View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Digital.CSIC

frontiers in Plant Science

TECHNOLOGY REPORT
published: 27 June 2016
doi: 10.3389/fpls.2016.00878

# PlantFuncSSR: Integrating First and Next Generation Transcriptomics for Mining of SSR-Functional Domains Markers

*Gaurav Sablok[1]*, Antonio J. Pérez-Pulido[2], Thac Do[1], Tan Y. Seong[3], Carlos S. Casimiro-Soriguer[2], Nicola La Porta[4,5,6], Peter J. Ralph[1], Andrea Squartini[7], Antonio Muñoz-Merida[8] and Jennifer A. Harikrishna[3]*

[1] Plant Functional Biology and Climate Change Cluster (C3), University of Technology, Sydney, NSW, Australia, [2] Centro Andaluz de Biología del Desarrollo (CABD-CSIC), Universidad Pablo de Olavide, Sevilla, Spain, [3] Centre for Research in Biotechnology for Agriculture and Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia, [4] Department of Sustainable Agro-Ecosystems and Bioresources, Research and Innovation Centre, Fondazione Edmund Mach, Trento, Italy, [5] MountFOR Project Centre, European Forest Institute, Trento, Italy, [6] Consiglio Nazionale delle Ricerche, Istituto per la Valorizzazione del Legno e delle Specie Arboree, Florence, Italy, [7] Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Padova, Italy, [8] CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Vila do Conde, Portugal

OPEN ACCESS

**Edited by:**
*Xiaowu Wang, Chinese Academy of Agricultural Sciences, China*

**Reviewed by:**
*Jianjun Zhao, Hebei Agricultural University, China*
*Kui Lin, Beijing Normal University, China*

***Correspondence:**
*Gaurav Sablok*
*sablokg@gmail.com*

**Specialty section:**
*This article was submitted to Plant Genetics and Genomics, a section of the journal Frontiers in Plant Science*

**Received:** *08 April 2016*
**Accepted:** *03 June 2016*
**Published:** *27 June 2016*

**Citation:**
*Sablok G, Pérez-Pulido AJ, Do T, Seong TY, Casimiro-Soriguer CS, La Porta N, Ralph PJ, Squartini A, Muñoz-Merida A and Harikrishna JA (2016) PlantFuncSSR: Integrating First and Next Generation Transcriptomics for Mining of SSR-Functional Domains Markers. Front. Plant Sci. 7:878. doi: 10.3389/fpls.2016.00878*

Analysis of repetitive DNA sequence content and divergence among the repetitive functional classes is a well-accepted approach for estimation of inter- and intra-generic differences in plant genomes. Among these elements, microsatellites, or Simple Sequence Repeats (SSRs), have been widely demonstrated as powerful genetic markers for species and varieties discrimination. We present PlantFuncSSRs platform having more than 364 plant species with more than 2 million functional SSRs. They are provided with detailed annotations for easy functional browsing of SSRs and with information on primer pairs and associated functional domains. PlantFuncSSRs can be leveraged to identify functional-based genic variability among the species of interest, which might be of particular interest in developing functional markers in plants. This comprehensive on-line portal unifies mining of SSRs from first and next generation sequencing datasets, corresponding primer pairs and associated in-depth functional annotation such as gene ontology annotation, gene interactions and its identification from reference protein databases. PlantFuncSSRs is freely accessible at: http://www.bioinfocabd.upo.es/plantssr.

**Keywords: short tandem repeats (STRs), NGS, gene ontology (GO), inter-pro, functional domains markers**

## INTRODUCTION

Identification of repetitive patterns in genomic DNA has proved to be a powerful approach to reveal diversity and to discriminate plant populations and individuals within species. Microsatellites or Simple Sequence Repeats (SSRs) formed as a result of the strand-slippage mechanism (Schlötterer and Harr, 2001) have been used widely as functional genetic markers (Studer et al., 2010), for testing genetic fidelity, genetic variability (Rahman and Rajora, 2002; Schellenbaum et al., 2008) and for population genetic studies (Sim et al., 2009). However, the previously described approaches such as by screening the small insert genomic DNA libraries (Shokeen et al., 2007) are time

consuming and not so cost effective. Furthermore, SSRs identified by such approaches have no certainty of association to the functional domains. Leveraging the computational advances, *in silico* mining approaches using transcriptomics have filled a major gap in the development of these functional classes of markers (Sablok and Shekhawat, 2008; Sablok et al., 2011), which could be potentially used for developing the markers harboring the functional domains for marker assisted gene selection, genotyping, and anchoring quantitative trait localization (QTL; Parida et al., 2010; Kujur et al., 2013) mainly due to the associative nature of the mined SSRs to the coding region variations and the associated functional variations.

