Revealing Nature's Cellulase Diversity: The Digestion Mechanism of *Caldicellulosiruptor bescii* CelA

Roman Brunecky,[1] Markus Alahuhta,[1] Qi Xu,[1] Bryon S. Donohoe,[1] Michael F. Crowley,[1] Irina A. Kataeva,[2] Sung-Jae Yang,[2] Michael G. Resch,[1] Michael W. W. Adams,[2] Vladimir V. Lunin,[1] Michael E. Himmel,[1] Yannick J. Bomble[1]*

Most fungi and bacteria degrade plant cell walls by secreting free, complementary enzymes that hydrolyze cellulose; however, some bacteria use large enzymatic assemblies called cellulosomes, which recruit complementary enzymes to protein scaffolds. The thermophilic bacterium *Caldicellulosiruptor bescii* uses an intermediate strategy, secreting many free cellulases that contain multiple catalytic domains. One of these, CelA, comprises a glycoside hydrolase family 9 and a family 48 catalytic domain, as well as three type III cellulose-binding modules. In the saccharification of a common cellulose standard, Avicel, CelA outperforms mixtures of commercially relevant exo- and endoglucanases. From transmission electron microscopy studies of cellulose after incubation with CelA, we report morphological features that suggest that CelA not only exploits the common surface ablation mechanism driven by general cellulase processivity, but also excavates extensive cavities into the surface of the substrate. These results suggest that nature's repertoire of cellulose digestion paradigms remain only partially discovered and understood.

In nature, most cellulolytic enzymes systems are of two general types: those with noncomplexed cellulases and hemicellulases produced by aerobic fungi and most bacteria (*1*), and those in which polysaccharidases self-assemble onto a common protein scaffold to form large macromolecular assemblies called cellulosomes (*2*, *3*). Only a few anaerobic bacteria and fungi are known to produce cellulosomes. In both cases, the enzymes secreted are primarily equipped with a single catalytic domain. An alternate enzymatic system, in a sense midway between that of the two previous models, is used by the thermophile *Caldicellulosiruptor bescii* (previously *Anaerocellum thermophilum*) (*4*–*6*), and some other bacteria in which the individual cellulases secreted are multimodular, containing multiple binding and catalytic domains. The most abundant enzymes secreted by *C. bescii* are not only multimodular, but possess catalytic domains with different activities (multifunctional). Lochner and co-workers conducted an extensive characterization of the *C. bescii* secretome and concluded that CelA is the dominant cellulase (*5*). CelA is a complex, thermally stable enzyme containing an *N*-terminal family 9 glycoside hydrolase (GH9) endo-β-1,4-glucanase domain, three family 3 carbohydrate-binding modules (CBM3), and a C-terminal GH48 exo-β-1,4-glucanase domain. family 9 and 48 catalytic domains have been considered to be highly synergistic (*7*, *8*). CelA was first isolated and partially characterized in terms of its activity by Zverlov and co-workers (*9*), who reported that CelA demonstrated activity on cellulose, as well as weak activity on xylan. However, they did not report the performance of CelA on biomass and did not propose its mode of action or the complementarity of the two catalytic domains.

C. bescii can operate at temperatures up to 90°C (*5*, *6*), which may explain the advantage of several CBMs in counteracting the loss of binding due to increased temperature. Although *C. bescii* has been well studied at the microbial level, the structural features of its highly thermally stable enzymes have not been explained owing to the lack of structural information for enzymes from the *Caldicellulosiruptor* clade. Currently, large-scale biomass saccharification in bioconversion processes relies exclusively on fungal enzymes operating at 50° to 55°C. *C. bescii* CelA may have several advantages over other fungal and bacterial cellulases for use in biofuels production, namely, its high specific activity and stability at elevated temperatures.

