



FastPCR: An *in silico* tool for fast primer and probe design and advanced sequence analysis



Ruslan Kalendar^{a,*}, Bekbolat Khassenov^b, Yerlan Ramankulov^b, Olga Samuilova^c, Konstantin I. Ivanov^d

^a PrimerDigital Ltd, FIN-00710 Helsinki, Finland

^b RSE “National Center for Biotechnology” under the Science Committee, Ministry of Education and Science of the Republic of Kazakhstan, Astana, Kazakhstan

^c Department of Biochemistry, I.M. Sechenov First Moscow State Medical University, 8-2 Trubetskaya st., Moscow 119991, Russian Federation

^d Department of Biophysics, Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow 119234, Russian Federation

ARTICLE INFO

Article history:

Received 8 October 2016

Received in revised form 9 May 2017

Accepted 11 May 2017

Available online 12 May 2017

Keywords:

Polymerase chain reaction

Isothermal amplification of nucleic acids

DNA primers

Nucleic acid hybridization

Primer binding site

PCR primer and probe analysis

Degenerate PCR

Probe

Genetic engineering tools

DNA fingerprinting

ABSTRACT

Polymerase chain reaction (PCR) is one of the most important laboratory techniques used in molecular biology, genetics and molecular diagnostics. The success of a PCR-based method largely depends on the correct nucleic acid sequence analysis *in silico* prior to a wet-bench experiment. Here, we report the development of an online Java-based software for virtual PCR on linear or circular DNA templates and multiple primer or probe search from large or small databases. Primer or probe sensitivity and specificity are predicted by searching a database to find sequences with an optimal number of mismatches, similarity and stability. The software determines primer location, orientation, efficiency of binding and calculates primer melting temperatures for standard and degenerate oligonucleotides. The software is suitable for batch file processing, which is essential for automation when working with large amounts of data. The online Java software is available for download at <http://primerdigital.com/tools/pcr.html>.

Accession numbers for the sequences resulting from this study:

EU140956

EU177767

EU867815

EU882730

FJ975775–FJ975780

HM481419

HM481420

KC686837–KC686839

KM262797

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Polymerase chain reaction (PCR) is fundamental to molecular biology and is the most important laboratory technique used for nucleic acid amplification. PCR employs temperature cycling to synthesize large amounts of progeny DNA from small amounts of original template DNA and involves the following repeated temperature-dependent steps: (1) heat denaturation of double-stranded DNA to single-stranded DNA, (2) annealing of primers to the single-stranded DNA and (3) primer elongation by a thermostable DNA polymerase [1,2].

A number of isothermal techniques have also been developed that do not rely on thermocycling to drive the amplification reaction. One

such technique is LAMP (Loop-Mediated Isothermal Amplification) [3], in which template DNA is mixed with oligonucleotide primers and a polymerase with high strand displacement activity, and the mixture is held at a constant temperature of 60–65 °C. Other isothermal DNA amplification techniques similarly depend on the strand displacement activity of particular DNA polymerases. Examples of such techniques include Strand Displacement Amplification (SDA) [4], Chimeric Displacement Reaction (CDR), Rolling Circle Amplification (RCR) [5], Isothermal Chimeric Amplification of Nucleic Acids, SMart Amplification Process (SMAP) [6], Transcription-based amplification system (TAS) [7], Self-sustained sequence replication reaction (3SR) [8], Helicase-dependent amplification (HDA) [9], Single primer isothermal amplification (SPIA) [10] and cross-priming amplification (CPA) [11–13,2,14,15].

The success of a PCR reaction heavily depends on correct design of oligonucleotide primers. The forward primer anneals to the first DNA strand, while the reverse primer anneals to the second, complementary

* Corresponding author.

E-mail addresses: ruslan.kalendar@primerdigital.com (R. Kalendar), khassenov@biocenter.kz (B. Khassenov), ramankulov@biocenter.kz (Y. Ramankulov).

strand. If DNA sites to which the primers anneal are separated by an appropriate distance, the DNA fragment between these sites, known as PCR product or “amplicon”, can be copied by DNA polymerase, approximately doubling in abundance with each temperature cycle. Because stable hybridization of a primer with the template is an absolute prerequisite for primer extension by DNA polymerase, correct selection and validation of primers is critical to the success of a PCR reaction.

Primer specificity is one of the most important factors for a successful PCR. Well-designed primers should anneal only to the target DNA sequences, particularly when complex genomic DNA is used as a template. Ideally, all DNA sequences from the target set would be exactly complementary to the primers and would be amplified, while no DNA sequence from the background set would match the primers. In other words, primers have to be very specific to amplify only the target DNA sequence. Furthermore, primer melting temperatures need to be very similar to allow proper annealing to the target DNA sequence at the same temperature. Problems may arise when primers are complementary to repetitive sequences (retrotransposons, DNA transposons or tandem repeats) [16] and alternative amplification products may be generated when primers are complementary to inverted repeats. However, generation of inverted repeats is exploited in generic DNA fingerprinting methods. Often only one primer is used in these PCR reactions and the product ends contain inverted repeats complementary to the primer sequence.

For eukaryotes, the inter-repeat amplification polymorphism techniques such as inter-retrotransposon amplified polymorphism (IRAP) have exploited the presence of highly abundant dispersed repeats such as the long terminal repeats (LTRs) of retrotransposons and SINE-like sequences [17,18]. The association of these sequences with each other makes it possible to generate a series of PCR products (DNA fingerprints) using primers homologous to the repeats. One, two or more primers pointing outwards from an LTR are used to amplify DNA sequences between the two nearby retrotransposons. The primers may be homologous to sequences from the same retrotransposon family or from different families. The PCR products and the corresponding fingerprint patterns result from amplification of hundreds to thousands of target sequences in the genome [19,20].

