Genome Sequence and Analysis of the Soil Cellulolytic Actinomycete

**Thermobifida fusca YX**

Athanasios Lykidis,1* Konstantinos Mavromatis,1 Natalia Ivanova,1 Iain Anderson,1 Miriam Land,2 Genevieve DiBartolo,1 Michele Martinez,1 Alla Lapidus,1 Susan Lucas,1 Alex Copeland,1 Paul Richardson,1 David B. Wilson,3 and Nikos Kyrpides1

Department of Energy, Joint Genome Institute, Walnut Creek, California 94598; Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831; and Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853

Received 15 December 2006/Accepted 24 December 2006

**Thermobifida fusca** is a moderately thermophilic soil bacterium that belongs to *Actinobacteria*. It is a major degrader of plant cell walls and has been used as a model organism for the study of secreted, thermostable cellulases. The complete genome sequence showed that *T. fusca* has a single circular chromosome of 3,642,249 bp predicted to encode 3,117 proteins and 65 RNA species with a coding density of 85%. Genome analysis revealed the existence of 29 putative glycoside hydrolases in addition to the previously identified cellulases and xylanases. The glycosyl hydrolases include enzymes predicted to exhibit mainly dextran/starch- and xylan-degrading functions. *T. fusca* possesses two protein secretion systems: the sex general secretion system and the twin-arginine translocation system. Several of the secreted cellulases have sequence signatures indicating their secretion may be mediated by the twin-arginine translocation system. *T. fusca* has extensive transport systems for import of carbohydrates coupled to transcriptional regulators controlling the expression of the transporters and glycosylhydrolases. In addition to providing an overview of the physiology of a soil actinomycete, this study presents insights on the transcriptional regulation and secretion of cellulases which may facilitate the industrial exploitation of these systems.

**MATERIALS AND METHODS**

**Genome sequencing and assembly.** The complete genome of *T. fusca* was sequenced at the Joint Genome Institute using a combination of 3-kb and fosmid (40-kb) libraries. Library construction, sequencing, finishing, and automated annotation steps were performed as described at the JGI web page (http://www.jgi.doe.gov/sequencing/index.html). Predicted coding sequences (CDSs) were manually analyzed and evaluated using an Integrated Microbial Genomes (IMG) annotation pipeline (http://img.jgi.doe.gov).

**Genome analysis.** Comparative analysis of *T. fusca* with related organisms was performed using a set of tools available in IMG. Unique and orthogonal *T. fusca* genes were identified by using BLASTp (cutoff scores of E < 10−7 and 20% identity and reciprocal hits with cutoff scores of E < 10−5 and 30% identity, respectively). Signal peptides were identified using the SignalP 3.0 (3) and TMHMM (25) at default values. Whole-genome comparisons were performed using MUMmer (27).

**Nucleotide sequence accession numbers.** The sequence data described here have been deposited in GenBank (CP000088).

**RESULTS AND DISCUSSION**

**Genome features and comparative Actinobacteria genomics.** The *T. fusca* genome consists of a single circular chromosome with 3,642,249 bp. The GC content is 67.5%, and there are 3,117 predicted CDSs in the genome. The overall genome statistics are listed in Fig. 1. Among the predicted genes, 68% have been assigned a function. Twenty-six percent (830 genes) display sequence similarity to other organisms in the database with no known function, and 106 genes (3.3%) appear to be unique in *T. fusca*. There are four *rrn* loci arranged in 5S-23S-16S operons.

According to Suhre and Claverie (47), thermophilic proteomes exhibit an increased content of charged (Asp, Glu, Lys, and Arg) versus polar (Asn, Gln, Ser, and Thr) residues. This
trend is a common feature for organisms with optimal growth temperatures higher than 55°C. *T. fusca* exhibits the same trend with the striking exception of Lys residues, which appear to be close to the minimum for bacteria in IMG. The reverse trend is observed for Ala, which is elevated in *T. fusca* while most thermophiles have fewer Ala residues (Fig. 2).

*T. fusca* has 412 (~13%) unique genes when compared to the 32 *Actinobacteria* genomes present in IMG. From these 412 genes only 83 CDSs have InterPro hits, and the rest are hypothetical proteins with no functional hits. Comparisons between representatives of the five major *Actinobacteria* genera, *T. fusca*, *Streptomyces coelicolor*, *Mycobacterium tuberculosis*, *Nocardia farcinica*, and *Corynebacterium diphtheriae*, indicate that there are 1,101 genes (~30%) genes shared between these five organisms. *T. fusca* has 660 unique genes (20%) compared to the above five genomes. Overall, comparisons between these five genomes both in terms of gene similarity (Table 1) and synteny (Fig. 3) indicate that *T. fusca* is most closely related to *S. coelicolor* and *N. farcinica*.

