Potato J (2017) 44 (1): 74-78

GENOME SEQUENCING OF POTATO LATE BLIGHT PATHOGEN, *PHYTOPHTHORA INFESTANS* A2 MATING TYPE

Youvika Singh[1], HC Rawal[2], TR Sharma[2], BP Singh[1], Pradeep K Shukla[3], Sanjeev Sharma[1], VU Patil[1], SK Chakrabarti[1], Shashi Rawat[1]

KEYWORDS: CAZy, genome sequencing, late blight, *Phytophthora infestans*, potato

Phytophthora infestans, is considered to be the most devastating pathogen for potato crop worldwide. The annual losses due to late blight are estimated to exceed €12 billion per annum worldwide (Haverkort, 2009). *P. infestans* has been replaced by new aggressive strains characterized as A2 mating type which were originally recognized in central Mexico (Dyer *et al.*, 1993; Gallegly *et al.*, 1958; Arora *et al.*, 2014). First report of A2 mating type outside Mexico was in Europe (Hohl and Iselin, 1984) and from there it spread to rest of the world (Dyer *et al.*, 1993; Goodwin, 1997). Occurrence of the new A2 mating type strain in India was first reported by Singh *et al.* (1994). A2 mating type enables sexual reproduction in *Phytophthora* that contributes to genetic diversity and increase metalaxyl resistance (Deahl *et al.*, 1995; Drenth *et al.*, 1995). To the best of our knowledge, only one genome of this species is available in databases. Therefore, acquiring the genome sequence of an Indian strain would further increase our understanding towards behavioral activity and designing management strategies for the catastrophic pathogen. This sequence information is the first study, which covers the preliminary information on Indian strain of A2 mating type of *P. infestans*. In this study, we report 152 Mb of sequenced and assembled genome (out of the estimated 240 Mb genome size) of *P. infestans* HP-10-31 strain of A2 mating type, isolated in growing season of the year 2010 from Shimla, Himachal Pradesh, India. The sequenced genome of *P. infestans* T30-4 (Haas *et al.*, 2009) was used as reference for assembly in this work.

Genomic DNA was extracted from freshly harvested mycelium of *P. infestans* grown in rye broth medium (Caten and Jinks, 1968) for 15 days at 18°C using GenElute Plant Genomic DNA isolation kit as per manufacturer's instructions. The quality of DNA was checked by running on 1% agarose gel as well as by NanoDrop 2000 to determine the A260/A280 ratio. Sequencing libraries were constructed and size selected using 500 ng of DNA as described by Roche 454 Rapid Library preparation method. The genome of *P. infestans* was sequenced using next generation sequencing platform, Genome sequencer FLX instrument following the manufacturer's protocol (Roche Applied Science). The reads were reference assembled using G S Reference Mapper v2.8 and *P.*

[1]ICAR-Central Potato Research Institute, Shimla- 171001, Himachal Pradesh, India
Email: uvika85@gmail.com
[2]ICAR-National Research Centre on Plant Biotechnology, New Delhi, India
[3]Department of Biological Science, School of Basic Sciences, SHIATS, Allahabad, India

