Research Article

UniPrimer: A Web-Based Primer Design Tool for Comparative Analyses of Primate Genomes

Nomin Batnyam,¹ Jimin Lee,¹ Jungnam Lee,¹ Seung Bok Hong,² Sejong Oh,¹ and Kyudong Han¹

¹ Department of Nanobiomedical Science and WCU Research Center, Dankook University, Cheonan 330-714, Republic of Korea ² Department of Clinical Laboratory Science, Juseong University, Cheongwon 363-794, Republic of Korea

Correspondence should be addressed to Sejong Oh, sejongoh@dankook.ac.kr and Kyudong Han, kyudong.han@gmail.com

Received 19 January 2012; Accepted 15 March 2012

Academic Editor: Jinchuan Xing

Copyright © 2012 Nomin Batnyam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Whole genome sequences of various primates have been released due to advanced DNA-sequencing technology. A combination of computational data mining and the polymerase chain reaction (PCR) assay to validate the data is an excellent method for conducting comparative genomics. Thus, designing primers for PCR is an essential procedure for a comparative analysis of primate genomes. Here, we developed and introduced UniPrimer for use in those studies. UniPrimer is a web-based tool that designs PCR-and DNA-sequencing primers. It compares the sequences from six different primates (human, chimpanzee, gorilla, orangutan, gibbon, and rhesus macaque) and designs primers on the conserved region across species. UniPrimer is linked to RepeatMasker, Primer3Plus, and OligoCalc softwares to produce primers with high accuracy and UCSC *In-Silico* PCR to confirm whether the designed primers work. To test the performance of UniPrimer, we designed primers on sample sequences using UniPrimer and manually designed primers for the same sequences. The comparison of the two processes showed that UniPrimer was more effective than manual work in terms of saving time and reducing errors.

1. Introduction

The field of comparative genomics has emerged as a result of several whole-genome sequencing projects. At present, six primate whole-genome sequences (human, chimpanzee, gorilla, orangutan, gibbon, and rhesus macaque) are available at the UCSC genome browser (http://www.genome.ucsc .edu/) [1–4]. Based on these data sets, several insertions/ deletions (INDELs) and coy number variations (CNVs) have been studied by comparing primate genome sequences [5– 10]. However, the computational data analysis output should be experimentally verified using the polymerase chain reaction (PCR), quantitative PCR, comparative genome hybridization array, or single nucleotide polymorphism genotyping array. Among the wet-bench methods, PCR is the most popular and easily accessible skill in molecular biology. Primer selection is very important in PCR-based systems because a specific pair of primers should amplify only a single target from a whole genome. In other words, the properties of the primers determine the specificity of PCR.

Several web-based tools for primer design such as Primer3 [11], Primer3Plus [12], PDA [13], PRIMO [14], and PrimeArray [15] have been developed and upgraded. Along with these software, OligoCalc [16] and Oligo Analysis tools (http://www.operon.com/tools/oligo-analysis-tool.aspx/) are available to calculate the molecular weight, GC content, melting temperature, intermolecular self-hybridization, and intramolecular hairpin loop formation of oligomers or primers. These web-accessible engines are particularly useful for manually selecting PCR primers and optimizing the PCR assay. Here, we introduce a novel web-based primer design tool, UniPrimer, which compares multiple primate sequences and designs primers for the conserved sequences. Before designing primers, the UniPrimer is linked to the repetitive



FIGURE 1: UniPrimer work flow chart. The flowchart represents an overall process for UniPrimer. The sequential steps are denoted as red alphabetical order and numbers. Start and End symbols, pictured as circles, indicate the "submit entry" and "receive product", respectively.

DNA annotation utility RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker/) to eliminate the building of any candidate primers containing repetitive elements. The candidate primers are also linked to OligoCalc for users to easily and rapidly access the properties of the designed primers. "UniPrimer" unites all the necessary algorithms and applications needed for designing good quality primers. Users are able to save considerate amount of time and energy that, otherwise, would have been spared on finding each of these web tools separately, submitting sequence over and over again if several primate genomes are compared and validating primers manually one by one. UniPrimer is an easy-to-use tool at hand that combines preeminent primer designing algorithms available on the internet so far and it is accessible at http://biosw.dankook.ac .kr/UniPrimer/.

2. Materials and Methods

2.1. Sources. UniPrimer incorporates the search results of popular web-based tools and software to produce output. The following is a list of programs and their brief introductions that we utilized for our study.

BLAT (see [17]). It is a popular and one of the most powerful homology search tools used to look up the location of a sequence in the genome or determine the exon structure of an mRNA. It is designed to quickly find DNA sequences of 92% and higher similarity with lengths of 40 bases or more. It also searches protein sequences of 80% and greater similarity.

BLAT's speed and sensitivity surpass other tools of its kind; this algorithm is much faster and more accurate.

RepeatMasker. (http://www.repeatmasker.org/cgi-bin/WEB-RepeatMasker/) It is a program that screens DNA sequences for repetitive elements and low complexity DNA sequences and delivers a detailed annotation of the repeats that are present in the query sequences. Currently, it is more popular than other similar programs and summarizes repetitive elements found in the primate genomic DNA sequences.

Primer3Plus (see [12]). It is a web interface for Primer3. It is a program for designing PCR primers, as well as hybridization of oligomers and sequencing primers. While designing primers, it takes into consideration many criteria, such as PCR product size, oligonucleotide melting temperature, and GC content and all these criteria are user specifiable. As a result, the user can get as accurate primer design as possible.

OligoCalc (see [16]). It is a web-based oligonucleotide properties calculator that computes single or double-stranded DNA and RNA properties, including molecular weight, solution concentration, melting temperature, estimated absorbance coefficients, self-complementarity, and hairpin loop formation.

UCSC In-Silico PCR. (http://genome.csdb.cn/cgi-bin/hg-Pcr/) searches a sequence database with a pair of PCR primers. It is fast in performance as indexing strategy is used for search.



