the **"Distinct Amplicon"** option. The amplicons generated could be as close as one base pair apart:

> **(100–250) -TILLING[0100]**,

where both **N1** and **N2** positive values do not allow overlapping amplicons and distance between amplicons ranges from 0 to 100 bases.

5.6 Bisulfite Modified DNA

The "**C >>T bisulphite conversion**" option allows the design of specific PCR primers for in silico bisulfite conversion for both

strands—only cytosines not followed by guanidine (CpG methylation) will be replaced by thymines:

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

5.7 PCR Primer Design

The PCR primer design algorithm generates a set of primers with a high likelihood of success in any amplification protocol. All PCR primers designed by FastPCR can be used for PCR, sequencing experiments or isothermal amplification.

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 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 an additional PCR product. 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 as a measure of the oligonucleotide quality and the thermodynamic stability of the 3′ and 5′ terminal bases are evaluated.

The proposal of primer pairs and the selection of the best pairs are both possible. The user can vary the product size or design primer pairs for the whole sequence without specifying parameters by using default or predesigned parameters. The predesigned parameters are specified for different situations: for example, for sequences with low CG content or long distance PCR, or degenerate sequences, or for manual input. A list of 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 primer design inside each sequence, whilst PCR design will be performed independently for different targets. Multiplex PCRs can be performed simultaneously within a single sequence with multiple amplicons as well as for different sequences, or combinations of both (Fig. 1).

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. Results are generated by the program showing the suggested primers and primer pairs in tabulated format for Excel or Open Office. The spreadsheets show the following properties: 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.

5.8 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 cross interactions. The multiplex PCR algorithm is based on the fast non-recursion method, with the software performing checks on product size compatibility (if necessary), the melting and annealing (T_a) temperatures, dG compatibility and cross-dimer interaction 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 the input sequences and/or multiplex targets within each sequence. PCR conditions may need to be adjusted; for example, the annealing temperature increased or lowered so that all products are amplified with equal efficiency. To achieve this, most existing multiplex primer design packages use primer melting temperature. In practical terms, the design of nearly identical T_a's and T_m's is important. The melting temperatures of the PCR products are also important, since these are related to annealing temperature values. The T_m of a PCR product directly depends on its GC content and length; short products are more efficiently amplified at low PCR annealing temperatures (100 bp, 50–55 °C) than long products (>3000 bp, 65–72 °C). For most multiplex PCRs, there is usually a small variation (up 5 °C) between the optimal T_a's of all primer pairs and PCR products. The annealing temperature must be optimal in order to maximize the likelihood of amplifying the target genomic sequences while 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_m's. Once prompted, FastPCR calculates multiplex PCR primer pairs for given target sequences. The speed of calculation depends on the number 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 also select input options for the PCR products such as the minimum product size difference between the amplicons. One can set primer design conditions either individually for each given sequence or use common values. The individual setting has a higher priority for PCR primer or probe design than do the general settings. The results include primers for individual sequences, primers compatible together, the 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 with information

including 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 PCR reactions by changing the filtering conditions as mentioned above. For example, a conventional multiplex PCR for gel-electrophoresis analysis 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 the need to avoid same-sized amplicons, 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 allows the selection of 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 also allows the user to design not only compatible pairs of primers, but also compatible single primers for different targets or sequences.

The input sequence can be made of either a single with 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 oligonucleotides designing parameters.

On the **PCR Primer Design** tab the user chooses the **Multiplex PCR** option and selects the limit for multiple PCR compatible combination of pair primers (default 100 primer pairs), minimal difference between multiplex PCR products (default 10 bp) and maximal difference between T_a/dG of multiplex PCR products (default ±5 °C). After specifying inputs and PCR primers design options, the user can execute the PCR primers design task. Once the primer set's design is complete, the result will appear in two **Result** text editors: **PCR primer design result** and **Multiplex PCR compatible pair primers**. Figure 4 shows the access to PCR primer design output. The result text editors **PCR primer design result** displays the individual PCR primers design data, including the primers list and the compatible primer pairs for all the sequences and their internal tasks where suitable primers are found. Second **Multiplex PCR compatible pair primers** text editors collects final search results that are presented as a list of the sets of the compatible primer pairs for multiplex PCR.