**Cellulases and glycoside hydrolases.** The genome of *T. fusca* encodes a total of 45 hydrolytic enzymes predicted to act on oligo- and/or polysaccharides as identified by the CAZy ModO database (http://afmb.cnrs-mrs.fr/CAZY/) (Table 2). These enzymes include 36 glycoside hydrolases, 9 carbohydrate esterases, and 2 polysaccharide lyases. The glycoside hydrolases are distributed in 22 families (GH families), with the GH_13 family being the most abundant with six individual members represented in the genome. Computational analysis of signal peptides predicts 16 glycoside hydrolases are secreted. The majority of these secreted proteins (13 out of 16) have a signal sequence resembling the twin-arginine translocation (TAT) signal in their N terminus. Analysis of the secretion systems will be presented in a subsequent section.

Fourteen enzymes have been isolated and studied (including six enzymes with various cellulolytic activities and cellulases E1 to E6, as well as a β-mannanase and an endo-xylanase) from *T. fusca* (Table 2). In addition to these enzymes, 28 more putative glycosyl hydrolases and enzymes potentially involved in plant cell wall degradation were identified in the genome (mainly with dextran/starch- and xylan-degrading functions). Two enzymes were identified to be similar to endochitinase and exochitinase (Tfu0580 and Tfu0868, respectively). The absence of *N*-acetyl-hexosaminidase implies that chitin degradation proceeds either with an alternative pathway or by gradual degra-
Deacetylation of the substrate by an exochitinolytic enzyme. Deacetylases of chitin, xylan, and N-acetyl-glucosamine were identified, suggesting that the organism can utilize these substrates for energy production. *T. fusca* contains several enzymes involved in the degradation of xylan: endo-1,4-xylanase (Tfu1213, Tfu2791, and Tfu2923), xylosidase (Tfu1613), -L-arabinofuranosidase (Tfu1616), and several CDSs with putative acetylxylan esterase activity. In addition to the above-mentioned carbohydrases, *T. fusca* possesses two CDSs with significant similarity to pectate lyases, Tfu0153 and Tfu2168 (Table 2). These two genes encode secreted proteins that possess a signal peptide that may be interacting with the TAT system. Pectate lyases (EC 4.2.2.2) catalyze the eliminative cleavage of de-esterified pectin. However, *T. fusca* does not appear to possess a pectin methylesterase or a pectin acetylesterase gene. These observations agree with experimental evidence showing that *T. fusca* cannot grow on pectin but does grow on its de-esterified derivative, polygalacturonic acid (pectate), which is the major pectin polymer in the cell walls of most plants. Besides pectate lysases, *T. fusca* has a gene (Tfu2009) predicted to be a rhamnogalacturonan lyase, which may be involved in the degradation of rhamnogalacturonan, a constituent of pectin and pectate.

