

RESEARCH ARTICLE

Open Access

# Genome-wide analysis of *acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system

Bareket Dassa<sup>1</sup>, Ilya Borovok<sup>2</sup>, Raphael Lamed<sup>2</sup>, Bernard Henrissat<sup>3</sup>, Pedro Coutinho<sup>3</sup>, Christopher L Hemme<sup>4</sup>, Yue Huang<sup>4</sup>, Jizhong Zhou<sup>4</sup> and Edward A Bayer<sup>1\*</sup>

## Abstract

**Background:** Microbial degradation of plant cell walls and its conversion to sugars and other byproducts is a key step in the carbon cycle on Earth. In order to process heterogeneous plant-derived biomass, specialized anaerobic bacteria use an elaborate multi-enzyme cellulosome complex to synergistically deconstruct cellulosic substrates. The cellulosome was first discovered in the cellulolytic thermophile, *Clostridium thermocellum*, and much of our knowledge of this intriguing type of protein composite is based on the cellulosome of this environmentally and biotechnologically important bacterium. The recently sequenced genome of the cellulolytic mesophile, *Acetivibrio cellulolyticus*, allows detailed comparison of the cellulosomes of these two select cellulosome-producing bacteria.

**Results:** Comprehensive analysis of the *A. cellulolyticus* draft genome sequence revealed a very sophisticated cellulosome system. Compared to *C. thermocellum*, the cellulosomal architecture of *A. cellulolyticus* is much more extensive, whereby the genome encodes for twice the number of cohesin- and dockerin-containing proteins. The *A. cellulolyticus* genome has thus evolved an inflated number of 143 dockerin-containing genes, coding for multimodular proteins with distinctive catalytic and carbohydrate-binding modules that play critical roles in biomass degradation. Additionally, 41 putative cohesin modules distributed in 16 different scaffoldin proteins were identified in the genome, representing a broader diversity and modularity than those of *Clostridium thermocellum*. Although many of the *A. cellulolyticus* scaffoldins appear in unconventional modular combinations, elements of the basic structural scaffoldins are maintained in both species. In addition, both species exhibit similarly elaborate cell-anchoring and cellulosome-related gene- regulatory elements.

**Conclusions:** This work portrays a particularly intricate, cell-surface cellulosome system in *A. cellulolyticus* and provides a blueprint for examining the specific roles of the various cellulosomal components in the degradation of complex carbohydrate substrates of the plant cell wall by the bacterium.

**Keywords:** Cellulosomics, *Clostridium thermocellum*, Scaffoldin, Cohesin, Dockerin

## Background

Plant cell walls are composed of different types of recalcitrant polysaccharides, notably cellulose, which together with lignin form a rigid, stable composite material. Microbial degradation of these polysaccharides and its conversion to sugars is a key step in the carbon cycle, and its subsequent conversion to ethanol is a vital objective for

society [1]. One of the major paradigms for efficient degradation of cellulose is a supramolecular, multi-enzyme complex called the cellulosome, which was demonstrated in various bacteria [2-7]. The cellulosome harbors a multiplicity of carbohydrate-active enzymes, i.e., glycoside hydrolases (GHs), carbohydrate esterases (CEs) and polysaccharide lyases (PLs). These include multiple endoglucanases, cellobiohydrolases, xylanases and other degradative enzymes which work synergistically to attack heterogeneous, insoluble cellulose substrates [8-11]. These enzymes are very similar in their mode of action to those

\* Correspondence: ed.bayer@weizmann.ac.il

<sup>1</sup>Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel

Full list of author information is available at the end of the article

of the free enzyme systems of other bacteria and fungi, except that the cellulosomal enzymes contain a dockerin module in place of a carbohydrate-binding module (CBM), which would target the individual enzymes to the substrate. Scaffoldin (Sca), a major cellulosomal subunit, is responsible for organizing the cellulolytic subunits into the complex. The dockerin-borne enzyme subunits are integrated into the scaffoldin subunit via the tenacious protein-protein interaction with multiple copies of cohesin modules. The scaffoldin subunit also contains a single CBM that attaches the entire enzymatic complex (as well as the parent bacterial cell) to the cellulose substrate, thereby enabling efficient synergistic degradation of the substrate.

*Acetivibrio cellulolyticus* is a mesophilic, anaerobic, gram-positive bacterium, known both for its efficient degradation of crystalline cellulose [12-15] and for its distinct protuberant cell surface ultrastructure [16]. A gene cluster of four cellulosomal scaffoldin proteins (ScaA-ScaD) from *A. cellulolyticus* ATCC 33288 was studied during the past decade [17-19]. The primary scaffoldin, ScaA (previously termed CipV), contains a singular intrinsic family-9 glycoside hydrolase (GH) and mediates direct incorporation of the dockerin-containing enzymes through its seven type-I cohesins. It is bound to the cell surface via its C-terminal X-module/dockerin dyad (XDoc) to at least two additional scaffoldins. Thus, ScaA can either interact directly with the ScaD surface-anchoring scaffoldin or it may bind to the ScaC scaffoldin indirectly through a ScaB adaptor scaffoldin [18,20,21]. ScaC and ScaD serve as anchoring scaffoldins, owing to their C-terminal S-layer homology (SLH) modules, but unlike any other scaffolding yet described, the ScaD protein harbors two different types of cohesin (types I and II), which exhibit two divergent dockerin-binding specificities [19]. Thus, only four scaffoldin proteins of the bacterium have been recognized and analyzed prior to sequencing of its genome [22].

Despite the limited genomic information available at the time, a putative model of the cellulosome architecture was proposed, suggesting alternative modes of interactions among the *A. cellulolyticus* scaffoldin components and mechanisms of attachment to the cell surface. Still, the exact model and stoichiometry of the cellulosome arrangement is currently unknown. Original experiments indicated the presence of additional putative cellulosomal enzyme components [18] and scaffoldins [19] which were probed by the ScaC cohesin but were never fully identified.

The expansion of genome sequencing efforts during the past decade has also provided information regarding several cellulosome-producing bacteria [23-26], and their genome-wide comparison has spawned the field of cellulomics [5], i.e., a general overview of cellulosome-

related constituents of a given bacterium. The recent sequencing of the *A. cellulolyticus* genome [22] has thus enabled identification and analysis of numerous additional cellulosomal components, gene regulatory elements, and cell anchoring modules in the bacterium, as documented in this communication. The interrelationship of the *A. cellulolyticus* cellulosome components was further explored by genome-wide comparison of its cellulosomal architecture and subunits with those of *Clostridium thermocellum*.

