Unearthing the genes of plant-beneficial marine yeast - *Wickerhamomyces anomalus* strain MSD1

Radhesh Krishnan S[1], Prabhakaran N[1], Sengali Ragunath K[1], Srinivasan R[1], Keerthana Ponni K[1], Balaji G[1], Gracy M, Brindha C, Lakshmi Narayanan and Latha K[1*]
[1]R & D Division Extension, T. Stanes and Company Limited, Coimbatore-641018, TN.
*Correspondence: testinglab@t-stanes.com

Abstract

The *de novo* genome of unique marine yeast, *Wickerhamomyces anomalus* isolated from seaweed along Indian coast is presented. The genome assembly was carried out using MaSurCA assembler that generated a data size ~14.3 mb from short and long reads obtained from Illumina Hiseq 4000 and GridION-X5 respectively. This assembled genome data were used for predicting genes using Augustus gene prediction tool that reported 6720 genes and proteins. The gene sequences were used to unravel the metabolic pathway analysis using KAAS database. The protein sequences were used for secondary analysis to predict the presence of signal peptides using SignalP tool, predicting protein family, domains using Pfam tool and prediction of transmembrane helices in proteins using TMHMM tool. Presence of genes involved in plant growth–promotion and regulation (PGPR) including siderophore and IAA production, iron and sulfur transformation, zinc and phosphate solubilization, nitrogen fixation, synthesis of anti-bacterial and volatile organic compound (VOCs), were assigned. Additionally, acid and alkaline phosphatases, ACC deaminases and lytic enzymes such as β-glucanases, proteases and chitinases involved in pathogen suppression, are also reported. The study elucidates comprehensive understanding of PGP attributes of MSD1 and its potential use in agriculture as bio-fertilizer /bio-stimulant.

Key words

Industrial Microorganism, Comparative genomics, Next-generation sequencing, Genome annotation, Genes, *Wickerhamomyces anomalus*

Introduction

Wickerhamomyces anomalus, formerly known as *Pichia anomala, Hansenula anomala, Candida pelliculosa* was recently assigned to the *genus Wickerhamomyces* based on phylogenetic analysis of gene sequences, which has caused major changes in the classification of yeasts. This species has been frequently isolated from grapes and wines. *W. anomalus* is a biotechnologically relevant yeast species with food, environmental, industrial, and medical applications [1].

Wickerhamomyces anomalus has many different roles in agriculture and the food industry. *W. anomalus* is often among the “film-forming” yeasts associated with beer spoilage [2,3], as well as the spoilage of bakery products [4]. In contrast, *W. anomalus* is among the consortium of yeasts and other microorganisms that are necessary for the fermentation of cocoa and coffee beans, which includes degradation of pectin from the surrounding plant tissue (*Masoud and Jespersen 2006, Schwan and Wheals 2003*). *W. anomalus* has been tested extensively for biocontrol of mold growth that develops during postharvest storage of apples and airtight-storage of grain [5,6]. As summarized by [1], *W. anomalus* can grow under conditions of extreme environmental stress, including anaerobiosis, which makes it strongly competitive with spoilage molds under storage conditions.

The yeast has been reported for its glycosidase [7], volatile organic compound production [1] and antimicrobial [8,9] properties. This species has gained considerable importance for the wine industry since it enhances the flavor of wine [2, 5] and produces bioethanol [10]. The current report ascertains de novo genomic DNA of the isolate confirmed as (NCBI Accession number- MF174856, Safe deposit Accession number-NAIMCC-SD-0004) that was present as an epiphyte on the seaweed Sargassum, Mandapam Beach Park, Tamil Nadu, India. Additionally, a patent has been filed for this yeast and its use in agriculture as a Plant growth promoting yeast (Indian Patent application no. 202041036012). Annotation analysis of the whole genome sequencing (WGS) leads us in prediction and identification of key genes that are responsible for the PGPR activity of the strain MSD1.

Materials and methods

The marine yeast *W. anomalus* strain MSD1 was isolated from the marine macroalgae (Sargassum sp.,) collected from Mandapam Beach Park, Rameswaram, Tamil Nadu, India [11]. MSD1 was one among the potential seaweed associated microbes (our published research [12]) possessing plant growth promoting microbe like character (Data not disclosed here).

DNA Isolation, Genome Sequencing, and Assembly & Variation Identification and Genome Diversity Analysis

The yeast was cultured in Zobel Marine Broth (HiMedia, Mumbai, India) for 48 h at 30°C with constant shaking (150 rpm). The high quality DNA from the sample was sequenced at Genotypic Technology Pvt Ltd. India, using Hiseq 4000 (Illumina) and GridION-X5 (Oxford Nanopore Sequencing Technology). The short reads (Illumina) and long reads (Nanopore) data were demultiplexed using bcl2fastq and guppy [13] respectively. Hybrid assembly was performed using Illumina and nanopore reads by MaSurCA Hybrid Assembler [14] with standard parameters. The gene and protein sequence prediction from the assembled genome was performed using Augustus tool [15]. The secondary analysis of the protein was carried out using different protein analytical tools (signalp, tmhmm, PfamScan) [16–18]. The metabolic pathways were predicted using KAAS database.

Phylogenetic aanalysis

Here we used genome sequence data from 17 publicly available yeast genomes representing 17 known major lineages and 2 non-yeast fungal out groups to generate phylogenetic tree. *Wickerhamomyces anomalus* NRRL Y-366-8 (LWUN00000000.1), *Saccharomyces cerevisiae* (JRIV00000000.1), *Babjeviella inositovora* NRRL Y-12698 (LWKO00000000.1), *Suhomyces tanzawaensis* NRRL Y-17324 (LYME00000000.1), *Metschnikowia bicuspidata var. bicuspidata* NRRL

YB-4993 (LXTC00000000.1), *Hyphopichia burtonii* NRRL Y-1933 (LYBQ00000000.1), *Ascoidea rubescens* DSM 1968 (LYBR00000000.1), *Candida arabinofermentans* NRRL YB-2248 (LWUO00000000.1), *Tortispora caseinolytica* NRRL Y-17796 (LSKT00000000.1), *Cyberlindnera jadinii* NRRL Y-1542 (LTAD00000000.1), *Hanseniaspora valbyensis* NRRL Y-1626 (LXPE00000000.1), *Ogataea polymorpha* (AECK00000000.1), *Lipomyces starkeyi* NRRL Y-11557 (LSGR00000000.1), *Nadsonia fulvescens var. elongata* DSM 6958 (LXPB00000000.1), *Pachysolen tannophilus* NRRL Y-2460 (LZCH00000000.1), *Pichia membranifaciens* NRRL Y-2026 (AEHA00000000.1), *Saitoella complicata* NRRL Y-17804 (AEUO00000000.1), *Trichoderma reesei* (AAIL00000000), *Trichoderma harzianum* (JOKZ00000000.1).

Comparative Genomics

Assembled sequence was compared with the reference sequence to know the gene re-arrangements and genome coverage. We used BRIG to have circular genome representation and Mauve to visualize the synteny between reference genome and assembled genome.

Nucleotide Sequence Accession Number

This BioProject has been deposited in GenBank under accession number PRJNA556347. The sequences obtained in this project have been deposited in the NCBI Sequence Read Archive under the accession numbers SRR10092046, and SRR9822044. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA556347.

Results

General genome characteristics

A total of 6.71 million paired-end reads were generated for the marine yeast from Illumina and 0.37 million reads from Nanopore-GridION respectively. Read statistics are given in **Table 1** and **Table 2**.

The size of assembled genome generated was ~14.3 mb having 289 contigs and the longest contig was of ~0.2 mb length. Assembly was validated using blast alignment against nr database (**Table 3**).

The assembled genome sequence was used for predicting genes and protein sequences using Augustus gene prediction tool. A total of 6720 genes and proteins were predicted in the analysis. The GO annotation of the predicted genes was completed using Uniprot database and in-house scripts. Out of 6720 genes predicted, 6658 genes were annotated and 64 genes remained unannotated (**Figure 1**).

Comparative Genomics

Reference based whole genome sequencing of *Wickerhamomyces anomalus* was carried out using reference genome available at NCBI for of Wickerhamomyces anomalous, strain- NRRL Y-366-8. More than 140 X of sequencing coverage was achieved for the genome of approximate size 14MB. More than 99% of the reference genome was covered at 1X and >93% of the reference genome was covered at 20 X by good quality data which confirms the choice of reference and

the sufficiency of the data for reference based WGS. The consensus sequence resulted from the analysis was compared with the reference sequence in order to know genome significant rearrangements if any (**Figure 2**).
Assembled sequence was compared with the reference sequence to know the gene re-arrangements and genome coverage. We used BRIG to have circular genome representation and Mauve to visualize the synteny between reference genome and assembled genome (**Figure 3**).

Pathway analysis the yeast genome was carried for using KAAS database that provided functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database [19]. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways (**Supplementary file 1**). The secondary analysis of protein sequences obtained from Augustus was carried out using different tools- SignalP, Pfam-Scan, TMHMM. SignalP tool predicts the presence of signal peptides and the location of their cleavage sites in proteins. A total of 521 signal peptides were predicted out of which 304 had trans-membrane segments and 217 without trans-membrane segments. Pfam-Scan tool was used to predict protein family and domains present in the predicted protein sequences. A total of 8280 pfam annotation (which includes family, domain, repeat and motif) were predicted for 6720 proteins. Out of 8280 pfam annotation 5339 contained clan (group of related protein families) information and 2941 had no clan information. TMHMM tool predicts transmembrane helices in given proteins sequences. A total of 6720 proteins used for transmembrane helix prediction, out of which 1377 proteins contained transmembrane helices and remaining 5343 proteins were without transmembrane helices (**Supplementary file 2-4**).

Discussion

Genes identified from WGS of MSD1 related to PGPR traits

We identified genes in the MSD1 genome attributable to the production of IAA, solubilization of minerals like phosphate and zinc, synthesis of siderophores, acetoin and 2,3-butanediol, suppression of pathogenic fungi, resistance to oxidative stress, and ability to break down toxic compounds and other abiotic stresses.
Here, two proposed **IAA biosynthesis** pathways, amidase and aldehyde dehydrogenase pathways are identified in the genome of MSD1. In the indole-3-acetonitrile (IAN) pathway IAN can first be converted to indole-3-acetamide (IAM) by nitrile hydratase and then IAM is converted to IAA by amidase. In the IPyA pathway indole-3-pyruvate (IPyA) is converted to indole-3-acetaldehyde (IAAld) by indolepyruvate decarboxylase and then to IAA by aldehyde dehydrogenase. All of these genes responsible for IAA synthesis were present in MSD1 genome [20].

Gluconic acid (GA) is recognized as one of the major organic acids in most bacteria responsible for the solubilization of mineral phosphates. The synthesis of GA is catalyzed by glucose dehydrogenase (GDH) and its co-factor pyrrolo-quinolone quinine (PQQ)[21–23]. Accordingly, the MSD1 genome was searched for the presence for phosphate transporter genes. Gene IDs PHO84 and PHO87 that encodes for inorganic phosphate transport were predicted. In addition, 5 genes encoding mitochondrial thiamine pyrophosphate transporters (solute carrier proteins) were also predicted in the MSD1 genome.

MSD1 carrying the gene encoding for the synthesis of **siderophore** was identified. Genes encoding isochorismate domain containing protein, Gene IDs K08197 (7 copies) and K23503 (2 copies) that are annotated for siderophore-iron: H+ symporter and sideroflexin respectively were predicted. These indicate that although strain MSD1 cannot synthesis numerous sideropheres, it can heterologously obtain siderophores produced by other soil bacteria [20,24].

The MSD1 genome was predicted for the presence of a cascade of genes for Fe uptake/transport. Genes like K07243 (high-affinity iron transporter), K19791 (iron transport multicopper oxidase), K12346 (metal iron transporter), K22736 (vacuolar iron transporter family protein), K15113 (solute carrier family 25), K02304 (precorrin-2 dehydrogenase/sirohydrochlorin ferrochelatase), and K01772 (protoporphyrin/coproporphyrin ferrochelatase) were annotated from MSD1 genome.

