Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Whole Genome Sequence of *Wilsonomyces Carpophilus*, the Causal Agent of Shot Hole of Stone Fruits: Insights Into Secreted Proteins of a Necrotrophic Fungal Repository

Mahiya Farooq
Plant Virology and Molecular Pathology Laboratory, Division of Plant Pathology, SKUAST-K
Mehraj D. Shah (mehraj547@rediffmail.com)
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir
Bilal A. Padder
Plant Virology and Molecular Pathology Laboratory, Division of Plant Pathology, SKUAST-K
T.A. Sofi
Plant Virology and Molecular Pathology Laboratory, Division of Plant Pathology, SKUAST-K
Khalid k. Masoodi
Plant Virology and Molecular Pathology Laboratory, Division of Plant Pathology, SKUAST-K
Asha Nabi
Plant Virology and Molecular Pathology Laboratory, Division of Plant Pathology, SKUAST-K
Sumaira Hamid
Plant Virology and Molecular Pathology Laboratory, Division of Plant Pathology, SKUAST-K
Mushtaq Ahmad
Division of Biotechnology, FOH, SKUAST-K

Research Article

Keywords: Stigmina blight, Thyrostroma carpophilum, Illumina HiSeq, PacBio, Next Generation Sequencing (NGS)

Posted Date: May 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-541535/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Wilsonomyces carpophilus is a necrotrophic plant pathogenic fungus with a wide host range infecting all stone fruits such as peach, plum, apricot and cherry, and almonds among the nut crops. Necrotrophs are more devastating with a complex pathogenicity mechanism and least known effector repositories. Here, we report a 29.9 megabase draft genome assembly of *W. carpophilus*. We explored the hybrid technology of Illumina HiSeq and PacBio sequencing technologies to get the unbiased results of sequence reads. We aligned short Illumina reads against the long PacBio reads. A total of 10,901 protein-coding genes were predicted that includes varied set of genes such as HET genes, cytochrome-p450 genes, kinases etc. We mined 2851 simple sequence repeats (SSRs) in the genome assembly. We also predicted the diverse inventory of secretory proteins, transporters, primary and secondary metabolic enzymes. A total of 225 secreted proteins, hydrolases, polysaccharide-degrading enzymes, esterolytic, lipolytic and proteolytic enzymes were the most significant proteins reflecting the necrotrophic lifestyle of the *W. carpophilus*. We also identified 146 tRNAs and 52 rRNAs in the pathogen genome.

Introduction

The stone fruits that includes peach, plum, cherry, apricot, nectarine and, almonds among nut crops are important crops grown throughout the world. The foremost growing countries are America, Australia, Afghanistan, China, Iran, Italy, Greece, France, New Zealand, Portugal, India and Central Asian countries of earlier USSR [1]. Among the number of biotic factors affecting stone fruits among, shot hole disease caused by *Wilsonomyces carpophilus* is of paramount importance [2]. Shot hole disease is one of the major fungal foliar diseases in *Prunus* species worldwide [3]. The disease is reported from Africa, Asia, America (North, South, Central), Australia and Oceania [4]. Recently, [5] reported the shot hole disease of stone fruits (*Prunus* spp.) as a major threat to the wild-fruit forest of the Western Tianshan Mountains of China. The shot hole disease of *Prunus* spp. is also reported from California and Poland (https://nt.ars-grin.gov/ fungaldatabas-es/). In literature, a number of synonymous exits for the pathogen such as *Thyrostroma carpophilum, Stigmina carpophila* and *W.carpophilus*, however, recently [6] proposed *W. carpophilus* to name the pathogen. The intermittent outbreak of the disease causes notable yield losses ranging from 30 to 90% in cherry, and about 60.3% in apricot in Malatya province of Turkey [1]. The disease appear as small circular reddish or purplish lesion with yellow halo, the centre gradually enlarges and become necrotic that ultimately fall down leaving a shot hole appearance [7]. The fungus shows cross pathogenicity on different hosts under *in vitro* conditions [8] suggesting that the pathogen lacks specificity towards a particular host species and therefore causes disease in all the stone fruits, and almonds in nut crops. Such studies indicates broad host range of the pathogen and therefore needs profound study before devising a better management capsule. Existence of high pathological and molecular diversity in *W. carpophilus* hampers resistance breeding, a viable disease management alternative [1].

In literature, a number of synonym such as *Clasterosporium carpophilum* (Lev.), *Stigmina carpophila* (Lev.), *Thyrostroma carpophilum* and *W. carpophilus* exists for the pathogen [9], However, recently, [6]

considered *W. carpophilus* as a causal organism of shot hole disease of *Prunus* spp. *W. carpophilus* is an orphan plant pathogen in terms of studies conducted in *Prunus* - shot hole interface. Although high pathogen diversity and successful development of an ATMT protocol [1,10], for the fungus increased our understanding but many basic questions related to pathosystem warrants additional research.

Genome sequencing of plant pathogens has provided the insights on pathogen life style besides it changed conventional genetics to genomics. Over last few years, hundreds of plant pathogenic fungal genomes have been decoded and pangenomics currently provide deep insights on pathogenicity and life style characteristics of plant pathogens [11,12]. To gain the evolutionary insights about the poorly studied *W. carpophilus* fungus, it is necessary to exploit the high throughput sequencing toolbox for understanding its life style and evolutionary dynamics. Based on the feeding habit, *W. carpophilus* is a necrotroph and to infect diverse *Prunus* spp., the pathogen need to be equipped with pathogenicity arsenals that have the capacity to breach pathogen triggered as well as effector triggered immunity of *Prunus* hosts. Besides secreting the chain of enzymes to degrade the host tissues, nectrotrophs exploit the cell death machinery of the host [13]. Therefore, effectors, the pathogen-encoded secreted proteins play crucial roles in necrotrophs to evade host defense system. Since, it is necessary to understand the pathogenicity mechanism of *W. carpophilus* and studies are needed on how the pathogen is able to manipulate the host cell machinery. Decoding genomes coupled with the fine-tuned bioinformatics pipelines have increased our knowledge about the pathogenicity mechanisms and provided insights on role of secreted effector molecules to evade host defenses. Here we report the first genome draft of *W. carpophilus* with an aim to gain insights on *Prunus*- shot hole pathogen interaction and fungus pathogenicity mechanism. Decoding the *W. carpophilus* genome using hybrid NGS technology provided us clues about its host defense mechanism evading capabilities and the pathogenicity armory the pathogen is harboring in its genome that makes it a successful pathogen of all *Prunus* hosts.

Methods

Fungal culture preparation

The pathogen was isolated from the shot hole infected leaves of stone fruits *viz*., plum, peach, apricot, and cherry and almonds among nut crops grown in University orchard of SKUAST-K, Shalimar, Srinagar (J&K). The purified fungal culture was maintained on Asthana and Hawker's and potato dextrose agar (PDA) [10,14] media. On the basis of morpho-cultural characteristics, the pathogen was identified as *Wilsonomyces carpophilus* synonym *Thyrostroma carpophilum* Nabi [1,10]. The pathogenicity of these isolates was carried out by detached leaf technique on their respective hosts [15] followed by their cross infectivity on different stone fruits including almond.

DNA isolationfor whole genome sequencing

The most virulent isolate of the pathogen based on minimum incubation time and symptom development, was selected for whole genome sequencing. The DNA of the pathogen isolate was

extracted using XcelGen DNA isolation Kit (Xceleris, Ahmedabad, India) according to the manufacturer instructions. The quality and quantity of extracted DNA was checked using a Qubit 2.0 Fluorimeter (Life Technologies Ltd., Paisley, UK). The integrity of DNA (DIN) was checked using Bioanalyser 2100 (Agilent Technologies, Santa Clara, CA).

Library preparation and genome sequencing

The DNA Library was prepared using NEBNext Ultra DNA Library Prep Kit (Biolabs, England). The library preparation process was initiated with 200ng DNA. The adapters were ligated to both ends of the DNA fragments. These adapters contain sequences essential for binding dual-barcoded libraries to a flow cell for sequencing and PCR amplification. To ensure maximum yield from a limited amounts of starting material, a high-fidelity amplification step was performed using PCR Master Mix.

The whole genome of plant pathogenic fungus *W. carpophilus* was decoded using Illumina HiSeq and PacBio sequencing technologies. *De Novo* assembly of high quality paired end reads was accomplished using Velvet v1.2.10 and the assembly was optimized at Kmer-79 (Supplementary Table 2.) (Fig. 7). Further, scaffolding was performed on pre-assembled contigs taking long reads of PacBio using SSPACE-LongRead v1.1. We aligned Illumina short reads on PacBio long reads (a hybrid approach) using PBJelly software and GapCloser v1.12 to increase the precision of base calling.

