



UNIVERSITY OF HELSINKI



<https://helda.helsinki.fi>

Helda

FASTPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis

Kalendar, Ruslan

2014

Kalendar, R, Lee, D & Schulman, A 2014, 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, vol. 1116, Humana press, Springer Science+Business Media, New York, pp. 271-302. https://doi.org/10.1007/978-1-62703-764-8_18

<http://hdl.handle.net/10138/155666>

10.1007/978-1-62703-764-8_18

unspecified

submittedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Metadata of the chapter that will be visualized online

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

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

Keywords
(separated by “-”)

PCR primer design - Primer linguistic complexity - Sequence assembly - Software probe design - Ligase chain reaction - DNA primers

FastPCR Software for PCR, In Silico PCR, and Oligonucleotide Assembly and Analysis 2 3

Ruslan Kalendar, David Lee, and Alan H. Schulman 4

Abstract 5

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

Key words PCR primer design, Primer linguistic complexity, Sequence assembly, Software probe design, Ligase chain reaction, DNA primers 24 25

Abbreviation 26

OE-PCR	Overlap extension PCR	27
PCR	Polymerase chain reaction	28
RT-PCR	Real-time PCR	29
SSR	Simple sequence repeat	30

31 1 Introduction

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

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

66 2 Software, General Information

67 The FastPCR software (<http://primerdigital.com/fastpcr.html>) is
68 written in Microsoft Visual Studio 6.0 and compiled to an execu-
69 tive file that, after installation, can be used with any version of
70 Microsoft Windows. For Linux and Mac it requires “Wine”
71 (<http://www.winehq.org/>) as a compatibility layer for running
72 Windows programs. It is a completely free alternative implementa-
73 tion of the Windows API also for use with native Windows DLLs.

t1.1 **Table 1**
 t1.2 **Summary of the FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis**

t1.3	Features
t1.4	PCR tool provides comprehensive facilities for
t1.5	Design of primers for most PCR applications and their combinations, including standard, multiplex, long-distance, inverse, real-time, unique (specific primers for each from genetically related DNA sequences) or group-specific (universal primers for genetically related DNA sequences), linear-after-the-exponential (LATE)-PCR, bisulphite modification assays, polymerase extension PCR multi-fragment assembly cloning
t1.6	
t1.7	
t1.8	
t1.9	
t1.10	Design of long oligonucleotides for microarray analyses and dual-labeled oligonucleotides for probes such as molecular beacons
t1.11	
t1.12	Polymerase chain assembly (PCA) or oligo assembly—for automating the design of oligonucleotide sets for long sequence assembly by ligase chain reaction (LCR) and PCR
t1.13	
t1.14	In silico (virtual) PCR or multiple primer or probe searches, or in silico PCR against whole genome(s) or a list of predictions by chromosome of probable PCR products, and search for potential mismatching locations of the specified primers or probes
t1.15	
t1.16	
t1.17	Testing of individual primers, melting temperature calculation for standard and degenerate oligonucleotides including LNA and other modifications
t1.18	
t1.19	Evaluation of PCR efficiency, linguistic complexity, dimer and G/C-quadruplex detection, dilution and resuspension calculator
t1.20	
t1.21	Analysis of features of multiple primers simultaneously, including T_m , CG content, linguistic complexity, dimer formation; optimal T_a
t1.22	
t1.23	Identification of simple sequence repeat (SSR) loci by analyzing the low-complexity regions of input sequences
t1.24	
t1.25	Restriction digest analyses for Type I, II, and III restriction enzymes and homing endonucleases, finding or creating restriction enzyme recognition sites for coding sequences
t1.26	
t1.27	Searches for similar sequences (or primers)
t1.28	Translation of nucleotide (DNA/RNA) sequences to the corresponding peptide sequence in all six frames for standard and degenerate DNA and modifications (inosine, uridine)
t1.29	
t1.30	Determination of CG:(G-C)/(G+C), AT:(A-T)/(A+T), SW:(S-W)/(S+W), MK:(M-K)/(M+K), purine-pyrimidine (R-Y)/(R+Y) skews, CG% content and the melting temperature, primer quality and linguistic sequence complexity profiles
t1.31	
t1.32	
t1.32	

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

74
75
76
77
78

79

3 The Interface

3.1 Inputs to FastPCR

80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96

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

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

[AU2]