## Results and discussion

### Multiplicity of scaffoldins and cohesin-containing proteins

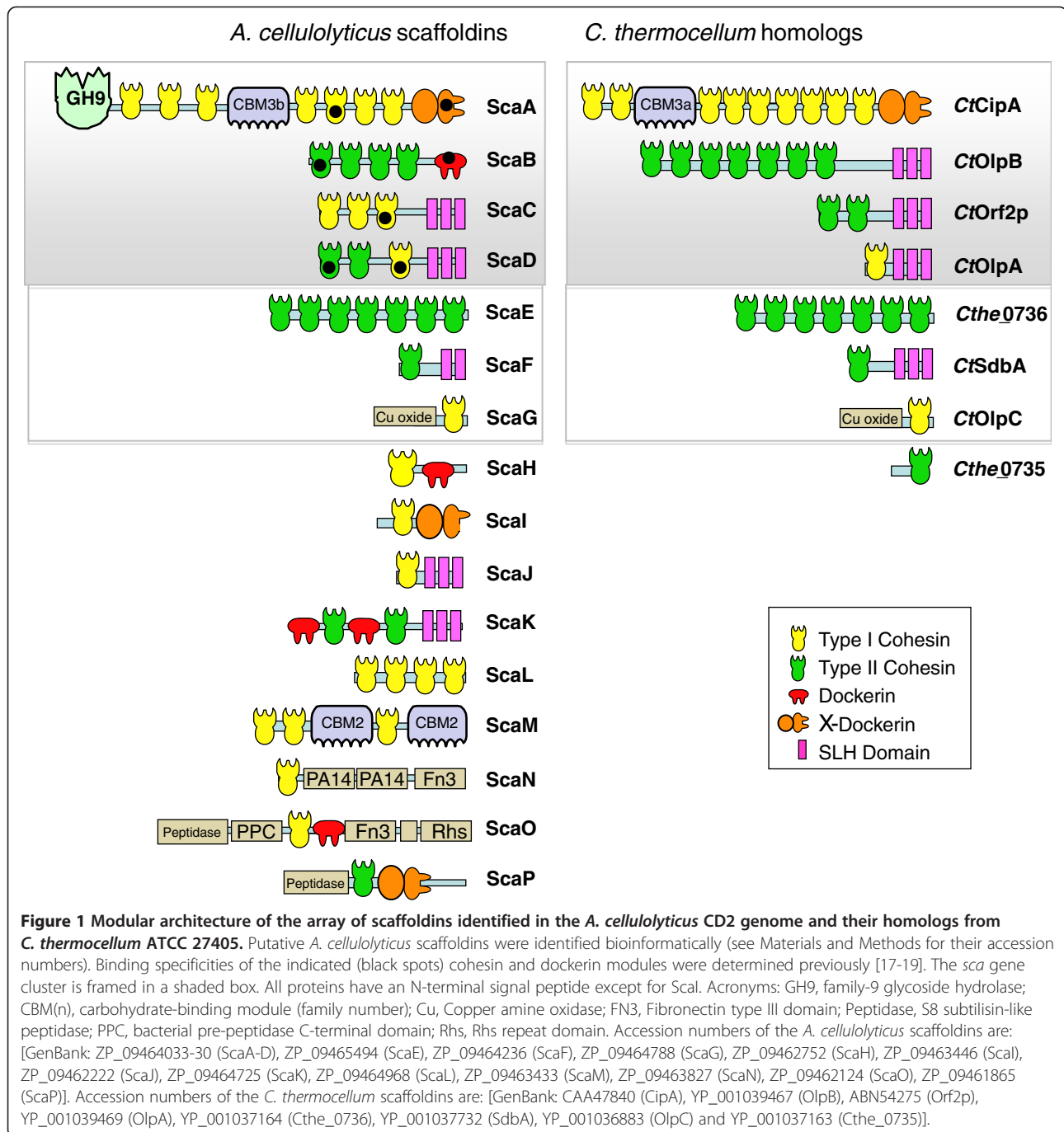
The *Acetivibrio cellulolyticus* CD2 genome [22] is the largest among the known cellulolytic bacteria (6.1 Mb). Analysis of its recent genome sequence revealed 41 putative cohesin modules, distributed in 16 scaffoldins, some of which have both cohesins and dockerins in the same polypeptide chain (Figure 1 and Additional file 1: Table S1). These include the four genes of the scaffoldin cluster (*scaA*, *scaB*, *scaC* and *scaD*), which were originally identified, sequenced and characterized in *A. cellulolyticus* ATCC 33288 [17-19].

The previous publications have indicated that this mesophilic bacterium harbors an intricate cellulosome system, which is characterized by several unique properties that distinguish *A. cellulolyticus* from the archetypical *C. thermocellum* cellulosome: The progression of the ScaA primary scaffoldin, the ScaB adaptor scaffoldin and the ScaC anchoring scaffoldin, with their resident cohesins (7, 4 and 3, respectively), suggests that the resultant fully occupied cellulosome complex would include up to 84 dockerin-containing proteins (enzymes) in addition to the intrinsic ScaA cellulase. The second type of cellulosome complex comprises a divergent anchoring scaffoldin, ScaD, which contains different cohesin specificities: two type-II cohesins that incorporate two ScaA subunits with their complement of dockerin-containing enzymes and a single type-I cohesin that binds a lone dockerin-containing protein.

Comparison of the original *A. cellulolyticus sca* genes which were individually sequenced by conventional methodology [17-19] to those of the newly sequenced genome shows only a few differences (two nucleotide substitutions out of 2601 in the *ScaB* gene [GenBank: ZP\_09464032]).

### Modular nature of the cohesin-containing proteins

In the present work, the sequenced *A. cellulolyticus* genome revealed 12 cohesin-containing proteins in addition to the previously known four major scaffoldins encoded by the *sca* gene cluster. Figure 1 presents their modular architecture. All of the proteins listed in the figure,



except for ScaI, contain a credible signal peptide, suggesting that these proteins would be secreted.

The cohesin modules exhibit a variety of intriguing sequence features. Like *C. thermocellum*, the 41 cohesins of *A. cellulolyticus* are classified into two types: type I (26 modules) and type II (15 modules). We examined the conservation of the cohesin sequences and compared copies of the various cohesin modules within a given scaffoldin protein, and among the different scaffoldins. The overall sequence similarity among the *A. cellulolyticus* cohesin

modules ranges from 41 to 97%. Some scaffoldins contain similar repeats of the same type of cohesin module, whereas others bear a single cohesin. ScaD alone contains a combination of two heterogeneous cohesin types on the same polypeptide chain. As has been experimentally documented [17-19], the cohesin type (i.e., type I or type II cohesin) does not necessarily indicate its binding specificity to a given dockerin. For example, the cohesins from ScaA and ScaC (Figure 1) are all type I according to their sequences, but they bind to different dockerins – the ScaA

cohesins bind to the dockerin-bearing enzymes, and the ScaC cohesins bind to the ScaB dockerin.

The combination of S-layer homology (SLH) modules with cohesin modules on the same polypeptide suggests a role for such proteins in anchoring the cellulosome assemblies or specific enzymes to the cell wall of the gram-positive bacterium [27,28].

In addition to the previously described anchoring scaffoldins, ScaC and ScaD, three more proteins which contain SLH modules are now revealed, i.e. ScaF [GenBank: ZP\_09464236], ScaJ [ZP\_09462222] and ScaK [ZP\_09464725]. Of the 37 SLH-containing proteins encoded in the *A. cellulolyticus* genome, ScaK was identified with an SLH module, two dockerins and two cohesin modules. This is the first example of such an architectural arrangement of a cell-surface anchoring scaffoldin that contains both types of cellulosome-related modules.

Uniquely, one cohesin-containing protein also contains two family 2 CBMs, interspersing its type-I cohesins (ScaM, [ZP\_09463433]). To our knowledge, this is the first description of a scaffoldin-borne CBM2; all previous CBM2s located on scaffoldins have been from family 3. CBM2s have been described as ancillary modules of enzymes and were shown to bind efficiently to cellulose and/or xylan. Thus, their appearance on a scaffoldin may serve to enhance the substrate-binding function of the dockerin-containing enzymes, which bind to this scaffoldin protein via its type-I cohesins. Other cohesins were identified in novel types of scaffoldins which bear FN3 (Fibronectin type III) repeats, PA14 (protective antigen) domain, peptidase or other extracellular modules.

#### **Relationship between cohesins of *A. Cellulolyticus* and *C. Thermocellum***

Complex cellulosome architectures were previously proposed for *A. cellulolyticus* and *C. thermocellum*, which are two phylogenetically related Clostridiales species, as implied from their 16 S rRNA analysis [29]. The *C. thermocellum* genome contains 8 cohesin-containing proteins (scaffoldins), whereas *A. cellulolyticus* has twice the number of scaffoldins. The cellulosome system of *C. thermocellum* was selected as the reference strain, since it is the first-identified and best-established multiple-scaffoldin system, which possesses clear similarities to that of *A. cellulolyticus* [4].