Several recent studies report that the optimal growth temperature for *C. bescii* is near 80°C (*5*, *6*). We examined the cellulolytic performance of purified CelA [supplementary material (SM) text S1 and S2] acting on Avicel, a model cellulose substrate that is generally used to evaluate cellulase action, at 60°, 75°, and 85°C. We compared CelA to a binary mixture containing *Trichoderma reesei* Cel7A exoglucanase, currently the most common exoglucanase found in commercial cellulase preparations (*10*, *11*), and *Acidothermus cellulolyticus* Cel5A endoglucanase (*11*), which mimics the two cellulolytic (endo- and exoglucanase) activities found in CelA. The percentage of glucan converted over a 7-day digestion study is shown in Fig. 1A. CelA retained high activity compared

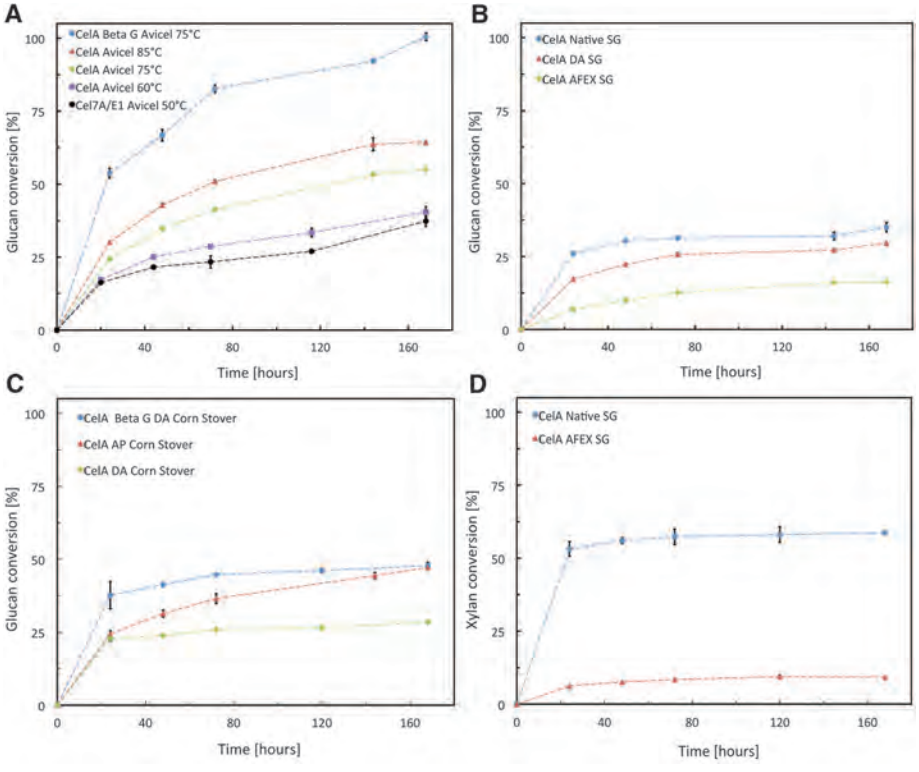

Fig. 1. Activity of CelA. (**A**) Avicel conversion by CelA at different temperatures (15 mg of CelA/g glucan or 14 mg of CelA + 1 mg β-glucosidase/g glucan) compared to Cel7A/E1 (12 mg/g + 3 mg/g glucan, respectively). (**B**) CelA conversion of switchgrass at 15 mg/g glucan. (**C**) Corn stover conversion by CelA at 20 mg/g (AP) or 15mg/g (DA) glucan and improvement of CelA by addition of β-D-glucosidase on DA corn stover (14 mg CelA + 1 mg β-glucosidase/g glucan). (**D**) CelA conversion of xylan from native switchgrass at 15 mg/g glucan. Error bars represent standard deviations from three different experiments.

[1]Biosciences Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, USA. [2]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA.

