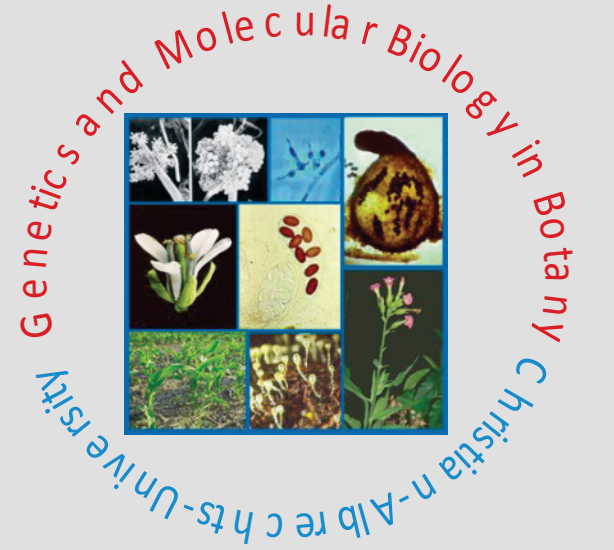


Genome sequencing, assembly and annotation of a marine fungal isolate of *Calcarisporium* using different next generation sequencing technologies



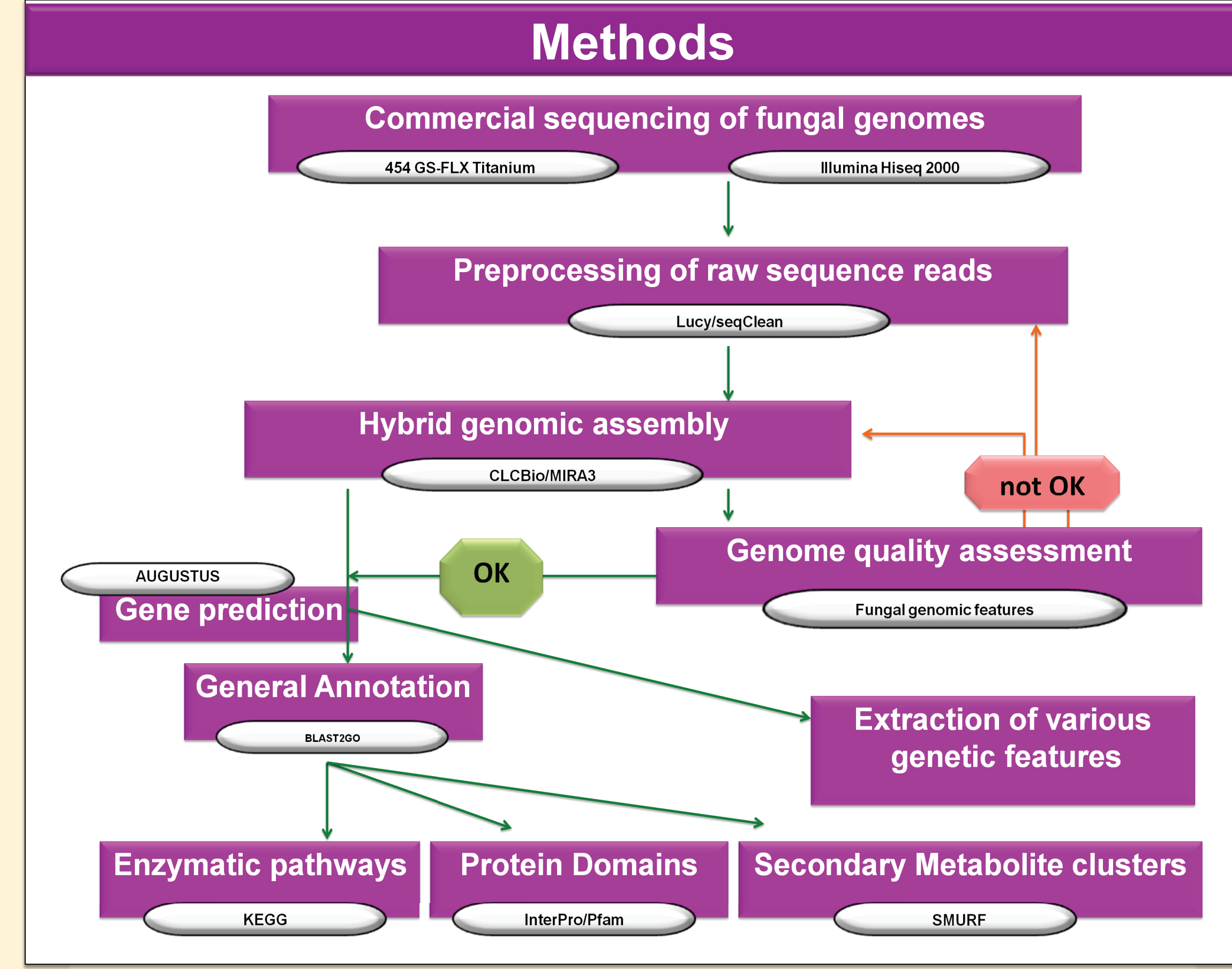
Abhishek Kumar & Frank Kempken

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 possess 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. *Calcarisporium* 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 has shown to produce various secondary metabolites with bioactivities (unpublished data).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. Now predicted genes are presently in process of validation using Illumina based RNA-seq.