Gene prediction and annotation

The assembled genome was subjected to gene prediction using Augustus v2.5.5 for the identification of coding sequences. The predicted protein coding genes were subjected to similarity search against NCBI's non-redundant (nr) database using *Uniprot, KOG* and *Pfam* database of BLASTP algorithm with an e-value threshold of 1e-5. Simultaneously, all the proteins were searched for similarity against BLASTP with an e-value threshold of 1e-5. Comparative analysis of gene annotation in different database was carried out using http://www.interactivenn.net/. Gene Ontology (GO) annotation was obtained using nr database through Blast2GO command line v-1.4.1. GO sequence distributions helps in specifying all the annotated nodes comprising of GO functional groups. Genes associated with the similar functions were assigned to same GO functional group. The GO sequence distribution was analyzed for all the three GO domains i.e. biological processes, molecular function and cellular components.

Secretome mining

The secretool was used to predict *W. carpophilus* secretome that enables secretome predictions out of amino acid sequence files (http://genomics.cicbiogune.es/SECRETOOL/Secretool.php). The Signal-P (v4.1) and WoLF PSORT (v0.2) were used to identify signal peptides and extracellular localizations in total of 10901 protein coding genes. The TMHMM (v2.0) and PredGPI (http://gpcr.biocomp.unibo.it/predgpi/pred.htm) were used to eliminate sequences with transmembrane domains, ER-retention signal and GPI (glycosylphosphatidyl inositol)-anchors, respectively (Fig. 8).

Simple sequence repeats

A high-throughput SSR search to identify mono- to hexa- nucleotide SSR motifs was performed using MIcroSAtellite (MISA) identification tool (http://pgrc.ipk-gatersleben.de/misa/download/misa.pl) with default parameters. The default parameters were used so that di-nucleotide pattern should appear at least six times, whereas tri-, tetra-, penta- and hexa- nucleotide motifs should appear five times.

Pathway analysis

Pathway analysis, ortholog assignment and mapping of genes to the biological pathways were performed using KEGG automatic annotation server (KAAS). All the gene sequences were compared against the KEGG database using BLASTP with threshold bit-score value of 60 (default).

Identification of tRNAs and rRNAs in the genome

To identify probable tRNA genes, we used tRNAscan-SE that allows detection of unusual tRNA species with accurate prediction of secondary structures. It includes both prokaryotic and eukaryotic selenocysteine tRNA genes, tRNA-derived repetitive elements and pseudogenes. The RNAmmer 1.2 was used for rRNA gene identification.

Collection of diseased planting material / samples

Necessary permission whenever required was obtained, and all the guidelines and legislation were followed for the collection of diseased planting material or samples from University orchard, SKUAST-K, Shalimar, Srinagar (J&K), India

Results

The introduction of novel sequencing technologies such as PacBio has revolutionized the genomic studies. Long read sequencing platforms are used to determine the complex genomic regions that are difficult to explore with short read length sequencing technologies. However, these long-read sequencing technologies are prone to higher error rates also, therefore, in a present study, we explored hybrid approach of Illumina HiSeq and PacBio to decipher the high quality whole genome sequence of *Wilsonomyces carpophilus,* a plant pathogenic fungus for its announcement for the first time. The genome assembly of *W. carpophilus* shows that the hybrid approach of sequencing is effective in constructing contigs with almost full length genes. To authenticate the genome assembly, we assessed the completeness of the gene space using different softwares and found that the most of the core eukaryotic conserved genes were represented in the assembled genome. It is also commendable that we found 10901 genes using gene prediction algorithm. When these genes were blasted against the nr database, most of the tophits were against the *Pyrenochaeta* spp. followed by *Ascochyta rabiei*. *Pyrenochaeta* genus is comprised of a wide range of species infecting plants and humans. We found that the plant pathogenic species of *Pyrenochaeta* behaves similarly as that of *W. carpophilus* and no sexual stage (mating) has been reported in *Pyrenochaeta lycopersici* (a plant pathogenic fungus) [16] as well as in *W. carpophilus* in the nature [17-19]. Interestingly, the pathogen genome also lacks MAT genes that are the

key factors to decide either sexual or asexual reproduction occurring in the fungus. This evidence clearly shows that the fungus is incompetent for sexual reproduction. Thus, the variability of fungus can be potentially due to the vegetative hyphal fusion or anastomosis, and is evident from the presence of HET genes in the *W. carpophilus* genome. The viable heterokaryon is formed by anastomosis only when individuals have same set of HET genotype, whereas individuals with different HET genotypes forms incompatible vegetative heterokaryon which later on undergoes programmed cell death [20]. The advantage of this is to limit the contamination of the pathogen and other lethal replicons between the strains [21]. The selective pressure is probably responsible for the broad diversification of HET genes and play a key role in the transfer of genetic information between the strains, and variability in the pathogen is indispensable for the adaptation to the environment and to overcome the host defense mechanisms. The other key genes deployed in the pathogen genome were antimicrobial peptide (AMP) binding genes that are used by the pathogen to surmount host defense. These AMPs are the part of plant's innate immunity system against the pathogen attack [22]. The pathogen has an ability to bind these genes in order to surpass plant defense system. AMPs are naturally synthesized low molecular products [up to 100 amino acids (AAs)] that are structurally and biochemically diverse in nature. The diversity in the AMPs suggests the diversity of AMP binding genes in the pathogen that ultimately explaining its diverse host range. We also found cytochrome P450, pKinases, sugar transporters etc. in the pathogen genome. The cytochrome P450 gene is a heme containing protein that are involved in the degradation of plant derived toxins and therefore plays an important role in fungal development eventually in pathogenesis [23]. Higher number of CYPs also indicates the wide host range of the pathogen requiring more toxins to overcome the phytoalexins. Similarly protein kinases play an important role in various key processes of the fungal life cycle such as growth direction, nutrient uptake, stress responses and reproduction [24], thus can be an important factor in the infection process of the pathogen. Secretory proteins also play a crucial role in the fungal pathogenesis and colonization, and a set of secretory proteins suggests the feeding habit of the pathogen as biotroph, hemibiotroph or necrotroph. The efficiency and hostility of the phytopathogens are often associated with the presence of cell wall degrading enzymes. The first and foremost obstacle to fungal pathogens in plants is the cell wall and its associated components. Plant pathogenic fungi secrete a concoction of hydrolytic enzymes known as carbohydrate-active enzymes (CAZymes), that are required to degrade the cell wall components of the host [25]. The presences of cell wall degrading enzymes such as glycosidases, glucanases, carboxylesterases, laccases, pectatelyases, cellulases etc. in the *W. carpophilus* genome are more often predicted in necrotrophs. The secretion and presence of these enzymes in the *W. carpophilus* showed its stronger resemblance to the necrotrophic plant pathogens. Unlike biotrophs, the necrotrophs have significantly expanded set of cell wall degrading enzymes, thus secretome of the *W. carpophilus* suggests necrotrophic behavior of the pathogen. The secretome of the fungus revealed some other proteins that also suggests its necrotrophic behavior such as FAD binding domains that have ability to catalyse various biochemical reactions and are mainly involved in electron transport chain [26]. We also found some other secretory proteins such as chaperone proteins also known as heat shock proteins that play an important role in the pathogenesis. They maintain the integrity of the pathogen in the adverse conditions, thus making it more viable for infection [27]. Laccasse precursor that

plays an important role in lignin depolymerization of infected host [28] and number of other enzymes such as endo-1, 4-beta-galactosidase, cutinase, rhamnogalacturonan lyase, pectin-esterase, pectate lyase, lipolytic protein etc. that play a crucial role in the pathogenicity. Surprisingly, some pathogenicity determinants that plays a role in disarming the host defense such as mycelial catalases that are known to degrade the hydrogen peroxide produced by the host as a result of oxidative burst to kill the pathogen [29]. The presence of these enzymes suggests that the pathogen secrete diverse proteins that co-ordinate in an organized manner to cause disease and evade the host defense. However, these pathogenicity determinants need further characterization and validation to get in depth insight of pathogenicity mechanism of the fungus. In our previous study, we were successful in transforming the fungus by using random ATMT method and the transformants were unable to cause the disease[10]. However, the present study has revealed the number of pathogenicity genes that can be easily targeted to render pathogen ineffective. The present study has opened new opportunities for the comprehensive genomic study of a variety of biological, metabolic and pathological aspects that make the *W. carpophilus* a successful necrotrophic pathogen. Therefore, it is an opportune time to go beyond the conventional neutral genetics by identifying, analyzing, site specific targeting of pathogenicity determinants and re-modelling the core effector repositories.