*Corresponding author. E-mail: yannick.bomble@nrel.gov

to Cel7A at all temperatures tested, converting 60% of glucan at 85°C compared to 28% glucan conversion by Cel7A at its optimal temperature of 50°C (Fig. 1A). The extent of conversion obtained in this study for Cel7A is consistent with those reported by several other research groups (*12*–*14*). Furthermore, the activity of this enzyme acting on Avicel, on a molar basis, is seven times as high as that of the common exo- and endocellulase standard mixture, Cel7A and Cel5A (SM text S3). This high activity may be due to the proximity of a chain-end–forming endoglucanase and an efficient cellobiohydrolase in the same molecule, thereby increasing the intramolecular synergy. These results illustrate that when acting on Avicel, CelA is a far more active single enzyme than the dominant enzyme in today's commercial cellulase formulations, *T. reesei* Cel7A. End-product inhibition of cellulases is a well-known characteristic (*15*), and CelA in particular is known to produce both cellobiose and glucose. To assess this issue, we tested the sensitivity of CelA to cellobiose by adding low loadings of a thermally stable β-D-glucosidase from *Thermotoga maritima* (*16*) to the CelA mix. The addition of β-D-glucosidase markedly improved the performance of CelA acting on Avicel, producing 100% conversion in 7 days (Fig. 1A).

The performance of CelA acting on biomass is predictably lower than when acting on Avicel; CelA performs best on alkaline peroxide (AP)–(*17*, *18*) pretreated corn stover (~50% conversion) (Fig. 1C). When tested on other substrates, such as dilute acid–pretreated switchgrass (DA) (*19*), native switchgrass (SG), and ammonia fiber expansion–treated switchgrass (AFEX) (*20*, *21*) (Fig. 1B), CelA performed better on the untreated substrate even though the entire rationale of pretreatment is to improve enzymatic hydrolysis of biomass substrates. CelA activity was lowest when acting on AFEX-pretreated switchgrass, where the glucan conversion was 20%. Also, the activity of purified CelA on biomass was similar to that of the *C. bescii* purified secretome (ExtP) fraction (fig. S6). The importance of removing end-product inhibition when CelA acts on DA-pretreated corn stover is shown by the results in Fig. 1C, where CelA alone converts only ~30% of the glucan content to sugars and the addition of β-D-glucosidase increases this conversion by almost 75%: a notable improvement, given the relatively low overall enzyme loading.

While performing these digestion experiments, we also observed that CelA could achieve 60% conversion of xylan in native switchgrass, which showcases its potential for use in an industrial process using mild or no pretreatment (Fig. 1D). We tested both catalytic modules from CelA and found that the GH48 is largely responsible for this xylan-degrading activity (SM text S4). To further examine the xylan-degrading ability of CelA, we crystallized the two catalytic domains (GH9 and GH48) of CelA (SM text S5). It is probable that the additional ability to degrade xylans as well as glucans is defined by small changes in the conformational properties of the active sites of glycoside hydrolases. However, closer examination of the CelA GH9 and GH48 structures shows no obvious features that would prevent or favor xylan binding over cellodextrins.