this figure will be printed in b/w

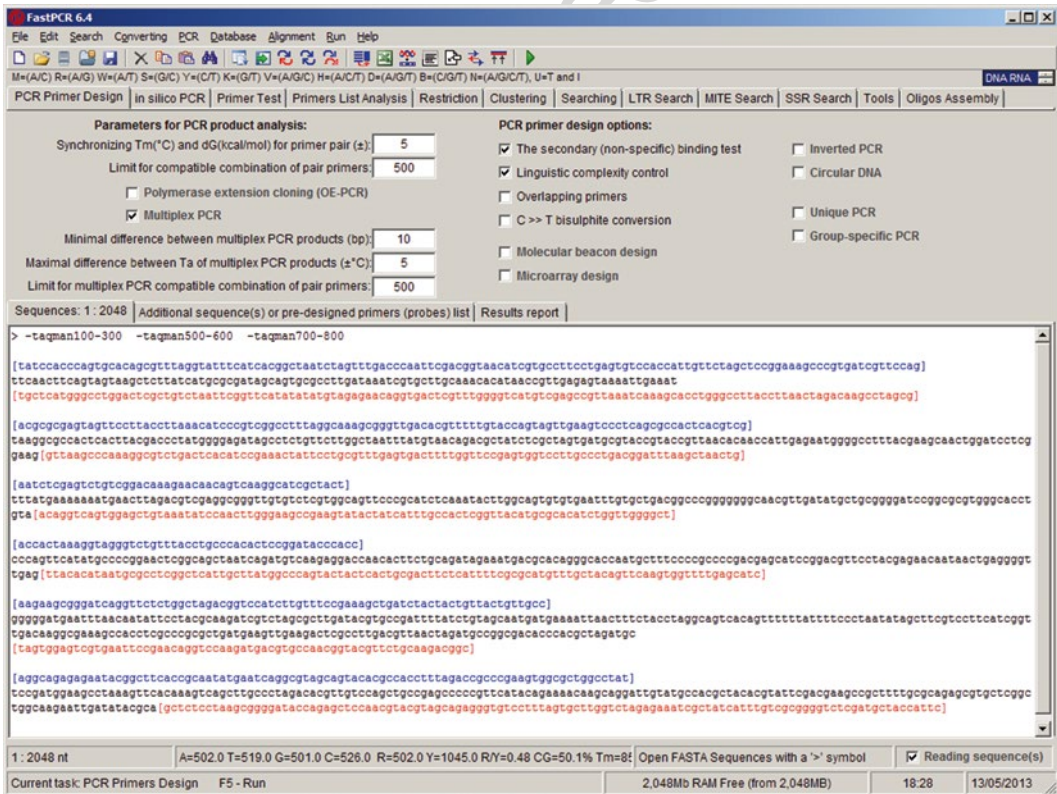


Fig. 1 The FastPCR sequence editor and user interface

- 3.2 Program Output** FastPCR automatically generates results at a third text editor, “Result report,” in tabulated format for transferring to spreadsheet software from the clipboard using copy–paste. Output results are easy to save as Excel worksheet (.xls) or as Rich Text Format (.rtf) text files, compatible with MS Excel or Open Office. The separated output of the primer design is a list of primers, a set of primer pair sequences with their theoretical PCR products, and, for multiplex PCR, the result of the calculation of multiple-PCR primers for given target sequences. In addition, the output shows optimal annealing temperature for each primer pair, the size of PCR product, and complete information for each designed primer and for each multiplex PCR product set.
- 3.3 Sequence Entry** Sequence data files are prepared using a text editor (Notepad, WordPad, MS Word), and saved in ASCII as text/plain format (.txt) or in .rtf. The program either takes a single sequence or accepts multiple separate DNA sequences in FASTA, tabulated format (two columns from MS Excel sheet or MS Word table), EMBL, MEGA, GenBank, MSF, DIALIGN, simple alignment, or BLAST Queue web alignment formats. The template length is not limited. The FastPCR clipboard allows the user to copy and paste text or tables from MS Word documents or MS Excel worksheets or other programs and to paste them into another Office document. It is important that all target sequences are prepared in the same format. Users can type or import from file(s) into “General Sequence(s)” or “Additional sequence(s) or pre-designed primers (probes) list” editors.
- FastPCR allows files to be opened in several ways: the original file can be opened as read-only for editing with text editors; files can be opened to memory without using text editors, which allows larger file(s), up to 200 Mb, to be analyzed; files within a folder can be selected and the files opened during task execution without the use of text editor program. Additionally, the program can open files within a selected folder in order to join all these files in a text editor. For example, this feature can be applied to convert all files from a selected folder into a single file of FASTA sequences. Alternatively this feature allows splitting FASTA sequences to individual files in a particular selected folder.
- When a sequence file is opened, FastPCR displays the information about the opened sequence and its format. The information status bar shows the number of sequences, the total sequence length (in nucleotides), the nucleotide composition, and the purine, pyrimidine, CG percentage, and the melting temperature. Files can be saved in various formats including .rtf, .xls, or txt from the text editor in use.
- 3.4 FASTA Format Description** FastPCR normally expects to read sequence files in FASTA format [10]. FASTA format has the highest priority and is simple, because the raw sequence is merely preceded by a definition line. The definition line begins with a “>” sign and optionally followed

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

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

160 **3.5 Alignment** 161 **Format Description**

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

173 **4 The PCR Primers or Probe Design Analysis Options**

174 **4.1 PCR Primer** 175 **Design Generalities**

176 Primer design is one of the key steps for successful PCR. For PCR
 177 applications, primers are usually 18–35 bases in length and should
 178 be designed such that they have complete sequence identity to the
 179 desired target fragment to be amplified. The parameters, either
 180 controllable by the user or selected automatically, are primer length
 181 (12–500 nt), melting temperature for short primers calculated by
 182 nearest neighbor thermodynamic parameters, theoretical primer
 183 PCR efficiency (quality at %) value, primer CG content, 3' end
 184 terminal enforcement, preferable 3' terminal nucleotide sequence
 185 composition in degenerated formulae, and added sequence tags at
 186 5' termini. The other main parameters used for primer selection
 are the general nucleotide structure of the primer such as linguistic
 complexity (nucleotide arrangement and composition), specificity,

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

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

t2.11 ^aNearest neighbor thermodynamic parameters according to SantaLucia [13]

t2.12 ^bAmbiguity codes

t2.13 ^cSequence linguistic complexity measurement calculated using the alphabet-capacity L -gram method

the melting temperature of the whole primer, the melting 187
temperature at the 3' and 5' termini, self-complementarity, and 188
secondary (nonspecific) binding. 189

The software can dynamically optimize the best primer length 190
for the entered parameters. All PCR primer (probe) design param- 191
eters are flexible and changeable according to the specifics of the 192
analyzed SEQUENCES and tasks. Primer pairs are analyzed for 193
cross-hybridization and specificity of both primers and, optionally, 194
selected with similar melting temperatures. Primers with balanced 195
melting temperatures (within 1–6 °C of each other) are desirable 196
but not mandatory. The default primer design selection criteria are 197
shown in Table 2. It is possible to use predesigned primers or 198
probes or, alternatively, predesigned primers can act as references 199
for the design of new primers. The program accepts a list of prede- 200
signed oligonucleotide sequences and checks the compatibility of 201
each primer with a newly designed primer or probe. 202

[AU3]

4.2 Melting Temperature Calculation

The melting temperature (T_m) is defined as the temperature at 203
which half the DNA strands are in the double-helical state and half 204
are in the “random-coil” state. The T_m for short oligonucleotides 205
with normal or degenerate (mixed) nucleotide combinations is 206
calculated in the default setting using nearest neighbor 207
thermodynamic parameters [12, 13]. The CG content of an 208
oligonucleotide is the most important factor that influences the T_m 209
value. The melting temperature for mixed bases is calculated by 210
averaging nearest neighbor thermodynamic parameters—enthalpy 211
and entropy values—at each mixed site; extinction coefficient is 212
similarly predicted by averaging nearest neighbor values at mixed 213
sites [2, 3]. Mismatched pairs can be taken into account since the 214
parameters provide for DNA/DNA duplexes and dangling ends, 215

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

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

$$229 \quad T_m(^{\circ}\text{C}) = 2(L + G + C),$$

230 for oligonucleotides longer than ten bases:

$$231 \quad T_m(^{\circ}\text{C}) = 64.9 + \frac{41([G + C] - 16.4)}{L}$$

232 or the formula [18]:

$$233 \quad T_m(^{\circ}\text{C}) = 77.1 + 11.7 \log_{10} [K^+] + \frac{41[G + C] - 528}{L}$$

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

$$239 \quad T_m(^{\circ}\text{C}) = 81.5 + 16.6 \log_{10} [K^+] + \frac{41[G + C] - 675}{L}$$

240 **4.3 Linguistic** 241 **Complexity of** 242 **Sequence and** 243 **Nucleotide-Skew** 244 **Analysis**

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

$$LC(\%) = \frac{100 \times \sum_{i=1}^L x_i}{E}, \tag{253}$$

where 254

$$E = \sum_{i=1}^L \begin{cases} s-i+1, & s < 4^i - 1 + i \\ 4^i, & s \geq 4^i - 1 + i \end{cases}, \tag{255}$$