The development of DNA sequencing technology and in particular the emergence of high-throughput sequencing (next generation sequencing) has led to progressive accumulation of huge amounts of raw genome sequence data. Currently, many prokaryotic and eukaryotic genomes have been sequenced and annotated in databases. For this reason, there is a growing demand for *in silico* approaches to extract useful information from raw sequence data, process the information using virtual tools and predict experimental results already at the planning stage. One such approach is *in silico* PCR. The usual goal of *in silico* PCR is to predict which PCR products are synthesized from one or multiple DNA templates using already designed oligonucleotide primers. Other common goals include template sequence location in the genome as well as primer/probe design and analysis, such as prediction of melting temperatures and secondary/tertiary structures including hairpins, G-quadruplexes, self-dimers and cross-dimers in primer pairs [21,22].

Currently, several web-based *in silico* PCR tools are available for molecular biologists [23–26,21,22,27–32]. One of them, PRIMEX, uses a k-mer lookup table to search for short sequence matches in whole genomes [31]. The web server “Electronic PCR” allows heuristic searches of predefined genomes with up to two mismatches (<https://www.ncbi.nlm.nih.gov/tools/epcr/>) [29]. Another web server, UCSC *in silico* PCR (<http://genome.ucsc.edu/cgi-bin/hgPcr>), uses an undocumented algorithm to search a predefined genome [24]. Finally, Primer-BLAST [33] is a web server that uses BLAST [26] as the underlying search method.

Most of the published *in silico* PCR algorithms are heuristic and few of them are available as stand-alone software packages. The notable exceptions are PRIMEX [31] and FastPCR for Windows [34,21,35]. Furthermore, the application of commonly used sequence similarity search

algorithms to *in silico* PCR has not been entirely successful. For example, BLAST creates local alignments which may fail to cover full primer sequences and it does not support searching for pairs of queries separated by arbitrary sequence of variable length. Degenerate primers cannot appear in alignment seeds and, therefore, require special handling to avoid false negatives. The parameters sensitive enough to find an acceptable number of matches may also generate a large number of false positives and extensive post-processing is required to identify valid hits. Because Primer-BLAST uses a word length of 7, sites must have an identical 7-mer match to a primer. Consider a primer of length 15 with differences in the 5th and 10th positions at a given site. The longest identical k-mer at that site would be 5 nt and BLAST would fail to find the match.

A good *in silico* PCR program must be able to handle degenerate primers or probes including those with 5′ or 3′ tail sequences and single nucleotide polymorphisms (SNPs). In addition, the program should be compatible with bisulfite-treated DNA containing only methylated cytosine and highly degraded and chemically modified DNA from ancient herbarium and mummies. Finally, the program must handle multiplex, nested or tiling PCR commonly used to amplify several DNA targets in a single reaction. These considerations motivated us to develop both stand-alone and online software that can handle all of the above tasks. Our aim was to create efficient and easy-to-use computational tool for advanced sequence analysis, *in silico* PCR on linear or circular DNA templates and multiple primer or probe search from large or small databases (Table 1). The software should be useful for quick selection of primers or probes from target sequences, for determining primer location, orientation, binding efficiency and calculating primer melting temperatures for standard and degenerate oligonucleotides. The software should also allow to validate existing primers, probes and their combinations as well as predict PCR products for linear and circular templates for standard or inverse PCR as well as multiplex PCR.

2. Methods

2.1. The search algorithm

The main purpose of the algorithm is to efficiently search for complementary sequences in the template DNA with less or equal than the allowed number of mismatches. These sequences have to be within a certain distance from each other, which corresponds to the maximal size of the expected PCR product.

Stable hybridization of a primer to a template DNA is essential for primer extension by DNA polymerase. Mismatches affect the stability of the primer-template duplex and the efficiency with which the polymerase extends the primer. Although any mismatch affects PCR specificity, mismatches at the 3′-end of a primer have a profound detrimental effect on primer extension. A two-base mismatch at the

Table 1
Summary of FastPCR software features for virtual PCR.

- Finds all possible primer pairs for conventional or multiplex PCR taking into account mismatches located within the specified primer or target sequence;
- Predicts primer location and PCR product sequences from chromosome lists, whole genomes or circular DNA;
- Performs advanced searching for two or more sequences linked to each other and located within a certain distance;
- Is compatible with *in silico* PCR fingerprinting based on repeats (SINE-PCR, IRAP, REMAP, ISSR or RAPD);
- Recognizes degenerate bases in primer and target sequences;
- Allows the addition of non-complementary sequences to the 5′-termini of primers or to the 3′- and 5′- termini of probes;
- Allows the use of query sequences as short as 4 nt;
- Calculates annealing and melting temperatures for oligonucleotide-target duplexes with standard and degenerate oligonucleotides, allowing the user to view stable guanine mismatches;
- Presents the results in the form of a sequence alignment which includes location and similarity data, predicted PCR product lengths, primer sequences and annealing temperatures.

3'-end of the primer results in PCR failure. Therefore, our algorithm pays particular attention to the 3'-end of the primer and calculates its similarity to the template with the user-defined level of stringency. A few mismatches can be tolerated, typically at the expense of reduced amplification efficiency. Therefore, we used data from actual PCR experiments to fine-tune how the algorithm handles the presence of mismatches.

The algorithm consists of three parts. The first one creates a hash table for all overlapping k-tuples for a primer set (query sequences). The second one performs sequence analysis using the above hash table. The third part predicts potential PCR products for linear or circular templates.

The hash-table for all overlapping k-tuples (k-mers, words of fixed length k) is stored in memory as a map structure consisting of a list of k-tuples and indexes for a helping array. The helping array creates a link to a certain primer and k-tuple coordinate, thus allowing the presence of similar, identical or repeated nucleotide sequences in the primer set. The k-tuple length can be 7, 9 or 12 nt depending on search sensitivity, type of task and primer length. The use of a long k-tuple of 12 nt results in only minor loss of sensitivity and is unlikely to generate false negatives because the algorithm allows up to one mismatch in each k-tuple. The ability of our algorithm to tolerate up to one mismatch in a k-tuple, together with degenerate nucleotide compatibility and detection of stable guanine mismatches (i.e. G•G, G•T and G•A), makes the algorithm different from other published approaches including BLAST [36,37]. Furthermore, our algorithm allows the use of short primers (≥ 4 nt) that can be shorter than the minimum k-tuple length of 7 nt.

