Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

3

Web-Based Primer Design Software for Genome-Scale Genotyping by Pyrosequencing®

Steven Ringquist, Christopher Pecoraro, Ying Lu, Alexis Styche, William A. Rudert, Panayiotis V. Benos, and Massimo Trucco

Summary

Design of locus-specific primers for use during genetic analysis requires combining information from multiple sources and can be a time-consuming process when validating large numbers of assays. Data warehousing of genomic DNA sequences and genetic variations when coupled with software applications for optimizing the generation of locus-specific primers can increase the efficiency of assay development. Selection of oligonucleotide primers for PCR and Pyrosequencing® (SOP[3]) software allows user-directed queries of warehoused data collected from the human and mouse genome sequencing projects. The software automates collection of DNA sequence flanking single-nucleotide polymorphisms (SNPs) as well as the incorporation of locus-associated functional information, such as whether the SNP occurs in an exon, intron, or untranslated region. SOP[3] software accepts three types of user-directed input consisting of gene locus symbols, SNP reference sequence numbers, or chromosomal physical location. For human polymorphisms, SOP[3] incorporates haplotype, ethnicity, and SNP validation attributes. The output is a list of oligonucleotide primers recommended for Pyrosequencing-based typing of genetic variations. SOP[3] is available at the Division of Immunogenetics computational server found at http://imgen.ccbb.pitt.edu.

Key Words: Bioinformatics; genetics; genomics; genotyping; sequencing.

1. Introduction

Computer software for designing primers for use during PCR are available from a number of sources *(1–3)*. A principle use of these applications is to aid assay design when evaluating the association between phenotypical changes and the inheritance of genetic markers, such as single-nucleotide polymorphisms (SNPs) including nucleotide insertions and deletions. Commercial software is also available for generating PCR and sequencing primers and for use

during DNA sequencing. For example, the manufacturer of the Pyrosequencing®
instrument Biotage, LLC provides assay design software aimed at developing
primer sets suitable for use during Pyrosequencing. These applications provide
access to PCR and sequencing primer design algorithms, but require user input
of FASTA-formatted sequences along with identification of polymorphic resi-
dues that are to be evaluated. For genetic studies where many polymorphisms
are being investigated, such as is necessary when investigating a set of candi-
date markers for association with a disease phenotype, this requirement creates
a bottleneck during assay development. Improved approaches to assay design
that provide integrated genomic sequence, genetic variant, and functional in-
formation increase the efficiency by which genotyping assays can be devel-
oped and validated by laboratory testing *(4)*.

SOP[3] is a web-based software application that has been developed for use in
Pyrosequencing assay design and is available at http://imgen.ccbb.pitt.edu *(5)*.
The computational webserver warehouses an integrated biological database
comprised of genetic variants, as well as DNA sequence of the human and
mouse genomes, including validation of human SNPs within the HapMap com-
parative genome project. The software provides a batch analysis tool that ac-
cepts the names of as many as 100 genetic loci, 100 SNP identifiers, or a 100-kb
section of genomic DNA under investigation. The application returns a set of
primers for PCR and Pyrosequencing for all corresponding genetic variants
listed in dbSNP along with flanking genomic DNA sequences contained within
the human and mouse genome resources. Local warehousing of genomic se-
quences, making them accessible to the application, enables the software to
provide the choice of locus-specific primer trios for PCR and Pyrosequencing
for use in rapid genotype analysis.

2. Materials

2.1. System Configuration

1. The SOP[3] application, genomic sequence, and genetic variation data from the
   human and mouse genome projects are warehoused on a computational webserver
   purchased from @Xi Computer (San Clemente, CA). The software application is
   written in preprocessor hypertext protocol v5.0.3 and is associated with a MySQL
   4.1 database developed on a Linux SuSE Enterprise Server 8 for the AMD64
   operating system with Apache web server v2.0.48. The computational server con-
   sists of dual AMD Opteron 246 64-bit processors with 1024 KB Cache, 8192 MB
   random access memory, and four 250 GB hard drives.