Discussion

The introduction of novel sequencing technologies such as PacBio has revolutionized the genomic studies. Long read sequencing platforms are used to determine the complex genomic regions that are difficult to explore with short read length sequencing technologies. However, these long-read sequencing technologies are prone to higher error rates also, therefore, in a present study, we explored hybrid approach of Illumina HiSeq and PacBio to decipher the high quality whole genome sequence of *Wilsonomyces carpophilus,* a plant pathogenic fungus for its announcement for the first time. The genome assembly of *W. carpophilus* shows that the hybrid approach of sequencing is effective in constructing contigs with almost full length genes. To authenticate the genome assembly, we assessed the completeness of the gene space using different softwares and found that the most of the core eukaryotic conserved genes were represented in the assembled genome. It is also commendable that we found 10901 genes using gene prediction algorithm. When these genes were blasted against the nr database, most of the tophits were against the *Pyrenochaeta* spp. followed by *Ascochyta rabiei*. *Pyrenochaeta* genus is comprised of a wide range of species infecting plants and humans. We found that the plant pathogenic species of *Pyrenochaeta* behaves similarly as that of *W. carpophilus* and no sexual stage (mating) has been reported in *Pyrenochaeta lycopersici* (a plant pathogenic fungus) [16] as well as in *W. carpophilus* in the nature [17-19]. Interestingly, the pathogen genome also lacks MAT genes that are the key factors to decide either sexual or asexual reproduction occurring in the fungus. This evidence clearly shows that the fungus is incompetent for sexual reproduction. Thus, the variability of fungus can be potentially due to the vegetative hyphal fusion or anastomosis, and is evident from the presence of HET genes in the *W. carpophilus* genome. The viable heterokaryon is formed by anastomosis only when individuals have same set of HET genotype, whereas individuals with different HET genotypes forms

incompatible vegetative heterokaryon which later on undergoes programmed cell death [20]. The advantage of this is to limit the contamination of the pathogen and other lethal replicons between the strains [21]. The selective pressure is probably responsible for the broad diversification of HET genes and play a key role in the transfer of genetic information between the strains, and variability in the pathogen is indispensable for the adaptation to the environment and to overcome the host defense mechanisms. The other key genes deployed in the pathogen genome were antimicrobial peptide (AMP) binding genes that are used by the pathogen to surmount host defense. These AMPs are the part of plant's innate immunity system against the pathogen attack [22]. The pathogen has an ability to bind these genes in order to surpass plant defense system. AMPs are naturally synthesized low molecular products [up to 100 amino acids (AAs)] that are structurally and biochemically diverse in nature. The diversity in the AMPs suggests the diversity of AMP binding genes in the pathogen that ultimately explaining its diverse host range. We also found cytochrome P450, pKinases, sugar transporters etc. in the pathogen genome. The cytochrome P450 gene is a heme containing protein that are involved in the degradation of plant derived toxins and therefore plays an important role in fungal development eventually in pathogenesis [23]. Higher number of CYPs also indicates the wide host range of the pathogen requiring more toxins to overcome the phytoalexins. Similarly protein kinases play an important role in various key processes of the fungal life cycle such as growth direction, nutrient uptake, stress responses and reproduction [24], thus can be an important factor in the infection process of the pathogen. Secretory proteins also play a crucial role in the fungal pathogenesis and colonization, and a set of secretory proteins suggests the feeding habit of the pathogen as biotroph, hemibiotroph or necrotroph. The efficiency and hostility of the phytopathogens are often associated with the presence of cell wall degrading enzymes. The first and foremost obstacle to fungal pathogens in plants is the cell wall and its associated components. Plant pathogenic fungi secrete a concoction of hydrolytic enzymes known as carbohydrate-active enzymes (CAZymes), that are required to degrade the cell wall components of the host [25]. The presences of cell wall degrading enzymes such as glycosidases, glucanases, carboxylesterases, laccases, pectatelyases, cellulases etc. in the *W. carpophilus* genome are more often predicted in necrotrophs. The secretion and presence of these enzymes in the *W. carpophilus* showed its stronger resemblance to the necrotrophic plant pathogens. Unlike biotrophs, the necrotrophs have significantly expanded set of cell wall degrading enzymes, thus secretome of the *W. carpophilus* suggests necrotrophic behavior of the pathogen. The secretome of the fungus revealed some other proteins that also suggests its necrotrophic behavior such as FAD binding domains that have ability to catalyse various biochemical reactions and are mainly involved in electron transport chain [26]. We also found some other secretory proteins such as chaperone proteins also known as heat shock proteins that play an important role in the pathogenesis. They maintain the integrity of the pathogen in the adverse conditions, thus making it more viable for infection [27]. Laccasse precursor that plays an important role in lignin depolymerization of infected host [28] and number of other enzymes such as endo-1, 4-beta-galactosidase, cutinase, rhamnogalacturonan lyase, pectin-esterase, pectate lyase, lipolytic protein etc. that play a crucial role in the pathogenicity. Surprisingly, some pathogenicity determinants that plays a role in disarming the host defense such as mycelial catalases that are known to degrade the hydrogen peroxide produced by the host as a result of oxidative burst to kill the pathogen

[29]. The presence of these enzymes suggests that the pathogen secrete diverse proteins that co-ordinate in an organized manner to cause disease and evade the host defense. However, these pathogenicity determinants need further characterization and validation to get in depth insight of pathogenicity mechanism of the fungus. In our previous study, we were successful in transforming the fungus by using random ATMT method and the transformants were unable to cause the disease[10]. However, the present study has revealed the number of pathogenicity genes that can be easily targeted to render pathogen ineffective. The present study has opened new opportunities for the comprehensive genomic study of a variety of biological, metabolic and pathological aspects that make the *W. carpophilus* a successful necrotrophic pathogen. Therefore, it is an opportune time to go beyond the conventional neutral genetics by identifying, analyzing, site specific targeting of pathogenicity determinants and re-modelling the core effector repositories.

Declarations

Acknowledgment

The authors are highly thankful to M/S Xcelris Labs Limited, Ahmedabad-380015, India, for their service and help provided during the present study.

Conflict of interest

The authors declare no conflict of interest during the present study

References

1. Nabi, A., Padder, B., Dar, M. & Ahmad, M. Morpho-cultural, pathological and molecular variability in Thyrostroma carpophilum causing shot hole of stone fruits in India. *European Journal of Plant Pathology.* **151**, 613–627 (2018).
2. Bird, G. W. *et al. Compendium of stone fruit diseases* (The American Phytopathological Society, 1995).
3. Youssefi, A. & Hajian Shahri, M. Shot hole disease, survival and pathogenicity of the causal agent on stone fruit trees in Northeast Iran. *Journal of Crop Protection.* **3**, 563–572 (2014).
4. Väcäroju, C., Zalä, C. & Cristea, S. in *scientific conferences with international participation durable agriculture-Agriculture of future, 4th edition. Craiova (Romania): Faculty of Agriculture, University of Craiova.*
5. Ye, S. *et al.* DNA Phylogeny, and Pathogenicity of Wilsonomyces carpophilus Isolate Causing Shot-Hole Disease of Prunus divaricata and Prunus armeniaca in Wild-Fruit Forest of Western Tianshan Mountains, China. *Forests.* **11**, 319 (2020).
6. Marin-Felix, Y. *et al.* Genera of phytopathogenic fungi: GOPHY 1. *Studies in mycology.* **86**, 99–216 https://doi.org/10.1016/j.simyco.2017.04.002 (2017).

