

Review

Open Access

Recent developments in primer design for DNA polymorphism and mRNA profiling in higher plants

Xiaohan Yang^{*1,2}, Brian E Scheffler³ and Leslie A Weston¹

Address: ¹Department of Horticulture, Cornell University, Ithaca, NY 14853, USA, ²Department of Plant Sciences, University of Tennessee, 2431 Joe Johnson Drive, Knoxville, TN 37996, USA and ³USDA-ARS-CGRU, MSA Genomics Laboratory, 141 Experiment Station Rd., Stoneville, MS 38776, USA

Email: Xiaohan Yang^{*} - xyang6@utk.edu; Brian E Scheffler - bscheffler@msa-stoneville.ars.usda.gov; Leslie A Weston - law20@cornell.edu

^{*} Corresponding author

Published: 01 March 2006

Received: 14 January 2006

Plant Methods 2006, **2**:4 doi:10.1186/1746-4811-2-4

Accepted: 01 March 2006

This article is available from: <http://www.plantmethods.com/content/2/1/4>

© 2006 Yang et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Primer design is a critical step in the application of PCR-based technologies in gene expression and genetic diversity analysis. As more plant genomes have been sequenced in recent years, the emphasis of primer design strategy has shifted to genome-wide and high-throughput direction. This paper summarizes recent advances in primer design for profiling of DNA polymorphism and mRNA in higher plants, as well as new primer systems developed for animals that can be adapted for plants.

Introduction

mRNA profiling is very important for identifying new genes, determining gene function and elucidating genetic networks. Differential display [1,2], cDNA amplified fragment length polymorphism (cDNA-AFLP) [3-5], and microarray [6] technologies have been employed extensively for profiling of plant mRNAs. DNA polymorphism profiling is essential for gene mapping, marker-assisted selection of crop plants, and molecular diversity studies. PCR-based techniques such as AFLP and microsatellites or simple sequence repeats (SSRs) have also played important roles in plant DNA profiling. Primers are essential components of PCR-based systems as well as modern microarray systems which utilize appropriate probes that are obtained by PCR amplification. This paper summarizes recent advances in primer design for profiling of DNA polymorphism and mRNA in higher plants, as well as new primer systems developed for animals that can be adapted for plants.

PCR primer design in general

Understanding of primer properties is very important for primer design. The major aspects of primer properties include specificity, melting temperature (T_m), and intra-primer or inter-primer homology. Primer specificity is mostly determined by the 3'-end sequences. It was reported that single internal mismatches had no significant effect on PCR product yield while the 3'-terminal mismatches, especially the A:A, A:G, G:A, and C:C mismatches, markedly reduced overall PCR product yield [7]. Khabar et al. [8] assessed the annealing specificity of primers in PCR reactions under different annealing temperatures (35°C, 40°C, and 45°C). They found that there were perfect matches between at least eight bases at the 3' end of the 5' primers and the target region, whereas mispriming occurred only toward the 5' end. Therefore it is critical to include 8–10 unique bases in the 3'-end of the primer. The site-specificity of the primer can be checked by performing a sequence homology search (e.g. blastn) through all known template sequences in the public genome database such as National Center for Biotechnology Information (NCBI) [9]. To ensure specific annealing

Table 1: Free programs for PCR primer design.

Name	Web source	Ref.	Note*
<u>General primer design</u>			
Primer3	http://frodo.wi.mit.edu/cgi-bin/primer3/primer3 WWW.cgi	[63]	W
GeneFisher	http://bibiserv.techfak.uni-bielefeld.de/genefisher/	[64]	W
Primo Pro 3.4	http://www.changbioscience.com/primo/primo.html	[65]	W
PRIMO	http://bioweb.pasteur.fr/seqanal/interfaces/primo.html	[66]	W
FastPCR	http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm	[67]	S
<u>Primer design based upon multi-alignments</u>			
Primaclade	http://doust.umsi.edu/cgi-bin/primaclade.cgi	[68]	W
CODEHOP	http://blocks.fhcrc.org/blocks/codehop.html	[69]	W
PriFi	http://cgi-www.daimi.au.dk/cgi-chili/PriFi/main	[70]	W
<u>Primer design for single-nucleotide polymorphism (SNPs)</u>			
IRA-PCR	http://cedar.genetics.soton.ac.uk/public_html/primer2.html	[71]	W
SNP_Primers	http://www2.eur.nl/fgg/kgen/primer/SNP_Primers.html	[72]	W
Primo SNP 3.4	http://www.changbioscience.com/primo/primosnp.html	[65]	W
<u>Primer design for Simple Sequence Repeat (SSR)</u>			
SSR Finder	http://bioinfo.agri.gov.il/cgi-bin/GE_SSR_Finder.pl	[73]	W
<u>Primer design for Microarrays</u>			
ProbeWiz	http://www.cbs.dtu.dk/services/DNAarray/probewiz.php	[74]	W
ROSO	http://pbil.univ-lyon1.fr/roso/Home.php	[75]	W
OligoWiz 2	http://www.cbs.dtu.dk/services/OligoWiz2/	[76]	S
Picky	http://www.complex.iastate.edu/download/Picky/	[77]	S
<u>Oligonucleotide properties calculation</u>			
Oligo Analyzer	http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/	[78]	W
Poland server	http://www.biophys.uni-duesseldorf.de/local/POLAND/poland.html	[79]	W
NetPrimer	http://www.premierbiosoft.com/netprimer/	[80]	W
dnaMATE	http://dna.bio.puc.cl/cardex/servers/dnaMATE/index.html	[81]	S/W

