

# SCIENTIFIC REPORTS



OPEN

## Draft whole genome sequence of groundnut stem rot fungus *Athelia rolfsii* revealing genetic architect of its pathogenicity and virulence

M. A. Iquebal<sup>1</sup>, Rukam S. Tomar<sup>2</sup>, M. V. Parakhia<sup>2</sup>, Deepak Singla<sup>1</sup>, Sarika Jaiswal<sup>1</sup>, V. M. Rathod<sup>2</sup>, S. M. Padhiyar<sup>2</sup>, Neeraj Kumar<sup>1</sup>, Anil Rai<sup>1</sup> & Dinesh Kumar<sup>1</sup>

Groundnut (*Arachis hypogaea* L.) is an important oil seed crop having major biotic constraint in production due to stem rot disease caused by fungus, *Athelia rolfsii* causing 25–80% loss in productivity. As chemical and biological combating strategies of this fungus are not very effective, thus genome sequencing can reveal virulence and pathogenicity related genes for better understanding of the host-parasite interaction. We report draft assembly of *Athelia rolfsii* genome of ~73 Mb having 8919 contigs. Annotation analysis revealed 16830 genes which are involved in fungicide resistance, virulence and pathogenicity along with putative effector and lethal genes. Secretome analysis revealed CAZY genes representing 1085 enzymatic genes, glycoside hydrolases, carbohydrate esterases, carbohydrate-binding modules, auxiliary activities, glycosyl transferases and polysaccharide lyases. Repeat analysis revealed 11171 SSRs, LTR, Gypsy and COPIA elements. Comparative analysis with other existing ascomycotina genome predicted conserved domain family of WD40, CYP450, Pkinase and ABC transporter revealing insight of evolution of pathogenicity and virulence. This study would help in understanding pathogenicity and virulence at molecular level and development of new combating strategies. Such approach is imperative in endeavour of genome based solution in stem rot disease management leading to better productivity of groundnut crop in tropical region of world.

Groundnut (*Arachis hypogaea* L.) is an important oil seed crop, which like other legumes, harbor symbiotic nitrogen-fixing bacteria in root nodules<sup>1,2</sup> improving soil fertility. It has global annual production of 43.92 million tonnes in 2014. India is the second largest producer of groundnut with 6.56 million tonnes (FAO, 2014) after China (16.55 million tonnes). One of the major biotic constraints in production is stem rot/ *Sclerotium* wilt disease caused by *Sclerotium* (*Athelia*) *rolfsii*. *Athelia rolfsii* is the sexual state of the fungus, but it is rarely seen, in contrast to the asexual state, known as *Sclerotium rolfsii*. It causes average loss of 25%<sup>3</sup> which goes upto 80% in severe cases<sup>4</sup>. This soil borne pathogen has vast range of host (>500 plant species) due to its ability of high growth rate along with production of cell degrading enzymes<sup>5</sup>.

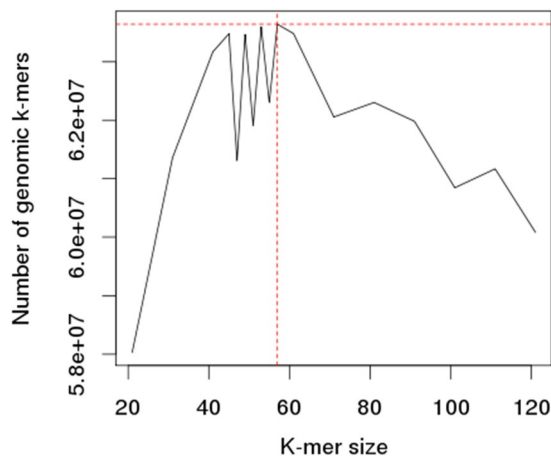
The fungal mycelium covering plant stem near soil produces organic acid which is toxic to plants<sup>6</sup>. With the help of fungal secretory system, mycelia invade in stem causing rotting of the tissue due to toxicity that further leads to necrosis<sup>7</sup>. The characteristic symptom of the disease is due to abundance of white mycelium along with small brown spherical sclerotia<sup>8</sup>. Biocontrol of *A. rolfsii* has been reported by *Trichoderma* fungi, and chemical control by fungicides such as tebuconazole, pentachloronitrobenzene (PCNB) and flutolanil<sup>9</sup>. Biocontrol has compromised efficacy and chemical control by fungicide has issues of bio-magnification in food chain and associated environmental problems<sup>10</sup>.

Genome sequencing can reveal role of each individual gene and their networks responsible for plant pathogen interaction, growth, evolutionary relationship and genes for pathogenicity. Whole genome sequencing of *Athelia rolfsii* is imperative not only to study the host-pathogen (HP) interaction but such knowledge discovery may lead

<sup>1</sup>Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, Pusa, New Delhi, 110012, India. <sup>2</sup>Junagadh Agricultural University, Junagadh, 362 001, Gujarat, India. M. A. Iquebal and Rukam S. Tomar contributed equally to this work. Correspondence and requests for materials should be addressed to D.K. (email: [Dinesh.Kumar@icar.gov.in](mailto:Dinesh.Kumar@icar.gov.in))

	Sequences > 1 bp	Sequences > 500 bp	Sequences > 1 kb	Sequences > 5 kb	Sequences > 10 kb
Total contigs	8,919	7,733	5,222	2,692	1,833
Average contigs length	8,204	9,402	13,586	24,108	32,064
N50 (bp)	32,103	32,380	33,175	36,179	39,338
L50	670	663	636	549	468
Assembly length(bp)	73,180,279	72,709,682	70,949,542	64,900,867	58,773,395

**Table 1.** Results obtained after CAP3 assembly.



**Figure 1.** Estimation of genome size and *k*-mer value using *kmergenie* software.

to more effective disease combating strategy. Annotated genes/ predicted proteins can be directly used as new targets in fungicides development using computational approach<sup>10</sup>.