Interestingly, three pairs of scaffoldins from both species have the same basic modular organization. Thus, two homologous scaffoldins, *A. cellulolyticus* ScaE [GenBank: ZP\_09465494] and *C. thermocellum* Cthe\_0736, each consist of seven consecutive type-II cohesins (Figure 1). Likewise, ScaF [GenBank: ZP\_09464236] and *C. thermocellum* (Ct) SdbA have a similar architecture comprising a single type-II cohesin followed by an SLH

module. Finally, ScaG [GenBank: ZP\_09464788] and the cell-surface Ct OlpC [30] both possess a single type-I cohesin, following a unique domain annotated as copper amine oxidase-like [Pfam: PF07833].

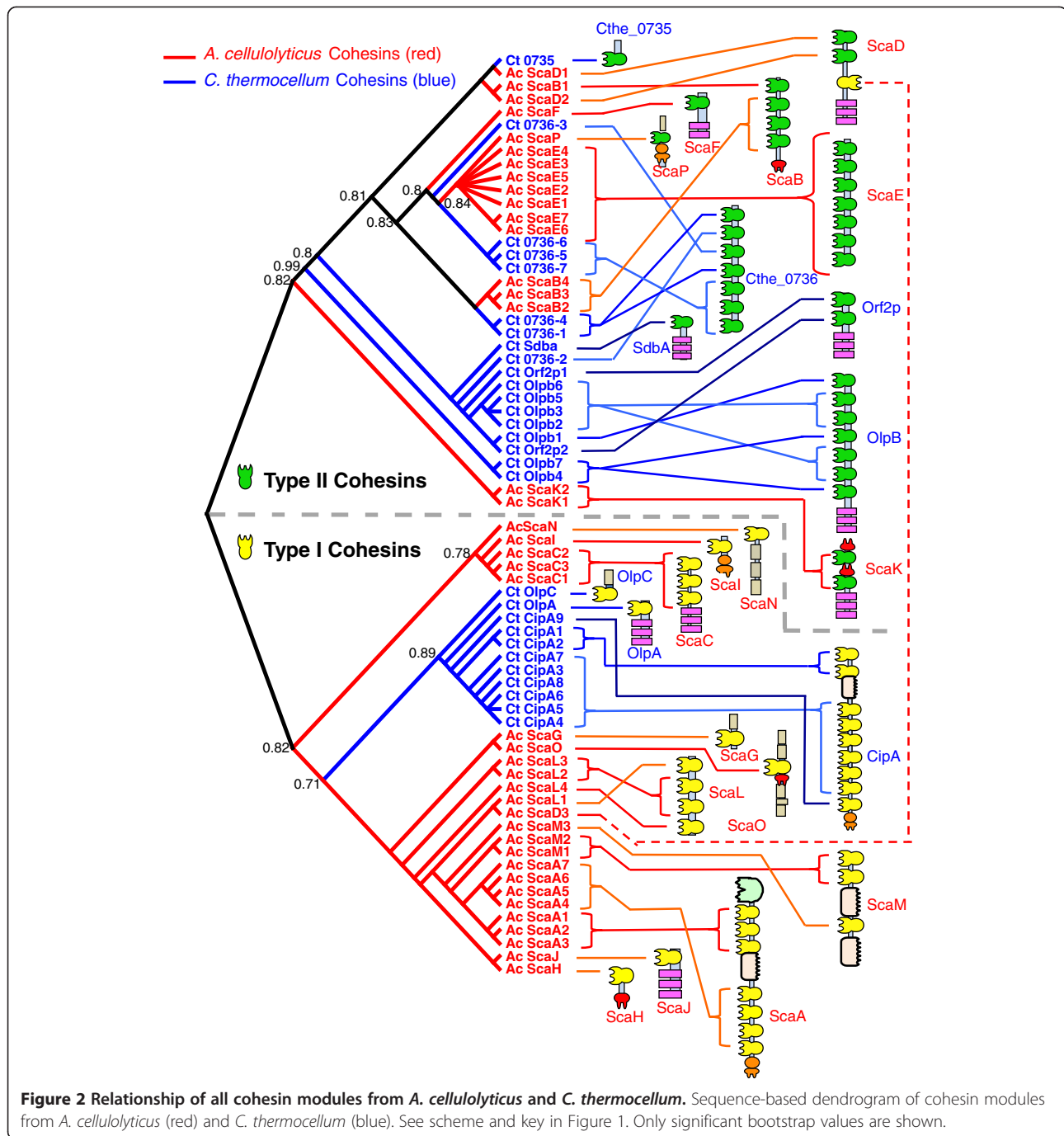
It is important to examine the phylogenetic relationship among the different cohesins within and between the two species, in order to reveal clues regarding their divergence (Figure 2). For example, all seven of the *A. cellulolyticus* ScaE cohesins are similar to each other and are thus clustered together on a single branch of the phylogenetic tree. In contrast, the seven Cthe\_0736 cohesins are interwoven on different branches, such that cohesins 1 and 4 are closely related, as are cohesins 5 to 7, indicating domain duplication events in the evolution of this protein. Further diversification of Cthe\_0736 is evident in the acquisition of cohesin 2 which bears similarity to divergent type-II cohesins of other *C. thermocellum* anchoring scaffoldins. The seven *A. cellulolyticus* ScaE cohesins appear to be most similar to Cthe\_0736 cohesins 3 and 5–7, which presumably suggests a common origin.

The cellulosomes of both species harbor several anchoring proteins, composed of one or more cohesins with SLH modules. For example, ScaF and Ct SdbA have a single type-II cohesin followed by SLH repeats. Yet, their cohesins are clustered on very different branches on the tree (Figure 2), suggesting that their parent proteins are the product of different evolutionary pathways. The ScaF cohesin is closely related to those of ScaE and the above-mentioned Cthe\_0736 cohesins, whereas that of Ct SdbA is more similar to those of the other *C. thermocellum* anchoring scaffoldins. In a similar manner, each of the anchoring scaffoldins, ScaJ and Ct OlpA, harbors a single type-I cohesin, located on divergent branches of the phylogenetic tree. As opposed to the type-II cohesins, the relationship among type-I cohesins is more straightforward, where cohesins from each species are clustered on separate branches of the tree.

#### **Abundance of dockerins in the *A. Cellulolyticus* genome**

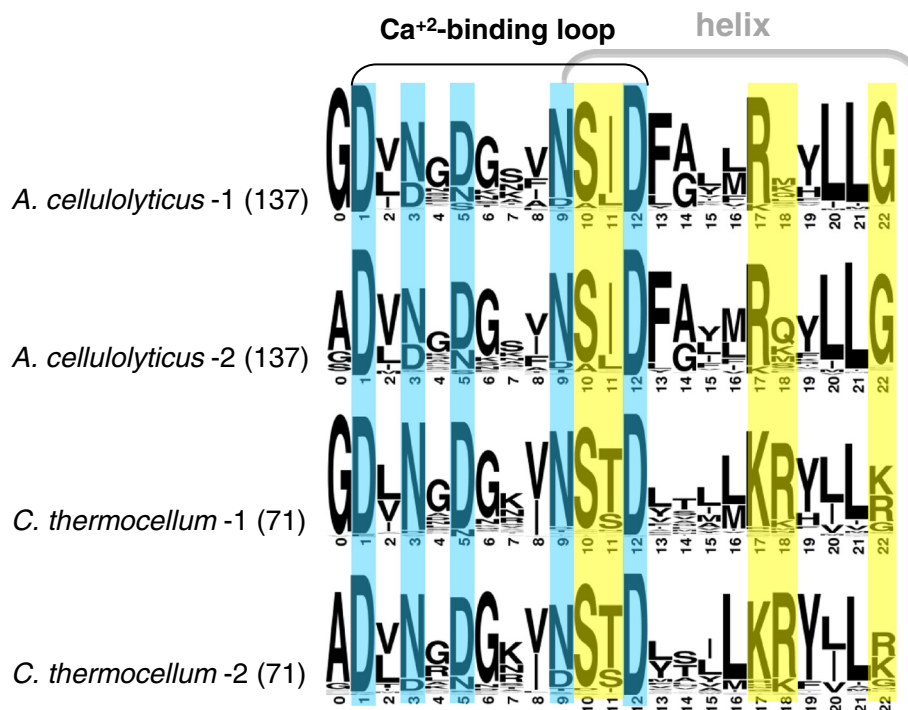
The *A. cellulolyticus* genome is particularly enriched with dockerin-containing genes, and 143 genes that contain putative dockerin modules were identified. Therefore, *A. cellulolyticus* contains almost twice the number of dockerins as other Clostridial bacteria, such as *Clostridium cellulolyticum* (>60 dockerins) or *Clostridium thermocellum* (>70 dockerins) [23,31,32]. Only the genome of *Ruminococcus flavefaciens* FD-1 is known currently to contain more dockerin-containing genes (>220) [26,33]. Unlike the *R. flavefaciens* dockerins, which are classified into 6 major groups and 11 subgroups [33], the *A. cellulolyticus* dockerins are highly similar, with the exception of six dockerins located downstream of an X module. These latter dockerins