*Note: W = Web-based; S = Standalone.

of primer to the DNA template, it is also important to avoid 4 or more G's or C's in a row in the 3'-end. T_m is determined by primer length, GC-content and nucleotide composition. Ideally the primer will have a T_m in the range of 50 – 65°C, random nucleotide composition, a 40–60% GC-content, and be 18 – 30 bases long. The

intra-primer or inter-primer homology should be kept as low as possible to avoid formation of hairpin structures (>3 bp complementarity within primer) or primer dimers (>3 bp complementarity between primers) which will interfere with annealing of primer to the DNA template [10]. Up to date a lot of programs have been developed for

Table 3: The 3' eight mRNA specific bases for the 5' primer set for Multiplex DD [15].

ID	Sequence	ID	Sequence	ID	Sequence	ID	Sequence
1	TACTCCCT	37	CAAAGGAC	73	TCTTGGGT	109	CCAAGAGT
2	ATCTCCGA	38	GATAGGAG	74	ACATGGAC	110	GCTAGAAC
3	TCTTCCGA	39	CAAAGGAG	75	CAATGGAG	111	ACCAGAAC
4	TCATCCGA	40	TCTAGGAG	76	CTTTGGTC	112	AGCAGAAC
5	AGATCCGA	41	AAGAGGTC	77	AGTTGGTC	113	GAGAGAAG
6	GATTCCGT	42	CAAAGGTC	78	AAGTGGTG	114	CAGAGAAG
7	ATGTCCGT	43	AGAAGGTC	79	GAATGGTG	115	GGAAGAAG
8	AACTCCGT	44	AAGAGGTG	80	AACTGGTG	116	GTGAGATC
9	GAATCCAC	45	ACAAGGTG	81	AGTTGGTG	117	CTCAGATC
10	TTCTCCAC	46	AAGAGCCA	82	AAGCACCA	118	ACCAGATC
11	ATCTCCAC	47	CAAAGCCA	83	TTCCACCA	119	GCAAGATG
12	GTTTCCAG	48	ATCAGCCA	84	CATCACCA	120	GTCCATCA
13	ATCTCCAG	49	AGTAGCCA	85	AACCACCT	121	GAGCATCT
14	AAGTCCTC	50	TTCAGCCT	86	CTTCACCT	122	GCTCATCT
15	GATTCCTC	51	CTTAGCCT	87	TCTCACCT	123	GCTCATGT
16	ATGTCCTC	52	ACAAGCCT	88	CAACACGA	124	AGCCATGT
17	GTTTCCTC	53	AAGAGCGA	89	CTTCACGA	125	GACCATAC
18	AACTCCTC	54	TACAGCGA	90	ACACACGA	126	CTCCATAC
19	AGATCCTC	55	TTCAGCAC	91	CTTCACGT	127	CACCATTC
20	TTCTCCTG	56	ATCAGCAG	92	GAACACAC	128	GCTCATTG
21	AACTCCTG	57	TACAGCAG	93	GTTACAG	129	CTCCATTG
22	ATCAGGCA	58	AGAAGCAG	94	CTTCACTC	130	GAGAGTCA
23	AAGAGGCT	59	AAGAGCTC	95	CTTCACTG	131	CCAAGTCT
24	ATGAGGCT	60	ATGAGCTC	96	CCTACACT	132	GTCAGTCT
25	CAAAGGCT	61	ATCAGCTC	97	CCAACAGA	133	GAGAGTGT
26	TTGAGGCT	62	TTGAGCTC	98	TGGACAGT	134	GTGAGTGT
27	TCTAGGCT	63	ACAAGCTC	99	CTCACAAC	135	CTCAGTAC
28	TCAAGGCT	64	TCAAGCTG	100	GTCACAAC	136	GGTAGTTC
29	AGAAGGCT	65	CATAGCTG	101	GGAACATC	137	CTGAGTTC
30	ATGAGGGA	66	CTTTGGCA	102	CACACATC	138	GTCAGTTC
31	TCAAGGGA	67	AGATGGCA	103	GTGACATG	139	GCTAGTTG
32	AAGAGGGT	68	GAATGGCT	104	CCAAGACA	140	CTCAGTTG
33	ATGAGGGT	69	TCATGGCT	105	ACCAGACT	141	TCCAGTTG
34	TTCAGGGT	70	AGATGGCT	106	GGAAGAGA	142	AGCAGTTG
35	TCTAGGGT	71	AGTTGGCT	107	GGTAGAGA		
36	GATAGGAC	72	CAATGGGT	108	CAGAGAGT		

