

Genome sequence analysis of *Chrysosporium lucknowense* C1: a filamentous fungus of biotechnological importance

Hans Visser^a, Sandra Hinz^a, Martijn Koetsier^a, Vivi Joosten^a, Scooter Willis^b, Bruce Pascal^b and Jan Wery^a

^aDyadic Netherlands, Nieuwe Kanaal 7-S, 6709 PA Wageningen, The Netherlands. ^bThe Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458, USA

Introduction

Dyadic owns and develops the fungus *Chrysosporium lucknowense* C1 as a platform for the hyper production of a broad variety of enzymes. Cellulases are the main class of extracellular enzymes secreted by C1 production strains. Here, we describe preliminary results of the sequencing and analysis of the C1 genome.

Genome sequencing of C1

The genome of **C1 wild type strain** VKM F-3500-D was initially sequenced in 2005 by Sanger-sequencing at a 6.3x coverage. In order to improve the genome sequence data, C1 was re-sequenced recently at a 12.6x coverage by a paired-end approach using the 454 pyro-sequencing technology. Both datasets were assembled into the new **C1 draft genome**. The genome size was approximately **38.5 Mbp** of which 37.2 Mbp were assigned to the 10 largest scaffolds. The overall GC content was 48.6%, whereas protein coding regions consisted of 61.9% GC.

Automated gene prediction and annotation

The assembled genome sequence was assessed for protein encoding genes. Three different **gene prediction algorithms** were applied: GeneMark-ES (1), Geneid (2) and GlimmerHMM (3). Most putative genes were predicted by all 3 algorithms simultaneously, indicating a high confidence to those sequences actually being genes. Some clear differences were also observed (Table 1).

Table 1. Preliminary data on C1 gene predictions.

Feature	Genemark	Geneid	Glimmer
Total predicted genes	9807	16012	8545
Uniprot hits	3151	4234	3831
TrEMBL hits	5007	3813	2856
Average gene length ¹ (bp)	2223	2271	2313
Average number of exons per gene ¹	3.47	3.66	3.70
Average exon length ¹ (bp)	549	499	503
Average number of introns per gene ¹	2.79	3.10	2.91
Average number of introns per gene (incl. 0) ¹	2.47	2.66	2.70
Average intron length ¹ (bp)	130	169	169
Average protein length ¹ (aa)	636	609	621

¹, number of genes sampled: GeneMark 3151, Geneid 4234, Glimmer 3831. These correspond to predicted genes having a blast hit score of 1E-10 (or less) in all uniprot annotated fungal proteins.

Taken together, **9499 unique genes** were identified. In order to appoint putative gene function, all predicted genes were blasted against the UniProt, TrEMBL and Pfam databases. This was referred to as the **automated annotation** of the C1 genome. About 100 previously manually annotated genes were used to randomly monitor the correctness of the genome assembly, gene calling and automated annotation processes. A C1 genome database was subsequently built using the **Gmod / Gbrowse** software (4) (Figure 1).