Here, we report the first draft assembly of *A. rolfsii* genome and its analysis revealing genes involved in fungicide resistance, virulence and pathogenicity along with putative effector and lethal genes. Our analysis of fungal proteome resulted in the prediction of various secretory proteins involved in carbohydrate metabolism mediating the HP interaction. Evolutionary insight has been elucidated by comparative study of this genome with other ascomycotina fungi. This work would be useful for understanding HP interactions and designing of the strategies for controlling the *A. rolfsii* infection.

## Results and Discussion

**Genome sequencing and assembly.** Genome sequencing was performed using Ion torrent PGM and 3.5 Gb of sequence data was generated with average read length 303 bp. High quality reads obtained after trimming of raw reads were subjected for *de novo* genome assembly using MIRA v4.0.2 followed by CAP3 (Table 1). The estimated genome size of ~63.64 Mb was obtained by *k*-mergenie software with optimal *k*-mer size of 57 (Figure 1). An assembly size of 75.37 Mb was obtained using MIRA assembler having 15281 contigs with N50 value of 26586 bp and largest contig length being 249590 bp. Further this assembly was improved using CAP3 software. Finally, we obtained an assembly of size 73.18 MB having 8919 contigs with N50 value of 32103 bp. This improvement of assembly was in magnitude of 41% (reduction in contigs number from 15281 to 8919) and increase in average contig length by ~66% (Table 1). Fungal genome size observed in the study was found to be 73.18 Mb. The potential reasons for difference in estimated and assembled genome size could be: amount of repetitiveness, heterozygosity, read error rate, type of chemistry used, coverage, compromise between contig length and number of errors in assembly<sup>11</sup>. In ascomycotina fungi genome size range varies from 15 (*Babjeviella inositovora*) to 152 (*Zopfia rhizophila*) Mb<sup>12</sup>. Our findings are similar to other reported ascomycotina genomes like *Ascobolus immersus* (59 Mb) and *Terfezia boudieri* (63 Mb)<sup>12</sup>.

Genemark-ES suite used for gene prediction resulted in a total of 16830 genes. We used 248 core eukaryotic genes (CEGs) using CEGMA pipeline to access the completeness of the genome assembly which revealed hits for 198 CEGs (~80%) with complete match and 229 (~92%) with partial match. Blast2GO Pro version 4.0.7 was used for functional characterization of genes using blastx tool resulted 15292 genes (90.8%) with Blast hits of which 10712 genes (63%) and 5623 genes (33%) were mapped and annotated. Average number of genes in fungal genome are in range of 11129 which varies from 4657 to 27529<sup>12</sup>. The predicted number of genes in our study (16830) is very similar to other ascomycotina like *Anthostoma avocetta* genome (56.23 Mb) having 15755 genes<sup>12</sup>. We observed that ~99% blast hits were obtained from fungal species among which agaricomycetes was in top species and rest (0.96%) were from non-fungal. The maximum number of hits were from *Plicaturopsis crispa* (4884 hits) followed by *Piloderma croceum*, *Serpula lacrymans* and *Jaapia argillacea* (Figure 2).

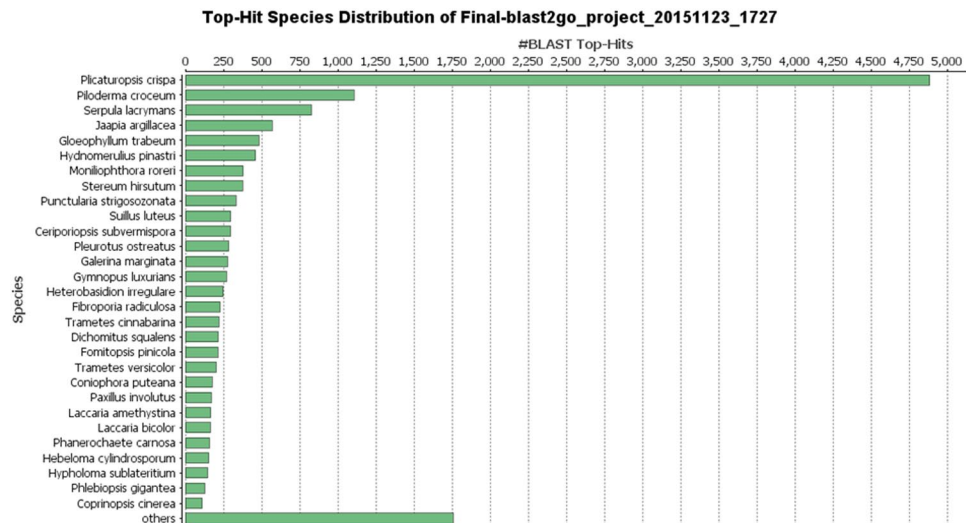


Figure 2. Distribution of blast hits among the different fungal species.

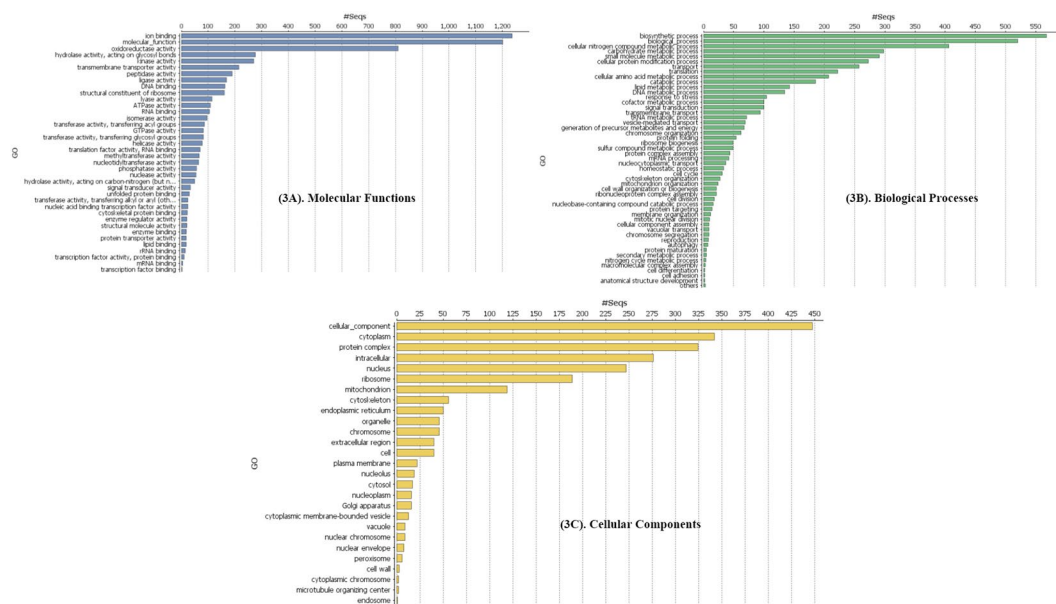
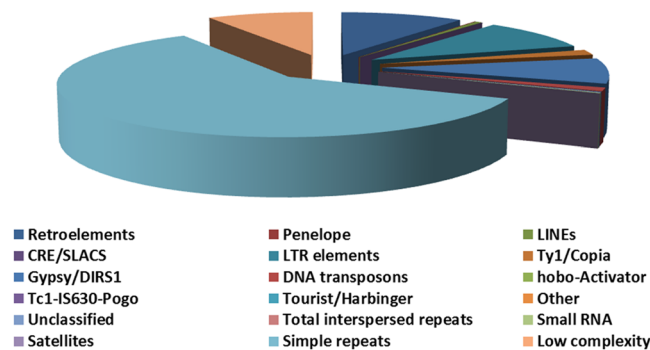