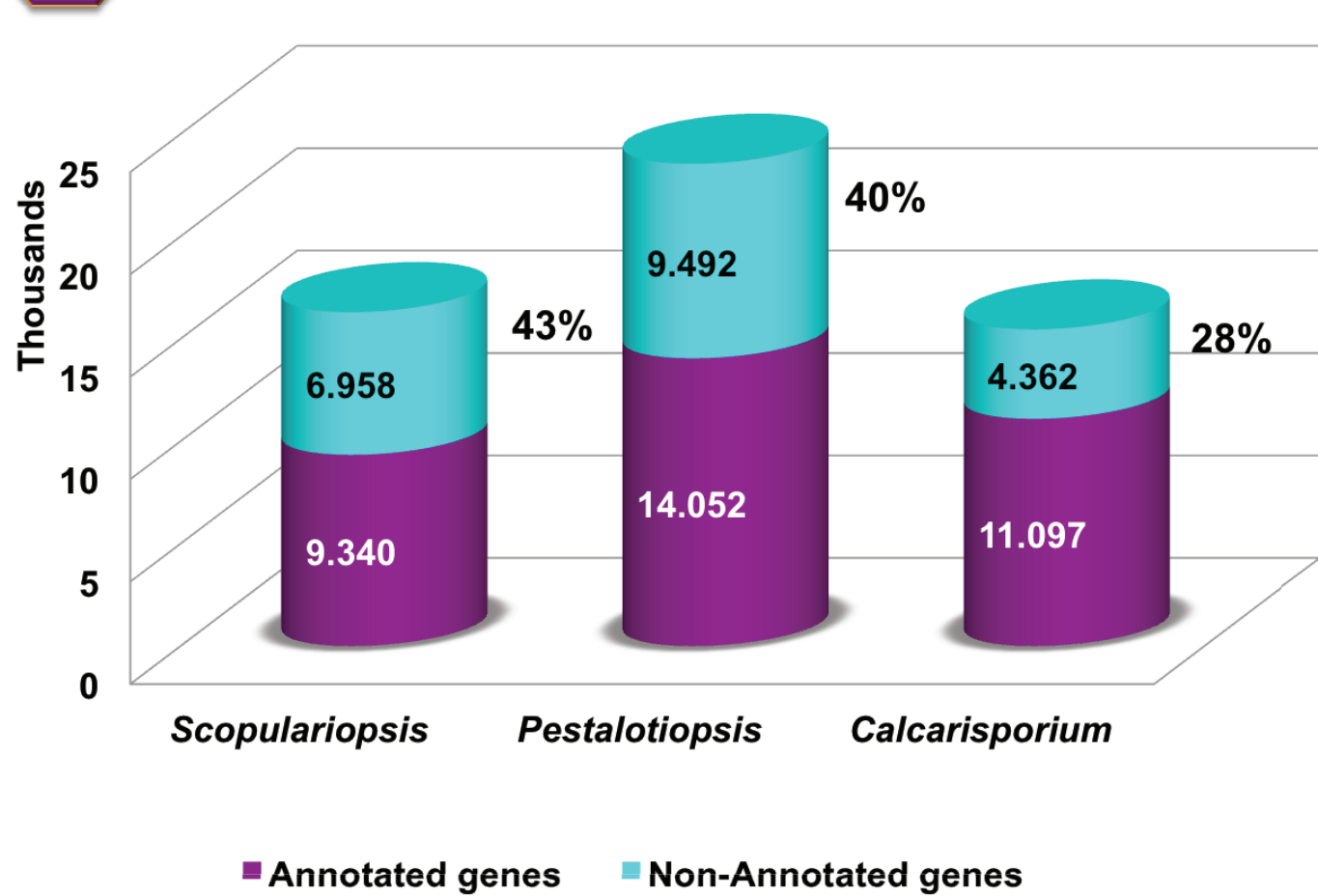


Results

1 Summary of genome assembly

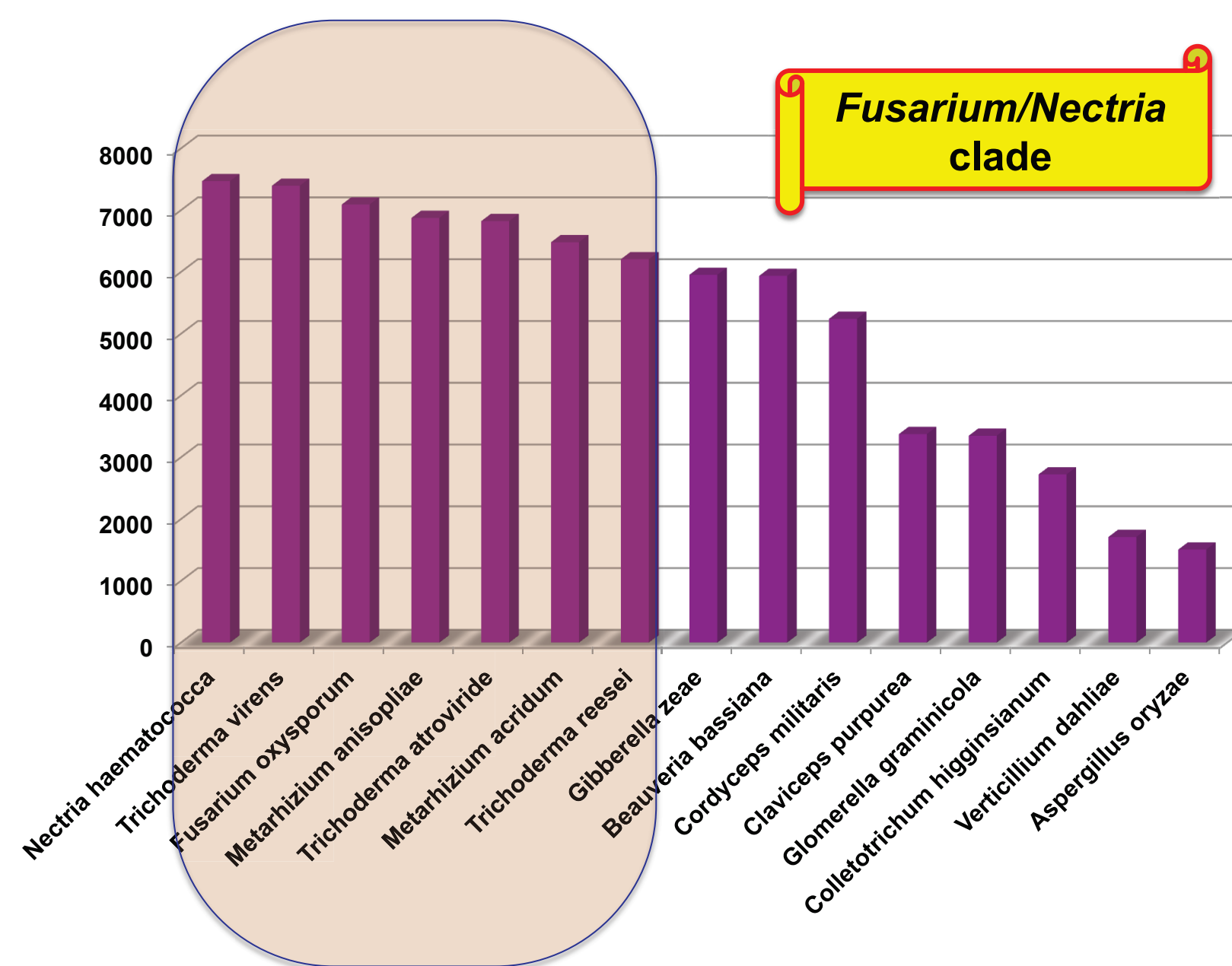
Characteristics	<i>Calcarisporium</i> Sp.
Assembled genome size (Mb)	35
Number of contigs	2464
N50 (kb)	91.9
Largest contig (kb)	420.22
Average contig (kb)	14.62
%GC content	50.7
Number of Genes	15455
Average intron length	120.7
Average intron/gene	2.1