The genes Associated with Plant Growth Promotion Traits

Previous phenotypical and PGP abilities, observed in pure culture and in plant experiments under salt stress, was supported by the MSD1 genome content (**Table 4**). The MSD1-detailed genomic profile of their confirmed PGP abilities and other possible mechanisms involved in plant promotion were analyzed and described here. It has been reported that PGPR may produce compounds such as **phenazine** and 4-hydroxybenzoate which function as antibiotics and suppress plant pathogenic microbes [24]. UbiD, involved in 4-**hydroxybenzoate** synthesis, and PhzC-PhzF, involved in phenazine synthesis, were identified in the MSD1 genome. Moreover, a homologue of the gene coding for **chitinase** enzyme was identified that can potentially dissolve the cell wall of pathogenic fungal and insect pests [24]. In addition to these, the genes gabD and gabT which are responsible for the production of pest/disease inhibiting **γ -aminobutyric acid (GABA)** in the genome was identified [20]. This suggests that the synthesis of the three antimicrobial compounds is a widespread pathway in MSD1.

In addition to the above PGP traits, two growth-promoting **volatile organic compounds** (VOCs), acetoin and butanediol, were reported to promote plant growth by stimulating root formation and increasing systemic disease resistance and drought tolerance in some other very efficient PGPR[5,25–29]. Genes encoding enzymes including **acetolactate synthase** and **acetoin dehydrogenase** (Table 1) which are involved in acetoin and butanediol synthesis, were detected in the genome of MSD1 [24,28,30]. Two pyruvate molecules condensed into acetolactate is catalyzed by acetolactate synthase, and which is converted to acetoin by acetolactate decarboxylase and finally acetoin is converted to **2,3-butanediol** catalyzed by acetoin reductase [24,28,30].

Nitrogen fixation

Nitrogenase is the enzyme central to nitrogen fixation and it consists of Fe-protein encoded by *nifH* and MoFe-protein encoded by *nifDK*. Full assembly of the nitrogenase complex needs the products of at least twelve nif genes, especially for the processing of catalytic stability and nitrogenase metalloclusters (nifMZ, nifUS, and *nifW*) and for synthesis of a particular molybdenum cofactor (MoCFC). Many microbial gene families are responsible for organic N decomposition, metabolism, and biosynthesis in soil. Here, five gene families directly related with N cycling processes were extracted and analyzed, including ***nao*** (nitroalkane oxidase), ***nmo*** (nitronate monooxygenase), gdh (glutamate dehydrogenase), *ureC* (urease) and **GS** (glutamine synthetase)[21].

MSD1 is able to grow on nitrogen-free medium (data not shown) and this indicates that the strain is able to fix atmospheric nitrogen. The MSD1 genome contains nif genes together with the *NifU* and nitronate genes which are the positive/negative regulatory proteins for *nif* genes [21].

ACC Deaminase

One of the mechanisms of PGPR to alleviate salt stress is the synthesis of the enzyme 1-aminocyclo-propane-1-carboxylate (ACC) deaminase or its homologue D-cysteine desulfhydrase encoded by acdS or dcyD, respectively. Both enzymes lower ethylene accumulation in stressed plants by cleaving ACC, an immediate precursor of ethylene in plants, to form ammonia and α-ketobutyrate. This reaction is pyridoxal phosphate dependent, and both ACC deaminase and D-cysteine desulfhydrase belong to the pyridoxal phosphate-dependent enzyme family PALP. In the MSD1 genome, neither acdS genes nor dcyD genes are present but eight CDSs containing genes encoding genes belonging to the PALP domain (Table 1) was found [24,31,32]. Of these genes, presence of Cys_K 1 and 2 for Cysteine desulfurase (KO ID: K04487, 5 copies of iscS), Tryptophan synthase beta chain (2 copies) and L-threonine ammonia-lyase (6 copies) both show lyase activity and potentially perform ammonia synthesis similarly to the enzymes encoded by acdS and dcyD[30].

Genes of Central Metabolism and Cellular Processes

Carbohydrate degradation pathways **Emden-Meyerhof pathway** and **Entner-Doudoroff pathway** for glucose, arabinose, mannose, trelose, mannitol, and the respective transport systems have been detected. All genes of the TCA cycle were present (**Table 1**). Exo- and polysaccharide biosynthesis and the respective transporter have been discovered (**Table 4**). A comprehensive list of detected genes is given in the **supplementary files**.

Genes putatively involved in salt tolerance

MSD1 can grow well in 0–12% NaCl (data not shown). Analysis of the genome reveals that strain MSD1 has a number of genes related to salt tolerance. For example, trehalose can act as an osmoprotectant under environmental stresses such as high salt or drought, low temperature or osmotic stress in many organisms. Trehalose accumulates in transgenic rice and enhances plant abiotic stress tolerance. So far five trehalose biosynthetic pathways have been found in bacteria including treS, otsA/otsB, treP, treT and treY/treZ40[30,33]. Here, two trehalose biosynthesis pathway **otsA/B**, were identified in the MSD1 genome. In the otsA/otsB pathway both glucose-6-phosphate and UDP-glucose can synthesize trehalose-6-phosphate catalyzed by trehalose-6-phosphate synthase (otsA) activity. **Trehalose-6-phosphate** is then formed from trehalose catalyzed by trehalose-6-phosphate phosphatase (**otsB**) activity. Eventually, trehalose may be hydrolyzed by trehalase (2 copies) with the generation of two glucose molecules (**Table 4**). This pathway has been recognized as a universal pathway present in microorganisms and contributes to the survival under harsh environmental conditions [30,33].

Moreover, a number of **osmoregulation receptors** and transport systems were determined in the MSD1 genome. These genes can encode up to 24 two component systems (TCSs), among which 21 TCSs can be functionally assigned based on the KEGG database (**Table 4**) [30,33].

Of those 21 assigned TCSs, 3 belong to the SSK1 (response regulator) family, two to the YPD1 (phosphorelay intermediate protein) family, 16 belong to the SLN1 (sensor histidine kinase) family and one to the SKN7 (response regulator) family. The eight remaining TCS genes are annotated as sensor histidine kinase (**Table 4**).

In addition, genes encoding transport systems such as K+ transport systems for K+ accumulation and H+/Na+ antiporters (nha) for importing H+ and pumping out Na+ have also been found to resist hyperosmotic (Aft1 domain) stress in the genome of MSD1 (**Table 4**).

The MSD1 genome carries heat shock genes dnaJ, dnaK, groES, groEL, htpG, and grpE (**Table 4**). Moreover, the clpB gene, a **heat shock protein**, specified to be upregulated during salt stress in marine bacteria is also contained. The MSD1 genome also carries CDSs encoding for **peroxidases**, **superoxidase**, and glutathione S-transferase (**Table 1**). These genes play a role in the protection of cell **oxidative stress** caused by salt stress[30,33,34].

The genome sequence of marine isolate *W. anomalus* strain MSD1 presented in this paper is a plant growth promoting yeast isolated from the seaweed [12]. This study showed MSD1 has potential traits such as Zinc and phosphate-solubilizing, iron and sulfate transformation capability, production of ACC deaminase, siderophore, and VOCs; making it as an effective PGP yeast. Considering a variety of complex conditions that occur in rhizospheres [35], the environmental adaptability of PGPR in in situ rhizosphere became an important factor for improved plant growth-promoting capacity. In addition, initial studies focusing on the functional properties of PGPR have led to interest in the comparative analyses of pan-/core-genomes of these bacteria, which are of ecological importance for elucidating the fundamental genotypic features of PGPY [36,37].

Conclusions

The genetic information obtained for *W. anomalus* strain MSD1 will enable us to interpret the expressed traits of the yeast and further provide insights into the practical applications of the strain as a bio-stimulant/PGPR for agriculture use or agri-input.

Data availability

This whole genome sequence of the biosample SAMN12347843 has been deposited at GenBank/NCBI under the accession number SRR10092046 and BioProject number PRJNA556347. The associated Illumina HiSeq 4000 subreads are available under the SRA accession number SRR9822044. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA556347.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The work is an outcome of the collaborative BIRAC project between T. Stanes & Company Limited and DBT, India.

References

1. Passoth V, Fredlund E, Druvefors UÄ, et al. Biotechnology, physiology and genetics of the yeast Pichia anomala. FEMS Yeast Res. 2006; 6:3–13
2. Limtong S, Into P, Attarat P. Biocontrol of rice seedling rot disease caused by curvularia lunata and Helminthosporium oryzae by epiphytic yeasts from plant leaves. Microorganisms 2020; 8:
3. Tay ST, Lim SL, Tan HW. Growth inhibition of Candida species by Wickerhamomyces anomalus mycocin and a lactone compound of Aureobasidium pullulans. BMC Complement. Altern. Med. 2014; 14:1–11
4. Coda R, Cassone A, Rizzello CG, et al. Antifungal activity of Wickerhamomyces anomalus and Lactobacillus plantarum during sourdough fermentation: Identification of novel compounds and long-term effect during storage of wheat bread. Appl. Environ. Microbiol. 2011; 77:3484–3492
5. Satora P, Tarko T, Sroka P, et al. The influence of Wickerhamomyces anomalus killer yeast on the fermentation and chemical composition of apple wines. FEMS Yeast Res. 2014; 14:729–740
6. Wang X, Glawe DA, Kramer E, et al. Biological Control of Botrytis cinerea : Interactions with Native Vineyard Yeasts from Washington State. Phytopathology

2018; 108:691–701

7. Padilla B, Gil J, Manzanares P. Challenges of the Non-Conventional Yeast Wickerhamomyces anomalus in Winemaking. Fermentation 2018; 4:68

8. De Ingeniis J, Raffaelli N, Ciani M, et al. Pichia anomala DBVPG 3003 Secretes a Ubiquitin-Like Protein That Has Antimicrobial Activity. Appl. Environ. Microbiol. 2009; 75:1129–1134

9. Schneider J, Rupp O, Trost E, et al. Genome sequence of Wickerhamomyces anomalus DSM 6766 reveals genetic basis of biotechnologically important antimicrobial activities. FEMS Yeast Res. 2012; 12:382–386

10. Swangkeaw J, Vichitphan S, Butzke CE, et al. The characterisation of a novelPichia anomala β-glucosidase with potentially aroma-enhancing capabilities in wine. Ann. Microbiol. 2009; 59:335–343

11. Singh RP, Reddy CRK. Seaweed-microbial interactions: Key functions of seaweed-associated bacteria. FEMS Microbiol. Ecol. 2014; 88:213–230

12. Srinivasan R, Ragunath K, Karuppiah V, et al. Isolation and screening of seaweed associated microbes for development of marine based Agri -inputs. Seaweed Res. Util. 2017; 39:8

13. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019; 20:129

14. Zimin A V., Marçais G, Puiu D, et al. The MaSuRCA genome assembler. Bioinformatics 2013; 29:2669–2677

15. Stanke M, Keller O, Gunduz I, et al. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006; 34:W435–W439

16. Krogh A, Larsson B, von Heijne G, et al. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. J. Mol. Biol. 2001; 305:567–580

17. Petersen TN, Brunak S, von Heijne G, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 2011; 8:785–786

18. Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. Nucleic Acids Res. 2010; 38:D211–D222

19. Moriya Y, Itoh M, Okuda S, et al. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007; 35:W182–W185

20. Luziatelli F, Ficca AG, Cardarelli M, et al. Genome sequencing of pantoea agglomerans C1 provides insights into molecular and genetic mechanisms of plant growth-promotion and tolerance to heavy metals. Microorganisms 2020; 8:1–19

21. Tsurumaru H, Okubo T, Okazaki K, et al. Metagenomic analysis of the bacterial community associated with the taproot of sugar beet. Microbes Environ. 2015; 30:63–69

22. Shrivastava M, Rajpurohit YS, Misra HS, et al. Survival of phosphate-solubilizing bacteria against DNA damaging agents. Can. J. Microbiol. 2010; 56:822–830

23. Choi O, Kim J, Kim JG, et al. Pyrroloquinoline quinone is a plant growth promotion factor produced by Pseudomonas fluorescens B16. Plant Physiol. 2008; 146:657–668

24. Gupta A, Gopal M, Thomas G V., et al. Whole genome sequencing and analysis of plant growth promoting bacteria isolated from the rhizosphere of plantation crops coconut, cocoa and arecanut. PLoS One 2014; 9:

25. Schneider J, Rupp O, Trost E, et al. Genome sequence of Wickerhamomyces anomalus DSM 6766 reveals genetic basis of biotechnologically important antimicrobial activities. FEMS Yeast Res. 2012; 12:382–386

26. Bailly A, Weisskopf L. Mining the volatilomes of plant-associated microbiota for new biocontrol solutions. Front. Microbiol. 2017; 8:1–12

27. Gong AD, Li HP, Shen L, et al. The Shewanella algae strain YM8 produces volatiles with strong inhibition activity against Aspergillus pathogens and aflatoxins. Front. Microbiol. 2015; 6:1–12

28. Fincheira P, Quiroz A. Microbial volatiles as plant growth inducers. Microbiol. Res. 2018; 208:63–75

29. Parafati L, Vitale A, Restuccia C, et al. Performance evaluation of volatile organic compounds by antagonistic yeasts immobilized on hydrogel spheres against gray, green and blue postharvest decays. Food Microbiol. 2017; 63:191–198

30. Liu W, Wang Q, Hou J, et al. Whole genome analysis of halotolerant and alkalotolerant plant growth-promoting rhizobacterium Klebsiella sp. D5A. Sci. Rep. 2016; 6:20–22

31. Khan MS, Zaidi A, Musarrat J. Phosphate solubilizing microorganisms: Principles and application of microphos technology. Phosphate Solubilizing Microorg. Princ. Appl. Microphos Technol. 2014;

32. Ma W, Sebestianova SB, Sebestian J, et al. Prevalence of 1-aminocyclopropane-1-carboxylate deaminase in Rhizobium spp. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 2003; 83:285–291

33. Suarez C, Ratering S, Hain T, et al. Complete genome sequence of the plant growth-promoting bacterium hartmannibacter diazotrophicus strain E19T. Int. J. Genomics 2019; 2019:

34. Zarei A, Brikis CJ, Bajwa VS, et al. Plant glyoxylate/succinic semialdehyde reductases: Comparative biochemical properties, function during chilling stress, and subcellular localization. Front. Plant Sci. 2017; 8:1–13

35. Berg G, Smalla K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. FEMS Microbiol. Ecol. 2009; 68:1–13

36. Khunnamwong P, Lertwattanasakul N, Jindamorakot S, et al. Evaluation of antagonistic activity and mechanisms of endophytic yeasts against pathogenic fungi causing economic crop diseases. Folia Microbiol. (Praha). 2020; 65:573–590

37. Kumla J, Nundaeng S, Suwannarach N, et al. Evaluation of Multifarious Plant Growth Promoting Trials of Yeast Isolated from the Soil of Assam Tea (Camellia sinensis var. assamica) Plantations in Northern Thailand. Microorganisms 2020; 8:1168

Tables

Table1. Illumina Read Statistics

Sample	Read Count
Total raw reads (in million)	6713813 (Read1, Read2)
Total processed reads (in million)	6417015(Read1, Read2)

Table2. Nanopore Read Statistics

Parameters	**Read Stat**
Reads Generated	374070
Maximum Read Length	61005
Minimum Read Length	94
Average Read Length	1177.6
Median Read Length	2311
Total Reads Length	440506966
Total Number of Non-ATGC Characters	0
Percentage of Non-ATGC Characters	0
Reads >= 100 bp	374067
Reads >= 200 bp	372318
Reads >= 500 bp	241259
Reads >= 1 Kbp	125627
Reads >= 10 Kbp	1861
N50 value	1847

Table 3. Assembly statistics

Parameter	**Statistical data**
Contigs Generated	289
Maximum Contig Length	275125
Minimum Contig Length	3374
Average Contig Length	49645.9
Median Contig Length	51021
Total Contigs Length	14347675
Total Number of Non-ATGC Characters	0
Percentage of Non-ATGC Characters	0
Contigs >= 1 Kbp	289
Contigs >= 10 Kbp	259
N50 value	83236

Table 4. Genes related to Plant Growth Promoting traits that are annotated to be present in *Wickerhamomyces anomalus* strain MSD1

Attributes	**Gene or protein ID**	**Gene Annotation**
Indole-3-Acetic acid	K01426	Amidase
	K00128	Aldehyde dehydrogenase
	K01568	pyruvate decarboxylase

GABA production		
gabD	g6155.t1	succinate-semialdehyde dehydrogenase (NAD+) activity [GO:0004777]; succinate-semialdehyde dehydrogenase [NAD(P)+] activity [GO:0009013]; cellular response to oxidative stress [GO:0034599]; gamma-aminobutyric acid catabolic process [GO:0009450]; glutamate decarboxylation to succinate [GO:0006540]
gabT	g4127.t1	cytosol [GO:0005829]; 4-aminobutyrate transaminase activity [GO:0003867]; pyridoxal phosphate binding [GO:0030170]; gamma-aminobutyric acid catabolic process [GO:0009450]
	g5717.t1	cytosol [GO:0005829]; 4-aminobutyrate transaminase activity [GO:0003867]; pyridoxal phosphate binding [GO:0030170]; gamma-aminobutyric acid catabolic process [GO:0009450]
Antimicrobial		
Phenazine	g6439.t1	Phenazine biosynthesis-like protein
Acetoin and 2,3 butanediol synthesis	K01653	Acetolactete synthase 1
	K01652	Acetolactete synthase
	K00004	BDH; (R,R)-butanediol dehydrogenase / meso-butanediol dehydrogenase / diacetyl reductase [EC:1.1.1.4 1.1.1.- 1.1.1.303]
Phosphate	K08176	PHO84; MFS transporter, PHS family, inorganic phosphate transporter
	K14430	PHO87_91; phosphate transporter
	K15108	SLC25A19; solute carrier family 25 (mitochondrial thiamine pyrophosphate transporter), member 19
	K14684	SLC25A23S; solute carrier family 25 (mitochondrial phosphate transporter), member 23/24/25/41
	K15102	SLC25A3; solute carrier family 25 (mitochondrial phosphate transporter), member 3
Antimicrobial compound		
Chitinase production	K01183	Putative chitinase II
GABA	K00135	succinate-semialdehyde dehydrogenase
Other PGPR fitness conferring genes		
	K01480	speB; agmatinase
	K00797	speE; spermidine synthase
	K00802	SMS; spermine synthase [EC:2.5.1.22]
Resistance to antifungal drugs	K03327	TC.MATE; multidrug resistance protein, MATE family
	K08157	TPO1; MFS transporter, DHA1 family, multidrug resistance protein
	K08165	ATR1; MFS transporter, DHA2 family, multidrug resistance protein
	K08158	MDR1; MFS transporter, DHA1 family, multidrug resistance protein
Siderophore production (Iron homeostasis)		
	g2983.t1	catalytic activity [GO:0003824]
	g4122.t1	nicotinamidase activity [GO:0008936]
	K08197	ARN; MFS transporter, SIT family, siderophore-iron:H+

		symporter
	K23503	SFXN5; sideroflexin-5
Fe transport	K07243	FTR; high-affinity iron transporter
	K19791	FET3_5; iron transport multicopper oxidase
	K12346	SMF; metal iron transporter
	K22736	VIT; vacuolar iron transporter family protein
	K07243	FTR; high-affinity iron transporter
	K15113	SLC25A28_37; solute carrier family 25 (mitochondrial iron transporter), member 28/37
	K02304	MET8; precorrin-2 dehydrogenase / sirohydrochlorin ferrochelatase [EC:1.3.1.76 4.99.1.4]
	K01772	hemH; protoporphyrin/coproporphyrin ferrochelatase [EC:4.99.1.1 4.99.1.9]
Sulfur metabolism	g1937.t1	Sulfate adenylyltransferase (EC 2.7.7.4) (ATP-sulfurylase) (Sulfate adenylate transferase) (SAT)
	g5561.t1	Adenylyl-sulfate kinase (EC 2.7.1.25)
	g6660.t1	Glutathione synthetase (GSH-S) (EC 6.3.2.3)
Potassium	g381.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g2155.t1	K+ potassium transporter
	g4374.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g139.t1	potassium ion transport [GO:0006813]; protein insertion into mitochondrial inner membrane from matrix [GO:0032979]; proton transmembrane transport [GO:1902600]
	g1052.t1	cellular potassium ion homeostasis [GO:0030007]; positive regulation of mitochondrial translation [GO:0070131]; potassium ion transport [GO:0006813]; protein insertion into mitochondrial inner membrane from matrix [GO:0032979]; proton transmembrane transport [GO:1902600]
	g1718.t1	potassium ion transport [GO:0006813]; protein insertion into mitochondrial inner membrane from matrix [GO:0032979]; proton transmembrane transport [GO:1902600]
Zinc transport	K14709	SLC39A1_2_3; solute carrier family 39 (zinc transporter), member 1/2/3
	K14688	SLC30A1; solute carrier family 30 (zinc transporter), member 1
	K14713	SLC39A7; solute carrier family 39 (zinc transporter), member 7
	K07238	TC.ZIP; zinc transporter, ZIP family
	K14692	SLC30A5_7; solute carrier family 30 (zinc transporter), member 5/7
Resistance to oxidative stress		
Peroxidase	K00432	gpx; glutathione peroxidase
	K03564	BCP; peroxiredoxin Q/BCP
Catalase	K03781	katE; catalase
superoxide dismutase	K04564	SOD2; superoxide dismutase, Fe-Mn family
	K04565	SOD1; superoxide dismutase, Cu-Zn family
glutathione S-transferase	K00799	GST; glutathione S-transferase
Hydroperoxide	K03386	PRDX2_4; peroxiredoxin (alkyl hydroperoxide reductase subunit C)
Heat shock	K03687	GRPE; molecular chaperone GrpE
	K03686	dnaJ; molecular chaperone DnaJ
	K04043	dnaK; molecular chaperone DnaK

Rhodamese	K11996	MOCS3; adenylyltransferase and sulfurtransferase [EC:2.7.7.80 2.8.1.11]
Genes involved in the N cycle		
Nitrogenase complex	K04487	iscS; cysteine desulfurase [EC:2.8.1.7]
Genes involved in salt tolerance		
Trehalose Metabolism	K16055	TPS; trehalose 6-phosphate synthase/phosphatase [EC:2.4.1.15 3.1.3.12]
	K22337	TSL1; trehalose 6-phosphate synthase complex regulatory subunit
	K00697	otsA; trehalose 6-phosphate synthase [EC:2.4.1.15 2.4.1.347]
	K01194	TREH; alpha,alpha-trehalase [EC:3.2.1.28]
Genes involved in Na^+ and K^+ transport		
	K03316	TC.CPA1; monovalent cation:H+ antiporter, CPA1 family
	K01507	ppa; inorganic pyrophosphatase [EC:3.6.1.1]

Genes involved in pH wide adaptation		
Acid expressed	K05389	KCNKF; potassium channel subfamily K, other eukaryote
	K14429	SLC12A9; solute carrier family 12 (potassium/chloride transporters), member 9
	K13754	SLC24A6; solute carrier family 24 (sodium/potassium/calcium exchanger), member 6
	K04078	groES; chaperonin GroES
	K04077	groEL; chaperonin GroEL
	K04043	dnaK; molecular chaperone DnaK
	K03695	clpB; ATP-dependent Clp protease ATP-binding subunit ClpB
	K03544	clpX; ATP-dependent Clp protease ATP-binding subunit ClpX
Alkaline expressed	K07300	chaA; Ca2+:H+ antiporter
	K03316	TC.CPA1; monovalent cation:H+ antiporter, CPA1 family
	K23541	TMEM165; Ca2+/H+ antiporter, TMEM165/GDT1 family
	K03316	TC.CPA1; monovalent cation:H+ antiporter, CPA1 family
	K08744	CRLS; cardiolipin synthase (CMP-forming) [EC:2.7.8.41]
	K08744	CRLS; cardiolipin synthase (CMP-forming) [EC:2.7.8.41]
	K20498	DSD1; D-serine ammonia-lyase [EC:4.3.1.18]
	K17989	SDS; L-serine/L-threonine ammonia-lyase [EC:4.3.1.17 4.3.1.19]
Resistance to heavy metals		
Divalent	g2210.t1	integral component of membrane [GO:0016021]; mitochondrion [GO:0005739]; pyrimidine nucleotide transmembrane transporter activity [GO:0015218]; divalent metal ion transport [GO:0070838]; mitochondrial genome maintenance [GO:0000002]; regulation of mitochondrial membrane potential [GO:0051881]
sensor kinase	g3166.t1	ATP binding [GO:0005524]; phosphorelay sensor kinase activity [GO:0000155]
	g4634.t1	histidine phosphotransfer kinase activity [GO:0009927]; osmosensor activity [GO:0005034]; phosphorelay sensor kinase activity [GO:0000155]
Arsenic	K01551	arsA; arsenite/tail-anchored protein-transporting ATPase