infestans T30-4 as reference genome. Instances of repetitive regions were identified in the assembled genome using MapRep (MolQuest V2.2.0 www.molquest.com/molquest.phtml; MolQuest –Bioinformatics Toolbox for analysis of biomedical data) using default parameters with at least 80% homology against Fungi Repbase database (Jurka *et al.*, 2005). Total repeat content was quantified by summing all classes of repeats including transposons. SignalP4.0 (http://www.cbs.dtu.dk/services/signal) (Petersen *et al.*, 2011) was used to perform signal peptide cleavage site prediction in *P. infestans*. Transmembrane helices in proteins were predicted using TMHMM2.0 (http://www.cbs.dtu.dk/services-2.0/) (Krogh *et al.*, 2001). The proteins that contained signal peptide cleavage site and no transmembrane helices were selected as effectors. The proteins with lengths less than 200 amino acids were omitted. RxLR effectors were identified using the approach described by Win *et. al*. (2007). Candidate RxLR effectors were selected from the secreted translations using custom Perl scripts. Secreted translations with RxLR position between 30-60 residues from signal peptide, RxLR position downstream of signal peptide cleavage site of less than 30 amino acids were selected as candidate RxLR effectors. To predict the genes in assembled genome we employed FGENESH module of MolQuest V2.2.0 (www.molquest.com/molquest.phtml; MolQuest –Bioinformatics Toolbox for analysis of biomedical data) with default parameters. For functional annotation, predicted genes were compared against NR database of NCBI using standalone BLASTx algorithm to filter out the significant hits with identity % ≥60% and E-value ≤ e^{-20}. The associated hits were searched for their respective Gene Ontology (GO) terms manually. To identify proteins involved in carbohydrate metabolism we used CAZyme annotation program (http://www.cazy.org/) with default parameters. Search and functional annotations for carbohydrate-active modules and ligninolytic enzymes, including glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activities (AAs), were performed. To identify the potential pathogenesis related genes, we BLAST searched the predicted genes against PHI-database V3.6 (http://www.phi-base.org) (pathogen host interaction database) (Winnenburg *et al.*, 2006) and classified the genes with significant hits (E value≤ e^{-20}) into different categories like reduced virulence, lethal, unaffected pathogenicity, loss of pathogenicity etc.

A total of 157,910,676 reads (5.5 Gb) with average read length of 450 base pair were generated using Roche 454 sequencing technology. The final assembly consists 47,871 contigs with an N_{50} of 12,952 bp. Instances of repetitive regions were identified in the assembled genome that reveals that 42.8 Mb or one third of genome is repetitive.

Secreted proteins or secretome are important for interaction with host plant (Muller*et al.*, 2008; Doehlemann *et al.*, 2009). Some secreted proteins may incapacitate plant defense and disrupt cellular processes. Computational tools are available that predict whether a protein is likely to be secreted or not (Soanes *et al.*, 2008). Following such criteria 713 protein-coding genes were predicted to be secreted. RxLR effectors were identified using approach described by Win method (Win *et al.*, 2007). Candidate RxLR effectors were selected from these secreted proteins using custom perl scripts. Secreted proteins with RxLR position between 30-60 residues from signal peptide, RxLR position downstream of signal peptide cleavage site was selected as candidate RxLR effectors (Supp. S1).

A total of 21,118 protein-coding genes (>150 aa) were predicted using FGENESH module of MolQuest V2.2.0 (http://www.molquest.com) out of which 15,291 genes (72% of predicted genes) were annotated against the NR database using BLASTx hits with Identity % ≥60% and E-value ≤ e^{-20} (Supp. S2). Among these annotated genes, 8,162 protein coding genes were distributed into 22 different GO terms: 53.91% associated with binding, 12.63% with metabolic activities, 9.88% with catalytic activity, 6.4% with transporter activity and oxidoreductase activity. 6.4% genes were related to various functional categories like pathogenesis, structural molecule activity, isomerase and lyase activity, spore germination etc. (**Fig. 1**).

Identification of pathogenicity related gene is essential to understand pathogenicity mechanism. Pathogen host interaction database catalogues experimentally verified pathogenicity, virulence and effector genes from fungal and oomycete pathogens which infect animal, plant, fungus. The information could be helpful in future research targeted on development of disease controlling strategies. Pathogenesis related genes were predicted using PHI database v3.6 (Winnenburg *et al.*, 2006) with significant hits (E value ≤ e^{-20}). A total of 2,377 (11.2%) genes were involved in pathogen host interactions (Supp. S3). Genes in *P. infestans* HP-10-31 spanning across 58 different fungal and oomycete species. Highest number of homologs was found in *Magnaporthe oryzae* (1258 genes) followed by *Gibberella zeae* (250 genes), *Candida albicans* (106 genes), *Ustilago maydis* (97 genes), *P. infestans* (88 genes), *Aspergillus fumigatus* (78 genes) and other species (506 genes) (Supp. S3).