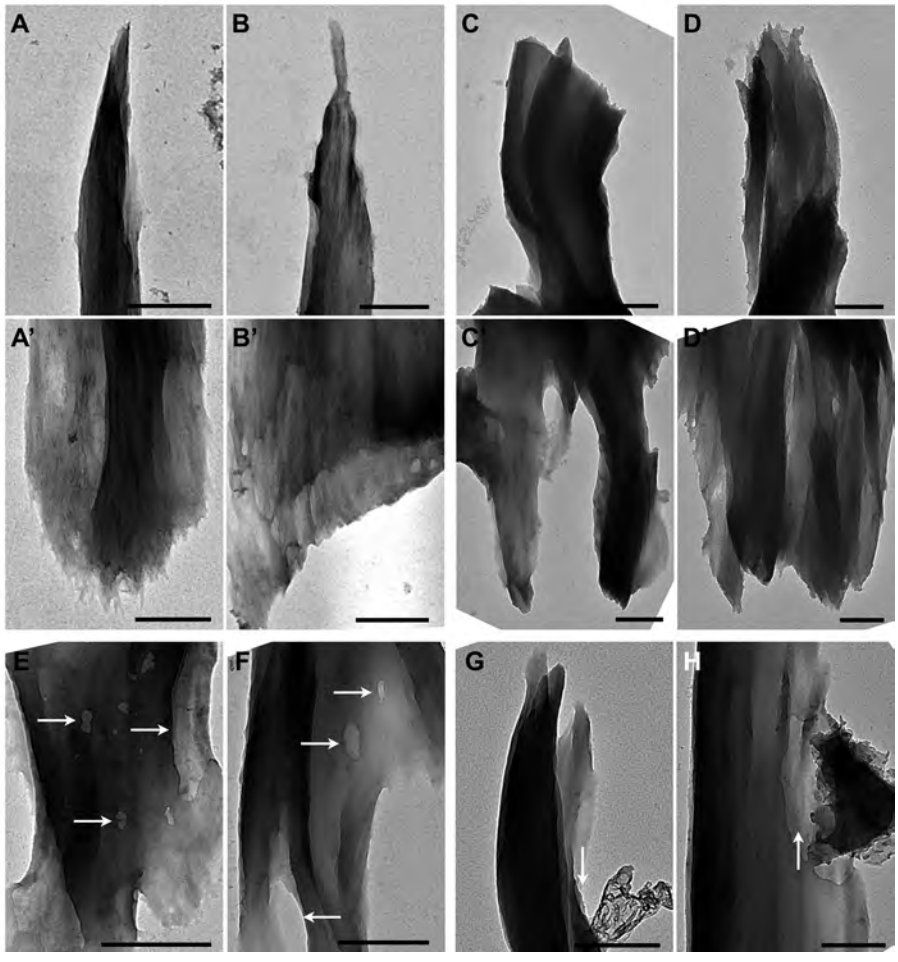

Fig. 2. TEM micrographs of partially digested small Avicel PH101 particles. Those digested to ~60% conversion by a CTec2 cellulase mixture (primarily comprising *T. reesei* Cel7A) display tip sharpening (**A** and **B**) on one end and blunted morphology on the other (**A′** and **B′**) Particles digested to ~65% conversion with CelA display slightly narrowed, tapered, blunt ends (**C** and **D**) and irregular blunt or angled ends (**C′** and **D′**), as well as cavities of various sizes on the surface [arrows, (**E**) and (**F**)]. Occasionally, some of the material being removed to form the cavities appears to remain attached to the cavity edge (**G** and **H**). All scale bars are 500 nm.

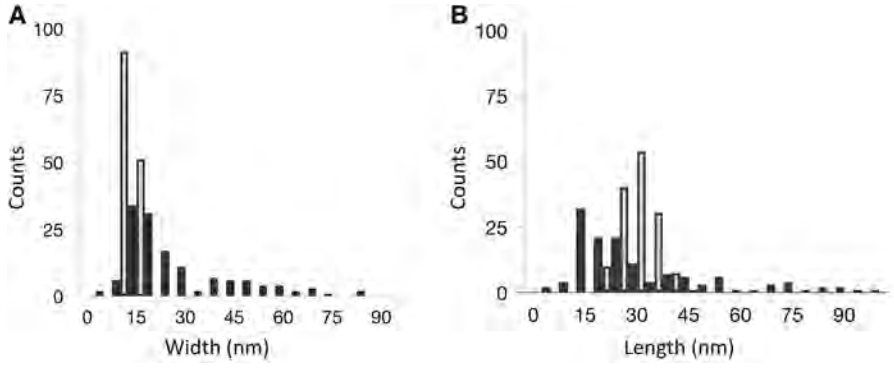

Fig. 3. Evidence that CelA fits in the cavities. (**A** and **B**) The histograms show dimensions of the cavities created by CelA (black), extracted from the TEM micrographs in Fig. 2, E and F, and the calculated dimensions of CelA itself (gray) over the course of a 40-ns molecular dynamics simulation, with the smallest dimension in (A) and largest dimension in (B).

Despite these compelling properties exhibited by CelA, several limitations may also be attributed to its multimodular architecture. Although the levels of glucan conversion achieved by CelA acting on highly crystalline cellulose (Avicel) are much higher than that of *T. reesei* Cel7A, the overall performance of CelA acting on commercially relevant feedstocks, such as switchgrass and corn stover, are lower, even with the use of a β-D-glucosidase (Fig. 1C). There are several important differences between biomass feedstocks and Avicel. Biomass is structurally and chemically more complex than Avicel. Smaller enzymes, such as Cel7A, may be better able to penetrate into the plant cell walls, even after pretreatment, whereas the larger CelA, with its multiple CBMs, may be too large and be more prone to nonproductive binding. Some of these characteristics may be enhanced by pretreatments that markedly change the structural integrity of the biomass, perhaps reflected by our observation that CelA works best with the AP-pretreated biomass, a pretreatment that substantially removes lignin (*22*, *23*) (Fig. 1C).