have distinctive sequence features compared to the rest of the *A. cellulolyticus* dockerins. Their X-modules are of family X60 [34], which display significant sequence similarity with the X-module at the C-terminus of the *C. thermocellum* CipA scaffoldin. Indeed, several of these X-dockerin pairs are found at the C-terminus of *A. cellulolyticus* scaffoldins (ScaA, ScaP and ScaI). Interestingly, ScaI protein contains an X-dockerin modular dyad with a truncated type-II dockerin at its C-terminus.

The characteristic sequence conservation profile [35-37] of the *A. cellulolyticus* dockerin module is shown in Figure 3. The sequence similarity among *A. cellulolyticus* dockerin modules is 53% on average (73% for the most similar dockerins pairs, with no two identical dockerins). Like the dockerins in *C. thermocellum* and unlike those of *R. flavefaciens*, each *A. cellulolyticus* dockerin module contains two canonical Ca<sup>2+</sup> binding repeats, followed by putative helices and linkers. Examination of the



**Figure 3 Sequence conservation pattern of dockerin modules.** The two internal dockerin repeats of *A. cellulolyticus* (based on 137 sequences) and *C. thermocellum* (71 sequences) are represented by sequence logos. Positions of calcium binding residues are shown in cyan, and putative recognition residues are shown in yellow.

putative “recognition” residues of the dockerins, which may participate in their tight binding interface with cohesins, shows a conserved pattern of the two repeated segments wherein S(I/L) residues occupy positions 10 and 11, R(X) positions 17 and 18, and a highly conserved G in position 22 (Figure 3, in yellow). The corresponding positions in the *C. thermocellum* dockerins are S(T/S), K(R/K) and K/R/G, respectively. Position 18 is much less conserved in the *A. cellulolyticus* dockerins than those of *C. thermocellum*, whereas the reverse is true for position 22. Some modifications are evident in position 11 of the *A. cellulolyticus* dockerin sequences. For example, the ScaK scaffoldin contains an N-terminal dockerin with an Asn residue in position 11 of its first dockerin repeat. ScaB dockerin contains Asn residues in both repeats, and instead of the conserved Asn in position 9 it contains a positively charged Lys or Arg residue. In the case of ScaB, these modifications lead to different specificity characteristics, as the dockerin binds selectively to the cohesins of ScaC [18].

#### Diversity of dockerin-containing enzymes

*A. cellulolyticus* grows on amorphous and crystalline forms of cellulose, xylans, and cellobiose [38,39]; the bacterium can also be adapted to grow on glucose and xylose [13,40]. Consequently, it was presumed in these early

works that the bacterium produces endoglucanases, exoglucanases,  $\beta$ -glucosidases and xylanase activities. Indeed, the present study reveals an intricate array of cellulolytic and hemicellulolytic enzymes in the *A. cellulolyticus* genome, capable of hydrolyzing diverse cellulosic substrates to reducing sugars.

The sequence features of the dockerin-containing enzymes of *A. cellulolyticus* were assessed using the following approach: (i) Like the cohesin-bearing proteins, the dockerin-containing proteins are multimodular in nature, composed of more than one type of module (catalytic, structural, etc.) and sometimes more than one repeat of the same module. The different modular types were therefore enumerated, in order to determine their general distribution among the *A. cellulolyticus* proteins. (ii) Where appropriate, we distinguished between cellulosomal (i.e., those that harbor a dockerin) and non-cellulosomal (without a dockerin) proteins. (iii) We compared the *A. cellulolyticus* proteins with those of *C. thermocellum*.

Among the 143 dockerin-containing proteins, about half (63 proteins) contain one or more known carbohydrate-active CAZyme module(s) [41], and their composition is presented in Table 1 and Additional file 1: Table S1. Because of the multimodular nature of the proteins, some of them contain more than one type of catalytic

**Table 1 Comparative distribution of dockerin-containing CAZyme modules in *A. cellulolyticus* vs. *C. thermocellum***

<b>A. Glycoside Hydrolases families</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>5</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>13</b>	<b>15</b>	<b>16</b>	<b>18</b>	<b>19</b>	<b>23</b>	<b>26</b>	<b>30</b>	<b>39</b>	<b>43</b>	<b>44</b>	<b>48</b>	<b>51</b>	<b>53</b>	<b>59</b>	<b>74</b>	<b>77</b>	<b>81</b>	<b>94</b>	<b>105</b>	<b>116</b>	<b>124</b>	<b>Total</b>
<i>A. cellulolyticus</i>	Genome-wide	2	1	3	16	4	21	4	1	3	1	1	5	2	2	5	3	—	4	1	2	—	1	1	1	1	1	3	1	1	1	92
	Dockerin-containing proteins	—	1	—	12	3	19	4	1	—	—	1	—	—	—	4	4	—	4	1	1	—	1	1	1	—	1	—	1	1	1	62
<i>C. thermocellum</i>	Genome-wide	2	1	2	10	1	16	6	1	2	1	2	4	—	2	3	2	1	6	1	2	1	1	—	1	—	1	3	—	—	1	73
	Dockerin-containing proteins	—	1	—	8	1	15	3	1	—	—	1	1	—	—	3	2	1	5	1	1	—	1	—	1	—	1	—	—	—	1	48
<b>B. Polysaccharide Lyases families</b>		<b>1</b>	<b>9</b>	<b>11</b>																												
<i>A. cellulolyticus</i>	Genome-wide	1	1	3																			5									
	Dockerin-containing proteins	1	1	3																			5									
<i>C. thermocellum</i>	Genome-wide	2	1	1																			4									
	Dockerin-containing proteins	2	1	1																			4									
<b>C. Carbohydrate Esterases families</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>12</b>	<b>14</b>																					
<i>A. cellulolyticus</i>	Genome-wide	2	1	3	4	1	—	1	1	7	—															20						
	Dockerin-containing proteins	2	1	3	1	1	—	1	—	4 (6)	—															13						
<i>C. thermocellum</i>	Genome-wide	3	1	2	3	—	1	1	1	2	1															15						
	Dockerin-containing proteins	3	1	1 (2)	1	—	—	1	—	1 (2)	—															8						

Numbers represent proteins which contain one or more modules of the different protein families (glycoside hydrolases, polysaccharide lyases and carbohydrate esterases) as were identified by CAZy. The number of proteins are compared between cellulosomal and non-cellulosomal (genome-wide) proteins. Data are provided for both species. Numbers of modules which appear more than once in the same protein are shown parenthetically.

module, therefore the total sum of catalytic modules in the 63 enzymes is 80 in Table 1 (62 GH-, 13 CE- and/or 5 PL-containing enzymes). Of the 92 GHs, about two-thirds are equipped with dockerins, suggesting that they are recruited to the cellulosome and may thus play a critical role in biomass degradation. Interestingly, the percentage of dockerin-containing GHs in the *A. cellulolyticus* genome is almost identical to that of *C. thermocellum*. The 62 dockerin-containing GHs belong to 19 different families according to the CAZy database (Table 1). As in all known cellulosomes produced by other species, the *A. cellulolyticus* cellulosome contains a single distinctive GH48 enzyme. As in *C. thermocellum*, the *A. cellulolyticus* genome also codes for a second, non-cellulosomal GH48-containing cellulase, as opposed to other characterized cellulosome-producing species that possess only one cellulosomal enzyme. The most abundant GH family is represented by the GH9 enzymes, again like in the *C. thermocellum* cellulosome. This is followed by the GH5 enzymes which are also numerous in both cellulosome-producing species. Of the 21 GH9 enzymes, 10 exhibit a GH9-CBM3 motif that would potentially modulate the activity as in *C. thermocellum* and other cellulolytic bacteria [42-45]. In addition, there are three enzymes that show an extended GH9-CBM3-CBM3 motif, compared to two such enzymes in *C. thermocellum* [46].