It was estimated that DD can detect approximately 96% of expressed genes in a cell utilizing the three different oligo-dT primers in combination with 80 arbitrary primers [16]. This estimation was based on the hypothesis that genes have random nucleotide distributions. However, experimental approaches, as well as computer analysis of genomic sequences, have revealed that there is large variation in base composition between regions in the same genome or between different genomes [18]. The nucleotide distribution within genes is not random [19]. Therefore, random design is not the best approach for creating DD primers. It would be more logical to use a bioinformatical approach for custom design of DD primers based on mRNA sequence information. We designed a set of eight-base sequences targeting plant mRNAs [15]. Specifically, an initial pool of 1,292 eight-base sequences was established based on the analysis of codon usage in eight plant species that included four dicots (*Arabidopsis thal-*

iana, *Lycopersicon esculentum*, *Medicago sativa*, *Nicotiana tabacum*) and four monocots (*Oryza sativa*, *Sorghum bicolor*, *Triticum aestivum*, *Zea mays*). The initial pool of 5' primers was screened against the database At (11,583 *A. thaliana* mRNA sequences) for perfect matches between primers and mRNA sequences in the region of 400 – 1,500 nt from the 3' end. The mRNAs were divided into two groups, with each mRNA having one, and more than one primer matches, respectively. The sequences matching the mRNAs that had one primer match each were selected into the primer set for mRNA profiling, which contained 142 primers (Table 3). The selected primer set was tested on databases At800 (9,370 *A. thaliana* mRNA sequences derived from database At by removing sequences with a length of ≤800 nt), Odi800 (650 mRNA sequences of >800 nt from the three dicots: *L. esculentum*, *M. sativa*, *N. tabacum*), and Mon800 (1,081 mRNA sequences of >800 nt from the four monocots: *O. sativa*,

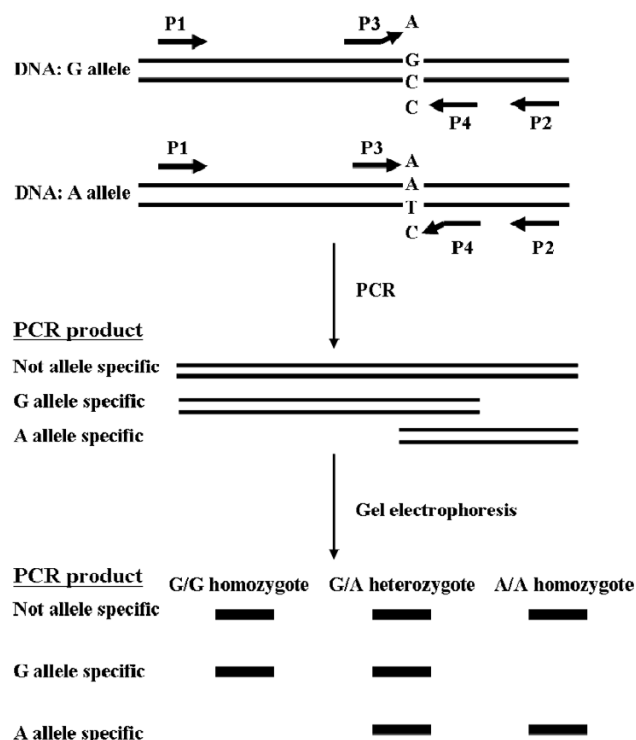


Figure 1
Diagrammatic presentation of the tetra-primer ARMS-PCR method for SNP identification (Redrawn from [58]). Two allele-specific amplicons are generated using two pairs of primers, one pair (P1 and P4) producing an amplicon representing the G allele and the other pair (P2 and P3) producing an amplicon representing the A allele. By positioning the two outer primers (P1 and P2) at different distances from the polymorphic nucleotide, the two allele-specific amplicons differ in length, allowing them to be discriminated by gel electrophoresis.

S. bicolor, *T. aestivum*, *Z. mays*) by a search for perfect matches between primers and mRNA sequences in the region of 400 – 1,500 nt from the 3' end. The primer set matched ~91% of mRNAs in each of the three databases. Since these databases contain mRNA sequences of diverse plant species including both monocots and dicots, it is also likely that this primer set would generate good coverage of mRNAs in a variety of other plant species as well. There was no major difference in the number of primer matches per mRNA among the three databases tested, with an average of about three primer matches per mRNA. This sampling redundancy is close to an estimation given by calculations using equation $P(0) = e^{-\mu}$, where μ represents the sampling redundancy and $P(0)$ represents the probability of missing an mRNA by primers [20]. According to this equation, each mRNA species needs to be matched (or sampled) 2.4 [= $-\ln(1-0.91)$] times on average by the oligos to achieve at least 91% coverage of an

mRNA database. Also, there was no significant difference observed in the mRNA-matching frequency per primer among the three databases evaluated. On average each primer matched ~2% of mRNAs in each of the databases.

Primers for cDNA-AFLP

In cDNA-AFLP, cDNA samples are digested with two different restriction enzymes, adapters are attached to the specific ends of the resulting fragments, and the fragments are amplified using primers homologous to the adaptors with an extension of additional selective nucleotides. Thus, for each selective primer pair only the fragments whose ends match the primer extensions get amplified and these fragments form a pool. Finally, the fragments in each pool are separated by electrophoresis [21]. The primer design for cDNA-AFLP depends on the choice of restriction enzymes and 3' selective sequences. Unfortunately, one pair of enzymes does not in practice produce a fragment for every cDNA molecule that could be amplified and detected by electrophoresis. The fragments generated from a particular cDNA can be too long or too short to be revealed by electrophoresis. One pair of restriction enzymes generally covers up to two-thirds of the transcripts in a species [4,21,22]. In cDNA-AFLP performed with plant samples, two selective bases on the 3' end of each primer are required to give a scorable banding pattern, giving a total of 256 (16×16) possible primer combinations [22].