$$L = \left\lceil \log_4 \left(\frac{s}{3} \right) + 1 \right\rceil. \tag{256}$$

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

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

**4.4 Primer Quality
 (Virtual PCR
 Efficiency)
 Determination**

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

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

292
293
294
295
296
297

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

298 **4.5 Hairpin (Loop)** 299 **and Dimer Formation**

300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338

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

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

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

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

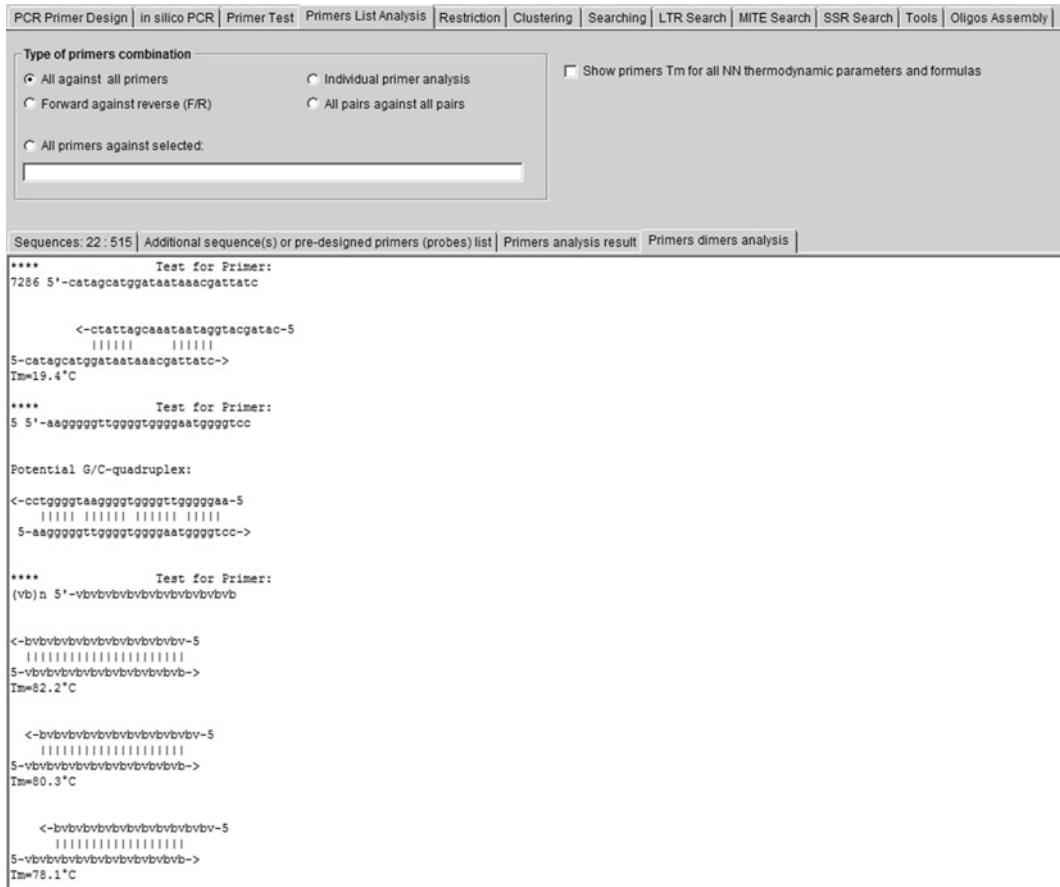


Fig. 2 One of the FastPCR duplex formation results

Intermolecular G-quadruplex-forming sequences are detected 339
 according to the formula $\dots G_{m1} X_n G_{m2} \dots$, where m is the number 340
 of G residues in each G-tract ($m_1, m_2 \geq 3$); the gap X_n ($n \leq 2 \cdot$ minimal 341
 ($m_1; m_2$)) can be any combination of residues, including G 342
 [16]. The gap sequences (X_n) may have varying lengths, and a relatively 343
 stable quadruplex structure may still be formed with a loop 344
 more than seven bases long, but in general, increasing the length 345
 of the gap leads to a decrease in structure stability. It is also possible 346
 for one of the gaps to be zero length when there are long poly-G 347
 tracts of >6 bases. 348

4.6 Calculation of Optimal Annealing Temperature

The optimal annealing temperature (T_a) is the temperature, 349
 generally stated as a range, where efficiency of PCR amplification is 350
 maximal but nonspecific products minimal. The most important 351
 values for estimating the T_a are the primer quality, the T_m of the 352
 primers, and the length of PCR fragment. Primers with high T_m 353
 (>60 °C) can be used in PCRs with a wide T_a range compared to 354

355 primers with low T_m (<50 °C). The optimal annealing temperature
356 for PCR is calculated directly as the value for the primer with the
357 lowest T_m (T_m^{\min}). However, PCR can work in temperatures up to
358 10° higher than the T_m of the primer so as to favor primer target
359 duplex formation:

360 $T_a(^{\circ}\text{C}) = T_m^{\min} + \ln L$, where L is the length of the PCR
361 fragment.

362 **4.7 The Secondary** 363 **Nonspecific Binding** 364 **Test; Alternative** 365 **Amplification Products**

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

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

394 The mismatches at the 3' end of the primers affect target
395 amplification much more than mismatches at the 5' end. A two-
396 base mismatch at the 3' end of the primer prevents amplification.
397 A single-base mismatch at the 3' end as well as several mismatches
398 at the 5' end of the primer permits amplification, albeit with
399 reduced efficiency. However, the presence of multiple primer bind-
400 ing sites does not necessarily lead to alternative amplification prod-
401 ucts because, for successful amplification, the priming sites for both
402 primers must be both located close to each other, in correct orien-
403 tation, and sufficiently match the primer sequences.

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

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

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

5 Methods

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

5.1 Execution of the Selected Task

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

5.2 PCR Primer Design Options

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

Sequences: 3 : 844		Additional sequence(s) or pre-designed primers (probes) list			PCR primers design result				
PrimerID	Sequence (5'-3')	Length (bp)	Tm (°C)	dG (kcal/mol)	Tm 3'end (°C)	CG (%)	Linguistic_Complexity (%)	Primer_Quality (%)	
example1									
1F1_1_18-39	togtattcaggcgtaacctctg	22	56.5	-27.3 41.2	50.0 88	88			
1F2_1_42-62	gocctcgggtgcgcttactctg	21	64.0	-30.0 40.3	66.7 85	85			
1F3_1_53-73	cgctactctcgttcggatagg	21	56.6	-26.3 41.4	52.4 87	87			
1F4_1_65-85	goggataggaactatctcgtg	21	57.3	-26.3 37.2	57.1 85	83			
1F5_1_78-97	tctcgggtgtttctcactgc 20	55.4	-25.0	36.3 50.0	76	73			
1F6_1_99-118	tggctcccgagctcaatgc 20	61.1	-27.9	43.4 60.0	92	92			
1R1_1_348-368	atgtgggtttgccttacaagc	21	56.7	-26.1 40.2	47.6 92	90			
1R2_1_264-286	tgttggttactctcogtaaggt	23	55.3	-27.4 38.8	43.5 80	80			
1R3_1_161-180	atgatgcctcatgtccgta 20	55.5	-25.1	37.7 50.0	82	82			
1R4_1_136-157	tgttttatcgacacctcgtcc	22	55.2	-26.6 42.8	45.5 85	80			
Forward_PrimerID	Sequence (5'-3')	Tm (°C)	Primer_Quality (%)	Reverse_PrimerID	Sequence (5'-3')	Tm (°C)	Primer_Quality (%)		
	PCR_Fragment_Size (bp)	Topt (°C)							
1F1_1_18-39	togtattcaggcgtaacctctg	56.5	88	1R1_1_348-368	atgtgggtttgccttacaagc	56.7	90	351 62.0	
1F1_1_18-39	togtattcaggcgtaacctctg	56.5	88	1R2_1_264-286	tgttggttactctcogtaaggt	55.3	80	269 61.0	
1F1_1_18-39	togtattcaggcgtaacctctg	56.5	88	1R3_1_161-180	atgatgcctcatgtccgta 55.5	82	163	61.0	
1F1_1_18-39	togtattcaggcgtaacctctg	56.5	88	1R4_1_136-157	tgttttatcgacacctcgtcc	55.2	80	140 60.0	
1F3_1_53-73	cgctactctcgttcggatagg	56.6	87	1R1_1_348-368	atgtgggtttgccttacaagc	56.7	90	316 62.0	
1F3_1_53-73	cgctactctcgttcggatagg	56.6	87	1R2_1_264-286	tgttggttactctcogtaaggt	55.3	80	234 61.0	
1F3_1_53-73	cgctactctcgttcggatagg	56.6	87	1R3_1_161-180	atgatgcctcatgtccgta 55.5	82	128	60.0	
1F3_1_53-73	cgctactctcgttcggatagg	56.6	87	1R4_1_136-157	tgttttatcgacacctcgtcc	55.2	80	105 60.0	
1F4_1_65-85	goggataggaactatctcgtg	57.3	83	1R1_1_348-368	atgtgggtttgccttacaagc	56.7	90	304 62.0	
1F4_1_65-85	goggataggaactatctcgtg	57.3	83	1R2_1_264-286	tgttggttactctcogtaaggt	55.3	80	222 61.0	
example2									
2F1_1_9-28	gagcacataggaacgctcca 20	57.3	-25.8	44.5 55.0	92	92			
2F2_1_58-77	taatcggcgcaacggaggt 20	60.8	-28.1	47.6 60.0	84	80			
2F3_1_84-105	gtgctacaacacacggtaagtc	22	56.8	-27.0 40.6	50.0 88	88			
2F4_1_96-115	acggtaagtctcactcctcca 20	55.3	-25.2	40.3 50.0	82	80			
2R1_1_96-115	ttggacgtgagacttacgct 20	55.3	-25.2	37.6 50.0	82	82			
2R2_1_84-105	gactaacctgggtgttagcac	22	56.8	-27.0 38.3	50.0 88	86			
2R3_1_67-88	agcacatattacctcctcgtg	22	55.2	-26.4 41.8	45.5 90	80			
2R4_1_56-75	ctccgtgtgcgcgattagg 20	61.7	-27.9	46.4 65.0	87	87			
2R5_1_16-35	cgagctctggagcgttctcta 20	58.8	-26.6	42.0 60.0	84	84			
2R6_1_5-25	agcgttctatgtgctccgca	21	60.7	-28.9 46.1	57.1 87	87			
Forward_PrimerID	Sequence (5'-3')	Tm (°C)	Primer_Quality (%)	Reverse_PrimerID	Sequence (5'-3')	Tm (°C)	Primer_Quality (%)		
	PCR_Fragment_Size (bp)	Topt (°C)							
2F1_1_9-28	gagcacataggaacgctcca	57.3	92	2R1_1_96-115	ttggacgtgagacttacgct 55.3	82	107	60.0	
2F1_1_9-28	gagcacataggaacgctcca	57.3	92	2R2_1_84-105	gactaacctgggtgttagcac	56.8	86	97	61.0

