CHAPTER 1

MANY PATHS UP THE MOUNTAIN: TRACKING THE EVOLUTION OF CELLULOSE BIOSYNTHESIS*

DAVID R. NOBLES, JR. AND R. MALCOLM BROWN, JR.**

Section of Molecular Genetics and Microbiology, The University of Texas at Austin, Austin, TX 78712

Abstract

Available evidence supports a common ancestry for all cellulose synthases. These enzymes appear to have been a bacterial invention acquired by various eukaryotes via multiple lateral gene transfers. However, the proteins associated with regulation of cellulose biosynthesis and polymer crystallization seem to have evolved independently. Sequence divergence of eukaryotic cellulose synthases and the presence of multiple gene clusters associated with bacterial cellulose synthases are discussed in relation to the possible evolutionary pathways of cellulose biosynthesis.

Keywords

bacterial cellulose, cellulose biosynthesis, cellulose synthase, cyanobacteria, lateral gene transfer, synteny.

1   INTRODUCTION

Cellulose biosynthesis is a phenomenon observed in bacteria (proteobacteria, firmicutes, and cyanobacteria) (Mühlethaler 1949; Deinema and Zevenhuizen 1971; Napoli et al. 1975; Brown, Jr. et al. 1976; Roberts 1991; Ross et al. 1991; Nobles et al. 2001; Zogaj et al. 2001, 2003; Spiers et al. 2002) and eukaryotes including plants (Glaucophyceae, Rhodophyceae, and Chlorophyceae) (Brown, Jr. 1985), animals (urochordates) (Kimura et al. 2001b), stramenopiles (Brown, Jr.

1

et al. 1969; Yamada and Miyazaki 1976) the myceteozoan *Dictyostelium discoideum* (Blanton et al. 2000), and *Acanthamoeba* spp. (Linder et al. 2002). The economical, ecological, and biological significance of cellulose production, as well as its widespread bacterial and eukaryotic distribution, have inspired a great deal of research into mechanisms of regulation and synthesis. The subject of this chapter is a much less researched area of cellulose research: specifically, the origins and evolutionary pathways of cellulose biosynthesis. In order to address this issue, it is first necessary to consider whether the trait of cellulose production among the wide range of organisms in which the process has been observed is a result of homology (similarity due to common descent) or homoplasy (similarity due to convergent evolution from independent origins).

Biosynthesis of the most common crystalline allomorph, cellulose I, requires the distinct processes of polymerization and crystallization (Benziman et al. 1980; Brett 2000; Saxena and Brown, Jr. 2005). These processes are coupled at highly ordered, membrane spanning, multienzyme terminal complexes (TCs) (Roelofsen 1958; Preston 1974). Such complexes have been observed in vascular plants (Mueller and Brown, Jr. 1980), algae (Brown, Jr. and Montezinos 1976; Giddings et al. 1980; Tsekos 1999; Schüaler et al. 2003; Okuda et al. 2004), *Gluconacetobacter xylinum* (synonym *Acetobacter xylinum*) (Zaar 1979; Kimura et al. 2001a), urochordates (Kimura et al. 2001b), and *Dictyostelium discoideum* (Grimson et al. 1996). A highly organized complex is believed to be necessary to produce the metastable parallel glucan chain orientation of the crystalline cellulose I allomorph and to prevent the biosynthesis of noncrystalline material and/or the folding of the nascent glucan chains into cellulose II – the most thermodynamically stable allomorph of cellulose (Brown, Jr. 1996).

The comparative morphology of TCs has been utilized as a tool for constructing an evolutionary history of cellulose biosynthesis of vascular plants and algae (Hotchkiss and Brown, Jr. 1988; Tsekos 1999). With regard to eukaryotic terminal complexes in general, it is not possible to know the validity of such inferences since the structural components of terminal complexes have not been identified. Although comparisons between such disparate groups as plants, stramenopiles, ascidians, and *Dictyostelium discoideum* are questionable, this methodology may have some merit when considering the evolution of eukaryotic TCs within related groups. The structural proteins comprising prokaryotic TCs are also unknown. However, the TC-associated proteins responsible for export and secretion of cellulose in bacteria with gram negative cell envelope architecture almost certainly have no relationship to their functional counterparts in eukaryotes. Therefore, the biosynthesis of cellulose I in bacteria and eukaryotes is in all likelihood a result of convergent evolution.

Polymerization of the β-1,4-glucan chain is catalyzed by cellulose synthase enzymes. All known cellulose synthases are family 2 processive glycosyltransferases, a ubiquitous family of enzymes which also includes chitin synthases, hyaluronan synthases, and NodC proteins (Saxena et al. 2001). Cellulose synthase sequences share a highly conserved catalytic region containing the D, D, D,

QXXRW (associated with regions U1, U2, U3, and U4, respectively), a motif characteristic of processive β glycosyltransferases (Saxena et al. 1995). Cellulose synthases undoubtedly share a common ancestry. Therefore, unlike the process of crystallization, synthesis of the β-1,4-glucan homopolymer in bacteria and eukaryota is a homologous process that forms an evolutionary link between all cellulose producing organisms.

