RESEARCH NOTE

Database of predicted SCAR markers in five fruit and three vegetable crops

BALAKRISHNAN VASANTHAKUMARI PREMKRISHNAN[1,2] and VADIVEL ARUNACHALAM[1,3]*

[1]*Central Plantation Crops Research Institute, Kasaragod 671 124, India*
[2]*Present address: SciGenom Labs Pvt. Ltd., Cochin Special Economic Zone, Kakkanad 682 307, India*
[3]*Horticultural Science, ICAR Research Complex for Goa, Ela Old Goa 403 402, India*

[Premkrishnan B. V. and Arunachalam V. 2016 Database of predicted SCAR markers in five fruit and three vegetable crops. *J. Genet.* **95**, 171–175]

Introduction

A wide range of molecular markers—amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), simple sequence repeats (SSR), single-nucleotide polymorphism (SNP)—are available (review by Khlestkina 2014) for genetics and breeding research. Molecular markers work on the principle of variations in sequences for any of the following, namely (i) presence or absence of priming site of the primer (RAPD); (ii) recognition site of a restriction enzyme (RFLP, AFLP); (iii) length of repeat unit in a microsatellite locus (SSR); (iv) substitution (SNP) at or insertion or deletion (INDEL) of a single nucleotide.

RAPD markers are widely used for identifying genetic diversity in living organisms (Williams *et al.* 1990). RAPDs are generated from a single random primer usually only 10 bases long, in a PCR reaction. SCAR (sequence characterized amplified region; Paran and Michelmore 1993) markers were introduced to overcome the disadvantages of RAPD. It is easy to design SCAR primers if the target sequence is known. SCAR primers are longer (18–25 bases) than RAPD primers (10 bases) hence specific. They are usually generated by cloning and sequencing RAPD bands of interest. An algorithm, eRAPD (Li *et al.* 2006), was proposed to identify the priming sites of given RAPD primers in the target genome. The predicted RAPD primers were more likely to give amplification than randomly chosen primers. The software (Marker express 2.0) designs SCAR primers for shortlisted RAPD/ISSR primers. SCAR markers identified by use of SCAR primers designed *in silico* are called iSCAR markers. SCAR markers are deployed in a range of crops (Marczewski *et al*. 2001; Koveza and Gostimsky 2005; Joseph *et al*. 2014) for marker–trait associations. Genome sequences of major fruits important vegetables are now available. Whole genomes offer scope for marker-aided genetic improvement of major horticultural crops. SCAR markers continue to be used along with high-throughput SSR/SNP markers even in species where complete genome sequences are available because they can be used as diagnostic markers to discriminate specific trait differences (e.g. male versus female, Joseph *et al*. 2014) in a species.

In this study we chose five fruit (apple, banana, cacao, papaya and strawberry) and three vegetable crops (potato, tomato and melon) commercially cultivated and/or commonly consumed worldwide. Whole genome, core nucleotide, expressed-sequence tag and genomic survey sequences for the eight target species were retrieved and *in silico* SCAR markers were predicted. The objective of the study was to establish a free online database for iSCAR primers in these eight plants.

Materials and methods

Different types of sequences (whole genome, core nucleotide, expressed sequence tags and genome survey sequences) of the eight target plant species were retrieved from GenBank. Selected primer sequences of RAPD were obtained from www.operon.com and published reports. Selected ISSR primer sequences were obtained from published reports of the University of British Columbia (UBC) set primer and other ISSR markers.

Different iSCAR primers were generated using the software (Premkrishnan and Arunachalam 2012) and input files of nucleotide sequences of each of the target species and RAPD and ISSR primer sequences. The output files of iSCAR primer sequences were parsed and converted to the input format using a set of Perl scripts. A web interface was

*For correspondence. E-mail: vadivelarunachalam@yahoo.com; v.arunachalam@icar.gov.in.

Keywords. sequence characterized amplified region markers; random amplified polymorphic DNA markers.

implemented using HTML, PHP, CSS and JavaScript. The database, called FViSCARdb, was created using MySQL.

Different input tables were constructed for species, sequence type and primers. The result table was designed to store the predicted iSCAR primer sequences and fields like sequence id, product size, etc. Schema of the database was designed to facilitate rapid data retrieval with minimal redundancy. Indexing of the table values was done to make quick retrieval from the database. A flowchart of the database is described in figure 1. A snapshot view of sample output is depicted in figure 2.