FIGURE 2: UniPrimer input interface and main screen. (a) Main menu, (b) search result pages, (c) type of input sequence, (d) direct input of a sequence, (e) sequence upload from a file, (f) sequence position within the query sequence, (g) adding extra 5' and 3' sequences from the query, (h) genome assembly and their identity percentage, (i) Target locus default and user defined values.

TABLE 1: UniPrimer development environment.

Items	Tools
Script language	Java server pages (JSP)
User interface	HTML, jQuery
Servlet container	Apache tomcat 6.0
Browser	Chrome, Mozilla, Safari

2.2. Development Environment. Table 1 shows the development environment for UniPrimer. It was developed on Java server pages based on Apache Tomcat 6.0 and the user interface was written on HTML and jQuery. It is well compatible with recent versions of Mozilla, Safari, and Chrome browsers and is accessible from any computer that has access to the internet.

2.3. Work Flowchart of UniPrimer (see Figure 1)

User Interface. The first column depicts the user interface and options followed by columns of processing steps. Only first column contents are visible to user.

BLAT (see [17]). When user submits a sequence, UniPrimer processes BLAT (step A in Figure 1), where it searches for sequences of high similarity to the user's input. Percent identity can be either a user defined value or a default (>93% depending on the divergence of humans).

RepeatMasker. [http://www.repeatmasker.org/cgi-bin/WEB-RepeatMasker/] In the next step (step B in Figure 1), the program identifies repetitive elements in query sequences and returns their detailed annotation.

Primer3Plus (see [12]). Primers can be designed on any region of query sequence (step C-1 in Figure 1). However, if user wants to get better PCR result, he should take into account the outputs of BLAT (step C-2 in Figure 1) and RepeatMasker (step C-3 in Figure 1).

OligoCalc (see [16]). UniPrimer is able to check the selfcomplementarity of candidate primers using OligoCalc (step D in Figure 1). If the candidate primers do not contain a potential hairpin formation, 3' complementary, and all potential self-annealing sites, it returns "NONE". Thus, the