Chemical species other than glucan present in biomass, such as lignin and hemicelluloses, may trap CelA in nonproductive binding states. Naturally, all cellulases will experience some level of nonproductive binding on biomass. However, owing to the larger molecular weight of CelA and because enzyme loadings are done on an enzyme weight basis, the molar enzyme loading of CelA is much lower than it is for Cel7A. Therefore, the activity of a CelA would be affected to a greater extent by nonproductive adsorption than would an equivalent mixture of cellulases with lower molecular weights. This might explain the results on AFEX-pretreated biomass, where lignin-carbohydrate complexes (LCCs) are redeposited on plant cell walls after such pretreatment (*22*, *23*). LCCs represent a barrier for all enzymes, but the problem seems amplified for a complex enzyme such as CelA.

Further comparison of CelA to its fungal counterparts reveals another fascinating aspect of this cellulase. Transmission electron microscopy (TEM) imaging results indicate that CelA has a mode of action distinct from that ascribed to fungal enzymes, such as those used in the commercial formulation CTec2 (Novozymes).

We examined changes in the surface and internal structure of Avicel PH101 particles treated with CelA. These particles were recovered from digestions carried out to ~65% cellulose conversion and then compared to Avicel particles digested to ~60% conversion using CTec2 (CTec2 is composed primarily of Cel7A). Our analysis focused on the most electron-translucent particles within each sample, where individual cellulose microfibrils could often be identified within the particles (Fig. 2). Particles digested by the Cel7A-containing formulation displayed morphology previously reported where one end of the particle was finely tapered to a narrow point (Fig. 2, A and B) and an opposite end displayed a blunt edge with a slight angle from the long axis (Fig. 2, A' and B') (*24*). CelA-digested particles, by contrast, displayed narrowed, irregular, but not finely tapered morphology on one end (Fig. 2, C and D) and an irregular, scalloped, angled morphology on the opposite end (Fig. 2, C' and D'). In addition to a surface ablation activity typical of fungal enzymes that seem to work only on the surface of the substrate (Fig. 2, A and B), CelA appears to excavate down into the layers of the substrate, generating cavities (Fig. 2, E to H, and fig. S8). TEM evidence for this excavation mechanism was notable: CelA-digested Avicel particles displayed surfaces marked by irregularly spaced cavities distributed along the length of the particles (Fig. 2, E and F). The cavities vary widely in cross-sectional area at the surface between 25 and ~1000 nm^2. Most cavities are less than 500 nm^2 in cross-sectional area and have widths of 15 to 20 nm and lengths of 15 to 30 nm (Fig. 3). No cavities were observed on the surface of Avicel particles digested by CTec2. Additionally, on the basis of the predicted conformations that CelA can attain, as inferred by molecular dynamics simulations, it is clear that CelA fits into these cavities (Fig. 3). From these simulations, the effective size of CelA is estimated to be between 10 and 35 nm. These spatial dimensions for CelA correlate very well with the smaller cavity dimensions (diameter) in the range of 15 to 30 nm (Fig. 3). The schematic in Fig. 4 summarizes the differences in the digestion mechanism of CelA and Cel7A.

To better understand the digestion mechanism of CelA, we constructed a kinetic model of enzymatic digestion of cellulose to test the hypothesis that CelA has a specific mode of action that leads to cavity formation, a mode not found in simpler cellulase enzymes (SM text S6). In the model, CBMs are designed to bind and unbind from cellulose; catalytic domains have binding, engaging, and processive digestion functionality, with enhanced binding when a linker-attached module is bound. The model reproduces the surface ablation mode of digestion found in the commercial formulations when applied to enzyme with only a single CBM linked to a single catalytic domain (fig. S13C). The same digestion, dominated by single-layer ablation, is observed if identical functionality is used with the complex CelA-type model (fig. S10A). The model shows that the cavity formation occurs (fig. S10B) when the additional constraint is put into the complex model so that, once fully bound, the processive digestion is slowed owing to competing digestion directions and increased binding affinity, which inhibits repeated unbinding and rebinding in other locations.