Besides the glycohydrolases mentioned above, the *T. fusca* genome encodes a set of four CDSs that possess a CBM domain without any identifiable glycohydrolase catalytic domain. Tfu0644 is predicted to have one predicted transmembrane helix, exposing the CBM2 domain to the extracellular face of the membrane. Tfu1268 is a secreted protein containing a CBM33 chitin binding domain. Tfu1665 is a secreted protein that contains two CBMs: a chitin binding domain at its N terminus (CBM33) and a C-terminal cellulose binding module (CBM2). We hypothesize that these proteins probably participate in plant cell wall degradation by providing scaffold services to hydrolytic enzymes and facilitating their action. Tfu1643 is an intracellular protein with a ricin-type...
| Locus tag | Family | Description | Binding module | Signal peptide? | Potential TAT signal? | Reference(s) |
|-----------|--------|-------------|----------------|-----------------|---------------------|--------------|
| Tfu0082 CE_1 | Acetyl xylan esterase | Yes | Yes, RRIPARVWVALAVGLGAAALLGITTALAPR AEAA |
| Tfu2990 CE_3 | Acetyl xylan esterase | Yes | No |
| Tfu1621 CE_4 | Acetyl xylan esterase/chitin deacetylase | Yes | No |
| Tfu2788 CE_4 | Acetyl xylan esterase/chitin deacetylase | Yes | Yes, PRRSRLRLLVALCALGLAFTSAATAHAQV |
| Tfu2789 CE_4 | Acetyl xylan esterase/chitin deacetylase | Yes | No |
| Tfu2473 CE_9 | 6P-NAG deacetylase/6P-NAGal deacetylase | No |
| Tfu0084 CE_14 | GlcNAc-Pl deacetylase | No |
| Tfu0449 CE_14 | GlcNAc-Pl deacetylase | No |
| Tfu0486 CE_14 | GlcNAc-Pl deacetylase | No |
| Tfu0937 GH_1 | β-Glucosidase/cellubiose | No | 46 |
| Tfu1629 GH_1 | Glucosidase/galactosidase | Yes | No |
| Tfu0915 GH_2 | Galactosidase | No |
| Tfu1607 GH_3 | Xylosidase/acetylhexosaminidase | Yes | No |
| Tfu2486 GH_3 | Xylosidase/acetylhexosaminidase | Yes | No |
| Tfu2768 GH_4 | Glucuronidase/galactosidase/glucosidase | No |
| Tfu0900 GH_5 | β-Mannanase (EC 3.2.1.78) | Yes | Yes, RRRKLVAAAATVALLASVFALTQPAANAT |
| Tfu0901 GH_5 | Endocellulase (E5) | Cellulose, CBM2 | Yes | Yes, RKGCPVPATVATLALLIALLSPGV |
| Tfu2712 GH_5 | Cellulase | Cellulose, CBM3 | Yes | No |
| Tfu0620 GH_6 | 1,4-Exocellulase (E3) | Cellulose, CBM2 | Yes | Yes, RRSWRRGGLAASG |
| Tfu1074 GH_6 | 1,4-Endoglucanase (E2) | Cellulose, CBM2 | Yes | No |
| Tfu1627 GH_9 | 1,4-Endoglucanase (E1) | Cellulose/amorphous/ xylan, CBM2, CBM4 | Yes | Yes, RRPRSPSLVAL |
| Tfu2176 GH_9 | Cellulase (E4) | Cellulose, CBM2, CBM3 | Yes | Yes, PRRRGHRSSRRRF |
| Tfu2791 GH_10 | Xylanase | No |
| Tfu2923 GH_10 | Xylanase | Cellulose, CBM2 | Yes | Yes, RHRPSRRRSLRLLLTSALTAAGILVTAAP AQAES |
| Tfu1213 GH_11 | Endo 1,4-β-xyylanase | Yes | No |
| Tfu0852 GH_12 | Amylase/pullulanase | No |
| Tfu0854 GH_13 | Amylase | No |
| Tfu0833 GH_13 | Amylase/pullulanase | No |
| Tfu0985 GH_13 | Amylase/pullulanase | Starch, CBM20 | Yes | Yes, RRSLAALLAALGCATSVLATVAASPA |
| Tfu1891 GH_13 | Amylase/pullulanase | No |
| Tfu0046 GH_15 | Amylase/dextranase | No |
| Tfu1345 GH_15 | Amylase/dextranase | No |
| Tfu0580 GH_18 | Chitinase | No |
| Tfu0868 GH_18 | Exochitinase | Yes | Yes, RRRFAPTWWVLLVAAGVVA |
| Tfu0480 GH_23 | Lyszyme | No |
| Tfu2594 GH_23 | Lyszyme | No |
| Tfu1613 GH_31 | Amylase/isoamylase | No |
| Tfu1615 GH_42 | Galactosidase | No |
| Tfu1616 GH_43 | Xylanase/galactan galactosidase | Yes |
| Tfu1959 GH_48 | β-1,4-Exocellulase (En-celF) | Cellulose, CBM2 | Yes | Yes, RRWRDSLAAALAAALVLSPGAHAA |
| Tfu3044 GH_65 | Trehalase/maltose phosphate | No |
| Tfu1612 GH_74 | Xyloglucanase, endo-β-1,4-glucanase | Yes | Yes, RRRGIIARLTCAAAAATVAAV |
| Tfu2205 GH_77 | Amylomaltsase or 4-glucanotransferase | No |
| Tfu2130 GH_81 | Glucanase | Yes | Yes, RRWRRAATSAATAALLGCAATTS |
| Tfu1614 GH_95 | Fucosidase | No |
| Tfu0153 PL_1 | Pectate-lyase | Yes | Yes, MRRAATLGVVALPITLAPPASSALQ |
| Tfu2168 PL_1 | Pectate-lyase | Yes | Yes, VRSSRICTEAATLATAAVVVAGTGLTVSPA SAQT |
| Tfu0644 Pl_1 | Cellulase, CBM2 | No |
| Tfu1268 Pl_1 | Chitin, CBM33 | Yes | No |
| Tfu1643 Pl_1 | Xylan/mannose/galactose, CBM13 | Yes | No |
| Tfu1665 Pl_1 | Cellulose/chitin, CBM2, CBM3 | Yes | No |
| Tfu2009 Pl_1 | Rhamnogalacturonan lyase | Cellulose, CBM2 | Yes | Yes, RRRPVRGAALAAAFVGLGATGAAALFP |

**Table 2.** Summary of glycoside hydrolases and CBM-containing proteins in *T. fusca*
The transcriptional regulation of six T. fusca cellulase genes (Tfu0620, Tfu0901, Tfu1074, Tfu1627, Tfu1593, and Tfu2176) is mediated by the CelR gene product (Tfu0938), which binds to a 14-bp inverted repeat (5′-TGGGAGCGCTCCCG) in their 5′-upstream regions (44, 45). CelR acts as a repressor, and physiological concentrations of cellobiose (the major end product of cellulosomes) cause dissociation of the CelR-DNA complex. The 14-bp inverted repeat mentioned above is found in an additional five locations in the T. fusca genome.