ITGIGAT TIGACA TIGICA TIGICA TIGICA GGAGAT ACTCGT CACCAG TITITT AATCTI	Result TICTCT6696C CAAG9CTGTC TGTGTCATGG ATCCCA6AAC CAGCA696AC TGCAGTGAAG TCCCA6CTCT GCTTCA6CT TAATG6AGTC	САТІІСАСІА АGIATCСІСІ АGIIACACAG СААСАGCAIA ІАТGIGAIG АGAIGAGAGIG IICTCAAGAG ICTCICIGIC	GTCATGACTG GTTGAGAAAG TCATATGTAA TGCCCCAGGAGA ATGGCTGTIG CATGCATGAG AATCCAGAAA	CTCAGACATG CCACIGIICC CTCCAGIGGG GCCCCAGGGT AICIAGGGGC ACGACGAIGC CACAGGCCCA	50 100 150 200 250 300 350
TGTGAT TTGACA TTGACA TTGCA TGGCAC AGATGC GGAGAT ACTCGT CACCAG TTTTTT AATCTT	Result TTCTCT666C CAA66CTGTC TGTGTCAT66 ATCCCA6AAC TGCA5GAAG TGCA5GAAG TCCCA6CTCT GTCTTCA6CT TAATG6AGTC	CATTICACTA AGTATCCICT AGTIACACAG CAACAGCATA IATGIGAAIG AGAAIGGIGC IICICAAGAG ICTCICIGIC	GTCATGACTG GTTGAAGAAAG TCCATATGTAA TGCCCCAGAG TCCCCAGGAGA ATGGCTGTTG CATGCATGAG AATCCAGAAA	CTCAGACATG CCACTGTICC CTCCAGIGGG GCCCCAGGGG ACCACGACGATGC CACGACGATGC	50 100 150 200 250 300 350
ed TGIGAT TIGACA TIGICA TGGCAC AGAIGC GGAGAT ACTCGT CACCAG TITITT AAICTT	TICICIGGGC CAAGGCIGIC IGIGICAIGG AICCCAGAAC CAGCAGGGAC IGCAGIGAAG ICCCAGCICI GICIICAGCI TAAIGGAGIC	CATTICACTA AGTATCCICI AGTIACACAG CAACAGCAIA IATGIGAAIG AGAAGAGAGTG IICICAAGAG IICICAAGAG	GTCATGACIG GTIGAGAAAG ICATAIGTAA IGCCCCAGGAG ICCCAGGAGA AIGGCIGTIG CATGCAIGAG AAICCAGAAA	CTCAGACATG CCACTGTTCC CTCCAGTGGG GCCCCAGGGT ATCTAGGGGC ACGACGATGC CACAGGCCCA	50 100 150 200 250 300 350
TGTGAT TTGACA TTGTCA TGGCAC AGATGC GGAGAT ACTCGT CACCAG TTTTTT AATCTT	TTCTCTGGGC CAAGGCTGTC TGTGTCAIGG ATCCCAGAAC CAGCAGGGAC TGCAGTGAAG TCCCAGCTCT GTCTTCAGCT TAATGGAGTC	CATTICACTA AGIATCCICI AGITACACAG CAACAGCATA IATGIGAAIG AGAAIGAGAGI AGATIGIIGC IICTCAAGAG ICTCICIGIC	GTCATGACTG GTTGAGAAAG TCATATGTAA TGCCCCAGAGG TCCCAGGAGA ATGGCTGTTG CATGCATGAG AATCCAGAAA	CTCAGACATG CCACTGTICC CTCCAGTGGG GCCCCAGGGG ATCTAGGGGC ACGACGATGC CACAGGCCCA	50 100 150 200 250 300 350
TGTGAT TTGACA TTGTCA TGGCAC AGATGC GGAGAT ACTCGT CACCAG TTTTTT AATCTT	TICTCIGGGC CAAGGCIGTC IGIGICAIGG AICCCAGAAC CAGCAGGGAC IGCAGIGAAG ICCCAGCICI GICTICAGCI TAAIGGAGIC	CATITCACTA AGTATCCTCT AGTIACACAG CAACAGCATA TATGTGAATG AGAAGAGATG AGATTGTTGC TICTCAAGAG ICTCTCTGTC	GTCATGACTG GTTGAGAAAG TCATATGTAA TGCCCCAGAAG TCCCAGGAGA ATGGCTGTTG CATGCATGAAG AATCCAGAAA	CTCAGACATG CCACIGITCC CTCCAGTGGG GCCCCAGGGT ATCTAGGGGC ACGACGATGC CACAGGCCCA	50 100 150 200 250 300 350
TIGACA TIGTCA IGGCAC AGAIGC GGAGAT ACTCGT CACCAG TITTTT AAICTT	CAAGGCIGIC IGIGICAIGG ATCCCAGAAC CAGCAGGGAC IGCAGIGAAG ICCCAGCICI GICTICAGCI TAAIGGAGIC	AGTATCCTCT AGTTACACAG CAACAGCATA TATGTGAATG AGAAGAGATG AGATTGTTGC TTCTCAAGAG TCTCTCTGTC	GTTGAGAAAG TCATATGTAA TGCCCCAGAG TCCCAGGAGA ATGGCTGTTG CATGCATGAG AATCCAGAAA	CCACTGTTCC CTCCAGTGGG GCCCCAGGGT ATCTAGGGGC ACGACGATGC CACAGGCCCA	100 150 200 250 300 350
TIGTCA TGGCAC AGATGC GGAGAT ACTCGT CACCAG TITTTT AATCTT	IGTGTCATGG ATCCCAGAAC CAGCAGGGAC IGCAGTGAAG ICCCAGCICT GICTICAGCI TAAIGGAGIC	AGTTACACAG CAACAGCATA TATGTGAATG AGAAGAGATG AGATTGTTGC TTCTCAAGAG TCTCTCTGTC	TCATATGTAA TGCCCCAGAG TCCCAGGAGA ATGGCTGTTG CATGCATGAG AATCCAGAAA	CTCCAGTGGG GCCCCAGGGT ATCTAGGGGC ACGACGATGC CACAGGCCCA	150 200 250 300 350
TGGCAC AGATGC GGAGAT ACTCGT CACCAG TTTTTT AATCTT	ATCCCAGAAC CAGCAGGGAC TGCAGTGAAG TCCCAGCTCT GTCTTCAGCT TAATGGAGTC	CAACAGCATA TATGTGAATG AGAAGAGATG AGATTGTTGC TTCTCAAGAG TCTCTCTGTC	TGCCCCAGAGA TCCCAGGAGA ATGGCTGTTG CATGCATGAG AATCCAGAAA	GCCCCAGGGT ATCTAGGGGC ACGACGATGC CACAGGCCCA	200 250 300 350
AGATGC GGAGAT ACTCGT CACCAG TITTTT AATCTT	CAGCAGGGAC IGCAGTGAAG ICCCAGCTCT GTCTTCAGCT TAATGGAGTC	TATGTGAATG AGAAGAGATG AGATTGTTGC TTCTCAAGAG TCTCTCTGTC	TCCCAGGAGA ATGGCTGTTG CATGCATG <mark>A</mark> G AATCCAGAAA	ATCTAGGGGC ACGACGATGC CACAGGCC <mark>C</mark> A	250 300 350
GGA <mark>G</mark> AT ACTCGT CACCAG TTTTTT AATCTT	TGCAGTGAAG TCCCAGCTCT GTCTTCAGCT TAATGGAGTC	AGAAGAGATG AGATTGITGC TTCTCAAGAG TCTCTCIGTC	ATGGCTGTTG CATGCATG <mark>A</mark> G AATCCAGAAA	ACGACGATGC CACAGGCC <mark>C</mark> A	300 350
ACT <mark>C</mark> GT CACCAG TTTTTT AATCTT	TCCCAGCTCT GTCTTCAGCT TAATGGAGTC	AGATTGTTGC TTCTCAAGAG TCTCTCIGTC	CATGCATG <mark>A</mark> G AATCCAGAAA	CACAGGCCCA	350
CACCAG TTTTTT AATCTT	GTCTTCAGCT TAATGGAGTC	TTCTCAAGAG TCTCTCIGIC	AATCCAGAAA		
TTTTTT AATCTT	TAATGGAGTC	TCTCTCTGTC		CCAAGATACA	400
AATCTT			ACCCAGGCTG	GAGTGCAATG	450
	GGCTTACTGC	AACCTCTGTC	TCCCAGTTTC	AAGC G ATTGT	500
CCTCAG	CCTCCCAAGT	AGCTGGAATT	ACAGGCATGT	GCCACCAAAC	550
CTAATT	TTTGTATTTT	TAGTAGAGGC	GGGGTTTCAC	CATGTITGCC	600
TGGTCT	T GAACT C CTG	A T CT T AG G TG	ATCCACCTGC	CTCGGCCTCC	650
GTGCTA	GGATTACAGG	CATGAGCCAC	CATGCCTGGC	CCAGATACCA	700
ACTTAT	GTGAAAACTT	CTAAGTTTTA	AGTGGCAAGC	TCTTCATTCA	750
ааасаа	AACCAAACCT	GTGATTATCA	ATGCTGTGCT	атссааасаа	800
GGCCTG	TGGGTTTTCA	GTTCTGACTT	ATGGGTTGAA	GGTGTTTCAT	850
CCGAAC	TCATGGGIGC	CATTTATCTC	TCAGGGGGTCT	TAGACCTCCT	900
GCAGTT	ATAATTTTTC	CTTTAGTAGC	TATGAAGATT	TTCTTCCATC	950
ATGGGG	CTTATGACAT	TTTGATATGG	CCATTTAGCT	GCTGTAGTGA	100
CATATC	TATTTGGTGA	CCATAAAAAA	TAATTATIGA	ACTTARARA	105
TCAAGT	GATGCCTTAA	CCCTCACTCT	ATCCCTGACT	т	
	ACTTAT AAACAA GGCCTG CCGAAC GCAGII AIGGGG CATAIC ICAAGI	САСТАКТ GTGAAAACTT АААСАА ААССАААССТ GGCCTG TGGGTTTTCA CCGAAC TCATGGGIGC GCAGIT AIAAIIIITC AIGGGG CITAIGACAI CATAIC TAIIIGGIA ICAAGI GAIGCCITAA	CTANOTTAL CTGADALACT CTANOTTTAL ALCCALACCT GTGATATCA GTGATATCA GGCCTG TGGGTTTCA GTGCATTACA GGCCTG TGGGTTTCA GTTCTGACTT CGCAGT TCATGGGTGC CATTATCA GGCGTG TGGGTTTCA GTTCTGACTT GGCGTG TGGGTTTCA GTTCTGACTT GGCGTG TGGGATATCA CTTAGTGGGC GGAGGG TATTGGTGA CCCTAAJAJAA ICAAGT GATGCCTTAA CCCTCACTCT	АСТТАТ СТАРАЛАТТ СТАЛАСТТТА АЛТОССЛАСС АЛАСАЛА АЛССАЛАССТ СТАЛАСТТТА АЛТОССТАССТ СССАЛС ТОБОТТТСА СТАЛАСТТТА АЛТОСТОТСТ СССАЛС ТСАЛСТСКА СТАЛАСТСТ АЛТОССТСАЛ ССЛАГА АЛАЛТІТІС СПІТАСІЛС ГАЛСВАЗСТІ ПІТАЛТАТС ТІЛІТАСТАС ССЛАТАЛС АЛТІТІВСАТІ ССЛІТАЛАСТ САЛТАГ АЛТІТВСЯГА ССЛІТАЛАЛА ТАЛІТАТІВА ІСАЛБІ БАЛЕССТІЛА СССТСАСТСІ АЛСССІБАСІ	АСТТАТ СТАРАЛАСТТ СТАЛАГТТТА АЛТОССЛАСС ТСТГСАТТСА АЛАСАА АЛССТАСТ СТАЛАГТТТА АЛТОССЛАСС ТСТГСАТТСА АЛАСАА АЛССАЛАССТ СТАЛАГТТТА АЛТОССЛАССА АТССАЛАСАА ОССЛАСТ ТОБОТТТСА СТСТАЛСТТ АЛТОСОТОСТ АЛССАЛАСАА СОСЛАС ТСАЛБОБСТС СЛІТАЛСТС ТСЛОВОБСТС ТАВАССТССТ АЛВОВО СТТАЛБОСАТ СТІЛАБТАСС ТАГСАЛАБТА ТСГГСАЛС АЛТОС АЛТІТОСТАЛС СЛІТАЛАЛА ТАЛІТАЛТАВ ССАЛТАГС АЛТІТОВТАС СССТАЛАЛАЛ ТАЛІТАЛТАВ