2   SEQUENCE COMPARISONS

With the rapid growth of sequence databases, similarity searches such as BLAST have become integral tools for molecular biology and bioinformatics research. Although such searches are not suitable for inferring phylogenic relationships, they allow rapid identification of probable homologous sequences by utilizing pairwise alignments and give a statistical measurement of the significance of the similarity displayed by the two sequences. Pairwise alignments of cellulose synthase amino acid sequences yield interesting results. Although strong sequence similarity is displayed within related phyla (e.g., proteobacterial sequences are similar to other proteobacterial sequences and vascular plant sequences are similar to those of other vascular plants), cellulose synthases demonstrate little similarity when comparisons are made between more distantly related organisms (Blanton et al. 2000; Richmond 2000; Nobles et al. 2001). This is not necessarily surprising, as one might expect homologous sequences from *Arabidopsis thaliana* and *Escherichia coli* to display divergence. What is surprising, however, is that without exception, eukaryotic sequences display greater similarity to prokaryotic sequences than to their other eukaryotic counterparts (Table 1-1). Furthermore, in each case, the expectation values demonstrate that the similarity between eukaryotic and prokaryotic sequences is statistically significant. This is particularly noticeable in the results of pairwise sequence alignments of the

Table 1-1. Expectation values from pairwise BLAST alignments of eukaryotic and prokaryotic cellulose synthases

	IRX3	Ddis	Pram	Cint	Aory	N7120	Styp	Smel	Gxyl	Bthu
IRX3		8e-09	0.015	0.27	NSSF	*__2e-28__*	6e-06	0.27	0.35	0.006
Ddis	8e-09		4e-24	9e-24	2e-16	*__9e-42__*	2e-22	9e-19	3e-16	3e-23
Pram	0.015	4e-24		2e-13	3e-07	*__3e-28__*	3e-13	7e-10	3e-12	4e-14
Cint	0.27	9e-24	2e-13		7e-17	9e-33	8e-32	*__9e-38__*	3e-26	1e-36
Aory	NSSF	2e-16	3e-07	7e-17		2e-17	1e-19	*__4e-24__*	4e-20	3e-19

Results of alignments between two eukaryotic sequences are shaded. Expectations values from alignments involving bacterial sequences are unshaded. The expectation values demonstrating greatest sequence similarity are shown in bold and italicized. IRX3 – *Arabidopsis thaliana* (NP_197244.1), Ddis – *Dictyostelium discoideum* (AAF00200.1), Cint – *Ciona intestinalis* (BAD10864.1), Aory – *Aspergillis oryzae* (BAE64416.1), N7120 – *Nostoc* sp. PCC 7120 (NP_487797.1), Styp – *Salmonella typhimurium* (CAC86199.1), Smel – *Sinorhizobium meliloti* (NP_436917.1), Axyl – *Gluconacetobacter xylinus* (CAA38487.1), *Bacillus thuringiensis* serovar israelensis (ZP_00741731.1).

cellulose synthase from *Nostoc* sp. PCC 7120 with IRX3 (*A. thaliana* CesA) and the cellulose synthase from *Dictyostelium discoideum* (DcsA) in which expectation values of 2.5×10^{-20} and 2.25×10^{-18} times lower than the most similar eukaryotic sequences are obtained. Although no definite conclusions can be drawn from this data alone, it suggests that a mechanism other than vertical evolution is at work in the eukaryotic acquisition of cellulose synthesis.

3 EUKARYOTIC CELLULOSE SYNTHASES

3.1 The case for a cyanobacterial origin of plant cellulose synthases

Cellulose synthase amino acid sequences from various members of the Nostocales show striking similarity to plant cellulose synthase (CesA) and cellulose synthase-like protein (Csl) sequences (Nobles et al. 2001; Nobles and Brown, Jr. 2004). When a CesA (IRX3) from *A. thaliana* is compared with sequences from *Nostoc* sp. PCC 7120 (CcsA1), DcsA (the most similar nonplant eukaryotic sequence) and *Chloroflexus aurantiacus* J-10-fl (the most similar prokaryotic noncyanobacterial sequence), expectation values of 2e-28, 8e-09 and 7e-12 respectively, are generated. The significance of this similarity is augmented by multiple alignments of cellulose synthases which demonstrate sequence conservation within catalytic domains U1, U2, U3, and U4, but also reveal a large insertion region (first identified as the plant conserved and specific region or CR-P (Delmer 1999)) present in CesA, DcsA, and CcsA1 sequences that is absent in other prokaryotic sequences (Nobles et al. 2001; Roberts and Roberts 2004). Furthermore, protein trees generated by neighbor-joining, maximum likelihood, and maximum parsimony methods all demonstrate a sister grouping of cyanobacterial and vascular plant sequences similar to that observed with chloroplasts and cyanobacteria in 16s ribosomal trees (Olsen et al. 1994; Nobles et al. 2001; Nobles and Brown, Jr. 2004; Nakashima et al. 2004).