It is found upstream of Tfu0934 which, together with Tfu935 and Tfu936, constitutes a disaccharide ABC transport cassette. The chromosomal location of this cassette is immediately upstream of the intracellular β-glucosidase bgc (Tfu0937) (46) and the CelR-coding gene (Tfu0938). Spiridonov et al. characterized the operon Tfu935-Tfu938 and postulated that imperfect palindromes located upstream of Tfu935 may represent putative CelR binding sites. However, the existence of the perfect palindromic 5′-upstream of Tfu934 suggests that the set of Tfu0934-Tfu0938 may be transcriptionally coregulated by CelR. There are four additional genes with perfect CelR binding sites at their 5′ region: Tfu1135, Tfu1508, Tfu1665, and Tfu2844. Tfu1135 is a unique protein in T. fusca with no apparent homologs or similar genes in other organisms. Tfu1508 is a membrane protein of the major facilitator family, and it may be involved in the transport of the cellulose hydrolysis products into the cell. Tfu1665 is a secreted protein that contains three carbohydrate binding domains: an N-terminal chitin binding domain, a C-terminal cellulose binding domain (CBM_2 class), and a fibronectin type III binding region located in the middle of the protein. This structure suggests that Tfu1665 may act as a scaffold protein mediating the action of other carbohydrate-hydrolyzing enzymes. It does not contain a TAT signal, and presumably it is secreted via the sec pathway. Finally, Tfu2844 is an oxidoreductase of unknown specificity.

Besides the above-mentioned genes that contain perfect CelR binding palindromes, an additional secreted xylanase (Tfu2923) has an imperfect palindromic in its upstream region (5′-TGGGAGCGCTCCCG). Interestingly, this imperfect palindrome is located upstream of two secreted glycosylhydrolases (SCO0554, SCO0643, SCO0765, SCO1187, SCO0564, and SCO0763), a protein that contains a cellulose binding domain (SCO5786), and one lac-type transcriptional regulator (SCO2794). In the case of S. avermitilis, the palindrome is located upstream of two secreted glycosylhydrolases (Sav555 and Sav1854) as well as the cellobiose transporter (Sav5256, Sav5255, and Sav5254) and a lac-type transcriptional regulator (Sav5257).

There are six additional genes (Tfu2790, Tfu1620, Tfu1710, Tfu0834, Tfu1922, and Tfu0909) that have significant similarity to CelR and belong to the lactose repressor family (lac), and they might be involved in carbohydrate-induced transcriptional regulation. These genes may respond to different agonists than CelR. CelR is probably not a universal transcriptional regulator, since its levels are diminished when T. fusca grows on glucose or xylan. All of the above lac-type CDSs except Tfu1710 are localized on the chromosome adjacent to saccharide transporters and glycoside hydrolases (Fig. 4A). Based on this observation, we hypothesize that they regulate the expression of the respective genomic regions.

**Protein transport, secretion, adherence, and motility.** T. fusca has a complete set of proteins comprising a sec system for general protein secretion. It has homologs of the secA (Tfu2490 and Tfu0761), secY (Tfu2626), secE (Tfu2660), secG (Tfu2014), and yajC (Tfu2092) secretion components. A signal peptidase (Tfu0667) is also present. Most organisms have one copy of the SecA protein family, and recent evidence suggests that the second copy has distinct functions (5). In Actinobacteria a second copy of secA exists in Bifidobacterium longum DJO10A, all corynebacteria and mycobacteria sequenced to date, and S. avermitilis. T. fusca has all of the components of the sec-independent twin-arginine protein translocation (TAT) system: Tfu1768 (tatC), Tfu1769 (tatA/E), Tfu0381 (tatB), and Tfu0398 (tatD). The distinguishing feature of the TAT system is its ability to translocate fully folded proteins, and it may have significant implications for the secretion of functional cellulosomes. The genome does not contain any homologs of the type I and type III secretion systems. However, CDSs with similarity to components of the type II secretion system have been identified. Previous work demonstrated that the secretion of a Cel5 cellulase by Erwinia chrysanthemi was carried out by a type II protein secretion system (6). On the contrary, Faury et al. reported that the TAT export system mediates the secretion of an active xylanase from Streptomyces lividans (9). Moreover, recent work demonstrated that the TAT system is the major route for protein export in S. coelicolor (52). Therefore, we investigated whether the T. fusca genome encodes components of the type II secretion apparatus.