5.9 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 homological

genes or transposable elements 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 target sequences, following standard PCR design for the current sequence, and then testing complementarity of these primers to the other sequences. FastPCR performs either multiple sequence alignment or accepts alignment sequence input, giving it the flexibility to use a different strategy for primer design. Also, it can design both degenerate and nondegenerate PCR primers to amplify a conserved or polymorphic region of all related sequences.

The software 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 general options for primer design options. The FastPCR will work with any source of sequence as long as it is possible to find short (minimum 12 nt) consensus sequences among the sets. The quality of primer design is dependent on sequence relationships, phylogenetic 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.

On the **PCR Primer Design** tab the user chooses the **Group-specific PCR** option. After specifying inputs and PCR primers design options, the user can execute the PCR primers design task.

The program takes either multiple separate DNA sequences in either FASTA or alignment formats.

Once the primer set design is complete, the result will appear in the **Result** text editor: **PCR primer design result**. 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 an alignment has been input, the result text editors display only the group-specific PCR primer design set, including both degenerate and nondegenerate primers in the primer list, as well as compatible primer pairs for all the sequences.

5.10 Simple Sequence Repeat (SSR) Locus Search and PCR Primer Design

Simple sequence repeats (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 SSR searching is to analyse low complexity regions 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 flanks are used to design compatible forward and reverse primers for their amplification by PCR.

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 and 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 SSR, with command: **-ssr/200**. The software finds all SSR sites and then will design PCR primers and compatible primer pairs independently for each SSR loci.

5.11 Polymerase Cycling Assembly

The application to make long synthetic DNA molecules rely on the in vitro assembly of a set of short oligonucleotides, either by ligase chain reaction (LCR) [39] or by assembly PCR [40]. These oligonucleotides should be adjacent on the same strand and overlap the complementary oligonucleotides from the second strand. There are several major parameters to designing oligonucleotides for gene synthesis by LCR or assembly PCR: first, the oligonucleotides should share about similar T_m value; second, a given oligonucleotides sequence should be unique to avoid multiple nonspecific binding that conduct to incorrect assembly. The software must dynamically choose the length of the oligonucleotides to ensure both the specificity and the uniform T_m. Our algorithm 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 T_m of oligonucleotide has reached the T_m threshold. All oligonucleotides are designed without gaps between them. The other strand is used for design of the overlapping oligonucleotides with 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′ terminus in the oligonucleotide 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 the nonspecific target. Other complementary regions, apart from the 3′ terminus, are not important for assembling multiple fragments by PCR and ligation.

The input sequence 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. On the **Polymerase Cycling Assembly** tab, in the **Polymerase cycling assembly options**, the user can choose minimal (40–100 nt) and maximal (80–200 nt) oligonucleotide lengths, and minimal overlapping lengths (20–40 nt); by default the oligonucleotides' length ranges from 50 to 90 nt and 20 nt for overlapping length. The interface allows changing T_m calculation parameters (salt, Mg^{2+} and oligonucleotide concentration).

The searching process runs after pressing **F5** or from menu bar or from toolbox. The research result is presented as a list of oligonucleotides for both chains. On both strands, all oligonucleotides are overlapping with gap between neighbor oligonucleotides. Certain oligonucleotides will overlap two oligonucleotides from the complementary strand. The algorithm pays attention to avoid non-specific oligonucleotide hybridization to repeated regions. When it is not possible to design primers outside of repeated sequences, it will be 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. Figure 6 shows a sample result visualization window.

5.12 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) [41]. The efficiency depends on the T_m and on the length and uniqueness of the overlap. To achieve this, FastPCR 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 take special attention to the preparation of sequences for assembly.

The users can specify the locations for both forward and reverse primers design using "**[ ]**" to define each sequence ends. The defined regions will be used by the program for designing the overlapping primers.