Recently, several SSRs have been linked to putative functional domains; classifying them into a new class of functional markers called simple sequence repeats functional domains markers (SSR-FDMs) in model and non-model species (Yu et al., 2010; Bhattacharyya et al., 2014). Realizing the wide importance of SSRs, several online repositories and data mining tools have been developed to address the need for on-line mining of these markers in case of nuclear genomes such as PlantMarkers (Rudd et al., 2005), SSR Biome and SSR taxonomy (Jewell et al., 2006), UgMicroSatDb (Aishwarya and Sharma, 2008), MoccaDB (Plechakova et al., 2009), CicArMiSatDB (Doddamani et al., 2014), and for Coffee expressed sequence tags (ESTs) (Poncet et al., 2006) to assist the mining of the SSRs. However, there are some limitations to the previously developed tools that have restricted, in particular, the possibility to make comparisons across different datasets from different species as they either lack integration of the browsing platform with unified annotations or they are oriented toward specific species such as CicArMiSatDB (Doddamani et al., 2014), and FmMDb (B et al., 2013). In case of organelle genomes, we previously established ChloroMitoSSRDB (Sablok et al., 2013) and ChloroMitoSSRDB 2.00 (Sablok et al., 2015) to provide the large-scale access to the organelle derived markers.

Next generation sequencing (NGS) provides a cost-efficient way of transcript identification and facilitates the development of transcript based SSRs markers for model and non-model species, which has resulted in rapid increases in the data made available online. However, much of this data is scattered across numerous websites and has not been mined or annotated for the identification of functional SSRs. Recently, there have been some efforts to consolidate such data for example TropiTree[1] is a repository displaying the mined SSRs from NGS transcript assemblies for 24 tropical plants (Russell et al., 2014). Taking into account the limitations mentioned, we were motivated to develop PlantFuncSSRs, available at http://www.bioinfocabd.upo.es/plantssr, which is a unified functional SSRs portal displaying mined functional SSRs from 274 ESTs based transcript assemblies, and more than 100 NGS transcripts assemblies. PlantFuncSSRs also provides detailed primer pair information, functional annotations, and putative homologs to the transcript assemblies in Uniprot and curated SSR-FDMs in a single unified platform. We believe that the availability of the above resource will aid the rapid development of functional SSRs in non-model plant species.

[1]http://bioinf.hutton.ac.uk/tropiTree

# MATERIALS AND METHODS

## Data Resources for PlantFuncSSRs

To integrate previously published plant EST data, all Putative Unique Transcripts (PUT) representing 273 transcript assemblies were downloaded from PlantGDB (Version release 187) available from http://www.plantgdb.org/ (Dong et al., 2004). Additionally, version control 74 NGS transcriptomes available at PhytoMetaSync[2] (Facchini et al., 2012; Xiao et al., 2013), 14 medicinal plant transcriptomes available from medicinal plant genomics resource (MPGR)[3] (Góngora-Castillo et al., 2012; Góngora-Castillo and Buell, 2013) and 3 *Brachypodium sylvaticum* transcriptomes available from http://jaiswallab.cgrb.oregonstate.edu/genomics (Fox et al., 2013) were downloaded, representing a total of 364 plant species.

## SSRs Identification and Functional Assignments

For systematic identification of SSR, all the transcripts (ESTs as well as NGS) assemblies were first scanned for the presence of the homopolymer errors and sequence ambiguity was removed using the est_trimmer tool available at: http://pgrc.ipk-gatersleben.de/misa/download/est_trimmer.pl with the following settings: -amb = 2.50 -tr5 = *T*, 5.50 -tr3 = *A*, 5.50. Following the transcript ambiguity removal and trimming of the homopolymer runs, MISA (MIcroSAtellite identification tool) (Thiel et al., 2003) was deployed to identify the microsatellites. In the present version of the PlantFuncSSRs, we classified microsatellites as repetitive stretches of motifs of a minimum and 12-mer repetitive stretch as mono-, 6-mer repetitive stretches as di, 4-mer

[2]http://www.phytometasyn.ca

[3]http://medicinalplantgenomics.msu.edu

**TABLE 1 | Table describing the classified repeats types and embedded functional categories in PlantFuncSSRs.**

| Type of SSRs | Number of SSRs |
|---|---|
| P1 SSRs-FDMs | 221008 |
| P2 SSRs-FDMs | 200702 |
| P3 SSRs-FDMs | 1067949 |
| P4 SSRs-FDMs | 358245 |
| P5 SSRs-FDMs | 102593 |
| P6 SSRs-FDMs | 142452 |
| Compound SSRs-FDMs (C and C*) | 292472 |
| Functionally embedded SSRs annotations | |
| Gene names | 2278574 |
| Descriptions | 2332906 |
| Gene ontologies | 1986736 |
| Uniprot (keywords) | 2122976 |
| InterPro domains | 2172553 |