Figure 3. (A) GO-term for Molecular function. (B) GO-term for Cellular component. (C) GO-term for Biological function.

GO level annotation assigned the genes into three groups, viz., molecular function, cellular component and biological process. Based on molecular function, 3507 and 261 genes were annotated for catalytic and transporter activities, respectively (Figure 3A). Similarly, biological process classified 343 genes involved in regulation and 237 and 129 genes in stimulus and signaling, respectively (Figure 3B). However, cellular location assigned 795 genes that belonged to organelle and 736 genes were from the membrane (Figure 3C). Fungi are the rich source of secondary metabolites performing many biological functions. These metabolites are directly or indirectly affecting the growth of fungus species and mediation of pathogenicity. Our search for the gene involved in secondary metabolite biosynthesis resulted in the identification of two gene clusters, viz., terpene cyclases (TCs) and non-ribosomal peptides synthases (NRPSs) (Supplementary Figure S1)<sup>13</sup>. Both ascomycota and basidiomycota are known to produce various terpenoid compounds as secondary metabolites<sup>14</sup>. Fungal terpenoids play role as toxicant<sup>15</sup> and also acts for defensive purpose<sup>16</sup> in pathogenesis. NRPS genes play an important role in positive and negative control of transcription and post-translational events. NRPS gene has diverse gene family coding different peptides as secondary metabolite, such peptides has wide industrial applications like pigments, antibiotics, cytostatics, immunosuppressants, anticancer agents<sup>17,18</sup> and in crop protection<sup>19</sup>.

### Distribution of Repeat elements in *A. rolf sii* genome



**Figure 4.** Pie chart of repeats in *A. rolf sii* genome.

**Protein domain analysis.** Pfam domain search based on Pfam database resulted in the identification of 3308 domains distributed in 11624 (69.06%) genes. Maximum number of domains were from WD40 family (1250), followed by CYP450 (283), Pkinase (255), MFS\_1 (222), and ABC transporter (106). All of these were known to be involved in HP interaction. The role of these conserved gene families are well reported for example, WD40 in biotic stress tolerance<sup>20</sup>, fungal ABC transporters in virulence<sup>21</sup>, MAPK in fungal growth<sup>22</sup>, pathogenicity and virulence<sup>23</sup>, mechanical/enzymatic penetration in host tissues<sup>24</sup>.

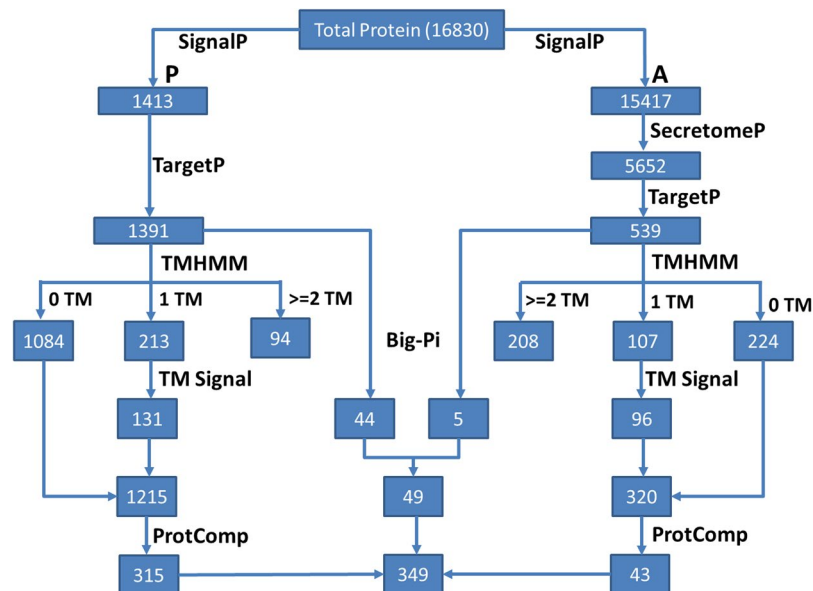
**Repetitive elements and SSR Marker analysis.** Repeats are the most frequently occurring region in case of eukaryotic genome. We used Repeat-masker for identification of repeats in *A. rolf sii* genome that accounts for only 3.73% of the total genome. Fungal genomes contain very less repetitive sequence as compared to other eukaryotes which very rarely exceeds 5% of the genome. The percentage observed in our study is quite consistent of repeat sequences reported in other ascomycete fungi<sup>25</sup>. Highest percentage of repeats were interspersed type (2.69%), small RNA (0.02%), simple repeats (0.85%) and low complexity regions accounting for 0.16% (Figure 4, Supplementary Table S1). LTR, Gypsy, and copia elements were present in higher amount (in 2.51%, 2.18% and 0.33%, respectively) in *A. rolf sii* genome.