The primary endosymbiotic capture of an ancestral cyanobacterium, its subsequent evolution into a plastid, and concomitant transfer of genes to the host nucleus provide the most parsimonious explanation for the observed results. Gene transfers from organelles occur frequently and have had a profound effect on host genomes (Archibald et al. 2003; Huang et al. 2003). Indeed, it has been estimated that approximately 18% of the protein coding genes in *A. thaliana* are of cyanobacterial origin (Martin et al. 2002). Although xenologous transfer (lateral transfer from a free living organism) cannot be dismissed, gene transfer from the ancestral plastid (synologous transfer) seems the most probable pathway for the integration of a cyanobacterial cellulose synthase into an ancestor of vascular plants.

3.2 Lateral transfer of cellulose synthase in the urochordates

The urochordates are unique in that they are the only animals known to produce cellulose. This ability is especially curious given their position as basal chordates. In order to explain this, one must make one of three assumptions: (1) An early

diverging ancestor of animals possessed the ability to produce cellulose which was subsequently lost by all animals except the urochordates; (2) Cellulose biosynthesis in urochordates is the result of convergent evolution; or (3) The ability to produce cellulose was obtained via lateral gene transfer of one or more of the components necessary for cellulose biosynthesis. The occurrence of the scenario described by the first assumption would be extraordinary indeed! So extraordinary in fact, that it can likely be dismissed as far too improbable to occur. Furthermore, the identification of cellulose synthase sequences from *Ciona intestinalis* and *Ciona savignyi* as family 2 processive glycosyltransferases (Dehel et al. 2003) suggests that the process of synthesizing a β-1,4-glucan homopolymer by urochordates is homologous to that of other cellulose synthesizing organisms. Therefore, the second assumption is also rather unlikely. Further examination of *Ciona* cellulose synthase (Ci-CesA) sequences by BLAST alignment demonstrates that they have significantly greater similarity to cellulose synthase sequences from firmicutes, cyanobacteria, and proteobacteria than to other eukaryotic cellulose synthases. Unfortunately, differences in the expectation values generated by comparisons of Ci-CesAs with sequences from these distinct bacterial phyla are equivocal. Therefore, it is not possible to identify a likely a point of origin for Ci-CesA based on sequence similarities. To date, phylogenetic analyses have also been unable to demonstrate a clear relationship of Ci-CesAs to any group of bacterial cellulose synthases. Thus, based on analysis of the glycosyltransferase, the identity of the donor organism remains a mystery (Nobles and Brown, Jr. 2004; Nakashima et al. 2004).

However, there is another piece to the Ci-CesA puzzle. Ci-CesA sequences have a unique feature: the C-terminus displays sequence similarity to bacterial family 6 glycosylhydrolases. The glycosylhydrolase regions of *Ciona savignyi* and *Ciona intestinalis* cellulose synthases are degenerate and therefore, probably retain no enzymatic activity (Matthysse et al. 2004). While the presence of this region strengthens the case for lateral gene transfer, it poses a problem for the identification of the sequence donor(s). A gene fusion of this type is most likely to occur as a result of the simultaneous transfer of adjacent coding regions rather than from independent acquisition of sequences and subsequent fusion. Although some species of *Streptomyces* (Actinobacteria) possess gene clusters containing putative cellulose synthases and family 6 glycosylhydrolases (Nakashima et al. 2004), these sequences show comparatively little similarity to the N and C termini of Ci-CesA, respectively. Therefore, even though all available evidence indicates that urochordates acquired their cellulose synthase through lateral transfer(s) from bacteria, it is not possible to determine the phylum of origin at this time.

3.3 The cellulose synthase of *Dictyostelium discoideum*

The cellulose synthase sequence (DcsA) from *D. discoideum* is far more similar to cellulose synthase sequences from *Nostoc* spp. (CcsA1) than to any other sequences in the current databases. Additionally, DcsA branches as a sister clade to CcsA1 in protein trees (Nobles and Brown, Jr. 2004; Nakashima et al. 2004).

These observations suggest a lateral transfer of cellulose synthase from cyanobacteria to *D. discoideum*. However, while the primary and secondary endosymbiotic events that led to the evolution of plastids in plants and algae provide a clear mechanism for the transfer of a cyanobacterial cellulose synthase to photosynthetic organisms, such a mechanism is lacking for *D. discoideum*. Cyanobacterial genes are known to exist in eukaryotes which have secondarily lost plastids. However, there is no evidence for the existence of an endosymbiotic relationship between ancestors of *D. discoideum* and a cyanobacterium. Therefore, if a lateral transfer occurred, it was likely xenologous, possibly via a food ratchet mechanism (Doolittle 1998).

