A Guide to Using *FASTPCR* Software for PCR, In Silico PCR, and Oligonucleotide Analysis

Ruslan Kalendar ⬤

Abstract

The FastPCR software is an integrated tool environment for PCR primer and probe design and for prediction of oligonucleotide properties. The software provides comprehensive tools for designing primers for most PCR and perspective applications, including standard, multiplex, long-distance, inverse, real-time with TaqMan probe, Xtreme Chain Reaction (XCR), group-specific, overlap extension PCR for multi-fragment assembling cloning, and isothermal amplification (Loop-mediated Isothermal Amplification). A program is available to design specific oligonucleotide sets for long sequence assembly by ligase chain reaction and to design multiplexed of overlapping and nonoverlapping DNA amplicons that tile across a region(s) of interest for targeted next-generation sequencing, competitive allele-specific PCR (KASP)-based genotyping assay for single-nucleotide polymorphisms and insertions and deletions at specific loci, among other features. The in silico PCR primer or probe search includes comprehensive analyses of individual primers and primer pairs. FastPCR includes various bioinformatics tools for analysis and searching of sequences, restriction I–II–III-type enzyme endonuclease analysis, and pattern searching. The program also supports the assembly of a set of contiguous sequences, consensus sequence generation, and sequence similarity and conservancy analysis. FastPCR performs efficient and complete detection of various repeat types with visual display. FastPCR allows for sequence file batch processing that is essential for automation. The software is available for download at https://primerdigital.com/fastpcr.html and online version at https://primerdigital.com/tools/pcr.html.

Key words PCR primer and probe design, PCR design software, Nucleic acid amplification technologies, Loop-mediated isothermal amplification, In silico PCR primer or probe search, Bioinformatic genome analysis

1 Introduction

The polymerase chain reaction (PCR) is a fundamental technique in genetic engineering and is the most important practical molecular technique for the research laboratory. The principle of this technique has been further applied in several other simple or complex approaches, known as nucleic acid amplification technologies (NAAT) [1, 2]. The utility of PCR is dependent on identifying

specific primer sequences and designing PCR-efficient primer(s). Primer design is a critical step in all types of DNA amplification methods and is necessary to ensure specific and efficient amplification of a target sequence [3, 4]. Although there are currently many online tools and commercial bioinformatic software programs available, primer design for DNA amplification methods is still not as convenient and practical as it might be for routine use [5].

A variety of thermocycling and isothermal techniques currently exist for amplification of nucleic acids [6–8]. Thermocycling techniques, such as PCR, use temperature cycling to drive repeated cycles of DNA synthesis leading to large amounts of new DNA being synthesized in proportion to the original amount of template DNA. A number of isothermal techniques have also been developed that are not based on thermal cycling to control the amplification reaction. One such method is Loop-mediated Isothermal Amplification (LAMP) [6, 9], in which the template DNA is mixed with oligonucleotide primers and polymerase with high chain displacement activity. The mixture is kept at a constant temperature of 60–75 °C. Other methods of isothermal DNA amplification [10, 11] likewise depend on the chain displacement activity of specific DNA polymerases. Such methods include strand displacement amplification (SDA), helicase-dependent amplification (HDA), recombinase polymerase amplification (RPA), rolling circle amplification (RCA), transcription-mediated amplification (TMA), self-sustained sequence replication (3SR), and nucleic acid sequence-based amplification (NASBA) [2]. Each of these targeted amplification strategies requires the use of oligonucleotide primers. The amplification process results in an exponential amplification of amplicons that include oligonucleotide primers at their 5′ ends and contain newly synthesized copies of the sequences located between the primers. The length and composition of the primer will depend on many factors, including melting temperature (T_m) and source and composition of the primer. For example, depending on the complexity of the sequence, the primer usually contains 12–50 nucleotides, although it may also contain more or fewer nucleotides.

Targeted sequencing of genomic regions is a widespread practice in biomedical research. There are several methods for targeted sequencing, each of which is suitable for a specific task. Targeted enrichment based on PCR is a key step, as different targeting regions can be freely combined or excluded by the selection of PCR primers. The main difficulty in enrichment based on PCR is to ensure representative amplification of all targeted regions and uniform coverage of subsequent sequencing as a result. This requires careful primer design, which considers conditions such as uniformity of primer length, comparable thermodynamic

parameters, and similar melting temperatures. The combination of several primers together in the same reaction inevitably creates the possibility of undesirable interactions between primers, which negatively affect amplification itself and, therefore, the coverage of the sequence. Thus, an important step for optimization is the allocation of primers to various pre-amplification and multiplex PCR reactions to avoid unwanted primer interactions. The key to every oligonucleotide-based assay is the specific binding between the oligonucleotides and their target DNA. The specificity of oligonucleotides is of great importance in nucleic acid technologies such as DNA amplification and detection [5]. Therefore, accurate determination of the melting temperature of both complementary DNA duplexes and duplexes with mismatches is necessary to optimize the design of primers and probes in silico using machine learning based on an extensive experimental dataset of DNA/DNA duplexes. 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 multiplex PCR, microarray analysis, Insulated Isothermal PCR (iiPCR) (https://www.genereach.com/index.php?func=technology) [12], Xtreme Chain Reaction (XCR) (https://fluoresentric.com/principal-of-xcr/), NanoString multiplex analysis for the nCounter platform (https://www.nanostring.com/scientific-content/technology-overview/ncounter-technology), Next-Generation Targeted Sequencing (NGS) based on PCR amplification, and related that methods require the use of mixtures or multiple oligonucleotides (primers or probes) in one tube.

In developing FastPCR and online Java Web tools [13–16], our aim was to create a 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 to design combinations of primer pairs for optimal amplification.