2 Summary of genome annotation

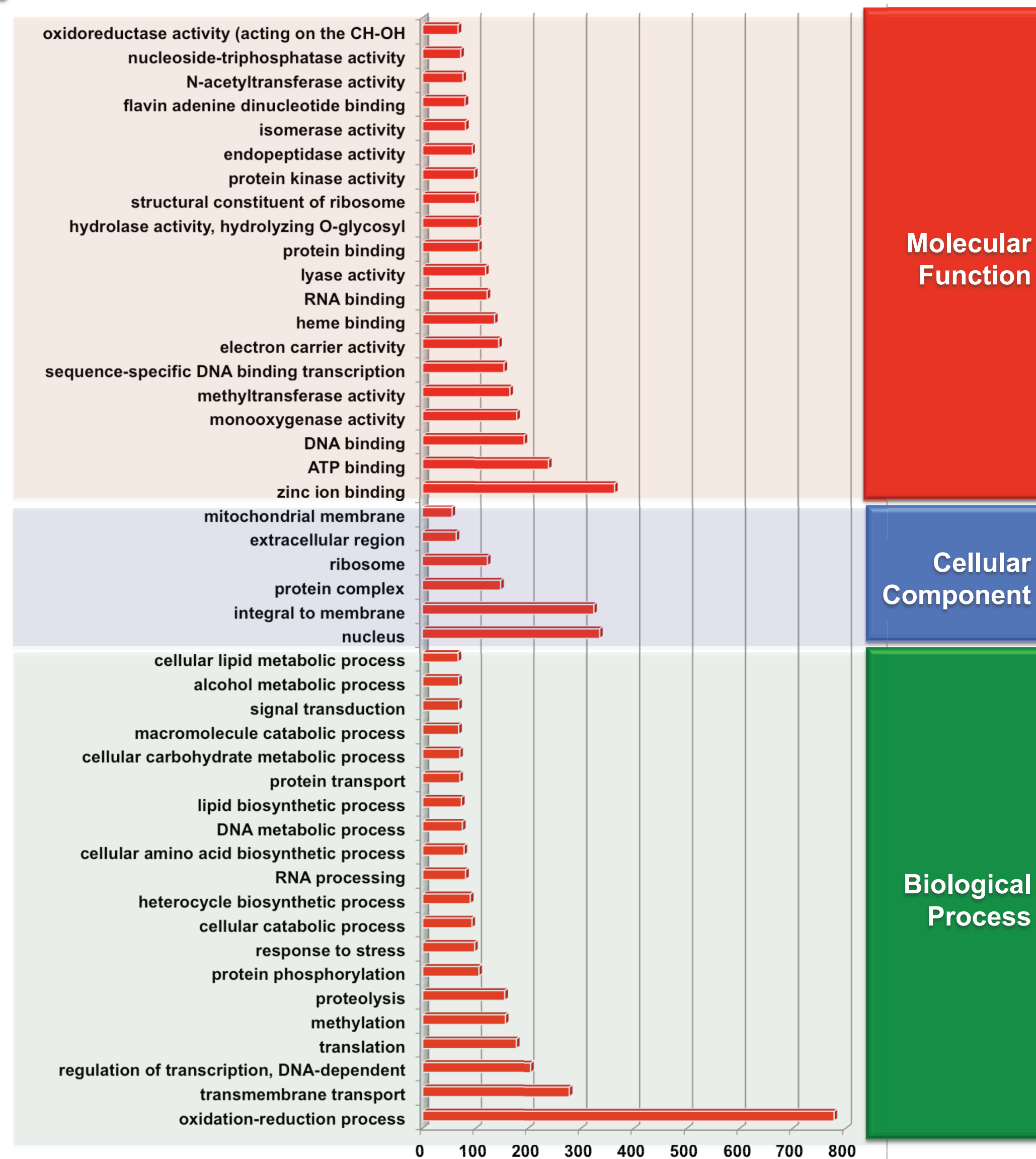


28% of genes are Non-Annotated in this genome

This fungus are members of sordariomycetes and are close to the clade of *Fusarium/Nectria* based on genome annotation using homology searches.



4 Top GO-terms assigned to encoded genes in three different categories.



6 Summary of secondary metabolite encoding genes

Secondary Metabolite encoding genes	#Genes
Non-ribosomal peptide synthetase (NRPS)	52
Polyketide synthase (PKS)	66
Hybrid NRPS/PKS	7
Fatty acid synthase (FAS)	1
Dimethyl-allyl-tryptophan synthase (DMATS)	0
Sesquiterpene cyclase (SesCyc)	1