FIGURE 3: BLAT search results. Matching and mismatching bases are in black and red, respectively.

user is able to choose primers with good quality for the PCR assay.

UCSC In-Silico PCR. [http://genome.csdb.cn/cgi-bin/hgPcr/] a final step (step E in Figure 1) is that UniPrimer searches for a sequence with a pair of PCR primers using UCSC In-Silico PCR. When successful, it returns a primer pair with the sequence lying between them. A more detailed review of the above steps is included in UniPrimer interface section.

3. Results and Discussion

3.1. UniPrimer Interface

3.1.1. Input. The screenshot of the input interface for UniPrimer is shown on Figure 2. The main menu is located on the top right corner of the figure (Figure 2a) where the user can find information about tools, contacts, and a tutorial. The topmost tabs (Figure 2b) in the main screen are pages that will contain search results for each step after a query sequence is submitted. After selecting the type of query sequence and its assembly (Figure 2c), the user pastes the target sequence directly into text area (Figure 2d) or uploads it from a file (Figure 2e). The position (Figure 2f) option can be used if the user wants to limit the query to a specific chromosome or region. In contrast, it is also possible to add extra sequences at the 5' and 3' ends of the query sequence (Figure 2g). Next go the genome types, their assembly versions, and identity percentages (Figure 2h)

where up to six-genome assembly similarities are searched at once. The last option in the input form is selecting the target locus (Figure 2i) in which user can select either default values (+500, -500) or define the positions that the PCR product should contain.

3.1.2. BLAT Search. BLAT step allows a user to find orthologous between a query sequence and several other genomes. A BLAT search result is shown in Figure 3. Identical bases between the query sequence and its corresponding sequences from other primates are colored black but mismatches between sequences are all combined and marked as red letters. However, below it, user can find links to a detailed comparison of a query sequence with each of selected genomes, separately.

It is possible to design primers on mismatching areas, but 100% conserved sequence among multiple species gives the best PCR result. Notably, UniPrimer's BLAT search compares the sequences from more than two species at the same time. As such, this approach is particularly helpful for users conducting PCR assay with more than two species.

3.1.3. RepeatMasker. RepeatMasker helps to eliminate the building of any candidate primer containing repetitive elements by masking them. Primers that are built on unmasked areas are more successful than those built on masked areas. The output result of RepeatMasker (Figure 4) returns a detailed annotation of repeat elements that are present in the query sequence. The annotation is shown in two formats; a table format contains a list of the repeats (Figure 4a), and repeat sequences are highlighted with light green color in the other format (Figure 4b). On the right side of the window, the user can find options to change the target locus and the type of primer (Figure 4c). After selecting the options, the user obtains a list of designed primers by clicking "Pick primer" on the top right corner of the main screen (step C-3 in Figure 1). Under the masked sequence, there are separate tables that describe masked areas of genomes, separately.