2   Software, General Information

The FastPCR software (https://primerdigital.com/fastpcr.html) can be used with any version of Microsoft Windows. The online version of FastPCR software (https://primerdigital.com/tools/pcr.html) is written in Java with NetBeans IDE (Apache) and requires the Java Runtime Environment (the Java SE 8 Platform) (https://www.oracle.com/java/technologies/javase-downloads.html). The program can be used with any operating system (a 64-bit OS is preferred).

3 The Interface

3.1 Inputs to FastPCR

The software contains menus, toolbars, and a ribbon and three text editors. The ribbon is designed to help the user quickly find the commands required to complete a task. Commands are organized in logical groups, which are collected 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 text file, performing copy–paste, or simply by typing.

There are three independent text editors under different tabs: "General Sequence(s)," "Additional sequence(s) or predesigned 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. The "Additional sequence (s) or predesigned primers (probes) list" text editor can be applied to special and additional sequences, such as predesigned primers, multiple query sequences, or the numbers for input.
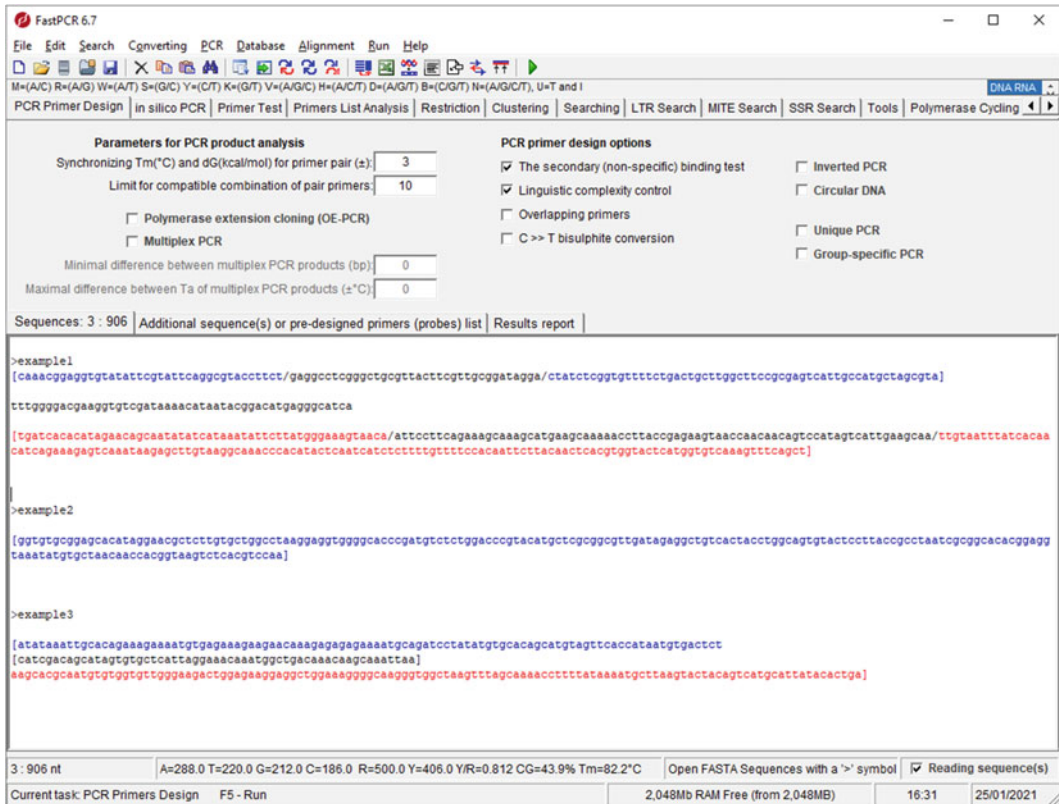

Fig. 1 FastPCR sequence editor and user interface

3.2 Program Output   FastPCR automatically generates results in the third text editor "Results report" in tabulated format. This allows transfer to a Microsoft Excel sheet via copy–paste. Alternatively, output results are easy to save as an .xls or .rtf text file, which is compatible with 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 the optimal annealing temperature (T_a) for each primer pair and the size of PCR product and complete information for each designed primer and for multiplex PCR product set.

3.3 Sequence Entry   The sequence data file is prepared using a text editor (Notepad++, WordPad, etc.) and saved in ASCII in a text/plain format or Rich Text Format (.rtf). The program takes either a single sequence or accepts multiple separate DNA sequences in FASTA (http://blast. ncbi.nlm.nih.gov/blast/fasta.shtml), tabulated format (two or three columns from an Excel sheet or table), EMBL (https:// www.ebi.ac.uk/services) [17], MEGA [18], GenBank and MSF (https://www.ncbi.nlm.nih.gov/books/NBK44863/), or simple alignment formats. For tabulated format, the software reads the first two columns, which includes a column for name and for sequence. For three and four columns, the software interprets them as primer pairs and probe. The template length is not limited.

The FastPCR clipboard allows the user to copy and paste operations of text or table from Microsoft Office documents or an Excel worksheet or other programs and paste them into another Office document. It is important that the entire target sequences are prepared with the same format.

The user can type or import from file(s) into the "General Sequence(s)" or "Additional sequence(s) or predesigned primers (probes) list" editors.

When a file is opened in FastPCR, users have several options for how the file opens. The user can open the original file as read-only for editing with text editors, or open to memory without opening to text editors, which allows the user to open larger file(s) up to 300 Mb. To work with files in a directory, the user points to any file in this directory or selects a list of files, or directs the program through a special menu to use all the current files in this directory. The program will open each file while executing the task without opening it to the text editor.

Additionally, the user can open all files from the selected folder and the program will join all files in the text editor. For example, this feature can be applied for converting all files from a selected folder to a single file of the list of FASTA sequences. Alternatively, this feature allows splitting FASTA sequences to individual files in the selected folder.