4 BACTERIAL GENE CLUSTERS

4.1 Introduction

A comprehensive phylogenetic analysis of bacterial cellulose synthases has not been performed to date. The few studies which include significant taxon sampling demonstrate that bacterial sequences included in cellulose synthase protein trees generally branch in a manner similar to that observed in species trees (Nobles and Brown, Jr. 2004; Nakashima et al. 2004). Unfortunately, phylogenetic studies have been unable to demonstrate the origin of cellulose synthase among the bacterial phyla (Nobles and Brown, Jr. 2004). However, the presence of cellulose synthases in firmicutes, actinobacteria, cyanobacteria, and proteobacteria indicates that the evolution of synthases cellulose likely predates the divergence of these groups.

The proliferation of complete genome sequences in public databases provides an additional means to track the evolution of cellulose in bacteria. Conservation of operons and/or gene clusters (synteny) can be used to trace not only the history of cellulose synthase, but also its associated proteins. The existence of a few these gene clusters has been well documented. In the sections below, I would like to give a brief review of the known gene organizations and introduce two novel ones which may be linked to the eukaryotic acquisition of cellulose biosynthesis.
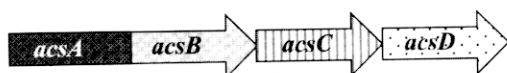
4.2 Characterized gene clusters

Gene clusters responsible for cellulose biosynthesis have only been extensively characterized in proteobacteria. Three archetypal gene organizations encoding proteins required for the synthesis of cellulose in the α, β, and γ subdivisions have been identified. It should be emphasized that variations of these archetypes exist and as such, the examples given here are meant to serve as paradigms to simplify the discussion of the general characteristics of gene clusters associated with cellulose biosynthesis (for a comprehensive review of these gene organizations, see Römling 2002). Based on sequence conservation, the three archetypal organizations can be divided into two groups (Figure 1-1): the Group I cluster found in the α, β, and γ subdivisions, encodes a cellulose synthase – A (*bcsAlacsA*), as well as the B (*bcsBlacsB*) and C (*bcsClacsC*) proteins within an

operon (Saxena et al. 1994; Zogaj et al. 2001). Additionally, a family 8 glyco-sylhydrolase, is encoded within or in close proximity to the cellulose synthase operon (Römling 2002). The Group II gene cluster is characteristically found in the α-Proteobacteria and has been most extensively studied in *Agrobacterium tumefaciens*. This organization consists of two adjacent directionally opposed operons. The first operon – *celABC* encodes homologs of the Group I *bcsA/acsA* (*celA*), *bcsB/acsB* (*celB*), and family 8 glycosyl hydrolase (*celC*). The second operon – *celDE* encodes proteins with no significant sequence similarity to the Group I proteins (Matthysse et al. 1995b). It should be noted however, that *celD* and *bcsC/acsC* share two conserved domains: COG3118 (thioredoxin containing proteins responsible for posttranslational modifications and protein turnover) and COG4783 (putative Zn-dependent proteases containing TPR repeats) (Marchler-Bauer et al. 2005) suggesting the possibility of a similar function.

Although all proteins encoded by Group I and II gene clusters are necessary for wild-type cellulose biosynthesis, only cellulose synthases have a known function.[1]
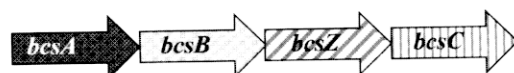
Gluconacetobacter xylinum ATCC 23769

Gluconacetobacter xylinum B42
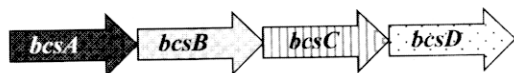
Salmonella spp.
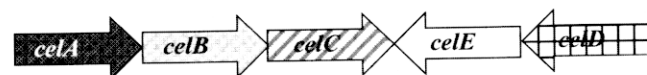
Agrobacterium tumefaciens A6

Figure 1-1. Characterized cellulose biosynthesis gene clusters of Group I (*Gluconacetobacter xylinum* ATCC 23769, *G. xylinum* B42, and *Salmonella* spp.) and Group II (*Agrobacterium tumefaciens* A6). Identical shades and patterns represent homologous sequences. Open reading frames are not drawn to scale. Note that *acsAB* comprises a single open reading frame in *G. xylinum* ATCC 23769 (synonym AY201). Although not shown here, fusions of *acsAB* are also observed in the cellulose synthesis oper-ons of *G. xylinum* ATCC 53582 (synonym NQ5) (Saxena et al. 1994) and *Azotobacter vinelandii* AvOP

---

[1] Previous research suggested that AcsB/BcsB had the ability to bind c-di-GMP which led to a general consensus that the B subunit had a regulatory function in cellulose biosynthesis (Amikam and Benziman 1989; Mayer et al. 1991). However, based on recent data, a regulatory role for AcsB/BcsB is questionable (Amikam and Galperin 2006).