2.2. Connecting to the Webserver

1. The SOP[3] website can be accessed over the internet at http://imgen.ccbb.pitt.edu.
   The application works best when viewed in Internet Explorer v6 or higher but can

also be viewed in Firefox as well as Netscape. Internet browsers are available for download from http://www.microsoft.com/windows/ie, http://browser.netscape.com, and http://www.mozilla.org, respectively.

3. Methods

The SOP[3] application has been designed for use with genotyping protocols and enables locus-specific PCR amplification of candidate SNPs for sequence analysis. Filtering of genetic variations for validated and HapMap-associated SNPs improves the efficiency of assay development for studying the role of DNA polymorphisms in determining the phenotype in models of genetic inheritance (**Table 1**). An advantage of the SOP[3] application is that design of genotyping studies can be developed using the assay design algorithm by filtering all available genetic variants for those specified by user-directed query, such as specifying human or mouse and using only validated nucleotide polymorphisms (*see* **Note 1**). The application accepts multiple queries of gene loci, SNPs, as well as chromosomal regions for analysis. SOP[3] provides multiple choices for selecting the most suitable genetic markers for evaluation during genotyping. The User Guide for operation of SOP[3] software is available online at http://imgen.ccbb.pitt.edu/sop3/userguide.

3.1. Basic Functions

1. Searching human or mouse genetic variations. The web interface to the SOP[3] application allows the user to specify whether to search for SNPs within the human or mouse genome. The default setting is for human genetic variants and is indicated on the website by the filled in button adjacent to the label for "Human." To specify searching of the mouse database the user should click the button next to "Mouse" provided on the website (**Fig. 1**).

2. Searching by locus name. To query the SOP[3] database for SNPs associated with up to 100 locus symbols the user can enter the symbol or a list of locus symbols into the textbox. Locus symbols can be typed directly into the text box, entered using the copy and paste function on the user's computer, or by uploading from a text file. Queries are restricted to no more than 100 entries per batch. Use the check boxes provided to specify whether the query is for human or mouse genomes. Click the "Search" button to activate the application. Results will appear at the bottom of the page. Selecting the "Clear" button will refresh the application to the default settings.

3. Searching by SNP identifier. To query the SOP[3] database by SNP identifier the user can enter either a single reference sequence number or a list of polymorphism reference sequence numbers. Queries are restricted to no more than 100 entries per batch. Use the check boxes provided to specify whether the query is for human or mouse genomes. Click the "Search" button to activate the application. Results will appear at the bottom of the page. Selecting the "Clear" button will refresh the application to the default settings.

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

Table 1
Classification of Human and Mouse Genetic Variants

Classification	Number of human SNPs	Number of mouse SNPs	Description
Validation status			
HapMap	35,126	NA	The genetic variant has been genotyped by the International HapMap Consortium and incorporated into the dbSNP resource.
Validated	384,831	464,137	Multiple submissions of the genetic variant has been reported within the dbSNP resource.
Function type			
Coding: syn. unknown	34	9	The genetic variant is within the coding region of a gene. The location of the polymorphism cannot be resolved because of an error in the alignment of the exon.
Intron	3,704,388	213,968	The variant is in the intron of a gene but not within two residues from the intron/exon boundry.
Nonsynonymous change	61,963	5478	The genetic variant is nonsynonymous for the codon within the gene. This class of allele is defined by a substitution and translation of the allele into the codon that results in a change to the amino acid specified by the exon reading frame.
No RNA_acc / Protein_acc	357,606	30,373	The polymorphism is within 2000 nucleotides 5' or 500 nucleotides 3' of a gene feature. The variant is not in the transcript for the gene.
Splice site	971	23	The variant is in the first two or last two nucleotides on the intron of a gene.
Synonymous change	48,702	9548	The genetic variant is synonymous for the codon within the gene. This class of allele is defined by a substitution and translation of the allele into the codon that results in no change to the amino acid specified by the exon reading frame.
Untranslated region	668,282	43,414	The nucleotide polymorphism is in the transcript of a gene but not in the coding region of the mRNA.