Google Scholar and NCBI Pubmed were used to search the published wet-lab reports on RAPD/ISSR markers for polymorphism/diversity or trait–marker association studies in these eight crops. Selected SCAR primers with proof from published reports for product size with original RAPD/ISSR primers were compared with those in the database FViSCARdb. Selected papers reporting primers matching with the predicted SCARs in the database were shortlisted.

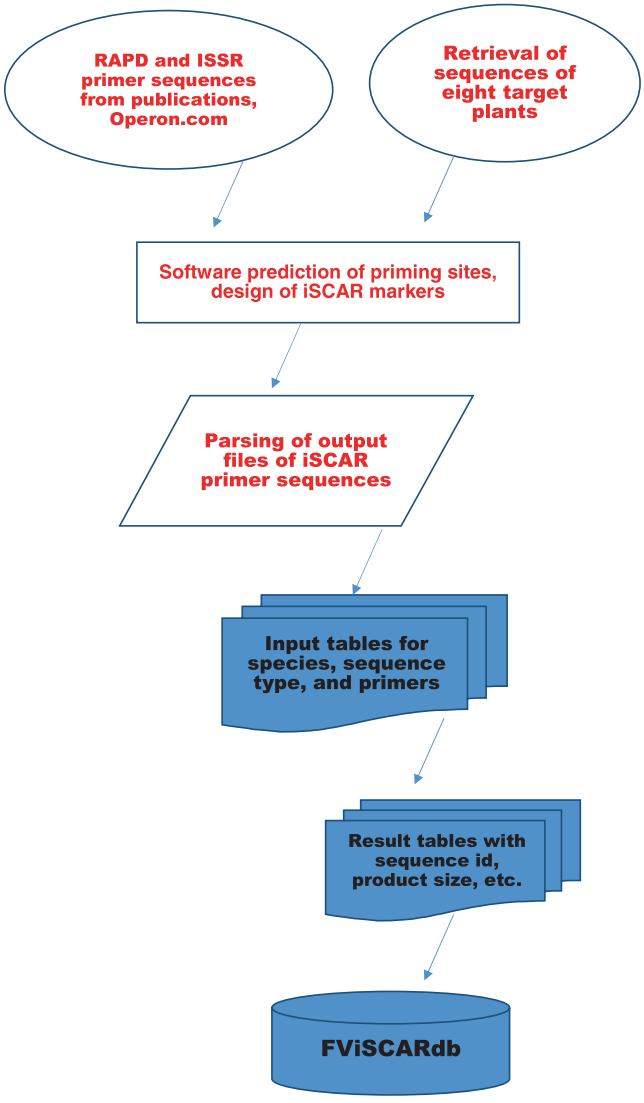

Figure 1. Flowchart of development of FViSCARdb.

Results and discussion

FViSCARdb has a very simple and user-friendly interface (figure 3). The database currently contains sequence information of eight fruit and vegetable plants (apple, banana, cacao, melon, papaya, potato, strawberry and tomato). The sequences for these plants include whole genome sequences, core nucleotide sequences, expressed sequence tag sequences and genome survey sequences. Three list boxes are provided on the home page of FViSCARdb (http://www.bioinfoindia.org/fv-iscardb) to the user to select desired options to search. The user can choose to see the iSCAR primers for a particular species (fruit or vegetable) or for a particular sequence type or for an existing RAPD/ISSR marker. There is also the option to try multiple combinations of searches. The following types of combination searches are possible: (i) species name + sequence type + primer name, (ii) species name + sequence type, (iii) species name + primer name, and (iv) sequence type + primer name. A search result table will be displayed if the search is successful. A short description about the search results can be found at the top of the result page in which the user can see the total number of search results and other information. If the combination search using the parameters that the user selected from the three list boxes is unsuccessful the program automatically tries other combinations using the available parameters. Other information can be viewed at the top of the search result page. After a successful search, a list of predicted iSCAR primers will be displayed on the page. The user can get information on species name, sequence id, sequence type, primer name and type, product size and the sequences of forward and reverse iSCAR primers.

