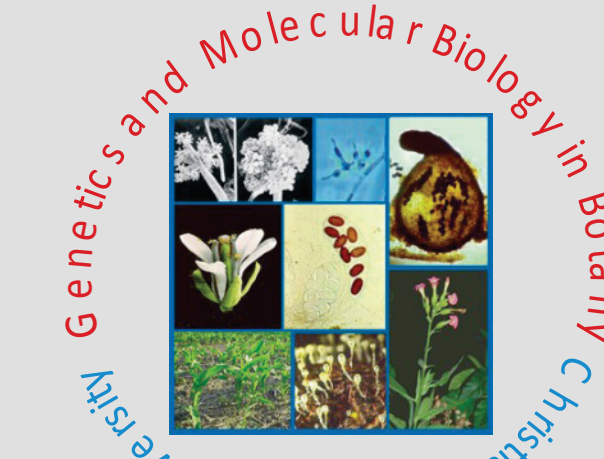


Genome sequencing, assembly and annotation of a marine fungal isolate of *Pestalotiopsis* using next generation sequencing technologies



Abhishek Kumar & Frank Kempken

Department of Genetics & Molecular Biology in Botany, Institute of Botany, Christian-Albrechts-University at Kiel, Olshausenstr. 40, D-24098 Kiel, Germany; [fkempken@bot.uni-kiel.de/akumar@bot.uni-kiel.de]



Introduction

Marine fungal isolates are characterized by the molecular diversity of their secondary metabolites [1]. The EU funded Marine fungi project aims to sequence genomes of selected