Identification of SSRs in *A. rolf sii* genome by whole genome scan revealed a total of 11171 SSRs, including 470 compound SSRs. Mononucleotide SSRs, which were most abundant (9197) represented approximately 82.32% of all SSRs. Among all compound SSRs, 439 were interrupted SSRs (93.4%) and 7 were uninterrupted SSRs. Mono repeats are most abundant in fungal genome<sup>20</sup>. Number of mono repeats varies with fungal species in ascomycotina for example, in *Aspergillus nidulans* (565), *Fusarium graminearum* (371), *Magnaporthe grisea* (3,063), *Neurospora crassa* (2,505), *Saccharomyces cerevisiae* (2,071) and *Schizosaccharomyces pombe* (2,069). In our study, SSR density was found to be 152 repeat length per Mb which is in the reported range of other ascomycotina like *Aspergillus nidulans* (80), *Fusarium graminearum* (80), *Magnaporthe grisea* (307), *Neurospora crassa* (377), *Saccharomyces cerevisiae* (255) and *Schizosaccharomyces pombe* (247)<sup>26</sup>.

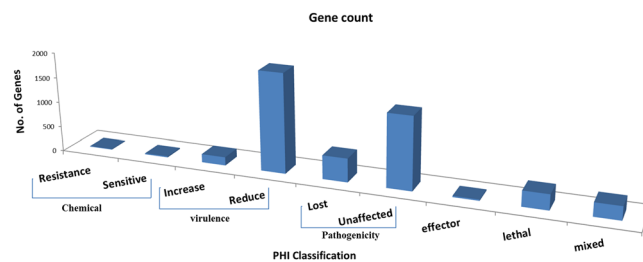
GC content of *A. rolf sii* genome was found to be 46.16% which is similar to wood degrading fungus *Phanerochaete chrysosporium* (57%)<sup>26,27</sup>. The GC and AT regions data can be used in studies related to fungal genome evolution and defense against transposon propagation which operated in AT rich region by repeat-induced point mutation (RIP). Such mechanism has also role in diversifying selection and mediation of host microbe disease interaction<sup>28</sup>. Our reported >11000 SSRs can be used in population structure and diversity analysis which is pivotal in designing strategies for better management of fungal disease. Such approach has been reported in management of leaf and fruit spot fungal disease of citrus<sup>29</sup>. These SSRs can also be used in fungal species/strain differentiation, rapid diagnostics with better sensitivity and specificity<sup>30</sup>.

**Secretome Analysis.** Fungal genome represents the rich source of secretory protein involved in various activities such as signalling, host-pathogen interactions etc. Secretome analysis of *A. rolf sii* genome resulted in the identification of 349 secretory proteins of which 306 were classical and 43 were non-classical secretory proteins (Figure 5). Further analysis of predicted secretome, we observed that 103 genes were assigned for GO-term for cellular location and 197 genes for biological process. Among the top GO-term biological process, cellular process and single cell organism process were highly abundant. Similarly, metabolic process is represented by catalytic and binding activities (Supplementary Figure S2). Under the cellular component, gene for membrane, cell, and cell part were found to be highly abundant. The enzyme annotation revealed four classes of enzymes, viz., oxidoreductases, transferases, hydrolases and isomerases. These enzymatic activities represent the high metabolic rate particularly involved in metabolism of host cell wall, an important phenomenon required for fungal invasion into host tissues (Supplementary Figure S3). All such enzymes along with virulence factors or toxins have been reported to degrade the plant cell wall in necrotrophic phase of fungal attack<sup>31</sup>. Pathogen hijacks this process compelling plant to alter cell wall making it more digestible<sup>32</sup>.

**Pathogenicity related genes.** Complete proteome of *A. rolf sii* was aligned to PHI database to reveal the pathogenicity related proteins. We observed a total of 4699 (27.99%) PHI genes were classified into different



**Figure 5.** Schematic representation of Secretome Analysis in *A. rolf sii*.



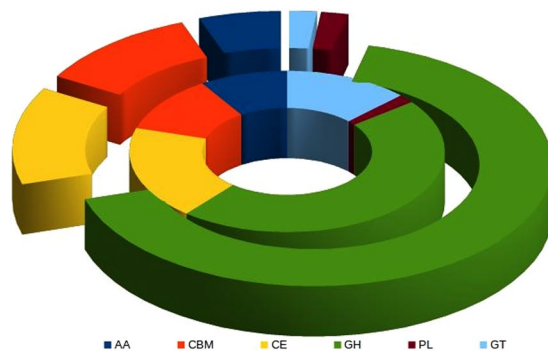
**Figure 6.** Distribution of *A. rolf sii* pathogenicity related genes in different classes.

classes such as chemical susceptibility, virulence, pathogenicity, effector, lethal and mixed. As shown in Figure 6, we observed 32 genes associated with chemical susceptibility having 11 resistant and 21 sensitive genes. Furthermore, 309 genes belonged to lethal and 167 to increased virulence class. Remaining three classes *viz.*, pathogenicity, effector and mixed were having 1919, 25 and 271 genes, respectively.

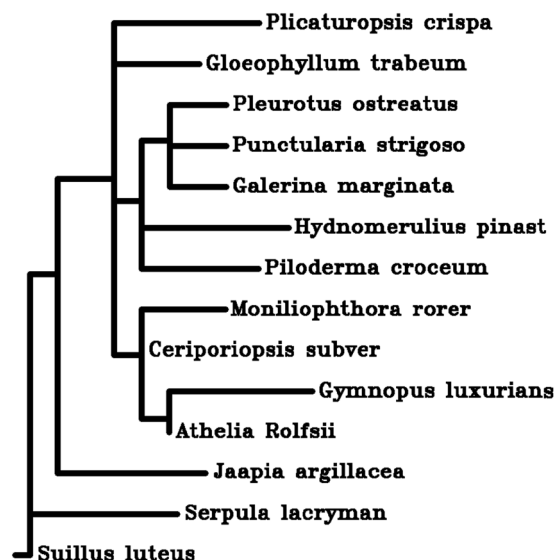
Pathogenicity related genes identified in this study have high relevance in future fungicide designing<sup>33</sup>. Mutation in these PR gene actually medicates the sensitivity and resistance of the fungal strain against fungicide<sup>34</sup>. These genes can be targeted for rapid and precise monitoring of fungicide resistance in the field<sup>35</sup>. Our enlisted lethal and virulence genes can be used in future research of fungal disease management, especially by designing of new generation genomic based fungicide<sup>36</sup>. Some of the virulence genes coding for extracellular proteins associated with carbohydrate metabolism have been reported to be involved in horizontal gene transfer (HGT). Such HGT leads to sudden outbreak with very high virulence<sup>37</sup>. The number of lethal genes in our study was found to be 309. In other ascomycotina, for example in *S. cerevisiae*, it was found to have 900 lethal genes<sup>38</sup>.