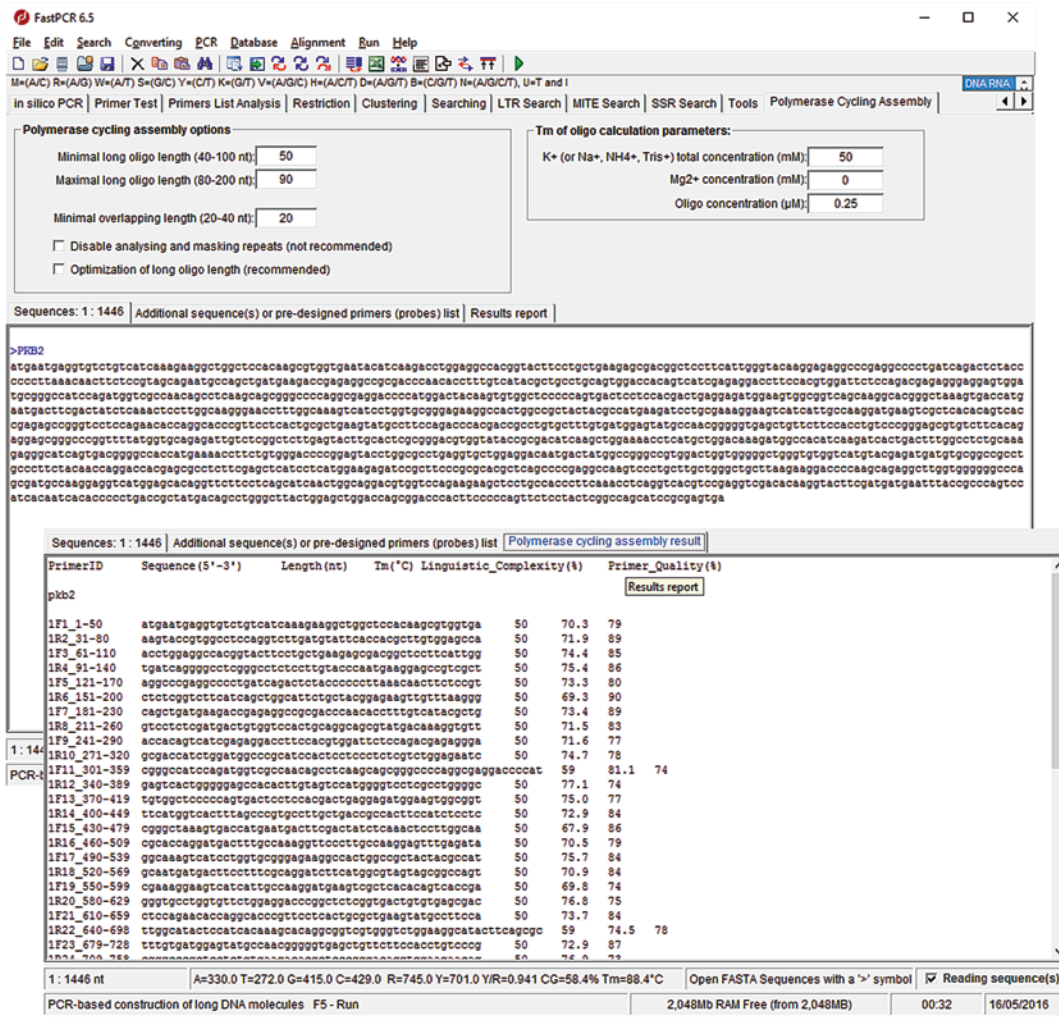

Fig. 6 An example of polymerase cycling assembly for fragment assembly result

The program selects the overlapping area so that the primers from overlapping fragments are similar in size and in their optimal annealing temperature. 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 dimers within the appropriate primer pair.

On the **PCR Primer Design** tab the user chooses the **Polymerase extension cloning (OE-PCR)** option and the limit for multiple PCR compatible combination of pair primers (default is 100). After specifying inputs and PCR primers design options, the user can execute the searching task. Once the primer set's design is complete, the result will appear in two **Result** text

editors: **PCR primer design result** and **PCR fragments assembling compatible pair primers**.

