Piperpep – a database of experimentally generated peptides from black pepper (*Piper nigrum* L.)

P. Umadevi, Vivek Srivastav, M. Anandaraj, Johnson K. George and Santhosh J. Eapen

The Piperpep database is a compilation of peptide data from mass spectrometry analysis of a medicinally important export oriented crop – black pepper, developed by the fact that any mass spectrometry derived data would be an excellent source for genome annotation in future. This database has 1001 peptides, including 36 antimicrobial peptides annotated from the leaf of black pepper. The user friendly browsing and search tools would help the researcher with crop improvement in black pepper and also in other crops via peptide or protein data. The information on antimicrobial peptides in this database is expected to cater towards development of new drugs for agricultural and medicinal uses.

Black pepper (*Piper nigrum* L.) belongs to the Piperaceae family. It is believed that Europeans ventured into the new world primarily because of this spice crop[1]. The genomic and proteomic information on this crop is scanty but is important, as it would aid in understanding the molecular basis of the crop response to various biotic and abiotic factors. We attempted label free proteomics to generate peptide data in this crop. Mass spectrometry of tryptic digests, LC-MS/MS analysis provides the high resolution mass spectral data with accurate mass values. This list of MS/MS spectra is used to search proteins in the databases to identify them. The tandem mass spectrometry (MS/MS) data are generally partial sequences of a peptide, referred to as peptide sequence tags. These sequence tags have enough information to identify the protein from databases as large as the *Arabidopsis* genome database[2]. MS derived data is also essential for annotation of nucleic acid and protein sequence databases since any protein based data will provide a better prediction accuracy rate. MS as the proteomic tool of choice, is poised to contribute greatly to information on *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster* and human genome draft databases[3]. The MS derived data was used to correct genome annotations and finding novel proteins in *Anopheles gambiae*[4]. Many such reports make MS-derived peptide sequence data more valuable for gene prediction. Hence peptide-based bioinformatics tools are of immense importance. Presently, the number of plant peptide databases is scanty. With the annotation of 1001 peptides, we generated the Piperpep database. Currently, the database has information on peptide sequences and its physicochemical properties from this medicinally important spice, but variation in expression of peptides during the biotic and abiotic stress will be included during upgradation of the database. Our database with annotated peptides is a ready resource for whole genome sequencing work by future researchers. This database will also be a source of reference for finding similar peptides from other plants. Apart from data on peptide sequences of proteins, the antimicrobial peptides (AMPs) from this crop were also included. AMPs are low molecular weight peptides synthesized using normal ribosomal protein synthesis machinery, and that show broad spectrum antimicrobial activity. Plant AMPs are classified as cyclotides, defensins, thionins, lipid transfer proteins, snakins and hevein-like proteins. Recent reports show that these antimicrobial peptides are multitasking, involved in regulating the plant development and acts as a defense against biotic/abiotic stress[5].

IISR-Shakthi, an improved variety derived from open pollinated progeny of Perambramundi, was used to produce the peptide data. This variety is resistant to *Phytophthora* foot rot, and is suitable for plains and a high range rain fed conditions. Protein was extracted from leaf[6] followed by label-free proteomics, using tryptic digested protein extract in LC-MS/MS (LTQ-Orbitrap) with the resolution 300–2000 mass range and mass accuracy <10 ppm.

The ion source was ESI; the charge state of the peptides used to generate MS/MS data was +2, +3. The raw data were analysed in Mascot search engine against Swissprot, TrEMBL, AMP database (APD) and the peptide sequences were annotated using specific parameters. With annotated sequences the database was developed. Piperpep database platform was designed using the open sources WAMP server (http://www.wampserver.com/en/) on Windows 8 with MySQL 5.6.17 (https://mysql.com) and Apache 2.4.9 (www.apache.org/) in back-end and front-end of web-interface implemented with PHP 5.5.12 (https://www.php.net/). The database is freely accessible at http://220.227.138.213/piperpep/

The database is in-built with browsing and searching features with various physicochemical properties and predicts similar peptide from the Piperpep database. The physicochemical properties include accession, protein name, origin, peptide, peptide ion, score, hits, mass and modification point in the amino acid. The database reports peptides from more than 1000 biologically important peptides along with the 36 antimicrobial peptides. Figure 1 *a* shows the percentage of individual amino acid modifications in the peptides. The antimicrobial peptides were grouped into different families and the statistics on the groups of AMPs are shown in Figure 1 *b*. The length distribution of peptides was analysed and the frequency of peptides with particular length is given in Figure 1 *c*. Piperpep database has three in-built tools for identifying input peptide queries (Piper BLAST, MAP Tool) and determining of physicochemical properties (Pepcal) of the query peptide. Piperpep database is in-built with different types of tabs such as browse and search that explore the data from the database. Browse-by option allows users to explore the tab against PiperID, accession, protein name,