The program takes either a single sequence or accepts multiple separate DNA sequences in FASTA (http://blast.ncbi.nlm.nih.gov/blastcgihelp.shtml) or text file format (tab delimited). The FASTA format is preferred; it is simply the raw sequence preceded by a 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.

Degenerate DNA sequences are accepted as IUPAC code (https://www.qmul.ac.uk/sbcs/iubmb/misc/naseq.html), which is an extended vocabulary of 11 letters that 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 user can type or import data from file(s) into the "General Sequence(s)," "Additional sequence(s), or predesigned primers (probes) list" editors. In FastPCR software, users have several options for how to open a file when starting the program. The user can open the original file as read-only to work with text editors or open a file to memory without opening it to text editors, which allows the user to open a larger file(s).

For genome analysis, the user can open all files from a selected folder and the program will open each file while executing the task without opening it to a text editor. When a sequence file is open, the FastPCR software displays the information about the opened sequences and the sequence formats. The information status bar shows the number of sequences, total sequence length (in nucleotides), nucleotide composition, and the purine, pyrimidine, and GC content.

When saving a file from the current text editor, the user clicks the desired file format to save in, such as Rich Text Format (.rtf), Excel worksheet (.xls), or text/plain format (.txt).

4 The PCR Primers or Probe Design Analyze 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, controllable either by the user or automatically, are primer length (12–500 nt), melting temperature (T_m) for short primers (calculated by nearest neighbor thermodynamic parameters), the theoretical primer PCR efficiency (quality at %) value, primer GC content, 3′ end melting temperature, preferable 3′ terminal

Table 1
Default primer design selection criteria

Criteria	Default
length (nt)	20–23
T_m range (°C)[a]	53–55
T_m[a] 12 bases at 3′-end	25–40
3′-end composition (5′-nnn-3′)	swh, ssw, wsh, sww, www
Sequence linguistic complexity (LC, %)[b]	>75
Sequence quality (PQ, %)	>75

[a]Nearest neighbor thermodynamic parameters SantaLucia [20]
[b]Sequence linguistic complexity measurement was performed using the alphabet-capacity *l*-gram method

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 in %); specificity; the melting temperature of the entire primer and 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 sequence and task. Primer pairs are analyzed for cross-hybridization, specificity of both primers, and, optionally, for similar melting temperatures. Primers with balanced melting temperatures (within 1–3 °C of each other) and the thermodynamic free energy (*d*G) (within 1–3 kcal/mol of each other) are desirable but not mandatory. The default primer design selection criteria are shown in Table 1. 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.

We consider that the GC content parameter for primer evaluation is outdated and unnecessary. Basic thermodynamic parameters (the thermodynamic free energy and the melting temperature) of primer indirectly already include GC content [19]. Therefore, this parameter was unplugged from the user's control.

4.2 Melting Temperature Calculation

The melting temperature (T_m) is defined as the temperature at which half the DNA strands are in a 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 [20, 21]. The T_m is calculated using a formula

based on nearest neighbor thermodynamic theory with unified dS and dH parameters:

$$T_m(^{\circ}C) = \frac{d\text{H}}{d\text{S} + R \ln\left(\frac{c}{f}\right) + 0.368(L-1)\ln([K^+])} - 273.15$$

where dH is enthalpy for helix formation (cal/M), dS is entropy for helix formation (cal/K M), R is molar gas constant (1.987 cal/K M), and c is the nucleic acid molar concentration (250×10^{-9} M), $[K^+]$ is salt molar concentration (default value is 0.05 M), and $f = 4$ when the two strands are different and $f = 1$ when self-hybridization takes place.

The T_m for mixed bases is calculated by averaging the nearest neighbor thermodynamic parameters (enthalpy and entropy values) at each mixed site; the extinction coefficient is similarly predicted by averaging the nearest neighbor values at mixed sites. Mismatched pairs can be considered since the parameters provide for DNA/DNA duplexes and dangling ends (unmatched terminal nucleotides) [22]. The T_m for primer (probe) self or cross-dimers and for in silico PCR experiments with oligonucleotides with mismatches to the target is calculated using values for the thermodynamic parameters for a nucleic acid duplex.

4.3 Primer Quality (Virtual PCR Efficiency) Determination

The primer nucleotide composition and T_m of the 12 bases at the 3′-terminal position of the primers are important factors for PCR efficiency. The composition of the sequence at the 3′ terminus is important; primers with two terminal C/G bases are recommended for increased PCR efficiency [23]. 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 polymerization efficiency. More than 3 G's or C's should be avoided in the last six bases at the 3′ end of the primer. The last 12 bases of the 3′ end of the hybridizing part of the primer should preferably not exceed 42 °C; the last 12 bases of the 5′ end of the hybridizing part of the primer should preferably have at least 42 °C. Polynucleotide stretches should be avoided; runs of three or more guanine residues in particular can cause problems due to intermolecular stacking. Primers with T_m in the range of 55–60 °C generally produce the best results.

We specify an abstract parameter called Primer Quality (PQ) that can help estimate the efficiency of primers for PCR. PQ is calculated as the lower value of the following parameters: total sequence complexity and the T_m of the whole primer and of the terminal 3′ and 5′. Self-complementarity, which gives rise to potentially stable dimer and hairpin structures, and long runs of a single base reduce the final value. PQ tries to describe the likelihood of PCR success of each primer; this value varies from 100% (best primer) to 0% (worst primer).

To meet multiplexing demands, it is possible in the program to select the best primer with an optimal temperature range, which allows for the design of qualified primers or probes for any target sequence with any GC and repeat content. PQ values ≥80 allow for rapid choice of the best PCR primer pair combination. No adverse effects due to modification of the reaction buffer, chosen thermostable polymerases, or variations in annealing temperature have been observed in the reproducibility of PCR amplification using primers with high PQ.