7. Ivanová, H., Kalocaiová, M. & Bolvansky, M. Shot-hole disease on Prunus persica-the morphology and biology of Stigmina carpophila. *Folia Oecologica.* **39**, 21 (2012).
8. Raabe, R. Shot hole of Catalina Cherry.*Phytopathology***47** (1957).
9. Ellis, M. B. Clasterosporium and some allied Dematiaceae-Phragmosporae, I, II. Mycological Papers(1959).
10. Rasool, R. S. *et al.* Thyrostroma carpophilum insertional mutagenesis: A step towards understanding its pathogenicity mechanism. *J Microbiol Methods.* **171**, 105885 https://doi.org/10.1016/j.mimet.2020.105885 (2020).
11. Mahmoud, M., Zywicki, M., Twardowski, T. & Karlowski, W. M. Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. *Genomics.* **111**, 43–49 (2019).
12. Steiner, C. C., Putnam, A. S. & Hoeck, P. E. & Ryder, O. A. Conservation genomics of threatened animal species. *Annu. Rev. Anim. Biosci.* **1**, 261–281 (2013).
13. Liu, Z. *et al.* SnTox3 acts in effector triggered susceptibility to induce disease on wheat carrying the Snn3 gene. *PLoS Pathog.* **5**, e1000581 https://doi.org/10.1371/journal.ppat.1000581 (2009).
14. Eddleman, H. Making bacteria media from potato.*Indiana Biolab*(1998).
15. Nabi, A. *et al.* Morphological characterisation and media preferences in Wilsonomyces carpophilus, the causal agent of shot hole disease of stone fruits in Kashmir. *Journal of Pharmacognosy and Phytochemistry.* **7**, 1326–1331 (2018).
16. Aragona, M. *et al.* De novo genome assembly of the soil-borne fungus and tomato pathogen Pyrenochaeta lycopersici. *BMC genomics.* **15**, 313 (2014).
17. Samuel, G. On the shot-hole disease caused by Clasterosporium carpophilum and on the'shot-hole'effect. *Annals of Botany.* **41**, 375–404 (1927).
18. Adaskaveg, J., Shaw, D. & Ogawa, J. A mist generator and environmental monitoring system for field studies on shothole disease of almond. *Plant disease.* **74**, 558–562 (1990).
19. Wilson, E. E. The shot-hole disease of stone-fruit trees (No. Folleto1106). (1937).
20. Hall, C., Welch, J., Kowbel, D. J. & Glass, N. L. Evolution and diversity of a fungal self/nonself recognition locus. *PloS one.* **5**, e14055 (2010).
21. Tuite, M. F. & Serio, T. R. The prion hypothesis: from biological anomaly to basic regulatory mechanism. *Nature reviews Molecular cell biology.* **11**, 823–833 (2010).
22. Breen, S., Solomon, P. S., Bedon, F. & Vincent, D. Surveying the potential of secreted antimicrobial peptides to enhance plant disease resistance. *Frontiers in plant science.* **6**, 900 (2015).
23. Shin, J., Kim, J. E., Lee, Y. W. & Son, H. Fungal cytochrome P450s and the P450 complement (CYPome) of Fusarium graminearum. *Toxins.* **10**, 112 (2018).
24. Turrà, D., Segorbe, D. & Di Pietro, A. Protein kinases in plant-pathogenic fungi: conserved regulators of infection. *Annual Review of Phytopathology.* **52**, 267–288 (2014).
25. Condon, B. J. *et al.* Comparative genome structure, secondary metabolite, and effector coding capacity across Cochliobolus pathogens. *PLoS Genet.* **9**, e1003233 (2013).

26. Pigné, S. *et al.* A flavoprotein supports cell wall properties in the necrotrophic fungus Alternaria brassicicola. *Fungal biology and biotechnology.* **4**, 1–13 (2017).
27. Neckers, L. & Tatu, U. Molecular chaperones in pathogen virulence: emerging new targets for therapy. *Cell host & microbe.* **4**, 519–527 (2008).
28. Edens, W. A., Goins, T. Q., Dooley, D. & Henson, J. M. Purification and characterization of a secreted laccase of Gaeumannomyces graminis var. tritici. *Applied and Environmental Microbiology.* **65**, 3071–3074 (1999).
29. Garre, V., Tenberge, K. B. & Eising, R. Secretion of a fungal extracellular catalase by Claviceps purpurea during infection of rye: putative role in pathogenicity and suppression of host defense. *Phytopathology.* **88**, 744–753 (1998).

Tables

Table 1: Genomic characteristics of *Wilsonomyces carpophilus*

Characteristics	Value
Total no of scaffolds	130
Total genome size including gaps ('Ns')	29992279
Total genome size without gaps ('Ns')	29984629
Scaffolds N50	662300
Average scaffold length	117586
Maximum scaffold length	2185089
GC content	49.77%

Table 2: Distribution of secreted proteins throughout the *Wilsonomyces carpophilus* genome

Secreted protein	Scaffold no.	Gene no.	e-value	Species
FAD binding domain-containing protein	Scaffold 43	gene 9152	1.00E-83	*Magnaporthe oryzae*
Ricin B lectin	Scaffold 46	gene 9390	9.00E-13	*Clostridium thermocellum*
Protein disulfide-isomerase tigA precursor	Scaffold 47	gene 9460	2.00E-110	*Neurospora crassa*
Protein disulfide isomerase	Scaffold 47	gene 9460	1.00E-52	*Dictyostelium discoideum*
Para-nitrobenzyl esterase	Scaffold 48	gene 9543	2.00E-138	*Magnaporthe oryzae*
Probable galactose oxidase precursor	Scaffold 50	gene 9571	4.00E-125	*Neurospora crassa*
Kelch repeat-containing protein	Scaffold 50	gene 9571	2.00E-29	*Nostoc punctiforme*
Probable exported protease	Scaffold 55	gene 9807	4.00E-15	*Mycobacterium tuberculosis*
Laccase precursor	Scaffold 65	gene 10095	4.00E-127	*Neurospora crassa*
Laccase precursor	Scaffold 89	gene 10486	9.00E-132	*Neurospora crassa*
Endoglucanase IV precursor	Scaffold 98	gene 10521	2.00E-22	*Neurospora crassa*
Alpha-Glucosidase precursor	Scaffold 112	gene 10729	2.00E-168	*Neurospora crassa*
G-D-S-L lipolytic protein	Scaffold 1	gene 106	4.00E-13	*Clostridium thermocellum*
G-D-S-L lipolytic protein	Scaffold 1	gene 106	6.00E-06	*Clostridium thermocellum*
Probable DFG5 protein	Scaffold 1	gene 360	2.00E-92	*Neurospora crassa*
Phospholipase /carboxylesterase	Scaffold 2	gene 657	2.00E-16	*Trichodesmium erythraeum*
Probable DFG5 protein	Scaffold 2	gene 661	4.00E-102	*Neurospora crassa*
Beta-Hexosaminidase	Scaffold 2	gene 715	3.00E-25	*Xylella fastidiosa*
CATB_EMENI Catalase B	Scaffold	gene	0	*Aspergillus nidulans*

	2	809		
Mycelial catalase Cat1	Scaffold 2	gene 809	0	*Aspergillus fumigatus*
Catalase 3 precursor	Scaffold 2	gene 809	0	*Neurospora crassa*
Chaperone protein DnaJ	Scaffold 2	gene 963	3.00E-50	*Xylella fastidiosa*
Endo-1, 4-beta-galactosidase	Scaffold 4	gene 1832	6.00E-127	*Neurospora crassa*
Glycosyl hydrolase family 53	Scaffold 4	gene 1832	9.00E-55	*Enterococcus faecium*
Protein ATG-5, isoform c	Scaffold 6	gene 2389	3.00E-16	*Caenorhabditis elegans*
Cutinase	Scaffold 6	gene 2543	8.00E-08	*Mycobacterium bovis*
Vacuolar protease A precursor	Scaffold 6	gene 2597	6.00E-21	*Neurospora crassa*
AGAP003277-PA	Scaffold 6	gene 2597	2.00E-18	*Anopheles gambiae*
Acetylornithine deacetylase	Scaffold 6	gene 2686	2.00E-29	*Actinobacillus pleuropneumoniae*
Phytase	Scaffold 8	gene 3185	2.00E-08	*Nostoc punctiforme*
Rhamnogalacturonan acetylesterase precursor	Scaffold 8	gene 3288	2.00E-72	*Neurospora crassa*
G-D-S-L lipolytic protein	Scaffold 8	gene 3288	4.00E-19	*Clostridium thermocellum*
G-D-S-L lipolytic protein	Scaffold 8	gene 3288	8.00E-08	*Clostridium thermocellum*
Glycerophosphoryl diester phosphodiesterase	Scaffold 9	gene 3772	6.00E-12	*Nostoc punctiforme*
Glycerophosphoryl diester phosphodiesterase	Scaffold 9	gene 3772	1.00E-10	*Nostoc punctiforme*
L-asparaginase II	Scaffold 10	gene 4114	2.00E-65	*Yersinia pestis*
Type II L-Asparaginase	Scaffold 10	gene 4114	1.00E-49	*Azotobacter vinelandii*
Rhamnogalacturonan lyase	Scaffold 12	gene 4368	7.00E-21	*Aspergillus nidulans*