		[EC:7.3.2.7 7.3.-.-]
	K01551	arsA; arsenite/tail-anchored protein-transporting ATPase [EC:7.3.2.7 7.3.-.-]
	K03325	ACR3; arsenite transporter
	K03325	ACR3; arsenite transporter
	K03325	ACR3; arsenite transporter
Copper	K19791	FET3_5; iron transport multicopper oxidase
	K19791	FET3_5; iron transport multicopper oxidase
	K19791	FET3_5; iron transport multicopper oxidase
	K14686	SLC31A1; solute carrier family 31 (copper transporter), member 1
	K19791	FET3_5; iron transport multicopper oxidase
Cobalt	g1946.t1	cellular cobalt ion homeostasis [GO:0006877]; cellular manganese ion homeostasis [GO:0030026]; cobalt ion transport [GO:0006824]; manganese ion transport [GO:0006828]
	g3821.t1	cellular cobalt ion homeostasis [GO:0006877]; cellular detoxification of cadmium ion [GO:0098849]; cellular zinc ion homeostasis [GO:0006882]; zinc ion import into endoplasmic reticulum [GO:0140209]
Mercury	g3176.t1	SCF ubiquitin ligase complex [GO:0019005]; cellular response to methylmercury [GO:0071406]; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process [GO:0031146]
Molybdenum	g78.t1	MoCF_biosynth, Probable molybdopterin binding domain
	g78.t1	MoCF_biosynth, Probable molybdopterin binding domain
	g1469.t1	Molybdopterin oxidoreductase
	g1469.t1	Molybdopterin oxidoreductase
	g3082.t1	MoCF_biosynth, Probable molybdopterin binding domain
	g3845.t1	Mob_synth_C, Molybdenum Cofactor Synthesis C
	g4014.t1	Oxidored_molyb, Oxidoreductase molybdopterin binding domain
Cadmium	g269.t1	fungal-type vacuole membrane [GO:0000329]; integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; ATPase activity [GO:0016887]; ATPase-coupled transmembrane transporter activity [GO:0042626]; bilirubin transmembrane transporter activity [GO:0015127]; cadmium ion transmembrane transporter activity [GO:0015086]; vacuole fusion, non-autophagic [GO:0042144]
	g387.t1	mitochondrion [GO:0005739]; thioredoxin peroxidase activity [GO:0008379]; cell redox homeostasis [GO:0045454]; cellular response to oxidative stress [GO:0034599]; response to cadmium ion [GO:0046686]
	g859.t1	Cdc48p-Npl4p-Vms1p AAA ATPase complex [GO:0036266]; cytosol [GO:0005829]; Doa10p ubiquitin ligase complex [GO:0000837]; Hrd1p ubiquitin ligase ERAD-L complex [GO:0000839]; nucleus [GO:0005634]; RQC complex [GO:1990112]; VCP-NPL4-UFD1 AAA ATPase complex [GO:0034098]; ATP binding [GO:0005524]; ATPase activity [GO:0016887]; identical protein binding [GO:0042802]; protein phosphatase regulator activity [GO:0019888]; ubiquitin binding [GO:0043130]; ATP metabolic process [GO:0046034]; cellular protein complex disassembly [GO:0043624]; cytoplasm protein quality control by the ubiquitin-proteasome system

		[GO:0071629]; endoplasmic reticulum membrane fusion [GO:0016320]; ER-associated misfolded protein catabolic process [GO:0071712]; mitochondria-associated ubiquitin-dependent protein catabolic process [GO:0072671]; mitotic spindle disassembly [GO:0051228]; negative regulation of telomerase activity [GO:0051974]; nonfunctional rRNA decay [GO:0070651]; nuclear protein quality control by the ubiquitin-proteasome system [GO:0071630]; piecemeal microautophagy of the nucleus [GO:0034727]; positive regulation of histone H2B ubiquitination [GO:2001168]; positive regulation of mitochondrial fusion [GO:0010636]; positive regulation of protein localization to nucleus [GO:1900182]; protein transport to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway [GO:0043328]; retrograde protein transport, ER to cytosol [GO:0030970]; ribophagy [GO:0034517]; ribosome-associated ubiquitin-dependent protein catabolic process [GO:1990116]; SCF complex disassembly in response to cadmium stress [GO:1990171]; sister chromatid biorientation [GO:0031134]; stress-induced homeostatically regulated protein degradation pathway [GO:0120174]; ubiquitin-dependent ERAD pathway [GO:0030433]
	g1517.t1	cytoplasm [GO:0005737]; adenylosuccinate synthase activity [GO:0004019]; GTP binding [GO:0005525]; magnesium ion binding [GO:0000287]; sulfinylpropanyl adenylate synthase [GO:0061483]; 'de novo' AMP biosynthetic process [GO:0044208]; cellular response to cadmium ion [GO:0071276]; fumarate metabolic process [GO:0006106]
	g2030.t1	integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]; metal ion transmembrane transporter activity [GO:0046873]; solute:proton symporter activity [GO:0015295]; cadmium ion transport [GO:0015691]; cellular cadmium ion homeostasis [GO:0006876]; cellular copper ion homeostasis [GO:0006878]; cellular manganese ion homeostasis [GO:0030026]; copper ion transport [GO:0006825]; iron ion transport [GO:0006826]; manganese ion transport [GO:0006828]
	g2030.t1	integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]; metal ion transmembrane transporter activity [GO:0046873]; solute:proton symporter activity [GO:0015295]; cadmium ion transport [GO:0015691]; cellular cadmium ion homeostasis [GO:0006876]; cellular copper ion homeostasis [GO:0006878]; cellular manganese ion homeostasis [GO:0030026]; copper ion transport [GO:0006825]; iron ion transport [GO:0006826]; manganese ion transport [GO:0006828]
	g2626.t1	fungal-type vacuole membrane [GO:0000329]; integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; ATPase activity [GO:0016887]; ATPase-coupled glutathione S-conjugate transmembrane transporter activity [GO:0015431]; ATPase-coupled phytochelatin transmembrane transporter activity [GO:0044604]; ATPase-coupled transmembrane transporter activity [GO:0042626]; bilirubin transmembrane transporter activity [GO:0015127]; cell redox homeostasis [GO:0045454]; cellular detoxification of cadmium ion [GO:0098849]; glutathione metabolic process [GO:0006749]; glutathione transmembrane import into vacuole [GO:0071996]; phytochelatin 2 import into vacuole [GO:0036246]; vacuole fusion, non-autophagic [GO:0042144]
	g3821.t1	endoplasmic reticulum membrane [GO:0005789]; fungal-type vacuole membrane [GO:0000329]; integral component of

		membrane [GO:0016021]; zinc ion transmembrane transporter activity [GO:0005385]; cellular cobalt ion homeostasis [GO:0006877]; cellular detoxification of cadmium ion [GO:0098849]; cellular zinc ion homeostasis [GO:0006882]; zinc ion import into endoplasmic reticulum [GO:0140209]
	g4593.t1	nuclear SCF ubiquitin ligase complex [GO:0043224]; identical protein binding [GO:0042802]; protein binding, bridging [GO:0030674]; ubiquitin binding [GO:0043130]; DNA replication initiation [GO:0006270]; protein polyubiquitination [GO:0000209]; regulation of DNA-dependent DNA replication initiation [GO:0030174]; regulation of transcription involved in G1/S transition of mitotic cell cycle [GO:0000083]; response to arsenic-containing substance [GO:0046685]; response to cadmium ion [GO:0046686]; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process [GO:0031146]
	g6370.t1	nuclear SCF ubiquitin ligase complex [GO:0043224]; identical protein binding [GO:0042802]; protein binding, bridging [GO:0030674]; ubiquitin binding [GO:0043130]; DNA replication initiation [GO:0006270]; protein polyubiquitination [GO:0000209]; regulation of DNA-dependent DNA replication initiation [GO:0030174]; regulation of transcription involved in G1/S transition of mitotic cell cycle [GO:0000083]; response to arsenic-containing substance [GO:0046685]; response to cadmium ion [GO:0046686]; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process [GO:0031146]
	g6660.t1	glutathione synthase complex [GO:0036087]; ATP binding [GO:0005524]; glutathione binding [GO:0043295]; glutathione synthase activity [GO:0004363]; magnesium ion binding [GO:0000287]; protein homodimerization activity [GO:0042803]; cellular detoxification of cadmium ion [GO:0098849]; phytochelatin biosynthetic process [GO:0046938]; phytochelatin-metal complex formation [GO:0090423]

482

Aromatic Compounds Degradation cleavage		
Catechol	g3300.t1	catechol 1,2-dioxygenase activity [GO:0018576]; ferric iron binding [GO:0008199]; catechol-containing compound metabolic process [GO:0009712]
	g3300.t1	Catechol dioxygenase N terminus
Nitrilase	g5205.t1	nitrilase activity [GO:0000257]; nitrogen compound metabolic process [GO:0006807]
	g6098.t1	nitrilase activity [GO:0000257]; nitrogen compound metabolic process [GO:0006807]
Phenol hydrolase	g3741.t1	Phenol hydroxylase, C-terminal dimerisation domain
	g5896.t1	Phenol hydroxylase, C-terminal dimerisation domain
	g6214.t1	Phenol hydroxylase, C-terminal dimerisation domain
Ferridoxin	K22071	FDX2; ferredoxin-2, mitochondrial
	K22071	FDX2; ferredoxin-2, mitochondrial
Hydrolase	g381.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g381.t1	haloacid dehalogenase-like hydrolase
	g4374.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g4584.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit

	g5107.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g6016.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
Biofilm formation		
Cell adhession	g4902.t1	cell adhesion [GO:0007155]; cellular response to nitrogen starvation [GO:0006995]; establishment of mitotic spindle orientation [GO:0000132]; fungal-type cell wall assembly [GO:0071940]; invasive growth in response to glucose limitation [GO:0001403]; negative regulation of translation [GO:0017148]; positive regulation of filamentous growth of a population of unicellular organisms in response to starvation [GO:1900436]; positive regulation of gluconeogenesis [GO:0045722]; positive regulation of macroautophagy [GO:0016239]; positive regulation of pseudohyphal growth [GO:2000222]; replicative cell aging [GO:0001302]; response to unfolded protein [GO:0006986]; single-species surface biofilm formation [GO:0090606]
	g5629.t1	cell adhesion involved in single-species biofilm formation [GO:0043709]; chromatin silencing [GO:0006342]; negative regulation of chromatin silencing at rDNA [GO:0061188]; negative regulation of chromatin silencing at silent mating-type cassette [GO:0061186]; negative regulation of chromatin silencing at telomere [GO:0031939]; negative regulation of transcription by RNA polymerase II [GO:0000122]; positive regulation of transcription from RNA polymerase II promoter in response to heat stress [GO:0061408]; regulation of invasive growth in response to glucose limitation [GO:2000217]
	g5643.t1	cell adhesion involved in single-species biofilm formation [GO:0043709]; chromatin silencing [GO:0006342]; negative regulation of chromatin silencing at rDNA [GO:0061188]; negative regulation of chromatin silencing at silent mating-type cassette [GO:0061186]; negative regulation of chromatin silencing at telomere [GO:0031939]; negative regulation of transcription by RNA polymerase II [GO:0000122]; positive regulation of transcription from RNA polymerase II promoter in response to heat stress [GO:0061408]; regulation of invasive growth in response to glucose limitation [GO:2000217]
Quorum sensing		
Homoserine		
	K17069	MET17; O-acetylhomoserine/O-acetylserine sulfhydrylase [EC:2.5.1.49 2.5.1.47]
	K17069	MET17; O-acetylhomoserine/O-acetylserine sulfhydrylase [EC:2.5.1.49 2.5.1.47]
	K17069	MET17; O-acetylhomoserine/O-acetylserine sulfhydrylase [EC:2.5.1.49 2.5.1.47]
	K17069	MET17; O-acetylhomoserine/O-acetylserine sulfhydrylase [EC:2.5.1.49 2.5.1.47]
	K00003	hom; homoserine dehydrogenase [EC:1.1.1.3]
	K00003	hom; homoserine dehydrogenase [EC:1.1.1.3]
	K00641	metX; homoserine O-acetyltransferase/O-succinyltransferase [EC:2.3.1.31 2.3.1.46]
	K00641	metX; homoserine O-acetyltransferase/O-succinyltransferase [EC:2.3.1.31 2.3.1.46]
	K00872	thrB1; homoserine kinase [EC:2.7.1.39]
Biocontrol agents and plant growth regulators related-genes		
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-

		semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K01657	trpE; anthranilate synthase component I [EC:4.1.3.27]
	g2983.t1	Isochorismatase
	g4122.t1	Isochorismatase
	g1968.t1	3-hydroxyanthranilate 3,4-dioxygenase (EC 1.13.11.6) (3-hydroxyanthranilate oxygenase) (3-HAO) (3-hydroxyanthranilic acid dioxygenase) (HAD) (Biosynthesis of nicotinic acid protein 1)
	scf7180000000928.g1968.t1	00380 Tryptophan metabolism