Organisms with the Group I gene organization are believed to carry out cellulose biosynthesis without the use of lipid-linked intermediates in a process which is upregulated by the allosteric activator cyclic diguanosine monophosphate (c-di-GMP) (Aloni et al. 1982; Ross et al. 1986; Saxena et al. 1994; Römling 2002; García et al. 2004). In the case of bacteria with Group II gene organizations, the presence of lipid linked intermediates and regulation by c-di-GMP are matters of some debate (Amikam and Benziman 1989; Matthysse et al. 1995a; Ausmees et al. 2001). Despite clear differences between the gene clusters of Groups I and II, the universal presence of a family 8 glycosylhydrolase and the positioning of the *acsB*/*bcsB*/*celB* genes adjacent to cellulose synthases suggests a common ancestry for these gene clusters and important roles for these proteins that transcend possible differences in mechanisms of synthesis.

5   NOVEL GENE CLUSTERS

5.1   Introduction

Although cellulose synthase sequences within closely related bacterial groups (e.g., within γ-proteobacteria or actinobacteria) generally display relatively high sequence conservation, they are often divergent when compared across phyla. Examination of sequenced genomes reveals the presence of alternative gene clusters that coincide with the sequence divergence of various groups of cellulose synthases. Such novel gene organizations exist in cyanobacteria, actinobacteria, chloroflexales, as well as proteobacteria and, unlike the Group I and II gene clusters described above, gene organizations are conserved across phyla. The characterization of novel gene organizations has the potential to inform current knowledge of the components necessary for cellulose biosynthesis, broaden our definitions of what constitutes cellulose, and ultimately provide a map of routes taken by organisms to utilize the β-1,4-homopolymer.

5.2   Group III

The Group III gene cluster is found in orders chroococales and nostocales of cyanobacteria and in the α and β subdivisions of proteobacteria. These clusters have not been experimentally shown to be responsible for cellulose biosynthesis and therefore, the designation of these glycosyltransferases as cellulose synthases is a putative one. Group III gene clusters encode a membrane fusion protein (MFP) of the AcrA/EmrA/HylD family adjacent and upstream of the cellulose synthase (Figure 1-2). This organization suggests the possibility of a three component system consisting of the cellulose synthase, a MFP, and an outer membrane protein OMP. In such a system, the cellulose synthase would be linked to an OMP pore by the membrane-bound periplasmic MFP and thus, form a continuous channel for export and secretion of the glucan polymer. Alternatively, some Group III clusters contain genes encoding ATP binding
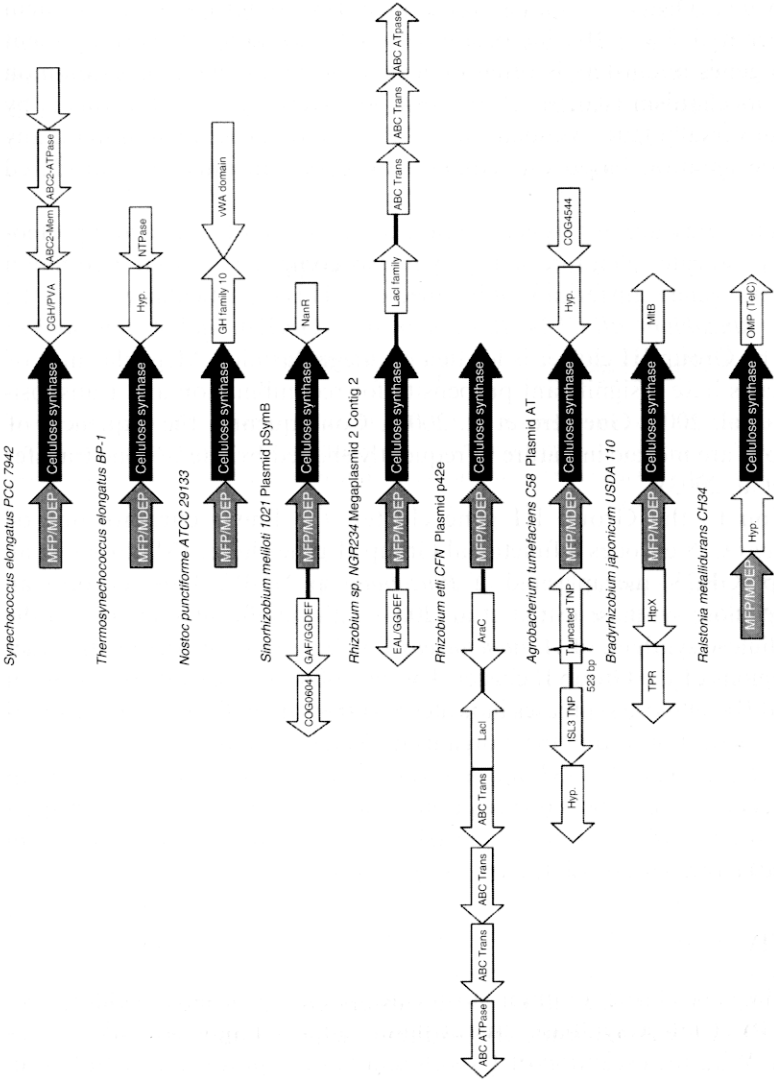
Figure 1-2. Group III gene clusters. MFP/MDEP—Membrane fusion protein/Multidrug efflux pump, CGH–Chololylglycine hydrolase, PIN–Nucleotide binding domain, GH–Glycosylhydrolase, vWA–von Willibrand factor, GGDEF–Diguanylate cyclase, NanR–NanR family transcriptional repressor, LacI–LacI family transcriptional repressor, AraC–AraC family transcriptional regulator, TNP–Transposase, TPR–TPR repeat region, HtpX – Heat shock protein, MltB–Lytic murein transglycosylase, Hyp.–Hypothetical protein, OMP–Outer membrane protein. The regions depicted here are not drawn to scale