Expansion

7 Modules of selected polyketide synthase (PKS)

Sequence name	Gene	KS	AT	DH	ER	KR	ACP
contig_100.g2642	PKS1	KS	AT	DH	ER	KR	ACP
contig_296.g9341	PKS2	KS	AT	DH	ER	KR	ACP
contig_286.g8897	PKS3	KS	AT	DH	ER	KR	ACP
contig_367.g10065	PKS4	KS	AT	DH	ER	KR	ACP
contig_102.g2810	PKS5	KS	AT	DH	ER	KR	ACP
contig_461.g13594	PKS6	KS	AT	DH	ER	KR	ACP
contig_280.g8728	PKS7	KS	AT	DH	ER	KR	ACP
contig_201.g6947	PKS8	KS	AT	DH	ER	KR	ACP
contig_233.g7691	PKS9	KS	AT	DH	ER	KR	ACP
contig_427.g11119	PKS10	KS	AT	DH	ER	KR	ACP
contig_346.g9813	PKS11	KS	AT	DH	ER	KR	ACP
contig_413.g10757	PKS12	KS	AT	DH	ER	KR	ACP
contig_101.g4694	PKS13	KS	AT	DH	ER	KR	ACP
contig_358.g9930	PKS14	KS	AT	DH	ER	KR	ACP
contig_467.g11618	PKS15	KS	AT	DH	ER	KR	ACP
contig_85.g3018	PKS16	KS	AT	DH	ER	KR	ACP
contig_401.g4694	PKS17	KS	AT	DH	ER	KR	ACP
contig_365.g14604	PKS18	KS	AT	DH	ER	KR	ACP
contig_27.g633	PKS19	KS	AT	DH	ER	KR	ACP
contig_231.g4821	PKS20	KS	AT	DH	ER	KR	ACP
contig_343.g2085	PKS21	KS	AT	DH	ER	KR	ACP
contig_380.g10325	PKS22	ER	AT	DH	ER	KR	ACP
contig_459.g11546	PKS23	KS	AT	DH	ER	KR	ACP
contig_338.g14524	PKS24	ER	AT	DH	ER	KR	ACP
contig_380.g10326	PKS25	KS	AT	DH	ER	KR	ACP
contig_233.g7693	PKS26	KR	AT	DH	ER	KR	ACP
contig_282.g8782	PKS27	ER	AT	DH	ER	KR	ACP
contig_740.g13966	PKS28	ACP	AT	DH	ER	KR	ACP
contig_607.g13201	PKS29	ACP	AT	DH	ER	KR	ACP
contig_109.g5926	PKS30	KS	AT	DH	ER	KR	ACP
contig_73.g2409	PKS31	KS	AT	DH	ER	KR	ACP
contig_422.g10998	PKS32	KS	AT	DH	ER	KR	ACP
contig_59.g1824	PKS33	KS	AT	DH	ER	KR	ACP
contig_39.g1187	PKS34	KS	AT	DH	ER	KR	ACP
contig_274.g8653	PKS35	KS	AT	DH	ER	KR	ACP
contig_414.g10861	PKS36	KS	AT	DH	ER	KR	ACP
contig_82.g2859	PKS37	KS	AT	DH	ER	KR	ACP
contig_73.g2414	PKS38	KS	AT	DH	ER	KR	ACP
contig_282.g8784	PKS39	KS	AT	DH	ER	KR	ACP
contig_189.g6530	PKS40	KS	AT	DH	ER	KR	ACP
contig_205.g6974	PKS41	KS	AT	DH	ER	KR	ACP
contig_272.g8530	PKS42	KS	AT	DH	ER	KR	ACP
contig_130.g4803	PKS43	KS	AT	DH	ER	KR	ACP
contig_907.g14405	PKS44	KS	AT	DH	ER	KR	ACP
contig_116.g4191	PKS45	KS	AT	DH	ER	KR	ACP