Geographic population

Geographic population			
Cent. Asia	46	NA	Samples from Russia and its satellite republics and from nations bordering the Indian Ocean between East Asia and the Persian Gulf regions.
Cent/S. Africa	30	NA	Samples from nations south of the Equator, Madagascar, and neighboring island nations.
Cent/S. America	11	NA	Samples from Mainland Central and South America and island nations of the western Atlantic, Gulf of Mexico, and Eastern Pacific.
E. Asia	7447	NA	Samples from eastern and south eastern Mainland Asia and from Northern Pacific island nations.
Europe	535	NA	Samples from Europe north and west of Caucasus Mountains, Scandinavia, and Atlantic islands.
Multinations	13,004	NA	Samples that were designated to maximize measures of heterogeneity or sample human diversity in a global fashion.
N. America	11,438	NA	All samples north of the Tropic of Cancer, including defined samples of US caucasians, African Americans, Hispanic Americans, and the National Human Genome Research Initiative (NHGRI) polymorphism discovery resource.
N.E. Africa and Mid. East	27	NA	Samples collected from North Africa (including the Sahara desert), East Africa (south to the Equator), Levant, and the Persian Gulf.
Pacific	144	NA	Samples from Australia, New Zealand, Central and Southern Pacific Islands, and Southeast Asian peninsular/island nations.
Unknown	479	NA	Samples with unknown geographic provinces that are not global in nature.
W. Africa	174	NA	Sub-Saharan nations bordering the Atlantic north of the Congo River and central/southern Atlantic nations.

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place
Compositor: MBL
Date: 06.06.06
Revision: First Pass

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

Fig. 1. User interface to the SOP³ web-based software application. The example illustrates a search for polymorphisms associated with the human glucokinase locus, GCK. The query is limited to single-nucleotide polymorphisms linked to the HapMap project, filtered for genetic variants found in populations within North America, and for polymorphisms associated with intron elements of the RNA transcript.

AU: Is GCK referring to a gene in Figs. 1 and 2?

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

Table 2
Maximum Allowable Upper Limit for Screening Chromosome Range

Human chromosome	Maximum range	Mouse chromosome	Maximum range
1	245,442,500	1	195,198,653
2	242,818,021	2	181,685,801
3	199,450,386	3	160,571,871
4	191,400,945	4	154,132,574
5	180,837,593	5	149,217,787
6	170,972,421	6	149,553,910
7	206,556,958	7	133,031,105
8	146,272,185	8	128,674,990
9	138,428,984	9	124,140,960
10	135,412,916	10	130,567,357
11	134,451,003	11	121,607,694
12	132,389,146	12	114,933,529
13	114,127,336	13	116,456,691
14	106,360,250	14	117,011,917
15	100,337,960	15	104,102,711
16	88,821,548	16	98,800,952
17	80,652,345	17	93,538,550
18	76,116,152	18	90,878,013
19	63,806,020	19	60,667,351
20	62,435,629	X	160,070,598
21	46,943,948	Y	47,759,179
22	49,534,318		
X	154,823,225		
Y	57,700,652		

Values for maximum range were determined using the highest chromosomal position number associated with a SNP on the human or mouse genomes using build 35 or 33, respectively.

4. Searching by chromosomal location and range. To query the SOP[3] database by chromosomal range use the check boxes to specify whether the search will focus on human or mouse genomes. Next, use the "Chromosome Number" dropdown menu to specify which chromosome is to be used. Enter lower and upper limits for the nucleotide positions that should be screened. Queries are restricted to regions no greater than 100,000 nucleotides. The maximum allowable upper limit for screening a range of nucleotides on each chromosome corresponds to the length of the chromosome and is indicated in **Table 2**. Click the "Search" button to activate the application. Results will appear at the bottom of the page. Selecting the "Clear" button will refresh the application to the default settings.

5. Filtering the results to show only validated SNPs. Selection of the "Validated SNPs" search filter directs the application to screen genetic polymorphisms for inclusion in the list of validated polymorphisms provided by dbSNP. There are

TABLE 2

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

32   Uncorrected Proof Copy   *Ringquist et al.*

currently 384,831 human and 464,137 mouse polymorphisms associated with this attribute. Validation status is defined as a genetic variant that has been reported by multiple submissions, is linked to frequency data for the SNP in a population, or was submitter validated by an updated submission from the original report.