cassette (ABC) transporter domains. This type of organization is characteristic of the bacterial ABC capsular polysaccharide exporter family (CPSE) in which secretion of polysaccharides is accomplished via the concerted actions of an ABC transporter, a membrane periplasmic auxiliary protein (MPA2 – analogous to the MFP associated with Type I bacterial secretion), an outer membrane auxiliary protein (OMA), and an as yet unidentified, outer membrane protein (OMP) (Silver et al. 2001). It is important to note that although the arrangement of bacterial genes in clusters is often indicative of components of a common pathway or mechanism (Korbel et al. 2004; Guerrero et al. 2005); this is by no means universally true. As such, in the absence of experimental data, any functional designation based on sequence organization must be considered speculative.

Although the presence of this gene organization in cyanobacteria and proteobacteria may indicate retention of key synthesis components from a common ancestor, lateral gene transfer cannot be ruled out. This is particularly true in the instances of *Rhizobium etli* CFN, *Rhizobium* sp. NGR234, and *A. tumefaciens* C58 where the Group III cluster is located on megaplasmids. Megaplasmids of the Rhizobiales have a significant propensity for recombination and transposition (Streit et al. 2004; Guerrero et al. 2005). Consequently, the sequences of these replicons are mosaic in nature – frequently shaped by lateral gene transfer (González et al. 2003).

In addition to the Group III gene cluster, the linear chromosome of *A. tumefaciens* C58 encodes a functional Group II gene cluster (Matthysse et al. 2005). The pNGR234 megaplasmid of *Rhizobium* sp. NGR234 also encodes an additional cellulose synthase (Streit et al. 2004) with significant similarity to the Group II cellulose synthase of *A. tumefaciens* C58 but does not possess the other conserved regions of the Group II cluster. The lack of sequence similarity between Group II and III cellulose synthases indicates that they are unlikely to be the products of gene duplication within these organisms. Rather, significant similarity with *Nostoc punctiforme* ATCC 29133 and *Synechococcus elongatus* PCC 7942 Group III cellulose synthase sequences (expectation values of 2e-56 – 1e-89 lower respectively, than observed when Group II and III sequences within the same organism are compared) indicates a possible lateral gene transfer.

5.3 Group IV

The Group IV gene cluster, found in nostocales, actinobacteria, and chloroflexales, encodes a cellulose synthase, an antisigma factor antagonist, and an antisigma factor. With the exception of those found in the chloroflexales, all Group IV clusters identified also contain an antisigma factor antagonist phophatase (Figure 1-3). The ancillary proteins surrounding the cellulose synthases in this group are homologous to SpoIIAA (antisigma factor antagonist), SpoIIAB (antisigma factor), and SpoIIE (antisigma factor antagonist phophatase) present in the regulatory clusters involved in the stage II sporulation of *Bacillus subtilis*.
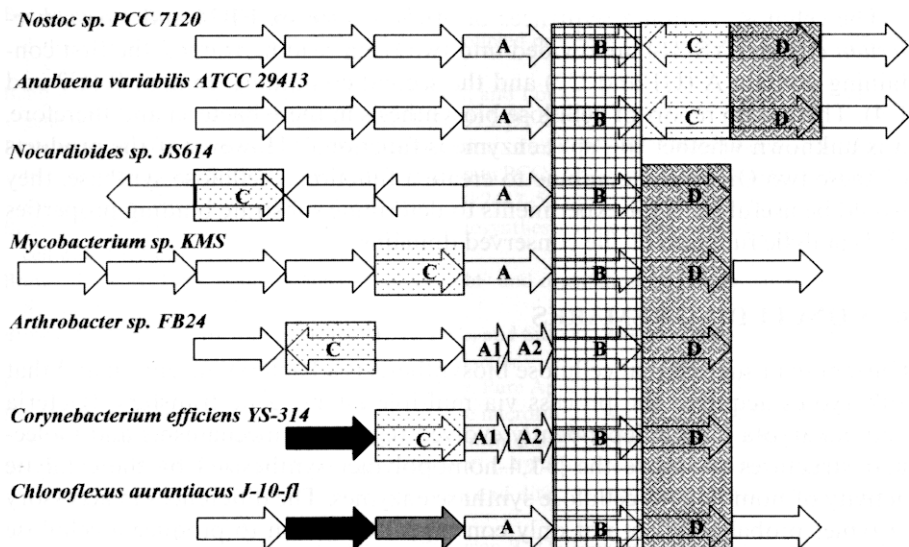
Figure 1-3. Group IV gene clusters shared by Nostocales, Actinobacteria, and Chloroflexales. **A** – Cellulose synthases or putative cellulose synthases, **B** – Antisigma factor antagonist, **C** – Sigma factor phosphatase, **D** – Antisigma factor regulation. Black arrows indicate putative diguanylate cyclases and gray arrow indicates bcsB homolog. The regions represented here are not drawn to scale