Recently, Wang and Bughara [5] found that for *Festuca* species, restriction enzyme *NspI* coupled with *TaqI* generated a much higher number of transcript-derived fragments than the commonly used enzyme pair of *EcoRI* and *TaqI*. The enzyme *NspI* has two degenerate bases in its recognition sequence (RCATGY). This enzyme can cut the cDNA more frequently than *EcoRI*, and generate more bands in the cDNA-AFLP gels. An additional advantage of using enzyme pair of *NspI* and *TaqI* is that the possible selective primer combinations were 128 (8×16), only half of those for *EcoRI/TaqI* (16×16). To achieve a 90% coverage of mRNAs in a species, it is necessary to increase the number of enzyme combinations up to 4 [21]. In a cDNA-AFLP system with 4 enzyme combinations and two selective bases on the 3' end of each primer, there are 1,024 (4×256) possible primer combinations. PCR analysis of all the 1,024 possible primer combinations would be time-consuming and costly. In addition, a portion of the mRNA species could be sampled by two or more enzyme combinations, resulting in a somewhat wasteful usage of resources. Fortunately, several computer programs have recently been developed to perform *in silico* simulation of cDNA-AFLP using available sequencing data [21,23,24]. With the assumption that the real target genome has roughly the same characteristics as the sequence data available [21], it is possible to find appro-

Table 4: Oligonucleotide sequences of 12 URP primers (Adapted from [62]).

Primers	Sequences (5'-3')
URP1F	ATCCAAGGTCCGAGACAACC
URP2F	GTGTGCGATCAGTTGCTGGG
URP2R	CCCAGCAACTGATCGCACAC
URP4R	AGGACTCGATAACAGGCTCC
URP6R	GGCAAGCTGGTGGGAGGTAC
URP9F	ATGTGTGCGATCAGTTGCTG
URP13R	TACATCGCAAGTGACACAGG
URP17R	AATGTGGGCAAGCTGGTGGT
URP25F	GATGTGTTCTTGAGCCTGT
URP30F	GGACAAGAAGAGGATGTGGA
URP32F	TACACGTCTCGATCTACAGG
URP38F	AAGAGGCATTCTACCACCAC

priate enzyme combinations to increase mRNA coverage while decreasing selective primer combinations by reducing redundant coverage of the same mRNA species. This could be achieved by simulating cDNA-AFLP *in silico* for the available genome sequencing data in *A. thaliana* [25], *O. sativa* [26-28] and *Populus trichocarpa* [29] as well as plant EST data in the public domains.

Primers for microarray probes

DNA microarrays provide powerful tools for global mRNA profiling. Despite widespread use, recent studies have demonstrated discordance among data produced by different microarray platforms and approaches [30-33]. For example, Tan et al. [32] reported that from a set of 185 common genes in PANC-1 cells, only four behaved consistently on three major commercial microarray platforms from Affymetrix, Agilent and Amersham. One major reason for this is due to the fact that probes have not generally been designed in the past for specificity with gene-splice variants. It is encouraging that companies are now beginning to make arrays specific to different splice variants [31,34,35]. The discordance among different microarray platforms can also be caused by cross-hybridization of highly similar sequences [34]. Possible choices of probe types include spotted cDNA sequences or PCR products, several hundred to thousand base pairs in length, short (25-30 mer) oligonucleotides or longer (60-70 mer) oligonucleotide reporters [36]. Based on theoretical considerations that were confirmed experimentally, it appears that 150-mer is the optimal probe length for expression measurement [37], and thus PCR primers can be designed to amplify the 150-mer gene-specific probes.

Cross-species comparisons of gene expression are important for identifying functionally related genes, because if a set of genes displays similar expression patterns in several species, the probability that the genes are functionally related, rather than co-expressed by chance, increases. Microarray experiments using probes covering a whole

transcriptome are expensive to initiate, and a major part of the cost arises from the synthesis of gene-specific PCR primers or hybridization probes [38]. Andersson et al. [38] developed a method to reduce the number of primers required to amplify the genes of two different genomes. In this method, regions of high sequence similarity were identified, and from these regions PCR primers shared between the genomes were selected, such that either one or, preferentially, both primers in a given PCR were used for amplification from both genomes. This method could be used to design PCR primers for amplification of microarray probes shared by *A. thaliana* and *P. trichocarpa*, or *A. thaliana* and *O. sativa*, or *P. trichocarpa* and *O. sativa*.

Primer design for DNA profiling

Primers for sequence-related amplified polymorphism (SRAP)

Recently, a series of SRAP primers were designed by Li and Quiros [39] for the amplification of open reading frames (ORFs). This system is based on two-primer amplification. The 17 or 18-mer primers consist of the following elements: core sequences, which are 13 - 14 bases long, where the first 10 or 11 bases starting at the 5' end are sequences of no specific constitution ("filler" sequences) followed by the sequence CCGG in the forward primer and AATT in the reverse primer. The core is followed by three selective nucleotides at the 3' end. The filler sequences of the forward and reverse primers must be different from each other and can be 10 or 11 bases long. The SRAP system has been successfully utilized to profile DNA polymorphism in turf grass species [40], tomato (*Lycopersicon esculentum* L. Mill.) [41], and squash (*Cucurbita moschata*) [42].

