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

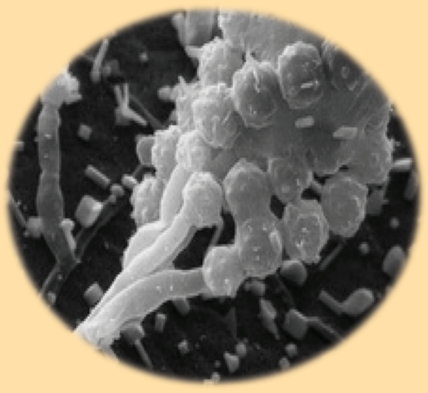
Abhishek Kumar & Frank Kempken

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

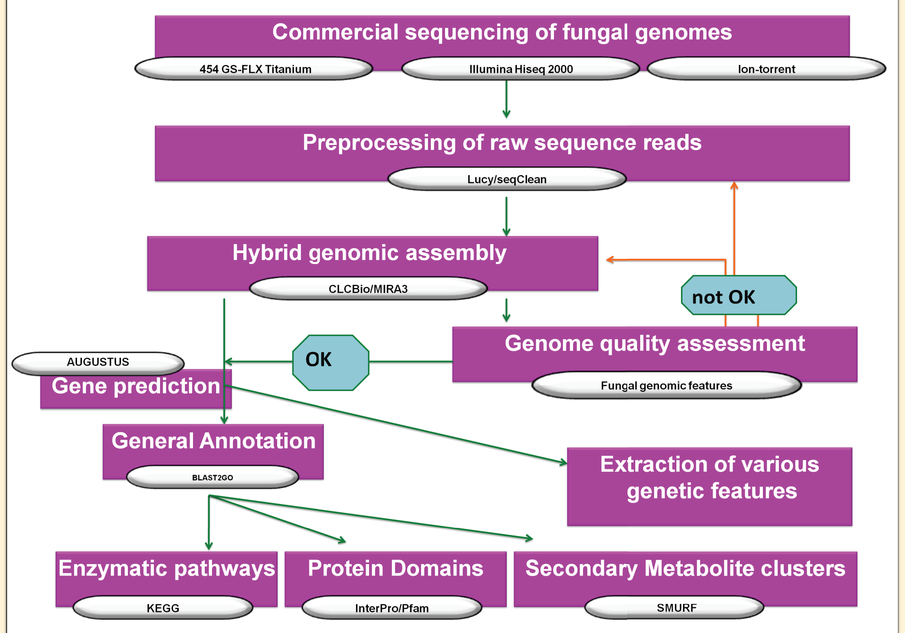
Enormous biodiversity of marine fungal isolates is mirrored by the molecular diversity of their secondary metabolites [1]. Over 100 terrestrial fungal genomes using Sanger method have been sequenced, and recently the *Sordaria macrospora* genomic sequences became available using next-generation sequencing [2]. In the EU funded Marine fungi project, we aim to sequence genomes of selected marine isolates of fungi, which possess genes encoding for secondary metabolites with potential roles in cancer treatment. *Scopulariopsis brevicaulis* is known to produce the cyclic peptides Scopularide A and B [3] and we used marine isolate of this fungi to sequence genome. We have established the genomic sequence of this fungi using three different next-generation sequencing methods (Roche 454, Illumina and ion-torrent) and predicted genes are presently in process of validation using Illumina based RNA-seq.

MARINE FUNGI



Scopulariopsis brevicalis

Methods

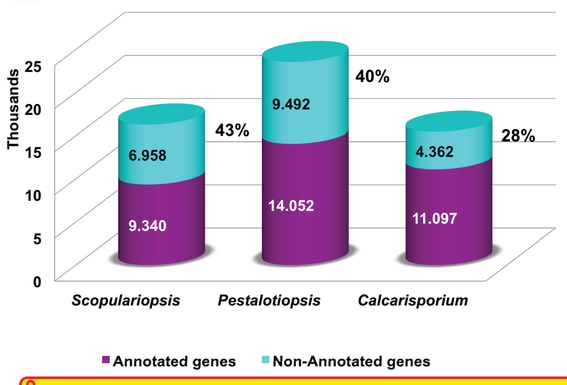


Results

1 Summary of genome assembly

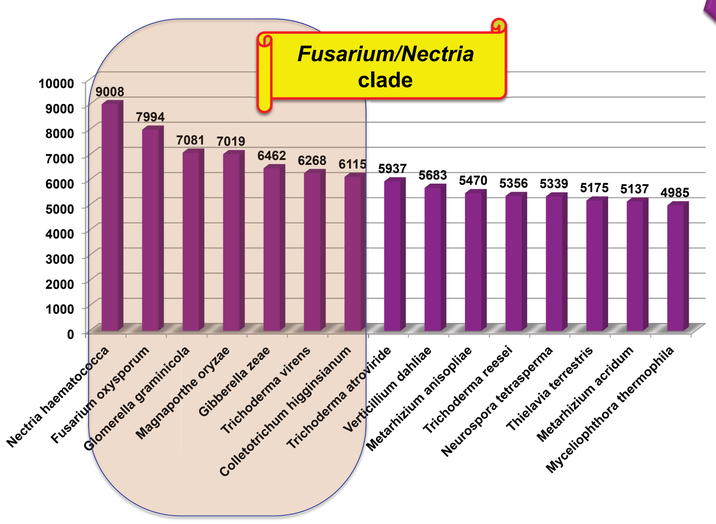
Characteristics	*S. brevicaulis*
Assembled genome size (Mb)	32.2
Number of contigs	935
N50 (kb)	88
Largest contig (kb)	342
Average contig (kb)	34.44
%GC content	54.5
Number of Genes	16298
Average intron length	129.4
Average intron/gene	3.09