The algorithm searches for primer target sites using a sliding window that moves along the DNA sequence at one-nucleotide intervals. The algorithm does not create a hash table for the template sequence; instead, it continuously tries to find matches using the primer hash table. One mismatch is tolerated and long sequence gaps are ignored. Individual k-tuples are extended in both directions to find regions of similarity until their length becomes equal to that of the primer or until a critical amount of mismatches is reached. Both primer and target sequences may contain degenerate nucleotides and the user may specify the number of mismatches near the 3'-end of the primer, with a default setting of one mismatch in the last seven nucleotides. The adjustable 3'-mismatch threshold makes the algorithm compatible with oligonucleotide hybridization probes (molecular beacons) lacking complementarity to the target at their 3' ends. The algorithm also detects stable guanine mismatches: G•G, G•T and G•A [38], which represent alternatives to Watson-Crick base-pairing.

The user can define the desired size of the PCR product. By default, the distance between forward and reverse primers ranges from 50 to 5000 bases. PCR product size prediction is possible for linear or circular templates using standard, inverse or multiplex PCR, or when using bisulphite-treated DNA as a template. The algorithm accepts single or multiple DNA sequences as primers or amplification targets.

Primer melting temperatures (T_m) for *in silico* PCR experiments with oligonucleotides that have mismatches to their targets are calculated using the averaged nearest neighbor thermodynamic parameters [39–42]. The optimal PCR annealing temperature (T_a) is calculated as the value for the primer with the lowest T_m plus natural logarithm of the PCR product length [22].

The algorithm implemented in FastPCR saves memory by creating a hash table only for primer sequences and is computationally efficient, approaching linear time in the database size. The memory and execution time requirements of FastPCR were determined on a computer with the Intel® Core™ i7-4700HQ (2.4 Ghz) processor using genomic DNA sequences of different sizes and primer sequences with universal or degenerate bases. Because the entire DNA sequence was loaded into RAM, the amount of memory required for analysis was directly proportional to the sequence length. In absolute terms, the RAM requirement was about 4 GB for the analysis of the complete human genome. Software execution time was proportional to the number of identified amplicons and the target genome size, ranging between 10 s for the

0.1 Gbp genome of *Arabidopsis thaliana* to about 3 min for the 3.1 Gbp human genome (see Table 2).

The algorithm was validated by estimating the number of false positives and false negatives in a complex computational model. For this purpose, we used the EMBL-EBI Eukaryotes database of genome sequences (<http://www.ebi.ac.uk/genomes/eukaryota.html>) as a target set and retrotransposon primer and probe sequences as query sets. The use of subsets of retrotransposon sequences allowed us to assess how the software handles complex tasks such as *in silico* degenerate PCR and linked searching. The results of software validation are presented in Supplemental data together with query sequences.

2.2. The Interface

The software has a user-friendly interface. It contains menus, toolbars, a ribbon and three text editor tabs. Getting started with a basic project is easy; the user can open and save files, type, copy or paste text. Detailed description of the FastPCR user interface is available at the website (<http://primerdigital.com/tools/pcr.html>) [21,34,35].

3. Results

3.1. *In silico* PCR analysis

To demonstrate the capabilities and potential of our software, we performed *in silico* PCR analysis of several plant genomes using a degenerate primer pair corresponding to a highly conserved peptide sequence of plant Copia-type reverse transcriptase (RT) [43]:

```
>RT+(QMDVK)
5'-CARATGGAYGTNAARAC
>RT-(YVDDML)
5'-CATRTCRCNACRTA
```

Results of the Copia-type RT analysis are represented in Table 2. As expected, no amplicons were predicted for the human genome due to the lack of sequence similarity between plant and animal reverse transcriptase genes. The Copia-type RT analysis provides an example of a complex computational task, which can be successfully handled by very few software packages such as PRIMEX and the one described in this study. Less complex tasks, such as conventional *in silico* PCR without degenerate primers, are easier to perform and their examples are available for independent study by the user under the File menu. The software is compatible with both single-primer methods based on sequence repeats such as SINE-PCR, IRAP, REMAP, ISSR, RAPD, etc. and methods utilizing multiple primers such as LAMP, multiplex PCR, etc. It is also compatible with retroelement-based genotyping and DNA fingerprinting methods. Primer and probe lengths can be set between 12 and 500 nucleotides and the maximal PCR product (amplicon) length is limited only by the polymerase processivity and not by the software. The results of the analysis include PCR product sequences, their lengths and T_a , and are presented separately for nuclear DNA, plastids and mitochondrial DNA. Another application of the software is detection of plasmid sequences based on the ends of an inserted sequence. For this purpose, sequence fragments flanking an insert should be used instead of primer sequences and the “probe search” option should be selected in the software.

3.2. Linked searching

In silico PCR is an example of sequence similarity searching, in which primer sequences are located at a certain distance from each other and are oriented towards each other. A more general method called linked (associated, programmed) searching allows advanced searching of primer-template binding sites in a variety of scenarios, including that of *in silico* PCR. In linked searching, the search criteria are based only on the distance between annealing sites without taking into account primer orientation and query sequences can be as short as 4 nt. The

Table 2

Software execution times and total number of hits returned by the search of several eukaryotic genomes using a degenerate primer pair corresponding to a highly-conserved peptide sequence of the plant Copia-type reverse transcriptase (RT).