The result text editors **PCR primer design result** displays the individual PCR primer design data, including the primer list and compatible primer pairs for all sequences where suitable primers are found. The **PCR fragments assembling compatible pair primers** text editor collects the final search result, which is presented as a list of sets of the compatible primer pairs for individual fragments amplification and assembly.

5.13 Primer Analyses

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

For LNA modifications the four symbols: dA = E, dC = F, dG = J, dT = L are used. Both programs perform analyses on-type, which allow 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 dry oligonucleotide.

All primers are analyzed for intra- and inter-primer interactions to form dimers. Primer(s) can efficiently hybridize using the 5′ end or middle of the sequences. Even though such interactions are not efficiently extended by DNA polymerase, their formation reduces the effective primer concentration available for binding to the target and their presence can strongly inhibit PCR, since double-stranded DNA at high concentrations is a strong inhibitor of DNA polymerase (Fig. 7).

Example for a primer complementarity test including mismatches: when it is necessary to determine the annealing stability at the binding site and the T_m for primers with mismatches, the user may use the in silico PCR task. Another way to solve this problem is by analyzing the two primers in the task **Primer Test**, where one primer is analyzed, and the second complementary strand of it. To do so, it is necessary to convert the primer sequence into complementary and inverse chain and exclude the mismatches. The program will identify self-dimer and calculate the T_m including mismatches:

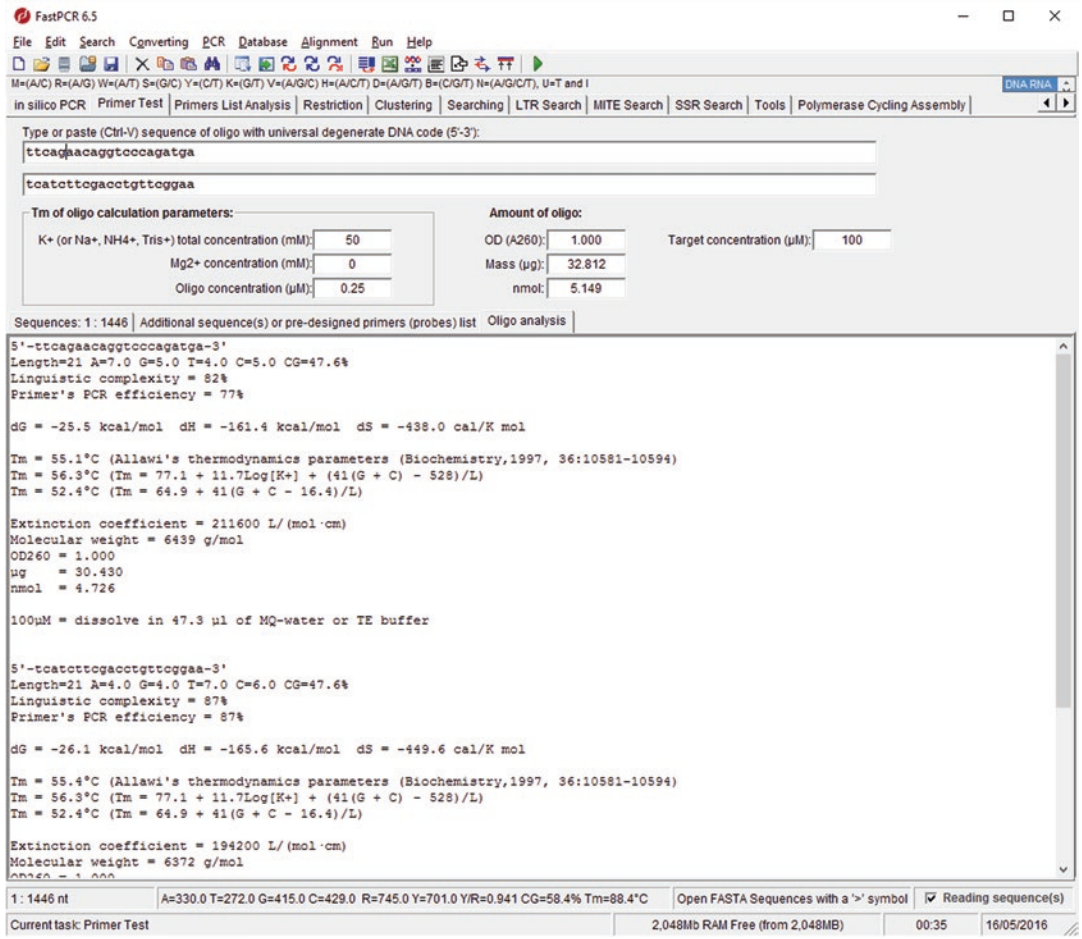

Fig. 7 Example result of the oligonucleotide analysis

```
Tm = 38.8 °C
5'-ttcagaacaggtcccagatga-3'
Length=21 A=7.0 G=5.0 T=4.0 C=5.0 CG=47.6%
Linguistic complexity = 82%
Primer's PCR efficiency = 77%
Tm = 55.1 °C

5'-tcatcttcgacctgttcggaa-3'
Length=21 A=4.0 G=4.0 T=7.0 C=6.0 CG=47.6%
Linguistic complexity = 87%
Primer's PCR efficiency = 87%
Tm = 55.4 °C

Dimers found between primers:

<-aaggcttgtccagcttctact-5
  |||:|||||||| ||||||
```

```
5-ttcagaacaggtcccagatga->
Tm=38.8 °C
```

Acknowledgments

Web tools are available free, provided for noncommercial research and education use only. They may not be reproduced or distributed for commercial use. This work was supported by the companies PrimerDigital Ltd.

References

1. Walker-Daniels J (2012) Current PCR methods. Mater Methods 2:119. doi:10.13070/mm.en.2.119
2. Tisi LC, Gandelman O, Kiddle G, Mcelgunn C (2010) Nucleic acid amplification. Canada Patent CA2417798
3. Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, Hase T (2000) Loop-mediated isothermal amplification of DNA. Nucleic Acids Res 28(12):e63. doi:10.1093/nar/28.12.e63
4. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386
5. Kalendar R, Lee D, Schulman AH (2014) FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis. In: Valla S, Lale R (eds) DNA cloning and assembly methods, Methods in molecular biology, vol 1116. Humana, New York, pp 271–302. doi:10.1007/978-1-62703-764-8_18
6. Kalendar R, Lee D, Schulman AH (2011) Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. Genomics 98(2):137–144. doi:10.1016/j.ygeno.2011.04.009
7. Marshall OJ (2004) PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics 20(15):2471–2472. doi:10.1093/bioinformatics/bth254
8. Owczarzy R, Tataurov AV, Wu Y, Manthey JA, McQuisten KA, Almabrazi HG, Pedersen KF, Lin Y, Garretson J, McEntaggart NO, Sailor CA, Dawson RB, Peek AS (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res 36:163–169. doi:10.1093/nar/gkn198
9. Bekaert M, Teeling EC (2008) UniPrime: a workflow-based platform for improved universal primer design. Nucleic Acids Res 36(10):e56. doi:10.1093/nar/gkn191
10. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (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
11. Smykal P, Kalendar R, Ford R, Macas J, Griga M (2009) Evolutionary conserved lineage of Angela-family retrotransposons as a genome-wide microsatellite repeat dispersal agent. Heredity (Edinb) 103(2):157–167. doi:10.1038/hdy.2009.45
12. 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
13. Gadberry MD, Malcomber ST, Doust AN, Kellogg EA (2005) Primaclade--a flexible tool to find conserved PCR primers across multiple species. Bioinformatics 21(7):1263–1264. doi:10.1093/bioinformatics/bti134
14. Nomenclature Committee of the International Union of Biochemistry (NC-IUB) (1984) Nomenclature for incompletely specified bases in nucleic acid sequences. http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html
15. Allawi HT, SantaLucia J Jr (1997) Thermodynamics and NMR of internal G.T mismatches in DNA. Biochemistry 36(34):10581–10594. doi:10.1021/bi962590c
16. SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A 95(4):1460–1465
17. Le Novere N (2001) MELTING, computing the melting temperature of nucleic acid duplex. Bioinformatics 17(12):1226–1227. doi:10.1093/bioinformatics/17.12.1226
18. Bolton ET, McCarthy BJ (1962) A general method for the isolation of RNA complementary to DNA. Proc Natl Acad Sci U S A 48(8):1390–1397
19. Guedin A, Gros J, Alberti P, Mergny JL (2010) How long is too long? Effects of loop size on G-quadruplex stability. Nucleic Acids Res 38(21):7858–7868. doi:10.1093/nar/gkq639
20. Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K (1979) Hybridization

