

Summary of Conserved Primer Design for Wheat SNP Discovery

Objectives

Development of conserved PCR primers for test

Materials

EST sequences:

For the test purpose, the four sets of EST sequences were selected for primer design:

1. Mapped contigs and singletons: The total of 4186 contigs (unigenes) and 1859 singletons mapped to wheat chromosomes were obtained by a query from the EST databases (GrainGene database).
2. Mapped single 5' EST sequences: 6426.
3. Mapped single 3' EST sequences: 5814.

Rice sequences:

A total of 4074 rice sequences with greater than 10kb and less than 1Mbp were downloaded from RiceGAAS(<http://ricegaas.dna.affrc.go.jp/>). All sequences have 545,430,227 bp long.

Pipeline of Conserved PCR primer design

1. The pipeline of conserved primer design is illustrated in Figure 1.
2. *Blastn* search against rice sequence database:
Parameters: -b 10 -W 15 -E 3 -G 1 -U -e 1e10
3. Intron/exon analysis (alignment of introns/exons):
 - a. Find all matches with bit score > 50 or expect value > 1e-10. Each match corresponds to an exon in wheat EST or rice sequence.
 - b. Exclude the exon(s) from ESTs in which the same slice of sequence has more than one matches (repetitive elements)
 - c. Exclude the EST contigs in which slices of sequences have wrong order of matches against rice sequences. This is probably due to the incorrect EST assembly.
 - d. Exclude the ESTs with no exon or only one exon found from blastn search.
 - e. Calculate the length of exons and introns, and find the coordinates of each exon and intron.
 - f. Create a new sequence for each EST or contig, in which the intron sequences were inserted between two exons, but intron sequences were replaced by "N", e.g., "...GATCGGTTTACNNNNNNNN....NNNNGGTTCAATT...". This new sequence was used to design primer in *Primer3* program. The advantages to use this new sequence rather than the original EST sequence are that (1) primers can surely be picked only from exon regions, and (2) the product size, and number and length of introns and exons included in product can be estimated, (3) in

primer analysis, we can check if the primers picked using Primer3 come from different exons or from the same exon region.

4. Primer design:

Parameters used in Primer3 program:

| <i>Parameters</i> | <i>Min</i> | <i>Optimum</i> | <i>Max</i> |
|-------------------|------------|----------------|------------|
| Product size | 400 | 800 | 1500 |
| Tm | 55 | 60 | 65 |
| Primer length | 18 | 20 | 25 |
| GC content | 30 | | 70 |

Only one primer set was picked for each EST or contig.

5. Primer analysis:

- Assign the corresponding EST sequence name, accession number, contig name (if any), chromosome position, number of match exons, coordinates of exons in sequences, match score of each exon, etc. to each primer set.
- Find out the exons the primer sequences were picked from. If the primers are picked from the same exon, the primer set will be removed and re-picked manually later.
- Estimate the product size, number of introns and exons included in the amplifying product, length of introns and exons included in the amplifying product.
- Transfer all related information into databases.

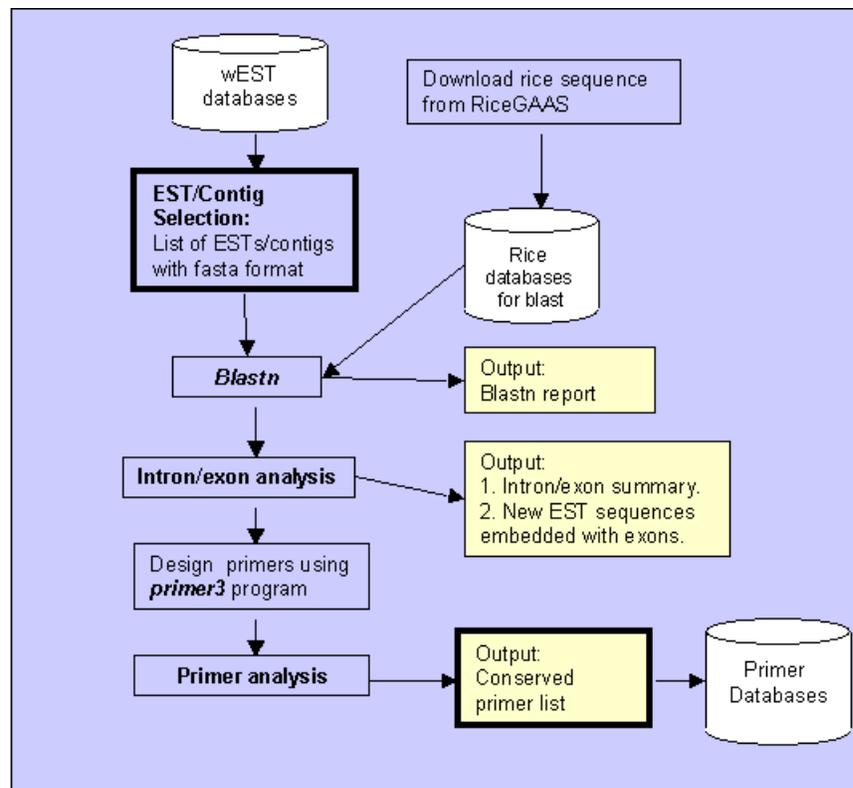


Figure 1: Pipeline of conserved PCR primer design for STS amplification

Results:

Pipeline of conserved primer design:

Several Java and Perl programs have been developed. Those programs along with *blastn* and *primer3* programs, as well as rice sequence database compose a pipeline of the conserved primer design. Currently, the primer design is carried out by running several separate programs. The pipeline has not been generalized yet.