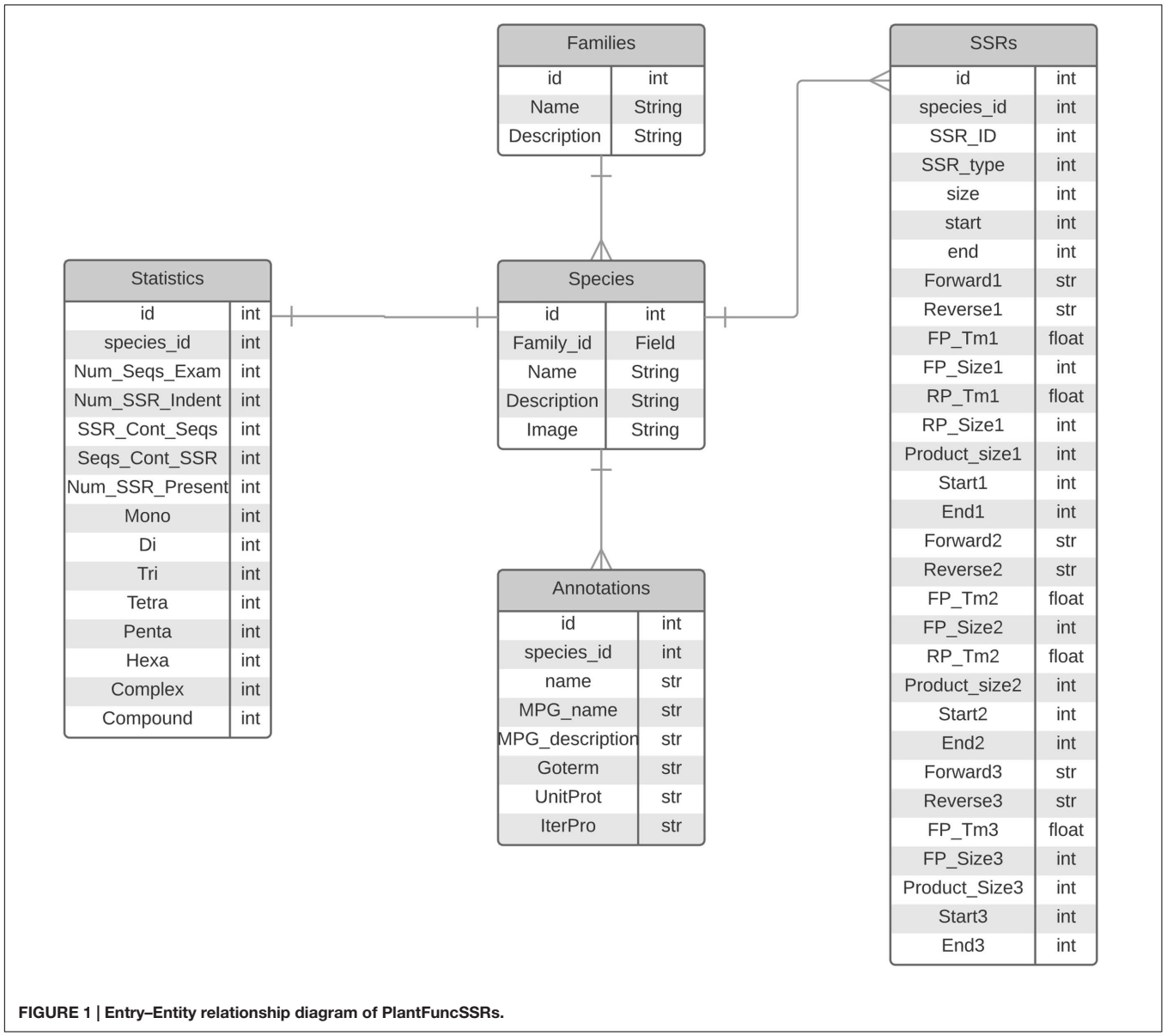



**FIGURE 1 | Entry–Entity relationship diagram of PlantFuncSSRs.**

repetitive stretches of tri- and tetra-, and a minimum of 3-mer repetitive stretch as penta- and hexa-nucleotide. Additionally, the identified SSRs have been classified into perfect and compound repeats, with compound repeats interrupted by a minimum of 100 bp as previously described (Victoria et al., 2011). Primer pairs were designed for all of the identified SSRs using primer3 available from primer3.sourceforge.net (Untergasser et al., 2012) using the settings as described in MISA (Thiel et al., 2003).

Following SSRs identification, in-depth functional annotation of the identified SSRs was carried out using the standalone annotator Sma3s (Muñoz-Mérida et al., 2014), which uses the plant taxonomic division set in the Uniprot database[4], including both Swiss-Prot and TrEMBL sections to enrich the final annotation. The annotations gave the found Gene Ontology (GO) terms which were subsequently linked to their GO_SLIM terms using the plant GO slim available from www.geneontology.org, in order to simplify the GO terms and allow cross-comparison. In this way, each SSRs sequence was identified with the more probable gene name and description, as well as both GO terms from the existing three categories and Swiss-Prot keywords, all of them for cataloging the SSRs and assigning functional domains. The IntAct annotations and Interactions were cross-linked using the IntAct resources available from EBI at: http://www.ebi.ac.uk/intact/. The functional SSRs annotation also includes putative InterPro domains (Quevillon et al., 2005, pathways from UniProt to have more details of the involved biological processes. PlantFuncSSRs presents only those SSRs, which have functional annotations appended to them and are thus termed as SSR-functional markers.