Rhamnogalacturonase	Scaffold 12	gene 4368	3.00E-19	
Alpha-1,6-mannosyltransferase	Scaffold 12	gene 4398	2.00E-91	*Magnaporthe oryzae*
Histidine kinase	Scaffold 12	gene 4401	2.00E-15	*Cytophaga hutchinsonii*
Calcium-related spray protein	Scaffold 12	gene 4405	3.00E-29	*Neurospora crassa*
Carboxypeptidase A4 preproprotein	Scaffold 12	gene 4449	7.00E-06	*Homo sapiens*
Beta (1-3) glucanosyltransferase	Scaffold 12	gene 4500	5.00E-148	*Neurospora crassa*
Dolichyl-phosphate-mannose-protein mannosyltransferase 1	Scaffold 16	gene 4860	1.00E-31	*Magnaporthe oryzae*
Phospholipase/carboxylesterase	Scaffold 17	gene 5083	1.00E-12	*Nostoc punctiforme*
Possible conserved lipoprotein LpqP	Scaffold 17	gene 5083	2.00E-08	*Mycobacterium tuberculosis*
Pectin-esterase	Scaffold 18	gene 5419	1.00E-29	*Clostridium thermocellum*
Cathepsin D	Scaffold 18	gene 5432	2.00E-30	*Dictyostelium discoideum*
Aspartic endopeptidase Pep2	Scaffold 18	gene 5432	9.00E-28	*Aspergillus fumigatus*
Beta-glucosidase	Scaffold 19	gene 5495	6.00E-46	*Neurospora crassa*
Glycosyl Hydrolase Family 88	Scaffold 20	gene 5660	2.00E-49	*Enterococcus faecium*
Crystal protein	Scaffold 21	gene 5861	1.00E-23	*Dictyostelium discoideum*
Carboxypeptidase A2 (pancreatic)	Scaffold 22	gene 6186	1.00E-18	*Homo sapiens*
Carboxypeptidase A4 preproprotein	Scaffold 22	gene 6186	3.00E-18	*Homo sapiens*
Galactose oxidase precursor	Scaffold 22	gene 6227	4.00E-47	*Neurospora crassa*
Pectate lyase precursor	Scaffold 27	gene 7051	3.00E-48	*Aspergillus nidulans*
Pectate lyase	Scaffold 27	gene 7051	1.00E-18	*Cytophaga hutchinsonii*

Oryzin	Scaffold 29	gene 7186	1.00E-104	*Magnaporthe oryzae*
Acetylxylan esterase precursor	Scaffold 32	gene 7974	9.00E-82	*Neurospora crassa*
Aliphatic nitrilase	Scaffold 32	gene 8021	2.00E-07	*Neurospora crassa*
Proteinase T precursor	Scaffold 34	gene 8276	4.00E-18	*Neurospora crassa*
Autophagic serine protease Alp2	Scaffold 34	gene 8276	5.00E-118	*Aspergillus fumigatus*
Proteinase T precursor	Scaffold 34	gene 8277	3.00E-68	*Neurospora crassa*
Autophagic serine protease Alp2	Scaffold 34	gene 8277	9.00E-67	*Aspergillus fumigatus*
Endoglucanase (carboxymethyl cellulase)	Scaffold 35	gene 8314	7.00E-40	*Mycobacterium tuberculosis*
Carbonic anhydrase	Scaffold 35	gene 8495	4.00E-15	*Vibrio cholerae*
Carbonic anhydrase	Scaffold 35	gene 8495	1.00E-17	*Oenococcus oeni*
Actin-like protein 3	Scaffold 38	gene 8752	1.00E-17	*Aspergillus nidulans*
Actin-like protein 3	Scaffold 38	gene 8752	2.00E-176	*Fusarium graminearum*
Actin-like protein	Scaffold 38	gene 8752	1.00E-168	*Dictyostelium discoideum*
Beta-glucuronidase isoform 1 precursor	Scaffold 41	gene 8920	9.00E-120	*Homo sapiens*
Beta-D-glucuronidase	Scaffold 41	gene 8920	1.00E-24	*Haemophilus somnus*
FAD binding domain-containing protein	Scaffold 41	gene 8926	8.00E-22	*Magnaporthe oryzae*
FAD binding domain-containing protein	Scaffold 41	gene 8926	9.00E-93	*Magnaporthe oryzae*

Table 3: SSR distribution across the *Wilsonomyces carpophilus* genome

SSR distribution	Count
Total number of sequences examined	130
Total size of examined sequences (bp)	29987235
Total number of identified SSRs	2851
Number of SSR containing sequences	91
Number of sequences containing more than 1 SSR	77
Number of sequences with 150bp flanking region	2592
Number of dinucleotide motifs	937
Number of trinucleotide motifs	1362
Number of tetranucleotide motifs	265
Number of pentanucleotide motifs	141
Number of hexanucleotide motifs	146

Table 4: Pfam gene annotation of *Wilsonomyces carpophilus* genome

Pfam domain	*Total*
Pkinase	86
p450	69
adh_short	61
Sugar_tr	52
AMP-binding	37
HET	27
COesterase	24
AAA	23
Aldo_ket_red	21
DEAD	21

Table 5: Diverse metabolic processes of the *Wilsonomyces carpophilus* predicted using KEGG pathway analysis

Metabolic pathway	count
Carbohydrate metabolism	328
Energy metabolism	117
Lipid metabolism	173
Nucleotide metabolism	73
Amino acid metabolism	287
Metabolism of other amino acids	68
Glycan biosynthesis and metabolism	81
Metabolism of cofactors and vitamins	128
Metabolism of terpenoids and polyketides	27
Biosynthesis of other secondary metabolites	48
Xenobiotics biodegradation and metabolism	72
Transcription	97
Translation	264
Folding, sorting and degradation	215
Replication and repair	135
Membrane transport	11
Signal transduction	377
Transport and catabolism	300
Cell motility	17
Cell growth and death	296
Cellular community - eukaryotes	39
Cellular community - prokaryotes	15
Development and regeneration	14
Environmental adaptation	67
Protein families: metabolism	364
Protein families: genetic information processing	1930
Protein families: signaling and cellular processes	436
Unclassified: metabolism	83

Unclassified: genetic information processing	4
Unclassified: signaling and cellular processes	13
Poorly characterized	33

Figures

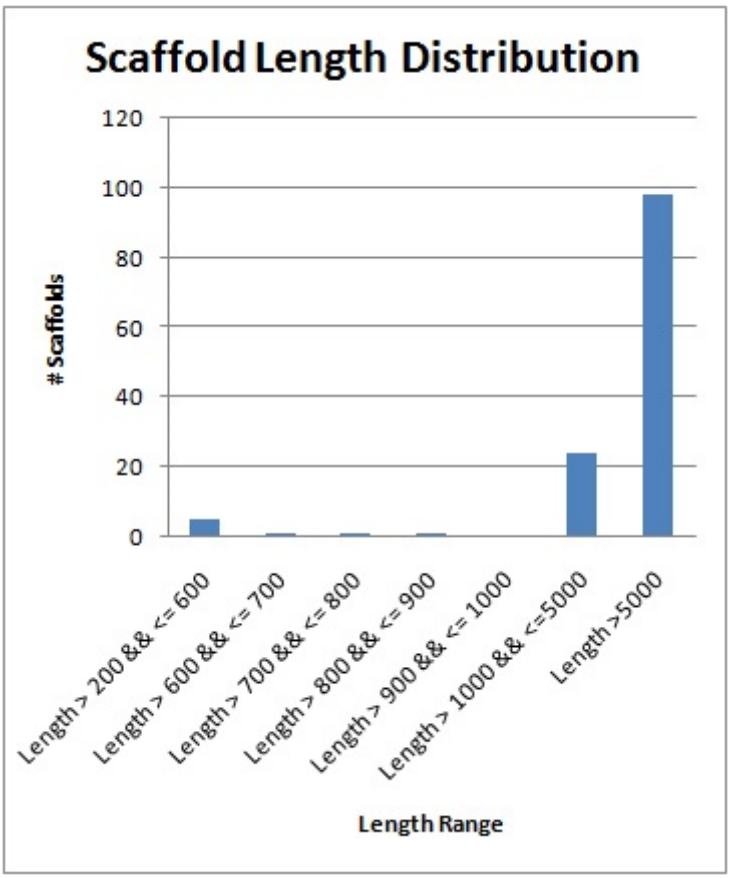

Figure 1

It represents scaffold length distribution across genome and maximum number of scaffolds has length above 5 Kb. The least represented scaffold length is in between less than 1kb to less than equal to 5kb.