Details about the table and options are listed below.

How to read the table result in Figure 4a:

This section shows the output of RepeatMasker (http:// www.repeatmasker.org/cgi-bin/WEBRepeatMasker/).

SW score: Smith-Waterman score of the match.

Perc. div.: percentage of substitutions in the matching region compared to the consensus.

Perc. del.: percentage of bases opposite a gap in the query sequence (deleted bp).

Perc. ins.: percentage of bases opposite a gap in the repeat consensus (inserted bp).

Query sequence: name of input sequence.

Position in query: beginning and ending are starting and ending positions for the match in the query sequence respectively. (left) indicates the number of bases in query sequence past the end position of the match.

Input	Blat	Search	Rep	eatMask	er Pick Pri	mer Result									Pick	Primer
►Re	peat	Maske	er Resu	ilt							(a)	► Options				(c)
sw	perc	perc p	erc que	ry	posit	ion in query m	atching rep	eat	р	ositio	on in repeat	Target Locus: 2	00	- 900		
scor	e div.	del. i	ns. seq	uence	begin	end (left) re	epeat cla	ss/fam	ily be	egin	end (left) ID	RepeatMasker				
481	24.3	4.6 1	9.2 Unn	amedSe	quence 253	400 (691) C M	ER102a DN	A/hAT-	Charlie (6	6)	275 125 1	Exclude Repe	at			
2135	12.7	0.7 0	0.0 Unn	amedSe	quence 401	691 (400) C A	uSx SIN	IE/Alu	(1	9)	293 1 2	Include Repe	at in Forwa	rd or Re	verse Pri	mer
481	22.8	5.5 1	9.8 Unn	amedSe	quence 692	823 (268) C M	ER102a DN	A/hAT-	Charlie (2	06)	124 12 1	- Perc div. >	0.0	rd and F	Roverce E	rimor
												- Perc div. >	0.0		(6461361	
1	CCTT	TGTGAT	TTCTC	TGGGC	CATTTCACTA	GTCATGACTG	CTCAGACAT	G 50	(b)							
51	TGCT	TTGAC	A CAAGO	CIGIC	AGTATCCTCT	GTTGAGAAAG	CCACTGTTC	2 100								
101	CAGG	TIGIC	A TGTGI	CATGG	AGTTACACAG	TCATATGTAA	CTCCAGTGG	G 150								
151	TCAC	TGGCA	ATCCO	AGAAC	CAACAGCATA	TGCCCCAGAG	GCCCCAGGG	200								
201	GGTC	AGATGO	CAGC3	GGGAC	TATGTGAATG	TCCCAGGAGA	ATCTAGGGG	250								
251	TIGT	GGAGAI	TGCAG	TGAAG	AGAAGAGATG	ATGGCTGTTG	ACGACGATG	300								
301	AGUU	REICGI	CTCTT	GCICI	AGAIIGIIGC	CATGCATGAG	CACAGGUUU	4 350								
401	TTTT	TTTTTT	TAATO	GAGTC	TCTCTCTGTC	ANTCONGRAM	GEGEGGEEET	450								
451	GCAC	AATCTI	GGCTT	ACTGC	AACCTCTGTC	TCCCAGTTTC	AAGCGATTG	500								
501	CCTG	CCTCA	CCTCC	CAAGT	AGCTGGAATT	ACAGGCATGT	GCCACCAAA	550								
551	CTGG	CTAATI	TTTGT	TTTTA	TAGTAGAGGC	GGGGTTTCAC	CATGTTTGC	600								
601	AGGA	TGGTCI	TGAAC	TCCTG	ATCTTAGGTG	ATCCACCTGC	CTCGGCCTC	650								
651	TGAA	GTGCT	GGATI	ACAGG	CATGAGCCAC	CATGCCTGGC	CCAGATACCI	700								
701	AGGT	ACTTAI	GTGAA	AACTT	CTAAGTTTTA	AGTGGCAAGC	TCTTCATTC	750								
751	CACG	AAACAJ	AACCA	AACCT	GTGATTATCA	ATGCTGTGCT	ATCCAAACAI	800								
801	AGCA	GGCCTO	TGGGI	TTTCA	GTTCTGACTT	ATGGGTTGAA	GGTGTTTCA	850								
851	TGGC	CCGAAO	C TCATG	GGTGC	CATTTATCTC	TCAGGGGTCT	TAGACCTCC	000								
⊳F	Repe	at M	asker	of Ch	imp Sequ	<u>ience</u>										(d)
S۱	N	perc	perc	perc	query		positi	on in	query		matching	repeat	positio	n in r	epeat	
SC	ore	div.	del.	ins.	sequend	ce	begin	end	(left)		repeat	class/family	begin	end	(left)	ID
45	50	23.4	2.9	2.9	Unname	dSequenc	e 253	397	(995)	С	MER102a	DNA/hAT-Charlie	(66)	275	120	1
20	94	12.8	1.0	0.0	Unname	dSequenc	e 401	690	(702)	С	AluSx	SINE/Alu	(18)	294	2	2
21	90	12.2	2.0	0.0	Unname	dSequenc	e 691	992	(400)	С	AluSx	SINE/Alu	(4)	308	1	3
⊳F	Repe	eat Ma	asker	of Ora	angutan S	Sequence										
S١	N	perc	perc	perc	query		positi	on in	query		matching	repeat	positio	on in	repea	t
SC	оге	div.	del.	ins.	sequenc	ce	begin	end	(left)		repeat	class/family	begin	end	(left)	ID
42	26	25.9	4.1	3.6	Unname	dSequenc	e 203	397	(1291) C	MER102a	DNA/hAT-Charlie	(14)	327	120	1
22	200	11.0	0.7	0.0	Unname	dSequenc	e 401	690	(998)	С	AluSx	SINE/Alu	(19)	293	2	2

FIGURE 4: RepeatMasker search result. (a) Detailed annotation of repeats in a table format. (b) Sequence format with masked repeats; masked areas are highlighted in light green. (c) Extra options for selecting the target locus and primer type. (d) Separate tables summarize repetitive elements found in each of selected primate genomes (orthologous region).