[4]http://www.uniprot.org

**FIGURE 2 | PlantFuncSSRs: Schematic view of the PlantFuncSSRs and the browsing options implemented in PlantFuncSSRs.** The structure of PlantFuncSSRs allows for the browsing of the functional SSRs either according to the species or according to the family.

# RESULTS AND DISCUSSION

## PlantFuncSSRs Architecture and Visualization

Expressed sequence tags and NGS based Transcriptome reconstruction represent the functional portion of the genome and have been widely used as resources to mine and develop functional markers. Developing an efficient browsing system for the mining of repeats is an important task, as this can be widely applied to a wide range of on-going plant breeding and crop improvement research. To develop an efficient browsing system, PlantFuncSSRs architecture has been developed using Ruby Rails and MySQL, which provides faster integration and query based searches to the users. The current version of the PlantFuncSSRs presents more than 2 million SSRs and SSR-FDMs from 364 species for easy access and browsing of transcript derived plant SSRs across the plant kingdom (**Table 1**). These species are ranging from important crops to wild species, from

A

PlantFuncSSRs Welcome Overview of Functional SSRs Browse Microsatellites Tutorial Contact Us

Search by Species: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

B

Copy Excel CSV Show 10 entries

| Specie Name | Num Seqs Exam | Size Exam Seqs | Num SSR Indent | SSR Cont Seqs | Seqs Cont SSR | Num SSR Present | Get Repeat |
|---|---|---|---|---|---|---|---|
| Abies_balsamea_B1 | 82816 | 78979398 | 13201 | 10521 | 1959 | 923 | Get Repeat |
| Acacia_victoriae_RI1 | 139945 | 97632297 | 20324 | 16796 | 2719 | 1603 | Get Repeat |
| Actinidia_chinensis | 18734 | 11687951 | 7289 | 5408 | 1401 | 1112 | Get Repeat |

Abies_balsamea_B1

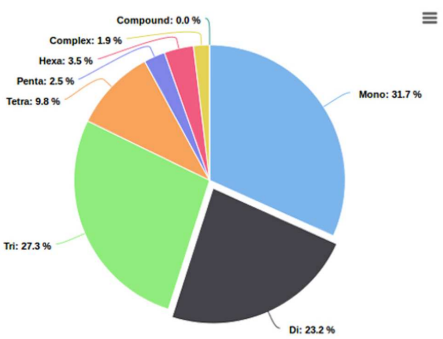
Summary Statistics

| | |
|---|---|
| Total number of sequences examined: | 82816 |
| Total size of examined sequences (bp): | 78979398 |
| Total number of identified SSRs: | 13201 |
| Number of SSR containing sequences: | 10521 |
| Number of sequences containing more than 1 SSR: | 1959 |
| Number of SSRs present in compound formation: | 923 |

Summary Perfect Repeats

| SSR Type | Numbers | Details |
|---|---|---|
| Mono | 130 | Click for Repeats |
| Di | 361 | Click for Repeats |
| Tri | 3210 | Click for Repeats |
| Tetra | 1694 | Click for Repeats |
| Penta | 481 | Click for Repeats |
| Hexa | 733 | Click for Repeats |
| Complex | 446 | Click for Repeats |
| Compound | 22 | Click for Repeats |
| **Total** | **7077** | Click for Repeats |

C




Copy Excel CSV Show 50 entries

| Sl_No | SSR ID | SSR Type | SSR | SSR Size | SSR Start | SSR End | Primer | Annotation |
|---|---|---|---|---|---|---|---|---|
| 1 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp100101_c0_seq4 | Hexa | (AGGAGC)3 | 18 | 98 | 115 | Primers | Annotation |
| 2 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp101010_c2_seq2 | Tri | (TTA)4 | 12 | 364 | 375 | Primers | Annotation |
| 3 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp101917_c0_seq2 | Hexa | (ATGGCA)4 | 24 | 502 | 525 | Primers | Annotation |
| 4 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp101972_c0_seq1 | Hexa | (CAGCAA)3 | 18 | 6327 | 6344 | Primers | Annotation |
| 5 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp102029_c0_seq1 | Tri | (TCT)4 | 12 | 780 | 791 | Primers | Annotation |
| 6 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp106177_c0_seq1 | Tri | (AGA)4 | 12 | 413 | 424 | Primers | Annotation |
| 7 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp106177_c0_seq1 | Tri | (GAG)5 | 15 | 714 | 728 | Primers | Annotation |
| 8 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp108456_c0_seq1 | Tetra | (AATA)3 | 12 | 1442 | 1453 | Primers | Annotation |
| 9 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp40044_c0_seq1 | Tetra | (TGCC)3 | 12 | 1337 | 1348 | Primers | Annotation |
| 10 | gnl\|MAGPIE\|aba.ABB1JB_Trinity--comp67326_c0_seq1 | Tetra | (GGAT)3 | 12 | 3664 | 3675 | Primers | Annotation |