6. Filtering the results to show only HapMap-genotyped SNPs. Selection of the "HapMap" search filter directs the application to screen genetic polymorphisms for those validated polymorphisms that are also included in the International HapMap Consortium haplotyping mapping project incorporated into the dbSNP database. There are currently 35,126 human polymorphisms associated with this attribute.

7. Filtering the results by limiting the population criteria. Selection of the "Limit By Population Criteria" filter directs the application to screen genetic polymorphisms by geographic origin of the sample. Selection of more than one population filter will return those polymorphisms that meet either of the criteria (i.e., a Boolean OR operation). The number of polymorphisms associated with each population group is indicated on the application's homepage.

8. Limiting the search by function type. Genetic variants may be filtered by whether they are near a gene (locus region), in a UTR (untranslated region), in an intron (intron), or in a splice site (splice site). If the variation is in a coding region, then the functional class of the variation depends on how each allele may affect the translated peptide sequence, e.g., nonsynonymous change or synonymous change. Selection of more than one of the function-type filters will return those polymorphisms that meet either of the criteria (i.e., a Boolean OR operator). A definition of these filters as well as the number of polymorphisms associated with each function type, for human and mouse, is indicated in **Table 1**.

9. Initiating design of primers for PCR and Pyrosequencing. Selecting the "Search" button will activate the application. Locus-specific primers for PCR and Pyrosequencing will be returned for those SNPs that meet the filtering criteria selected in the user input section of the web interface.

10. Reporting the results. After submitting a search the application will present the results at the bottom frame of the website. Each report consists of four sections, attribute bar, candidate primer bar, DNA sequence flanking the polymorphism, and computer-simulated Pyrosequencing data for heterozygous and each homozygous genotype (**Fig. 2**).

11. The locus attribute bar provides information linked to the polymorphism consisting of reference sequence number, identity of the polymorphism, locus symbol, physical location on the chromosome, heterozygosity value, and function attribute. If the filters for "Validated SNPs," "HapMap," or "Limit By Population Criteria" were selected that information will also appear in the attribute bar.

12. The candidate primer bar contains the oligonucleotide sequence (written 5' to 3') for the candidate PCR forward and reverse primers as well as the Pyrosequencing primer (*see* **Note 2**). The nomenclature used for designating a forward PCR primer is that it is on the same side of the SNP as the Pyrosequencing primer. The nomenclature for designating a reverse PCR primer is that it is on the opposite side of the SNP as the Pyrosequencing primer. The reverse PCR primer should be synthesized with a 5' biotin-TEG modification during the Pyrosequencing protocol.