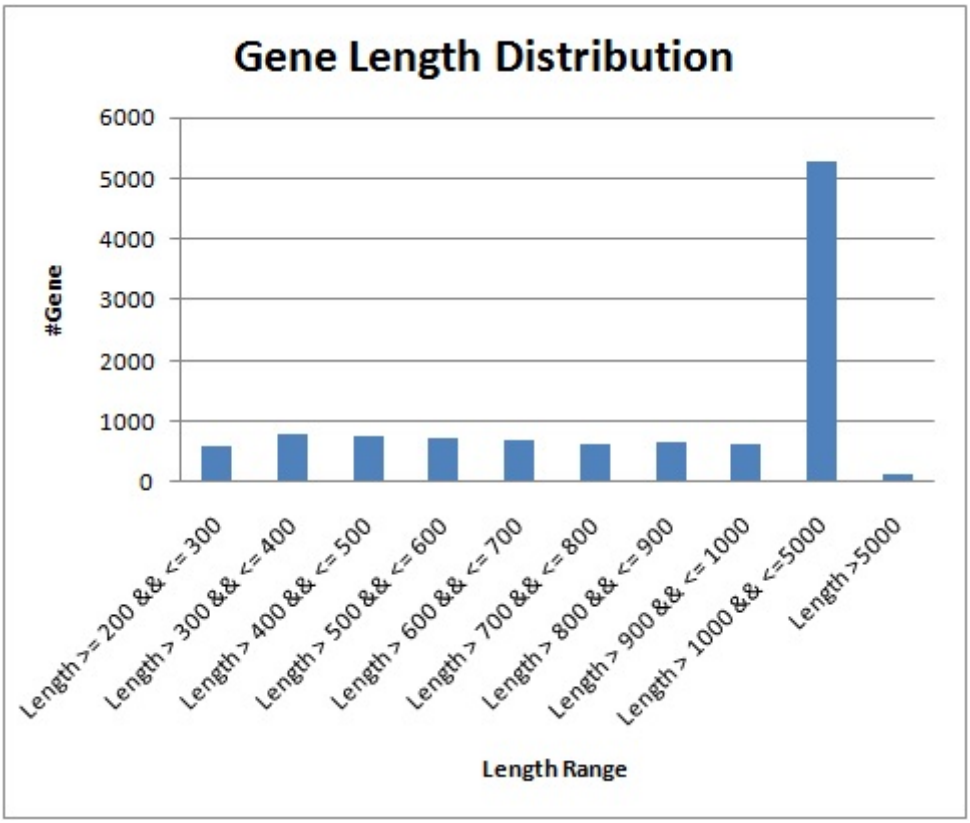

Figure 2

It signifies the correspondence of number of genes to the gene length and we found maximum number of genes have gene length in the range of 1Kb to 5 Kb and very less number of genes have gene length above 5kb.

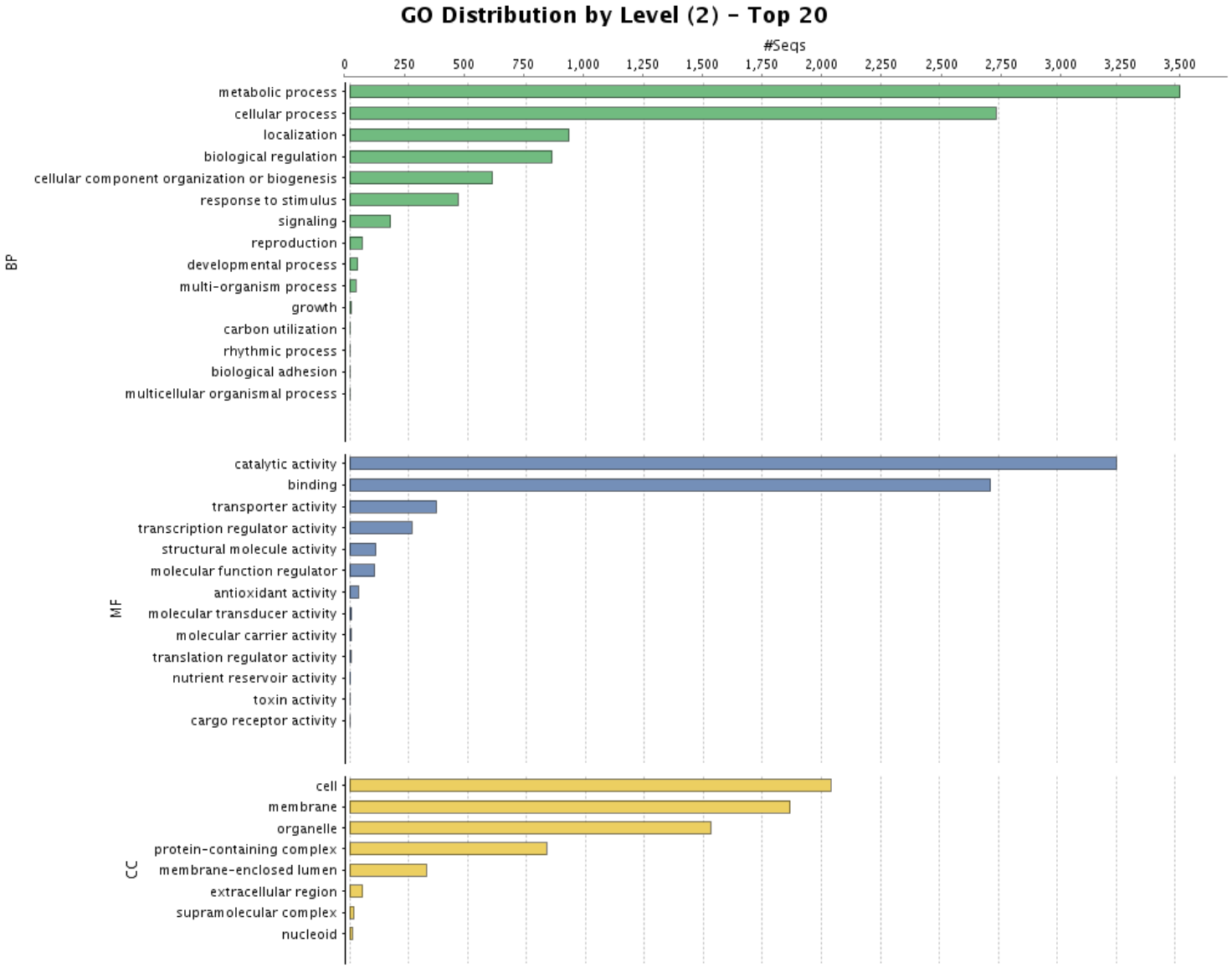

Figure 3

Gene annotation for all the three GO domains i.e. Biological Processes (BP), Molecular Function (MF) and Cellular Components (CC). The bars represent the number of genes involved in the biological, cellular and molecular functions of the Wilsonomyces carpophilus genome.