Fig. 3 Primer design result window

this figure will be printed in b/w

PCR Primers or Probe Design Options

Length range (12-500 bases)

Min: Max:

Tm range (20-72°C)

Min: Max:

Tm at 3' end (10-50°C)

Min: Max:

CG content range (20-75%)

Min: Max:

Special settings for primers or probes

Default primers

Best primers

Long-distance PCR

Quantitative PCR

Microarray design

Low CG% sequence

Difficult sequence

Degenerate sequence

Any primers

Tm of oligo calculation parameters

K⁺ (and Na⁺, NH₄⁺, Tris⁺) concentration (mM):

Mg²⁺ concentration (mM):

DMSO concentration (%):

Oligo concentration (µM):

Ends nucleotide composition

5'-...-3'

Minimal size is 2 letter (maximum - primer length), 5'-nn-3' is any, 5'-sw-3' - all variants for 5'-(C/G)(A/T)(A/T)-3'

Non-specific primer(probe) binding test parameters

Load dataset file with reference DNA sequence(s):

Synchronizing with dataset by sequence names

Excluding synchronizing with dataset by sequence names

Without synchronizations

Primer quality limit (0-100%):

Primer linguistic complexity limit (0-100%):

Primer's Tm optimisation

Fig. 4 The “PCR Primers or Probes Design Options” window

“{-ln20-23 -tm55-68 -3tm37-50 -cg41-70 -q70 446
 -lc75 -npr400 -c5[NN] -c3[SWW SSW SWS WSS]},” 447
 where “-ln20-23” determines the range of primer length (20– 448
 23 bases), “-tm55-68” determines the range of primer T_m (55– 449
 68 °C), “-3tm37-50” determines the range of primer T_m at 3’ 450
 end (37–50 °C), “-cg41-70” determines the range of primer 451
 CG% contents (between 41 and 70 %), “-npr400” shows the 452
 maximum number of primers (400) designed to each target, 453
 “-c5[NN]” denotes a primer having no specific sequence pattern 454
 for 5’ ends, and “-c3[SWW SSW SWS WSS]” specifies primers 455
 that conform to particular patterns of ambiguity, such as that 456
 shown here for example. 457

**5.3 Examples for
 Primer Selection
 Region**

Users can specify, individually for each sequence, multiple locations 458
 for both forward and reverse primer placement with the commands: 459
 “-FpdN1-N2” for forward primers and “-RpdN1-N2” for reverse 460
 primers, where from N1 to N2 are bases from the selected regions; 461
 “-pdN1-N2” (see more at: [http://primerdigital.com/soft/pcr_](http://primerdigital.com/soft/pcr_help.html) 462
[help.html](http://primerdigital.com/soft/pcr_help.html)). Alternatively, users can specify multiple locations for 463
 both forward and reverse primers using [“and”] inside each 464
 sequence: the software allows multiple and independent locations 465
 for both forward and reverse primers inside each of the sequences. 466
 Whilst PCR primer design will be performed independently for 467
 different targets, multiplex PCR primer design can be performed 468
 simultaneously with multiple amplicons within a single sequence as 469
 well as for different sequences, i.e., all possible combinations of 470
 [“and”] inside one or more sequences. By default, the software 471
 designs primers within the entire sequence length. Optionally, 472
 users can specify, individually for each sequence, multiple locations 473
 for both forward and reverse primers with the commands: 474

1. The same location for both forward and reverse primers will be 475
 designed in the central [nnnnnnnnnn] part (“[]” is used 476
 only once): 477
 [nnnnnn] 478
2. Different locations for forward and reverse primers; forward 479
 primers will be chosen inside the “[1nnnnnn]” location and 480
 reverse primers inside “[2nnnnnn]” location (twice “[]”): 481
 [1nnnnnn] [2nnnnnn] 482
3. Primers must flank the central “[]nnnnnn[]”; forward primers 483
 will be chosen from [I to “A]” bases and reverse primers will be 484
 chosen from base “[C” to the end of sequence:A] 485
 nnnnnn[C 486
4. Forward and reverse primers have an overlapping part 487
 “[nnnnnn]”; forward primers will be chosen from “[A to n]” 488
 bases and reverse primers will be chosen from “[n base to C]”: 489
 [A [nnnnnn] C] 490

491 The software allows the selection of any number of independent
 492 PCR primer (or probe) designing tasks for each sequence using mul-
 493 tiple combinations of “[. .]” and $-FpdN1-N2$, $-RpdN1-N2$, or
 494 $-pdN1-N2$ commands. Multiplex PCR can be carried out simulta-
 495 neously within a single sequence with multiple tasks as well as for
 496 different sequences with multiple tasks or a combination of both. [AU4]

497 All possible combinations of “[]” (forward) with “[]”
 498 (reverse) within the sequence(s):

- 499 1. []
- 500 2.] [
- 501 3. [] []
- 502 4. [[]]
- 503 5. ([] []) n or/and ([[]]) n .

504 **5.4 User-Defined** 505 **PCR Product Size**

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

512 **5.5 PCR Set-up** 513 **Examples with** 514 **Individual Commands**

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

```
517 "-fpr [ggagagtagcttacctcgct cggttaaggttct-tcatgc]  

  518 -npd"
```

519 will analyze the selected sequence between the two primers (5'
 520 ggagagtagcttacctcgct and 5' cggttaaggttcttcatgc).

- 521 2. Design primers with a difference in T_m of about 10° , e.g., for
 522 LATE-PCR:

```
523 "-Ftm50-55 -Rtm64-68 -pTMs10."
```

- 524 3. Design primers with a specific restriction enzyme site at the 3'
 525 end (“-z3eNameEnzyme,” “-Fz3eNameEnzyme,” “-Rz3e
 526 NameEnzyme,” where “NameEnzyme” is the name of the
 527 restriction enzyme: “-z3eXceI.” The alternative command
 528 (“-c3NN”) is also used for a special primer location. For exam-
 529 ple, “-c3YCATGR” is the same as “-z3eXceI”: both com-
 530 mands will design primers with the recognition site for *Xce I*
 531 endonucleases 5'-YCATGR-3' at the 3' end of the primers.