Case study: in silico RAPD-SCAR marker for salt tolerance in banana

We demonstrate use of the database using a published report (Miri *et al.* 2009) of an RAPD marker for salinity tolerance in banana as a case study. The report showed the presence of an RAPD band OPA02, of 240 bp size, in only the salt-tolerant banana clone. We queried the database FViSCARdb to list iSCARprimers for OPA02 with banana as target species and whole genome sequence as query input. The predicted primer pairs (20) are shown in figure 2.

Two primer pairs from the listed primer sequences, which amplified a region of nearly 250 bases, were analysed *in silico* using primer BLAST option of tool menu of the banana genome hub (Droc *et al.* 2013). Two hits were found — chr3: 18,709,636 to 18,710,108, 472 base long; chr7: 17,969,872 to 17,970,338, 466 base long.

The sequence at chr7 matched with a domain of carbon catabolite repressor protein on BLAST search. These two

Species Name	Sequence ID	Sequence Type	Primer Name	Primer Type	Product Size	ForwardSCAR	ReverseSCAR
Banana	>gi|401022641|emb|HE813979.1|	Whole Genome Sequences	OPA-02	RAPD	340	TGCCGAGCTACGCAAGCAG	GCTCGGCAGCCAGAGCCT
Banana	>gi|401022641|emb|HE813979.1|	Whole Genome Sequences	OPA-02	RAPD	1237	ACGGCTCGGTACGACAAA	CGAGCCGTACAAGATTTC
Banana	>gi|400265854|emb|HE813976.1|	Whole Genome Sequences	OPA-02	RAPD	646	ACGGCTCGGCGCGTTCGT	CGAGCCGTCCGCGACACG
Banana	>gi|400265854|emb|HE813976.1|	Whole Genome Sequences	OPA-02	RAPD	709	TGCCGAGCTTCTTCTCTCC	AGCTCGGCAACTGCAGTAG
Banana	>gi|415426013|emb|HE813977.1|	Whole Genome Sequences	OPA-02	RAPD	372	ACGGCTCGTTGACCCAGC	CGAGCCGTAGGTCGGGAA
Banana	>gi|415426013|emb|HE813977.1|	Whole Genome Sequences	OPA-02	RAPD	1027	TGCCGAGCTCCTTGACCGC	GCTCGGCAGCCATGATAG
Banana	>gi|400265855|emb|HE813978.1|	Whole Genome Sequences	OPA-02	RAPD	316	ACGGCTCGGTCGCACCCA	TCGAGCCGTGTCCGGCCTT
Banana	>gi|415426042|emb|HE813979.1|	Whole Genome Sequences	OPA-02	RAPD	431	ACGGCTCGTCGTCCTAGC	CGAGCCGTAGGTCGAGAA
Banana	>gi|415426042|emb|HE813979.1|	Whole Genome Sequences	OPA-02	RAPD	1820	TGCCGAGCTGAGAACGAGGA	AGCTCGGCACTCTCCTCTG
Banana	>gi|415426063|emb|HE813980.1|	Whole Genome Sequences	OPA-02	RAPD	1745	ACGGCTCGAAGCCTTCTGT	TCGAGCCGTGCTGACACCT
Banana	>gi|415426063|emb|HE813980.1|	Whole Genome Sequences	OPA-02	RAPD	265	TGCCGAGCTACGTCTCTCT	GCTCGGCATGTCTACCAC
Banana	>gi|400271221|emb|HE813981.1|	Whole Genome Sequences	OPA-02	RAPD	266	ACGGCTCGAAAGTTGCGCC	CGAGCCGTCCACATGCAG
Banana	>gi|400271221|emb|HE813981.1|	Whole Genome Sequences	OPA-02	RAPD	827	TGCCGAGCAACTAGACTC	GCTCGGCAAACAGGTTCA
Banana	>gi|400265856|emb|HE813982.1|	Whole Genome Sequences	OPA-02	RAPD	403	ACGGCTCGGGAGGCTTAC	CGAGCCGTTGGCCATGAA
Banana	>gi|400265856|emb|HE813982.1|	Whole Genome Sequences	OPA-02	RAPD	1387	TGCCGAGCGTCTGGATCT	GCTCGGCAACCTCACGAG
Banana	>gi|415426079|emb|HE813983.1|	Whole Genome Sequences	OPA-02	RAPD	779	ACGGCTCGGGCGGGCAGG	CGAGCCGTGTCCCTCGGC
Banana	>gi|415426079|emb|HE813983.1|	Whole Genome Sequences	OPA-02	RAPD	752	TGCCGAGCGATGAAAGGG	GCTCGGCAACTGCATTGA
Banana	>gi|400265857|emb|HE813984.1|	Whole Genome Sequences	OPA-02	RAPD	333	ACGGCTCGCGTCTCGATG	CGAGCCGTGCGGCATTCT
Banana	>gi|400265857|emb|HE813984.1|	Whole Genome Sequences	OPA-02	RAPD	1830	TGCCGAGCTGATCACAGCAC	CAGCTCGGCATGCATGGAAC
Banana	>gi|415426089|emb|HE813985.1|	Whole Genome Sequences	OPA-02	RAPD	1248	ACGGCTCGATCGTACGTGG	CGAGCCGTTATAATGCTG