Genome	Size in Mb, (total number of files)	Time in seconds	Number of 270–281 bp amplicons, which were found using search criteria with the 400 bp limit.
<i>Arabidopsis thaliana</i>	126 (9)	10	1
<i>Brachypodium distachyon</i>	275 (6)	17	5
<i>Oryza sativa</i>	380 (12)	25	13
<i>Sorghum bicolor</i>	545 (10)	35	10
<i>Glycine max</i>	965 (20)	60	18
<i>Solanum pennellii</i>	1006 (14)	70	14
<i>Zea mays</i>	2100 (18)	135	22
<i>Homo sapiens</i>	3140 (24)	195	0

idea behind linked searching is similar to that behind AGREP (<https://github.com/Wikinaut/agrep>), which stands for approximate GREP for fast fuzzy string searching. AGREP works similarly to a traditional UNIX GREP utility by searching input files for occurrences of a particular string. Linked searching, however, can perform a broader variety of tasks ranging from conventional sequence matching to *in silico* PCR

and general tasks involving DNA sequence analysis with approximate matching.

To give a better idea of how linked searching works in practice, we will use the above example of *in silico* PCR with two degenerate Copia-type RT primers. Let us convert the primer sequences into a single line, where forward primer sequence (5'-CARATGGAYGTNAARAC) is

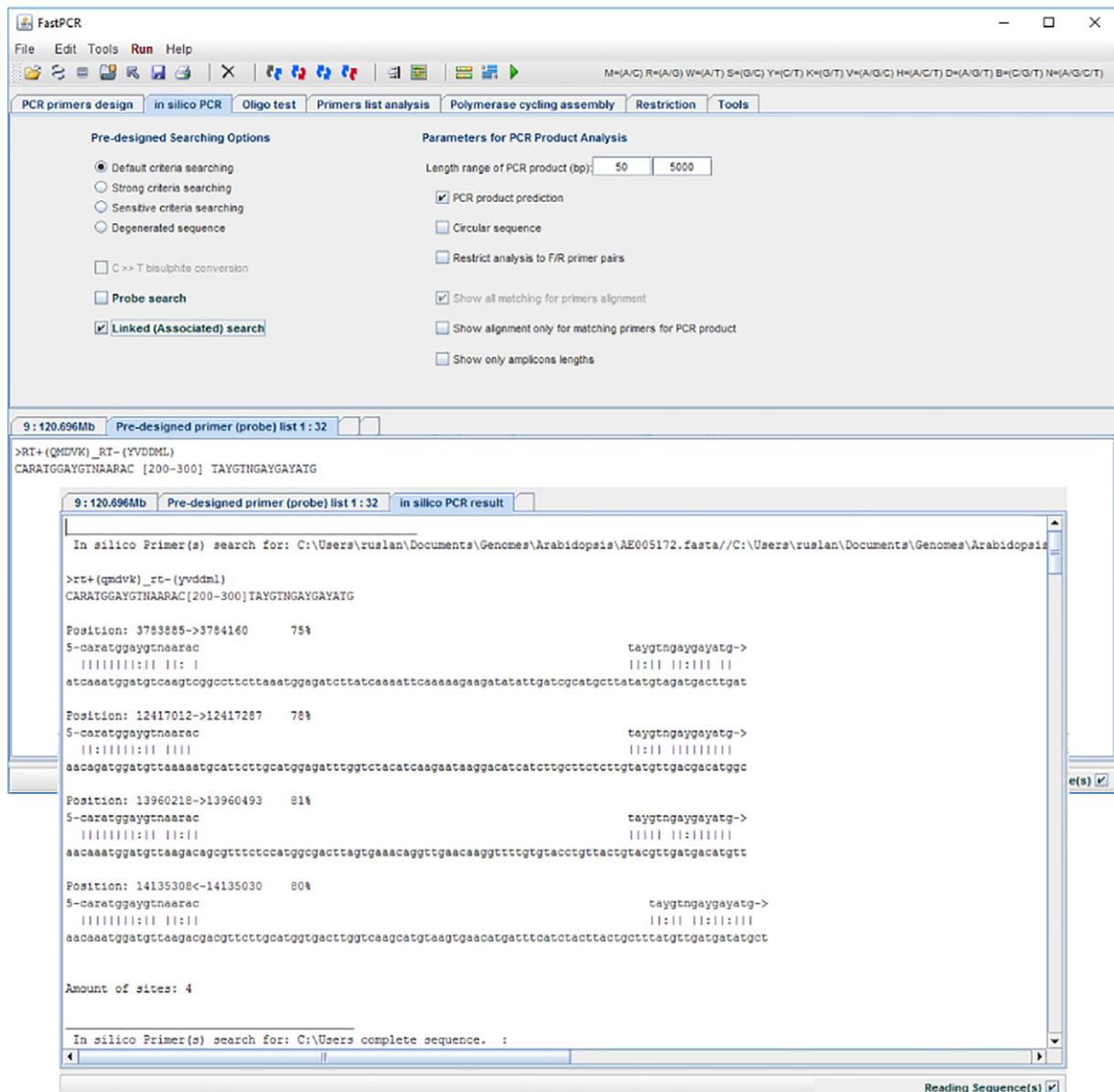


Fig. 1. Results of a linked search using two degenerate Copia-type RT primers against the genome of *Arabidopsis thaliana*.

followed by expected distance between primer binding sites (200–300 nt) and a complementary sequence (**TAYGTNGAYGAYATG**) of the reverse primer (5'-CATRTCRCNACRTA):

```
>RT+(QMDVK)_RT-(YVDDML)
CARATGGAYGTNAARAC [200–300] TAYGTNGAYGAYATG.
```

The result of a linked search using the above query against the genome of *Arabidopsis thaliana* is shown below as local alignment of query and target DNA sequences (Fig. 1):

```
In silico primer(s) search for:
C:\Users\Genomes\Arabidopsis\AE005172.fasta\Arabidopsis thaliana chromosome
1 top arm, complete sequence:
```

```
>RT+(QMDVK)_RT-(YVDDML)
CARATGGAYGTNAARAC[200–300]TAYGTNGAYGAYATG
```

```
Position: 3,783,885 → 3,784,160          75%
5-CARATGGAYGTNAARAC          TAYGTNGAYGAYATG→
|||||||:| | | | |              ||:| | | | |
atcaaatggatgcaagtcggcctctt/..tgatcgcagccttatatgtagatgacttgat
```