4.4 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 is avoided. In particular, complementarity at the 3′ ends of primers from where the polymerase will extend should be avoided. Stable primer dimer formation is very effective at inhibiting PCR, since the dimers formed are amplified efficiently and will compete with the intended target.

Primer dimer prediction is based on analysis of nongap 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 based on nucleotide composition and the presence of at least 5 G/C bases at the 3′ end. Tools calculate the 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 [20, 21, 24]. In addition to Watson–Crick base pairing, there are a variety of other hydrogen bonding configurations possible that FastPCR can detect, such as G-quadruplexes or G-T base pairs. Mismatch stability is the following (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, cytosine is the most discriminating base, since it forms the strongest pair and the three weakest mismatches [25]. Therefore, the tools also assess stable guanine mismatches (G-G, G-T, and G-A). Guanine-rich nucleic acid sequences can fold into four-stranded DNA structures that contain stacks of G quartets. 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 G. FastPCR predicts the presence of putative G-quadruplexes in primer sequences. Intermolecular guanine-quadruplex-forming sequences are detected according to the $d(G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+})$, where N is any nucleotide base (including guanine) [22, 26]. The gap sequences (N_{1-7}) may have varying lengths, and a relatively stable quadruplex structure may still be

formed with a loop seven bases long. However, increasing the length of the gap generally leads to a decrease in structure stability. It is also possible for one of the gaps to be zero length when there are long polyguanine tracts of >6 bases.

4.5   Secondary Nonspecific Binding Test; Alternative Amplification

Oligonucleotide specificity is one of the most critical factors for PCR efficiency. 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 interspersed repetitive sequences or to homologous sequences [27, 28]. Alternative and unexpected product amplification can also occur when primers are complementary to or to homologous sequences. This is unlikely when primers have been designed using specific DNA sequences. However, the generation of interspersed inverted repetitive sequences is exploited in the Random Amplified Polymorphic DNA (RAPD) [29, 30] or the Inter-Simple Sequence Repeat (ISSR) methods [31, 32]. Similar PCR-based DNA fingerprinting techniques such as Inter-Retrotransposon Amplification Polymorphism (IRAP), Retrotransposon-Microsatellite Amplification Polymorphisms (REMAP) [33, 34], Inter-MITE (IMP) [35] or Inter-SINE amplification polymorphism (ISAP) [36], and Inter-Primer Binding Site (iPBS) amplification polymorphism [37] have exploited these highly abundant interspersed repeats as the target [38–42]. However, primers complementary to interspersed repetitive sequences may produce many nonspecific bands in a single-primer amplification and compromise performance of targeted PCRs [42, 43]. A homology search of the primer sequence, for example, using 'blastn' against all sequences in GenBank, will determine whether the primer is likely to interact with interspersed repeats.

By default, 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 against a database of choice. Primers that bind to more than one location on current sequences will be rejected. Although the nonspecific primer binding test is performed as a default for all primers, the user may fully control this function [42].

5   Methods

FastPCR provides various execution features once the input files are selected or the sequence is copy-pasted to the **General Sequence(s)** text editor. Figure 1 shows a design from a user's perspective for primer design.
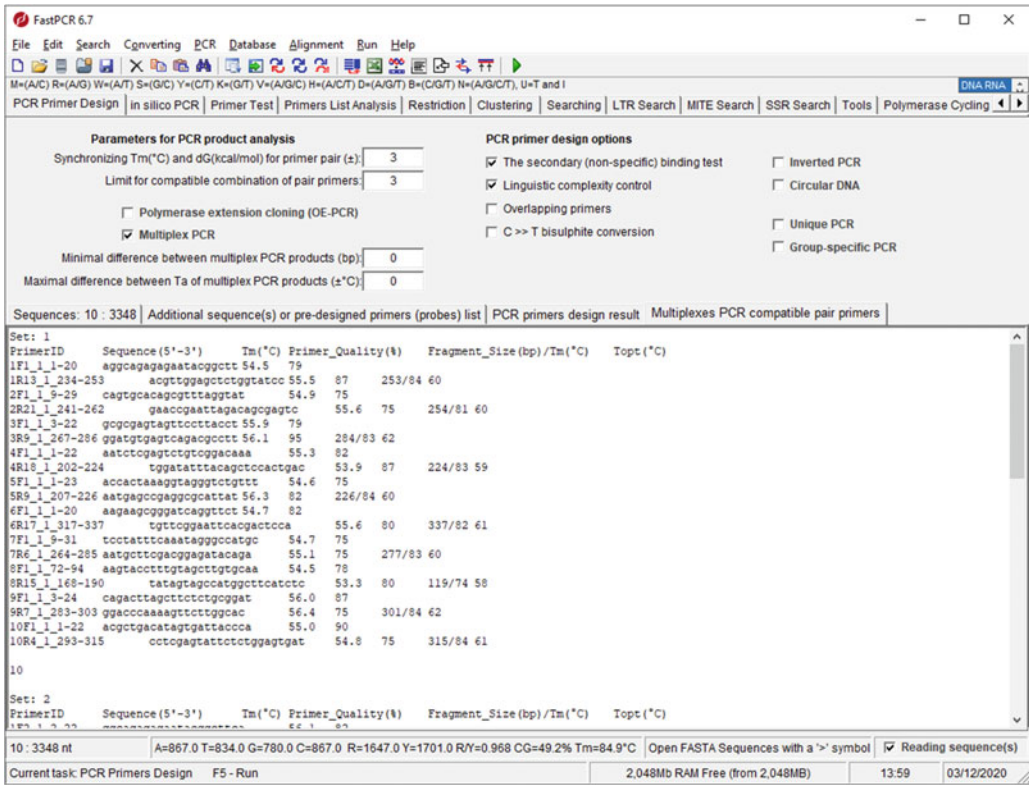
Fig. 2 Result of the multiplex set for the mPCR design is shown at the Result report

5.1 Execution Selected Task

The user selects the ribbon with a task required. The program will only perform the selected task. Depending on which task the user selects, pressing **F5** will execute the current task (Fig. 2). The user either presses **F5** or clicks on the toolbar to execute the current task.