Figures

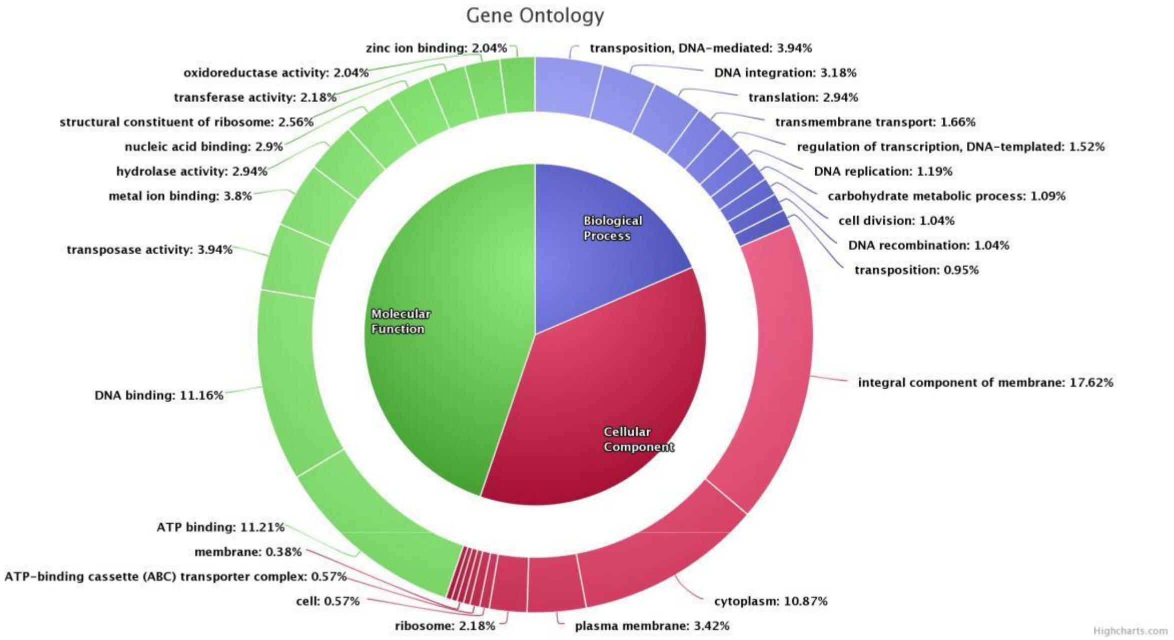


Figure 1. GO Annotation graph

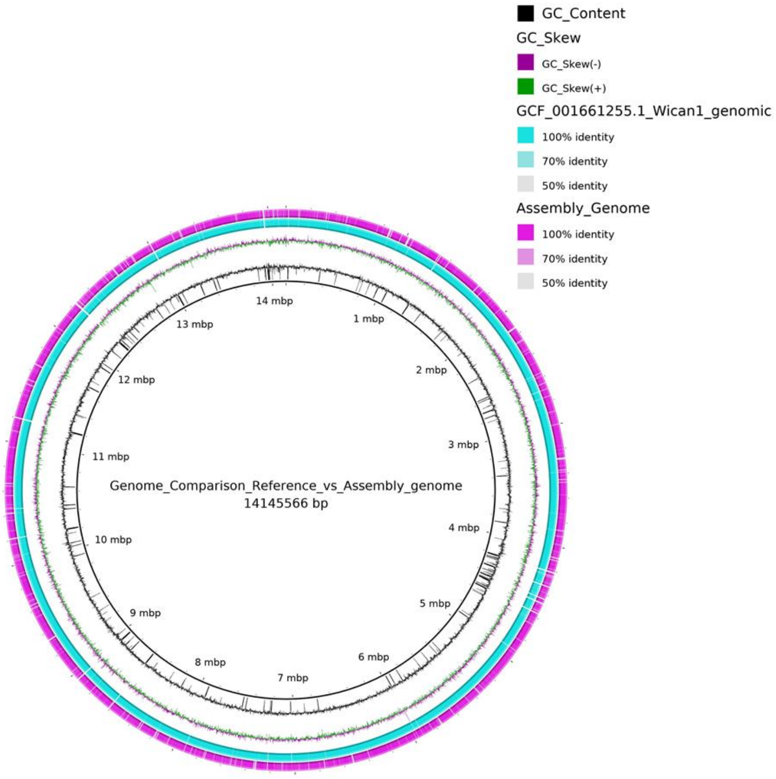


Figure 2. Genome comparison of reference (NRRL Y-366-8) and assembled genome (MSD1)

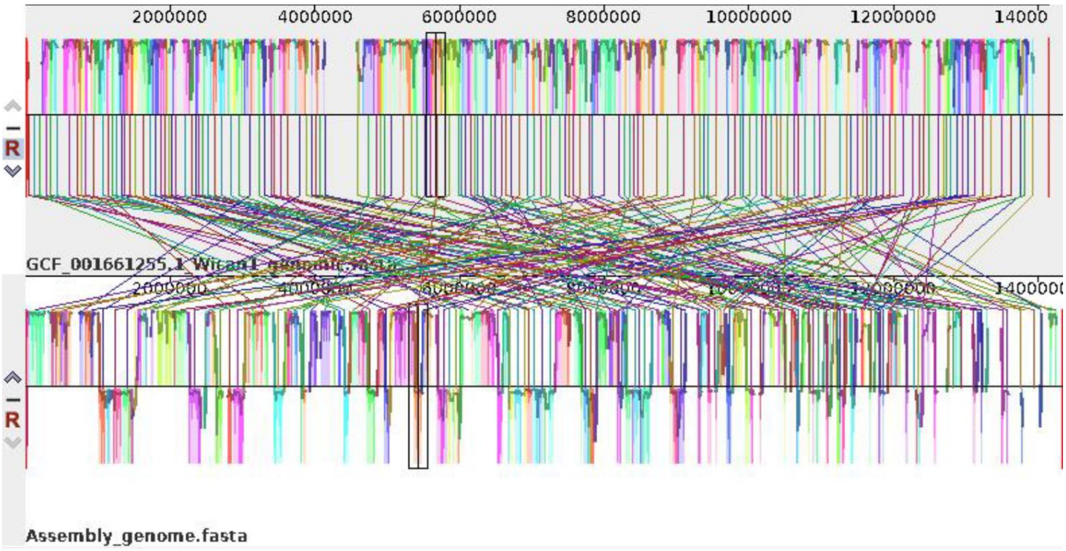


Figure 3. Synteny map of reference (NRRL Y-366-8) and assemble genome (MSD1)

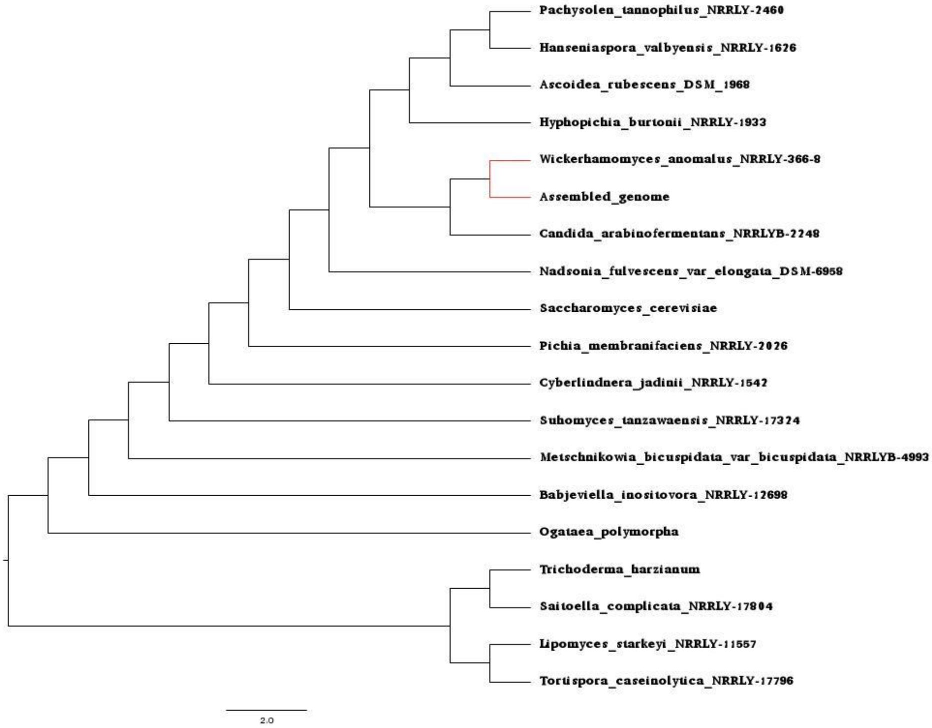


Figure 4. Phylogenetic tree visualizing the comparative genome analysis of MSD1 (Assembled genome) along with other fungal and yeast taxa.

Table1. Illumina Read Statistics

Sample	Read Count
Total raw reads (in million)	6713813 (Read1, Read2)
Total processed reads (in million)	6417015(Read1, Read2)

Table2. Nanopore Read Statistics

Parameters	**Read Stat**
Reads Generated	374070
Maximum Read Length	61005
Minimum Read Length	94
Average Read Length	1177.6
Median Read Length	2311
Total Reads Length	440506966
Total Number of Non-ATGC Characters	0
Percentage of Non-ATGC Characters	0
Reads >= 100 bp	374067
Reads >= 200 bp	372318
Reads >= 500 bp	241259
Reads >= 1 Kbp	125627
Reads >= 10 Kbp	1861
N50 value	1847

Table 3. Assembly statistics

Parameter	**Statistical data**
Contigs Generated	289
Maximum Contig Length	275125
Minimum Contig Length	3374
Average Contig Length	49645.9
Median Contig Length	51021
Total Contigs Length	14347675
Total Number of Non-ATGC Characters	0
Percentage of Non-ATGC Characters	0
Contigs >= 1 Kbp	289
Contigs >= 10 Kbp	259
N50 value	83236

Table 4. Genes related to Plant Growth Promoting traits that are annotated to be present in *Wickerhamomyces anomalus* strain MSD1