Matching repeat: name of the matching interspersed repeat.

Repeat class/family: the class of the repeat.

Position in repeat: beginning and end are starting and ending positions for the match in the database sequence. (left) indicates the number of bases in the repeat consensus sequence prior to beginning the match.

How to adjust options in Figure 4c:

Target locus: the minimum region of PCR amplification.

Exclude repeat: avoid designing primers for repeat sequences.

Include repeat in forward primer or reverse primer: one could include repeat sequences for either forward or reverse primers.

Include repeat in forward primer and reverse primer: could include repeat sequences for both the forward and reverse primers. Additionally, the user can set the minimum perc div. of repetitive elements that belong to the primer. Higher perc div. leads to better results.

3.1.4. Pick Primer. Pick primer is linked to Primer3Plus, which selects the primer pairs that fit best to the selected parameters and orders them by quality. The Pick primer output is shown in Figure 5. The left side table (Figure 5a) will contain detailed information of all possible primers present in the query sequence, as well as the whole sequence with BLAT and RepeatMasker output results shown together; the left primers are highlighted in purple, the right primers are highlighted in yellow, the BLAT mismatches are marked in red, and the masked repeats are highlighted in light green.

Similar to the RepeatMasker step, if the user wants more specified primers he can recalculate the Pick primer step after customizing with additional options (Figure 5b) on the right-hand side of the page.

How to read a table result (Figure 5a):

Primer seq.: LEFT and RIGHT primers are highlighted in purple and yellow colors, respectively.

ut	Blat Search	RepeatMasker	Pick Primer	Result						Pick
Prin	ner Result									►Options
nere a	are 5-Primers, fo	or SAMPLE SEQ	JENCE.						(a)	-Number to return: 5 (b
		Primer Seq.	Start	Length	TM	GC	ANY	SELF]`´	-Max 3' Stability: 9.0
LEFT	GCACA	TCCCAGAACCAAC	IG 157	20	61.5	55.0	2.0	1.0	1	-Max Repeat Mispriming: 12.0
RIGH	T CCCCA	TGAGAGATGGAAGI	A 950	20	60.0	50.0	4.0	0.0	1	-Pair Max Repeat Mispriming: 24.0
		Product Size:	804 / Pair Any:	4.0 / Pair E	nd: 0.0				1	-Max Template Mispriming: 12.0
1	CCTTTGTGAT	TICTCTGGGC	CATTTCACTA	GTCATGA	CTG (CTCAGAC	'A T G	50	1	-Pair Max Template Mispriming. 24.0
51	TGCTTTGACA	CAAGGCTGTC	AGT <mark>ATC</mark> CICT	GTTGAGA	AAG (CCACTGI	TCC	100		-Primer size -Primer TM
101	CAGGTTGTCA	TGTGTCATGG	AGTTACACAG	TCATATG	TAA (CTCCAGT	GGG	150		Min: 18 Min: 57.0
151	TCACTGGCAC	ATCCCAGAAC	CAACAGCATA	TGCCCCA	GAG (GCCCCAG	GGT	200		Opt 20 Opt 60.0
201	GGTCAGATGO	CAGCAGGGAC	TATGTGAATG	TCCCAGG	AGA J	ATCTAGG	GGC	250		Max: 27 Max: 63.0
251	TT <mark>GTGGAGAT</mark>	TGCAGTGAAG	AG <mark>A</mark> AGAGAT <mark>G</mark>	ATGGCTG	TG	ACGACGA	TGC	300		-Product TM -GC%
301	AGCCACTCGI	TCCCAGCTCT	AGATTGTIGC	CATGCAT	GAG	CACAGGO	CCA	350		Opt 0.0 Opt 0.0
351	GTGTCACCAG	GTCTTCAGCT	TTCT <mark>C</mark> AAGAG	AATCCAG	AAA	CCAAGAT	ACA	400		Max 0.0 Max 80.0
401	TTTTTTTTTT	TAATGGAGTC	TCTCTGTC	ACCCAGG	CIG	GAGTGCA	ATG	450		max. 0.0 max. 00.0
451	GCACAATCTI	GGCTTACTGC	AACCTCTGTC	TCCCAGT	TTC	AAGC <mark>G</mark> AT	TGT	500		Mispriming/Repeat Library:
501	CCTGCCTCAG	CCTCCCAAGT	AGCTGGAATT	ACAGGCA	IGT	GCCACCA	AAC	550		NONE
551	CTGGCTAATI	TITGTATTTT	TAGTAGA66 <mark>C</mark>	GG G GTTT(CAC	C <mark>A</mark> TG T TT	GCC	600		
601	AGGATGGTCI	TGAACTCCTG	ATCTTAG6TG	ATCCACC	rgc (CTCGGCC	TCC	650		Mismatch base Range(begin with 5'): 0
651	TGAAGTGCTA	GGATTACAGG	CATGAGCCAC	CATGCCT	GC	CCAGATA	CCA	700		The number of mismatch base: 0
701	AGGTACTTAT	GTGAAAACTT	CTAAGTTTTA	AGTGGCA	AGC 1	ICTTCAT	TCA	750		Save
751	CACGAAACAA	AACCAAACCT	GTGATTATCA	ATGCTGT	SCT J	АТССААА	CAA	800		
801	AGCAGGCCTG	G TGGGTTTTCA	GTTCTGACTT	ATGGGTT	GAA (GGTGTTT	CAT	850		
851	TGGCCCGAAC	TCATGGGTGC	CATTTATCTC	TCAGGG	ICT 1	TAGACCT	CCT	900		
901	GGGAGCAGTI	ATAATTTTTC	CTTTAGTAGC	TATGAAG.	ATT	ITCTTCC	ATC	950		
951	TCTCATGGGG	CTTATGACAT	TTTGATATGG	CCATTTA	GCT (GCTGTAG	TGA	1000		
100	1 TATTCATATC	TATTTGGTGA	ССАТАААААА	TAATTAT	IGA J	ACTTAAA	AAA	1050		
105	1 AAAGTCAAGT	GATGCCTTAA	CCCTCACTCT	ATCCCTG.	ACT	г				