Once the executive task is complete, the result is shown in the **Result report** text editor. Figure 2 shows a sample result visualization window.

5.2 PCR Primer Design Options

To perform PCR analysis, select PCR -> PCR options from the menu bar. The PCR Primer or Probe Design Options dialog box appears (Fig. 3). The PCR Primer or Probe Design Options dialog box allows the user to designate basic parameters for the PCR reaction and the primers that are generated. The "**PCR Primer or Probe Design Options**" dialog contains various execution options for selection of PCR type and the most important PCR parameters. Figure 3 shows "**PCR Primers or Probe Design Options**" on the option panel. 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.

Fig. 3 "PCR Primers or Probes Design Options" window

For individual PCR primer design options for each sequence, the user can type special commands at the header of the sequence. Typically, the user does not need to use commands to manage PCR primer design; these commands are used optionally and only for advanced tasks. The user can type in the text editor this help command: '/?' and software replaces it with default global parameters for primer design:

```
// -ln21-25 -tm54-56 -3tm25-40 -q75 -lc75 -npr5000 -c5[nn] -c3
[swh ssw wsh sww www] -dmr = 1 (20-5000) //,
```

where '-ln21-25' determines the range of primer length (21–25 bases);'-tm54-56' determines the range of primer T_m (54–56 °C);'-3tm25-40' determines the range of primer T_m at 3'end (25–40 °C);'-npr5000' shows the limit for maximal primer amount designed for each target (5000);'-dmr=1' sensitivity for primer dimer detection, default value is 1 (as the strictest criterion), the higher the value, the lower the detection sensitivity, if the value is 0, the program will not assess for the presence of primer dimers, the optimal value of this parameter is 2, as the most realistic; (20–5000) boundaries of the size of PCR amplicons, the exact size is allowed, in this case, the size of the PCR amplicon will be in the range of 20–5000 base pairs;'-c5[nn]' primer has no specific sequence pattern for 5′ ends;'-c3[swh ssw wsh sww www]' specifies for primers 3'ends with these patterns with three bases per pattern.

Adding non-template DNA sequences to primer ends: adding sequence to the 5′ end with command: '-5e[NN]' or adding sequence to the 3′ end with command: '-3e[NN]', where 'NN' is a sequence of one or more bases, for example: '-F5e[CGACG]-R5e[TTTTTT]', adds sequence 'CGACG' to forward primers and sequence 'TTTTTT' to reverse primers at 5′ ends. For tasks in which it is necessary to specify several allelic variants with individual added tail sequences, which can be indicated with a separator, the forward slash character "/" is used to denote each tail variant. For example, for biallelic scoring of single-nucleotide polymorphisms (SNPs) and insertions with deletions (Indels) at specific loci in relation to the development of competitive allele-specific PCR (KASP) genotyping assays (https://www.biosearchtech.com/support/education/kasp-genotyping-reagents), two (up to four) variants of the added tail sequences like that:

```
-p5e[GAAGGTGACCAAGTTCATGCT/GAAGGTCGGAGTCAACGGATT].
```

5.3 Examples for Primer Selection Region

The user can specify individually for each sequence multiple locations for both forward and reverse primer designs using '[' and ']' inside each sequence. The software allows multiple and independent locations of both forward and reverse primer designs inside each of the sequences, while PCR design will be performed independently for different targets. Multiplex PCRs can be performed simultaneously within a single sequence with multiple amplicons, for different sequences, or combinations of both, such as all possible combinations of '(' and ')' inside the sequence(s). By default, the software designs primers within the entire sequence length.

The excluded region list(s) denotes locations where primers and probe must not bind. Multiple excluded regions may be defined per sequence. This feature can be used to avoid unwanted regions (such as introns and SNPs). For example, if the excluded region includes bases 500–1000, for the target selection the following command should be used: -exclude500-1000. An alternative is to use two '/' signs for the start and end of the excluded region (this is possible multiple times):

```
>example
[gtcccgagaacctgagtatgcatcacccggatcgcttcttcc/gggaggtgttgggggg/
ctatctcggtgtttttctgactgcttggcttccgcgagtcattgccatgctagcgta]
[attgcaataaccggagcgagatgatgcacc/ccccc/ccttgacaagcgccaatac-
cacgcactattaagagtaaaaaaaa]
```

5.4 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, isothermal amplification, or sequencing experiments. The program can 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 cycler sequences. It will check if either primers or probes have secondary binding sites in the input sequences that may yield 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 as a measure of the oligonucleotide quality, and the thermodynamic stability of the $3'$ and $5'$ terminal bases is evaluated.

Primer pairs are suggested, and selection of the best pairs is 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. These include sequences with low GC content, long-distance PCR, XCR, TaqMan probes design, microarray or degenerated sequences, or manual input. For thermodynamic compatibility of primers, two options can be specified by the user. Default is primer pair compatibility based on similar melting temperatures (Primer design with T_m priority) for both primers, while second option is based on similar Gibbs free energy (Primer design with dG priority). Designing primers with the same dG will render more efficient primers pairs, and matching T_ms is a less accurate approach than matching dGs.

Results show the sorted list of best primer candidates and all compatible primer pairs that are optimal for PCR. The program generates primer pairs (and probes, or both primers with probe) 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 and show 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, GC content, molecular weight, molar extinction coefficient, linguistic complexity, and PQ. For compatible primer pairs, the annealing temperature and PCR product size with the T_m of the PCR product are also provided.