	Gene or protein ID	**Gene Annotation**
Indole-3-Acetic acid	K01426	Amidase
	K00128	Aldehyde dehydrogenase
	K01568	pyruvate decarboxylase
GABA production		
gabD	g6155.t1	succinate-semialdehyde dehydrogenase (NAD+) activity [GO:0004777]; succinate-semialdehyde dehydrogenase [NAD(P)+] activity [GO:0009013]; cellular response to oxidative stress [GO:0034599]; gamma-aminobutyric acid catabolic process [GO:0009450]; glutamate decarboxylation to succinate [GO:0006540]
gabT	g4127.t1	cytosol [GO:0005829]; 4-aminobutyrate transaminase activity [GO:0003867]; pyridoxal phosphate binding [GO:0030170]; gamma-aminobutyric acid catabolic process [GO:0009450]
	g5717.t1	cytosol [GO:0005829]; 4-aminobutyrate transaminase activity [GO:0003867]; pyridoxal phosphate binding [GO:0030170]; gamma-aminobutyric acid catabolic process [GO:0009450]
Antimicrobial		
Phenazine	g6439.t1	Phenazine biosynthesis-like protein
Acetoin and 2,3 butanediol synthesis	K01653	Acetolactete synthase 1
	K01652	Acetolactete synthase
	K00004	BDH; (R,R)-butanediol dehydrogenase / meso-butanediol dehydrogenase / diacetyl reductase [EC:1.1.1.4 1.1.1.- 1.1.1.303]
Phosphate	K08176	PHO84; MFS transporter, PHS family, inorganic phosphate transporter
	K14430	PHO87_91; phosphate transporter
	K15108	SLC25A19; solute carrier family 25 (mitochondrial thiamine pyrophosphate transporter), member 19
	K14684	SLC25A23S; solute carrier family 25 (mitochondrial phosphate transporter), member 23/24/25/41
	K15102	SLC25A3; solute carrier family 25 (mitochondrial phosphate transporter), member 3
Antimicrobial compound		
Chitinase production	K01183	Putative chitinase II
GABA	K00135	succinate-semialdehyde dehydrogenase
Other PGPR fitness conferring genes		
	K01480	speB; agmatinase

	K00797	speE; spermidine synthase
	K00802	SMS; spermine synthase [EC:2.5.1.22]
Resistance to antifungal drugs	K03327	TC.MATE; multidrug resistance protein, MATE family
	K08157	TPO1; MFS transporter, DHA1 family, multidrug resistance protein
	K08165	ATR1; MFS transporter, DHA2 family, multidrug resistance protein
	K08158	MDR1; MFS transporter, DHA1 family, multidrug resistance protein
Siderophore production (Iron homeostasis)		
	g2983.t1	catalytic activity [GO:0003824]
	g4122.t1	nicotinamidase activity [GO:0008936]
	K08197	ARN; MFS transporter, SIT family, siderophore-iron:H+ symporter
	K23503	SFXN5; sideroflexin-5
Fe transport	K07243	FTR; high-affinity iron transporter
	K19791	FET3_5; iron transport multicopper oxidase
	K12346	SMF; metal iron transporter
	K22736	VIT; vacuolar iron transporter family protein
	K07243	FTR; high-affinity iron transporter
	K15113	SLC25A28_37; solute carrier family 25 (mitochondrial iron transporter), member 28/37
	K02304	MET8; precorrin-2 dehydrogenase / sirohydrochlorin ferrochelatase [EC:1.3.1.76 4.99.1.4]
	K01772	hemH; protoporphyrin/coproporphyrin ferrochelatase [EC:4.99.1.1 4.99.1.9]
Sulfur metabolism	g1937.t1	Sulfate adenylyltransferase (EC 2.7.7.4) (ATP-sulfurylase) (Sulfate adenylate transferase) (SAT)
	g5561.t1	Adenylyl-sulfate kinase (EC 2.7.1.25)
	g6660.t1	Glutathione synthetase (GSH-S) (EC 6.3.2.3)
Potassium	g381.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g2155.t1	K+ potassium transporter
	g4374.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g139.t1	potassium ion transport [GO:0006813]; protein insertion into mitochondrial inner membrane from matrix [GO:0032979]; proton transmembrane transport [GO:1902600]
	g1052.t1	cellular potassium ion homeostasis [GO:0030007]; positive regulation of mitochondrial translation [GO:0070131]; potassium ion transport [GO:0006813]; protein insertion into mitochondrial inner membrane from matrix [GO:0032979]; proton transmembrane transport [GO:1902600]

	g1718.t1	potassium ion transport [GO:0006813]; protein insertion into mitochondrial inner membrane from matrix [GO:0032979]; proton transmembrane transport [GO:1902600]
Zinc transport	K14709	SLC39A1_2_3; solute carrier family 39 (zinc transporter), member 1/2/3
	K14688	SLC30A1; solute carrier family 30 (zinc transporter), member 1
	K14713	SLC39A7; solute carrier family 39 (zinc transporter), member 7
	K07238	TC.ZIP; zinc transporter, ZIP family
	K14692	SLC30A5_7; solute carrier family 30 (zinc transporter), member 5/7
Resistance to oxidative stress		
Peroxidase	K00432	gpx; glutathione peroxidase
	K03564	BCP; peroxiredoxin Q/BCP
Catalase	K03781	katE; catalase
superoxide dismutase	K04564	SOD2; superoxide dismutase, Fe-Mn family
	K04565	SOD1; superoxide dismutase, Cu-Zn family
glutathione S-transferase	K00799	GST; glutathione S-transferase
Hydroperoxide	K03386	PRDX2_4; peroxiredoxin (alkyl hydroperoxide reductase subunit C)
Heat shock	K03687	GRPE; molecular chaperone GrpE
	K03686	dnaJ; molecular chaperone DnaJ
	K04043	dnaK; molecular chaperone DnaK
Rhodamese	K11996	MOCS3; adenylyltransferase and sulfurtransferase [EC:2.7.7.80 2.8.1.11]
Genes involved in the N cycle		
Nitrogenase complex	K04487	iscS; cysteine desulfurase [EC:2.8.1.7]
Genes involved in salt tolerance		
Trehalose Metabolism	K16055	TPS; trehalose 6-phosphate synthase/phosphatase [EC:2.4.1.15 3.1.3.12]
	K22337	TSL1; trehalose 6-phosphate synthase complex regulatory subunit
	K00697	otsA; trehalose 6-phosphate synthase [EC:2.4.1.15 2.4.1.347]
	K01194	TREH; alpha,alpha-trehalase [EC:3.2.1.28]
Genes involved in Na^+ and K^+ transport		
	K03316	TC.CPA1; monovalent cation:H+ antiporter, CPA1 family
	K01507	ppa; inorganic pyrophosphatase [EC:3.6.1.1]

Genes involved in pH wide adaptation		
Acid expressed	K05389	KCNKF; potassium channel subfamily K, other eukaryote
	K14429	SLC12A9; solute carrier family 12 (potassium/chloride transporters), member 9
	K13754	SLC24A6; solute carrier family 24 (sodium/potassium/calcium exchanger), member 6
	K04078	groES; chaperonin GroES
	K04077	groEL; chaperonin GroEL
	K04043	dnaK; molecular chaperone DnaK
	K03695	clpB; ATP-dependent Clp protease ATP-binding subunit ClpB
	K03544	clpX; ATP-dependent Clp protease ATP-binding subunit ClpX
Alkaline expressed	K07300	chaA; Ca2+:H+ antiporter
	K03316	TC.CPA1; monovalent cation:H+ antiporter, CPA1 family
	K23541	TMEM165; Ca2+/H+ antiporter, TMEM165/GDT1 family
	K03316	TC.CPA1; monovalent cation:H+ antiporter, CPA1 family
	K08744	CRLS; cardiolipin synthase (CMP-forming) [EC:2.7.8.41]
	K08744	CRLS; cardiolipin synthase (CMP-forming) [EC:2.7.8.41]
	K20498	DSD1; D-serine ammonia-lyase [EC:4.3.1.18]
	K17989	SDS; L-serine/L-threonine ammonia-lyase [EC:4.3.1.17 4.3.1.19]
Resistance to heavy metals		
Divalent	g2210.t1	integral component of membrane [GO:0016021]; mitochondrion [GO:0005739]; pyrimidine nucleotide transmembrane transporter activity [GO:0015218]; divalent metal ion transport [GO:0070838]; mitochondrial genome maintenance [GO:0000002]; regulation of mitochondrial membrane potential [GO:0051881]
sensor kinase	g3166.t1	ATP binding [GO:0005524]; phosphorelay sensor kinase activity [GO:0000155]
	g4634.t1	histidine phosphotransfer kinase activity [GO:0009927]; osmosensor activity [GO:0005034]; phosphorelay sensor kinase activity [GO:0000155]
Arsenic	K01551	arsA; arsenite/tail-anchored protein-transporting ATPase [EC:7.3.2.7 7.3.-.-]
	K01551	arsA; arsenite/tail-anchored protein-transporting ATPase [EC:7.3.2.7 7.3.-.-]
	K03325	ACR3; arsenite transporter
	K03325	ACR3; arsenite transporter
	K03325	ACR3; arsenite transporter
Copper	K19791	FET3_5; iron transport multicopper oxidase
	K19791	FET3_5; iron transport multicopper oxidase
	K19791	FET3_5; iron transport multicopper oxidase

	K14686	SLC31A1; solute carrier family 31 (copper transporter), member 1
	K19791	FET3_5; iron transport multicopper oxidase
Cobalt	g1946.t1	cellular cobalt ion homeostasis [GO:0006877]; cellular manganese ion homeostasis [GO:0030026]; cobalt ion transport [GO:0006824]; manganese ion transport [GO:0006828]
	g3821.t1	cellular cobalt ion homeostasis [GO:0006877]; cellular detoxification of cadmium ion [GO:0098849]; cellular zinc ion homeostasis [GO:0006882]; zinc ion import into endoplasmic reticulum [GO:0140209]
Mercury	g3176.t1	SCF ubiquitin ligase complex [GO:0019005]; cellular response to methylmercury [GO:0071406]; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process [GO:0031146]
Molybdenum	g78.t1	MoCF_biosynth, Probable molybdopterin binding domain
	g78.t1	MoCF_biosynth, Probable molybdopterin binding domain
	g1469.t1	Molybdopterin oxidoreductase
	g1469.t1	Molybdopterin oxidoreductase
	g3082.t1	MoCF_biosynth, Probable molybdopterin binding domain
	g3845.t1	Mob_synth_C, Molybdenum Cofactor Synthesis C
	g4014.t1	Oxidored_molyb, Oxidoreductase molybdopterin binding domain
Cadmium	g269.t1	fungal-type vacuole membrane [GO:0000329]; integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; ATPase activity [GO:0016887]; ATPase-coupled transmembrane transporter activity [GO:0042626]; bilirubin transmembrane transporter activity [GO:0015127]; cadmium ion transmembrane transporter activity [GO:0015086]; vacuole fusion, non-autophagic [GO:0042144]
	g387.t1	mitochondrion [GO:0005739]; thioredoxin peroxidase activity [GO:0008379]; cell redox homeostasis [GO:0045454]; cellular response to oxidative stress [GO:0034599]; response to cadmium ion [GO:0046686]
	g859.t1	Cdc48p-Npl4p-Vms1p AAA ATPase complex [GO:0036266]; cytosol [GO:0005829]; Doa10p ubiquitin ligase complex [GO:0000837]; Hrd1p ubiquitin ligase ERAD-L complex [GO:0000839]; nucleus [GO:0005634]; RQC complex [GO:1990112]; VCP-NPL4-UFD1 AAA ATPase complex [GO:0034098]; ATP binding [GO:0005524]; ATPase activity [GO:0016887]; identical protein binding [GO:0042802]; protein phosphatase regulator activity [GO:0019888]; ubiquitin binding [GO:0043130]; ATP metabolic process [GO:0046034]; cellular protein complex disassembly [GO:0043624]; cytoplasm protein quality control by the ubiquitin-proteasome system [GO:0071629]; endoplasmic reticulum membrane fusion [GO:0016320]; ER-associated