Primers for sequence-specific amplification polymorphism (SSAP)

SSAP is a multiplex amplified fragment length polymorphism (AFLP)-like technique that displays individual retrotransposon insertions as bands on a sequencing gel. Retrotransposons are mobile genetic elements that accomplish transposition via an RNA intermediate that is

reverse transcribed before integration into a new location within the host genome. They are ubiquitous in eukaryotic organisms and constitute a major portion of the nuclear genome (often more than half of the total DNA) in plants [43]. It has been successfully used for profiling of DNA polymorphism in many crops such as barley (*Hordeum vulgare*) [44], *M. sativa* [45], and sweetpotato (*Ipomoea batatas* (L.) Lam.) [46]. To adapt conventional SSAP method to high-throughput situations, Tang et al. [47] developed and optimized a fluorescent multiplex PCR system for simultaneous selective amplification of Ty1-copia retrotransposon-based SSAPs, followed by capillary electrophoresis.

Primers for simple sequence repeat (SSR)

Recently, Robinson et al. [48] developed a computer program to identify and design PCR primers for amplification of SSR loci based on available DNA sequence information. SSR primers have been designed using publicly available expressed sequence tags (ESTs) in barley [49,50], almond (*Prunus communis* Fritsch.) [51], peach (*P. persica* (L.) Batsch.) [51], *T. aestivum* [52], and *O. sativa* [52]. These SSRs are useful as molecular markers because their development is inexpensive, they represent transcribed genes and their putative function can often be deduced by a homology search [53]. SSRs have been the backbone to creating molecular maps for a number of years.

Chung and Staub [54] developed a set of consensus chloroplast primer pairs for simple sequence repeats (ccSSRs) from *N. tabacum* chloroplast sequences. All primer pairs produced amplicons after PCR employing chloroplast DNA from members of the *Cucurbitaceae* (six species) and *Solanaceae* (four species). Sixteen, 22 and 19 of the initial 23 primer pairs were successively amplified by PCR using template DNA from species of the *Apiaceae* (two species), *Brassicaceae* (one species) and *Fabaceae* (two species), respectively. Twenty of the 23 primer pairs were also functional in three monocot species of the *Liliaceae* (onion and garlic), and the *Poaceae* (oat). ccSSR primers were strategically "recombined" and were referred to correctly as recombined consensus chloroplast primers (RCCP) for PCR analysis of cucumber DNA.

Target-specific PCR primers

Gawel et al. [55] developed a semi-specific PCR system in which primers were designed to target the semi-conservative sequences of the intron-exon junction. The most informative primers were selected from among the exon targeting (ET) and intron targeting (IT) primers, 12 to 18 bases in length. Also, Holland et al. [56] developed PCR primer pairs that target exons, introns, promoter regions in *Z. mays* and introns as well as repeat sequences in *Avena sativa*. Most recently, Hu and Vick [57] developed a primer system called target region amplification polymorphism

(TRAP). This system uses 2 primers of 18 nucleotides each to generate markers. One of the primers, the fixed primer, is designed from the targeted EST sequence in the database; the second primer, the arbitrary primer, is an arbitrary sequence with either an AT- or GC-rich core to anneal with an intron or exon, respectively. The TRAP technique, taking advantage of the availability of sequence information, should be useful in plant genomics research involved in marker-trait association [57].

Primers for single-nucleotide polymorphism (SNPs)

Ye et al. [58] established an efficient procedure for genotyping single nucleotide polymorphisms, named tetra-primer ARMS-PCR, which employs two primer pairs to amplify, respectively, the two different alleles of a SNP in a single PCR reaction. ARMS-PCR has been used for barley SNP genotyping [59]. The ARMS-PCR primer system is illustrated in Figure 1. Also, Kota et al. [50] developed SNP primer pairs for barley based on available EST database. Recently, a computer program was developed to automate the primer design process for SNP analysis [60]. Recently, Rudd et al. [61] created a database resource, PlantMarkers, to predict, analyze and display various molecular markers including SNP and SSR for over 50 plant species. This database will greatly facilitate primer design for profiling of DNA polymorphisms using SNP and SSR in the future.

Universal rice primer (URP)

Repeat-based PCR strategies such as microsatellites are also potentially very useful for DNA polymorphism profiling. Recently, Kang et al. [62] developed a primer system, referred to as the universal rice primer (URP), based on a repetitive DNA fragment (pKRD) in rice. Forty 20-mer primers were randomly designed from the entire pKRD fragment, with the idea that short oligomers complementary to primers are well dispersed within the rice genome. Twelve primers listed in Table 4 produced characteristic fingerprints from diverse genomes of 14 plant species: *O. sativa*, *Z. mays*, barley, bamboo (*Phyllostachys spp.*), oat (*A. sativa*), soybean (*Glycine max* L.), chinese cabbage (*Brassica rapa* var. *pekinensis*), pumpkin (*Cucurbita pepo* L.), cucumber (*Cucumis sativa* L.), spinach (*Spinaceae oleracea* L.), pepper (*Capsicum annuum* L.), garlic (*Allium sativum* L.), *N. tabacum*, *A. thaliana*, 7 animal species and 6 microbial species, indicating its universal applicability.