In one third of the dockerin-containing proteins (46 proteins) we identified modules which are predicted to be associated with extra-cellular proteins (i.e., FN3 modules, Leu-rich repeats, RhsA and PKD domains, see Table 2). Some of these modules are conserved in sequence, but their function is still unknown; some may represent a yet undiscovered enzyme. In this regard, a *C. thermocellum* dockerin-containing protein of previously unknown function was recently demonstrated to be a cellulase [47]. The dockerin-containing proteins of *A. cellulolyticus* are more enriched with such structural and unknown modules than those of *C. thermocellum* (Table 2).

Many of the GH or CE catalytic modules in the multi-modular proteins are associated with CBMs. In the case of a non-cellulosomal protein, a CBM may serve to deliver the parent catalytic module to a preferred site on the polysaccharide substrate.

Otherwise, an appended CBM may serve to modulate directly the hydrolytic properties of the catalytic module. Table 3 shows the number and distribution of such proteins in the genomes of both bacteria, *A. cellulolyticus* and *C. thermocellum*. Interestingly, 38 of the dockerin-containing enzymes in *A. cellulolyticus* consist of both a catalytic module and a CBM, most of the latter mostly families 3 and 6 (Table 3). In addition, another 12 non-cellulosomal enzymes contain an appended CBM. Although *A. cellulolyticus* contains approximately double the number of dockerin-containing proteins as *C. thermocellum*, the two species have the same number of CBM-appended enzymes (Table 2), and their distribution into different CAZy families largely overlaps.

Even more intriguing are the 10 multi-functional enzymes of *A. cellulolyticus*, which harbor a combination of at least two catalytic modules, including one or two GHs, CEs, PLs and/or glycosyl transferases (GTs), on the same polypeptide (Table 4). In *A. cellulolyticus*, some of these enzymes do not contain a dockerin module. In contrast, *C. thermocellum* codes for 8 multi-functional dockerin-containing enzymes, and *Ruminococcus flavefaciens* FD-1 codes for 18 dockerin-containing multi-functional enzymes. As stated in an earlier section, both genomes encode for two GH48 enzymes – one cellulosomal and another non-cellulosomal. In *C. thermocellum*, there are two separate non-cellulosomal enzymes – Cel48Y (GH48-CBM3b) and the other Cel9I (GH9-CBM3c-CBM3b), whereas in *A. cellulolyticus* the two catalytic modules are fused together into a single polypeptide chain that share a single cellulose-binding CBM3b, thus forming a multi-functional non-cellulosomal enzyme (GH48-GH9-CBM3c-CBM3b, [GenBank:ZP\_09464448]).

#### Putative cellulosome-related regulatory elements

It is clear that such an elaborate cellulosome system in *A. cellulolyticus* would require a regulatory mechanism by which the bacterium controls expression of its cellulosomal genes. One possible regulator may be inherent in the two types of cohesin modules (i.e., type I and type II), which, like in *C. thermocellum*, signifies

**Table 2 Summary of protein modules in cellulosomal proteins**

	Cohesins	Dockerin-containing proteins <sup>a</sup>	Modules in dockerin-containing proteins				
			Catalytic modules <sup>b</sup>	Structural modules <sup>c</sup>	CBMs <sup>b</sup>	Cohesins	Others <sup>d</sup>
<i>A. cellulolyticus</i>	41	143 (5)	74	46	53	13	36
<i>C. thermocellum</i>	29	73 (3)	51	25	47	9	3

<sup>a</sup> Numbers in parenthesis indicate the number of X60-dockerin modular pairs in the given species.

<sup>b</sup> Catalytic modules, such as: GH, PL, CE, and CBMs according to CAZy (<http://www.cazy.org/>).

<sup>c</sup> Structural domains are defined in Pfam (<http://pfam.sanger.ac.uk/>), such as: FN3, GDSL, SNGH, CoH, TRX-like, Kelch-like, SLH, RshA, LRR, PKD.

<sup>d</sup> Others include Pfam domains such as peptidases, serpins, DUF303, DUF1565, DUF3237.



**Table 3 Genome-wide co-occurrence of CBMs together with either GH or CE modules in *A. cellulolyticus* vs. *C. thermocellum***

CBM families	3	4	6	9	11	13	22	23	27	30	32	34	35	42	44	48	50	54	62	
<b>GH 5</b>	3/1		2/0		0/1	1/0					1/1									1/0
<b>GH 9</b>	19/10	0/2								0/1					1/0					
<b>GH 10</b>			2/1	2/2			4/5													
<b>GH 11</b>			1/1																	
<b>GH 13</b>												0/1				2/1				
<b>GH 16</b>		0/4																		0/1
<b>GH 18</b>	1/0																			2/2
<b>GH 26</b>					0/1			1/0	1/0				3/2							
<b>GH 30</b>			1/1											0/1						
<b>GH 39</b>													0/2							
<b>GH 43</b>			2/3			1/1								0/3						
<b>GH 44</b>										0/1					1/1					
<b>GH 48</b>	1/1																			
<b>CE 1</b>			2/1																	
<b>CE 4</b>			1/1																	
<b>CE 6</b>			1/0			1/0														1/0
<b>CE 12</b>														4/2						

The number of proteins with the combination of the specified modules is noted in the genomes of *A. cellulolyticus* (left) vs. *C. thermocellum* (right).

at least two divergent specificities of cohesin-dockerin interaction in this species.

Recently, a distinctive system of cellulosome gene regulation was proposed. A carbohydrate-sensing mechanism was described in *C. thermocellum* [48-50], suggesting that a set of putative  $\sigma$  and anti- $\sigma$  factors are activated by extracellular polysaccharides. Thus, the different components of the cellulosic biomass would be detected extracellularly by corresponding RsgI-borne binding elements (CBMs, GHs, etc.), and appropriate signals are transmitted intracellularly. This in turn was proposed to disassociate the interaction between the intracellular portions of the RsgI-like proteins and complementary  $\sigma^1$ -like factors, resulting in the release of the  $\sigma^1$ s, followed by their association with RNA polymerase and transcription of corresponding genes involved in cellulose utilization. Interestingly, analysis of the other known cellulosome-producing bacterial genomes (e.g., *C. cellulolyticum* and *C. cellulovorans*) revealed only a single RsgI-like protein, which lacks a recognizable C-terminal binding element. It therefore appeared that an extensive RsgI-mediated carbohydrate-sensing mechanism is restricted to *C. thermocellum*.

It was thus of interest to evaluate the status of the RsgI-like proteins in *A. cellulolyticus*. Indeed, analysis of the genome revealed multiple copies of genes coding for  $\sigma^1$ -like factors and their cognate membrane-associated

RsgI-like (anti- $\sigma^1$ ) factors, which may be involved in regulatory mechanisms of cellulosomal and related cellulase genes. Twelve putative  $\sigma^1$ /RsgI-like proteins were detected in the *A. cellulolyticus* genome (Table 5), as opposed to the eight in *C. thermocellum*. The *A. cellulolyticus* RsgI-like proteins contain predicted C-terminal modules such as CBM3, CBM42, CBM35, PA14-like, but none appeared to contain a GH module like the ones detected in *C. thermocellum* [50]. Significantly, most of the putative  $\sigma^1$ -like proteins of *A. cellulolyticus* have orthologs in *C. thermocellum*, some of which have been validated experimentally.