		misfolded protein catabolic process [GO:0071712]; mitochondria-associated ubiquitin-dependent protein catabolic process [GO:0072671]; mitotic spindle disassembly [GO:0051228]; negative regulation of telomerase activity [GO:0051974]; nonfunctional rRNA decay [GO:0070651]; nuclear protein quality control by the ubiquitin-proteasome system [GO:0071630]; piecemeal microautophagy of the nucleus [GO:0034727]; positive regulation of histone H2B ubiquitination [GO:2001168]; positive regulation of mitochondrial fusion [GO:0010636]; positive regulation of protein localization to nucleus [GO:1900182]; protein transport to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway [GO:0043328]; retrograde protein transport, ER to cytosol [GO:0030970]; ribophagy [GO:0034517]; ribosome-associated ubiquitin-dependent protein catabolic process [GO:1990116]; SCF complex disassembly in response to cadmium stress [GO:1990171]; sister chromatid biorientation [GO:0031134]; stress-induced homeostatically regulated protein degradation pathway [GO:0120174]; ubiquitin-dependent ERAD pathway [GO:0030433]
	g1517.t1	cytoplasm [GO:0005737]; adenylosuccinate synthase activity [GO:0004019]; GTP binding [GO:0005525]; magnesium ion binding [GO:0000287]; sulfinylpropanyl adenylate synthase [GO:0061483]; 'de novo' AMP biosynthetic process [GO:0044208]; cellular response to cadmium ion [GO:0071276]; fumarate metabolic process [GO:0006106]
	g2030.t1	integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]; metal ion transmembrane transporter activity [GO:0046873]; solute:proton symporter activity [GO:0015295]; cadmium ion transport [GO:0015691]; cellular cadmium ion homeostasis [GO:0006876]; cellular copper ion homeostasis [GO:0006878]; cellular manganese ion homeostasis [GO:0030026]; copper ion transport [GO:0006825]; iron ion transport [GO:0006826]; manganese ion transport [GO:0006828]
	g2030.t1	integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]; metal ion transmembrane transporter activity [GO:0046873]; solute:proton symporter activity [GO:0015295]; cadmium ion transport [GO:0015691]; cellular cadmium ion homeostasis [GO:0006876]; cellular copper ion homeostasis [GO:0006878]; cellular manganese ion homeostasis [GO:0030026]; copper ion transport [GO:0006825]; iron ion transport [GO:0006826]; manganese ion transport [GO:0006828]
	g2626.t1	fungal-type vacuole membrane [GO:0000329]; integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; ATPase activity [GO:0016887]; ATPase-coupled glutathione S-conjugate transmembrane

		transporter activity [GO:0015431]; ATPase-coupled phytochelatin transmembrane transporter activity [GO:0044604]; ATPase-coupled transmembrane transporter activity [GO:0042626]; bilirubin transmembrane transporter activity [GO:0015127]; cell redox homeostasis [GO:0045454]; cellular detoxification of cadmium ion [GO:0098849]; glutathione metabolic process [GO:0006749]; glutathione transmembrane import into vacuole [GO:0071996]; phytochelatin 2 import into vacuole [GO:0036246]; vacuole fusion, non-autophagic [GO:0042144]
	g3821.t1	endoplasmic reticulum membrane [GO:0005789]; fungal-type vacuole membrane [GO:0000329]; integral component of membrane [GO:0016021]; zinc ion transmembrane transporter activity [GO:0005385]; cellular cobalt ion homeostasis [GO:0006877]; cellular detoxification of cadmium ion [GO:0098849]; cellular zinc ion homeostasis [GO:0006882]; zinc ion import into endoplasmic reticulum [GO:0140209]
	g4593.t1	nuclear SCF ubiquitin ligase complex [GO:0043224]; identical protein binding [GO:0042802]; protein binding, bridging [GO:0030674]; ubiquitin binding [GO:0043130]; DNA replication initiation [GO:0006270]; protein polyubiquitination [GO:0000209]; regulation of DNA-dependent DNA replication initiation [GO:0030174]; regulation of transcription involved in G1/S transition of mitotic cell cycle [GO:0000083]; response to arsenic-containing substance [GO:0046685]; response to cadmium ion [GO:0046686]; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process [GO:0031146]
	g6370.t1	nuclear SCF ubiquitin ligase complex [GO:0043224]; identical protein binding [GO:0042802]; protein binding, bridging [GO:0030674]; ubiquitin binding [GO:0043130]; DNA replication initiation [GO:0006270]; protein polyubiquitination [GO:0000209]; regulation of DNA-dependent DNA replication initiation [GO:0030174]; regulation of transcription involved in G1/S transition of mitotic cell cycle [GO:0000083]; response to arsenic-containing substance [GO:0046685]; response to cadmium ion [GO:0046686]; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process [GO:0031146]
	g6660.t1	glutathione synthase complex [GO:0036087]; ATP binding [GO:0005524]; glutathione binding [GO:0043295]; glutathione synthase activity [GO:0004363]; magnesium ion binding [GO:0000287]; protein homodimerization activity [GO:0042803]; cellular detoxification of cadmium ion [GO:0098849]; phytochelatin biosynthetic process [GO:0046938]; phytochelatin-metal complex formation [GO:0090423]

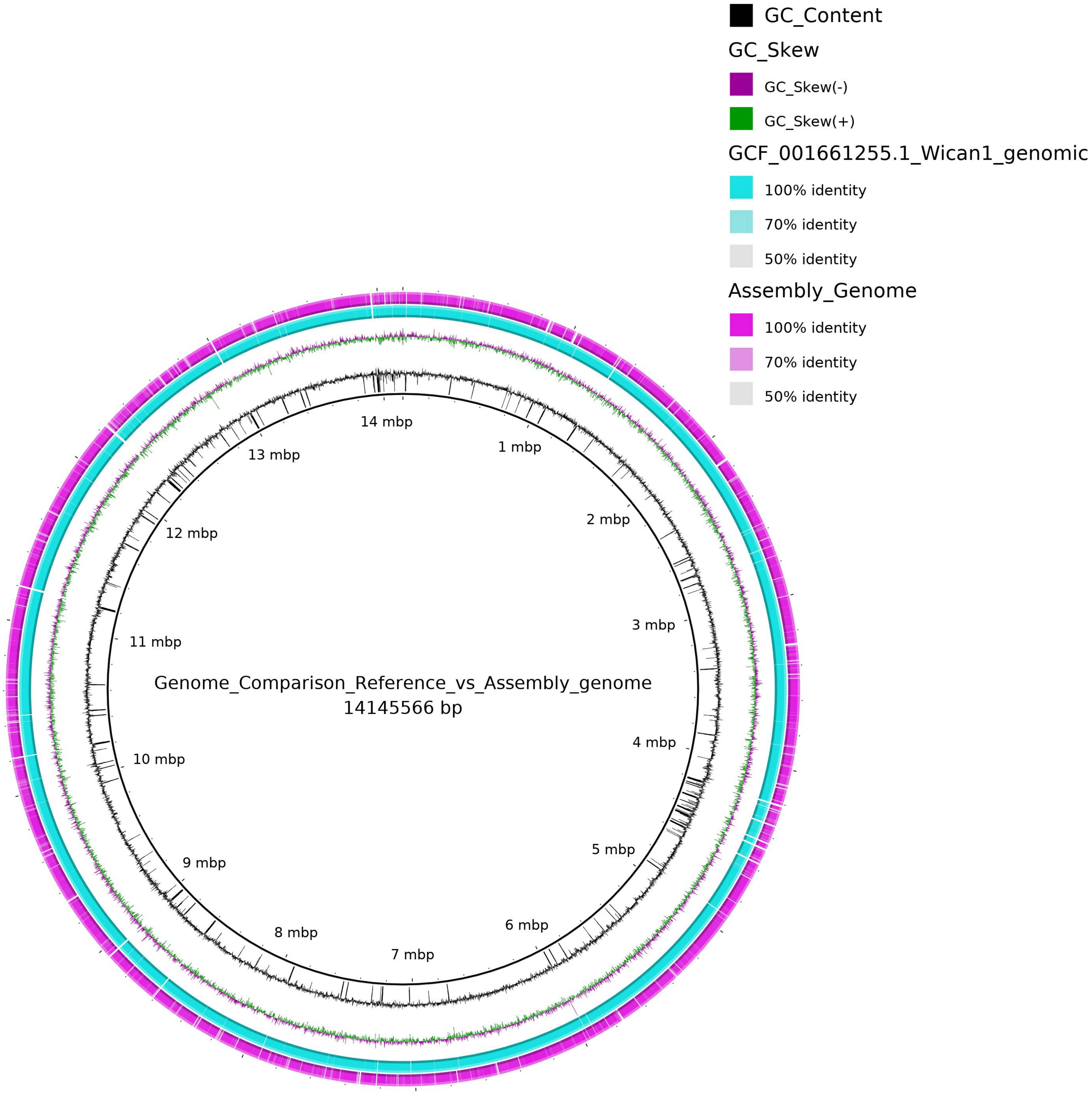
Aromatic Compounds Degradation cleavage

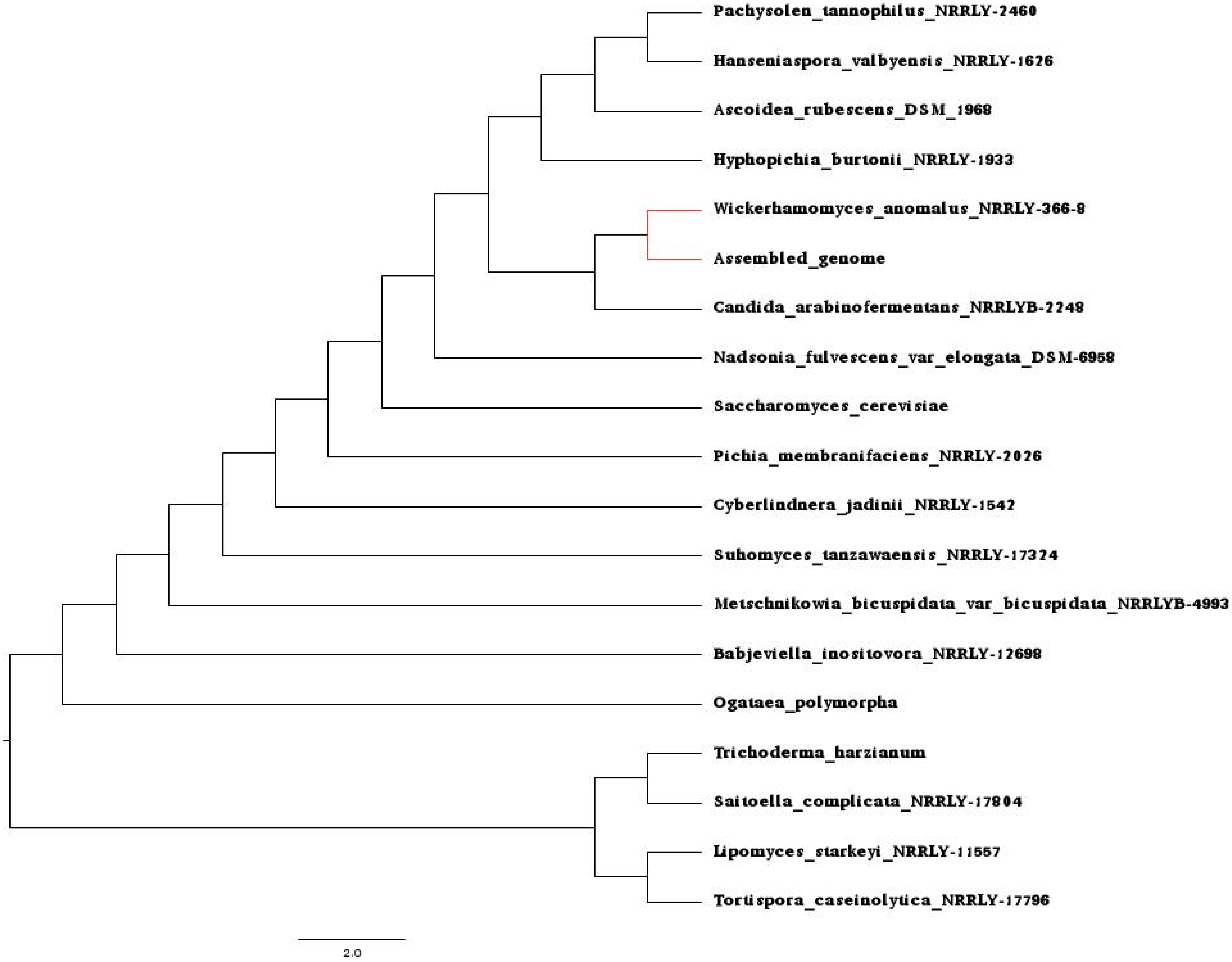
Catechol	g3300.t1	catechol 1,2-dioxygenase activity [GO:0018576]; ferric iron binding [GO:0008199]; catechol-containing compound metabolic process [GO:0009712]
	g3300.t1	Catechol dioxygenase N terminus
Nitrilase	g5205.t1	nitrilase activity [GO:0000257]; nitrogen compound metabolic process [GO:0006807]
	g6098.t1	nitrilase activity [GO:0000257]; nitrogen compound metabolic process [GO:0006807]
Phenol hydrolase	g3741.t1	Phenol hydroxylase, C-terminal dimerisation domain
	g5896.t1	Phenol hydroxylase, C-terminal dimerisation domain
	g6214.t1	Phenol hydroxylase, C-terminal dimerisation domain
Ferridoxin	K22071	FDX2; ferredoxin-2, mitochondrial
	K22071	FDX2; ferredoxin-2, mitochondrial
Hydrolase	g381.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g381.t1	haloacid dehalogenase-like hydrolase
	g4374.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g4584.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g5107.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
	g6016.t1	Putative hydrolase of sodium-potassium ATPase alpha subunit
Biofilm formation		
Cell adhesion	g4902.t1	cell adhesion [GO:0007155]; cellular response to nitrogen starvation [GO:0006995]; establishment of mitotic spindle orientation [GO:0000132]; fungal-type cell wall assembly [GO:0071940]; invasive growth in response to glucose limitation [GO:0001403]; negative regulation of translation [GO:0017148]; positive regulation of filamentous growth of a population of unicellular organisms in response to starvation [GO:1900436]; positive regulation of gluconeogenesis [GO:0045722]; positive regulation of macroautophagy [GO:0016239]; positive regulation of pseudohyphal growth [GO:2000222]; replicative cell aging [GO:0001302]; response to unfolded protein [GO:0006986]; single-species surface biofilm formation [GO:0090606]
	g5629.t1	cell adhesion involved in single-species biofilm formation [GO:0043709]; chromatin silencing [GO:0006342]; negative regulation of chromatin silencing at rDNA [GO:0061188]; negative regulation of chromatin silencing at silent mating-type cassette [GO:0061186]; negative regulation of chromatin silencing at telomere [GO:0031939]; negative regulation of transcription by RNA polymerase II