Ability to penetrate the physical barrier of plant cell wall is essential need for pathogenesis. Cell wall degrading enzymes are extracellular effectors secreted by pathogen to assist in degradation of wide variety of complex cross-linked polysaccharides and glycoprotein. Carbohydrate active enzymes (CAZyme) play a central role in synthesis and breakdown of plant cell wall. In *P. infestans* CAZyme can serve as pathogenicity factors by unambiguously targeting carbohydrates of plant cell wall. CAZy classes' glycoside hydrolases (GHs), glycosyl transferases, (GTs),

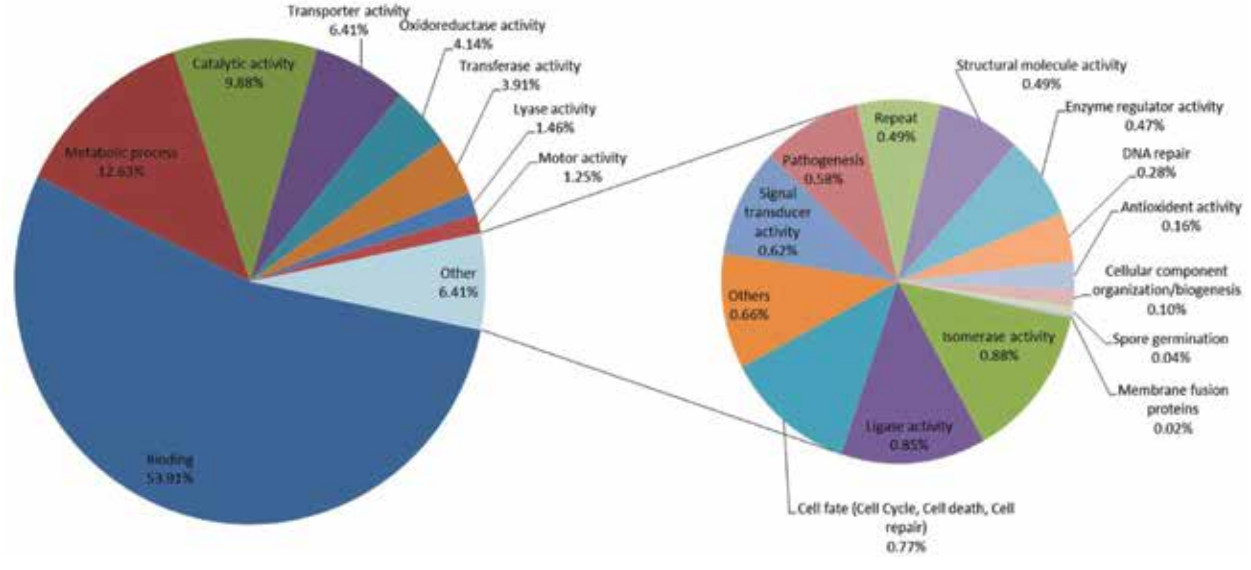

Fig. 1: General annotation of P. infestans HP-10-31 genome.

polysaccharide lyases (PLs), carbohydrate estearases (CEs), auxiliary activities (AAs), and carbohydrate binding modules (CBMs) were searched from CAZy database. Predicted proteins were subjected to dbCAN web server (Yin *et al.*, 2012) to mine CAZy families from the genome. We predict 406 CAZy coding genes in *P. infestans* HP-10-31.In this study, we predict few expanded CAZy genes that were not reported previously in any *Phytophthora* genomes (Supp. S4). The GHs are involved in hydrolysis of glycosidic bond between or within carbohydrate molecules. The number of GH12 has been increased in *P. infestans* HP-10-31 (16 homologues) however, in *P. infestans* T30-4 (10 homologues), *P. sojae* (14 homologues) and *P. ramorum* (7 homologues) (Ospina-Giraldo *et al.*, 2010) probably due to duplication events.

Further, genome sequencing of this strain to achieve opportunities for comparative analysis with earlier genome will clear our understanding of host-pathogen interaction. Genomic level information will provide ways to regulate this havoc pathogen. As a result of genome analysis, efforts are in progress to validate the function of all pathogenicity related genes.

Nucleotide sequence accession numbers: The genome sequence has been deposited at DDBJ/GenBank/EMBL under accession no. LYVM00000000. The version described in this paperisversion LYVM01000000.