As evidenced from Fig. 1, the results of linked search and those of *in silico* PCR differ only in presentation to the user. For example, in the case of Copia-type RT analysis, the complete sequence of *Arabidopsis* reverse transcriptase pseudogene (see below) can be obtained by linked searching but not by *in silico* PCR analysis.

```
atCAATGGATGTCAGTCggcctctttaatggagatcttgaagaagaagttacattgagcaaccacaagg
ctacatagctcaagggtgaagaagacaagcttgaggctaaaaaggtgctttatggattaaacaagcccaagag
cttggatctcaagtgacaagatttcaaggagaagatttcatcaagtgccatgatgagcagccttatcaaa
attcaaaaagaagatatattgatcgcagccttatTATGTAGATGACTTGat
```

Another task where linked searching is useful is analysis of two or more nested nucleotide sequences. In the classical *in silico* PCR, the necessary condition for amplicon search is existence of two primer-binding sites on complementary DNA strands, which are located at a certain distance and orientation relative to each other. In LAMP, the strand displacement-type of isothermal DNA amplification, template nucleic acid strands are mixed with three or four nested primer pairs [25,3]. Linked searching allows the user to analyse primer-template matching during LAMP and predict whether existing primer pairs can amplify a newly discovered sequence variant. Linked searching can be used to quickly determine if the designed LAMP primers can amplify related, consensus or repeated sequences. Such analysis cannot be easily performed with conventional virtual PCR software.

Yet another compelling example of linked searching is the *in silico* extraction of LTR-retrotransposon sequences from plant genomes [44]. The LTR-retrotransposon Cassandra universally carries conserved 5S rRNA sequences and associated RNA polymerase III promoters in its long terminal repeat (LTR) and is found in all vascular plants. Linked searching for conserved Cassandra sequences can detect new copies of this retrotransposon in plant genomes and reveal the presence of Cassandra in plant species where it has not been previously identified.

5S rRNA sequences from the Cassandra retrotransposon contains two conservative regions, boxA (RGTTAAGYRHGY) and boxC (RRRATRGGTRACY), separated by 18 nucleotides. Furthermore, the Cassandra retrotransposon contains a conserved sequence (TGGTATCAGAGC) called primer binding site or PBS. In ferns, PBS is located at a distance of 8 nt from the 5S rRNA sequence and in *Brassica* species the distance is much longer (173 bp). Therefore, the following query can be used to search both fern and seed plant genomes for the Cassandra LTR-retrotransposon sequences:

```
>Cassandra_LTR_PBS.
RGTTAAGYRHGY [15–25] RRRATRGGTRACY [5–200]
TGGTATCAGAGC.
```

The following genomes were searched with the above query: *Arabidopsis thaliana*, *Brachypodium distachyon*, *Glycine max*, *Zea mays*, *Oryza sativa*, *Sorghum bicolor*, *Vitis vinifera*, *Medicago truncatula*, *Saccharum hybrid*, *Panicum virgatum*, *Solanum lycopersicum* and a

Table 3

Total numbers of hits returned by the linked search of several eukaryotic genomes for the highly-conserved sequence of the plant LTR retrotransposon Cassandra.

Genome	Size in Mb	Predicted copy number
<i>Arabidopsis thaliana</i>	126	4
<i>Brachypodium distachyon</i>	275	5
<i>Oryza sativa</i>	380	69
<i>Medicago truncatula</i>	391	32
<i>Vitis vinifera</i>	420	2
<i>Sorghum bicolor</i>	545	64
<i>Glycine max</i>	965	3
<i>Zea mays</i>	2100	817
<i>Homo sapiens</i>	3140	0

shotgun sequence of *Citrus clementina*. The linked search results are summarized in Table 3 and in Supplemental data. They include target fragments, sequence alignments and detailed information for each query sequence including genomic position and sequence similarity (Fig. 2).

The largest number of complete Cassandra retrotransposon sequences (701 in total) were detected in the *Zea mays* genome (release B73) and the smallest in the genomes of *Arabidopsis thaliana* (4) and *Brachypodium distachyon* (5) (see Table 3). As expected, no Cassandra sequences were found in the human genome. These results correlate well with the experimental data on the copy number of the Cassandra retrotransposons in the studied genomes. Analysis of the genomes of *Sorghum bicolor*, *Vitis vinifera*, *Medicago truncatula*, *Saccharum hybrid*, *Panicum virgatum* and *Solanum lycopersicum* identified a large number of previously unknown Cassandra retrotransposons. Several novel consensus sequences of complete Cassandra retrotransposons were submitted to the NCBI database (accession numbers EU140956, EU177767, EU867815, EU882730, FJ975775-FJ975780, HM481419, HM481420, KC686837-KC686839 and KM262797). Thus, using the linked search algorithm of our software we were able to identify new Cassandra LTR retrotransposons *in silico* without relying on expensive and time-consuming wet lab methods such as molecular cloning.

3.2.1. Comparison with other software

Currently, several web-based software tools are freely available for *in silico* PCR analysis. Among the most commonly used are NCBI/Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>), the web server “*In silico* simulation of molecular biology experiments” (<http://insilico.ehu.es/>, [28]), WebPCR (<http://webpcr.appspot.com/>) and UCSC In-Silico PCR (<http://genome.ucsc.edu/cgi-bin/hgPcr>). Table 4 compares these web-based software tools with FastPCR (<http://primerdigital.com/tools/pcr.html>). Unlike FastPCR, other web-based software does not go beyond the classical PCR approach with two primers. For example, NCBI/Primer-BLAST is unable to search for pairs of queries separated by arbitrary sequences of variable length, is incompatible with degenerate nucleotide sequences, does not take into account sequence repeats, does not accept one or several primers, does not allow the use of primers with the same sequence and is incompatible with methods employing molecular beacons. Maximum primer length is capped at 36 nt while it is unlimited in our software. If the primer has multiple binding sites at repeated sequences, NCBI/Primer-BLAST locates and displays only the most remote ones. Finally, primer order is fixed in NCBI/Primer-BLAST; forward and reverse primers must be entered according to the following criteria: “forward primer (5'→3' on plus strand)” and “reverse primer (5'→3' on minus strand)”. Some of the above limitations are experimentally demonstrated in Supplemental materials (S1).