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
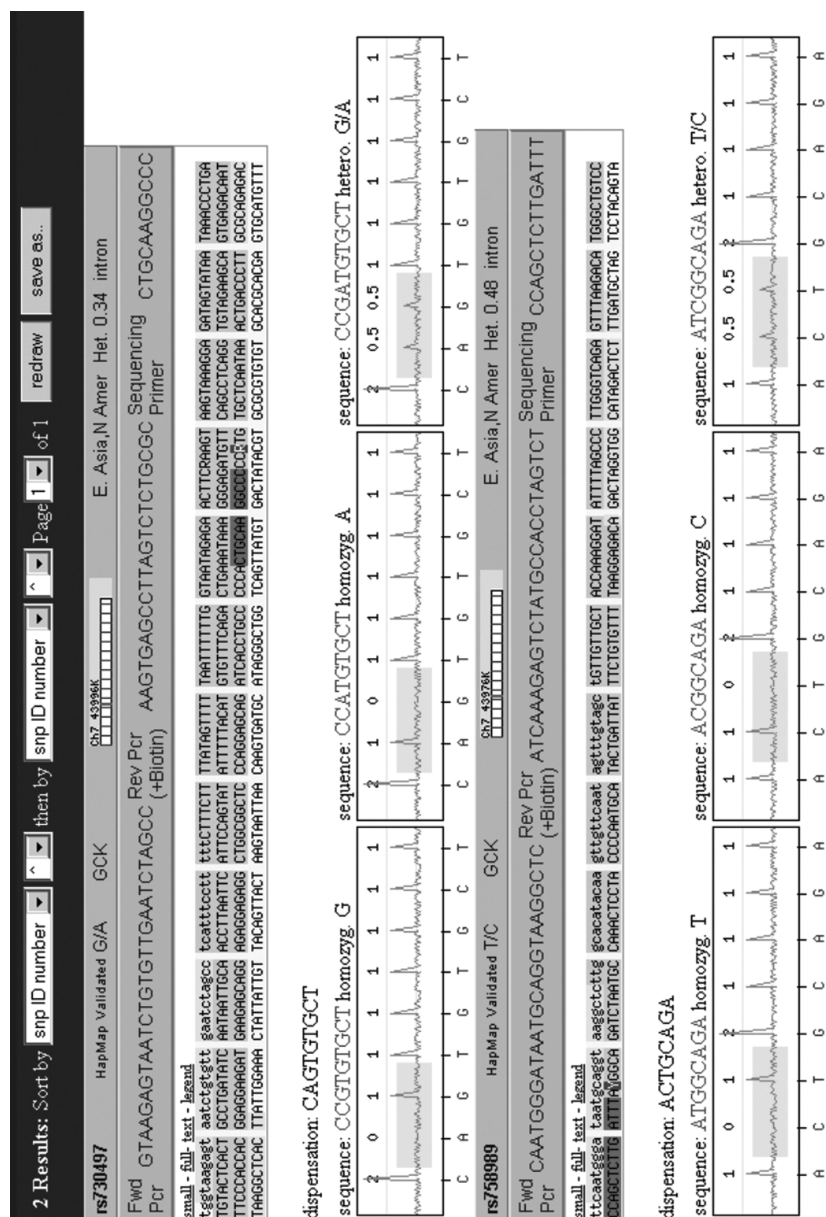Date: 06.06.06
Revision: First Pass

Fig. 2. Results of the query for single-nucleotide polymorphisms (SNPs) linked to the human glucokinase locus, GCK. There are two entries in SOP³ v2 corresponding to the query. The attribute bar, candidate primer bar, DNA sequence, and simulated Pyrosequencing® data are shown. The color code in the DNA sequence box indicates the location of the PCR primers (yellow), Pyrosequencing primer (red), region of amplified DNA (orange), and location of the SNP (blue).

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

Uncorrected Proof Copy

13. The DNA sequence flanking the SNP is indicated in the third panel of the results frame. The user may view the "Full" sequence that was evaluated, consisting of the polymorphism and 1000 flanking residues. The DNA sequence may also be viewed in FASTA format by selecting the "Text" button. The option for viewing the flanking sequence as "Small" will display only the sequence amplified by the locus-specific PCR primers (*see* **Note 3**). The color code in the bottom frame shows PCR primers indicated in yellow, Pyrosequencing primer in red, the SNP is colored blue, the amplified region in orange, and nearby polymorphisms shown in gray. Lowercase characters indicate the location of repeat masker-identified genomic regions. A "Legend" button providing a key to the color coding is also provided.

14. Simulated Pyrosequencing data is displayed in the bottom panel of each result. A simulated Pyrogram is drawn for each genotype, i.e., heterozygous and both homozygous genotypes. A nucleotide dispensation order for evaluating these data in the Pyrosequencing software is illustrated above the Pyrogram charts. Informative dispensations for distinguishing the genotypes are indicated in yellow (*see* **Note 4**).

15. Minimization and maximization of the input/output frames can be accomplished by clicking the "Minimize/Maximize" button located in the upper right-hand corner of the web interface. A full-screen view of the results of the SOP[3] application is available, allowing easier browsing of the results frame.