5.5 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 required; multiplex PCR thus increases throughput efficiency. The design of

multiplex PCR assays is based on the nonrecursion fast method, with the software performing checks on product size compatibility, thermodynamic 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 (thermodynamic compatibility) to their targets. FastPCR can quickly design several sets of multiplex PCR primers for all the input sequences, multiplex targets, or both within each sequence. In practical terms, designing primers with almost identical Gibbs free energy and T_m and with optimal annealing temperatures (T_a) is preferably. The T_m of the PCR products is also important as 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 (<1000 bp, 50–55 °C) than long products (>3000 bp, 60–72 °C). The optimal annealing temperature for PCR is calculated directly as the value for the primer with the lowest T_m (T_m^{min}) and taking into account the length of the PCR fragment (L): $T_a(°C) = T_m^{min} + \ln L$ [14].

For most multiplex PCRs, there is usually a small variation (up 3–5 °C) between the optimal annealing temperatures of all primer pairs and PCR products. The annealing temperature must be optimal 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 Gibbs free energy.

An alternative way to design compatible multiplex PCR primer pairs is to use predesigned primers as a compatible reference list for the design of new primers. The user can also select input options for the PCR products, such as the minimum product size differences between the amplicons.

The user 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, product sizes, and annealing temperatures. As 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 the alternative compatible primer pair combinations that provide the desired product sizes. The user can select predesigned primer pairs from a target for their desired type of PCR reaction 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 the need to avoid amplicons of identical 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 allows the selection of primer pairs that give the most likelihood of producing only the amplicons of the target sequences by choosing sequences that avoid repeats or other motifs. The input sequence can be generated by either a single with a minimum of two internal tasks or many sequences with or without internal tasks. Most parameters on the interface are self-explanatory. Optionally, the user is asked to provide the sequence and select the parameters for oligonucleotide design.

On the **PCR Primer Design** tab, the user chooses **Multiplex PCR** and chooses the limit for multiple PCR-compatible combinations of primer pair sets (default three sets), minimal difference between multiplex PCR products (default is 0, so ignoring this parameters), and maximal difference between T_a of multiplex PCR products (default is 0). After specifying inputs and PCR primer design options, the user can execute the PCR primers design task. Once the primers set design is complete, the result will appear in two **Result** text editors: **PCR primer design result** and **Multiplexes PCR-compatible pair primers**. Figure 2 shows the access to PCR primer design output. For the results text **PCR primer design result**, the window displays the individual PCR primers design data, including primer list and compatible primer pairs for all sequences and their internal tasks whose primers are found. The second **Multiplexes PCR-compatible pair primers** window collects the final search result that is presented as a list of the sets of the compatible primer pairs for multiplex PCR.

5.6 In Silico PCR Predicting the hybridization of primers and probe to targeted annealing sites is the only way to evaluate PCR products [4]. The last 10–12 bases at the 3′ end of primer are critical for binding stability; single mismatches can significantly reduce PCR efficiency, and the effect increases with proximity to the 3′ end. FastPCR allows simultaneous testing of a single primer or a set of primers designed for multiplex target sequences or genome. For in silico PCR, a quick gapless alignment for detection of primer locations on the target sequence is performed by analyses on both strands using a hash index of 7-, 9-, or 12-mers (containing up to one mismatch within) and by calculating the local similarity and the T_m for the primer sequence. The parameters can be set 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 1 mismatch) at the 3′ end of the primer. The program can also handle degenerate primer or probe sequences, including those with 5′ or 3′ tail sequences.

Probable PCR products of linear and circular templates can be found using standard or inverse PCR and multiplex PCR or bisulfite-treated DNA sequence. This in silico tool is useful for quickly analyzing primers or probes against target sequences, for determining primer location, orientation, and binding efficiency and for calculating T_m and T_a in PCR.

The user must input the preexisting primer's list into a second **Additional sequence(s) or pre-designed primers (probes) list** text editor. The amount of preexisting primers is not limited on the primer pair; it can be as much as the user requires. The target sequences either multiple separate DNA sequences or opening files from the entered folder. The user must specify a directory for input for in silico PCR against the whole genome(s) or a list of chromosomes. The program performs a consistent, file-by-file check of each file to the DNA sequence position of the primers. In the **In silico PCR** tab, the user can execute the search task with **F5** or, alternatively, can specify search options such as stringency and PCR product detection options. For the stringency options, the user can specify the number of mismatches that are allowed at the primer 3′ end. Once the primer sets design is complete, the result will appear in the following two **Result** text editors: **In silico PCR Result**. The results in the **In silico PCR Result** text editor reports the primer specificity (locations, including target position, similarity, and T_m) and summary of primer pairs in relation to the PCR template, and detailed information on each primer pair, its length, and T_a. Target-specific primers are shown if found, and the actual targets are listed with detailed alignments between primers and targets (Fig. 4).