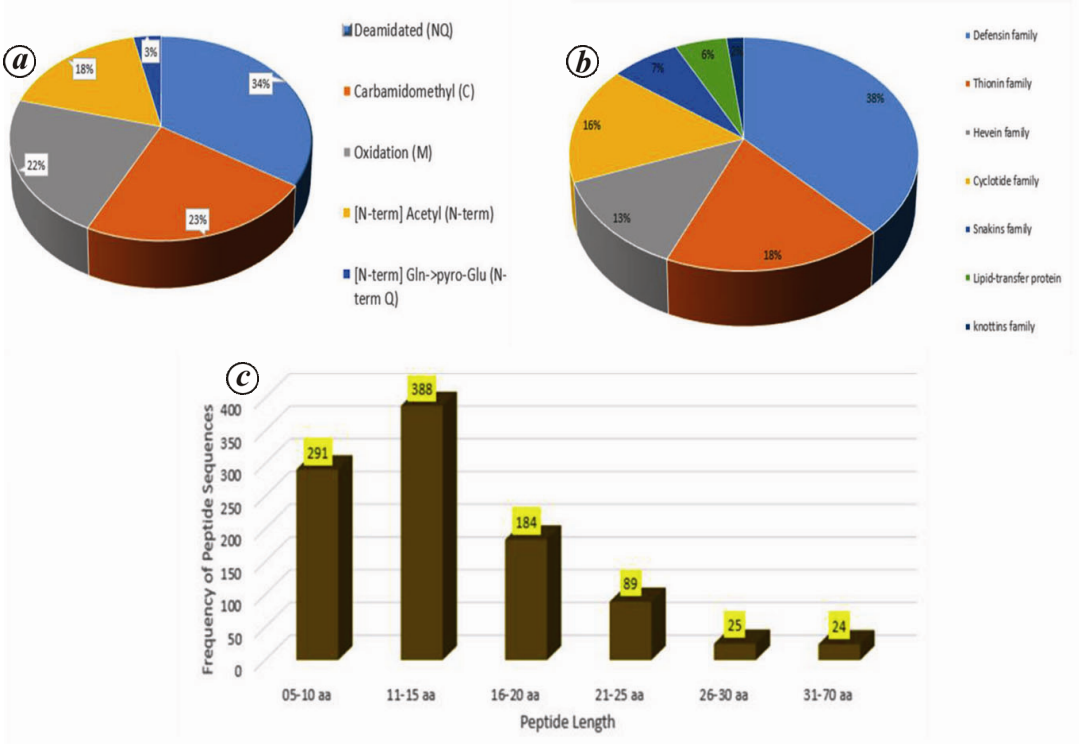

Figure 1. ***a***, Percentage of peptide with the modification in the amino acid. ***b***, Classification of anti microbial peptides in the Piper pep database. ***c***, Peptide length distribution pattern of the peptide entries in the database.