**Type II secretion systems generally consist of 12 proteins and share many components with the type IV pilus biogenesis machinery (35).** Examination of the genome revealed the existence of CDSs similar to type II and IV secretion components distributed between two operons. The first operon contained three CDSs with homologies to pilus assembly protein TadB (Tfu0126 and Tfu0127) and the cytoplasmic ATPase of the PulE type (Tfu0128). The second operon, located between Tfu2271 and Tfu2276, contains proteins with homologies to the pilus assembly proteins CpaB (Tfu2271), TadG (Tfu2273), and GspII_F (Tfu2275 and Tfu2276). It also contains two predicted ATPases, Tfu2274, which contains a pulE-like domain (Pfam GspII_E), and Tfu2272, which is also an ATPase containing a receiver domain at its amino-terminal region. However, we were unable to identify any homologs of the secretin (T2SD), which is the component forming channels for protein efflux, in T. fusca and other Actinobacteria. On the contrary, the gene similarities and structures of the respective chromosomal regions resemble the tight adherence (tad) locus (Fig. 4B) identified in the gram-negative bacterium Actinobacillus actinomycetemcomitans (20, 37,
The tad locus has been identified as an important mechanism for the production of pili that mediate nonspecific adherence to surfaces and formation of biofilms (21). However, both T. fusca and S. coelicolor lack homologs of either RepA (a secretin family protein) or prepilins. We hypothesize that these particular functions have been replaced by nonorthologous genes in T. fusca. Overall, based on the above observations, we propose that T. fusca lacks a type II secretion system, and the above-mentioned CDSs constitute a system for nonspecific adherence to surfaces analogous to the one characterized in Actinobacillus actinomycetemcomitans.
temcomitans (36, 37). Consequently, we hypothesize that the secretion of cellulases in T. fusca is mediated by the TAT translocation system and the general secretion system.

**Transport systems.** Approximately 6% (180 CDSs) of the T. fusca genome encodes genes for transport. T. fusca has the core phosphotransferase system components E1 (Tfu2489) and Hpr (Tfu2487), as do all published Actinobacteria except Tropheryma (reduced genome) and mycobacteria. However, unlike other Actinobacteria, it has no identifiable homologs of PTS transporters. Therefore, we hypothesize that T. fusca does not have an active PTS transport system, and the E1 and Hpr components function in signaling pathways. This is a major difference with the carbohydrate transport system of S. coelicolor, which utilizes the PTS system to transport fructose and N-acetylgalcosamine (33, 34).

T. fusca has an extensive set of transporters for carbohydrate uptake. Eight ABC disaccharide transporter cassette were identified in the genome, and the majority of them (five) are localized on the chromosome next to glycoside hydrolases (Fig. 4A). The transport system (Tfu0934-Tfu0936) that clusters with the CelIR transcription factor (see above) exhibits high similarity to the cellobiose/cellotriose transporter system of Streptomyces reticuli (40). This system lacks an ATPase component in both organisms. However, T. fusca harbors a homolog of the Ms1K (multiple sugar import) protein, Tfu1763. Ms1K has been characterized in Streptomyces, and it was shown to assist both in cellobiose and maltose transport (41). The T. fusca maltose transport system consists of three CDSs, Tfu0830, Tfu0831, and Tfu0832. It is clustered on the chromosome with Tfu0833, which is an intracellular 1,4-α-D-glucanohydrolase, probably involved in the hydrolysis of maltose. The adjacent CDS (Tfu0834) is a lacI-type transcriptional regulation and may be involved in the transcription of the transporter-glycoside hydrolase operon. The chromosomal region between Tfu1609 and Tfu1620 (12 CDSs) contains two transport systems, a set of two secreted and three intracellular glycosyl hydrolases, as well as a lac-type transcriptional regulator (Fig. 4A). Tfu1617, Tfu1618, and Tfu1619 are components of the xylobiose transporter based on their high similarity to the recently characterized xylobiose transporter from Streptomyces thermoviolaceus (49). Tfu1612 is the secreted xyloglucanase possessing a potential TAT targeting signal peptide characterized by Irwin et al. (17). Tfu1616 is a GH3 family protein encoding an uncharacterized secreted glycoside hydrolase activity, whereas Tfu1613, Tfu1614, and Tfu1615 are intracellular glycoside hydrolases. Tfu1609, Tfu1610, and Tfu1611 are components of a putative disaccharide transport system with unknown specificity. Tfu1620 is a lac-type transcriptional regulator, and we postulate that it regulates the expression of the gene set Tfu1609-Tfu1620.

The genome of T. fusca also has an ortholog (Tfu1668) of the GdeP sugar permease characterized in S. coelicolor as a major glucose uptake system (51). Overall, our analysis identified the presence of cellobiose/cellulotriose, maltose and xylobiose ABC transport systems, and a permease of the major facilitator superfamily for glucose uptake.