marine fungal isolates, which possesses genes encoding for secondary metabolites with potential roles in cancer treatment.

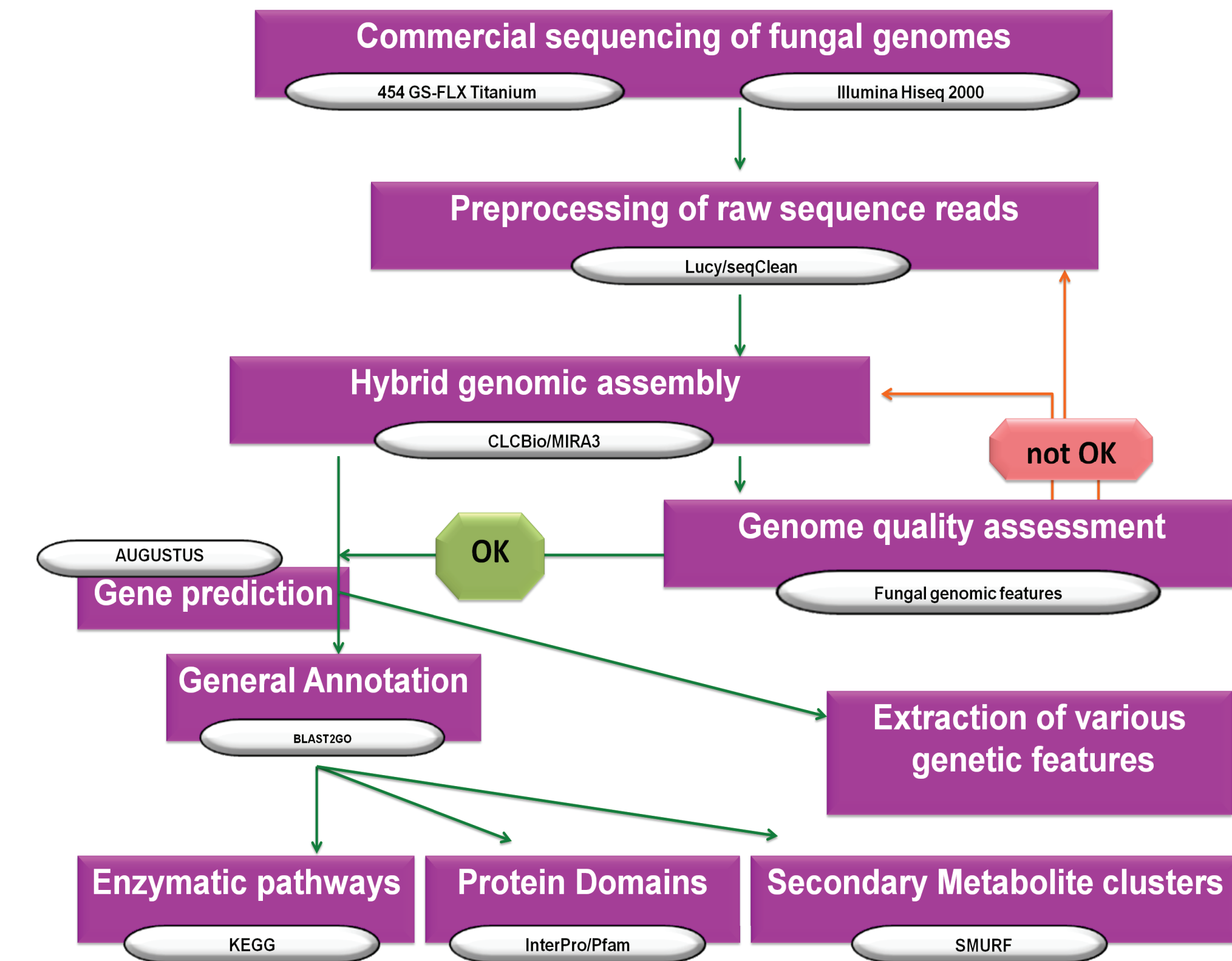
Pestalotiopsis sp. has shown to produce various secondary metabolites with bioactivities (unpublished data). We used this fungi to sequence genome to unravel genes encoding for these bioactive compound producer secondary metabolites. We have assembled and annotated the genomic sequence of this fungi using Roche 454 FLX+ and Illumina HiSeq 2000. Currently, predicted genes are presently in process of validation using Illumina based RNA-seq.