CAZymes represents the important component of fungal genome involved in HP interactions. Pathogenic fungus mediates the infection by its action on plant cell wall using the cell wall degrading enzymes. In this work, we identified a sum of 1085 CAZymes including 54 from the secretory proteins (Supplementary Table S2). As shown Figure 7, maximum number of enzymes were from glycoside hydrolases (GHs), followed by carbohydrate esterases (CEs), carbohydrate-binding modules (CBMs), auxillary activities (AAs), glycosyl transferases (GTs), and polysaccharide lyases (PLs).

Fungal genome often shows very high diversity in CAZymes profile due to their vast diversity in nutritional strategy and host specificity<sup>33</sup>. Large number of CAZymes are produced by fungi for degradation of plant polysaccharide in order to facilitate infection along with nutritional gain<sup>33</sup>. Our study revealed very high number (1085) of CAZymes which was obviously expected as *A. rolf sii* being a necrotrophic fungus. Necrotrophic fungi are reported to have more CAZymes than biotrophic and saprophytic fungi<sup>33</sup>. High abundance of pectinases are also reported in fungal pathogen of dicot like groundnut, with respect to monocot plant. This is due to high abundance of pectin in dicot. Previous phylogenetic studies based on more than 100 fungal species covering similar Basidiomycota along with Ascomycota, Chytridiomycota, and Zygomycota divisions has revealed that these CAZymes have complex history of lineage-specific expansions and attritions<sup>39</sup>.



**Figure 7.** Depicted the distribution of Caazy families into *A.rolfsii* secretome (inner circle) and whole proteome (outer circle). GH: Glycoside Hydrolases, GT: Glycosyl Transferases, PL: Polysaccharide Lyases, CE: Carbohydrate Esterases, CBM: Carbohydrate-Binding Modules, and AA: Auxillary Activities.



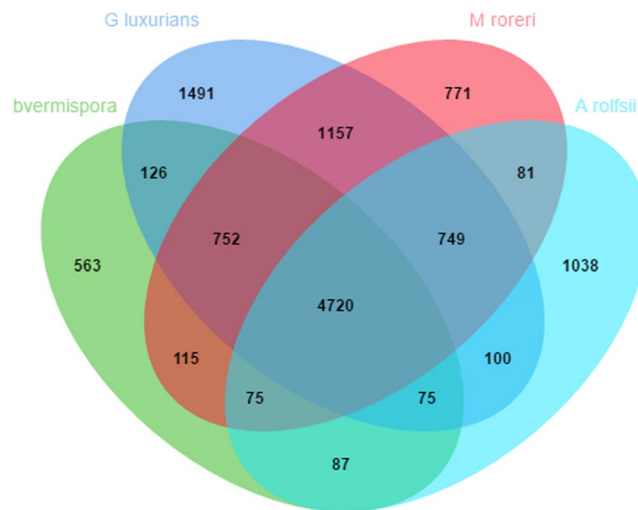
**Figure 8.** Phylogenetic tree showing the phylogenetic relationship amongst different fungal genomes belonging to class Agaricomycetes.

**Phylogenetic and Orthologous gene family Analysis.** Whole genome sequences of *A.rolfsii* along with 13 other fungal genomes from agaricomycetes class were used for phylogenetic analysis. As shown in Figure 8, *A.rolfsii* genome appeared to be closely related to *Gymnopus luxuricans* genome. Earlier, ascomycotina fungal species phylogenetic tree has been reported to cover evolutionary history of nutrition<sup>40</sup>. It is observed that in our constructed phylogenetic tree (Figure 8), only one species, *Suillus luteus* is non-saprophytic as it is ectomycorrhiza which is symbiotic<sup>41</sup>.

Orthologous gene families constitute a clade having *Athelia rolfsii*, *Gymnopus luxuricans*, *Ceriporiopsis subvermisporea* and *Moniliophthora roreri*. This comparative analysis predicted 4720 core orthologous groups (COGs) (Figure 9). Go enrichment analysis of the COGs identified tri-carboxylic acid (TCA) cycle genes significant at < 0.05 P-value. Among the COGs with significant hits, CYP-452 89A2 is the largest cluster followed by the Zinc finger CCCH domain-containing protein 62, UDP-glycosyltransferase 76E11, and probable voltage-gated potassium channel subunit beta.

The complete proteome of *A.rolfsii* comprises of 16830 genes were clustered into 6925 clustered of which 5887 were shared with other fungal genomes (Figure 9). Furthermore, Go-term enrichment analysis on 1038 unique clusters of *A.rolfsii* resulted in the cell surface (cellular component) and phosphatidylcholine 1-acylhydrolase activity (molecular function) as significant at P-value < 0.05 which plays role in fungal pathogenesis and virulence<sup>42</sup>.

Fungal genome is reported to have extremely variable (>300) Cytochrome P450 proteins (CYPs) genes due to diverse metabolism in their respective ecological niches as an adaptive response. Orthologous studies of CYP genes has high relevance in fungal systematics and classification<sup>35</sup>. Our analysis revealed largest hit with CYP-452 89A2. Agaricomycotina subdivision has been reported to have most dramatic expansion which might be due to gene duplication, adaptive divergence and horizontal gene transfer events<sup>43</sup>. CYP genes has potential to be evaluated for fungal diagnostics as reported in other fungi like *Candida spp*<sup>44</sup>. Conserved orthologous gene family



**Figure 9.** Venn diagram of orthologous gene family among the four fungal species.

UDP-glycosyltransferase 76E11 are reported to be associated with glyco-diversification of bioactive natural products having relevance in drug discovery. Similarly, the conserved gene family, probable voltage-gated potassium channel subunit beta is relevant in future studies on fungicide sensitivity<sup>45</sup>.