of synthetic oligodeoxyribonucleotides to ΦX 174 DNA: the effect of single base pair mismatch. Nucleic Acids Res 6(11): 3543–3558. doi:10.1093/nar/6.11.3543

21. von Ahsen N, Wittwer CT, Schutz E (2001) Oligonucleotide melting temperatures under PCR conditions: nearest-neighbor corrections for Mg2+, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. Clin Chem 47(11):1956–1961
22. Gabrielian A, Bolshoy A (1999) Sequence complexity and DNA curvature. Comput Chem 23(3–4):263–274. doi:10.1016/S0097-8485(99)00007-8
23. Orlov YL, Potapov VN (2004) Complexity: an internet resource for analysis of DNA sequence complexity. Nucleic Acids Res 32(Web Server issue):628–633. doi:10.1093/nar/gkh466
24. Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 72(3):1047–1069. doi:10.1016/S0006-3495(97)78756-3
25. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal a.A, C.C, G.G, and T.T mismatches. Biochemistry 38(12):3468–3477. doi:10.1021/bi9825091
26. Watkins NE Jr, SantaLucia J Jr (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. Nucleic Acids Res 33(19):6258–6267. doi:10.1093/nar/gki918
27. Sen D, Gilbert W (1992) Guanine quartet structures. Methods Enzymol 211:191–199
28. Il'icheva IA, Florent'ev VL (1992) Four-stranded complexes of oligonucleotides--quadruplexes. Mol Biol 26(3):512–531
29. Shing Ho P (1994) The non-B-DNA structure of d(CA/TG)n does not differ from that of Z-DNA. Proc Natl Acad Sci U S A 91(20):9549–9553
30. Kypr J, Kejnovska I, Renciuk D, Vorlickova M (2009) Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res 37(6):1713–1725. doi:10.1093/nar/gkp026
31. SantaLucia J Jr, Hicks D (2004) The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct 33:415–440. doi:10.1146/annurev.biophys.32.110601.141800
32. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. Nucleic Acids Res 18(22):6531–6535. doi:10.1093/nar/18.22.6531
33. Welsh J, Mcclelland M (1990) Fingerprinting genomes using pcr with arbitrary primers. Nucleic Acids Res 18(24):7213–7218. doi:10.1093/nar/18.24.7213
34. Kalendar R, Schulman A (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nat Protoc 1(5):2478–2484. doi:10.1038/nprot.2006.377
35. 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(5):773–781. doi:10.1007/s001220051709
36. Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci U S A 86(17):6686–6690
37. Sinnett D, Deragon JM, Simard LR, Labuda D (1990) Alumorphs—human DNA polymorphisms detected by polymerase chain reaction using Alu-specific primers. Genomics 7(3):331–334
38. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110(1–4):462–467. doi:10.1159/000084979
39. Heckman KL, Pease LR (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc 2(4):924–932. doi:10.1038/nprot.2007.132
40. Higasa K, Hayashi K (2002) Ordered catenation of sequence-tagged sites and multiplexed SNP genotyping by sequencing. Nucleic Acids Res 30(3):E11
41. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4(7):e6441. doi:10.1371/journal.pone.0006441