Conclusion

In the post-genomics era, recent availability of DNA sequence data has fostered the further development of DNA/mRNA profiling technologies that exhibit enhanced genome-wide coverage and improved targeting accuracy. Recent trends related to primer design include the following:

- 1) Optimization of primer structure or probe properties for increased specificity of primer (or probe) hybridization to the target sequence in PCR reactions (or microarray analysis).
- 2) Increased genome-wide coverage with minimum primer numbers and reduced sampling redundancy.
- 3) Novel primer design for amplification of microarray probes specific to gene-splice variants for more accurate mRNA profiling.

Recent innovations have led to more cost-effective and successful profiling studies and greater ease in subsequent purification of gene fragments. This review is intended not only to help scientists to update their knowledge of primer design for DNA polymorphism and mRNA profiling in higher plants, but also to increase their interests in making technical improvements using plant genomics and bioinformatics approach.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

XY carried out the design of the modified primers for differential display, and drafted the manuscript. BS and LW conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Roselee Harmon for assistance with laboratory analyses performed to obtain data related to primer utilization and specificity.

References

1. Liang P: **A decade of differential display.** *Biotechniques* 2002, **33**(2):338-44, 346.
2. Liang P, Pardee AB: **Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction.** *Science* 1992, **257**(5072):967-971.
3. Bachem CWB, Oomen R, Visser RGF: **Transcript imaging with cDNA-AFLP: A step-by-step protocol.** *Plant Molecular Biology Reporter* 1998, **16**(2):157-173.
4. Volkmut W, Turk S, Shapiro A, Fang Y, Kiegle E, van Haaren M, Donson J: **Technical advances: genome-wide cDNA-AFLP analysis of the Arabidopsis transcriptome.** *Omics* 2003, **7**(2):143-159.
5. Wang JP, Bughrara SS: **Detection of an Efficient Restriction Enzyme Combination for cDNA-AFLP Analysis in Festuca mairei and Evaluation of the Identity of Transcript-Derived Fragments.** *Mol Biotechnol* 2005, **29**(3):211-220.
6. Kuhn E: **From library screening to microarray technology: Strategies to determine gene expression profiles and to identify differentially regulated genes in plants.** *Annals of Botany* 2001, **87**(2):139-155.
7. Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ: **Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type I model studies.** *Nucleic Acids Res* 1990, **18**(4):999-1005.
8. Khabar KSA, Dhalla M, Bakheet T, Sy C, al-Haj L: **An integrated computational and laboratory approach for selective amplification of mRNAs containing the adenylate uridylylate-rich element consensus sequence.** *Genome Res* 2002, **12**(6):985-995. [<http://www.ncbi.nlm.nih.gov/>]. National Center for Biotechnology Information
10. Sharrocks AD: **The design of primers for PCR.** In *PCR technology: current innovations* Edited by: Griffin HG, Griffin AM. London, CRC Press; 1994:5-11.
11. Jorgensen M, Bevoort M, Kledal TS, Hansen BV, Dalgaard M, Leffers H: **Differential display competitive polymerase chain reaction: an optimal tool for assaying gene expression.** *Electrophoresis* 1999, **20**(2):230-240.
12. Crawford DR, Kochheiser JC, Schools GP, Salmon SL, Davies KJ: **Differential display: a critical analysis.** *Gene Expr* 2002, **10**(3):101-107.
13. Hwang IT, Kim YJ, Kim SH, Kwak CI, Gu YY, Chun JY: **Annealing control primer system for improving specificity of PCR amplification.** *Biotechniques* 2003, **35**(6):1180-1184.
14. Kim YJ, Kwak CI, Gu YY, Hwang IT, Chun JY: **Annealing control primer system for identification of differentially expressed genes on agarose gels.** *Biotechniques* 2004, **36**(3):424-6, 428, 430.
15. Yang XH: **Development of new technology for isolation of key bioherbicidal genes in sorghum root hairs.** PhD thesis. Cornell University, Department of Horticulture; 2003.
16. Liang P: **Factors ensuring successful use of differential display.** *Methods* 1998, **16**(4):361-364.
17. Cho YJ, Prezioso VR, Liang P: **Systematic analysis of intrinsic factors affecting differential display.** *Biotechniques* 2002, **32**(4):762-4, 766.
18. Gautier C: **Compositional bias in DNA.** *Current Opinion in Genetics & Development* 2000, **10**(6):656-661.
19. Som A, Sahoo S, Chakrabarti J: **Coding DNA sequences: statistical distributions.** *Mathematical Biosciences* 2003, **183**(1):49-61.
20. McClelland M, Welsh J: **RNA fingerprinting by arbitrarily primed PCR.** *PCR Methods Appl* 1994, **4**(1):S66-81.
21. Kivioja T, Arvas M, Saloheimo M, Penttilä M, Ukkonen E: **Optimization of cDNA-AFLP experiments using genomic sequence data.** *Bioinformatics* 2005.
22. Bachem CW, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RG: **Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development.** *Plant J* 1996, **9**(5):745-753.
23. Rombauts S, Van De Peer Y, Rouze P: **AFLP in Silico, simulating AFLP fingerprints.** *Bioinformatics* 2003, **19**(6):776-777.
24. Qin L, Prins P, Jones JT, Popeijus H, Smant G, Bakker J, Helder J: **GenEST, a powerful bidirectional link between cDNA sequence data and gene expression profiles generated by cDNA-AFLP.** *Nucleic Acids Res* 2001, **29**(7):1616-1622.
25. The Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.** *Nature* 2000, **408**(6814):796-815.
26. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhattacharjee S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalima T, Oliphant A, Briggs S: **A draft sequence of the rice genome (Oryza sativa L. ssp. japonica).** *Science* 2002, **296**(5565):92-100.
27. Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Li J, Liu Z, Qi Q, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Zhao W, Li P, Chen W, Zhang Y, Hu J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Tao M, Zhu L, Yuan L, Yang H: **A draft sequence of the rice genome (Oryza sativa L. ssp. indica).** *Science* 2002, **296**(5565):79-92.