Pestalotiopsis



Methods



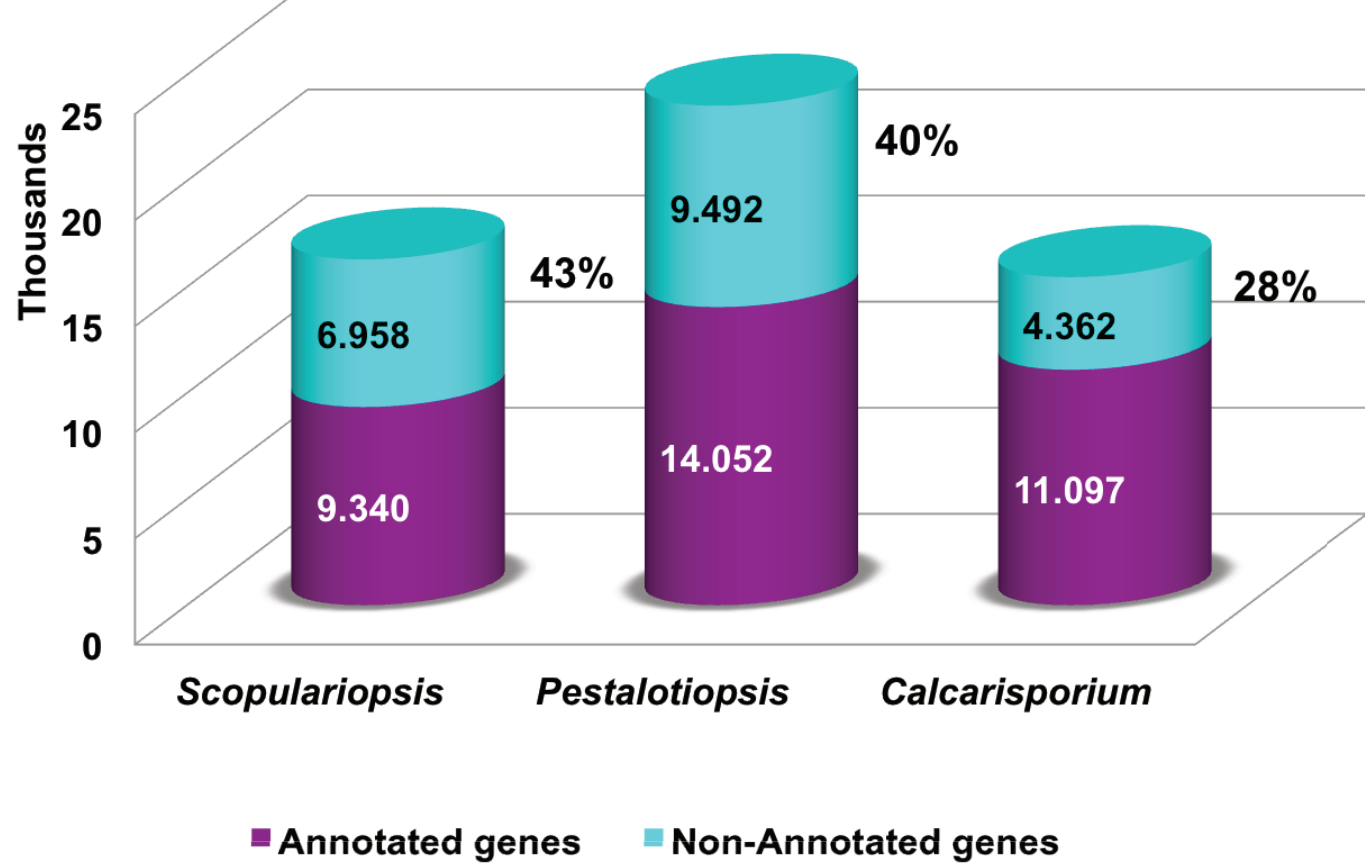
Results

1 Summary of genome assembly

Characteristics	<i>Pestalotiopsis</i> Sp.
Assembled genome size (Mb)	46
Number of contigs	4186
N50 (kb)	71.9
Largest contig (kb)	395.9
Average contig (kb)	11.15
%GC content	52.2
Number of Genes	23566
Average intron length	126.8
Average intron/gene	2.2