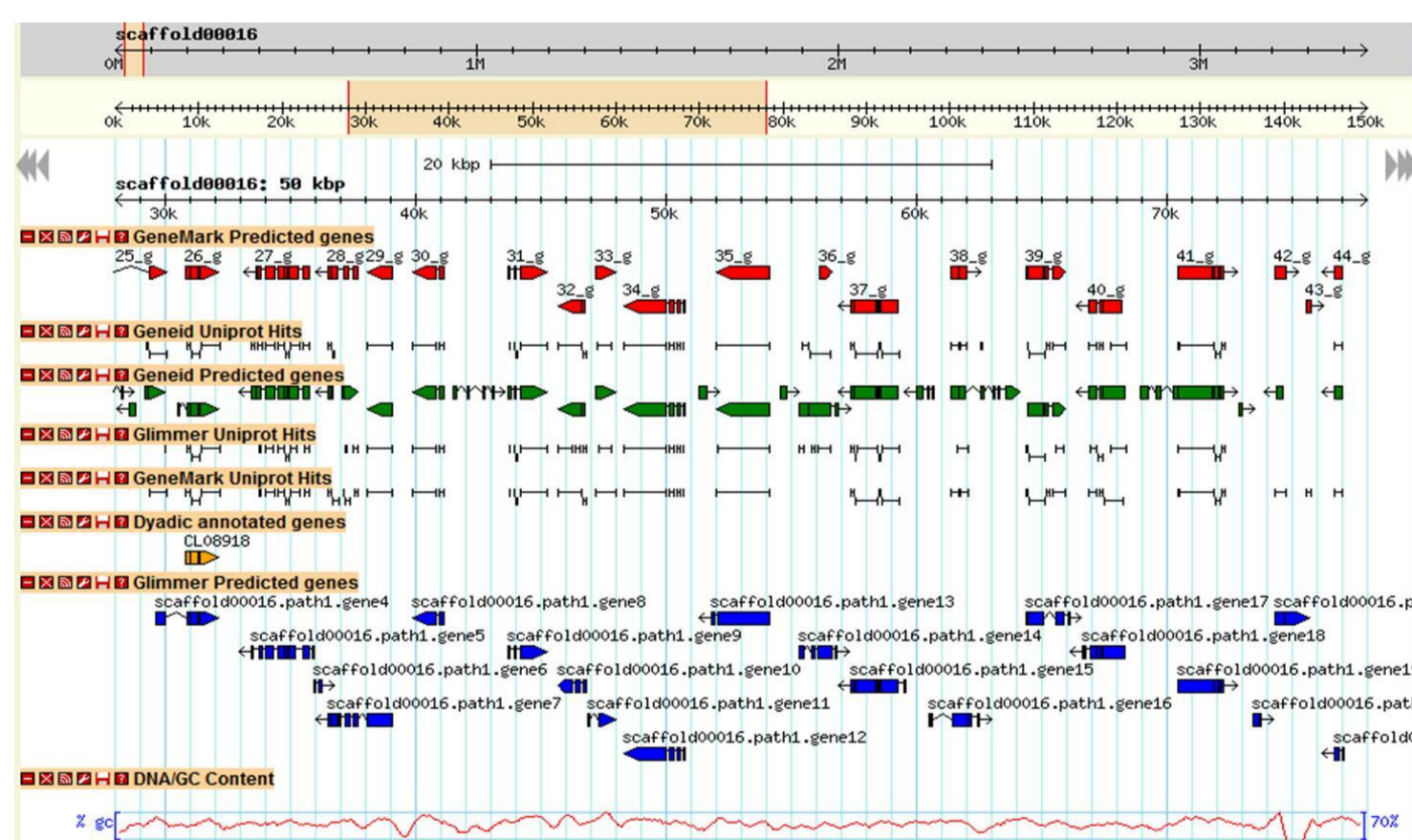


Figure 1. Example of a Gbrowse track view of the C1 genome database.

Genome mining for industrial enzymes

The C1 genome was mined for genes encoding **plant cell wall degrading enzymes**, which could possibly be used in the saccharification of complex (hemi-) cellulosic plant biomass. More than 200 genes were identified encoding **cellulases**, **hemi-cellulases** and **accessory enzymes**. Comparison of the plant cell wall degrading capacity of C1 with that of *Aspergillus niger* and *Trichoderma reesei*, revealed interesting similarities and differences (Table 2). C1 and *A. niger* are comparable with respect to the number of hemi-cellulases, while *T. reesei* has notably less.

C1 distinguishes itself from *A. niger* and *T. reesei* by the presence of a relatively high number of (glucurono) arabinoxylan degrading enzymes. Thus, C1 is particularly **rich in (hemi-)cellulolytic enzymes** (5). Furthermore, several C1 enzymes show high thermostability and broad pH optima, which make them promising candidates for **white biotechnology** applications (6,7).

Table 2. Comparison of the C1, *A. niger* and *T. reesei* genomes with respect to the number of a subset of putative (hemi-)cellulase encoding genes.

	C1	<i>A. niger</i> ^a	<i>T. reesei</i> ^b
Cellulases (GH3, 5, 6, 7, 12 and 45)	30	35	27
Cellulases (GH61)	24	7	9
Cellulose binding domains (CBM1)	46	8	11
Xylanases	13	5	5
Arabinofuranosidases / arabinases	14	13	3
Esterases (Axe, Fae)	13	15	2

^a, from the CAZy database (8). ^b, from the JGI database (9) and (10). GH, CAZy glycoside hydrolase family number.

In addition to these carbohydrate active enzymes, other enzyme classes with industrial potential were mined and identified. Preliminary results indicated the presence of about e.g. 50 **glycosyl transferases**, 150 **proteases** (excluding “proteasome” proteases), 700 **oxido-reductases** and 75 **lipases / esterases**.