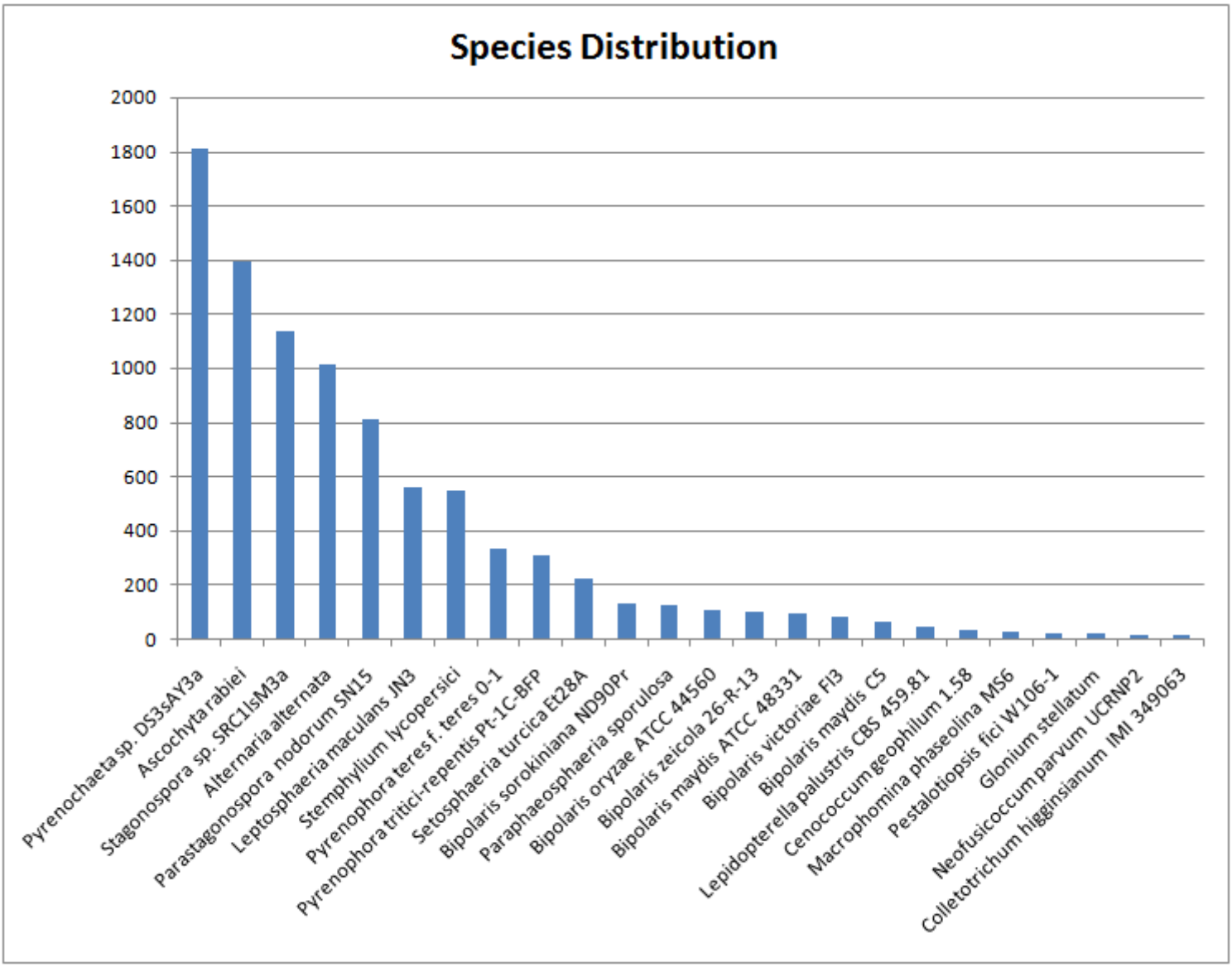

Figure 4

Relationship of other fungal species with Wilsonomyces carpophilus. The bars represent no. of hits attained in nrdata base against the Wilsonomyces carpophilus genome representing its similarity with other fungal plant pathogenic species with maximum similarity to Pyrenocheata species.

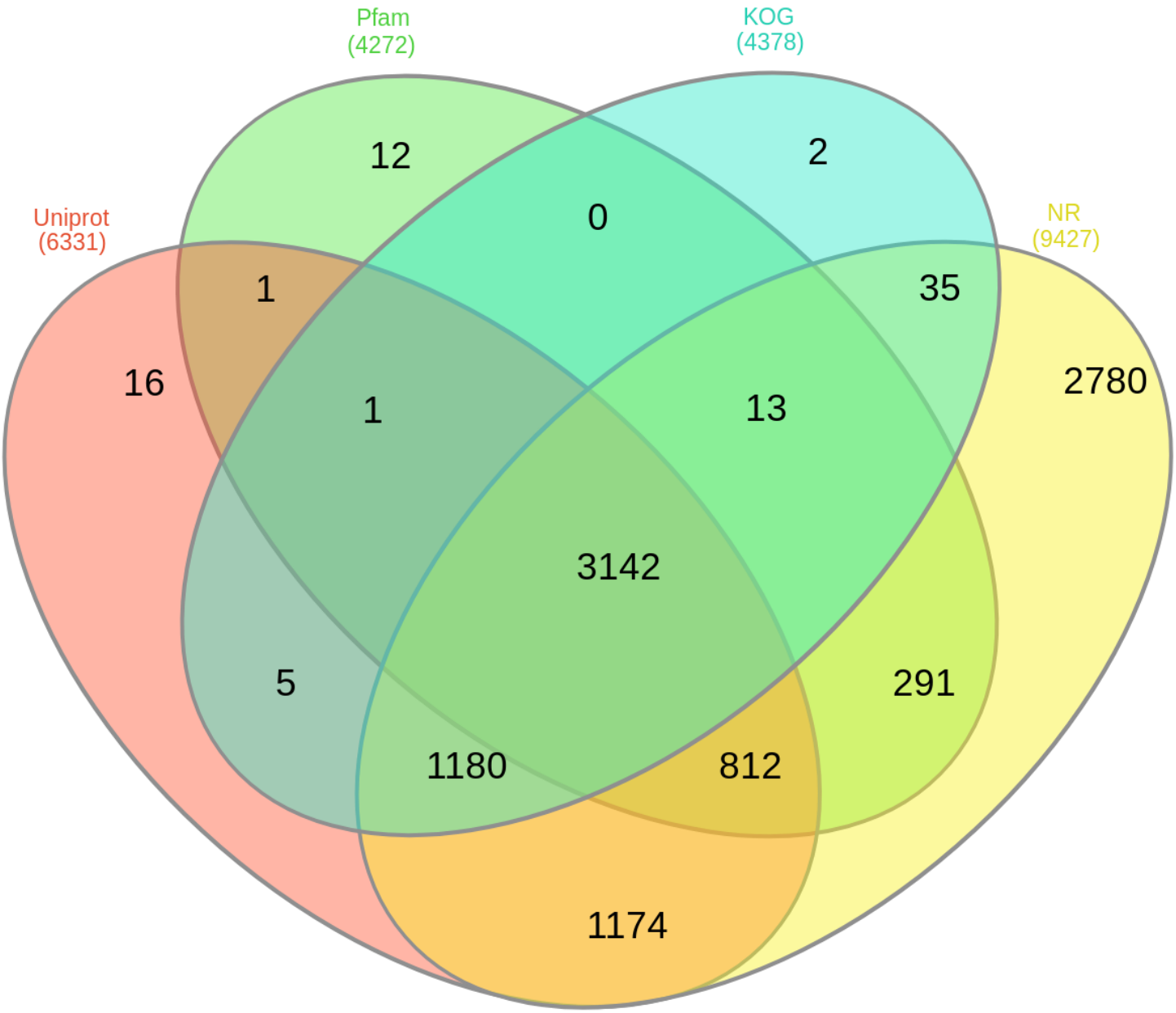

Figure 5

Venn diagram representing the number of genes that are common between and among the four data bases. The comparison between the data bases revealed that only 3142 genes are common to each data base out of 10901 genes.

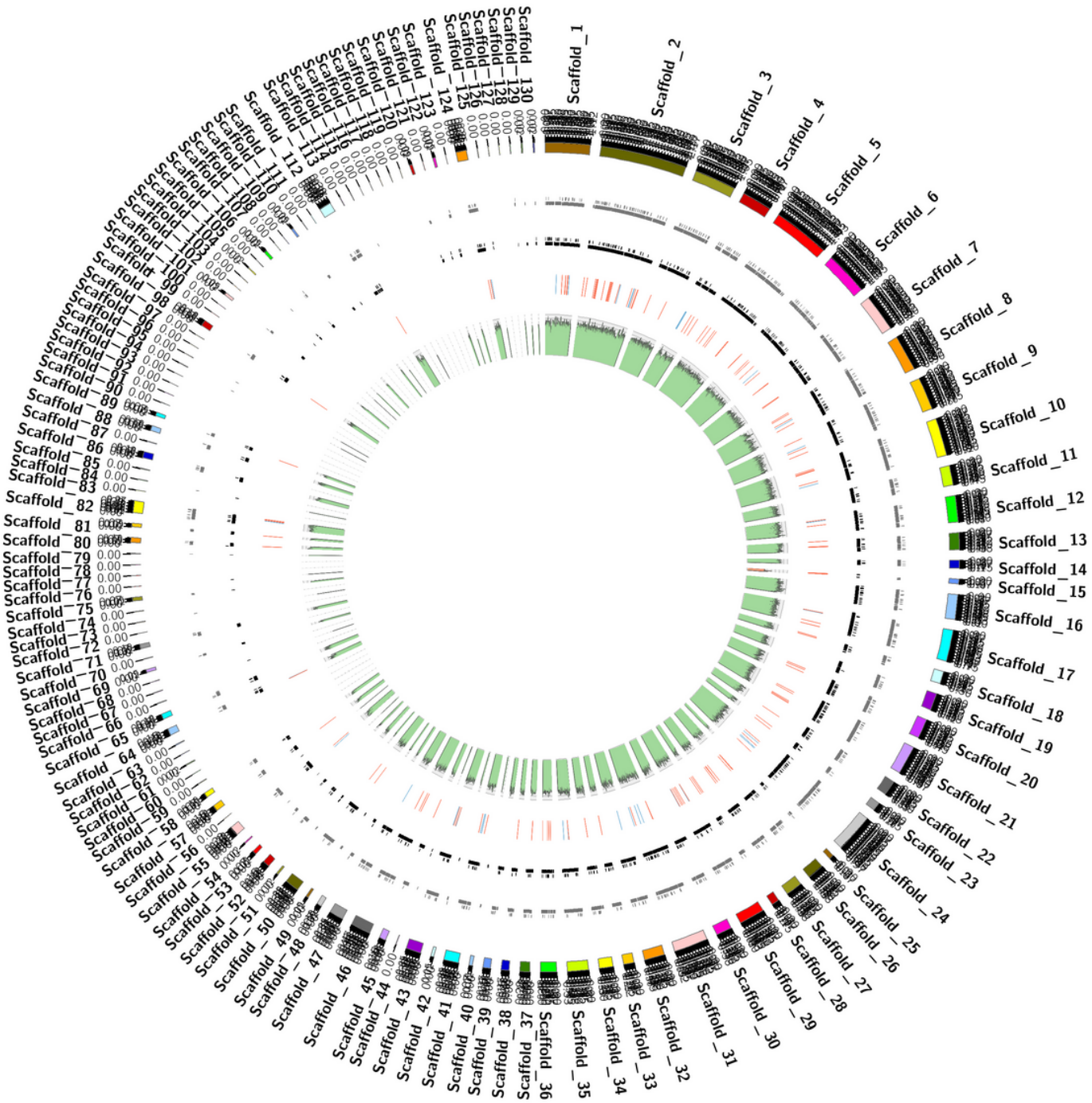

Figure 6

Schematic representation of genomic characteristics of Wilsonomyces carpophilus, Circle A and Circle B represents gene predicted in the positive frame. Circle C represents gene predicted in the negative frame, Circle D represents tRNA (red tiles) and rRNA (blue tiles). The inner E circle represents distribution of GC content .The red shade in the E circle represents regions of low GC contents (less than 35%).