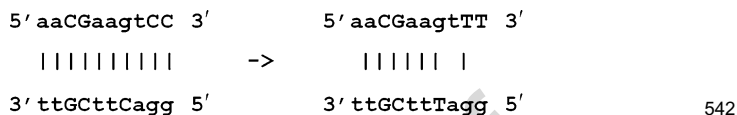
- 532 4. Additional bases can be added to primer ends using the com-
 533 mands “-5eNN” or “-3eNN,” where 5 or 3 denotes the end at

[AU5]

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

5.6 Bisulphite-Modified DNA

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



5.7 Uniqueness of Primers

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

5.8 PCR Primer Design

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

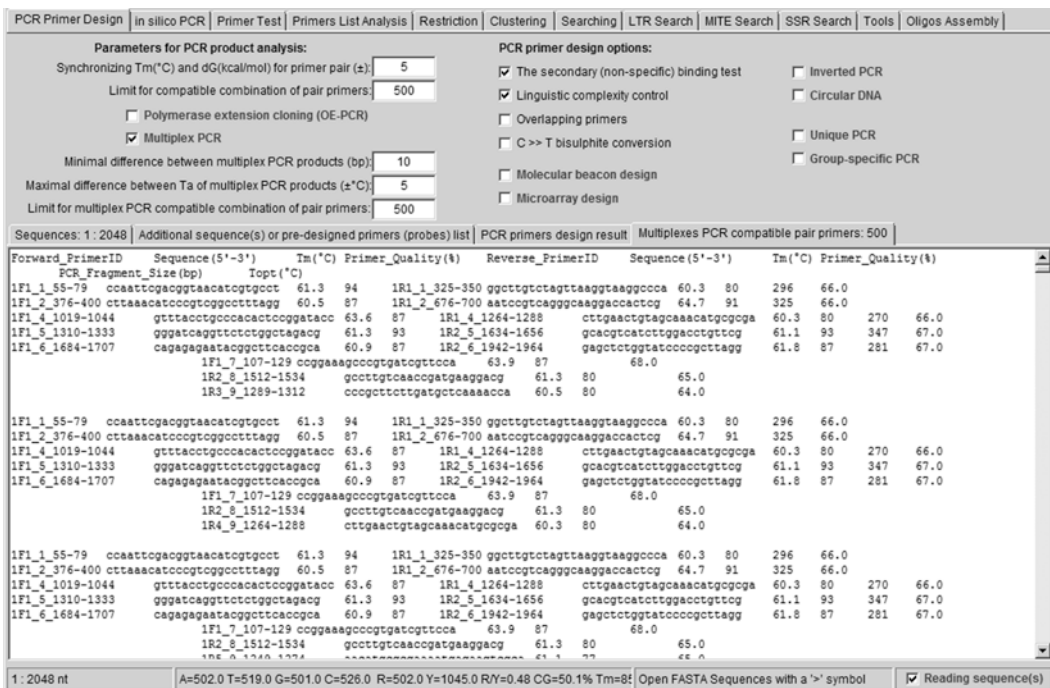


Fig. 5 An example of multiplex PCR result

575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598

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

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

Table 3
Program output of the primer design

Primer ID	Sequence (5'-3')	Length (nt)	T _m (°C)	dG, kcal/mol	T _m 3'end (°C)	CG %	LC	PQ
3.41F1_1_6-27	ggcggatcacttgaggtcagg	22	63.0	-29.7	40.0	63.6	88	87
3.51F2_1_107-132	ggcaggagaatcaactcaactggga	26	62.5	-33.6	40.2	53.8	89	89
3.61F3_1_133-154	ggcggagggtgcagtgaaacca	22	64.4	-31.3	43.9	63.6	90	90
3.71F4_1_153-175	gagaccggcactgcaactccag	23	67.2	-33.7	42.8	69.6	76	76
3.81F5_1_166-187	tgcaactcagcctgggcgacag	22	66.3	-32.1	49.5	68.2	85	85
3.91F6_1_176-197	cctggcgacagccttgagactc	22	64.8	-30.8	41.4	68.2	80	80
3.1dF1_2_859-880	cttacggaggccgagatgggca	22	63.4	-30.6	46.9	63.6	88	87
3.11F2_2_913-934	fggccaaatggtgaaacctg	22	60.9	-28.9	41.6	54.5	82	80
3.1dF3_2_956-977	aattagctgggcatggggcac	22	60.9	-29.1	47.8	54.5	80	73
3.1dF4_2_1013-1036	fggagctgaaccactgcactccag	24	63.2	-32.3	42.8	58.3	79	79
3.1dR1_1_412-434	accagaagagcctgagtgggca	23	63.9	-31.7	46.7	60.9	83	83
3.1dR2_1_309-330	ccttgctcagctctggccatcc	22	63.6	-30.0	44.8	63.6	80	80
3.1dR3_1_299-320	ctctggccatcccagttcaagc	22	61.5	-28.9	42.2	59.1	88	80
3.1dR1_2_1206-1231	agatggggtttaccatgttgccca	26	64.1	-34.8	44.6	53.8	80	80
3.1dR2_2_1166-1187	acctcaggtgatccactgctt	22	61.7	-29.5	44.2	59.1	82	80
3.1dR3_2_1150-1174	cacctgcctcagcttccc aaagtgc	25	65.3	-34.2	40.6	60.0	84	84
3.2dR4_2_1132-1157	caagtctggattacaggcgtgag	26	61.0	-33.0	42.4	50.0	93	93

The separated output of a set of primer pair sequences with their theoretical PCR products

(continued)

Table 3
(continued)

Primer ID	Sequence (5'-3')	Length (nt)	T_m (°C)	dG, kcal/mol	Reverse primer ID	T_m 3' end (°C)	CG %	LC	PQ	PCR Fragment Size (bp)	T_{opt} (°C)
t3.2f Forward primer ID	Sequence (5'-3')	T_m (°C)	PQ	Reverse primer ID	Sequence (5'-3')	T_m (°C)	PQ	PQ	PCR Fragment Size (bp)	T_{opt} (°C)	
t3.23											
t3.24 F1_1_8-27	gcgatcactgaggtcagg	58.9	87	IR3_1_301-320	ctctggccatcccagttcaa	57.6	80	313	63		
t3.2d F1_1_8-27	gcgatcactgaggtcagg	58.9	87	IR4_1_291-310	cccagttcaagccatccct	60.2	76	303	65		
t3.2d F2_1_49-68	caacgtggagctaggtatgg	56.0	80	IR1_1_409-429	gaagagcctgagttggcaca	60.0	85	381	62		
t3.2f F2_1_49-68	caacgtggagctaggtatgg	56.0	80	IR3_1_301-320	ctctggccatcccagttcaa	57.6	80	272	62		
t3.2d F2_1_49-68	caacgtggagctaggtatgg	56.0	80	IR4_1_291-310	cccagttcaagccatccct	60.2	76	262	62		
t3.2d F3_1_109-130	caggagaatcacttcaactcagg	56.4	87	IR1_1_409-429	gaagagcctgagttggcaca	60.0	85	321	62		
t3.3d F3_1_109-130	caggagaatcacttcaactcagg	56.4	87	IR3_1_301-320	ctctggccatcccagttcaa	57.6	80	212	62		
t3.3f F3_1_109-130	caggagaatcacttcaactcagg	56.4	87	IR4_1_291-310	cccagttcaagccatccct	60.2	76	202	62		
t3.3d F4_1_134-154	gaggaggttcagtgaaaccga	62.6	90	IR1_1_409-429	gaagagcctgagttggcaca	60.0	85	296	66		
t3.3f F4_1_134-154	gaggaggttcagtgaaaccga	62.6	90	IR2_1_311-333	tgtccttgcctcagctctggccat	62.3	83	200	68		
t3.3d F4_1_134-154	gaggaggttcagtgaaaccga	62.6	90	IR3_1_301-320	ctctggccatcccagttcaa	57.6	80	187	63		
t3.3d F4_1_134-154	gaggaggttcagtgaaaccga	62.6	90	IR4_1_291-310	cccagttcaagccatccct	60.2	76	177	65		
t3.3d F5_1_153-172	gagaccgcgcaactgactc	64.3	73	IR2_1_311-333	tgtccttgcctcagctctggccat	62.3	83	181	68		
t3.3f F5_1_153-172	gagaccgcgcaactgactc	64.3	73	IR4_1_291-310	cccagttcaagccatccct	60.2	76	158	65		