FIGURE 5: Pick primer result. (a) Table showing primers, their detailed information, and position in the sequence. (b) Additional options for the Pick primer step.

put	Blat Sear	ch RepeatMaske	r Pick Primer	Result		
► Se	quence N	Name: SAMPLE_	SEQUENCE			
		Primer		Melting to	emp.	Self Comp.
	Left	GCACATCCCAG	ACCAACAG	61.5 (C	Oligo Calc
F	Right	CCCCATGAGAG	60.0 0	2	Oligo Calc	
			Product			
Ge	enome	Locatio	on	Lengt	h	PCR Result
Н	uman	chr15:46028034	-46028837	804 b	р	Product
С	himp	chr15:42885154	-42886258	1105 b	p	Product
G	orilla	nuíí		qd 0		
Ora	angutan	chr15:42197101	-42198501	1401 b	р	Product
G	ibbon	null		0 bp		
R	hesus	null		0 bp		

FIGURE 6: Final primer result.

Start: the position of the 5' base of the primer. For left primers it is the position of the leftmost base and for right primers it is the rightmost base.

Length: length of the primer.

TM: melting temperature of the primer.

GC: the percent of G and C bases in the primer.

ANY: self-complementary score of the primer.

SELF: 3' self-complementary of the primer.

How to adjust options for Pick primer (Figure 5b): This option is followed by Primer3 [11].

Number to return: the maximum number of primer pairs to return. Setting this parameter to a large value will increase running time.



FIGURE 7: Oligo self-complementary check window.

Max 3' stability: the maximum stability for the last five 3' bases of a left or right primer. Larger numbers mean more stable 3' ends.

Max Repeat Mispriming: the maximum allowed weighted similarity with any sequence in the Mispriming Library. Default is 12.

Pair Max Repeat Mispriming: the maximum allowed sum of similarities of a primer pair (one similarity for each primer) with any single sequence in the Mispriming Library.

Max Template Mispriming: the maximum allowed similarity to ectopic sites in the sequence from which you are designing the primers.

Pair Max Template Mispriming: the maximum allowed summed similarity of both primers to ectopic

biosw.dankook.ac.k	a/primer/target	Pop.jsp?genome	=Human&left=	CACTGGCACAT	CCCAG
		Product	t		
	genome:	Human			
	left primer:	CACTGGCACA	TCCCAGAAC		
	right prime	r:CCCCATGAGA	GATGGAAGAA		
1 CACTGGCACA	TCCCAGAACc	aacagcatat	gccccagagg	ccccagggtg	50
51 gtcagatgcc	agcagggact	atgtgaatgt	cccaggagaa	tctaggggct	100
101 tgtggagatt	gcagtgaaga	gaagagatga	tggctgttga	cgacgatgca	150
151 gccactcgtt	cccagctcta	gattgttgcc	atgcatgagc	acaggcccag	200
201 tgtcaccagg	tcttcagctt	tctcaagaga	atccagaaac	caagatacat	250
251 tttttttt	aatggagtct	ctctctgtca	cccaggctgg	agtgcaatgg	300
301 cacaatcttg	gcttactgca	acctctgtct	cccagtttca	agcgattgtc	350
351 ctgcctcagc	ctcccaagta	gctggaatta	caggcatgtg	ccaccaaacc	400
401 tggctaattt	ttgtatttt	agtagaggcg	gggtttcacc	atgtttgcca	450
451 ggatggtctt	gaactcctga	tcttaggtga	tccacctgcc	teggeeteet	500
501 gaagtgctag	gattacaggc	atgagccacc	atgcctggcc	cagataccaa	550
551 ggtacttatg	tgaaaacttc	taagttttaa	gtggcaagct	cttcattcac	600
601 acgaaacaaa	accaaacctg	tgattatcaa	tgctgtgcta	tccaaacaaa	650
651 gcaggcctgt	gggttttcag	ttctgactta	tgggttgaag	gtgtttcatt	700
701 ggcccgaact	catgggtgcc	atttatctct	caggggtctt	agacctcctg	750
751 ggagcagtta	taatttttcc	tttagtagct	atgaagattT	TCTTCCATCT	800
801 CTCATGGGG					

FIGURE 8: Polymerase chain reaction (PCR) product window.

sites in the sequence from which the primers are designed.

Primer Size: minimum, optimum, and maximum lengths of the primer oligo.

Primer TM: minimum, optimum, and maximum melting temperatures for the primer oligo in Celsius.

Product TM: minimum, optimum, and maximum melting temperatures for the amplicon.

GC %: minimum, optimum, and maximum percentages of Gs and Cs in any primer.

Mispriming/Repeat library indicates what mispriming library (if any) should be used to screen for interspersed repeats or for other sequence to avoid as a primer location.

We added a mismatch base range (begin with 5') and the number of mismatch bases options. A mismatched base at the 5' end of the primer is more tolerable than that at the 3' end to obtain a successful PCR amplification.

Mismatch base Range (begin with 5' end): it allows for a mismatched base from the 5' end.

The number of mismatch bases: it allows the number of mismatched base within the mismatched base range.