## Conclusion

This is the first report of whole genome sequence based draft assembly of *Athelia rolfsii*, causing important fungal disease, stem rot which affects groundnut productivity in many countries. Knowledge discovery from this studies will not only provide new insight in understanding the pathogenicity and virulence but will also lead to new dimension in development of disease combating strategies. Such more studies are required having transcriptomic and proteomic approach for genomic solution in the productivity management endeavour of its oil producing host crop.

## Methodology

**DNA Isolation.** *Athelia rolfsii* was isolated from an infected groundnut plant variety (GG-20) at Junagadh Agricultural University, Gujarat, India (21.51° N, 70.45° E). Potato Dextrose Agar (PDA) medium was used for growth and maintenance of isolates. These cultures were stored at 4 °C for long-term storage. *Athelia rolfsii* strains were grown in PDA broth medium for 6-days at 200 rpm at 28 °C. Mycelia were filtered and genomic DNA was extracted using Hipur ATM Fungal DNA purification kit (Himedia Cat. No. MB543) as per manufacturer's protocol.

**DNA library construction and Ion torrent PGM sequencing.** Library construction was conducted as per the Ion plus fragment library kit (Invitrogen Cat. No 4471252) for whole genome libraries. Total genomic DNA input was 100 ng which was fragmented using Ion shear<sup>TM</sup> enzyme mix II enzyme with an average of 400 bp DNA fragment sizes. The fragmented genomic library was cleaned using Agencourt Ampure XP Reagent (Beckman Coulter). The fragmented DNA was quantified using the Qubit DSDNA HsAssay Kit with the Qubit Fluorometer (Invitrogen). Purified DNA fragments were ligated with adapters specific to cleavage site of endonuclease enzyme, followed by the size selection of genomic library using E-gel 2% in order to get fragment size of 350–400 bp. The library was amplified using 10 cycles of PCR for enrichment of adapter ligated fragments and purified with Agencourt Ampure XP Reagent (Beckman Coulter). Template preparation for sequencing was conducted according to the OneTouch Ion<sup>TM</sup> Template Kit (Life Technologies). The selected PCR products were again used for emulsion PCR, followed by positive bead recovery. Ion Torrent sequencing was conducted using the Ion PGM<sup>TM</sup> 400 Sequencing Kit (Life Technologies) on an Ion Torrent Personal Genome Machine (PGM<sup>TM</sup>, Life Technologies) using a 318V<sup>2</sup>-chip (Ion 318<sup>TM</sup> chip, Life Technologies).

**Data pre-processing and genome assembly.** Reads were processed using Fastx toolkit v0.013 ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) and bases having phred score (Q) less than 20 were trimmed. High quality processed reads were used for estimation of genome size using kmergenie v1.69972 tool<sup>46</sup>. *De novo* assembly were performed using MIRA v4.0.2<sup>47</sup> assembler followed by CAP3 software<sup>48</sup> to obtain final draft genome assembly of *A. rolfsii*.

**Gene prediction and annotation.** RepeatMasker (<http://www.repeatmasker.org/>) was used to mask the repeats in contigs by selecting fungi as species and rmbblastn as the search engine with slow search option. To search the genes in draft genome, Genemark-ES suite v4.21<sup>49</sup> was used with following parameters –ES (self-training to predict genes); –fungus; –min-contig size: 500; –max-intron size: 3000; –min\_gene\_prediction: 120. Cufflinks was used for extracting coding and exon sequence in fasta format using –J option<sup>50</sup>. Presence of full length genes from the assembly was confirmed by using ORFfinder tool<sup>51</sup>. CEGMA v2.5<sup>52</sup> was used to assess the

completeness of the genome assembly and functional annotation of identified genes was done using Blast2GO Pro Ver. 4.0.7 software<sup>53</sup>. To access the domains in the genome of *Athelia rolfsii*, the assembly was analysed against Pfam<sup>54</sup> database at e-value cutoff 1e-05. The assembled genome was further used for searching genes involved in secondary metabolite biosynthesis using antiSMASH webserver<sup>55</sup>.

**Secretory protein prediction and its analysis.** All predicted proteins of *Athelia rolfsii* was used as input for identification and analysis of secretory proteins. SignalP v4.1<sup>56</sup> was used for the prediction of the signal peptide and those lacking the signal peptide were analyzed by SecretomeP v1.027<sup>57</sup>. Further, the output obtained from both tools were analysed using TargetP v1.1<sup>58</sup>. Subsequently, transmembrane domains and GPI (glycosylphosphatidyl inositol)-anchor were predicted using TMHMM v2.0<sup>59</sup> and big-PI FungalPredictor tool<sup>60</sup>, respectively.

Proteins having no transmembrane and one transmembrane domain within N-terminal Signal peptides were enlisted. Further, the proteins obtained from both classical and non-classical pipeline were subjected to sub-cellular localization prediction using ProtComp v9.0 based on LocDB and PotLocDB databases (<http://www.softberry.com>). Finally, GPI-anchor proteins were filtered out from the final set and rest of the sequences were the designated as secretome.

**Identification of the pathogenicity related genes.** PHI-base (a database of Host-Pathogen gene interactions)<sup>61</sup> was used to search the pathogenicity related gene in query sequence using Blastp. We also used dbCAN webserver<sup>62</sup> with default parameter to search the genes for their function as carbohydrate active enzyme (CAZY). This search was based on CAZY database which classified the hit into different families according to the type of reaction being catalyzed.

**Phylogenetic relationships.** Whole genome sequence of 13 fungal species belonging to same class, Agaricomycetes were downloaded from NCBI (<http://ncbi.nlm.nih.gov/>). Phylogenetic tree was constructed using RealPhy software<sup>63</sup> and visualized using the Drawgram module of Phylip tool v3.695<sup>64</sup>. This tool takes number of genomic sequences as input, uses bowtie2 for read mapping and constructs a phylip file containing the input genomic sequences<sup>65</sup>.