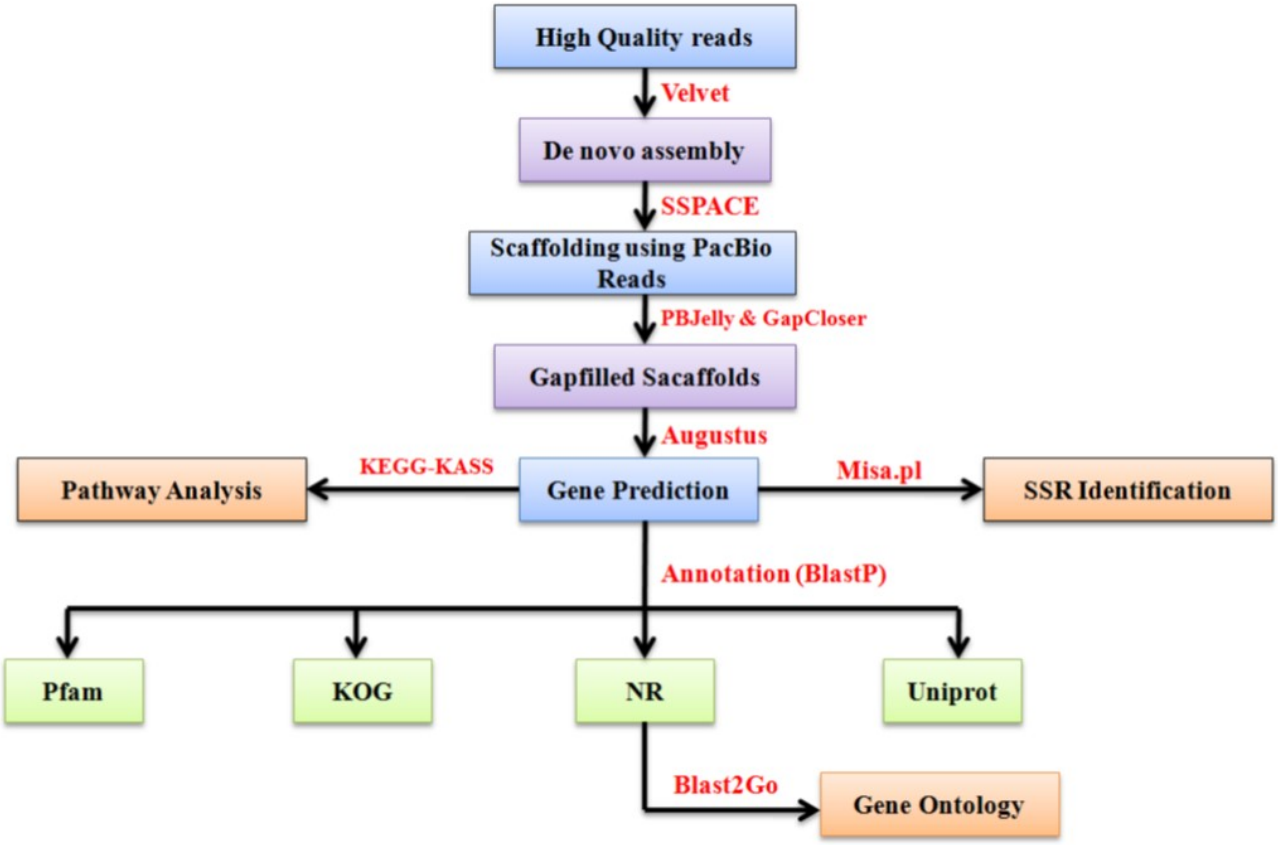

Figure 7

Schematic representation of steps employed in analyzing Wilsonomyces carpophilus genome using different software and databases.

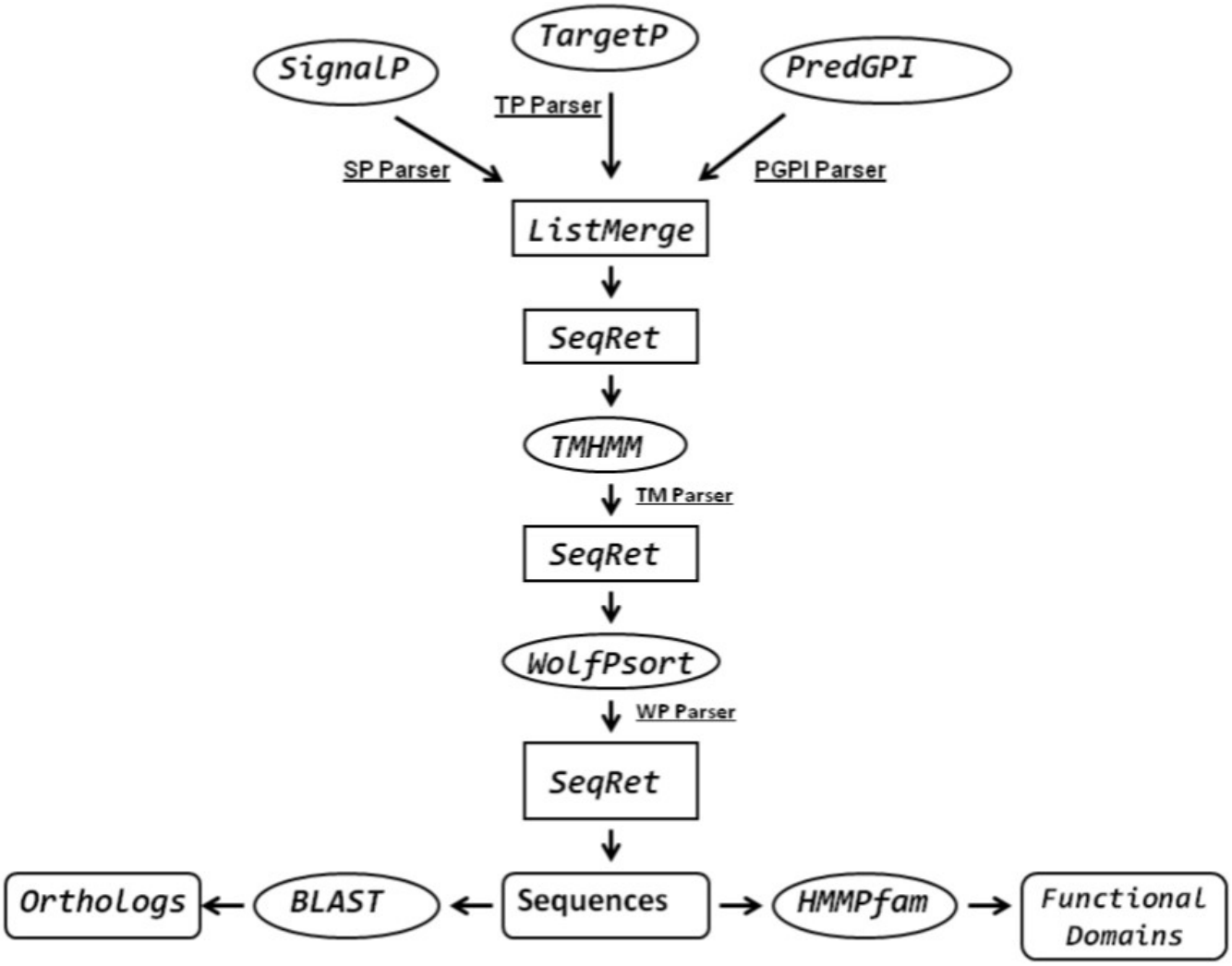

Figure 8

Bioinformatic pipeline used to find secreted proteins in the assembled genome of Wilsonomyces carpophilus

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables12andfig1.pdf