On the basis of these results, it appears that CelA and multifunctional cellulases represent a distinct paradigm for cellulose digestion and high activity created by combining complementary modules separated by long linker peptides on the same gene product. We propose that this hydrolysis mechanism could synergize with other cellulases representing the previously known paradigms—free enzymes and cellulosomes—because they offer distinctive mechanisms. For example, in nature, enzymes from mesophilic saprophytic fungi are unable to mix with enzymes secreted into high-temperature environments. We thus propose that considerable synergism can be afforded if it is possible to bring together enzymes from these natural cellulolytic paradigms.

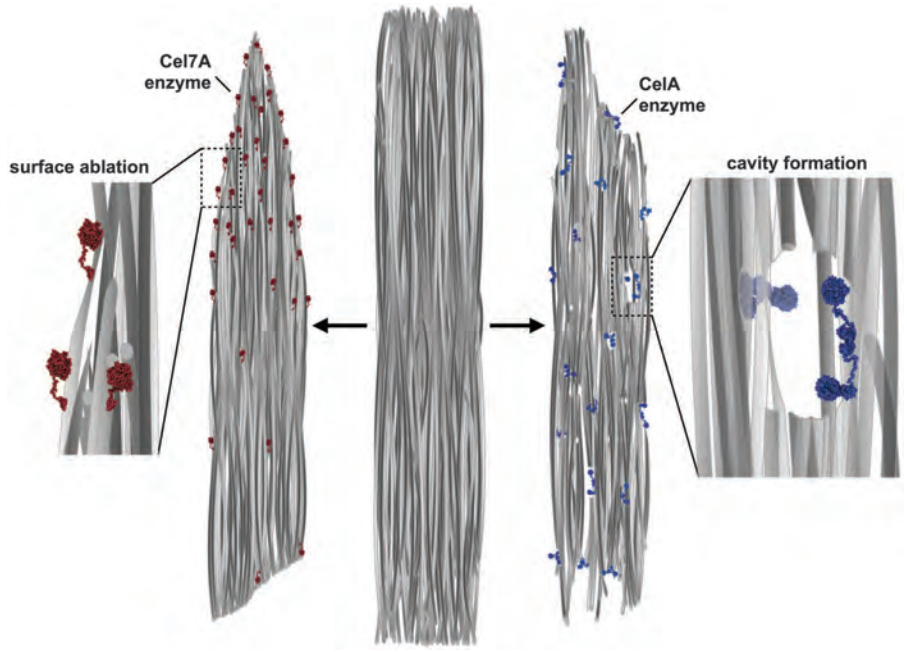

Fig. 4. Schematic representation of digested cellulose microfibril bundles. The diagrams contrast the surface ablation and reducing-end oriented mechanism of Cel7A (**left**) with the surface ablation and cavity-forming mechanism of CelA (**right**). This representation suggests how these distinct deconstruction mechanisms could be synergistic by specializing in different aspects of the nanoscale architecture of biomass materials and exposing new surfaces.