**Search for Orthologous gene family.** The complete proteome of the three fungal species i.e. *A. rolfsii*, *G. luxuricans*, *C. subvermispora*, *M. roleri* were downloaded from NCBI database. Orthovenn server<sup>66</sup> was used for identification of shared and unique orthologous gene families clusters in the above mentioned fungal genomes.

**Data Deposition.** The whole-genome sequence and annotation of *Athelia rolfsii* isolate MR10 have been deposited at NCBI (<https://www.ncbi.nlm.nih.gov/>) with accession JZWR00000000; BioSample SAMN03388249.

## References

- Bell, M. & Wright, G. The N<sub>2</sub>-fixing capacity of peanut cultivars with differing assimilate partitioning characteristics. *Aust. J. Agric. Res.* **45**, 1455 (1994).
- Toomsan, B., McDonagh, J. F., Limpinuntana, V. & Giller, K. E. Nitrogen fixation by groundnut and soyabean and residual nitrogen benefits to rice in farmers' fields in Northeast Thailand. *Plant Soil* **175**, 45–56 (1995).
- Jadon, K. S., Thirumalaisamy, P. P., Kumar, V., Koradia, V. G. & Padavi, R. D. Management of soil borne diseases of groundnut through seed dressing fungicides, doi:10.1016/j.cropro.2015.08.021 (2015).
- Doley, K. & Kaur JITE, P. Management of Stem-rot of Groundnut (*Arachis hypogaea* L.) Cultivar in Field. *Not Sci Biol* **5**, 316–324 (2013).
- Yaqub, F. & Shahzad, S. *In Vitro* Evaluation Of Microbial Antagonists Against Sclerotium Rolfsii. *Pak. J. Bot* **37**, 1033–1036 (2005).
- Wilson, C. *Preventing the Diseases of Peanut. United State Department of Agriculture Year Book* (1953).
- Baruah, P., Baruah, L. M. & Baruah, H. K. Analysis of Certain Fungi in Stored Food Products and their Toxic Effect: Zum Nachweis von bestimmten Pilzen in gelagerten Nahrungsmitteln und ihrer toxischen Wirkung. *Mycoses* **23**, 583–589 (2009).
- Mehrotra, R. S. & Aneja, K. R. *An Introduction to Mycology. Wilen Eastern Limited, New Delhi*, doi:10.1093/nar/gki394 (1990).
- Le, C. N., Kruijt, M. & Raaijmakers, J. M. Involvement of phenazines and lipopeptides in interactions between *Pseudomonas* species and *Sclerotium rolfsii*, causal agent of stem rot disease on groundnut. *J. Appl. Microbiol* **112**, 390–403 (2012).
- Fernández Acero, F. J. *et al.* Development of Proteomics-Based Fungicides: New Strategies for Environmentally Friendly Control of Fungal Plant Diseases. *Int. J. Mol. Sci.* **12**, 795–816 (2011).
- Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* btt310 (2013).
- Mohanta, T. K. & Bae, H. The diversity of fungal genome. *Biol. Proced. Online* **17**, 8 (2015).
- Burkhardt, I. *et al.* Mechanistic Characterisation of Two Sesquiterpene Cyclases from the Plant Pathogenic Fungus *Fusarium fujikuroi*. *Angew. Chemie Int. Ed.* **55**, 8748–8751 (2016).
- Quin, M. B., Flynn, C. M. & Schmidt-Dannert, C. Traversing the fungal terpenome. *Nat. Prod. Rep.* **31**, 1449–1473 (2014).
- Collado, I. G., Sánchez, A. J. M. & Hanson, J. R. Fungal terpene metabolites: biosynthetic relationships and the control of the phytopathogenic fungus *Botrytis cinerea*. *Nat. Prod. Rep.* **24**, 674–686 (2007).
- Avalos, J. & Limón, M. C. Biological roles of fungal carotenoids. *Current Genetics* **61**, 309–324 (2015).
- Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceed. Nat. Acad. Sci* **111**, 9259–9264 (2014).
- Bushley, K. E. & Turgeon, B. G. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol. Biol.* **10**, 26 (2010).
- Martínez-Núñez, M. A. & López, V. E. L. Nonribosomal peptides synthetases and their applications in industry. *Sust. Chem. Proc* **4**, 13 (2016).
- Liu, M.-M., Xing, Y.-M., Zhang, D.-W. & Guo, S.-X. Transcriptome analysis of genes involved in defence response in *Polyporus umbellatus* with *Armillaria mellea* infection. *Sci. Rep.* **5**, 16075 (2015).
- Roohparvar, R., Huser, A., Zwiers, L.-H. & De Waard, M. A. Control of *Mycosphaerella graminicola* on wheat seedlings by medical drugs known to modulate the activity of ATP-binding cassette transporters. *Appl. Environ. Microbiol.* **73**, 5011–9 (2007).
- Zhao, X., Mehrabi, R. & Xu, J.-R. Mitogen-activated protein kinase pathways and fungal pathogenesis. *Eukaryot. Cell* **6**, 1701–14 (2007).
- Leng, Y. & Zhong, S. The Role of Mitogen-Activated Protein (MAP) Kinase Signaling Components in the Fungal Development, Stress Response and Virulence of the Fungal Cereal Pathogen *Bipolaris sorokiniana*. *PLoS One* **10**, e0128291 (2015).