5.9 Multiplex PCR Primer Design

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

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

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

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

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

672 On the **PCR Primer Design** tab, the user clicks on the
673 **Multiplex PCR** option. The user then selects the limit for the
674 number of primer pairs (the default is 100), the minimal size dif-
675 ference between multiplex PCR products (the default 10 bp), and
676 the maximal difference between the T_s s of the PCR products (the
677 default is ± 5 °C). After specifying inputs and primer design options,
678 the user can execute the primer design task. Once the design of the
679 primer set is completed, the result will appear in two **Result** text
680 editors: **PCR primer design result** and **Multiplex PCR compat-**
681 **ible pair primers**. Figure 8 shows the access to the PCR primer
682 design output. The result text editor **PCR primer design result**
683 will display the individual PCR primer design data, including the
684 primer list and the compatible primer pairs for all the sequences
685 and their internal tasks. The second **Multiplex PCR compatible**
686 **pair primers** text editor collects final search results that are pre-
687 sented as a list of the sets of the compatible primer pairs for multi-
688 plex PCR.

689 **5.10 Group-Specific** 690 **PCR Primers**

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

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

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

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

5.11 Simple Sequence Repeat Locus Search and PCR Primer Design

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

[AU6]

PrimerID	Sequence (5'-3')	Length (bp)	Tm(°C)	dG(kcal/mol)	Tm_3'end(°C)	CG(%)	Linguistic_Complexity(%)	Primer_Quality(%)
F1_1_19-39	cgttatgcaaaagttccoggt	21	55.6	-25.9	41.3	47.6	92	80
F2_1_42-62	aaatgattctgagtcgggtg	21	55.3	-25.4	44.5	47.6	85	80
F3_1_51-73	ctgagtcggggaattggcaacg	23	63.3	-31.3	43.4	60.9	93	93
F4_1_81-103	tcagggtgaaacctggagcatcg	23	59.8	-29.8	41.9	52.2	93	93
F5_1_142-163	ttagagttccttgaaacctggca	22	56.3	-27.6	42.8	45.5	88	88
F6_1_153-172	tgaacctgctcctgattctg	20	55.5	-26.3	42.4	55.0	84	84
F7_1_191-213	ccatgaaagggtttaagaaaccga	23	57.1	-28.0	37.7	47.8	85	85
R1_1_1024-1044	agcagataagtcactctggct	21	55.4	-26.1	43.3	47.6	92	80
R2_1_1007-1027	cgctgtggaagtcgggtacat	21	58.1	-27.0	36.0	52.4	82	81
R3_1_996-1017	agtgggtacatgcaatgcaacg	22	57.9	-27.9	41.2	50.0	88	87
R4_1_979-998	acgcaactcagttaccgctgg	20	58.3	-26.5	43.0	55.0	95	95
R5_1_967-988	ttaccgctggtgctccggaacg	22	61.9	-29.9	43.3	59.1	88	87
R6_1_956-976	cctgtgacagcatttaacga	21	55.6	-25.8	36.6	47.6	97	82
R7_1_943-962	taacggaacctttgggcaat	20	55.1	-25.1	41.8	45.0	82	80
R8_1_922-941	tgccactctctcaagctggt	20	56.3	-25.6	37.6	50.0	84	84
R9_1_892-912	tcggttaacggaggtcttga	21	55.4	-26.0	37.9	47.6	90	85
R10_1_867-887	tcggagttttgtattggcct	21	56.9	-26.9	42.1	47.6	87	87
R11_1_834-853	gtagtagcaccattgggtc	20	58.8	-25.8	46.4	60.0	87	87
R12_1_811-831	acaattctgctctggcgag	21	56.0	-25.9	45.6	47.6	92	92

Forward_PrimerID	Sequence (5'-3')	Tm(°C)	Primer_Quality(%)	Reverse_PrimerID	Sequence (5'-3')	Tm(°C)	Primer_Quality(%)
F1_1_19-39	cgttatgcaaaagttccoggt	55.6	80	R1_1_1024-1044	agcagataagtcactctggct	55.4	80
F1_1_19-39	cgttatgcaaaagttccoggt	55.6	80	R2_1_1007-1027	cgctgtggaagtcgggtacat	58.1	81
F1_1_19-39	cgttatgcaaaagttccoggt	55.6	80	R10_1_867-887	tcggagttttgtattggcct	56.9	87
F1_1_19-39	cgttatgcaaaagttccoggt	55.6	80	R12_1_811-831	acaattctgctctggcgag	56.0	92
F2_1_42-62	aaatgattctgagtcgggtg	55.3	80	R1_1_1024-1044	agcagataagtcactctggct	55.4	80
F2_1_42-62	aaatgattctgagtcgggtg	55.3	80	R2_1_1007-1027	cgctgtggaagtcgggtacat	58.1	81
F2_1_42-62	aaatgattctgagtcgggtg	55.3	80	R3_1_996-1017	agtgggtacatgcaatgcaacg	57.9	87
F2_1_42-62	aaatgattctgagtcgggtg	55.3	80	R10_1_867-887	tcggagttttgtattggcct	56.9	87
F2_1_42-62	aaatgattctgagtcgggtg	55.3	80	R12_1_811-831	acaattctgctctggcgag	56.0	92
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R1_1_1024-1044	agcagataagtcactctggct	55.4	80
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R2_1_1007-1027	cgctgtggaagtcgggtacat	58.1	81
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R3_1_996-1017	agtgggtacatgcaatgcaacg	57.9	87
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R5_1_967-988	ttaccgctggtgctccggaacg	61.9	87
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R6_1_956-976	cctgtgacagcatttaacga	55.6	82
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R9_1_892-912	tcggttaacggaggtcttga	55.4	85
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R10_1_867-887	tcggagttttgtattggcct	56.9	87
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R11_1_834-853	gtagtagcaccattgggtc	58.8	87
F4_1_81-103	tcagggtgaaacctggagcatcg	59.8	93	R12_1_811-831	acaattctgctctggcgag	56.0	92
F5_1_142-163	ttagagttccttgaaacctggca	56.3	88	R1_1_1024-1044	agcagataagtcactctggct	55.4	80
F5_1_142-163	ttagagttccttgaaacctggca	56.3	88	R2_1_1007-1027	cgctgtggaagtcgggtacat	58.1	81

Fig. 6 An example of group-specific PCR result

742
743
744
745
746
747
748
749
750
751
752
753
754
755

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

756 **5.12 Oligonucleotide**
757 **Design for In Vitro**
758 **Long Sequence**
759 **Synthesis**
760
761
762

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

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

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

5.13 Polymerase Extension PCR for Fragment Assembly

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

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

824 Users can specify the locations for both forward and reverse
825 primers design using “[]” to bracket the region. The bracketed
826 sequences will be used by the program for designing the overlap-
827 ping primers. The program selects the overlapping area so that the
828 primers from overlapping fragments are similar in size with opti-
829 mal annealing temperatures. The program adds the required bases
830 so that the T_m of the overlap is similar to, or higher than, the T_m of
831 the initial primers. Primers are tested for dimer formation within
832 the appropriate primer pairs. The user chooses **Polymerase exten-
833 sion cloning (OE-PCR)** on the **PCR Primer Design** tab and
834 selects the limit for multiple-PCR-compatible combinations of
835 pair primers (default is 100). After specifying sequence inputs and
836 PCR primer design options, the user can execute the search task.
837 Once the design of the primer sets is complete, the result will
838 appear in two text editors: **PCR primer design result** and **PCR
839 fragments assembling compatible pair primers**. The text editor
840 **PCR primer design result** window displays the individual PCR
841 primer design data, including the primer list and the compatible
842 primer pairs for all sequences whose primers are found. The **PCR
843 fragments assembling compatible pair primers** text editor col-
844 lects the final search result and presents it as a list of sets of com-
845 patible primer pairs for individual fragment amplification and
846 assembly. Figure 7 shows a sample result visualization window.