Primer sets designed for test purpose:

Table 1: Primer sets designed for test

| <i>EST resources</i> | <i>Total sequences</i> | <i>Sequences matched 2 or more exons</i> | <i>Total primers picked</i> | <i>Chr 1</i> | <i>Chr 2</i> | <i>Chr 3</i> | <i>Chr 4</i> | <i>Chr 5</i> | <i>Chr 6</i> | <i>Chr 7</i> | <i>Unknown</i> |
|----------------------|------------------------|--|-----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|
| Mapped contig | 4186 | 2240 | 1958 | 254 | 295 | 285 | 334 | 319 | 226 | 245 | 0 |
| Mapped singletons | 1859 | 373 | 265 | 41 | 43 | 41 | 44 | 41 | 28 | 27 | 0 |
| Mapped single 5' EST | 6426 | 2092 | 1624 | 211 | 237 | 261 | 272 | 238 | 189 | 214 | 2 |
| Mapped single 3' EST | 5814 | 1020 | 705 | 95 | 103 | 113 | 117 | 119 | 77 | 81 | 0 |

Table 2: The success rates during primer design

| <i>Step</i> | <i>Mapped contig</i> | <i>Mapped singletons</i> | <i>Mapped 5' EST</i> | <i>Mapped 3'EST</i> |
|-------------------------------|----------------------|--------------------------|----------------------|---------------------|
| EST selected | 4186 | 1859 | 6426 | 5814 |
| Successful alignment of exons | 54% | 20% | 33% | 18% |
| Conserved primer | 87% | 71% | 78% | 69% |

Primer lists for test purpose:

The four primer lists are provided for test. These primers are saved in the four MS excel format files:

- contig_primer_list.xls
- singleton_primer_list.xls
- single_est_3_primer_list.xls
- single_est_5_primer_list.xls

Table 3: The meaning of each of 24 columns in the primer list files

| Column name | Meaning |
|----------------------|--|
| Primer Name | Primer is named for the accession number of an EST sequence or EST member of a contig plus “_cp”. The “.cp” means “conserved primer”. For a contig, only one of EST sequences in that contig was chosen to be primer name. The accession number of EST sequence is used for primer naming just because usually each EST sequence has 3’ and 5’ sequences but has only one EST sequence name. |
| EST Name | EST sequence name. The 3’ and 5’ sequences share one sequence name. If the primer is picked from a contig, the sequence name of one member of the contig is used. |
| Mapped Member Acc | Accession number of an EST sequence or EST member of a contig. This accession number is used for primer naming. |
| Contig Name | If a primer is picked from a contig, the contig name is listed here. Contig names vary with different assemblies or authors. Thus, contig names can not be used for primer naming. |
| EST Chrom | Chromosome that an EST or a contig has been mapped to. |
| Hit Rice Clone Acc | Accession number of rice sequence an EST sequence or a contig matches. |
| Hit Rice Clone Chrom | Chromosome of rice sequence an EST sequence or a contig matches. |
| Left Primer | Left primer sequence picked from a wheat exon. |
| Left Start | Start position of the left primer picked form the EST sequence embedded with intron sequence(s). |
| Left Len | Left primer length (bp) |
| Left Tm | Left primer melting temperature |
| Left GC | Left primer GC content |
| Left Exon Score | <i>Blastn</i> score of the exon from which left primer is picked |
| Right Primer | Right primer sequence picked from a wheat exon. |
| Right End | End position of the right primer picked form EST sequence embedded with intron sequence(s). |
| Right Len | Right primer length (bp) |
| Right Tm | Right primer melting temperature |
| Right GC | Right primer GC content |
| Right Exon Score | <i>Blastn</i> score of the exon from which right primer is picked |
| Product Size | Predicted product size calculated by right end minus left start |
| Included Introns | Predicted number of the introns in the amplified product |
| Matched No Exons | Number of the exons found by an EST sequence matching a rice sequence |
| Included Exons | Predicted number of the exons included in the amplified product |
| Included Intron Size | Predicted intron size (bp) in the amplified product |

Note: if a primer set has value 0s in the columns: left exon score, right exon score, included introns, matched no exons, includes exons, and included intron size, this primer set is picked from the same exon region. You don’t choose this primer for test because this primer will be redesigned manually later.