24. Hamel, L.-P., Nicole, M.-C., Duplessis, S. & Ellis, B. E. Mitogen-activated protein kinase signaling in plant-interacting fungi: distinct messages from conserved messengers. *Plant Cell* **24**, 1327–51 (2012).
25. Wöstemeyer, J. & Kreibich, A. Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Curr. Genet.* **41**, 189–198 (2002).
26. Karaoglu, H., Lee, C. M. Y. & Meyer, W. Survey of Simple Sequence Repeats in Completed Fungal Genomes. *Mol. Biol. Evol.* **22**, 639–649 (2004).
27. Martinez, D. *et al.* Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.* **22**, 695–700 (2004).
28. Testa, A. C., Oliver, R. P. & Hane, J. K. OcculterCut: A Comprehensive Survey of AT-Rich Regions in Fungal Genomes. *Genome Biol. Evol.* **8**, 2044–2064 (2016).
29. Moges, A. D. *et al.* Development of Microsatellite Markers and Analysis of Genetic Diversity and Population Structure of *Colletotrichum gloeosporioides* from Ethiopia. *PLoS One* **11**, e0151257 (2016).
30. Canfora, L. *et al.* Development of a method for detection and quantification of *B. brongniartii* and *B. bassiana* in soil. *Sci. Rep.* **6**, 22933 (2016).
31. Houston, K., Tucker, M. R., Chowdhury, J., Shirley, N. & Little, A. The Plant Cell Wall: A Complex and Dynamic Structure As Revealed by the Responses of Genes under Stress Conditions. *Front. Plant Sci.* **7**, 984 (2016).
32. Hok, S., Attard, A. & Keller, H. Getting the Most from the Host: How Pathogens Force Plants to Cooperate in Disease. *Mol. Plant-Microbe Interact* **23**, 1253–1259 (2010).
33. Cools, H. J. & Hammond-Kosack, K. E. Exploitation of genomics in fungicide research: current status and future perspectives. *Mol. plant pathol.* **14**, 197–210 (2013).
34. Deising, H. B., Reimann, S. & Pascholati, S. F. Mechanisms and significance of fungicide resistance. *Braz. J. Microbiol.* **39**, 286–95 (2008).
35. Lucas, J. A., Hawkins, N. J. & Fraaije, B. A. In *Advan. in app microbiol.* **90**, 29–92 (2015).
36. Seringhaus, M., Paccanaro, A., Borneman, A., Snyder, M. & Gerstein, M. Predicting essential genes in fungal genomes. *Genome Res.* **16**, 1126–35 (2006).
37. Sun, B.-F. *et al.* Multiple Interkingdom Horizontal Gene Transfers in Pyrenophora and Closely Related Species and Their Contributions to Phytopathogenic Lifestyles. *PLoS One* **8**, e60029 (2013).
38. Miklos, G. L. G. & Rubin, G. M. The Role of the Genome Project in Determining Gene Function: Insights from Model Organisms. *Cell* **86**, 521–529 (1996).
39. Zhao, Z., Liu, H., Wang, C. & Xu, J.-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **14**, 274 (2013).
40. Luo, J. *et al.* Phylogenomic analysis uncovers the evolutionary history of nutrition and infection mode in rice blast fungus and other Magnaporthales. *Scientific Rep.* **5**, 9448 (2015).
41. Kretzer, A., Li, Y., Szaro, T. & Bruns, T. D. Internal transcribed spacer sequences from 38 recognized species of *Suillus* sensu lato: phylogenetic and taxonomic implications. *Mycologia* **776**–785 (1996).
42. Köhler, G. A. *et al.* Phospholipase A 2 and phospholipase B activities in fungi. *Biochimica et Biophysica Acta (BBA)-Mol. and Cell Biol. Lipids* **1761**, 1391–1399 (2006).
43. Moktali, V. *et al.* Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. *BMC Genomics* **13**, 525 (2012).
44. Okhravi, N. *et al.* Polymerase chain reaction and restriction fragment length polymorphism mediated detection and speciation of *Candida* spp causing intraocular infection. *Invest. Ophthalmol. Vis. Sci.* **39**, 859–66 (1998).
45. Fernandes, A. S. *et al.* Modulation of fungal sensitivity to staurosporine by targeting proteins identified by transcriptional profiling. *Fungal Genet. Biol.* **48**, 1130–8 (2011).
46. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**, 31–37 (2014).
47. Chevreur, B. *et al.* Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**, 1147–59 (2004).
48. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–77 (1999).
49. Borodovsky, M. & Lomsadze, A. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr. Protoc. Bioinforma.* Chapter 4, Unit 4.6.1–10 (2011).
50. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–5 (2010).
51. Min, X. J., Butler, G., Storms, R. & Tsang, A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* **33**, W677–W680 (2005).
52. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
53. Conesa, A. & Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**, 619832 (2008).
54. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
55. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015).
56. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
57. Bendtsen, J. D., Jensen, L. J., Blom, N., von Heijne, G. & Brunak, S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* **17**, 349–356 (2004).
58. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
59. Möller, S., Croning, M. D. & Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–53 (2001).
60. Eisenhaber, B., Schneider, G., Wildpaner, M. & Eisenhaber, F. A Sensitive Predictor for Potential GPI Lipid Modification Sites in Fungal Protein Sequences and its Application to Genome-wide Studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J. Mol. Biol.* **337**, 243–253 (2004).
61. Urban, M. *et al.* The Pathogen-Host Interactions database (PHI-base): additions and future developments. *Nucleic Acids Res.* **43**, D645–D655 (2015).
62. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–51 (2012).
63. Bertels, F., Silander, O. K., Pachkov, M., Rainey, P. B. & van Nimwegen, E. Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads. *Mol. Biol. Evol.* **31**, 1077–1088 (2014).
64. Felsenstein, J. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
65. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9 (2012).
66. Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y. Q. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **43**, W78–W84 (2015).

## Acknowledgements

Authors are thankful to Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India for providing financial assistance in the form of CABIn grant (F. no. Agril. Edn.4-1/2013-A&P) as well as Advanced Super Computing Hub for Omics Knowledge in Agriculture (ASHOKA) facility at ICAR-IASRI, New Delhi, India created under National Agricultural Innovation Project, funded by World Bank.

## Author Contributions

D.K., R.S.T. and M.A.I. conceived theme of the study. R.S.T., M.V.P., V.M.R. and S.M.P. collected samples and generated data. D.S., S.J., M.A.I. and N.K. did the computational analysis of generated data. D.S., S.J., M.A.I. and M.V.P. drafted the manuscript. D.K., M.A.I., S.J., A.R. and R.S.T. edited the manuscript. All co-authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-05478-8](https://doi.org/10.1038/s41598-017-05478-8)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017