For example, the ability of  $\sigma^{11}$  of *C. thermocellum* to activate the promoters of *sigI1* and a family 48 cellulase, *celS*, was demonstrated in vitro [49]. In addition, the CBMs were shown to bind selectively to typical plant cell wall polysaccharides [48]. Interestingly, genes encoding the  $\sigma^1$ /RsgI regulatory systems are often found in genomic loci, where they are associated with other genes encoding dockerin- and cohesin-containing proteins (e.g., *celE*, *cel124*, *cel8A*, *scaF* etc.).

The multiple regulatory factors which we identified in *A. cellulolyticus* thus mirror the extensive regulatory system described previously in *C. thermocellum*, and may control the expression levels of cellulosomal and non-cellulosomal genes to reflect changes in the plant cell-wall substrates during the process of decomposition. Moreover, some of these factors may govern processes

**Table 4 Multifunctional proteins in *A. cellulolyticus* vs. *C. thermocellum***

<i>A. cellulolyticus</i>		<i>C. thermocellum</i>	
<b>A. Homologous cellulosomal enzymes</b>			
<b>GH11</b> -CBM6-Doc- <b>CE4</b>	ZP_09464944	<b>GH11</b> -CBM6-Doc- <b>CE4</b>	Cthe_2972;XynA/U
<b>PL1</b> -Doc- <b>PL9</b>	ZP_09465691	<b>PL1</b> -Doc-CBM35- <b>PL9</b>	Cthe_2179
<b>CE12</b> -Doc- <b>CBM35-CE12</b>	ZP_09463564	<b>CE12</b> -Doc- <b>CBM35-CE12</b>	Cthe_3141
<b>CE12</b> -Doc- <b>CBM35-CE12</b>	ZP_09465667		
<b>B. Non-homologous cellulosomal enzymes</b>			
<b>GH5</b> -CBM6-CBM13-CBM62-Doc- <b>CE6</b>	ZP_09463297	CBM30- <b>GH9-GH44</b> -Doc-CBM44	Cthe_0624;CelJ
<b>GH5</b> -Doc- <b>CE2</b>	ZP_09464730	CBM22- <b>GH10</b> -CBM22-Doc- <b>CE1</b>	Cthe_0912;XynY
<b>CE1</b> -CBM6-Doc- <b>GH10</b>	ZP_09465552	<b>GH26-GH5</b> -CBM11-Doc	Cthe_1472;CelH
		<b>GH30</b> -CBM42- <b>GH43</b> -Doc	Cthe_2139
		<b>CE3-CE3</b> -Doc	Cthe_0798
<b>C. Non-cellulosomal enzymes</b>			
<b>GH48-GH9</b> -CBM3c-CBM3b	ZP_09464448		
<b>GH18-CE4</b> -GT2	ZP_09465738		
GT84- <b>GH94</b>	ZP_09462312		

Domain architectures and their corresponding accession numbers or names are listed above. Catalytic modules are marked in bold. GH, glycoside hydrolase; PL, polysaccharide lyase; CE, carbohydrate esterase; CBM, carbohydrate-binding module; GT, glycosyl transferase; Doc, dockerin; numbers indicate family of the indicated module.

in the bacterium, which are not directly involved in plant cell wall degradation.

## Conclusions

Early electron microscopy observations of *A. cellulolyticus* demonstrated its particularly elaborate cell surface ultrastructure and its cellulose-degrading activities [16,51]. The availability of its genome sequence has enabled a better appreciation of the complex and modular nature of its cellulosome. Compared to *C. thermocellum*, the cellulosomal architecture of *A. cellulolyticus* is more extensive, encoding twice the number of cohesin- and dockerin-containing proteins, with previously undescribed combinations of protein modules. Yet, certain elements of the basic structural scaffolds, which dictate the assembly of the various functional carbohydrate-degrading enzymes, are maintained in both species. In addition, both species exhibit elaborate cell-anchoring and gene-regulation systems. Interestingly, the multiplicity of  $\sigma^1$ /RsgI-like proteins may be characteristic of cellulosome-producing bacteria that contain multiple- scaffoldin gene clusters, like *A. cellulolyticus* and *C. thermocellum*, as opposed to those like *C. cellulolyticum*, that contain enzyme-linked gene clusters.

This work provides a blueprint for understanding the cellulosome system of this intriguing cellulose-degrading bacterium and paves the way for studying the specific role of its cellulosomal protein components in the degradation of plant cell-wall carbohydrates. It is clear that

the bacterium utilizes a sophisticated system for efficient hydrolysis of crystalline cellulose of the plant cell wall. The cohesin-containing proteins of *A. cellulolyticus* present a broader diversity and modularity than those of *C. thermocellum*, where cohesins are associated in unconventional modular combinations, and their functional roles are yet to be defined.

## Methods

### Genomes source

Draft genome sequences of *Acetivibrio cellulolyticus* CD2 (DSM 1870, ATCC 33288) (30 Dec. 2011), and *Clostridium thermocellum* ATCC 27405 (16 Feb. 2007) were obtained from GenBank (accession: AEDB00000000 and CP000568, respectively). Assembly of *A. cellulolyticus* genome was approached by a combination of sequencing methods, using Sanger, 454-Titanium, 454 Titanium Paired-end and Solexa Paired-end technologies, as detailed in Hemme et al. [22]. The genome assembled into 112 contigs with an average coverage depth of  $\times 71.9 \pm 6.3$  (interval of depths 9 – 111). Protocols of the *A. cellulolyticus* sequencing methods, assemblies and annotation are detailed in Land et al. [52].

### Sequence identification of cohesins and dockerins

BLAST [53] searches were applied on *A. cellulolyticus* DNA contigs and predicted proteins, using sequences of known cohesin and dockerin modules as queries. All hits above E-value of  $10^{-4}$  were retrieved and inspected individually, by examining their characteristic sequence

**Table 5 Putative  $\sigma^I$  and anti- $\sigma^I$  regulatory factors in *Acetivibrio cellulolyticus* CD2**

<i>sigI</i> -like gene	<i>rsgI</i> -like pair	C-terminal sensing domain	Ortholog in <i>C. thermocellum</i>
ZP_09464729	ZP_09464728	CBM3	Cthe_0403
ZP_09464331	ZP_09464330	CBM3	Cthe_0058
ZP_09466014	ZP_09466013	CBM3	Cthe_0268
ZP_09463653	ZP_09463652	CBM3	Cthe_0058
ZP_09461804	ZP_09461805	CBM42	Cthe_1272
ZP_09463236	ZP_09463235	PA14, CBM35	Cthe_0315
ZP_09464238	ZP_09464237	PA14, PA14	Cthe_1272
ZP_09464575	ZP_09464574	unknown	Cthe_0403
ZP_09464240	ZP_09464239	S1/S6 peptidase	Cthe_0058
ZP_09463889	ZP_09463888	unknown	Cthe_2521
ZP_09466630	ZP_09466631	unknown	Cthe_2974
ZP_09465751	ZP_09465752	unknown	Cthe_2974

features. Obvious dockerin modules were expected to contain two Ca<sup>2+</sup>-binding repeats, putative helices and linker regions. Low-scoring hits of dockerins and cohesins were examined by comparing them against known dockerin or cohesin sequences, respectively. Sequence logos of dockerins were created with Weblogo v.2.8.2 (<http://weblogo.berkeley.edu/>) [54]. Multiple sequence alignment was obtained using CLASTALW [55], with manual corrections when needed.

The scaffoldin genes from *A. cellulolyticus* ATCC 33288 which were manually sequenced [17-19] are ScaA, [GenBank: AF155197]; ScaB, [GenBank: AY221112]; ScaC, [GenBank: AY221113], ScaD, [GenBank: AY221114]). The cohesin dendrogram was generated using PhyML algorithms (with LG substitution model, and default parameters of the Approximate Likelihood-Ratio test) [56] and visualized using TreeView [57].