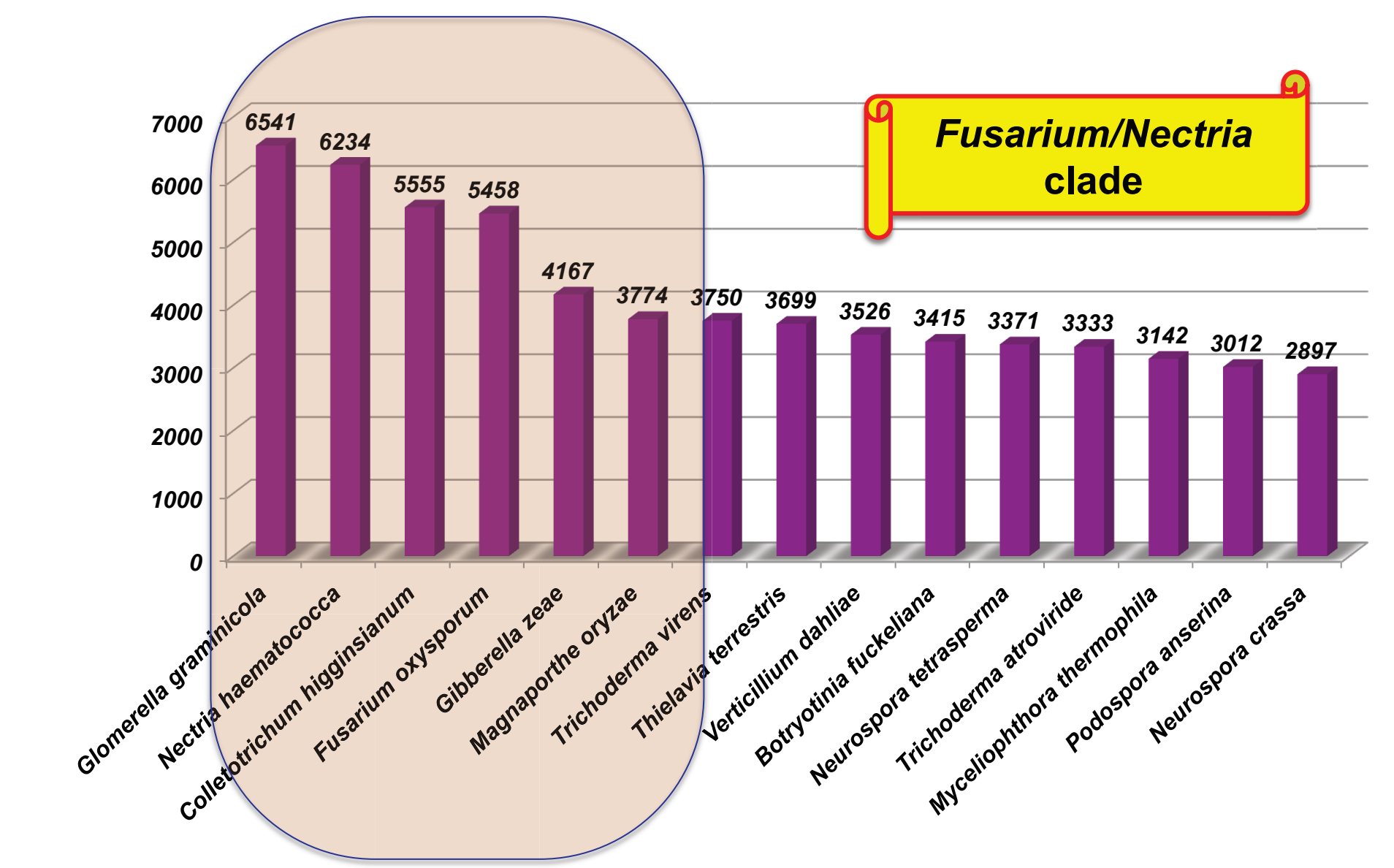
Higher number of genes hints for genome duplication events