Taxonomic re-classification of C1

In the early 1990-ies, C1 was isolated from forest alkaline soil in eastern Russia. It showed secreted neutral cellulase activity and it was characterized as a haploid filamentous fungus. Based on **morphological characteristics** the isolate was classified as *Chrysosporium lucknowense*. Recently, nucleotide sequence alignments were applied in order to re-evaluate C1’s taxonomical classification. Figure 2 displays an example of such an alignment. Based on these **molecular studies** it was concluded that C1 is not a *Chrysosporium lucknowense* species but a *Myceliophthora thermophila* species instead. C1 therefore belongs to the class of the **Sordariomycetes** instead of the Eurotiomycetes.

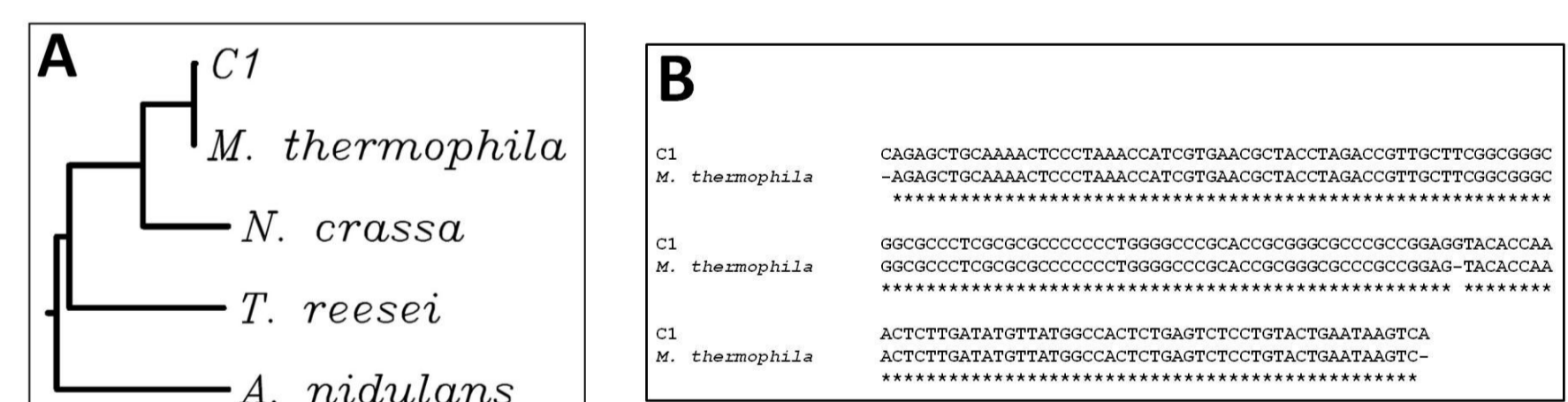


Figure 2. Molecular phylogenetic analysis of ITS1 regions of ribosomal RNA for determination of C1 taxonomy. A, phylogenetic tree (ClustalW). B, ITS1 sequence alignment. Accession numbers: Mt, see JGI genome project strain ATCC 42464. Nc, FJ360521. Tr, Z48932. An, AF138289.

Future work

The C1 genome database is continuously being improved in order to support C1 strain development and protein production R & D.

Conclusions

- C1 is rich in industrial enzymes, glycosyl hydrolases in particular.
- C1 enzymes show high thermostability and broad pH optima.
- An improved C1 draft genome sequence (38.5 Mbp) was determined.
- Approximately 9499 unique genes were predicted.
- A C1 genome database was constructed.
- C1 was re-classified as *Myceliophthora thermophila*.

Contact: Dr. H. Visser (hvisser@dyadic.nl) or Dr. J. Wery (jwery@dyadic.nl)

References

- (1), Ter-Hovhannisyan et al. (2003) Genome Res. 18:1979-1990. (2), Guigo (1998) J. Comput. Biol. 5:681-702. (3), Salzberg et al. (1998) Nucleic Acids Res. 26:544-548. (4), <http://gmod.org/wiki/Gbrowse>. (5), Hinz et al. (2009) J. Cereal Science 50: 318-323. (6), Gusakov et al. (2005) Enz. Microb. Tech. 36:57-69. (7), Ustinov et al. (2008) Enz. Microb. Tech. 43:56-65. (8), <http://www.cazy.org/>. (9), <http://www.jgi.doe.gov/genome-projects/>. (10), Martinez et al. (2008) Nat. Biotechnol. 26:553-560.