16. Sorting the results by reference SNP number, heterozygosity value, physical position on the chromosome, gene name, or function can be accomplished using the "Sort By" dropdown menus. The "Sort By" function is located in the horizontal bar found along the middle of the web interface. Results may be viewed in ascending "^" or descending "v" order. Results on the web interface are presented 20 at a time. When greater than 20 results are returned they can be viewed by choosing the "Page" dropdown menu. The total number of results and pages of results is also indicated. After selection of functions for sorting and page display the results can be refreshed by clicking the "Redraw" button located on the middle toolbar.

17. Results from the SOP[3] application can be saved in an Excel formatted text file. Select this function by clicking the "Save As." This will open a text box for designating a file name and a "Save" button for uploading the results to the SOP[3] server. The saved file can be viewed by selecting the "Open Saved Results" button located at the top of the page. Saved files are listed alphabetically and each file contains a tab-delimited list of PCR and sequencing primers, as well as SNP attributes such as reference sequence number, heterozygosity value, and function attribute. Selection of the "Save All" button will allow all the results to be transferred to the web server. Unselecting of the "Save All" button will allow only the page of results presently displayed in the bottom frame to be saved.

18. Selecting the "Clear" button located in the upper frame of the web interface will refresh the application to the default settings.

3.2. Advanced Options for PCR Primer Design

1. Selection of the "Advanced Options" button opens a list of PCR and sequencing primer design settings used in the assay design algorithm. This function allows

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

the user to specify customized settings for primer selection that are different from the default settings. Clicking the "Advanced Options" a second time will hide these values. Selecting the "Clear" button will refresh the application to the default settings.

2. The display for PCR primer design settings allows the user to adjust the primer selection criteria.

3. "Melting Temperature (T_m) ºC" default setting is 60ºC. T_m is calculated using the relationship ($T_m = 16.6*\log$[cation concentration] $+ 41*$[fraction of GC] $+ 81.5$, where the concentration of cation was estimated to be 0.1 *M*) described by Schildkraut and Lifson *(6)*. User-directed changes to this value instruct the application to design primers to the indicated minimum T_m value.

4. The "Minimum Foldover" filter is set at a default of five. This value reflects the minimum allowable number of contiguous base-paired residues that can occur within a candidate PCR primer. The function allows user-directed input in order to minimize the formation of potential primer secondary structure that can interfere with PCR efficiency.

5. The "Unique N'mer Maximum Length" filter default setting is six and specifies the maximum number of 3'-end residues that meet the requirement of only occurring once within 1000 nucleotides flanking either side of the polymorphism.

6. The "A/T Test" filter specifies that at least one A or T residue will occur in the last three 3' nucleotides of the PCR primer.

7. The "Residue Thresholds" filter specifies the frequency range for each A, T, C, G nucleotide residue in the candidate PCR primer. The default setting for this filter is minimum 14% and maximum 40% nucleotide composition.

8. The "Flank Length (Maximum 1000)" setting default setting is 1000 and denotes the length of DNA flanking the polymorphism that is used to screen for suitable PCR primers.

9. The filter "Use Repeat Masker" indicates whether the results of RepeatMasker (repeatmasker.org) will be used when evaluating the DNA sequence flanking the SNP. When selected, residues that are identified as being included in a repetitive region will be masked off and will not be considered when making the PCR primer, thus generating primers that will not anneal to regions of repetitive DNA. Using this option will generally decrease the number of primers found, but will better ensure that primers do not anneal to multiple sites.

10. The filter for "PCR Product Size" determines the range of length of an acceptable amplified PCR product. The default setting for this filter is minimum 200 bp and maximum 500 bp.

3.3. Advanced Options for Pyrosequencing Primer Design

1. User-directed changes to the Pyrosequencing primer design settings are accomplished using the three interactive boxes available on the lower panel of the "Advanced Options" window.

2. The "Melting Temperature (T_m) °C" default setting is 40°C. T_m is calculated using the relationship ($T_m = 16.6*\log$[cation concentration] $+ 41*$[fraction of GC] $+ 81.5$, where the concentration of cation was estimated to be 0.1 *M*) described

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

36   *Ringquist et al.*

by Schildkraut and Lifson *(6)*. User-directed changes to this value instruct the application to design each sequencing primer to the indicated minimum T_m value.