#### Annotation of dockerin-containing enzymes

Dockerin-containing proteins of *A. cellulolyticus* CD2 and *C. thermocellum* ATCC 27405 were annotated by CAZy database (<http://www.cazy.org>) [41], in order to bioinformatically analyze their catalytic modules. This includes identification of the catalytic modules and their classification into family types, according to sequence conservation, for glycoside hydrolases, carbohydrate esterases, polysaccharide lyases, carbohydrate-binding modules and glycosyl transferases. Additional conserved domains of the proteins were analyzed using the CD-search website (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and the Pfam database (<http://pfam.sanger.ac.uk/>). Putative cellulosome-related regulatory elements were identified by BLAST searches and sequence similarity using known elements from *C. thermocellum* as queries [48-50].

#### Additional file

**Additional file 1: Table S1.** Cellulosomal and non-cellulosomal CAZyme proteins in *A. cellulolyticus*. The modular architecture of the indicated proteins show *only* the CAZy-related modules: GH, glycoside hydrolase; PL, polysaccharide lyase; CE, carbohydrate esterase; CBM, carbohydrate-binding module; Doc, dockerin; Coh, cohesin, SLH, S-layer homology modules. Numbers indicate family of the indicated module. **A.** Cohesin-containing proteins. **B.** Dockerin-containing proteins. **C.** Non-cellulosomal CAZymes

#### Abbreviations

CAZy: Carbohydrate-active enzymes; GH: Glycoside hydrolase; CE: Carbohydrate esterase; PL: Polysaccharide lyase; CBM: Carbohydrate-binding module; GT: Glycosyl transferases; CBM: Carbohydrate-binding module; SLH: S-layer homology; X-doc: X-dockerin; Sca: Scaffoldin.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

We thank Yuval Hamberg for helpful discussions. Parts of this research were supported by grants from the Israel Science Foundation (Grant nos 966/09, 159/07 and 24/11), by The Israeli Centers of Research Excellence (I-CORE) program, (Center No. 152/11), by The Alternative Energy Research Initiative Bioenergy Consortium at the Weizmann Institute of Science and the China-Israel Scientific Research Cooperation. E.A.B. is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry.

#### Author details

<sup>1</sup>Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel. <sup>2</sup>Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv, Tel Aviv, Israel. <sup>3</sup>Architecture et Fonction des Macromolécules Biologiques, CNRS and Université Aix-Marseille I & II, Marseille, France. <sup>4</sup>Department of Botany and Microbiology, and Institute for Environmental Genomics, University of Oklahoma, Norman, OK, USA.

#### Authors' contributions

BD and EAB conceived of the project and wrote the manuscript. BD, IB, BH and PC analyzed the genome data. CLH, YH and JZ sequenced the genome. BD, RL and EAB wrote the paper. All authors read and approved the final manuscript.

Received: 21 February 2012 Accepted: 30 May 2012  
Published: 30 May 2012

## References

1. Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, et al: **How biotech can transform biofuels.** *Nature Biotechnol* 2008, **26**:169–172.
2. Himmel ME, Xu Q, Luo Y, Ding S-Y, Lamed R, Bayer EA: **Microbial enzyme systems for biomass conversion: Emerging paradigms.** *Biofuels* 2010, **1**:323–341.
3. Bayer EA, Chanzy H, Lamed R, Shoham Y: **Cellulose, cellulases and cellulosomes.** *Curr Opin Struct Biol* 1998, **8**:548–557.
4. Bayer EA, Belaich J-P, Shoham Y, Lamed R: **The cellulosomes: Multi-enzyme machines for degradation of plant cell wall polysaccharides.** *Annu Rev Microbiol* 2004, **58**:521–554.
5. Bayer EA, Lamed R, White BA, Flint HJ: **From cellulosomes to cellulosomes.** *Chem Rec* 2008, **8**:364–377.
6. Demain AL, Newcomb M, Wu JH: **Cellulase, clostridia, and ethanol.** *Microbiol Mol Biol Rev* 2005, **69**(1):124–154.
7. Doi RH, Kosugi A: **Cellulosomes: plant-cell-wall-degrading enzyme complexes.** *Nat Rev Microbiol* 2004, **2**(7):541–551.
8. Bayer EA, Shimon LJW, Lamed R, Shoham Y: **Cellulosomes: structure and ultrastructure.** *J Struct Biol* 1998, **124**:221–234.
9. Bayer EA, Shoham Y, Lamed R: **Cellulose-decomposing prokaryotes and their enzyme systems.** In *In: The Prokaryotes, Third Edition*. Edited by Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E. New York: Springer; 2006:578–617. vol. 2.
10. Raman B, Pan C, Hurst GB, Rodriguez M, McKeown CK, Lankford PK, Samatova NF, Mielenz JR: **Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis.** *PLoS One* 2009, **4**:e5271.
11. Schwarz WH, Zverlov VV, Bahl H: **Extracellular glycosyl hydrolases from *Clostridia*.** *Advan Appl Microbiol* 2004, **56**:215–261.
12. Khan AW: **Cellulolytic enzyme system of *Acetivibrio cellulolyticus*, a newly isolated anaerobe.** *J Gen Microbiol* 1980, **121**:499–502.
13. Patel GB, Khan AW, Agnew BJ, Colvin JR: **Isolation and characterization of an anaerobic cellulolytic microorganism, *Acetivibrio cellulolyticus*, gen. nov., sp. nov.** *Int J Syst Bacteriol* 1980, **30**:179–185.
14. Saddler JN, Khan AW: **Cellulase production by *Acetivibrio cellulolyticus*.** *Can J Microbiol* 1980, **26**:760–765.
15. Saddler JN, Khan AW: **Cellulolytic enzyme system of *Acetivibrio cellulolyticus*.** *Can J Microbiol* 1981, **27**:288–294.
16. Lamed R, Naimark J, Morgenstern E, Bayer EA: **Specialized cell surface structures in cellulolytic bacteria.** *J Bacteriol* 1987, **169**:3792–3800.
17. Ding S-Y, Bayer EA, Steiner D, Shoham Y, Lamed R: **A novel cellulosomal scaffoldin from *Acetivibrio cellulolyticus* that contains a family-9 glycosyl hydrolase.** *J Bacteriol* 1999, **181**:6720–6729.
18. Xu Q, Gao W, Ding S-Y, Kenig R, Shoham Y, Bayer EA, Lamed R: **The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell-surface anchoring protein.** *J Bacteriol* 2003, **185**:4548–4557.
19. Xu Q, Barak Y, Kenig R, Shoham Y, Bayer EA, Lamed R: **A novel *Acetivibrio cellulolyticus* anchoring scaffoldin that bears divergent cohesins.** *J Bacteriol* 2004, **186**:5782–5789.
20. Noach I, Alber O, Bayer EA, Lamed R, Levy-Assaraf M, Shimon LJW, Frolov F: **Crystallization and preliminary X-ray analysis of *Acetivibrio cellulolyticus* cellulosomal type II cohesin module: Two versions having different linker lengths.** *Acta Cryst* 2008, **F64**:58–61.
21. Noach I, Frolov F, Alber O, Lamed R, Shimon LJW, Bayer EA: **Inter-modular linker flexibility revealed from crystal structures of adjacent cellulosomal cohesins of *Acetivibrio cellulolyticus*.** *J Mol Biol* 2009, **391**:86–97.
22. Hemme CL, Mouttaki H, Lee YJ, Zhang G, Goodwin L, Lucas S, Copeland A, Lapidus A, Glavina del Rio T, Tice H, et al: **Sequencing of multiple clostridial genomes related to biomass conversion and biofuel production.** *J Bacteriol* 2010, **192**(24):6494–6496.
23. Blouzard JC, Coutinho PM, Fierobe HP, Henrissat B, Lignon S, Tardif C, Pagès S, de Philip P: **Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses.** *Proteomics* 2010, **10**:541–554.
24. Brown SD, Raman B, McKeown CK, Kale SP, He ZL, Mielenz JR: **Construction and evaluation of a *Clostridium thermocellum* ATCC 27405 whole-genome oligonucleotide microarray.** *Appl Biochem Biotechnol* 2007, **137**:663–674.
25. Tamaru Y, Miyake H, Kuroda K, Nakanishi A, Kawade Y, Yamamoto K, Uemura M, Fujita Y, Doi RH, Ueda M: **Genome sequence of the cellulosome-producing mesophilic organism *Clostridium cellulovorans* 743B.** *J Bacteriol* 2010, **192**:901–902.
26. Berg Miller ME, Antonopoulos DA, Rincon MT, Band M, Bari A, Akaikol T, Hernandez A, Kim R, Liu L, Thimmapuram J, et al: **Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1.** *PLoS One* 2009, **4**:e6650.
27. Chauvaux S, Matuschek M, Béguin P: **Distinct affinity of binding sites for S-layer homologous domains in *Clostridium thermocellum* and *Bacillus anthracis* cell envelopes.** *J Bacteriol* 1999, **181**:2455–2458.
28. Lemaire M, Ohayon H, Gounon P, Fujino T, Béguin P: **OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope.** *J Bacteriol* 1995, **177**:2451–2459.
29. Lin C, Urbance JW, Stahl DA: ***Acetivibrio cellulolyticus* and *Bacteroides cellulosolvans* are members of the greater clostridial assemblage.** *FEMS Microbiol Lett* 1994, **124**:151–155.
30. Pinheiro BA, Gilbert HJ, Sakka K, Fernandes VO, Prates JA, Alves VD, Bolam DN, Ferreira LM, Fontes CM: **Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*.** *Biochem J* 2009, **424**(3):375–384.
31. Fendri I, Tardif C, Fierobe HP, Lignon S, Valette O, Pagès S, Perret S: **The cellulosomes from *Clostridium cellulolyticum*: identification of new components and synergies between complexes.** *FEBS Journal* 2009, **276**:3076–3086. **The cellulosomes from *Clostridium cellulolyticum*: identification of new components and synergies between complexes.** *FEBS J* 2009, **276**:3076–3086.
32. Bayer EA, Henrissat B, Lamed R: **The cellulosome: A natural bacterial strategy to combat biomass recalcitrance.** In *In: Biomass Recalcitrance*. Edited by Himmel ME. London: Blackwell; 2008:407–426.
33. Rincon MT, Dassa B, Flint HJ, Travis AR, Jindou S, Borovok I, Lamed R, Bayer EA, Henrissat B, Coutinho PM, et al: **Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD1.** *PLoS One* 2010, **5**:e12476.
34. Adams JJ, Jang CJ, Spencer HL, Elliott M, Smith SP: **Expression, purification and structural characterization of the scaffoldin hydrophilic X-module from the cellulosome of *Clostridium thermocellum*.** *Protein Expr Purif* 2004, **38**(2):258–263.
35. Pagès S, Belaich A, Belaich J-P, Morag E, Lamed R, Shoham Y, Bayer EA: **Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: Prediction of specificity determinants of the dockerin domain.** *Proteins* 1997, **29**:517–527.
36. Mechaly A, Yaron S, Lamed R, Fierobe H-P, Belaich A, Belaich J-P, Shoham Y, Bayer EA: **Cohesin-dockerin recognition in cellulosome assembly: Experiment versus hypothesis.** *Proteins* 2000, **39**:170–177.
37. Mechaly A, Fierobe H-P, Belaich A, Belaich J-P, Lamed R, Shoham Y, Bayer EA: **Cohesin-dockerin interaction in cellulosome assembly: A single hydroxyl group of a dockerin domain distinguishes between non-recognition and high-affinity recognition.** *J Biol Chem* 2001, **276**:9883–9888. and Erratum 19678.
38. Khan AW, Meek E, Sowden LC, Colvin JR: **Emendation of the genus *Acetivibrio* and description of *Acetivibrio cellulosolvans* sp. nov., a nonmotile cellulolytic mesophile.** *Int J Syst Bacteriol* 1984, **34**:419–422.
39. Sanchez CR, Peres CS, Barbosa HR: **Growth and endoglucanase activity of *Acetivibrio cellulolyticus* grown in three different cellulosic substrates.** *Rev Microbiol* 1999, **30**:310–314.
40. Murray WD: ***Acetivibrio cellulosolvans* is a synonym for *Acetivibrio cellulolyticus*: Emendation of the genus *Acetivibrio*.** *Int J Syst Bacteriol* 1986, **36**:314–316.
41. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active Enzymes database (CAZy): an expert resource for glycogenomics.** *Nucl Acids Res* 2009, **37**:D233–238.
42. Gal L, Gaudin C, Belaich A, Pagès S, Tardif C, Belaich J-P: **CelG from *Clostridium cellulolyticum*: a multidomain endoglucanase acting efficiently on crystalline cellulose.** *J Bacteriol* 1997, **179**:6595–6601.