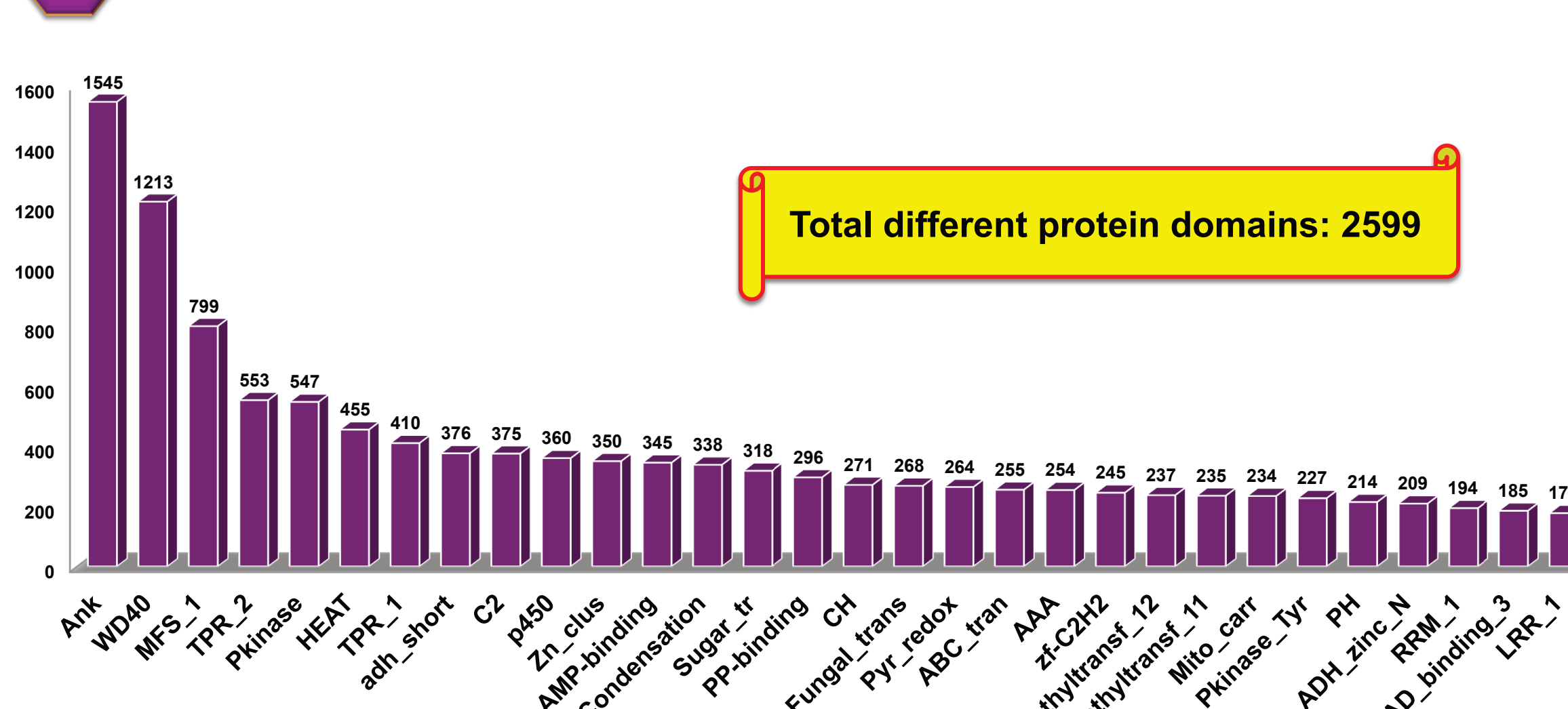
KS - ketosynthase, AT - acyl transferase, ACP - acyl carrier protein, KR - β-ketoreductase, ER - enoyl reductase, DH - dehydrogenase

8 Summary of repeats in this genome

Repeat type	Number of elements*	Length occupied (bp)	Percentage of sequence
Retroelements	170	69031	0.19
LINES:	35	26402	0.07
R1/LOA/Jockey	3	288	0.00
LTR elements:	135	42629	0.12
Ty1/Copia	52	10915	0.03
Gypsy/DIRS1	83	31714	0.09
DNA transposons	56	16380	0.05
hobo-Activator	4	236	0.00
Tc1-IS630-Pogo	31	10720	0.03
PiggyBac	1	77	0.00
Tourist/Harbinger	4	273	0.00
Unclassified:	2	187	0.00
Total interspersed repeats:		85598	0.24
Small RNA:	25	8425	0.02
Simple repeats:	7665	320027	0.89
Low complexity:	1042	52138	0.14

1.28% of total basepairs are repeats

5 Summary of top 30 protein domains



Total different protein domains: 2599

Discussion

1. We sequenced genome of *Calcarisporium* sp. from the marine environment.
2. The estimated genome size is ~35 Mb with 15455 genes and it possesses repeats as 1.28% of genome size.
3. This fungus is a member of sordariomycetes, close to clade of *Fusarium/Nectria*.
4. There is an expansion of secondary metabolites encoding genes in comparison to other fungal genomes.

Conclusion

We have sequenced the genome of *Calcarisporium* sp. from the marine environment with an estimated genome size of ~35 Mb (15455 genes) using two different DNA sequencing methods. This laid platform for various genetic studies using *Calcarisporium* sp.

References

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