This regulatory system acts to confine gene transcription to the prespore during asymmetric cell division (Stragier and Losick 1996; Yudkin and Clarkson 2005). The presence of these regulatory components in close association with cellulose synthases may indicate controlled expression of cellulose synthase biosynthesis in differentiated cells during asymmetric cell division in the actinobacteria (spores) and nostocales (heterocysts and/or akinetes). With regard to this possibility in the heterocysts of the nostocales, it is interesting to note the existence of alternative sigma factors which are expressed only under nitrogen limiting conditions (Brahamsha and Haselkorn 1992).

The presence of syntenic genes in cyanobacteria and actinobacteria is consistent with phylogenetic trees demonstrating a close relationship of cyanobacteria and actinobacteria (Olsen et al. 1994; Yu et al. 2005). Interestingly, the organization of the gene cluster of *C. aurantiacus* is a chimera which contains elements of the Group IV gene cluster combined with a BcsB homolog and a diguanylate cyclase. Since the chloroflexales are generally considered to have branched prior to the divergence of cyanobacteria, gram positive bacteria, and proteobacteria (Olsen et al. 1994), this organization could represent the prototypical organization for Groups I, II, and IV. However, since *C. aurantiacus* exists primarily as a photoheterotroph living in close association with cyanobacteria, the possibility of lateral gene transfer from a cyanobacterium to members of the chloroflexales cannot be discounted.

The cellulose synthase sequences of *Arthrobacter* sp. FB24 and *Corynebacterium efficiens* YS-314 are divided into two open reading frames: the first containing domain A (U1 and U2) and the second containing domain B (U3 and U4). There is no report of cellulose biosynthesis in these bacteria and therefore, it is unknown whether this split enzyme is functional. However, if the products of these two ORFs can combine to create a functional cellulose synthase, they would be useful tools for experiments to determine substrate binding properties and catalytic function of the conserved domains.

6 CONCLUDING REMARKS

Current data suggest that cellulose biosynthesis is a bacterial invention and that eukaryotes acquired the process via multiple lateral gene transfers. Bacteria and eukaryota have independently evolved regulatory mechanisms and molecular structures to utilize the β-1,4-homopolymer synthesized by the catalytic activity of homologous cellulose synthase enzymes. The differences in accessory enzymes probably reflect not only convergent evolution to produce a cellulose I crystalline allomorph, but also inventions of alternative products such as cellulose II, noncrystalline cellulose, or nematic ordered cellulose.

As sequence databases continue to grow, it is certain that new cellulose synthase sequences and gene clusters will be identified. To be sure, increasing the library of available sequences is essential to the development of our understanding of the origin of cellulose biosynthesis and the evolutionary pathways utilized for its distribution. However, the primary challenges for researchers will be to elucidate the function of cellulose synthase associated enzymes and to characterize the cellulosic products synthesized by organisms with disparate enzymes and gene organizations. Without a firm grasp of the relationship of synthesizing components to the characteristics of the cellulosic product, we cannot hope to understand the genesis of the varied mechanisms and product morphologies we discover nor the evolutionary context from which they arose.

REFERENCES

Aloni Y., Delmer D.P., and Benziman M. 1982. Achievement of high rates of *in vitro* synthesis of 1, 4-beta-D-glucan: activation by cooperative interaction of the *Acetobacter xylinum* enzyme system with GTP, polyethylene glycol, and a protein factor. Proc Natl Acad Sci USA 79(21):6448–6452.

Amikam D. and Benziman M. 1989. Cyclic diguanylic acid and cellulose synthesis in *Agrobacterium tumefaciens*. J Bacteriol 171(12):6649–6655.

Amikam D. and Galperin M.Y. 2006. PilZ domain is part of the bacterial c-di-GMP binding protein. Bioinformatics 22(1):3–6.

Archibald J.M., Rogers M.B., Toop M., Ishida K., and Keeling P. 2003. Lateral gene transfer and the evolution of plastid targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. PNAS 100(13):7678–7683.