28. International Rice Genome Sequencing Project: **The map-based sequence of the rice genome.** *Nature* 2005, **436**(7052):793-800.
29. Brunner AM, Busov VB, Strauss SH: **Poplar genome sequence: functional genomics in an ecologically dominant plant species.** *Trends Plant Sci* 2004, **9**(1):49-56.
30. Kuo WVP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS: **Analysis of matched mRNA measurements from two different microarray technologies.** *Bioinformatics* 2002, **18**(3):405-412.
31. Kothapalli R, Yoder SJ, Mane S, Loughran TPJ: **Microarray results: how accurate are they?** *BMC Bioinformatics* 2002, **3**(1):22.
32. Tan PK, Downey TJ, Spitznagel ELJ, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC: **Evaluation of gene expression measurements from commercial microarray platforms.** *Nucleic Acids Res* 2003, **31**(19):5676-5684.
33. Mah N, Thelin A, Lu T, Nikolaus S, Kuhbacher T, Gurbuz Y, Eickhoff H, Kloppel G, Lehrach H, Mellgard B, Costello CM, Schreiber S: **A comparison of oligonucleotide and cDNA-based microarray systems.** *Physiol Genomics* 2004, **16**(3):361-370.
34. Marshall E: **Getting the noise out of gene arrays.** *Science* 2004, **306**(5696):630-631.
35. Fehlbaum P, Guihal C, Bracco L, Cochet O: **A microarray configuration to quantify expression levels and relative abundance of splice variants.** *Nucleic Acids Res* 2005, **33**(5):e47.
36. Yauk CL, Berndt ML, Williams A, Douglas GR: **Comprehensive comparison of six microarray technologies.** *Nucleic Acids Res* 2004, **32**(15):e124.
37. Chou CC, Chen CH, Lee TT, Peck K: **Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression.** *Nucleic Acids Res* 2004, **32**(12):e99.
38. Andersson A, Bernander R, Nilsson P: **Dual-genome primer design for construction of DNA microarrays.** *Bioinformatics* 2005, **21**(3):325-332.
39. Li G, Quiros CF: **Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica.** *Theoretical and Applied Genetics* 2001, **103**(2-3):455-461.
40. Budak H, Shearman RC, Gaussoin RE, Dweikat I: **Application of sequence-related amplified polymorphism markers for characterization of turfgrass species.** *Hortscience* 2004, **39**(5):955-958.
41. Ruiz JJ, Garcia-Martinez S, Pico B, Gao MQ, Quiros CF: **Genetic variability and relationship of closely related Spanish traditional cultivars of tomato as detected by SRAP and SSR markers.** *Journal of the American Society for Horticultural Science* 2005, **130**(1):88-94.
42. Ferriol M, Pico B, de Cordova PF, Nuez F: **Molecular diversity of a germplasm collection of squash (*Cucurbita moschata*) determined by SRAP and AFLP markers.** *Crop Science* 2003, **44**(2):653-664.
43. Kumar A, Hirochika H: **Applications of retrotransposons as genetic tools in plant biology.** *Trends Plant Sci* 2001, **6**(3):127-134.
44. Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W: **Genetic distribution of Bare-I-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP).** *Mol Gen Genet* 1997, **253**(6):687-694.
45. Porceddu A, Albertini E, Barcaccia G, Marconi G, Bertoli FB, Veronesi F: **Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L.** *Mol Genet Genomics* 2002, **267**(1):107-114.
46. Berenyi M, Gichuki T, Schmidt J, Burg K: **Tyl-copia retrotransposon-based S-SAP (sequence-specific amplified polymorphism) for genetic analysis of sweetpotato.** *Theor Appl Genet* 2002, **105**(6-7):862-869.
47. Tang T, Huang J, Zhong Y, Shi S: **High-throughput S-SAP by fluorescent multiplex PCR and capillary electrophoresis in plants.** *J Biotechnol* 2004, **114**(1-2):59-68.
48. Robinson AJ, Love CG, Batley J, Barker G, Edwards D: **Simple sequence repeat marker loci discovery using SSR primer.** *Bioinformatics* 2004, **20**(9):1475-1476.
49. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.).** *Theor Appl Genet* 2003, **106**(3):411-422.
50. Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A: **Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.).** *Hereditas* 2001, **135**(2-3):145-151.
51. Xu Y, Ma RC, Xie H, Liu JT, Cao MQ: **Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region.** *Genome* 2004, **47**(6):1091-1104.
52. Yu JK, La Rota M, Kantety RV, Sorrells ME: **EST derived SSR markers for comparative mapping in wheat and rice.** *Mol Genet Genomics* 2004, **271**(6):742-751.