3. The filter for "T+A Percentage" evaluates the design of the candidate sequencing primer for % AT content ensuring that this value falls within the specified range. The default range is minimum 30% and maximum 65%.

4. The interactive box for designating the "Distance From SNP" provides a tool for specifying the minimum number of bases away from the SNP to initiate the design of the Pyrosequencing primer. The default range is minimum 1 and maximum 30 nucleotides. The software algorithm is designed to choose the closest Pyrosequencing primer to the SNP but with the range of distance designated by the "Distance From SNP" function.

3.4. Exploring the Website

1. General information regarding warehoused genomic and genetic variant data is indicated in the masthead of the web interface. For example, the build of dbSNP and human and mouse DNA sequences used for v2 of SOP[3] are build 123, 35, and 33, respectively. Updates to the SOP[3] database are performed twice a year and are indicated in the masthead.

2. The "Find Gene Name" button opens a link to the search page for determining the locus symbol for either human or mouse genes as specified by the user. The "Fine Gene Name" search box allows simple Boolean searches. It allows and, or, not. It uses "*" to indicate a wildcard character and quotation marks to designate a phrase. Queries are submitted using the "Submit Query" button.

3. The "Open Saved Results" button opens a link to the list of saved files available on the SOP[3] web server. Selection of a file provides a tab-delimited document consisting of the locus symbol, reference sequence number, allele, heterzygosity value, chromosome location, primer sequences, and size of expected PCR product. The tab-delimited document can be saved to the user's computer and opened in Microsoft Excel.

4. The "Print" button provides a link to the user's print command for use in obtaining a hardcopy of the results.

5. The "Upload" button opens a "Text Box," "Browse" button, "Submit" button, and "Cancel" button for uploading to the web server a text file containing a batch of locus symbols or reference sequence numbers from a file located on the user's computer. Selection of the "Browse" button prompts the user to indicate the location of a carriage return delimited text file. Selection of the "Submit" button uploads the query batch to "Search By SNP ID Or Locus Name," i.e., the application's main entry box. The application is initiated by selecting the "Search" button. Selection of the "Cancel" button removes the upload feature from the web interface and allows the user to query genomic SNPs using the basic functions available on the website.

6. The "View Validated Primers" button opens a link to the primer trios that have been made available for general use. They are listed in a tabulated format and include reference sequence, locus, forward PCR primer, reverse PCR primer, Pyrosequencing primer, and PCR product length.

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

3.5. Online User Guide

1. The "Guide Me" button opens a link to the online user guide for the SOP[3] appli-
cation. The user guide is updated along with changes to the application and the
warehoused database (*see* **Note 5**). Explanations of the choices available when
using the application are indicated in alphabetical order.
2. The "Info" buttons open information links for selected functions available on the
web interface. For example, selection of the "Info" button next to the "Limit By
Function Type" heading opens a new browser window to the online user guide
describing the use of this filter when selecting SNPs for genetic analysis.