Figure 2. Snapshot view of sample output for banana RAPDOPA02.

SCAR primers (OPA-02SCAR272 and OPA-02SCAR266) were tested with genomic DNA samples of banana in wet-lab PCR experiments initially with standard primers (figure 1 in electronic supplementary material at http://ias.ac.in/jgenet/) then by SCAR primer (figure 2 in electronic supplementary material).

The genome databases and references for the crops under study are listed in table 1. These databases currently do not have predicted SCAR markers for the eight crops. Plant Markers, an online database of predicted markers for plants (Rudd *et al.* 2005) does not support SCAR markers.

Figure 3. Start page of FViSCARdb.

Table 1. Genome databases and WGS report for species under study.

Crop	Genome database name / reference	URL	Whole genome sequence reference
Apple	Jung *et al.* (2014) GDR genome database for Rosaceae	http://www.rosaceae.org/	Velasco *et al.* (2010)
Banana	Droc *et al.* (2013) Banana genome hub	http://banana-genome.cirad.fr/	D'Hont *et al.* (2012)
Banana, cacao	Hamelin *et al.* (2013) TropgeneDB	http://tropgenedb.cirad.fr	
Cacao	Cacao genome database Cacao gen DB	http://www.cacaogenomedb.org/ http://cocoagendb.cirad.fr/	Argout *et al.* (2011)
Melon	International cucurbit genome database ICuGI	http://www.icugi.org/cgi-bin/ICuGI/index.cgi	Garcia-Mas *et al.* (2012)
Papaya, tomato, potato	Plant genome database	http://www.plantgdb.org/	Ming *et al.* (2008)
Potato, tomato	Pujar *et al.* (2013) SOL genomics-potato genome	http://solgenomics.net/ http://www.potatogenome.net/	Potato Genome Consortium (2011) Tomato Genome Consortium (2012)
Strawberry	Jung *et al.* (2014) GDR genome database for Rosaceae	http://www.rosaceae.org/	Shulaev *et al.* (2011)

Primers for RAPD and ISSR markers are short (10 bases) and hence it was not easy to perform BLAST in any genome database. The database described in the current report helps to convert them to pairs of 20-bases-long primer sequences. These primers can be verified in any primer BLAST analysis and validated in wet-lab experiments.

From published literature on RAPD/ISSRs in the eight crops used here, 32 publications reporting RAPD/ISSR markers were obtained (table 1 in electronic supplementary material). Twenty four of the predicted SCAR primers with published wet-lab reports of original RAPD/ISSR markers were submitted to the NCBI probe database with (accession numbers Pr032359357 to Pr032359372 and Pr032359658 to Pr032359665).

The FViSCARdb database currently includes 46,910 predicted primer pairs and is freely available at http://www.bioinfoindia.org/fv-iscardb.

Acknowledgements

We thank all the authors of published reports for providing access to data on genome sequences of the eight plants and primer sequences. The present authors acknowledge the help by Dr Tirathraj Singh, Assistant Professor (Bioinformatics) Jaypee University of Information Technology, Solan, India, in hosting the database in Bioinfoindia.org.

References

Argout X., Salse J., Aury J. M., Guiltinan M. J., Droc G., Gouzy J. *et al.* 2011 The genome of *Theobroma cacao*. *Nat. Genet.* **43**, 101–108.