Ausmees N., Mayer R., Weinhouse H., Volman G., Amikam D., Benziman M., and Lindberg M. 2001. Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. FEMS Microbiol Lett 204:163–167.

Benziman M., Haigler C.H., Brown, Jr. R.M., White A.R., and Cooper K.M. 1980. Cellulose biogenesis: polymerization and crystallization are coupled processes in *Acetobacter xylinum*. PNAS USA 77:6678–6682.

Blanton R.L., Fuller D., Iranfar N., Grimson M.J., and Loomis W.F. 2000. The cellulose synthase gene of *Dictyostelium*. Proc Natl Acad Sci USA 97:2391–2396.

Brahamsha B. and Haselkorn R. 1992. Identification of multiple RNA polymerase sigma factor homologs in the cyanobacterium *Anabaena* sp. strain PCC 7120: cloning, expression, and inactivation of the sigB and sigC genes. J Bacteriol 173(8):2442–50.

Brett C.T. 2000. Cellulose microfibrils in plants: biosynthesis, deposition, and integration into the cell wall. Int Rev Cytol 199:161–99.

Brown, Jr. R.M. 1985. Cellulose microfibril assembly and orientation: recent developments. J Cell Sci Suppl 2:13–32.

Brown, Jr. R.M., Franke W.W., Kleinig H., Falk H., and Sitte P. 1969. A cellulosic wall component produced by the golgi apparatus. Science 166:894–896.

Brown, Jr. R.M. 1996. The biosynthesis of cellulose. Pure Appl Chem 10:1345–1373.

Brown, Jr. R.M. and Montezinos D. 1976. Cellulose microfibrils: visualization of biosynthetic and orienting complexes in association with the plasma membrane. Proc Natl Acad Sci USA 73:143–147.

Brown, Jr. R.M., Willison J.H.M., and Richardson C.L. 1976. Cellulose biosynthesis in *Acetobacter xylinum*: 1. Visualization of the site of synthesis and direct measurement of the *in vivo* process. Proc Natl Acad Sci USA 73(12):4565–4569.

Dehal P., Satou Y., Campbell R.K., Chapman J., Degnan B., De Tomaso A., Davidson B., Di Gregorio A., Gelpke M., Goodstein D.M., Harafuji N., Hastings K.E., Ho I., Hotta K., Huang W., Kawashima T., Lemaire P., Martinez D., Meinertzhagen I.A., Necula S., Nonaka M., Putnam N., Rash S., Saiga H., Satake M., Terry A., Yamada L., Wang H.G., Awazu S., Azumi K., Boore J., Branno M., Chin-Bow S., DeSantis R., Doyle S., Francino P., Keys D.N., Haga S., Hayashi H., Hino K., Imai K.S., Inaba K., Kano S., Kobayashi K., Kobayashi M., Lee B.I., Makabe K.W., Manohar C., Matassi G., Medina M., Mochizuki Y., Mount S., Morishita T., Miura S., Nakayama A., Nishizaka S., Nomoto H., Ohta F., Oishi K., Rigoutsos I., Sano M., Sasaki A., Sasakura Y., Shoguchi E., Shin-i T., Spagnuolo A., Stainier D., Suzuki M.M., Tassy O., Takatori N., Tokuoka M., Yagi K., Yoshizaki F., Wada S., Zhang C., Hyatt P.D., Larimer F., Detter C., Doggett N., Glavina T., Hawkins T., Richardson P., Lucas S., Kohara Y., Levine M., Satoh N., and Rokhsar D.S. 2003. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298:2157–2167.

Deinema M.H. and Zevenhuizen L.P. 1971. Formation of cellulose fibrils by gram-negative bacteria and their role in bacterial flocculation. Arch Mikrobiol 78(1):42–51.

Delmer D.P. 1999. Cellulose biosynthesis: exciting times for a difficult field of study. Annu Rev Plant Phys Plant Mol Biol 50:245–276.

Doolittle W. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 14:307–311.

García B., Latasa C., Solano C., García-del Portillo F., Gamazo C., and Lasa I. 2004. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. Mol Microbiol 54(1):264–277.

Giddings J.R. TH, Brower D.L., and Staehelin L.A. 1980. Visualization of particle complexes in the plasma membrane of *Micrasterias denticulata* associated with the formation of cellulose fibrils in primary and secondary cell walls. J Cell Biol 84 (2):327–339.

González V., Bustos P., Ramirez-Romero M.A., Medrano-Soto A., Salgado H., Hernandez-Gonzalez I., Hernandez-Celis J.C., Quintero V., Moreno-Hagelsieb G., Girard L., Rodriguez O., Flores M., Cevallos M.A., Collado-Vides J., Romero D., and Davila G. 2003. The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. Genome Biol 4(6):R36.

Grimson M.J., Haigler C.H., and Blanton RL. 1996. Cellulose microfibrils, cell motility, and plasma membrane protein organization change in parallel during culmination in *Dictyostelium discoideum*. J Cell Sci 109 (Pt 13):3079–3087.