5.7 Primer(s) Analyses

Individual primers and primer sets can be evaluated using FastPCR software. The software calculates the primer T_m using default or other formulae for normal and degenerate nucleotide combinations, GC content, extinction coefficient, unit conversion (nmol per OD), mass (μg per OD), molecular weight, linguistic complexity, and primer PCR efficiency. Users can select either DNA or RNA primers with normal or degenerate oligonucleotides or modifications with different labels (e.g., inosine, uridine, or fluorescent dyes). The tools allow the choice of either nearest neighbor thermodynamic parameters or nonthermodynamic T_m calculation formulae. For locked nucleic acid (LNA) modifications, the four symbols dA = E, dC = F, dG = J, and dT = L are used. Both programs perform analyses on-type, which allow the user to view the results immediately on screen. The program can also calculate the volume of solvent required to attain a specific concentration from known mass (mg), OD, or moles of dry oligonucleotide. All
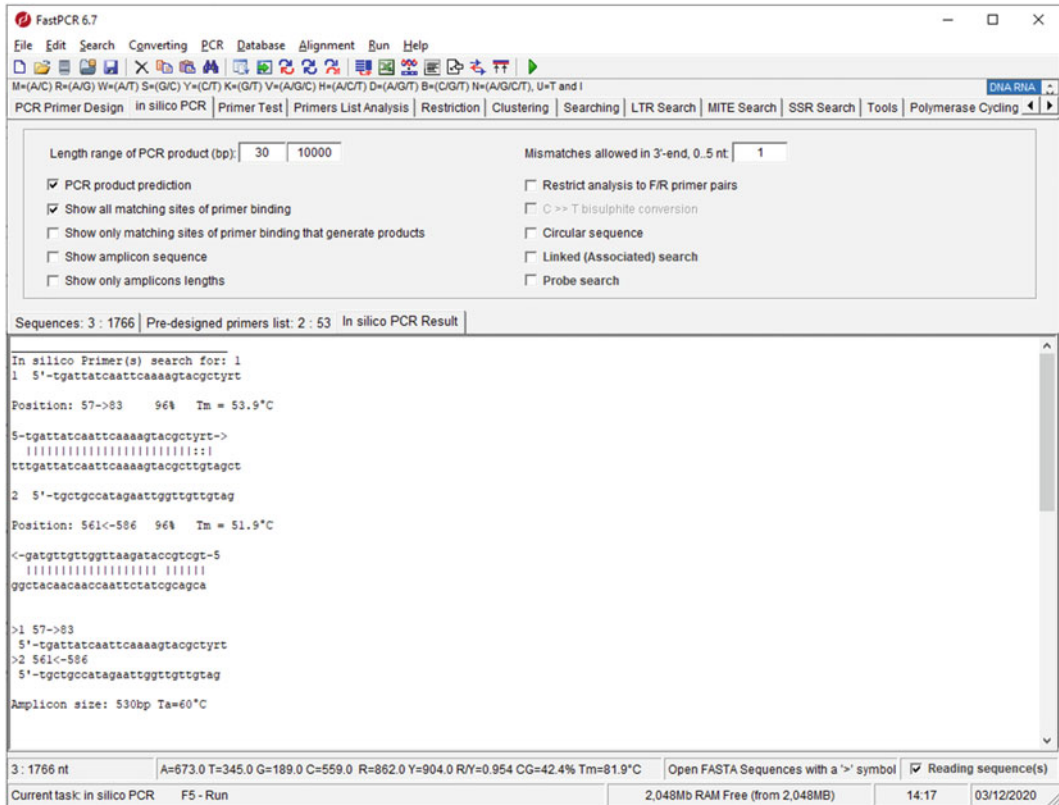
Fig. 4 In silico PCR the detailed result is shown

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.

6 Availability

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

The online version of the FastPCR software requires the Java Runtime Environment (the Java SE 8 Platform) (https://www.oracle.com/java/technologies/javase-downloads.html)or use OpenWebStart software (https://openwebstart.com/download/), an open-source reimplementation of the Java Web Start technology for latest Java SE 11 (LTS) or Java SE 16. Users should add the URL (https://primerdigital.com/) of this application to the

Exception Site List (https://www.java.com/en/download/help/exception_sitelist.html), which is located under the Security tab of the Java Control Panel (https://www.java.com/en/download/help/appsecuritydialogs.html). Users can download self-signed certificate files (https://primerdigital.com/j/primerdigital.cer) and import it to "Signer CA" (Certificate Authority) from the Java Control Panel. Finally, users should set "Security Level" to "High" under the Security tab of the Java Control Panel (as shown here: https://primerdigital.com/image/primerdigital_certificate_big.png).

Acknowledgments

References

1. Yasukawa K, Yanagihara I, Fujiwara S (2020) Alteration of enzymes and their application to nucleic acid amplification (Review). Int J Mol Med 46(5):1633–1643. https://doi.org/10.3892/ijmm.2020.4726

2. Gill P, Ghaemi A (2008) Nucleic acid isothermal amplification technologies: a review. Nucleosides Nucleotides Nucleic Acids 27 (3):224–243. https://doi.org/10.1080/15257770701845204

3. Bekaert M, Teeling EC (2008) UniPrime: a workflow-based platform for improved universal primer design. Nucleic Acids Res 36(10): e56. https://doi.org/10.1093/nar/gkn191

4. 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. https://doi.org/10.1186/1471-2105-13-134

5. Guo J, Starr D, Guo H, Wren J (2020) Classification and review of free PCR primer design software. Bioinformatics 36 (22-23):5263–5268. https://doi.org/10.1093/bioinformatics/btaa910

6. Shirato K (2019) Detecting amplicons of loop-mediated isothermal amplification. Microbiol Immunol 63(10):407–412. https://doi.org/10.1111/1348-0421.12734

7. Mayboroda O, Katakis I, O'Sullivan CK (2018) Multiplexed isothermal nucleic acid amplification. Anal Biochem 545:20–30. https://doi.org/10.1016/j.ab.2018.01.005

8. Kim J, Easley CJ (2011) Isothermal DNA amplification in bioanalysis: strategies and applications. Bioanalysis 3(2):227–239. https://doi.org/10.4155/bio.10.172

9. Tomita N, Mori Y, Kanda H, Notomi T (2008) Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. Nat Protoc 3 (5):877–882. https://doi.org/10.1038/nprot.2008.57

10. James A, Macdonald J (2015) Recombinase polymerase amplification: emergence as a critical molecular technology for rapid, low-resource diagnostics. Expert Rev Mol Diagn 15(11):1475–1489. https://doi.org/10.1586/14737159.2015.1090877

11. Qian J, Boswell SA, Chidley C, Lu Z-x, Pettit ME, Gaudio BL, Fajnzylber JM, Ingram RT, Ward RH, Li JZ, Springer M (2020) An enhanced isothermal amplification assay for viral detection. Nat Commun 11(1):5920. https://doi.org/10.1038/s41467-020-19258-y