2 Summary of genome annotation



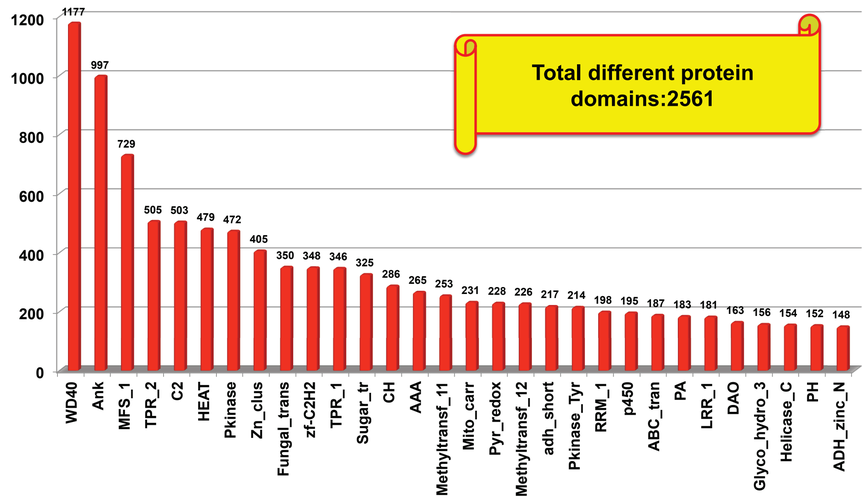
43% of genes are Non-Annotated in this genome

3 This fungus is close to the clade of *Fusarium/ Nectria* (sordariomycetes) based on genome annotation using homology searches.



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

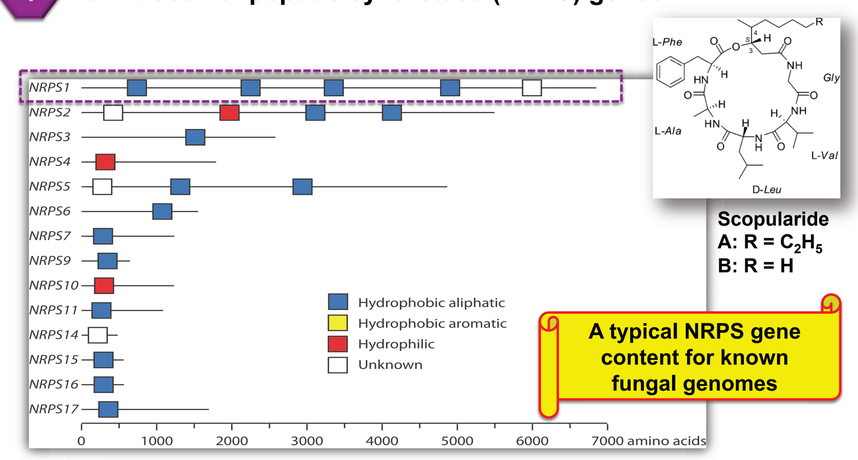
5 Summary of top protein domains in this genome



6 Summary of secondary metabolite encoding genes

Secondary metabolite encoding genes	*Number*
Non-ribosomal peptide synthetase (NRPS)	17
Polyketide synthase (PKS)	18
Hybrid NRPS/PKS	1
Fatty acid synthase (FAS)	2
Dimethyl-allyl-tryptophan synthase (DMATS)	0
Sesquiterpene cyclase (SesCyc)	0

7 Nonribosomal peptide synthetase (NRPS) genes



Scopularide
A: R = C_2H_5
B: R = H

A typical NRPS gene content for known fungal genomes

8 Summary of repeats in this genome

Repeat type	Number of elements*	Length occupied (bp)	Percentage of sequence
Retroelements	71	56092	0.18
Penelope	1	127	0.00
LINEs	11	4221	0.01
CRE/SLACS	3	674	0.00
LTR elements	60	51871	0.16
Ty1/Copia	10	659	0.00
Gypsy/DIRS1	50	51212	0.16
DNA transposons	81	53820	0.17
Tc1-IS630-Pogo	68	50313	0.16
Tourist/Harbinger	1	44	0.00
Other (Mirage, P-element,Transib)	1	87	0.00
Total interspersed repeats		109912	0.35
Small RNA	55	11897	0.04
Satellites	1	195bp	0.00
Simple repeats	5151	235633	0.75
Low complexity	1148	64645	0.20

1.33% of total basepairs are repeats

Discussion

1.By using NGS methods, we sequenced genome of *S. brevicaulis* from the marine environment.

2. The estimated genome size is ~32 Mb using three different next-generation sequencing technologies with 16298 genes and 1.33% of genome size is repeats.

3. This fungus is a member of sardariomycetes, close to clade of Fusarium/Nectria.

4. NRPS genes repository is also in agreement with other fungal genomes.

Conclusion

We have sequenced the genome of *S. brevicaulis* from the marine environment with an estimated genome size of ~32 Mb (16298 genes) using three different DNA sequencing methods. This laid platform for various genetic studies using *S. brevicaulis*.

References

1. König et al. (2006) Chembiochem 7(2):229-38.
2. Nowrousian *et al.* (2010) PLoS Genet 6(4): e1000891.
3. Zhiguo *et al.* (2008) *Journal of Natural Products 71* (6), 1052-1054

http://www.uni-kiel.de/Botanik/Kempken/fbkem.shtml