**FIGURE 3 | Alphabet sorting of the species names and search patterns (A); Species specific page showing the information on the identified Simple Sequence Repeats (SSRs) and also the functional SSRs.** "Click for repeats" pages are directly hyperlinked to the functional SSRs **(B)**; Weblayout describing the functional repeats identified in the respective plant species with information on type of repeat, classification of repeat, size, motif, start, and end coordinates and associated primers and functional annotation **(C)**.

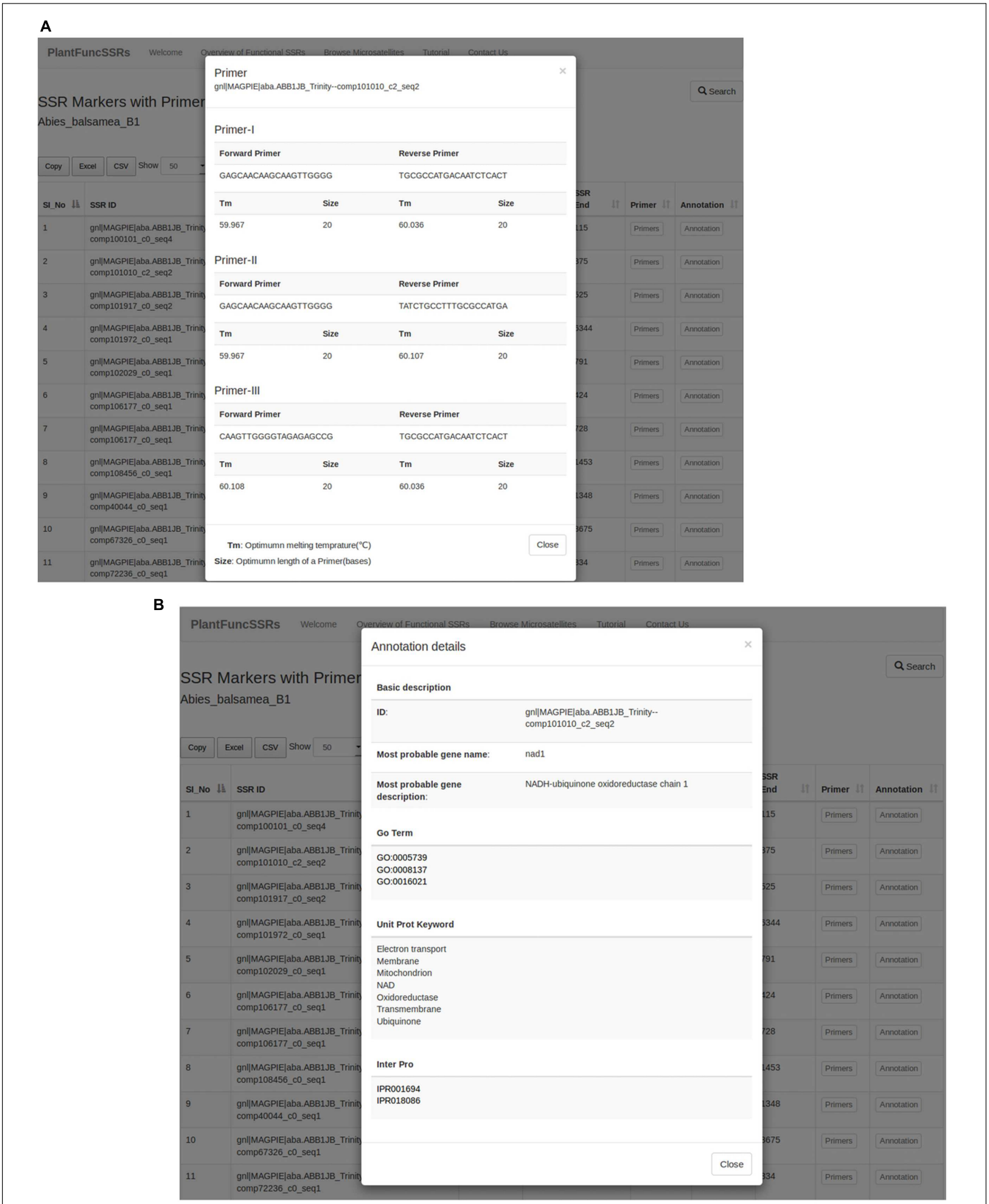



**FIGURE 4 | Pop-up Primer display window for the user selected functional SSRs (A); Pop-up window showing high throughput functional annotation for the user selected functional SSRs (B).**