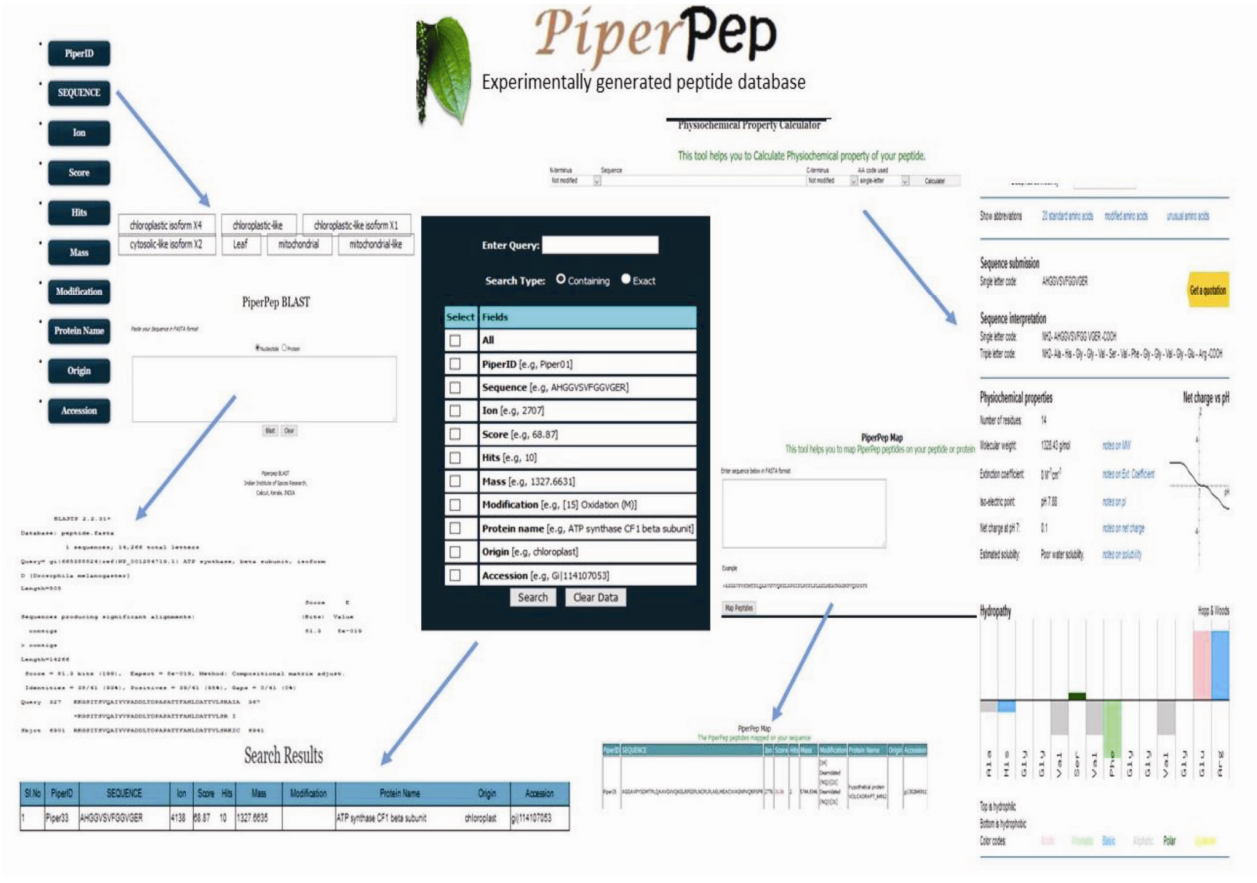

Figure 2. Piperpep user interface.

origin, peptide sequence, peptide ion, score, hits, mass and modification from the database. Search tab incorporates conditional search with user entry in the query form in the box and select any of the 12 fields against the field given in the database or keep the default '*all*' option to search the data from the database. The result displays the different categories and different physicochemical properties of the peptide from the database. The user interface on the available browse, search and tools of Piperpep is given in Figure 2.

Piperpep is an open source database of peptides. Our database with annotated peptides is a ready resource whole genome sequencing work by researchers. Piper BLAST: Blast tool in the Piperpep finds the peptide for the user provided sequences from the database[7]. The in-built link to NCBI helps identify the query peptide sequences with the proteins present in the NCBI database. The PiperPep MAP is a simple tool to align perfectly matching peptide available in Piperpep database. The output is displayed as matching Piper ID, origin, sequence and with amino acid modifications. Pepcal, an user friendly tool in-built in Piperpep database, assists the user to find the molecular weight, isoelectric point, pH, solubility and hydrophobicity. The results are provided with the link from http:www.innovagen.com. Apart from data on peptide sequences of proteins, this database is enriched with data on the antimicrobial peptides. The Piperpep database is expected to cater to the needs of researchers working in black pepper development programme in particular and provide an informative database on peptides for various plant sequence annotations.

1. Ravindran, P. N., In *Black Pepper* (*Piper nigrum* L.) (ed. Ravindran, K. C.), Harward Academic Publishers, 2000, pp. 1–22.
2. Kuster, R., *Proteomics*, 2001, **1**, 641–650.
3. Mann, M. and Pandey, A., *Trends Biochem. Sci*., 2001, **26**, 54–61.
4. Kalume, E. D., *BMC Genomics*, 2005, **6**, 128.
5. Goyal, R. K. and Matoo, A., *Plant Sci*., 2014, **228**, 135–149.
6. Umadevi, P. and Anandaraj, M., *Plant Omics*, 2015, **8**, 500–507.
7. Altschul, S. F., *J. Mol. Biol*., 1990, **215**, 403–410.

ACKNOWLEDGEMENTS. We acknowledge ICAR for funding through Network project on *Phytophthora*, *Fusarium* and *Ralstonia* diseases of horticultural crops.

P. Umadevi and Johnson K. George are in the Division of Crop Improvement, ICAR-Indian Institute of Spices Research, Kozhikode 673 012, India; Vivek Srivastav, M. Anandaraj and Santhosh J. Eapen are in the Division of Crop Protection, ICAR-Indian Institute of Spices Research, Kozhikode 673 012, India*
**e-mail: arajiisr@gmail.com*