Our online Java software is fully compatible with PCR fingerprinting techniques based on sequence repeats such as SINE-PCR, IRAP, ISSR or RAPD. It allows the use of single or multiple primers to virtually amplify hundreds to thousands of inter-repeat DNA sequences for genome fingerprinting. The software is able to quickly search for primers or probes, determine their location, orientation, melting temperatures and

Table 4
Comparison of several *in silico* PCR tools.

Features	FastPCR (http://primerdigital.com/ tools/pcr.html)	Primer-BLAST (http://www.ncbi.nlm.nih.gov/ tools/primer-blast/)	The web server: <i>in silico</i> experiments with complete genomes (http://insilico.ehu.es/)	WebPCR (http://webpcr.appspot.com/)
Input				
Whole genomes/maximal length of PCR template	+/any	+/50000	+/-	-/any
Multiple templates	+	-	-	-
Linear/circular templates	+/+	+/-	+/-	+/-
Bisulphite modification compatibility	+	-	-	-
Degenerate oligonucleotides in primer and template	+	-	-	-
Accepted number of primers	any	2	2	2
Minimal/maximal primer length	5(4)/any	4/36	10/35	10/35
Searching				
Singleplex/multiplex primers and probes	+/+	+/-	+/-	+/-
Mismatches accepted within primer and target	+	+	-	-
Addition of non-complementary sequences at one or both termini of probe or primer	+	-	-	-
Linked (associated) sequences searching	+	-	-	-
TaqMan assay, two primers and one probe	-	-	-	-
Exon/intron selection	-	+	-	-
<i>In silico</i> PCR-RFLP	-	-	+	-
DNA fingerprinting	+	-	+	-
<i>In silico</i> PCR for techniques based on repeats (SINE-PCR, IRAP, ISSR, RAPD etc.)	+	-	-	-
BLAST online search with genome database	-	+	-	-
Output				
Graphical interface	+	+	+	+
Annealing and melting temperatures calculated with standard and degenerate oligonucleotides for oligonucleotide-target duplexes	+	-	-	-
Calculation of optimal annealing temperature	+	-	-	+
Search for all variants of amplicons within repeated sequences	+	-	-	-
List of potential amplicons	+	-	-	-
Other				
Primer design parameters validated in lab	+	+	+	+
High-throughput compatibility	+	+	+	-
Relative calculation speed	Instant	Slow	Fast	Quick

+ supported feature, - unsupported feature.

However, our software is not without limitations. It lacks a graphical representation of the results and some technical aspects still require refinement. For example, two primers and a probe in the input may be confused for three primers, unless performing a linked search. The work is ongoing to improve FastPCR performance and functionality. In the future, we plan to extend the capabilities of the algorithm to make it compatible with more experimental techniques. For example, we may add DNA sequence analysis after restriction digestion or nickase treatment. An equally interesting alternative is to integrate searching against data from the Single Nucleotide Polymorphism database (dbSNP). Our software can be developed further to include additional functions and capabilities such as advanced features of LINQ (Language-Integrated Query), which extends query capabilities and provides easily-learned patterns for querying and updating data.

4.1. Availability

The online version of FastPCR is available at <http://primerdigital.com/tools/pcr.html>. The program manual, license agreement, and installation files are available at <http://primerdigital.com/fastpcr/>. YouTube tutorial videos are available at <http://www.youtube.com/user/primerdigital>.

The online software is written in Java with NetBeans IDE (Oracle) and requires the Java Runtime Environment (JRE 8, <http://www.oracle.com/technetwork/java/javase/downloads/>). It is compatible with any operating system running Java (64-bit OS is preferred for large chromosome files).

Before using the online jPCR software, a user has to add the URL <http://primerdigital.com/> to the Exception Site List (https://www.java.com/en/download/faq/exception_sitelist.xml) located under the Security tab of the Java Control Panel (<http://www.java.com/en/download/help/appsecuritydialogs.xml>). Optionally, the user can download self-signed certificates (<http://primerdigital.com/j/primerdigital.cer>) and import them to "Signer CA" (Certificate Authority) from the Java Control Panel. The user has to set "Security Level" to "High" under the Security tab of the Java Control Panel (as shown in the linked picture http://primerdigital.com/image/primerdigital_certificate_big.png).

The software can be downloaded and executed on a desktop computer using one of the following Java Web Start (JavaWS) commands:

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

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

Alternatively, the software can be launched from the following web page: <http://primerdigital.com/tools/pcr.html>

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ygeno.2017.05.005>.

Acknowledgments

Java Web tools are publicly available and may not be reproduced or distributed for commercial use. This work was supported by Primer Digital Ltd. and was partially supported by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (grant 5135/GF4).