LITERATURE CITED

Arora RK, Sharma S and Singh BP (2014) Late blight disease of potato and its management. *Potato J* **41**(1): 16-40

Caten CE and Jinks JL (1968) Spontaneous variability of single isolates of *Phytophthora infestans*. I. Cultural variation. *Can J Botany* **46**(4): 329-348

Deahl KL, DeMuth SP, Sinden SL and Rivera-Peña A (1995) Identification of mating types and metalaxyl resistance in North American populations of *Phytophthora infestans*. *Am J Potato Res* **72**(1): 35-49

Doehlemann G, van der Linde K, Assmann D, Schwammbach D, Hof A, Mohanty A, Jackson D and Kahmann R (2009) Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. *PLoS Pathog* **5**(2): e1000290

Drenth A, Janssen EM and Govers F (1995) Formation and survival of oospores of *Phytophthora infestans* under natural conditions. *Plant Pathol* **44**(1): 86-94

Dyer AT, Matuzak JM, Drenth A, Tooley PW, Sujkowski LS, Koh YJ, Cohen BA, Spielman LJ, Deahl KL and Inglis DA (1993) Historical and recent migrations of *Phytophthora infestans*: chronology, pathways, and implications. *Plant Dis* **77**: 653-661

Gallegly ME and Galindo J (1958) Mating types and oospores of *Phytophthora infestans* in nature in Mexico. *Phytopathology* **48**: 274-277

Goodwin SB (1997) The population genetics of *Phytophthora*. *Phytopathology* **87**: 462-473

Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AM, Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JI, Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, Fischbach MA, Fugelstad J, Gilroy EM, Gnerre S, Green PJ, Grenville-Briggs LJ, Griffith J, Grünwald NJ, Horn K, Horner NR, Hu CH, Huitema E, Jeong DH, Jones AM, Jones JD, Jones RW, Karlsson EK, Kunjeti SG, Lamour K, Liu Z, Ma L, Maclean D, Chibucos MC, McDonald H, McWalters J, Meijer HJ, Morgan W, Morris PF, Munro CA, O'Neill K, Ospina-Giraldo M, Pinzón A, Pritchard L, Ramsahoye B, Ren Q, Restrepo S, Roy S, Sadanandom A, Savidor A, Schornack S, Schwartz DC, Schumann UD, Schwessinger B, Seyer L, Sharpe T, Silvar C, Song J, Studholme DJ, Sykes S, Thines M, van de Vondervoort PJ, Phuntumart V, Wawra S, Weide R, Win J, Young C, Zhou S, Fry W, Meyers BC, van West P, Ristaino J, Govers F, Birch PR, Whisson SC, Judelson HS and Nusbaum C (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**(7262): 393-398

Haverkort AJ, Struik PC, Visser RGF and Jacobsen E (2009) Applied Biotechnology to Combat Late Blight in Potato Caused by *Phytophthora Infestans*. *Potato Res* **52**: 249–264

Hohl HR and Iselin K (1984) Strains of *Phytophthora infestans* with A2 mating type behaviour. *T Brit Mycol Soc* **83**: 529-530

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O and Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**(1-4): 462-467

Krogh A, Larsson B, von Heijne G and Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**(3): 567-580

Mueller O, Kahmann R, Aguilar G, Trejo-Aguilar B, Wu A and de Vries RP(2008) The secretome of the maize pathogen *Ustilago maydis*. *Fungal Genet Biol* **45**(1): S63-70

Ospina-Giraldo MD, Griffith JG, Laird EW and Mingora C (2010) The CAZyome of *Phytophthora* spp.: a comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*. *BMC Genomics* **11**: 525

Petersen TN, Brunak S, von Heijne G and Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**(10): 785-786

Singh BP, Roy S, Bhattacharyya SK (1994) Occurrence of the A2 mating type of *Phytophthora infestans* in India. *Potato Res* **37**: 227–231

Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, Rattray M, Hubbard SJ, Oliver SG and Talbot NJ (2008) Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. *PLoS One* **3**(6): e2300

Win J, Morgan W, Bos J, Krasileva KV, Cano LM, Chaparro-Garcia A, Ammar R, Staskawicz BJ and Kamoun S (2007) Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* **19**(8): 2349-2369

Winnenburg R, Baldwin TK, Urban M, Rawlings C, Köhler J and Hammond-Kosack KE (2006) PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res* **34** (Database issue): D459-464

Yin Y, Mao X, Yang J, Chen X, Mao F and Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**(W1): W445-W451

MS received: 22 February 2017; Accepted: 10 March 2017