mono- to di-cots, from annual to polyannual and wood species. Integration of visualization features with the rapid mining of the data is a key central feature that has been implemented in the PlantFuncSSRs. A schema of the database architecture in the form of entity-relationship is given in **Figure 1**. For the visualization of the SSRs and the associated information, several hierarchal levels of classified information have been inter-linked in PlantFuncSSRs (**Figure 2**). The front-end portal is user-friendly and allows the end-users to search SSRs as "species-wise", "family wise", or "advanced search menu" (**Figure 2**). A quick search implementation pattern displays the embedded species information in quick select "*species*" and "*families*", which are hyperlinked pages to the respective species and provide a quick view of the functional SSRs present in each species. **Figure 3** shows the webpage browsing of PlantFuncSSRs with detailed classification of the identified SSRs for user-selected species of interest. Alphabetical classification of the species provides an additional advantage for the users to quickly look for their species of interest (**Figure 3**).

Each record in the species displays the *Species_Name, Num_Seqs_Exam, Size_Exam_Seqs, Num_SSR_Ident, SSR_Cont_Seqs, Seqs_Cont_SSR,* and *Num_SSR_Present* providing summarized information on the number of the identified SSRs for that particular species of interest lined to the primer pair information and high throughput functional annotation (**Figure 3**). In PlantFuncSSRs, each species page has been hyperlinked to the corresponding repeat information pages that present detailed information on several statistics such as total number of sequences examined, total size of examined sequences (bp), total number of identified SSRs, number of SSR containing sequences, number of sequences containing more than one SSR and compound SSRs (**Figure 3**). In addition, to this summary information, each species classified page also details the types and distribution of the repeats in tabular format, which can be sorted "*on the fly*". An integral part of PlantFuncSSRs is to describe the associated primer pair information for each species to facilitate the development of functional SSRs for diversity analysis. To augment such capacity, each functional SSR has been associated with primer pages and detailed functional annotations, which describes the set of the "ready to use" primers for the functional validation of the corresponding SSRs (**Figure 4**).

## Functional SSRs and Functional Importance of PlantFuncSSRs

Microsatellites (SSRs) have been shown to be regulators of a number of plant genes demonstrating their importance as key players in regulating plant function (Faville et al., 2004). FuncPlantSSRs offers a wide variety of functional annotations for the identified SSRs such as GO terms, GO slim categories, pathways, descriptions to identify the sequences and comparing with putative homologues, and motif and domain modules to offer the domain architecture for the sequences. Recently, increasing interest toward the functional linkage of the markers to the domain association and function can be seen from several recent reports in plants such as *Ocimum basilicum* (Gupta et al., 2010), *Seasmum indicum* (Bhattacharyya et al., 2014), *Elaeis guineensis* (Tranbarger et al., 2012), and *Camellia sinensis* (Sahu et al., 2012) suggesting the role of the functional SSRs as important markers for developing the functional genic approaches for marker enrichment in plants. Nonetheless, established reports of the functional association of the repeats with the catalytic domains (Parida et al., 2010; Yu et al., 2010) has been widely developed. For quick advanced searches, PlantFuncSSRs offer several functionalities, such as searches customized and optimized on various hierarchal levels i.e., Family, Species, Type of Repeat, Number of Repeat, Functional annotation, GO annotation, and IPR annotations (**Figure 2**). Availability of the curated information provides end users with the flexibility to narrow their searches to functional SSRs linked to specific categories, motif types or functional annotations. Taking into account the vast amount of the species coverage and associated functional SSRs present in the PlantFuncSSRs, we believe that the PlantFuncSSRs provides access to the most comprehensive catalog available for the functional SSRs from plant transcriptomes.

# CONCLUSION

In the present version of the PlantFuncSSRs, we bring together under a unified portal the mining of the SSRs from the publically available first and second generation datasets. PlantFunctSSRs has been designed with an aim to serve as a stand-alone single access platform for the analysis of functional SSRs from first and NGS datasets for a large number of sequenced plant transcriptomes. In addition to providing the most comprehensive available resource for exploring and validating plant functional SSRs, the built in annotation platform will allow the users to have wide access to the functional relevance of the validated SSRs thus provides a valuable functional SSRs resource to support plant diversity, population and functional marker research.

# AUTHOR CONTRIBUTIONS

GS conceived and designed the research, identified SSRs and linked the SSRs to functions, AP and AM-M provided the annotation, TD build the database and the web-interface, TYS helped in the data integration, CSCS hosted the database, GS wrote the manuscript, NP, AS, PR, and JAH provided revisions. All authors have read and approved the manuscript.