References

- [1] J. Walker-Daniels, Current PCR methods, *Materials and Methods* 2 (119) (2012) <http://dx.doi.org/10.13070/mm.en.2.119>.
- [2] L.C. Tisi, O. Gandelman, G. Kiddle, C. Mcelgunn, *Nucleic acid amplification, Canada Patent CA2417798*, 2010.
- [3] T. Notomi, H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino, T. Hase, Loop-mediated isothermal amplification of DNA, *Nucleic Acids Res.* 28 (12) (2000), e63. <http://dx.doi.org/10.1093/nar/28.12.e63>.
- [4] G.T. Walker, M.S. Fraiser, J.L. Schram, M.C. Little, J.G. Nadeau, D.P. Malinowski, Strand displacement amplification—an isothermal, in vitro DNA amplification technique, *Nucleic Acids Res.* 20 (7) (1992) 1691–1696, <http://dx.doi.org/10.1093/nar/20.7.1691>.
- [5] J. Banér, M. Nilsson, M. Mendel-Hartvig, U. Landegren, Signal amplification of padlock probes by rolling circle replication, *Nucleic Acids Res.* 26 (22) (1998) 5073–5078, <http://dx.doi.org/10.1093/nar/26.22.5073>.
- [6] K. Tatsumi, Y. Mitani, J. Watanabe, H. Takakura, K. Hoshi, Y. Kawai, T. Kikuchi, Y. Kogo, A. Oguchi-Katayama, Y. Tomaru, H. Kanamori, M. Baba, T. Ishidao, K. Usui, M. Itoh, P.E. Cizdziel, A. Lezhava, M. Ueda, Y. Ichikawa, I. Endo, S. Togo, H. Shimada, Y. Hayashizaki, Rapid screening assay for KRAS mutations by the modified smart amplification process, *J. Mol. Diagn.* 10 (6) (2008) 520–526, <http://dx.doi.org/10.2353/jmoldx.2008.080024>.
- [7] D.Y. Kwoh, G.R. Davis, K.M. Whitfield, H.L. Chappelle, L.J. DiMichele, T.R. Gingeras, Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format, *Proc. Natl. Acad. Sci. U. S. A.* 86 (4) (1989) 1173–1177.
- [8] E. Fahy, D.Y. Kwoh, T.R. Gingeras, Self-sustained sequence replication (3SR): an isothermal transcription-based amplification system alternative to PCR, *PCR Methods Appl.* 1 (1) (1991) 25–33, <http://dx.doi.org/10.1101/gr.1.1.25>.
- [9] M. Vincent, Y. Xu, H. Kong, Helicase-dependent isothermal DNA amplification, *EMBO Rep.* 5 (8) (2004) 795–800, <http://dx.doi.org/10.1038/sj.embor.7400200>.
- [10] N. Kurn, P. Chen, J.D. Heath, A. Kopf-Sill, K.M. Stephens, S. Wang, Novel isothermal, linear nucleic acid amplification systems for highly multiplexed applications, *Clin. Chem.* 51 (10) (2005) 1973–1981, <http://dx.doi.org/10.1373/clinchem.2005.053694>.
- [11] R. Fang, X. Li, L. Hu, Q. You, J. Li, J. Wu, P. Xu, H. Zhong, Y. Luo, J. Mei, Q. Gao, Cross-priming amplification for rapid detection of Mycobacterium tuberculosis in sputum specimens, *J. Clin. Microbiol.* 47 (3) (2009) 845–847, <http://dx.doi.org/10.1128/JCM.01528-08>.
- [12] Y. Zhao, F. Chen, Q. Li, L. Wang, C. Fan, Isothermal amplification of nucleic acids, *Chem. Rev.* 115 (22) (2015) 12491–12545, <http://dx.doi.org/10.1021/acs.chemrev.5b00428>.
- [13] V.T. Katja Niemann, Isothermal amplification and quantification of nucleic acids and its use in microsystems, *Journal of Nanomedicine & Nanotechnology* 06 (03) (2015) <http://dx.doi.org/10.4172/2157-7439.1000282>.
- [14] M. Fakkruddin, K.S. Mannan, A. Chowdhury, R.M. Mazumdar, M.N. Hossain, S. Islam, M.A. Chowdhury, Nucleic acid amplification: alternative methods of polymerase chain reaction, *J. Pharm. Bioallied Sci.* 5 (4) (2013) 245–252, <http://dx.doi.org/10.4103/0975-7406.120066>.
- [15] W. Liu, D. Dong, Z. Yang, D. Zou, Z. Chen, J. Yuan, L. Huang, Polymerase Spiral Reaction (PSR): a novel isothermal nucleic acid amplification method, *Sci. Rep.* 5 (2015) 12723, <http://dx.doi.org/10.1038/srep12723>.
- [16] P. Smykal, R. Kalendar, R. Ford, J. Macas, M. Griga, Evolutionary conserved lineage of Angela-family retrotransposons as a genome-wide microsatellite repeat dispersal agent, *Heredity (Edinb)* 103 (2) (2009) 157–167, <http://dx.doi.org/10.1038/hdy.2009.45>.
- [17] R. Kalendar, A.H. Schulman, Transposon-based tagging: IRAP, REMAP, and iPBS, *Methods Mol. Biol.* 1115 (2014) 233–255, http://dx.doi.org/10.1007/978-1-62703-767-9_12.
- [18] R. Kalendar, A.J. Flavell, T.H. Ellis, T. Sjakste, C. Moisy, A.H. Schulman, Analysis of plant diversity with retrotransposon-based molecular markers, *Heredity* 106 (4) (2011) 520–530, <http://dx.doi.org/10.1038/hdy.2010.93>.
- [19] E. Hosid, L. Brodsky, R. Kalendar, O. Raskina, A. Belyayev, Diversity of long terminal repeat retrotransposon genome distribution in natural populations of the wild diploid wheat *Aegilops speltoides*, *Genetics* 190 (1) (2012) 263–274, <http://dx.doi.org/10.1534/genetics.111.134643>.
- [20] A. Belyayev, R. Kalendar, L. Brodsky, E. Nevo, A.H. Schulman, O. Raskina, Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat, *Mob. DNA* 1 (6) (2010) 1–16, <http://dx.doi.org/10.1186/1759-8753-1-6>.
- [21] R. Kalendar, D. Lee, A. Schulman, FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis, in: S. Valla, R. Lale (Eds.), *DNA Cloning and Assembly Methods*, Vol 1116. *Methods in Molecular Biology*, Humana Press, Clifton, N.J. 2014, pp. 271–302, http://dx.doi.org/10.1007/978-1-62703-764-8_18.