References and Notes

1. D. Martinez *et al.*, *Nat. Biotechnol.* **26**, 553–560 (2008).
2. E. A. Bayer, J. P. Belaich, Y. Shoham, R. Lamed, *Annu. Rev. Microbiol.* **58**, 521–554 (2004).
3. R. H. Doi, A. Kosugi, *Nat. Rev. Microbiol.* **2**, 541–551 (2004).
4. I. A. Kataeva *et al.*, *J. Bacteriol.* **191**, 3760–3761 (2009).
5. A. Lochner *et al.*, *Appl. Environ. Microbiol.* **77**, 4042–4054 (2011).
6. S. J. Yang *et al.*, *Appl. Environ. Microbiol.* **75**, 4762–4769 (2009).
7. Y. Vazana, S. Moraïs, Y. Barak, R. Lamed, E. A. Bayer, *Appl. Environ. Microbiol.* **76**, 3236–3243 (2010).
8. E. Berger, D. Zhang, V. V. Zverlov, W. H. Schwarz, *FEMS Microbiol. Lett.* **268**, 194–201 (2007).
9. V. Zverlov, S. Mahr, K. Riedel, K. Bronnenmeier, *Microbiology* **144**, 457–465 (1998).
10. S. Shoemaker *et al.*, *Biotechnology* **1**, 691–696 (1983).
11. J. O. Baker *et al.*, *Appl. Biochem. Biotechnol.* **45-46**, 245–256 (1994).
12. J. O. Baker, C. I. Ehrman, W. S. Adney, S. R. Thomas, M. E. Himmel, *Appl. Biochem. Biotechnol.* **70-72**, 395–403 (1998).
13. L. C. Textor *et al.*, *FEBS J.* **280**, 56–69 (2013).
14. L. P. Walker, C. D. Belair, D. B. Wilson, D. C. Irwin, *Biotechnol. Bioeng.* **42**, 1019–1028 (1993).
15. M. Gruno, P. Väljamäe, G. Pettersson, G. Johansson, *Biotechnol. Bioeng.* **86**, 503–511 (2004).
16. W. Liebl, J. Gabelsberger, K. H. Schleifer, *Mol. Gen. Genet.* **242**, 111–115 (1994).
17. M. M. Patel, R. M. Bhatt, *J. Chem. Technol. Biotechnol.* **53**, 253–263 (1992).
18. M. J. Playne, *Biotechnol. Bioeng.* **26**, 426–433 (1984).
19. C. E. Wyman *et al.*, *Bioresour. Technol.* **96**, 2026–2032 (2005).
20. H. Alizadeh, F. Teymouri, T. I. Gilbert, B. E. Dale, *Appl. Biochem. Biotechnol.* **124**, 1133–1141 (2005).
21. S. P. S. Chundawat *et al.*, *Bioresour. Technol.* **101**, 8429–8438 (2010).
22. R. Brunecky *et al.*, *Biotechnol. Bioeng.* **102**, 1537–1543 (2009).
23. S. P. S. Chundawat *et al.*, *Energ. Environ. Sci.* **4**, 973–984 (2011).
24. M. G. Resch *et al.*, *Energy Environ. Sci.* **6**, 1858 (2013).

Acknowledgments: This work was supported by the BioEnergy Science Center (BESC). BESC is a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the U.S. DOE Office of Science. We acknowledge colleagues at the Biomass Conversion Research Laboratory at Michigan State University for providing the AFEX-pretreated materials. Structures have been deposited in the Protein Data Bank with PDB codes 4DOD (GH9), 4DOE (GH9-CB), and 4EL8 (GH48).

Supplementary Materials

www.sciencemag.org/content/342/6165/1513/suppl/DC1
Materials and Methods
Figs. S1 to S17
Supplementary Text S1 to S6
Tables S1 to S8
References (*25*–*35*)

5 August 2013; accepted 8 November 2013
10.1126/science.1244273

Assembly and Validation of the Genome of the Nonmodel Basal Angiosperm *Amborella*

Srikar Chamala,[1]* Andre S. Chanderbali,[1,2]* Joshua P. Der,[3] Tianying Lan,[4] Brandon Walts,[1] Victor A. Albert,[4] Claude W. dePamphilis,[3] Jim Leebens-Mack,[5] Steve Rounsley,[6] Stephan C. Schuster,[7,8,9] Rod A. Wing,[10,11] Nianqing Xiao,[12] Richard Moore,[12] Pamela S. Soltis,[2,13] Douglas E. Soltis,[1,2,13] W. Brad Barbazuk[1,13]†

Genome sequencing with next-generation sequence (NGS) technologies can now be applied to organisms pivotal to addressing fundamental biological questions, but with genomes previously considered intractable or too expensive to undertake. However, for species with large and complex genomes, extensive genetic and physical map resources have, until now, been required to direct the sequencing effort and sequence assembly. As these resources are unavailable for most species, assembling high-quality genome sequences from NGS data remains challenging. We describe a strategy that uses NGS, fluorescence in situ hybridization, and whole-genome mapping to assemble a high-quality genome sequence for *Amborella trichopoda*, a nonmodel species crucial to understanding flowering plant evolution. These methods are applicable to many other organisms with limited genomic resources.