53. Varshney RK, Graner A, Sorrells ME: **Genic microsatellite markers in plants: features and applications.** *Trends Biotechnol* 2005, **23**(1):48-55.
54. Chung SM, Staub JE: **The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa.** *Theor Appl Genet* 2003, **107**(4):757-767.
55. Gawel M, Wisniewska I, Rafalski A: **Semi-specific PCR for the evaluation of diversity among cultivars of wheat and triticale.** *Cellular & Molecular Biology Letters* 2002, **7**(2A):577-582.
56. Holland JB, Helland SJ, Sharopova N, Rhyne DC: **Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat.** *Genome* 2001, **44**(6):1065-1076.
57. Hu JG, Vick BA: **Target region amplification polymorphism: A novel marker technique for plant genotyping.** *Plant Molecular Biology Reporter* 2003, **21**(3):289-294.
58. Ye S, Dhillion S, Ke X, Collins AR, Day IN: **An efficient procedure for genotyping single nucleotide polymorphisms.** *Nucleic Acids Res* 2001, **29**(17):E88-8.
59. Chiapparino E, Lee D, Donini P: **Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS-PCR.** *Genome* 2004, **47**(2):414-420.
60. Kaderali L, Deshpande A, Nolan JP, White PS: **Primer-design for multiplexed genotyping.** *Nucleic Acids Res* 2003, **31**(6):1796-1802.
61. Rudd S, Schoof H, Mayer K: **PlantMarkers--a database of predicted molecular markers from plants.** *Nucleic Acids Res* 2005, **33** Database Issue:D628-32.
62. Kang HW, Park DS, Go SJ, Eun MY: **Fingerprinting of diverse genomes using PCR with universal rice primers generated from repetitive sequence of Korean weedy rice.** *Mol Cells* 2002, **13**(2):281-287.
63. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132**:365-386.
64. Giegerich R, Meyer F, Schleiermacher C: **GeneFisher--software support for the detection of postulated genes.** *Proc Int Conf Intell Syst Mol Biol* 1996, **4**:68-77.
65. [<http://www.changbioscience.com/>]. Chang Bioscience
66. Li P, Kupfer KC, Davies CJ, Burbree D, Evans GA, Garner HR: **PRIMO: A primer design program that applies base quality statistics for automated large-scale DNA sequencing.** *Genomics* 1997, **40**(3):476-485.
67. Kalendar R: **FastPCR Ppd, DNA and protein tools, repeats and own database searches program.** 2006 [<http://www.bio-center.helsinki.fi/bi/Programs/fastpcr.htm>].
68. Gadberry MD, Malcomber ST, Doust AN, Kellogg EA: **Primaclade--a flexible tool to find conserved PCR primers across multiple species.** *Bioinformatics* 2005, **21**(7):1263-1264.
69. Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S: **Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences.** *Nucleic Acids Res* 1998, **26**(7):1628-1635.
70. Fredslund J, Schauser L, Madsen LH, Sandal N, Stougaard J: **PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs.** *Nucleic Acids Res* 2005, **33**(Web Server issue):W516-20.
71. Ke X, Collins A, Ye S: **PIRA PCR designer for restriction analysis of single nucleotide polymorphisms.** *Bioinformatics* 2001, **17**(9):838-839.
72. van Baren MJ, Heutink P: **The PCR suite.** *Bioinformatics* 2004, **20**(4):591-593.
73. **GreenExpress** [http://bioinfo.agri.gov.il/cgi-bin/green_page.pl]
74. Nielsen HB, Knudsen S: **Avoiding cross hybridization by choosing nonredundant targets on cDNA arrays.** *Bioinformatics* 2002, **18**(2):321-322.

75. Reymond N, Charles H, Duret L, Calevro F, Beslon G, Fayard JM: **ROSO: optimizing oligonucleotide probes for microarrays.** *Bioinformatics* 2004, **20(2)**:271-273.
76. Wernersson R, Nielsen HB: **OligoWiz 2.0--integrating sequence feature annotation into the design of microarray probes.** *Nucleic Acids Res* 2005, **33(Web Server issue)**:W611-5.
77. Chou HH, Hsia AP, Mooney DL, Schnable PS: **Picky: oligo microarray design for large genomes.** *Bioinformatics* 2004, **20(17)**:2893-2902.
78. **Integrated DNA Technologies (IDT)** [<http://idtdna.com/Home/Home.aspx>]
79. **Poland server** [<http://www.biophys.uniduesseldorf.de/local/POLAND/poland.html>]
80. **NetPrimer** [<http://www.premierbiosoft.com/netprimer/>]
81. Panjkovich A, Norambuena T, Melo F: **dnaMATE: a consensus melting temperature prediction server for short DNA sequences.** *Nucleic Acids Res* 2005, **33(Web Server issue)**:W570-2.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