3.1.5. Final Result. The last page of UniPrimer contains the primer information picked by the user (Figure 6). The upper part of the table shows the left and right primer sequences and their melting temperatures in Celsius. Figure 7 shows a pop-up window that the user can reach by clicking the "OligoClac" button on the Figure 6 and potential hairpin, 3' complementarity, and self-annealing sites of candidate primers are calculated and displayed. The lower part of the table in Figure 6 shows the genome type, the genomic position, and the expected size of the PCR product. Nationally, the user can access UCSC *In-Silico* PCR by clicking the "Product" button on the right side of Figure 6 (Figure 8). In

TABLE 2: Comparison of time consumed with UniPrimer and manual work measured in minutes.

UniPrimer	Manual work
2.5 minutes	8 minutes

the results of *In-Silico* PCR, primers that are written in capital letters and sequences that lie between them are written in small letters (Figure 8).

4. Performance Comparison

As mentioned before, UniPrimer is constructed to compare up to six primate sequences at once and design primers on the conserved regions among them. We believe that using this tool to design PCR primers would save great deal of time and reduce possible errors because of two reasons described below. First, UniPrimer is not required to submit the sequence over and over again to match it with each of selected genomes. Second, since masked and mismatching areas in the query sequence are marked accordingly on Pick primer step (Figure 5), user is able to determine a proper region for designing good primers.

To estimate the performance and scalability, we used Uni-Primer to design PCR primers using a sample sequence that is related to *Alu* recombination-mediated deletions in the human genome [8] as a query. In addition, we designed primers for the same sequence using a manual method to test the efficiency of UniPrimer. Table 2 shows the approximate time consumed by UniPrimer and the manual method to design the primers for the sample sequence. The result indicates that UniPrimer works much faster than that of the manual method. The user needs to open each web-based tool separately for the manual method and submit a query sequence on each window. The loading time required to return a result depends on the length of the query sequence and the number of genomes that are selected by user. Any error that occurs will add time.

5. Conclusions

We developed UniPrimer, a web-based tool to design PCRand DNA-sequencing primers. This tool is able to find conserved regions across different primate species and designs primers for the region. Then, users are allowed to select various options for picking the best primer for their purpose. UniPrimer was developed to reduce the time required for designing PCR primers and the errors that occur during the process. We conducted a performance test to determine whether this tool works as intended, and the result showed that UniPrimer was easy to use and saved time and effort. In conclusion, we believe that UniPrimer could be a useful tool for comparative analyses of primate genomes.

Acknowledgment

This research was supported by the WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology (R31-10069).

References

- T. C. S. A. Consortium, "Initial sequence of the chimpanzee genome and comparison with the human genome," *Nature*, vol. 437, pp. 69–87, 2005.
- [2] R. A. Gibbs, J. Rogers, M. G. Katze et al., "Evolutionary and biomedical insights from the rhesus macaque genome," *Science*, vol. 316, no. 5822, pp. 222–234, 2007.
- [3] E. S. Lander, L. M. Linton, B. Birren et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.
- [4] D. P. Locke, L. W. Hillier, W. C. Warren et al., "Comparative and demographic analysis of orang-utan genomes," *Nature*, vol. 469, pp. 529–533, 2011.
- [5] K. Han, J. Lee, T. J. Meyer et al., "Alu recombinationmediated structural deletions in the chimpanzee genome," *PLoS Genetics*, vol. 3, no. 10, pp. 1939–1949, 2007.
- [6] K. Han, S. K. Sen, J. Wang et al., "Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages," *Nucleic Acids Research*, vol. 33, no. 13, pp. 4040–4052, 2005.
- [7] J. Lee, R. Cordaux, K. Han et al., "Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons," *Gene*, vol. 390, no. 1-2, pp. 18–27, 2007.
- [8] S. K. Sen, K. Han, J. Wang et al., "Human genomic deletions mediated by recombination between Alu elements," *The American Journal of Human Genetics*, vol. 79, no. 1, pp. 41–53, 2006.
- [9] K. Han, J. Lee, T. J. Meyer, P. Remedios, L. Goodwin, and M. A. Batzer, "L1 recombination-associated deletions generate human genomic variation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 49, pp. 19366–19371, 2008.
- [10] R. Redon, S. Ishikawa, K. R. Fitch et al., "Global variation in copy number in the human genome," *Nature*, vol. 444, no. 7118, pp. 444–454, 2006.

- [11] S. Rozen and H. Skaletsky, "Primer3 on the WWW for general users and for biologist programmers," *Methods in Molecular Biology*, vol. 132, pp. 365–386, 2000.
- [12] A. Untergasser, H. Nijveen, X. Rao, T. Bisseling, R. Geurts, and J. A. Leunissen, "Primer3Plus, an enhanced web interface to Primer3," *Nucleic Acids Research*, vol. 35, pp. W71–W74, 2007.
- [13] S. H. Chen, C. Y. Lin, C. S. Cho, C. Z. Lo, and C. A. Hsiung, "Primer Design Assistant (PDA): a web-based primer design tool," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3751–3754, 2003.
- [14] P. Li, K. C. Kupfer, C. J. Davies, D. Burbee, G. A. Evans, and H. R. Garner, "PRIMO: a primer design program that applies base quality statistics for automated large-scale DNA sequencing," *Genomics*, vol. 40, no. 3, pp. 476–485, 1997.
- [15] G. Raddatz, M. Dehio, T. F. Meyer, and C. Dehio, "PrimeArray: genome-scale primer design for DNA-microarray construction," *Bioinformatics*, vol. 17, no. 1, pp. 98–99, 2001.
- [16] W. A. Kibbe, "OligoCalc: an online oligonucleotide properties calculator.," *Nucleic Acids Research*, vol. 35, pp. W43–W46, 2007.
- [17] W. J. Kent, "BLAT—the BLAST-like alignment tool," *Genome Research*, vol. 12, no. 4, pp. 656–664, 2002.