Amborella (*1*, *2*) has been identified as the single sister species to all other living angiosperms and is a pivotal reference for comparison to other angiosperms (*3*). However, *Amborella* is not a genetic model and has no existing genetic map, genetic resources, or genome sequence. Although next-generation sequencing (NGS) provides deep genomic sequence coverage at low cost, short-read assembly remains difficult, and assessing assembly accuracy is problematic without independently derived genomic maps. We produced a whole-genome assembly for *Amborella* from a mixed data set of 454, Illumina, and Sanger bacterial artificial chromosome (BAC)–end sequences, evaluated the assembly using fluorescence in situ hybridization (FISH), and improved contiguity using whole-genome mapping. FISH has broad utility (*4*), but has not been used in de novo genome assembly. Likewise, whole-genome mapping has been used to assemble bacterial genomes (*5*, *6*), but has only recently been applied to complex genomes of model organisms (*7*, *8*) to assist with scaffolding and correction of well-advanced genome assemblies.

More than 23 Gb of quality-filtered (*9*) DNA sequence comprising single-end (SE) 454-FLX, SE 454-FLX+ reads, 11-kb paired-end (PE) 454-FLX, 3-kb PE Illumina HiSeq, and Sanger-sequenced BAC-end reads (*10*) were combined and assembled (table S1). Assembly (*9*) resulted in 5745 scaffolds totaling 706 Mb (table S5) with a mean scaffold size of 123 kb and an N50 size of 4.9 Mb, and N90 scaffold metrics that indicate that 90% of our assembled sequence resides within 155 scaffolds greater than 1.1 Mb in length (table S5).

Flow cytometry was used to estimate the size of the *Amborella* genome at ~870 Mb (*11*), while our sequence-based size assessments (*9*, *10*, *12*, *13*) suggest that the *Amborella* genome size is closer to 748 Mb. Our high-quality sequence represents an average depth of coverage of ~31×, and the assembly covers >94% of the genome.

Long contig and scaffold assemblies are required to understand genome structure, enable gene identification, and support subsequent comparative, structural, and population genomics studies. We sought long continuous stretches of assembled sequence that represent all, or a major fraction of, the *Amborella* genome. Coverage of two finished BAC contigs (*10*) by assembled sequence contigs suggests that these two regions were faithfully represented in the assembly (figs. S9 and S10) (*9*), and all 155 of our N90 scaffolds incorporate physically mapped BAC-end sequences.

The accuracy of the genome assembly was further assessed by FISH analysis (*9*). BACs assembled in 104 scaffolds containing 430 Mb (68%) of the genome assembly were cytogenetically localized by FISH to assess scaffold integrity (Fig. 1, fig. S11, and table S8). This analysis confirmed contiguity across major regions (56%) of 66 scaffolds containing 306 Mb (44%) of the genome assembly. Notably, co-assembled BACs that were

[1]Department of Biology, University of Florida, Gainesville, FL 32611, USA. [2]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA. [3]Department of Biology and Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA. [4]Department of Biological Sciences, University at Buffalo (State University of New York), Buffalo, NY 14260, USA. [5]Department of Plant Biology, University of Georgia, Athens, GA 30602, USA. [6]Dow AgroSciences, 9330 Zionsville Road, Indianapolis, IN 46268, USA. [7]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA. [8]Center for Comparative Genomics and Bioinformatics, The Pennsylvania State University, University Park, PA 16802, USA. [9]Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, 637551 Singapore. [10]Arizona Genomics Institute, University of Arizona, Tucson, AZ 5721, USA. [11]School of Plant Sciences and BIO5 Institute for Collaborative Research, University of Arizona, Tucson, AZ 85721, USA. [12]OpGen, Inc., 708 Quince Orchard Road, Gaithersburg, MD 20878, USA. [13]Genetics Institute, University of Florida, Gainesville, FL 32610, USA.

*These authors contributed equally to this work.
†Corresponding author. E-mail: bbarbazuk@ufl.edu