4. Notes

1. Software applications for primer design are available from several sources. Com-
monly used web-based applications for designing PCR primers are available at
the Primer3 (http://www.primer3.org) and IDT Corporation (http://
www.idtdna.com) websites *(1)*. Both applications generate oligonucleotide se-
quences for locus-specific PCR amplification. They are limited, however, in that
they do not provide a sequencing primer for Pyrosequencing. Another applica-
tion, specific for Pyrosequencing, is Assay Design Software available commer-
cially from Biotage, LLC. This application provides access to a primer design
algorithm for locus-specific PCR and SNP-specific Pyrosequencing. The appli-
cation, however, requires user input of each SNP and DNA flanking sequence.
The Biotage Assay Design Software can be operated using a batch mode once the
sequences are arranged in a text file on the user's computer. These applications
and the SOP[3] software provide complimentary solutions for selecting robust prim-
ers for use during pyrosequencing-based typing of genetic variants.
2. Addition of a unique nucleotide motif to the 5'-end of the forward PCR primer
occurs in order to avoid formation of secondary structure in the biotinylated tem-
plate strand, which can lead to competing sequencing signal owing to self-prim-
ing during Pyrosequencing *(7–9)*. The nucleotide motif is chosen from a list of
all possible nucleotide combinations that do not occur within the DNA sequence
flanking the SNP.
3. Validating PCR primers designed by the computer application should be per-
formed by PCR amplification followed by analysis of the product with agarose
gel electrophoresis. Selection of primers assumes a standard PCR condition us-
ing an annealing temperature of 60°C. It is recommended that, when possible,
PCR product yield be examined at 1 and 2 mM $MgCl_2$ and over an annealing
temperature range from 54 to 65°C using a gradient thermal cycler instrument.
PCR primers are considered validated if they result in a single amplified DNA
product of the expected size as visualized by agarose gel electrophoresis.
4. Validating Pyrosequencing primers is accomplished by performing the
Pyrosequencing reaction using a control DNA of known quality and combined
with a set of negative control Pyrosequencing reactions. The negative control
reactions are designed to allow troubleshooting of the sequencing reaction by
analyzing the level of background signal associated with the sequencing primer,
biotinylated reverse PCR primer, and possible alternative PCR products. Recom-

Job: Marsh 645_8
Chapter: 03_Truc
Temp: MiMB_6x9_TempWS_place

Compositor: MBL
Date: 06.06.06
Revision: First Pass

38

Ringquist et al.

Uncorrected Proof Copy

mended negative control reactions are as follows: (1) PCR product without the addition of Pyrosequencing primer; (2) PCR negative control reaction with Pyrosequencing primer; (3) PCR negative control reaction without Pyrosequencing primer; (4) Biotinylated reverse PCR primer alone; (5) Pyrosequencing primer alone; (6) Biotinylated PCR primer with Pyrosequencing primer. Failure to positively genotype samples using SOP[3] designed primers for PCR and Pyrosequencing can often be the result of background signal associated with the negative control reactions. Isolation of the background signal can be helpful in redesigning assays for selection of new primer trio sets.

5. Contact information for web server administration. The SOP[3] application is maintained by the Division of Immunogenetics at the Children's Hospital of Pittsburgh. Inquiries concerning customized searches of the database or reports of errors in the application should be addressed to Steven Ringquist (email: smr73@pitt.edu).

Acknowledgments

References

1. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols,* (Krawetz, S. and Misener S., eds.), Humana Press, Totowa, NJ, pp. 365–386.

2. Varotto, C., Richly, E., Salamini, F., and Leister, D. (2001) GST-PRIME: a genome-wide primer design software for the generation of gene sequence tags. *Nucl. Acids Res.* **29,** 4373–4377.

3. Emrich, S. J., Lowe, M., and Delcher, A. L. (2003) PROBEmer: a web-based software tool for selection optimal DNA oligos. *Nucl. Acids Res.* **31,** 3746–3750.

4. Stein, L. D. (2003) Integrating biological databases. *Nat. Rev. Genet.* **4,** 337–345.

5. Ringquist, S., Pecoraro, C., Gilchrist, M. S., et al. (2005) SOP[3]v2: web-based selection of oligonucleotide primer trios for genotyping of human and mouse polymorphisms. *Nucl. Acids Res.* **33,** W548–W552.

6. Schildkraut, C. and Lifson, S. (1965) Dependence of the melting temperature of DNA on salt concentration. *Biopolymers* **3,** 195–208.

7. Ronaghi, M., Pettersson, B., Uhlen, M., and Nyren, P. (1998) PCR-introduced loop structure as primer in DNA sequencing. *Biotechniques* **25,** 876–878.

8. Ringquist, S., Alexander, A. M., Rudert, W. A., Styche, A., and Trucco, M. (2002) Pyrosequence based typing of alleles of the HLA-DQB1 gene. *Biotechniques* **33,** 166–175.

9. Utting, M. J., Hampe, J., Platzer, M., and Huse, K. (2004) Locking of 3' ends of single-stranded DNA templates for improved Pyrosequencing performance. *BioTechniques* **37,** 66–73.

Uncorrected Proof Copy