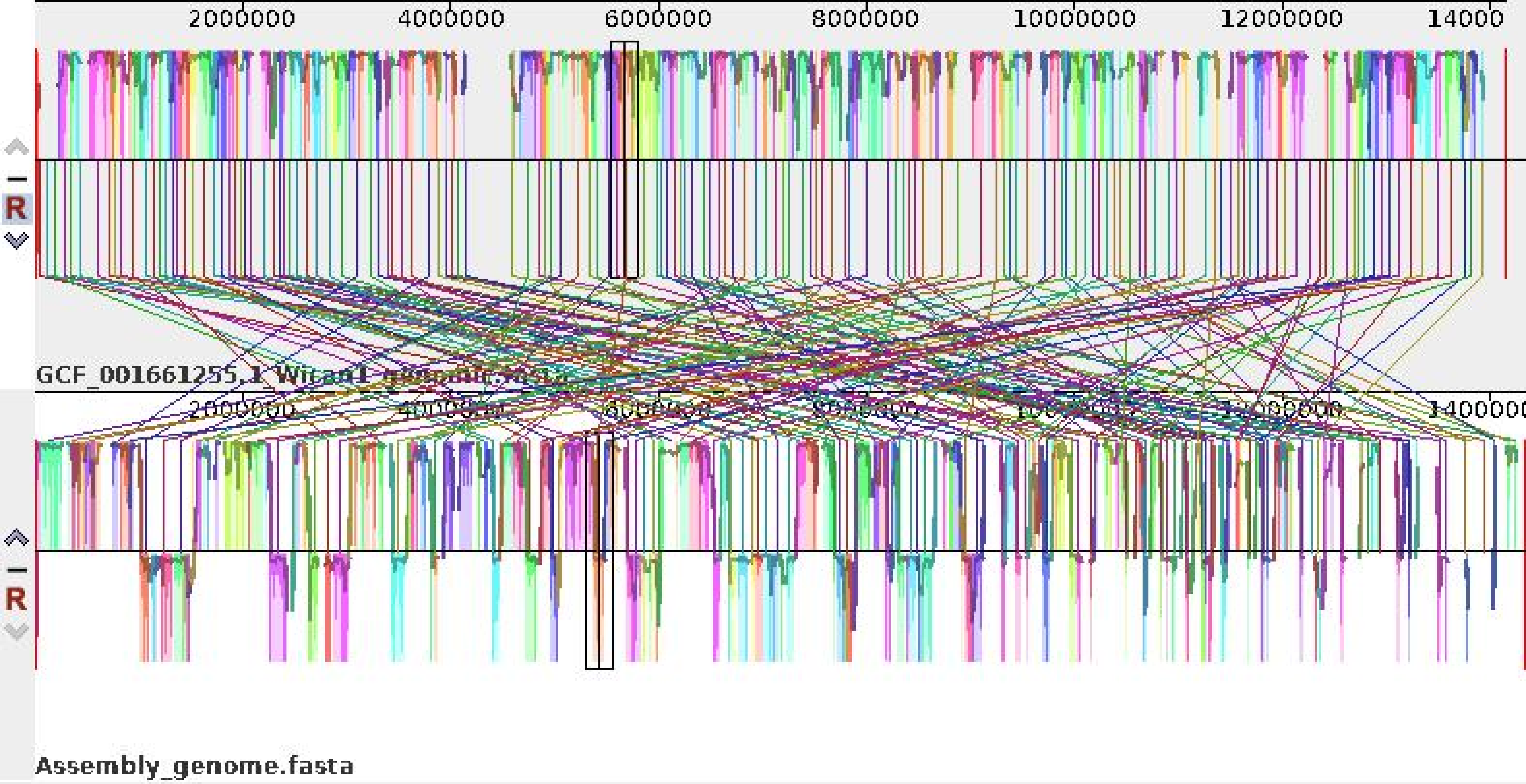
		[GO:0000122]; positive regulation of transcription from RNA polymerase II promoter in response to heat stress [GO:0061408]; regulation of invasive growth in response to glucose limitation [GO:2000217]
	g5643.t1	cell adhesion involved in single-species biofilm formation [GO:0043709]; chromatin silencing [GO:0006342]; negative regulation of chromatin silencing at rDNA [GO:0061188]; negative regulation of chromatin silencing at silent mating-type cassette [GO:0061186]; negative regulation of chromatin silencing at telomere [GO:0031939]; negative regulation of transcription by RNA polymerase II [GO:0000122]; positive regulation of transcription from RNA polymerase II promoter in response to heat stress [GO:0061408]; regulation of invasive growth in response to glucose limitation [GO:2000217]
Quorum sensing		
Homoserine		
	K17069	MET17; O-acetylhomoserine/O-acetylserine sulfhydrylase [EC:2.5.1.49 2.5.1.47]
	K17069	MET17; O-acetylhomoserine/O-acetylserine sulfhydrylase [EC:2.5.1.49 2.5.1.47]
	K17069	MET17; O-acetylhomoserine/O-acetylserine sulfhydrylase [EC:2.5.1.49 2.5.1.47]
	K17069	MET17; O-acetylhomoserine/O-acetylserine sulfhydrylase [EC:2.5.1.49 2.5.1.47]
	K00003	hom; homoserine dehydrogenase [EC:1.1.1.3]
	K00003	hom; homoserine dehydrogenase [EC:1.1.1.3]
	K00641	metX; homoserine O-acetyltransferase/O-succinyltransferase [EC:2.3.1.31 2.3.1.46]
	K00641	metX; homoserine O-acetyltransferase/O-succinyltransferase [EC:2.3.1.31 2.3.1.46]
	K00872	thrB1; homoserine kinase [EC:2.7.1.39]
Biocontrol agents and plant growth regulators related-genes		
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79

		1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K00135	gabD; succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase [EC:1.2.1.16 1.2.1.79 1.2.1.20]
	K01657	trpE; anthranilate synthase component I [EC:4.1.3.27]
	g2983.t1	Isochorismatase
	g4122.t1	Isochorismatase
	g1968.t1	3-hydroxyanthranilate 3,4-dioxygenase (EC 1.13.11.6) (3-hydroxyanthranilate oxygenase) (3-HAO) (3-hydroxyanthranilic acid dioxygenase) (HAD) (Biosynthesis of nicotinic acid protein 1)
	scf7180000000928.g1968.t1	00380 Tryptophan metabolism


Gene Ontology
zinc ion binding: 2.04%
oxidoreductase activity: 2.04%
transferase activity: 2.18%
structural constituent of ribosome: 2.56%
nucleic acid binding: 2.9%
hydrolase activity: 2.94%
metal ion binding: 3.8%
transposase activity: 3.94%
DNA binding: 11.16%
ATP binding: 11.21%
membrane: 0.38%
ATP-binding cassette (ABC) transporter complex: 0.57%
cell: 0.57%
ribosome: 2.18%
plasma membrane: 3.42%
cytoplasm: 10.87%
integral component of membrane: 17.62%
transposition, DNA-mediated: 3.94%
DNA integration: 3.18%
translation: 2.94%
transmembrane transport: 1.66%
regulation of transcription, DNA-templated: 1.52%
DNA replication: 1.19%
carbohydrate metabolic process: 1.09%
cell division: 1.04%
DNA recombination: 1.04%
transposition: 0.95%
Biological Process
Molecular Function
Cellular Component
Highcharts.com


GC_Content
GC_Skew
GC_Skew(-)
GC_Skew(+)
GCF_001661255.1_Wican1_genomic
100% identity
70% identity
50% identity
Assembly_Genome
100% identity
70% identity
50% identity
14 mbp
1 mbp
2 mbp
3 mbp
4 mbp
5 mbp
6 mbp
7 mbp
8 mbp
9 mbp
10 mbp
11 mbp
12 mbp
13 mbp
Genome_Comparison_Reference_vs_Assembly_genome
14145566 bp


Pachysolen_tannophilus_NRRLY-2460
Hanseniaspora_valbyensis_NRRLY-1626
Ascoidea_rubescens_DSM_1968
Hyphopichia_burtonii_NRRLY-1933
Wickerhamomyces_anomalus_NRRLY-366-8
Assembled_genome
Candida_arabinofermentans_NRRLYB-2248
Nadsonia_fulvescens_var_elongata_DSM-6958
Saccharomyces_cerevisiae
Pichia_membranifaciens_NRRLY-2026
Cyberlindnera_jadinii_NRRLY-1542
Suhomyces_tanzawaensis_NRRLY-17324
Metschnikowia_bicuspidata_var_bicuspidata_NRRLYB-4993
Babjeviella_inositovora_NRRLY-12698
Ogataea_polymorpha
Trichoderma_harzianum
Saitoella_complicata_NRRLY-17804
Lipomyces_starkeyi_NRRLY-11557
Tortispora_caseinolytica_NRRLY-17796
2.0


2000000
4000000
6000000
8000000
10000000
12000000
14000
R
GCF_001661255.1 Witan1
2000000
4000000
6000000
8000000
10000000
12000000
14000000
R
Assembly_genome.fasta


S
S
S
Fe^{3+}
Fe^{2+}
Protoporphyrin/ coproporphyrin ferrochelatase
Cysteine Desulfurase
assimilatory sulfur metabolism
Available Zn
ZnO
TC.ZIP (Zinc transporter)
Aldehyde dehydrogenase
Indole acetate
Indole acetaldehyde
Tryptophan Biosynthesis
Amidase
Tryptophan
Multiple antipathogen process
Antimicrobial activity
Multidrug resistance protein
Pathogen responsive gene elicitation in crops
Thaumatin (Tomato)
Acetolactate
Acetoin
Butanoate Metabolism
Diacetyl
Cysteine desulfurase
Butanediol
Acetoin
Nitrogenase Complex
ACC Deaminase, EPS, pH, Biofilm, Lytic enzymes, IAA, Volatiles, Cell Division
PO_4
Phosphonate
Glutamine
Histidine
HemH
PHO84 (Inorganic phosphate transporter)
PPA (inorganic pyrophosphatase)