D'Hont A., Denoeud F., Aury J. -M., Baurens F. -C., Carreel F., Garsmeur O. *et al.* 2012 The banana (*Musa cuminate*) genome and the evolution of monocotyledonous plants. *Nature* **488**, 213–217.

Droc G., Lariviere D., Guignon V., Yahiaoui N., This D., Garsmeur O. *et al.* 2013 The banana genome hub. Article ID bat035 (doi:10.1093/database/bat035).

Garcia-Mas J., Benjaka A., Sanseverinoa W., Bourgeoisa M., Mira G., Gonzálezb V. M. *et al.* 2012 The genome of melon (*Cucumismelo L.*) *Proc. Natl. Acad. Sci. USA* **109**, 11872–11877.

Hamelin C., Sempere G., Jouffe V. and Ruiz M. 2013 TropGeneDB, the multi-tropical crop information system updated and extended. *Nucleic Acids Res.* **41**, 1172–1175.

Joseph K., Murthy H. and Ravishankar K. 2014 Development of male-specific SCAR marker in *Garciniamorella.* (Gaertn.) Desr. *J. Genet.* **93**, 875–878.

Jung S., Ficklin S. P., Lee T., Cheng C. H., Blenda A., Zheng P. *et al.* 2014 The genome database for Rosaceae (GDR): year 10 update. *Nucleic Acids Res.* **42**, 1237–1244.

Khlestkina E. K. 2014 Molecular markers in genetic studies and breeding. *Russ. J. Genet. Appl. Res.* **4**, 236–244.

Koveza O. V. and Gostimsky S. A. 2005 Development and study of SCAR markers in pea (*Pisum sativum* L.) *Russ. J. Genet.* **41**, 1254–1261.

Li J. J., Pei G. L., Pang H. X., Bilderbeck A., Chen S. S. and Tao S. H. 2006 A new method for RAPD primers selection based on primer bias in nucleotide sequence data. *J. Biotechnol.* **126**, 415–423.

Marczewski W., Talarczyk A. and Hennig J. 2001 Development of SCAR markers linked to the *Ns* locus in potato. *Plant Breed* **120**, 88–90.

Ming R., Hou S., Feng Y., Yu Q., Dionne-Laporte A., Saw J. H. *et al.* 2008 The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya Linnaeus*). *Nature* **452**, 991–996.

Miri S. M., Mousavi A., Naghavi M. R., Mirzaii M., Talaei A. R. and Khiabani B. N. 2009 Analysis of induced mutants of salinity resistant banana (Musa cuminate cv. Dwarf Cavendish)using morphological and molecular markers. Iran. *J. Biotechnol.* **7**, 86–92.

Paran I. and Michelmore R. W. 1993 Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.* **85**, 985–993.

Premkrishnan B. V. and Arunachalam V. 2012 *In silico* RAPD priming sites in expressed sequences and iSCAR markers for oil palm. *Comp. Funct. Genomics* **2012**, 913709.

Pujar A., Menda N., Bombarely A., Edwards J. D., Strickler S. R. and Mueller L. A. 2013 From manual curation to visualization of gene families and networks across Solanaceae plant species. (doi:10.1093/database/bat028).

Potato Genome Sequencing Consortium 2011 Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195.

Rudd S., Schoof H. and Mayer K. 2005 PlantMarkers—a database of predicted molecular markers from plants. *Nucleic Acids Res.* **33**, D628–D632.

Shulaev V., Sargent D. J., Crowhurst R. N., Mockler T. C., Folkerts O., Delcher A. L. *et al.* 2011 The genome of woodland strawberry (*Fragariavesca*). *Nat. Genet.* **43**, 109–116.

Tomato Genome Consortium 2012 The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641.

Velasco R., Zharkikh A., Affourtit J., Dhingra A., Cestaro A., Kalyanaraman A. *et al.* 2010 The genome of the domesticated apple (*Malus* × *domestica* Borkh.) *Nat. Genet.* **42**, 833–839.

Williams J. G. K., Kubelik A. R., Livak K. J., Rafalski J. A. and Tingey S. V. 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**, 6531–6535.

Received 19 November 2014, in final revised form 28 February 2015; accepted 6 July 2015
Unedited version published online: 13 July 2015
Final version published online: 4 March 2016