There are four putative amino acid ABC transport systems (Tfu0304-Tfu0306, Tfu0910-Tfu0914, Tfu1182-Tfu1187, and Tfu2928-Tfu2930).

Genes encoding a siderophore transport system were identified (Tfu0656, Tfu0657, and Tfu1491-Tfu1494) as well as a siderophore biosynthesis cluster (Tfu1865-Tfu1867). In addition, there is a ferrous iron transport protein (Tfu0820). There are four additional heavy metal transport systems (Tfu0857-Tfu0858, Tfu0931-Tfu0932, Tfu1661-Tfu1663, and Tfu2312-Tfu2313).

**DNA repair and replication stress.** Compared to many other representatives of Actinobacteria, especially some pathogenic species, T. fusca has reduced DNA repair capabilities and fewer mechanisms of resistance to reactive oxygen species (Table 1). T. fusca has Mn-dependent and heme catalases (Tfu0886 and Tfu1649, respectively) and an Mn-dependent superoxide dismutase (Tfu0957), while other enzymes present in N. farcinica, Streptomyces spp., and Mycobacterium spp., such as catalase/peroxidase and Cu-Zn superoxide dismutase, are missing. Similarly, only a homolog of the organic hydroperoxide resistance protein Ohr (Tfu0697) is present (8), while peroxiredoxin reductase AhpD and peroxiredoxin AhpC are absent.

T. fusca is capable of repairing UV-induced cyclobutane pyrimidine dimers due to the presence of type I deoxyribozyme photolyase (Tfu0534). Deaminated bases are removed via a base excision mechanism (uracil-DNA glycosylase Tfu1341 and G/U-mismatch DNA glycosylase Tfu1918), while oxidized bases are removed by endonuclease III (Tfu0118), formamidopyrimidine-DNA glycosylase (Tfu0652), and A/G-specific adenine glycosylase (Tfu2875), and the resulting abasic sites are processed by the DNA (apurinic or apyrimidinic site) lyase activities of some of these enzymes or by endonuclease IV (Tfu1956) and endonuclease V (Tfu1400). The repertoire of pathways for repair of alkylated bases is limited in T. fusca to base excision by DNA-3-methyladenine glycosylase I (Tfu0498) and alkylation reversal via DNA-N1-methyladenine dioxygenase (Tfu1427) (23); no homologs of other methyladenine glycosylases or a suicidal protein–syntene S-methyltransferase AidA were found. Genes coding for the subunits of excinuclease ABC, an enzyme responsible for nucleotide excision repair, are also present in the genome (Tfu2024, Tfu1196, and Tfu2021). Similar to other Actinobacteria, the genes coding for a methyl-directed mismatch repair system (MutSLH proteins) are absent. T. fusca has only one error-prone DNA polymerase responsible for translesion synthesis (Tfu1096). Unlike many other representatives of Actinobacteria, no homologs of proteins responsible for double-strand break repair by nonhomologous end joining (3) were found in the genome of T. fusca (13).

One of the most striking features of T. fusca is the genetic organization of the recA gene (Tfu0803), which has two inteins (see the New England Biolabs intein database [http://www.neb.com/neb/inteins.html]) inserted at both the RecA-a and RecA-b sites (4). While intein-harboring recA genes appear to be quite common among both pathogenic and free-living mycobacteria (39), none of the genes sequenced so far has two intein insertions. Since it has been demonstrated that the intein-free RecA protein from Mycobacterium smegmatis and single intein-containing RecA protein from Mycobacterium tuberculosis display different activities in DNA strand exchange (10), RecA from T. fusca represents an interesting model to further clarify the mechanistic basis and the factors that contribute to the extent of DNA strand transfer in various Actinobacteria. Other features of recombinational DNA repair include the absence of a RecBCD complex or its AddAB equivol with DNA helicase and exonuclease activities.

To access the full article, visit [http://jb.asm.org/content/vol189/fig26](http://jb.asm.org/content/vol189/fig26) on April 26, 2019 by guest
ties. Some of the essential components of an alternative RecF recombination pathway operating in recBC mutants in E. coli (43) are also missing; these include helicase RecQ and 5’-3’-exonuclease RecJ. However, despite the apparent absence of any helicase or exonuclease activity typically associated with DNA recombination, other components of a recombination apparatus (24), such as RecF, RecO, RecR, and RecN proteins (Tfu0004, Tfu0852, Tfu0044, and Tfu2032, respectively) and enzymes responsible for Holliday junction resolution (Tfu2093, Tfu2094, Tfu2095, and Tfu0646), are present in the genome. Further analysis of the phylogenetic distribution of the genes coding for various components of recombination machinery in Actinobacteria reveals no correlation between the genetic structure of the recA gene and the presence of the proteins involved in RecBCD or RecF recombination pathways.

