

Genetics and population analysis

# KASPspoon: an *in vitro* and *in silico* PCR analysis tool for high-throughput SNP genotyping

Alsamman M. Alsamman<sup>1,\*</sup>, Shafik D. Ibrahim<sup>1</sup> and Aladdin Hamwiah<sup>2,\*</sup>

<sup>1</sup>Department of Genome Mapping, Molecular Genetics and Genome Mapping Laboratory, Agricultural Genetic Engineering Research Institute, Giza, Egypt and <sup>2</sup>Department of Biotechnology, International Center for Agricultural Research in the Dry Areas (ICARDA), Cairo, Egypt

\*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on March 22, 2018; revised on December 15, 2018; editorial decision on December 29, 2018; accepted on January 4, 2019

## Abstract

**Motivation:** Fine mapping becomes a routine trial following quantitative trait loci (QTL) mapping studies to shrink the size of genomic segments underlying causal variants. The availability of whole genome sequences can facilitate the development of high marker density and predict gene content in genomic segments of interest. Correlations between genetic and physical positions of these loci require handling of different experimental genetic data types, and ultimately converting them into positioning markers using a routine and efficient tool.

**Results:** To convert classical QTL markers into KASP assay primers, KASPspoon simulates a PCR by running an approximate-match searching analysis on user-entered primer pairs against the provided sequences, and then comparing *in vitro* and *in silico* PCR results. KASPspoon reports amplicons close to or adjoining genes/SNPs/simple sequence repeats and those that are shared between *in vitro* and *in silico* PCR results to select the most appropriate amplicons for gene discovery. KASPspoon compares physical and genetic maps, and reports the primer set genome coverage for PCR-walking. KASPspoon could be used to design KASP assay primers to convert QTL acquired by classical molecular markers into high-throughput genotyping assays and to provide major SNP resource for the dissection of genotypic and phenotypic variation. In addition to human-readable output files, KASPspoon creates Circos configurations that illustrate different *in silico* and *in vitro* results.

**Availability and implementation:** Code available under GNU GPL at (<http://www.ageri.sci.eg/index.php/facilities-services/ageri-sofwares/kaspspoon>).

**Contact:** smahmoud@ageri.sci.eg or a.hamwiah@cgiar.org

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

## 1 Introduction

Classically acquired quantitative trait loci (QTL) have been introduced using numerous field trials and laboratory experiments with significant correlations with economical useful traits. The expansion in genome sequencing avenues has provided public genome

repositories with massive amounts of biological information regarding different genome sequences and single nucleotide variations (Doddamani *et al.*, 2015). Simultaneously, advanced molecular marker technologies provide extremely high levels of assay robustness and accuracy with significant cost savings. An example of these technologies, the KASP assay (He *et al.*, 2014), has been efficiently

used to detect and validate nucleotide variations, single nucleotide polymorphisms (SNPs)/InDels, related to important traits across different organisms (Kolmer et al., 2018).

As a result, harvesting more gain from published QTL across different organisms requires correlations between genetic and physical positions of these loci, handling different experimental genetic data types, and ultimately converting them into positioning markers using a routine and fast computational tool. *In silico* PCR is one of the most widely used techniques to determine the physical position of these loci. *In silico* PCR is a computational procedure that estimates PCR results theoretically using a given set of primers to amplify DNA sequences from a sequenced genome or transcriptome (Lexa et al., 2001).

In recent years, a plethora of software programs have been developed to aid *in silico* PCR analysis, including Primer-BLAST (Ye et al., 2012), SNPCheck (<http://ngl.man.ac.uk>), FastPCR (Kalendar et al., 2011), Primersearch-EMBOSS (Rice et al., 2000) and PUNS (Boutros and Okey, 2004). Nevertheless, researchers have a long journey when decreasing and correlating prior or subsequent information regarding the genetic data of these markers to obtain conclusive results they can employ.

This study presents a PCR primer test application, called ‘KASPspoon,’ for routine manipulation and analysis of PCR primers. The final and main goal of KASPspoon is converting classically acquired QTL information into more comprehensive, routine and accurate molecular marker technologies such as a KASP assay. To reach this goal, KASPspoon can be used efficiently to compare *in vitro* (laboratory observed) and *in silico* (predicted) PCR results, and reports SNPs that are close to or adjoined by PCR amplimers by integrating a database for known SNPs. KASPspoon uses this information to design KASP assay primers to convert QTL markers into high-throughput genotyping to provide major SNP markers for dissection of genotypic and phenotypic variation. It also reports chromosomal-specific primers (anchored) and compares between physical and genetic maps, which could be useful for linkage and association mapping analysis. In addition, primer set genome coverage can support genome-walking PCR procedures that cover gene-rich genomic regions. Also, by processing simple sequence repeat (SSR) markers, KASPspoon can report amplimers that adjoin SSRs to analyze comparisons between observed and predicted motifs and to use their abundance in genome for more accurate primer selectivity.

Furthermore, KASPspoon produces different Circos configurations to illustrate *in silico* results, which can be easily handled through Circos software package (Krzywinski et al., 2009). This helps users to visualize results, unify biological data results for multiple analysis tools, magnify and exclude any part of the results, and develop further analysis techniques using simple programming (i.e. Circos scripting) without handling the original source code.