# ACKNOWLEDGMENTS

Gaurav Sablok thanks Plant Functional Biology and Climate Change Cluster (C3), University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia, for providing the computational facilities. An internal grant number to GS (2226018) supported this work. JAH and TYS were partially supported by High Impact Research Chancellory Grant UM.C/625/1/HIR/MOHE/SCI/19 from the University of Malaya.

## REFERENCES


Aishwarya, V., and Sharma, P. C. (2008). UgMicroSatdb: database for mining microsatellites from unigenes. *Nucleic Acids Res.* 36, D53–D56. doi: 10.1093/nar/gkm811

B, V. S., Muthamilarasan, M., Misra, G., and Prasad, M. (2013). FmMDb: a versatile database of foxtail millet markers for millets and bioenergy grasses research. *PLoS ONE* 8:e71418. doi: 10.1371/journal.pone.0071418

Bhattacharyya, U., Pandey, S. K., and Dasgupta, T. (2014). Identification of EST-SSRs and FDM in sesame (*Sesamum indicum* L.) through data mining. *Sch. J. Agric. Sci.* 4, 60–69.

Doddamani, D., Katta, M. A., Khan, A. W., Agarwal, G., Shah, T. M., and Varshney, R. K. (2014). CicArMiSatDB: the chickpea microsatellite database. *BMC Bioinformatics* 15:212. doi: 10.1186/1471-2105-15-212

Dong, Q., Schlueter, S. D., and Brendel, V. (2004). PlantGDB, plant genome database and analysis tools. *Nucleic Acids Res.* 32, D354–D359. doi: 10.1093/nar/gkh046

Facchini, P. J., Bohlmann, J., Covello, P. S., De Luca, V., Mahadevan, R., Page, J. E., et al. (2012). Synthetic biosystems for the production of high-value plant metabolites. *Trends Biotechnol.* 30, 127–131. doi: 10.1016/j.tibtech.2011.10.001

Faville, M. J., Vecchies, A. C., Schreiber, M., Drayton, M. C., Hughes, L. J., Jones, E. S., et al. (2004). Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). *Theor. Appl. Genet.* 110, 12–32. doi: 10.1007/s00122-005-1959-y

Fox, S. E., Preece, J., Kimbrel, J. A., Marchini, G. L., Sage, A., Youens-Clark, K., et al. (2013). Sequencing and de novo transcriptome assembly of *Brachypodium sylvaticum* (Poaceae). *Appl. Plant Sci.* 1:1200011. doi: 10.3732/apps.1200011

Góngora-Castillo, E., and Buell, C. R. (2013). Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Nat. Prod. Rep.* 30, 490–500. doi: 10.1039/c3np20099j

Góngora-Castillo, E., Childs, K. L., Fedewa, G., Hamilton, J. P., Liscombe, D. K., Magallanes-Lundback, M., et al. (2012). Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS ONE* 7:e52506. doi: 10.1371/journal.pone.0052506

Gupta, S., Shukla, R., Roy, S., Naresh, S., and Sharma, A. (2010). In silico SSR and FDM analysis through EST sequences in *Ocimum basilicum*. *Plant Omics J.* 3, 121–128.

Jewell, E., Robinson, A., Savage, D., Erwin, T., Love, C. G., Lim, G. A. C., et al. (2006). SSRPrimer and SSR taxonomy tree: biome SSR discovery. *Nucleic Acids Res.* 34, W656–W659. doi: 10.1093/nar/gkl083

Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2013). Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* 20, 355–373. doi: 10.1093/dnares/dst015

Muñoz-Mérida, A., Viguera, E., Claros, M. G., Trelles, O., and Pérez-Pulido, A. J. (2014). Sma3s: a three-step modular annotator for large sequence datasets. *DNA Res.* 21, 341–353. doi: 10.1093/dnares/dsu001

Parida, S. K., Pandit, A., Gaikwad, K., Sharma, T. R., Srivastava, P. S., Singh, N. K., et al. (2010). Functionally relevant microsatellites in sugarcane unigenes. *BMC Plant Biol.* 10:251. doi: 10.1186/1471-2229-10-251

Plechakova, O., Tranchant-Dubreuil, C., Benedet, F., Couderc, M., Tinaut, A., Viader, V., et al. (2009). MoccaDB - an integrative database for functional, comparative and diversity studies in the Rubiaceae family. *BMC Plant Biol.* 9:123. doi: 10.1186/1471-2229-9-123

Poncet, V., Rondeau, M., Tranchant, C., Cayrel, A., Hamon, S., de Kochko, A., et al. (2006). SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the Coffea genus. *Mol. Genet. Genom.* 6, 436–449. doi: 10.1007/s00438-006-0153-5

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33, 116–120. doi: 10.1093/nar/gki442

Rahman, M. H., and Rajora, O. P. (2002). Microsatellite DNA fingerprinting, differentiation, and genetic relationships of clones, cultivars, and varieties of six poplar species from three sections of the genus *Populus*. *Genome* 45, 1083–1094. doi: 10.1139/g02-077