12. Qiu J, Tsai Y-L, Wang H-TT, Chang H-FG, Tsai C-F, Lin C-K, Teng P-H, Su C, Jeng C-C,

Lee P-Y (2012) Development of TaqMan probe-based insulated isothermal PCR (iiPCR) for sensitive and specific on-site pathogen detection. PLoS One 7(9). https://doi.org/10.1371/journal.pone.0045278

13. Kalendar R, Khassenov B, Ramanculov E, Samuilova O, Ivanov KI (2017) FastPCR: an in silico tool for fast primer and probe design and advanced sequence analysis. Genomics 109 (3-4):312–319. https://doi.org/10.1016/j.ygeno.2017.05.005

14. 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. https://doi.org/10.1016/j.ygeno.2011.04.009

15. Kalendar R, Muterko A, Shamekova M, Zhambakin K (2017) In silico PCR tools for a fast primer, probe, and advanced searching. Methods Mol Biol 1620:1–31. https://doi.org/10.1007/978-1-4939-7060-5_1

16. Kalendar R, Lee D, Schulman AH (2014) FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis. Methods Mol Biol 1116:271–302. https://doi.org/10.1007/978-1-62703-764-8_18

17. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47(W1): W636–W641. https://doi.org/10.1093/nar/gkz268

18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547–1549. https://doi.org/10.1093/molbev/msy096

19. Benita Y, Oosting RS, Lok MC, Wise MJ, Humphery-Smith I (2003) Regionalized GC content of template DNA as a predictor of PCR success. Nucleic Acids Res 31(16):e99. https://doi.org/10.1093/nar/gng101

20. SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA 95(4):1460–1465. https://doi.org/10.1073/pnas.95.4.1460

21. Allawi HT, SantaLucia J Jr (1997) Thermodynamics and NMR of internal G.T mismatches in DNA. Biochemistry 36(34):10581–10594. https://doi.org/10.1021/bi962590c

22. 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. https://doi.org/10.1093/nar/gkq639

23. 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. https://doi.org/10.1016/S0006-3495(97)78756-3

24. Watkins NE Jr, SantaLucia J Jr (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. Nucleic Acids Res 33 (19):6258–6267. https://doi.org/10.1093/nar/gki918

25. SantaLucia J Jr, Hicks D (2004) The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct 33:415–440. https://doi.org/10.1146/annurev.biophys.32.110601.141800

26. Todd AK, Johnston M, Neidle S (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res 33 (9):2901–2907. https://doi.org/10.1093/nar/gki553

27. Jurka J (1998) Repeats in genomic DNA: mining and meaning. Curr Opin Struct Biol 8 (3):333–337. https://doi.org/10.1016/s0959-440x(98)80067-5

28. Kalendar R, Raskina O, Belyayev A, Schulman AH (2020) Long tandem arrays of cassandra retroelements and their role in genome dynamics in plants. Int J Mol Sci 21(8):2931. https://doi.org/10.3390/ijms21082931

29. Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18(24):7213–7218. https://doi.org/10.1093/nar/18.24.7213

30. 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. https://doi.org/10.1093/nar/18.22.6531

31. Sivolap IM, Kalendar RN, Chebotar SV (1994) The genetic polymorphism of cereals demonstrated by PCR with random primers. Cytol Genet 28(6):54–61. https://pubmed.ncbi.nlm.nih.gov/7701604/

32. Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20(2):176–183. https://doi.org/10.1006/geno.1994.1151

33. Kalendar R, Schulman A (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nat Protoc 1 (5):2478–2484. https://doi.org/10.1038/nprot.2006.377

34. Kalendar R, Grob T, Regina M, Suoniemi A, Schulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA

fingerprinting techniques. Theor Appl Genet 98(5):704–711. https://doi.org/10.1007/s001220051124

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. https://doi.org/10.1007/s001220051709

36. Seibt KM, Wenke T, Wollrab C, Junghans H, Muders K, Dehmer KJ, Diekmann K, Schmidt T (2012) Development and application of SINE-based markers for genotyping of potato varieties. Theor Appl Genet 125(1):185–196. https://doi.org/10.1007/s00122-012-1825-7

37. Kalendar R, Antonius K, Smykal P, Schulman AH (2010) iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. Theor Appl Genet 121(8):1419–1430. https://doi.org/10.1007/s00122-010-1398-2

38. Kalendar R, Amenov A, Daniyarov A (2019) Use of retrotransposon-derived genetic markers to analyse genomic variability in plants. Funct Plant Biol 46(1):15–29. https://doi.org/10.1071/fp18098

39. Kalendar R, Muterko A, Boronnikova S (2021) Retrotransposable elements: DNA fingerprinting and the assessment of genetic diversity. Methods Mol Biol 2222:263–286. https://doi.org/10.1007/978-1-0716-0997-2_15

40. Kalendar R, Schulman AH (2014) Transposon-based tagging: IRAP, REMAP, and iPBS. Methods Mol Biol 1115:233–255. https://doi.org/10.1007/978-1-62703-767-9_12

41. Hosid E, Brodsky L, Kalendar R, Raskina O, Belyayev A (2012) Diversity of long terminal repeat retrotransposon genome distribution in natural populations of the wild diploid wheat Aegilops speltoides. Genetics 190(1):263–412. https://doi.org/10.1534/genetics.111.134643

42. Kalendar R, Kospanova D, Schulman A (2021) Transposon-based tagging in silico using FastPCR software. Methods Mol Biol 2250:245–256. https://doi.org/10.1007/978-1-0716-1134-0_23

43. Kalendar R, Shustov AV, Seppänen MM, Schulman AH, Stoddard FL (2019) Palindromic sequence-targeted (PST) PCR: a rapid and efficient method for high-throughput gene characterization and genome walking. Sci Rep 9(1):17707. https://doi.org/10.1038/s41598-019-54168-0