- [22] R. Kalendar, D. Lee, A.H. Schulman, Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis, *Genomics* 98 (2) (2011) 137–144, <http://dx.doi.org/10.1016/j.ygeno.2011.04.009>.
- [23] M. Lexa, J. Horak, B. Brzobohaty, Virtual PCR, *Bioinformatics* 17 (2) (2001) 192–193, <http://dx.doi.org/10.1093/bioinformatics/17.2.192>.
- [24] B. Yu, C. Zhang, In silico PCR analysis, *Methods Mol. Biol.* 760 (2011) 91–107, http://dx.doi.org/10.1007/978-1-61779-176-5_6.
- [25] N.R. Salinas, D.P. Little, Electric LAMP: virtual loop-mediated isothermal AMPLification, *ISRN bioinformatics* 2012 (2012) 696758, <http://dx.doi.org/10.5402/2012/696758>.
- [26] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis, T.L. Madden, NCBI BLAST: a better web interface, *Nucleic Acids Res.* 36 (2008) (Web Server issue):W5–9, [10.1093/nar/gkn201](http://dx.doi.org/10.1093/nar/gkn201).
- [27] P.C. Boutros, A.B. Okey, PUNIS: transcriptomic- and genomic-in silico PCR for enhanced primer design, *Bioinformatics* 20 (15) (2004) 2399–2400, <http://dx.doi.org/10.1093/bioinformatics/bth257>.
- [28] J. Bikandi, R.S. Millán, A. Rementeria, J. Garaizar, In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction, *Bioinformatics* 20 (5) (2004) 798–799, <http://dx.doi.org/10.1093/bioinformatics/btg491>.
- [29] K. Rotmistrovsky, W. Jang, G.D. Schuler, A web server for performing electronic PCR, *Nucleic Acids Res.* 32 (2004) (Web Server issue):W108–112, [10.1093/nar/gkh450](http://dx.doi.org/10.1093/nar/gkh450).
- [30] S.N. Gardner, T. Slezak, Simulate_PCR for amplicon prediction and annotation from multiplex, degenerate primers and probes, *Bmc Bioinformatics* 15 (1) (2014) 1–6, <http://dx.doi.org/10.1186/1471-2105-15-237>.
- [31] M. Lexa, G. Valle, PRIMEX: rapid identification of oligonucleotide matches in whole genomes, *Bioinformatics* 19 (11) (2003) 2486–2488.
- [32] A.E. Pozhitkov, D. Tautz, An algorithm and program for finding sequence specific oligonucleotide probes for species identification, *BMC Bioinf.* 3 (1) (2002) 9, <http://dx.doi.org/10.1186/1471-2105-3-9>.
- [33] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, *BMC Bioinf.* 13 (2012) 134, <http://dx.doi.org/10.1186/1471-2105-13-134>.
- [34] R. Kalendar, A. Muterko, M. Shamekova, K. Zhambakin, In silico PCR tools a fast primer, probe and advanced searching, in: L. Domingues (Ed.), *PCR Methods and Protocols. Methods in Molecular Biology*, Vol. 1620. Springer-Verlag New York, Clifton, N.J. 2017, pp. 1–10, http://dx.doi.org/10.1007/978-1-4939-7060-5_1.
- [35] R. Kalendar, T. Tseluykh, Introduction on using the FastPCR software and the related Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis, in: L. Domingues (Ed.), *PCR Methods and Protocols. Methods in Molecular Biology*, Vol. 1620. Springer-Verlag New York, Clifton, N.J.) 2017, pp. 21–30, http://dx.doi.org/10.1007/978-1-4939-7060-5_2.
- [36] Z. Ning, A.J. Cox, J.C. Mullikin, SSAHA: a fast search method for large DNA databases, *Genome Res.* 11 (10) (2001) 1725–1729, <http://dx.doi.org/10.1101/gr.194201>.
- [37] E. Giladi, M.G. Walker, J.Z. Wang, W. Volkmut, SST: an algorithm for finding near-exact sequence matches in time proportional to the logarithm of the database size, *Bioinformatics* 18 (6) (2002) 873–877.
- [38] J. SantaLucia Jr., D. Hicks, The thermodynamics of DNA structural motifs, *Annu. Rev. Biophys. Biomol. Struct.* 33 (2004) 415–440, <http://dx.doi.org/10.1146/annurev.biophys.32.110601.141800>.
- [39] N. Peyret, P.A. Seneviratne, H.T. Allawi, J. SantaLucia Jr., Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches, *Biochemistry* 38 (12) (1999) 3468–3477, <http://dx.doi.org/10.1021/bi9825091>.
- [40] J. SantaLucia Jr., H.T. Allawi, P.A. Seneviratne, Improved nearest-neighbor parameters for predicting DNA duplex stability, *Biochemistry* 35 (11) (1996) 3555–3562, <http://dx.doi.org/10.1021/bi951907q>.
- [41] A.N. Lane, J.B. Chaires, R.D. Gray, J.O. Trent, Stability and kinetics of G-quadruplex structures, *Nucleic Acids Res.* 36 (17) (2008) 5482–5515, <http://dx.doi.org/10.1093/nar/gkn517>.
- [42] P. Shing Ho, The non-B-DNA structure of d(CA/TC)_n does not differ from that of Z-DNA, *Proc. Natl. Acad. Sci. U. S. A.* 91 (20) (1994) 9549–9553.
- [43] A.J. Flavell, E. Dunbar, R. Anderson, S.R. Pearce, R. Hartley, A. Kumar, Ty1-copigroup retrotransposons are ubiquitous and heterogeneous in higher plants, *Nucleic Acids Res.* 20 (14) (1992) 3639–3644, <http://dx.doi.org/10.1093/nar/20.14.3639>.
- [44] R. Kalendar, J. Tanskanen, W. Chang, K. Antonius, H. Sela, O. Peleg, A.H. Schulman, *Cassandra* retrotransposons carry independently transcribed 5S RNA, *Proc. Natl. Acad. Sci. U. S. A.* 105 (15) (2008) 5833–5838, <http://dx.doi.org/10.1073/pnas.0709698105>.