Central metabolism. The T. fusca genome encodes all the enzymes necessary to carry out glycolytic degradation of monosaccharides. Additionally, it has the Entner-Douodoroff pathway for glucose utilization.

All the enzymes of the tricarboxylic acid cycle are present. The presence of the glyoxylate cycle bypass from isotocitate to malate is indicative of its ability to grow on additional carbon sources such as acetate. The existence of homologs of glycerol kinase (Tfu0787), glycerol-3-phosphate dehydrogenase (Tfu0631), and triose phosphate isomerase (Tfu2015) indicates its ability to convert glycerol to glyceraldehyde-3-phosphate, an intermediate of glycolysis. Therefore, glycerol can also serve as a carbon source. It has the gluconeogenic enzymes pyruvate carboxylase (Tfu2554), phosphoenolpyruvate carboxykinase (Tfu0083), and fructose-1,6-bisphosphatase (Tfu0646) indicative of gluconeogenesis.

It has the machinery for the de novo biosynthesis of all amino acids except asparagine. However, it possesses a class II aspartyltRNA synthetase (Tfu2086), indicative of its ability to synthesize Asn-tRNA through transamidation of Asp-tRNA (32).

All the enzymes necessary for the biosynthesis of purines and pyrimidines have been identified. T. fusca has the enzymes for NAD, coenzyme A, riboflavin, pyridoxal, folate, cobalamin, and menaquinone biosynthesis. Two pathways have been described for the biosynthesis of quinolinate, an intermediate in NAD biosynthesis (26). The tryptophan-to-quinolinic pathway is present in T. fusca as well as in both Streptomyces strains sequenced to date, but it is absent in the rest of the Actinobacteria. T. fusca also utilizes the aspartate-to-quinolinic pathway, which is common to all Actinobacteria. T. fusca does not have a biotin pathway biosynthesis, since homologs of BioB, BioF, and BioD have not been identified. However, it possesses a transporter (Tfu2314) for scavenging biotin from the environment. The biosynthetic pathway of thiamine is also incomplete.

Lipid and cell wall metabolism. T. fusca is able to synthesize and metabolize saturated and unsaturated fatty acids, as well as other major lipid classes, such as phospholipids, glycolipids, and isoprenoids. T. fusca appears to have all the enzymes necessary for fatty acid biosynthesis. It utilizes exclusively a type II fatty acid synthesis system in common with Streptomyces and in contrast to Mycobacterium, Corynebacterium, and Nocardia, which utilize a type I fatty acid synthase in addition to the type II system. It generates unsaturated fatty acids by a fatty acid desaturase mechanism (Tfu0413). An interesting feature of the T. fusca genome lies in the existence of a PslX homolog in the absence of its PslY counterpart. PslX and PslY were recently shown to constitute a novel mechanism for phospholipid biosynthesis initiation through the formation of acylphosphates (30). T. fusca appears to lack both a PslB as well as a PslY homolog, raising the intriguing possibility that an as-yet-unidentified enzyme participates in the initiation of phospholipid biosynthesis.

Two secreted triacylglycerol lipases can be identified in the genome (Tfu0882 and Tfu0883). It has the enzymes for β-oxidation of both odd and even carbon number fatty acids. In contrast to Mycobacteria and Streptomyces, it does not have homologs of diacylglycerol acyltransferases and, therefore, it probably lacks the machinery for endocellular triacylglycerol accumulation.

Hence, these observations suggest that T. fusca can hydrolyze extracellular lipids, take up the resulting fatty acids, and utilize them subsequently as a carbon source. T. fusca has all the genes for the biosynthesis of phosphatidylglycerol, phosphatidylserine, phosphatidylethanolamine, and phosphatidylglycerol. Phosphatidylglycerol is further decorated with mannosyl groups, as evidenced by the existence of a phosphatidylmannosyltransferase homolog (Tfu2101). It utilizes a phospholipase A-type of cardiolipin synthase (Tfu2817). This type of enzyme catalyzes the synthesis of cardiolipin from two molecules of phosphatidylglycerol. In contrast, Mycobacterium and Nocardia appear to possess a “eukaryotic”-type reaction via a CDP-alcohol phosphatidyltransferase, which catalyzes cardiolipin formation from CDP-diacylglycerol and phosphatidylglycerol.

Additionally, T. fusca and both S. coelicolor and S. avermitilis have two novel classes of CDP-alcohol phosphatidyltransferases (Tfu2439 and Tfu2359), indicative of their ability to synthesize as-yet-identified phospholipids. The most attractive hypothesis is that these enzymes catalyze the formation of novel structures of phosphatidyl sugars.