## 2 Design and implementation

KASPspoon handles a different experimental genetic data type, correlates between the genetic and physical position of loci, and generates ultimate positioning markers for KASP assay. Snapshots of KASPspoon outputs are shown in [Supplementary File S1](#).

KASPspoon was developed as a stand-alone package for PCR primer analysis using both C and Perl programming languages. The Boyer–Moore–Horspool (Horspool, 1980) and Baeza–Yates–Perleberg (Baeza–Yates and Perleberg, 1992) string approximate-matching algorithms were used through C to search genomic sequences provided PCR primer-pair sequences were used as queries.

For primer genome coverage statistics and PCR-walking procedure, the overlap layout consensus algorithm with a user-defined gap between amplimers is used to report primer(s) covered area ([Supplementary File S1](#)).

KASPspoon uses common biological data formats as an input and only amplimers that do not exceed user-defined maximum primer mismatch or maximum total mismatch (forward primer mismatch + reverse primer mismatch), or those that do not have a mismatch in the first user-defined 3′ nucleotides are reported ([Supplementary File S1](#)). KASPspoon can compare *in silico* (predict PCR product size) and *in vitro* PCR results (observed PCR product size) by defining the maximum molecular weight mismatch between the *in silico* and *in vitro* amplimers if the *in vitro* PCR product length (in base pairs) is provided (approximately). For comparing *in silico* and *in vitro* PCR amplimers, KASPspoon generates a text file containing amplimers that exist in both and that contain (or do not) genes. In addition, MISA Perl script ([pgrc.ipk-gatersleben.de/misa/](http://pgrc.ipk-gatersleben.de/misa/)) is integrated inside KASPspoon tool to report all SSRs that lie between the PCR-amplified regions.

When genetic linkage map is provided, KASPspoon can compare between the physical (bp) and genetic (cM) positions for PCR markers, where information provided by *in silico* PCR analysis will be used to assign genetic linkage group(s) to physical chromosome(s).

If a list of SNP variations is provided, KASPspoon can generate KASP assay primers that can be used for SNP genotyping. These KASP primers are designed to target all SNPs that are nearby PCR-amplified chromosomal regions. Report files that contain all KASP-targeted genes and marker loci are generated. The KASP sequences are designed according to a KASP primer design manual published by LGC ([www.lgcgroup.com](http://www.lgcgroup.com)). Primer3 tool (Untergasser et al., 2012) was used to design two allele-specific forward primers, and a common reverse primer for allele-specific assays such as KASP assay. These primers designed by KASPspoon use a user-provided SNP database to create degenerate PCR primers to provide primers with minimal mismatches, where the target nucleotide is marked by ‘[ ]’ and untargeted nucleotides are masked according to IUPAC codes.

KASPspoon will generate different Circos configurations for *in silico* PCR statistical results, comparison between *in silico* and *in vitro* PCR data and linkages, and *in silico* (physical) maps.

The search returns a sequence output file in FASTA format, containing all sequences in the database that lie between, and include, the primer pair. FASTA header describes the region in the database and primer(s) names. Comma-separated output files generated by KASPspoon include:

1. *in silico* PCR-generated amplimer information
2. *in silico* amplimers near/adjoining genes
3. SNPs near *in silico* amplimers in VCF
4. *in silico* amplimers adjoining SSRs
5. *in vitro* and *in silico* amplimers acquired by the same PCR primer that share the same approximate band size
6. location of genes adjoining or close to PCR primer regions
7. genomic areas that are covered using this primer set
8. primer set coverage statistics with both base-pair and percentage scales (compared to length of total genome sequence covered and chromosomal sequence length)
9. chromosomal assignment for linkage genetic groups
10. KASP primer(s) sequence and information
11. final report files containing different information about this run in an abbreviated form.

**Table 1.** Comparison between KASPSpoon and published *in silico* PCR tools

Comparison	KASPSpoon	Primer-BLAST	SNPCheck	FastPCR	Primersearch
Anchored primer reporting	Yes	No	No	Yes	No
Unrestricted list of genomes	Yes	No	No	Yes	Yes
Noncommercial	Yes	Yes	No	No	Yes
Gene reporting	Yes	Yes	No	Yes	No
<i>In silico</i> <i>in vitro</i> PCR comparison	Yes	No	No	No	No
<i>In silico</i> PCR speed (primer/s) <sup>a</sup>	7	5	2	90	12
KASP primer design	Yes	No	No	No	No
Position-depend mispairing weight	Yes	No	No	No	No
Linkage/physical maps comparison	Yes	No	No	No	No
Local installation	Yes	Yes	No	Yes	Yes
Primer batch analysis	Yes	No	Yes	Yes	Yes
Primer mismatch parameters	Yes	Yes	No	Yes	Yes
Primer coverage	Yes	No	No	No	No
Output graphical illustration	Yes	No	Yes	No	No
SNP reporting	Yes	No	Yes	No	No
Source code availability	Yes	No	No	No	Yes
SSR reporting	Yes	No	No	Yes	No
Table-formatted outputs	Yes	No	Yes	Yes	No
GUI	Yes	Yes	Yes	Yes	Yes
Multiple operating systems	Yes	Yes	Yes	No	Yes

<sup>a</sup>Speed calculated by processing human Chr1 Mohajeri *et al.* (2016).