2 Summary of genome annotation

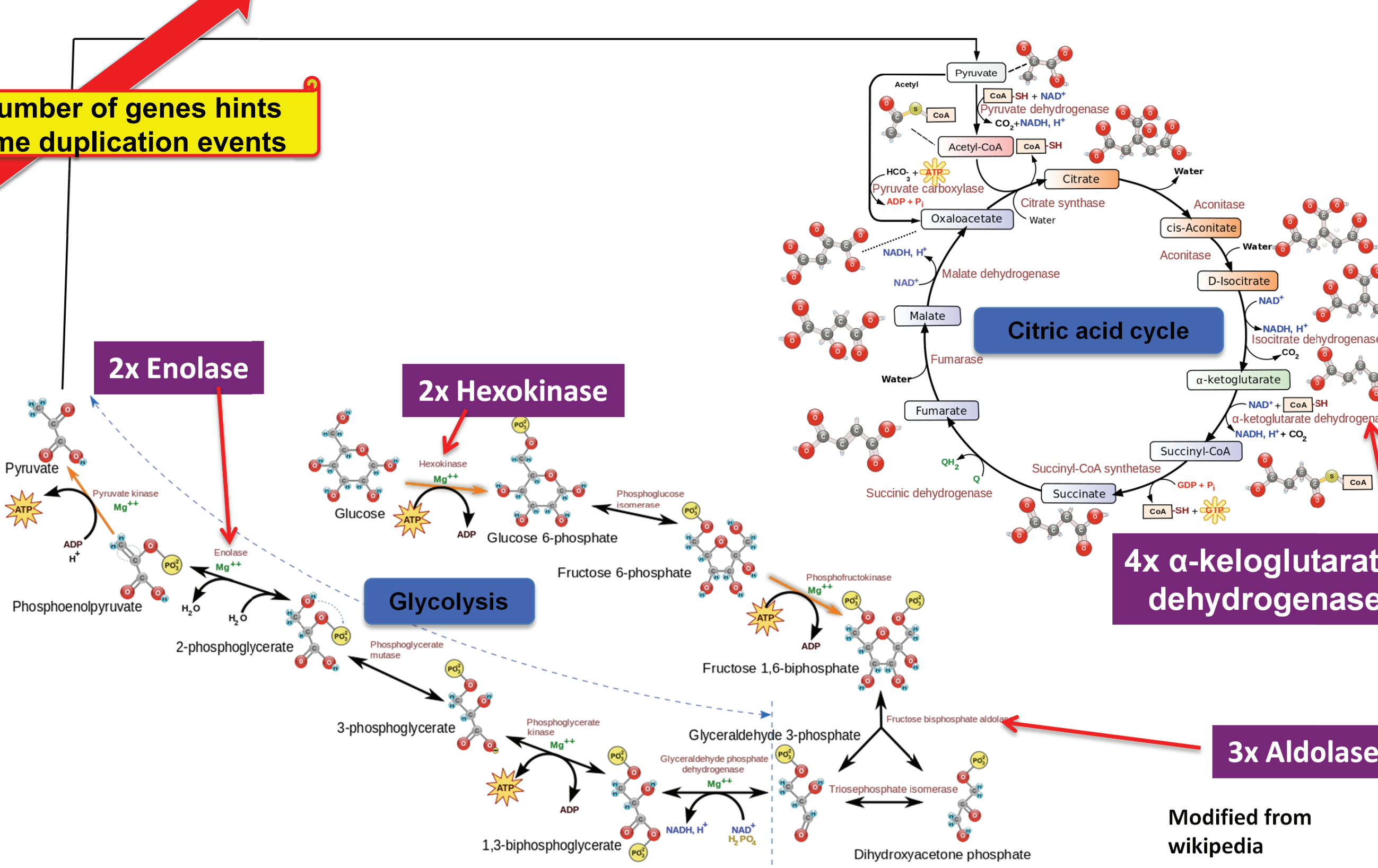


40% of genes are Non-Annotated in this genome

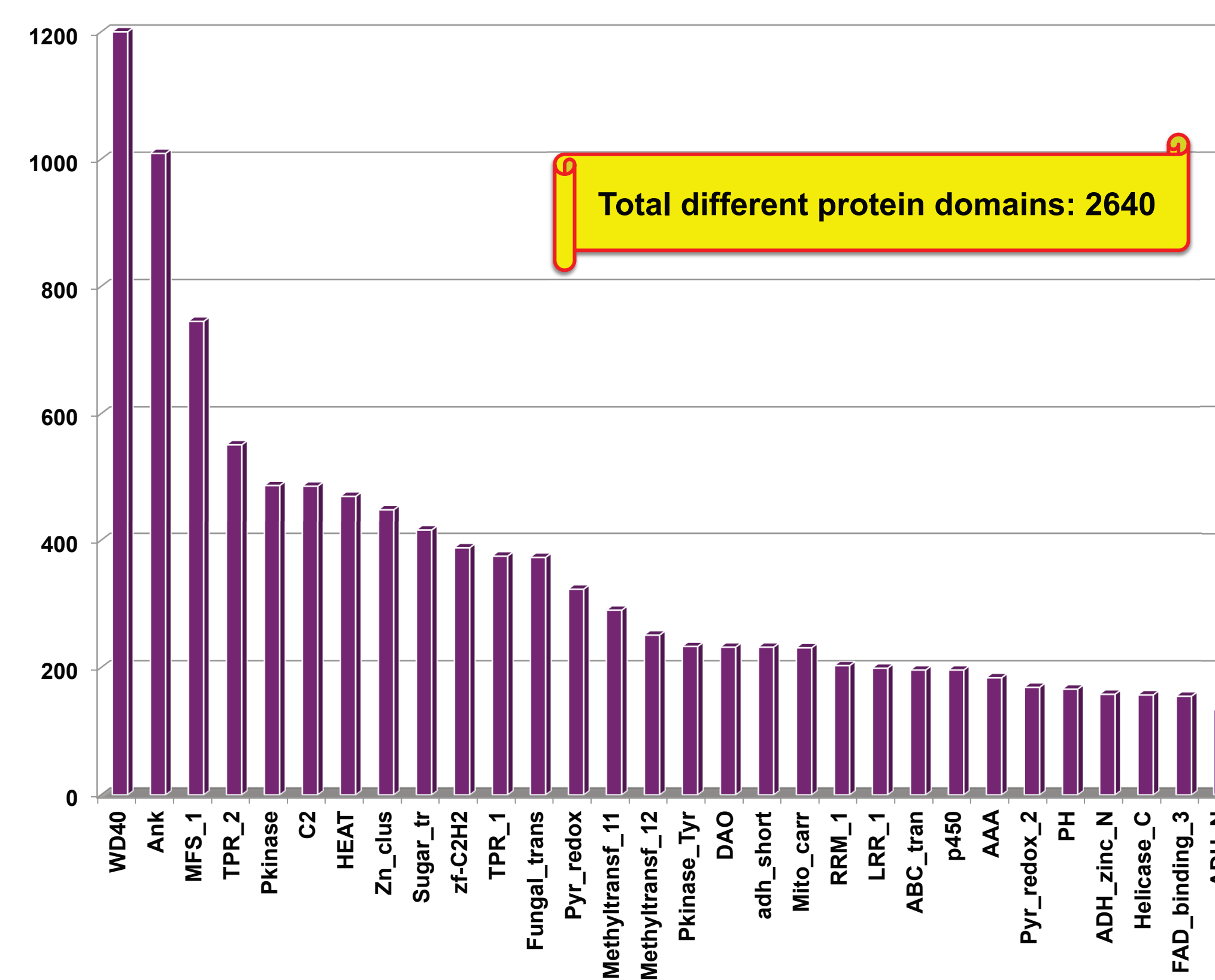
This fungus is member of sordariomycetes and is close to the clade of *Fusarium/Nectria* based on genome annotation using homolo/gly searches.



4 Partial genome duplication event evident from expansion of enzymes in glycolysis and citric acid cycle.



5 Summary of top 30 protein domains



6 Summary of secondary metabolite encoding genes

Secondary Metabolite encoding genes	#Genes
Non-ribosomal peptide synthetase (NRPS)	44
Polyketide synthase (PKS)	62
Hybrid NRPS/PKS	7
Fatty acid synthase (FAS)	3
Dimethyl-allyl-tryptophan synthase (DMATS)	5
Sesquiterpene cyclase (SesCyc)	1