43. Irwin D, Shin D-H, Zhang S, Barr BK, Sakon J, Karplus PA, Wilson DB: Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J Bacteriol* 1998, **180**:1709–1714.
44. Sakon J, Irwin D, Wilson DB, Karplus PA: Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Struct Biol* 1997, **4**:810–818.
45. Gilad R, Rabinovich L, Yaron S, Bayer EA, Lamed R, Gilbert HJ, Shoham Y: Cell, a non-cellulosomal family-9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. *J Bacteriol* 2003, **185**:391–398.
46. Jindou S, Xu Q, Kenig R, Shoham Y, Bayer EA, Lamed R: Novel architectural theme of family-9 glycoside hydrolases identified in cellulosomal enzymes of *Acetivibrio cellulolyticus* and *Clostridium thermocellum*. *FEMS Microbiol Lett* 2006, **254**:308–316.
47. Brás JL, Cartmell A, Carvalho AL, Verzé G, Bayer EA, Vazana Y, Correia MA, Prates JA, Ratnaparkhe S, Boraston AB, et al: Structural insights into a unique cellulase fold and mechanism of cellulose hydrolysis. *Proc Natl Acad Sci USA* 2011, **108**:5237–5242.
48. Kahel-Raifer H, Jindou S, Bahari L, Nataf Y, Shoham Y, Bayer EA, Borovok I, Lamed R: The unique set of putative membrane-associated anti- $\sigma$  factors in *Clostridium thermocellum* suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation. *FEMS Microbiol Lett* 2010, **308**:84–93.
49. Nataf Y, Bahari L, Kahel-Raifer H, Borovok I, Lamed R, Bayer EA, Sonenshein AL, Shoham Y: *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternate sigma factors. *Proc Natl Acad Sci USA* 2010, **107**:18646–18651.
50. Bahari L, Gilad Y, Borovok I, Dassa B, Kahel-Raifer H, Jindou S, Nataf Y, Shoham Y, Lamed R, Bayer EA: Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. *J Ind Microbiol Biotechnol* 2011, **38**:825–832.
51. Lamed R, Naimark J, Morgenstern E, Bayer EA: Scanning electron microscopic delineation of bacterial surface topology using cationized ferritin. *J Microbiol Methods* 1987, **7**:233–240.
52. Land M, Pukall R, Abt B, Goker M, Rohde M, Glavina Del Rio T, Tice H, Copeland A, Cheng JF, Lucas S, et al: Complete genome sequence of *Beutenbergia cavernae* type strain (HKI 0122). *Stand Genomic Sci* 2009, **1**(1):21–28.
53. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 1997, **25**:3389–3402.
54. Crooks GE, Hon G, Chandonia JM, Brenner SE: WebLogo: A sequence logo generator. *Genome Res* 2004, **14**:1188–1190.
55. Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994, **22**:4673–4680.
56. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 2010, **59**:307–321.
57. Page RD: TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 1996, **12**:357–358.

doi:10.1186/1471-2164-13-210

**Cite this article as:** Dassa et al.: Genome-wide analysis of *acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system. *BMC Genomics* 2012 **13**:210.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