847 **5.14 In Silico PCR**

848 Modelling the hybridization of primers to targeted annealing sites
849 is the only way to predict PCR products [7, 24, 40–44]. The last
850 10–12 bases at the 3' end of primers are important for binding
851 stability; single mismatches can reduce PCR efficiency, the effect
852 increasing with proximity to the 3' terminus. FastPCR allows
853 simultaneous testing of single primers or a set of primers designed
854 for multiplex PCR. It performs a fast, gapless alignment to test the
855 complementarity of the primers to the target sequences. For in
856 silico PCR, a quick alignment to detect primer locations on the
reference sequence is performed by analyses of both strands using

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

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

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

898 6 Primer Analyses

899 Individual and sets of primers are evaluated using FastPCR or the
900 online software. They calculate primer T_m s using default or other
901 formulae for features of the primers including normal and degen-
902 erate nucleotide combinations, CG content, extinction coeffi-
903 cient, unit conversion (nmol per OD), mass (μg per OD),
904 molecular weight, and linguistic complexity and consider primer
905 PCR efficiency. Users can select either DNA or RNA primers
906 (online: PrimerAnalyser, [http://primerdigital.com/tools/
907 PrimerAnalyser.html](http://primerdigital.com/tools/PrimerAnalyser.html)) with normal or degenerate oligonucleotides
908 or modifications with various labels (for example, inosine, uridine,
909 or fluorescent dyes). Tools allow the choice of other nearest
910 neighbor thermodynamic parameters or non-thermodynamic T_m
911 calculation formulae.

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

PCR Primer Design	In silico PCR	Primer Test	Primers List Analysis	Restriction	Clustering	Searching	LTR Search	MITE Search	SSR Search	Tools	Oligos Assembly
Type or paste (Ctrl-V) sequence of oligo with universal degenerate DNA code (5'-3'):											
catagcatggataataaaciattatc											
Tm of oligo calculation parameters:											
K+ (or Na+, NH4+, Tris+) total concentration (mM):	50	Amount of oligo:	OD (A260):	1	Target concentration (µM):	100					
Mg2+ concentration (mM):	0	Mass (µg):									
Oligo concentration (µM):	0.25	nmol:									
General Sequence(s) Additional sequence(s) or pre-designed primers (probes) list Oligo analysis											
5'-catagcatggataataaaciattatc-3'											
Length=26 A=11.0 G=3.0 T=7.0 C=4.0 CG=26.9%											
Linguistic complexity = 82%											
Primer's PCR efficiency = 25%											
dG = -25.0 kcal/mol dH = -184.0 kcal/mol dS = -512.5 cal/K mol											
Tm = 48.7°C (Allawi's thermodynamics parameters (Biochemistry,1997, 36:10581-10594)											
Tm = 52.6°C (Tm = 77.1 + 11.7Log[K+] + (41(G + C) - 528)/L)											
Tm = 50.1°C (Tm = 64.9 + 41(G + C - 16.4)/L)											
Extinction coefficient = 238000 L/(mol·cm)											
Molecular weight = 7971 g/mol											
OD260 = 1.000											
µg = 33.492											
nmol = 4.202											
100µM = dissolve in 42.0 µl of MQ-water or TE buffer											
Dimer:											
<pre> <-ctattaicaaataataggtacgatac-5 5-catagcatggataataaaciattatc-> Tm=22.5°C </pre>											

Fig. 9 Example result of the oligonucleotide analysis

926 7 Availability

927 The FastPCR software is available for download at [http://](http://primerdigital.com/fastpcr.html)
 928 primerdigital.com/fastpcr.html; the online version is available at
 929 <http://primerdigital.com/tools/pcr.html>. The program manual,
 930 licence agreement, and installation files can be found at [http://](http://primerdigital.com/fastpcr/)
 931 primerdigital.com/fastpcr/. YouTube tutorial videos have been
 932 placed at <http://www.youtube.com/user/primerdigital>. Web tools
 933 are accessible at <http://primerdigital.com/tools/>.

934 Acknowledgments

935 Web tools are available free to academic institutions, provided that
 936 they are used for noncommercial research and education only.
 937 They may not be reproduced or distributed for commercial use.
 938 This work was partially supported by the companies PrimerDigital
 939 Ltd. and Oligomer Ltd. and by the Academy of Finland, Project
 940 134079.