Rudd, S., Schoof, H., and Mayer, K. (2005). PlantMarkers–a database of predicted molecular markers from plants. *Nucleic Acids Res.* 33, D628–D632. doi: 10.1093/nar/gki074

Russell, J. R., Hedley, P. E., Cardle, L., Dancey, S., Morris, J., Booth, A., et al. (2014). TropiTree: an NGS-based EST-SSR resource for 24 tropical tree species. *PLoS ONE* 9:e102502. doi: 10.1371/journal.pone.0102502

Sablok, G., Luo, C., Lee, W. S., Rahman, F., Tatarinova, T. V., Harikrishna, J. A., et al. (2011). Bioinformatic analysis of fruit-specific expressed sequence tag libraries of *Diospyros kaki* Thunb: view at the transcriptome at different developmental stages. *Biotech* 1, 35–45.

Sablok, G., Mudunuri, S. B., Patnana, S., Popova, M., Fares, M. A., and Porta, N. L. (2013). ChloroMitoSSRDB: open source repository of perfect and imperfect repeats in organelle genomes for evolutionary genomics. *DNA Res.* 20, 127–133. doi: 10.1093/dnares/dss038

Sablok, G., Padma Raju, G. V., Mudunuri, S. B., Prabha, R., Singh, D. P., Baev, V., et al. (2015). ChloroMitoSSRDB 2.00: more genomes, more repeats, unifying SSRs search patterns and on-the-fly repeat detection. *Database (Oxford)* 2015:bav084.

Sablok, G., and Shekhawat, N. S. (2008). Bioinformatics analysis of distribution of microsatellite markers (SSRs)/single nucleotide polymorphism (SNPs) in expressed transcripts of *Prosopis juliflora*: frequency and distribution. *J. Comput. Sci. Syst. Biol.* 1, 87–91.

Sahu, J., Sarmah, R., Dehury, B., Sarma, K., Sahoo, S., Sahu, M., et al. (2012). Mining for SSRs and FDMs from expressed sequence tags of *Camellia sinensis*. *Bioinformation* 8, 260–266. doi: 10.6026/97320630008260

Schellenbaum, P., Mohler, V., Wenzel, G., and Walter, B. (2008). Variation in DNA methylation patterns of grapevine somaclones (*Vitis vinifera* L.). *BMC Plant Biol.* 8:78. doi: 10.1186/1471-2229-8-78

Schlötterer, C., and Harr, B. (2001). "Microsatellite instability," in *Encyclopedia of Life Sciences* (New York, NY: John Wiley & Sons, Ltd), 1–4.

Shokeen, B., Sethy, N. K., Kumar, S., and Bhatia, S. (2007). Isolation and characterization of microsatellite markers for analysis of molecular variation in the medicinal plant Madagascar periwinkle (*Catharanthus roseus* (L.) G. Don). *Plant Sci.* 172, 441–451. doi: 10.1016/j.plantsci.2006.10.010

Sim, S. C., Yu, J. K., Jo, Y. K., Sorrells, M. E., and Jung, G. (2009). Transferability of cereal EST-SSR markers to ryegrass. *Genome* 52, 431–437. doi: 10.1139/g09-019

Studer, B., Kölliker, R., Muylle, H., Asp, T., Frei, U., Roldán-Ruiz, I., (2010). EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (*Lolium* spp.). *BMC Plant Biol.* 10:177. doi: 10.1186/1471-2229-10-177

Thiel, T., Michalek, W., Varshney, R., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0

Tranbarger, T. J., Kluabmongkol, W., Sangsrakru, D., Morcillo, F., Tregear, J. W., Tragoonrung, S., et al. (2012). SSR markers in transcripts of genes linked to post-transcriptional and transcriptional regulatory functions during vegetative and reproductive development of *Elaeis guineensis*. *BMC Plant Biol.* 12:1. doi: 10.1186/1471-2229-12-1

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 40, e115. doi: 10.1093/nar/gks596

Victoria, F. C., da Maia, L. C., and de Oliveira, A. C. (2011). In silico comparative analysis of SSR markers in plants. *BMC Plant Biol.* 11:15. doi: 10.1186/1471-2229-11-15

Xiao, M., Zhang, Y., Chen, X., Lee, E. J., Barber, C., Chakrabarty, R., et al. (2013). Tanscriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* 166, 122–134. doi: 10.1016/j.jbiotec.2013.04.004

Yu, J. K., Paik, H., Choi, J. P., Han, J. H., Choe, J. K., and Hur, C. G. (2010). Functional domain marker (FDM): an in silico

demonstration in Solanaceae using simple sequence repeats (SSRs). *Plant Mol. Biol. Rep.* 28, 352–356. doi: 10.1007/s11105-009-0154-8

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2016 Sablok, Pérez-Pulido, Do, Seong, Casimiro-Soriguer, La Porta, Ralph, Squartini, Muñoz-Merida and Harikrishna. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*