Expansion

7 Modules of selected polyketide synthase (PKS)

Sequence name	Gene	KS	AT	ACP	ER	DH
contig_1034.g19915	PKS1	KS	AT	ACP	ER	DH
contig_298.g8813	PKS2	KS	AT	ACP	ER	DH
contig_120.g3509	PKS3	KS	AT	ACP	ER	DH
contig_739.g16724	PKS4	KS	AT	ACP	ER	DH
contig_571.g14790	PKS5	KS	AT	ACP	ER	DH
contig_38.g756	PKS6	KS	AT	ACP	ER	DH
contig_189.g5694	PKS7	KS	AT	ACP	ER	DH
contig_127.g3751	PKS8	KS	AT	ACP	ER	DH
contig_199.g6060	PKS9	KS	AT	ACP	ER	DH
contig_76.g1793	PKS10	KS	AT	ACP	ER	DH
contig_90.g2200	PKS11	KS	AT	ACP	ER	DH
contig_316.g9427	PKS12	KS	AT	ACP	ER	DH
contig_398.g11547	PKS13	KS	AT	ACP	ER	DH
contig_113.g3315	PKS14	KS	AT	ACP	ER	DH
contig_875.g18346	PKS15	KS	AT	ACP	ER	DH
contig_826.g17681	PKS16	KS	AT	ACP	ER	DH
contig_821.g17615	PKS17	KS	AT	ACP	ER	DH
contig_374.g11232	PKS18	KS	AT	ACP	ER	DH
contig_198.g6001	PKS19	KS	AT	ACP	ER	DH
contig_25.g482	PKS20	KS	AT	ACP	ER	DH
contig_125.g3668	PKS21	KS	AT	ACP	ER	DH
contig_1097.g20289	PKS22	KS	AT	ACP	ER	DH
contig_1227.g20816	PKS23	AT	DH	ER	KS	ACP
contig_93.g19005	PKS24	AT	DH	ER	KS	ACP
contig_608.g15143	PKS25	KS	AT	ACP	ER	DH
contig_177.g5224	PKS26	ER	KS	ACP	DH	AT
contig_2025.g22303	PKS27	ER	KS	ACP	DH	AT
contig_2052.g22326	PKS28	KS	AT	ACP	ER	DH
contig_3138.g23048	PKS29	AT	DH	ER	KS	ACP
contig_436.g12568	PKS30	ER	KS	ACP	DH	AT
contig_526.g14172	PKS31	ACP	AT	DH	ER	KS
contig_88.g2159	PKS32	ACP	AT	DH	ER	KS
contig_817.g17598	PKS33	ACP	AT	DH	ER	KS
contig_3560.g23291	PKS34	ACP	AT	DH	ER	KS
contig_330.g10048	PKS35	KS	AT	ACP	ER	DH
contig_208.g6183	PKS36	KS	AT	ACP	ER	DH
contig_169.g5018	PKS37	KS	AT	ACP	ER	DH

KS - ketosynthase, AT - acyl transferase, ACP - acyl carrier protein, ER - β-ketoreductase, DH - enoyl reductase, DH - dehydrogenase

8 Summary of repeats in this genome

Repeat type	Number of elements*	Length occupied (bp)	Percentage of sequence
Retroelements	254	103427	0.22
LINES	6	299	0.00
LTR elements	248	103128	0.22
Ty1/Copia	27	2681	0.01
Gypsy/DIRS1	221	100447	0.22
DNA transposons	33	2302	0.00
hobo-Activator	3	254	0.00
Tc1-IS630-Pogo	2	361	0.00
Tourist/Harbinger	6	367	0.00
Unclassified	1	127	0.00
Total interspersed repeats		105856	0.23
Small RNA	59	11106	0.02
Simple repeats	7520	302527	0.65
Low complexity	723	33149	0.07

0.97% of total basepairs are repeats

Discussion

1. We sequenced genome of *Pestalotiopsis* sp. from the marine environment.
2. The estimated genome size is ~46 Mb with 23492 genes and it possesses repeats as ~1% of genome size. Higher number of genes are due to partial genome duplication events.
3. This fungus is a member of sardariomycetes, close to clade of *Fusarium/Nectria*.
4. There is an expansion of secondary metabolites encoding genes in comparison to other known fungi.

Conclusion

We have sequenced the genome of *Pestalotiopsis* sp. from the marine environment with an estimated genome size of ~46 Mb (23492 genes) using two different DNA sequencing methods. This laid platform for various genetic studies using *Pestalotiopsis* sp.

References

1. König et al. (2006) Chembiochem 7(2):229-38.