References

941

- 942 1. Untergasser A et al (2012) Primer3—new 996
943 capabilities and interfaces. *Nucleic Acids Res* 997
944 40:e115. doi:10.1093/nar/gks596 998
945 2. Kalendar R, Lee D, Schulman AH (2009) 999
946 FastPCR software for PCR primer and probe 1000
947 design and repeat search. *Genes, Genomes and* 1001
948 *Genomics* 3:1–14 1002
949 3. Kalendar R, Lee D, Schulman AH (2011) Java 1003
950 web tools for PCR, *in silico* PCR, and oligo- 1004
951 nucleotide assembly and analysis. *Genomics* 1005
952 98:137–144 1006
953 4. Marshall OJ (2004) PerlPrimer: cross- 1007
954 platform, graphical primer design for standard, 1008
955 bisulphite and real-time PCR. *Bioinformatics* 1009
956 20:2471–2472 1010
957 5. Owczarzy R et al (2008) IDT SciTools: a suite 1011
958 for analysis and design of nucleic acid oligo- 1012
959 mers. *Nucleic Acids Res* 36:W163–9. 1013
960 doi:10.1093/nar/gkn198 1014
961 6. Bekaert M, Teeling EC (2008) UniPrime: a 1015
962 workflow-based platform for improved univer- 1016
963 sal primer design. *Nucleic Acids Res* 36:e56. 1017
964 doi:10.1093/nar/gkn191 1018
965 7. Ye J et al (2012) Primer-BLAST: a tool to 1019
966 design target-specific primers for polymerase 1020
967 chain reaction. *BMC Bioinformatics* 13:134. 1021
968 doi:10.1186/1471-2105-13-134 1022
969 8. Giegerich R, Meyer F, Schleiermacher C 1023
970 (1996) GeneFisher—software support for the 1024
971 detection of postulated genes. *Proc Int Conf* 1025
972 *Intell Syst Mol Biol* 4:68–77 1026
973 9. Gadberry MD et al (2005) Primaclade—a flex- 1027
974 ible tool to find conserved PCR primers across 1028
975 multiple species. *Bioinformatics* 21:1263– 1029
976 1264. doi:10.1093/bioinformatics/bti134 1030
977 10. National Center for Biotechnology 1031
978 Information, National Library of Medicine, 1032
979 Building 38A, Bethesda, MD, 20894. [http://](http://blast.ncbi.nlm.nih.gov/blastcgihelp.shtml) 1033
980 blast.ncbi.nlm.nih.gov/blastcgihelp.shtml 1034
981 11. Nomenclature for incompletely specified bases 1035
982 in nucleic acid sequences (1984) [http://www.](http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html) 1036
983 [chem.qmul.ac.uk/iubmb/misc/naseq.html](http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html) 1037
984 12. Allawi HT, SantaLucia J Jr (1997) 1038
985 Thermodynamics and NMR of internal G.T 1039
986 mismatches in DNA. *Biochemistry* 36:10581– 1040
987 10594. doi:10.1021/bi962590c 1041
988 13. SantaLucia J (1998) A unified view of poly- 1042
989 mer, dumbbell, and oligonucleotide DNA 1043
990 nearest-neighbor thermodynamics. *Proc Natl* 1044
991 *Acad Sci USA* 95:1460–1465 1045
992 14. Le Novère N (2001) MELTING, computing 1046
993 the melting temperature of nucleic acid duplex. 1047
994 *Bioinformatics* 17:1226–1227. doi:10.1093/ 1048
995 bioinformatics/17.12.1226 1049
15. Bolton ET, McCarthy BJ (1962) A general 996
method for the isolation of RNA complemen- 997
tary to DNA. *Proc Natl Acad Sci USA* 48: 998
1390–1397 999
16. Guedin A et al (2010) How long is too long? 1000
Effects of loop size on G-quadruplex stability. 1001
Nucleic Acids Res 38:7858–68. doi:10.1093/ 1002
nar/gkq639 1003
17. Wallace RB et al (1979) Hybridization of syn- 1004
thetic oligodeoxyribonucleotides to Φ X 174 1005
DNA: the effect of single base pair mismatch. 1006
Nucleic Acids Res 6:3543–57. doi:10.1093/ 1007
nar/6.11.3543 1008
18. von Ahsen N, Wittwer CT, Schutz E (2001) 1009
Oligonucleotide melting temperatures under 1010
PCR conditions: nearest-neighbor corrections 1011
for Mg^{2+} , deoxynucleotide triphosphate, and 1012
dimethyl sulfoxide concentrations with com- 1013
parison to alternative empirical formulas. *Clin* 1014
Chem 47:1956–1961 1015
19. Kypr J et al (2009) Circular dichroism and 1016
conformational polymorphism of DNA. 1017
Nucleic Acids Res 37:1713–25. doi:10.1093/ 1018
nar/gkp026 1019
20. Gabrielian A, Bolshoy A (1999) Sequence 1020
complexity and DNA curvature. *Comput* 1021
Chem 23:263–74. doi:10.1016/S0097-8485 1022
(99)00007-8 1023
21. Orlov YL, Potapov VN (2004) Complexity: an 1024
internet resource for analysis of DNA sequence 1025
complexity. *Nucleic Acids Res* 32:W628–33. 1026
doi:10.1093/nar/gkh466 1027
22. Gilson MK et al (1997) The statistical- 1028
thermodynamic basis for computation of 1029
binding affinities: a critical review. *Biophys J* 1030
72:1047–69. doi:10.1016/S0006-3495(97) 1031
78756-3 1032
23. Peyret N et al (1999) Nearest-neighbor ther- 1033
modynamics and NMR of DNA sequences 1034
with internal A.A, C.C, G.G, and T.T mis- 1035
matches. *Biochemistry* 38:3468–3477. 1036
doi:10.1021/bi9825091 1037
24. Watkins NE Jr, SantaLucia J Jr (2005) Nearest- 1038
neighbor thermodynamics of deoxyinosine 1039
pairs in DNA duplexes. *Nucleic Acids Res* 1040
33:6258–67. doi:10.1093/nar/gki918 1041
25. Sen D, Gilbert W (1992) Guanine quartet 1042
structures. *Methods Enzymol* 211:191–199 1043
26. Il'icheva IA, Florent'ev VL (1992) Four- 1044
stranded complexes of oligonucleotides- 1045
quadruplexes. *Mol Biol (Mosk)* 26:512–531 1046
27. Shing HP (1994) The non-B-DNA structure of 1047
 $d(CA/TG)_n$ does not differ from that of Z-DNA. 1048
Proc Natl Acad Sci USA 91:9549–9553 1049

- 1050 28. SantaLucia J Jr, Hicks D (2004) The thermody- 1083
 1051 namics of DNA structural motifs. *Annu Rev* 1084
 1052 *Biophys Biomol Struct* 33:415–40. doi:10.1146/ 1085
 1053 [annurev.biophys.32.110601.141800](https://doi.org/10.1146/annurev.biophys.32.110601.141800)
- 1054 29. Williams JGK et al (1990) DNA polymor- 1086
 1055 phisms amplified by arbitrary primers are use- 1087
 1056 ful as genetic-markers. *Nucleic Acids Res* 1088
 1057 18:6531–5. doi:10.1093/nar/18.22.6531
- 1058 30. Welsh J, McClelland M (1990) Fingerprinting 1089
 1059 genomes using PCR with arbitrary primers. 1090
 1060 *Nucleic Acids Res* 18:7213–8. doi:10.1093/ 1091
 1061 [nar/18.24.7213](https://doi.org/10.1093/nar/18.24.7213)
- 1062 31. Kalendar R, Schulman A (2006) IRAP and 1092
 1063 REMAP for retrotransposon-based genotyp- 1093
 1064 ing and fingerprinting. *Nat Protoc* 1:2478–84. 1094
 1065 doi:10.1038/nprot.2006.377
- 1066 32. Chang RY, O'Donoghue LS, Bureau TE 1095
 1067 (2001) Inter-MITE polymorphisms (IMP): a 1096
 1068 high throughput transposon-based genome 1097
 1069 mapping and fingerprinting approach. *Theor* 1098
 1070 *Appl Genet* 102:773–781
- 1071 33. Nelson DL et al (1989) Alu polymerase chain 1099
 1072 reaction: a method for rapid isolation of 1100
 1073 human-specific sequences from complex 1101
 1074 DNA sources. *Proc Natl Acad Sci USA* 86: 1102
 1075 6686–6690
- 1076 34. Sinnett D et al (1990) Alumorphs—human 1103
 1077 DNA polymorphisms detected by polymerase 1104
 1078 chain reaction using Alu-specific primers. 1105
 1079 *Genomics* 7:331–334
- 1080 35. Jurka J et al (2005) Repbase update, a database of 1106
 1081 eukaryotic repetitive elements. *Cytogenet Genom* 1107
 1082 *Res* 110:462–7. doi:10.1159/000084979
36. TREP, the Triticeae Repeat Sequence Database 1108
 (2008) [http://wheat.pw.usda.gov/ITMI/](http://wheat.pw.usda.gov/ITMI/Repeats/) 1109
[Repeats/](http://wheat.pw.usda.gov/ITMI/Repeats/) 1110
37. Landegren U et al (1988) A ligase-mediated gene 1111
 detection technique. *Science* 241:1077–1080
38. Higasa K, Hayashi K (2002) Ordered catena- 1112
 tion of sequence-tagged sites and multiplexed 1113
 SNP genotyping by sequencing. *Nucleic Acids* 1114
Res 30:E11
39. Quan J, Tian J (2009) Circular polymerase 1115
 extension cloning of complex gene libraries 1116
 and pathways. *PLoS One* 4:e6441. 1117
 doi:10.1371/journal.pone.0006441
40. Cao YF et al (2005) Information theory-based 1118
 algorithm for *in silico* prediction of PCR pro- 1119
 ducts with whole genomic sequences as tem- 1120
 plates. *BMC Bioinformatics* 6:190. doi:10.1186/ 1121
[1471-2105-6-190](https://doi.org/10.1186/1471-2105-6-190)
41. Rubin E, Levy AA (1996) A mathematical 1122
 model and a computerized simulation of PCR 1123
 using complex templates. *Nucleic Acids Res* 1124
 24:3538–45. doi:10.1093/nar/24.18.3538
42. Lexa M, Valle G (2003) PRIMEX: rapid iden- 1125
 tification of oligonucleotide matches in whole 1126
 genomes. *Bioinformatics* 19:2486–2488
43. Nishigaki K et al (2000) Whole genome 1127
 sequence-enabled prediction of sequences per- 1128
 formed for random PCR products of *Escherichia* 1129
coli. *Nucleic Acids Res* 28:1879–1884
44. Rotmistrovsky K, Jang W, Schuler GD (2004) 1130
 A web server for performing electronic PCR. 1131
Nucleic Acids Res 32:W108–12. doi:10.1093/ 1132
[nar/gkh450](https://doi.org/10.1093/nar/gkh450) 1133

Author Queries

Chapter No.: 18 0002067381

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