### 3 Results

Although KASPSpoon showed a medium processing speed compared to tools uses BLAST as a search engine such as Primer-BLAST and SNPCheck, KASPSpoon provides several additional advantages that are lacking in some or all published tools, such as being free, source code availability, primer coverage, *in silico* and *in vitro* PCR comparison, SNP or SSR or anchor primer reporting, physical and linkage map comparison, output graphical illustration, KASP primer design and human readable and easy manipulated outputs (Table 1).

The 462 previously published chickpea SSR markers (Supplementary File S2) produced 892 bands, covering 168 082 bp (0.048%) of the chromosomal genome, in which 89.6% were chromosome-specific and 24% had genes nearby (Supplementary Files S1 and S3). Published chickpea linkage map (Nayak *et al.*, 2010) was used to compare the genetic and *in silico* map position for 241 primers. Most of these markers were successfully assigned to corresponding chromosomes as assumed by Nayak *et al.* (2010), and others had a high number of markers belonging to other chromosomes. Chickpea SNP database (Doddamani *et al.*, 2015) was used to detect SNPs nearby and to design KASP primers (Supplementary File S4) and investigate SNP effects using SnpEff tool (Cingolani *et al.*, 2012). About 99.41% of the detected SNPs were 'MODIFIER.' Two had a 'HIGH' impact on two chickpea uncharacterized proteins.

### 4 Conclusion

KASPSpoon could be successfully integrated in different genomics procedures such as primer design, genome mapping, QTL fine mapping, genome wide association analysis, PCR-walking and SNP genotyping. Combining KASPSpoon with SNP selection programs such as SnpEff could decrease the number of SNPs used for KASP assay primer design. Adding more than one genome in one basket for *in silico* PCR could help to select potential polymorphic markers through genomes. Circos configurations will help to give a grand overview regarding QTL chromosomal position and closeness to genes or SNPs.

#### 4.1 Availability

Source code (Linux installer), Microsoft Windows installer, manual, sample data, and sample output are available for non-commercial purposes and can be downloaded from <http://www.ageri.sci.eg/index.php/facilities-services/ageri-softwares/kaspspoon>.

#### Acknowledgements

The authors would like to dedicate this work to the soul of Prof. Dr Sami Adawy. We also thank Mr Morad Mokhtar (Molecular Genetics and Genome Mapping Lab., Agricultural Genetic Engineering Research Institute, ARC, Egypt) and Abdulqader Jighly (Department of Economic Development, Jobs, Transport and Resources, Australia) for their valuable support during this study.

*Conflict of Interest:* none declared.

#### Funding

This work was funded by the Grain Legume and Dryland Cereals (GLDC) and Grain Research and Development Cooperation (GRDC).

#### References

- Baeza-Yates,R.A. and Perleberg,C.H. (1992) Fast and practical approximate string matching. In: Apostolico,A. *et al.* (eds) *Combinatorial Pattern Matching*. CPM 1992. Lecture Notes in Computer Science. Vol. 644. Springer, Berlin, Heidelberg.
- Boutros,P.C. and Okey,A.B. (2004) PUNS: transcriptomic-and genomic-*in silico* PCR for enhanced primer design. *Bioinformatics*, 20, 2399–2400.
- Cingolani,P. *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6, 80–92.
- Doddamani,D. *et al.* (2015) CicArVarDB: SNP and InDel database for advancing genetics research and breeding applications in chickpea. *Database*, 2015, 1–7.
- He,C. *et al.* (2014) SNP genotyping: the KASP assay. In: Fleury,D. and Whitford,R. (eds) *Crop Breeding. Methods in Molecular Biology (Methods and Protocols)*. Vol. 1145. Humana Press, New York, NY.
- Horspool,R.N. (1980) Practical fast searching in strings. *Software Pract. Exper.*, 10, 501–506.

- Kalendar, R. et al. (2011) Java web tools for PCR, *in silico* PCR, and oligo-nucleotide assembly and analysis. *Genomics*, **98**, 137–144.
- Kolmer, J.A. et al. (2018) Mapping and characterization of the new adult plant leaf rust resistance gene Lr77 derived from Santa Fe winter wheat. *Theor. Appl. Genet.*, **131**, 1553–1560.
- Krzywinski, M. et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.
- Lexa, M. et al. (2001) Virtual PCR. *Bioinformatics*, **17**, 192–193.
- Mohajeri, K. et al. (2016) Interchromosomal core duplicons drive both evolutionary instability and disease susceptibility of the Chromosome 8p23.1 region. *Genome Res.*, **26**, 1453–1467.
- Nayak, S.N. et al. (2010) Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. *Theor. Appl. Genet.*, **120**, 1415–1441.
- Rice, P. et al. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
- Untergasser, A. et al. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res.*, **40**, e115.
- Ye, J. et al. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.*, **13**, 134.