It possesses a complete nonmevalonate pathway for isoprenoid biosynthesis. The existence of genes with homology to phytoene synthase (Tfu3076), phytoene dehydrogenase (Tfu3075), and lycopene cyclase (Tfu3088) is indicative of a putative carotenogenesis process (Fig. 4C). A recent study established the ability of S. coelicolor to carry out light-induced synthesis of carotenoids (48).

All the genes (murC, murD, murE, and murF) that encode enzymes for the conversion of d-glutamate to UDP-MurNac pentapeptide are present, as are other enzymes in the peptidoglycan synthetic pathway, such as GlnU, MurA, and MurB.

Conclusions. The genome sequence of T. fusca provides the means for a detailed analysis of the cellular mechanisms controlling the expression and secretion of plant cell wall-degrading enzymes by this soil bacterium. T. fusca utilizes a variety of enzymes attacking cellulose, xylan, and pectin, major components of plant cell walls. Detailed genomic analysis provides evidence for the utilization of the TAT secretion system for the export of these enzymes to the extracellular space, the existence of multiple transcription factors regulating the expression of glycosylhydrolases, and oligo/poly saccharide transport systems. These observations open further research directions for understanding the mechanisms of plant cell wall hydrolysis and utilization by soil actinomycetes.
ACKNOWLEDGMENTS

This study was performed under the auspices of the U.S. Department of Energy’s Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231 and Lawrence Livermore National Laboratory under contract DE-AC02-06NA25396.

REFERENCES

1. Bachmann, S. L., and A. J. McCarthy. 2001. Purification and cooperative activity of enzymes constituting the xylan-degrading system of Thermomonospora fusca. Appl. Environ. Microbiol. 67:5257–5264.

2. Bellamy, W. D. 1977. Cellulose and lignocellulose digestion by thermophilic actinomycetes for single cell protein production. Dev. Ind. Microbiol. 18:249–254.

3. Bellamy, W. D., J. H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340:783–795.

4. Blanco, J., J. J. Coque, J. Velasco, and J. F. Martin. 1997. Cloning, expression in Streptomyces lividans and biochemical characterization of a thermophilic endo-

5. Blanco, J., J. J. Coque, J. Velasco, and J. F. Martin. 1997. Cloning, expression in Streptomyces lividans and biochemical characterization of a thermophilic endo-

6. Blanco, J., J. J. Coque, J. Velasco, and J. F. Martin. 1997. Cloning, expression in Streptomyces lividans and biochemical characterization of a thermophilic endo-

7. Cussiol, J. R., S. V. Alves, M. A. de Oliveira, and L. E. Netto. 2003. Improved DNA strand transfer function of Thermobifida fusca xylanase 10B. Can. J. Biochem. 81:4988–4997.

8. DeSalle, R., P. J. Planet, S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2001. Transfer RNA-dependent amino acid biosynthesis: an essential route for multiple reasons. Cell 105:13127–13132.

9. DeSalle, R., P. J. Planet, S. C. Kachlany, M. Tomich, D. H. Fine, and D. H. Figurski. 2006. Genetic analysis of the requirement for rcpB in Actinobacillus actinomycetemcomitans biofilm formation. J. Bacteriol. 188:6361–6375.

10. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

11. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

12. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

13. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

14. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

15. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

16. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

17. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

18. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

19. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

20. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

21. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

22. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

23. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

24. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

25. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.

26. DeSalle, R., P. J., S. C. Kachlany, D. H. Fine, R. DeSalle, and D. H. Figurski. 2003. The widespread colonization island of Actinobacillus actinomycetemcomitans. Nat. Genet. 34:193–198.
2004. Molecular characterization of a high-affinity xylobiose transporter of Streptomyces thermoviolaceus OPC-520 and its transcriptional regulation. J. Bacteriol. 186:1029–1037.

50. Van den Bogart, H. G., G. Van den Ende, P. C. Van Loon, and L. J. Van Griensven. 1993. Mushroom worker’s lung: serologic reactions to thermoophilic actinomycetes present in the air of compost tunnels. Mycopathologia 122:21–28.

51. van Wezel, G. P., K. Mahr, M. Konig, B. A. Traag, E. F. Pimentel-Schmitt, A. Willimek, and F. Titgemeyer. 2005. GlcP constitutes the major glucose uptake system of Streptomyces coelicolor A3(2). Mol. Microbiol. 55:624–636.

52. Widdick, D. A., K. Dilks, G. Chandra, A. Bottrill, M. Naldrett, M. Pohlschroder, and T. Palmer. 2006. The twin-arginine translocation pathway is a major route of protein export in Streptomyces coelicolor. Proc. Natl. Acad. Sci. USA 103:17927–17932.

53. Zhang, S., G. Lao, and D. B. Wilson. 1995. Characterization of a Thermomonospora fusca exocellulase. Biochemistry 34:3386–3395.