A standardized synthetic *Eucalyptus* transcription factor and promoter panel for re-engineering secondary cell wall regulation in biomass and bioenergy crops

Steven G. Hussey[1], Jacqueline Grima-Pettenati[2], Alexander A. Myburg[1], Eshchar Mizrachi[1], Siobhan M. Brady[3], Yasuo Yoshikuni[4], Samuel Deutsch[4]

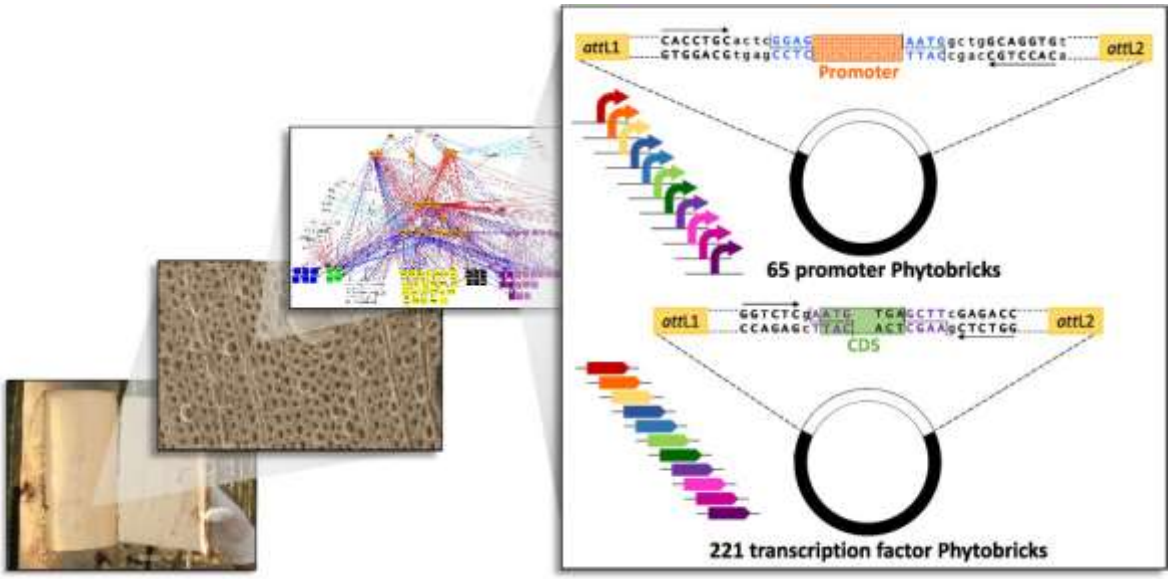
[1]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), University of Pretoria, Private Bag X28, Pretoria, South Africa 0002

[2]Laboratoire de Recherche en Sciences Végétales (LRSV), Université Toulouse, UPS, CNRS, BP 42617, F-31326, Castanet-Tolosan, France

[3]Department of Plant Biology and Genome Center, University of California, Davis, California 95616, United States

[4]US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, United States

Graphical abstract



Abstract

Re-engineering transcriptional networks regulating secondary cell wall formation may allow the improvement of plant biomass in widely grown plantation crops such as *Eucalyptus*. However, there is currently a scarcity of freely available standardized biological parts (e.g. Phytobricks) compatible with Type IIS assembly approaches from forest trees, and there is a need to accelerate transcriptional network inference in non-model biomass crops. Here we describe the design and synthesis of a versatile three-panel biological parts collection of 221 secondary cell wall-related *Eucalyptus grandis* transcription factor coding sequences and 65 promoters that are compatible with GATEWAY, Golden Gate, MoClo and GoldenBraid DNA assembly methods and generally conform to accepted Phytobrick syntaxes. This freely available resource is intended to accelerate synthetic biology applications in multiple plant biomass crops and enable reconstruction of secondary cell wall transcriptional networks using high-throughput assays such as DNA Affinity Purification sequencing (DAP-seq) and enhanced yeast one-hybrid (eY1H) screening.

Keywords: secondary cell wall, Phytobrick, *Eucalyptus*, transcriptional network

Transcription factors (TFs) regulating secondary cell wall (SCW) deposition and their associated gene promoters comprise a molecular toolbox for re-engineering SCW deposition[1]. By understanding the regulatory architecture of the transcriptional networks they form, novel network linkages may be engineered or transgenes expressed in a cell type-specific fashion such that plant biomass may be tailored for renewable biomaterials production and ever-diversifying biorefinery applications. Achieving this via synthetic biology (SynBio) is challenged by a limited availability of plant standard biological parts compatible with one-step multipartite DNA assembly approaches and cost-effective methods for elucidating SCW transcriptional networks in non-model crops. Following the standardization of biological parts in prokaryotic SynBio (e.g. BioBrick™), the plant SynBio community is beginning to adopt the Phytobrick standard characterised by Type IIS assembly methods such as Golden Gate, MoClo and GoldenBraid, coupled with a standard lexicon[2]. These assembly standards aim to assemble complex multipartite DNA constructs using a minimal set of restriction endonuclease in a single reaction vessel, beginning with separate standardized plasmids. Here, we describe the design and synthesis of a versatile three-panel biological parts collection of 221 SCW-related *Eucalyptus grandis* TF coding sequences (CDSs) and 65 SCW-related promoters that are compatible with various DNA assembly standards and generally conform to accepted plant SynBio syntaxes.

Commercial *Eucalyptus* plantations comprise the most widely planted hardwoods globally, and the genus has become a useful model for studying woody biomass accumulation[3]. This motivates the need to understand the transcriptional control of *Eucalyptus* wood formation using high-throughput assays such as DNA Affinity Purification sequencing (DAP-seq)[4] and enhanced yeast one-hybrid analysis (eY1H)[5]. The collection, distributed by the University of Pretoria, is freely accessible to the academic community through a US Department of Energy (DOE) Joint Genome Institute (JGI) material transfer agreement. More information on the constructs can be obtained from //fabinet.up.ac.za/index.php/fmg-projects/synthetic-biology.

Candidate SCW-related TFs from *E. grandis* were identified based on homology to known *Arabidopsis* SCW regulators[1] using the annotation files of the *E. grandis* v2.0 genome sequence (//phytozome.jgi.doe.gov/) and evidence implicating them in *E. grandis* SCW regulation from studies of MYB and NAC protein families and yeast two-hybrid screens[6-8]. The candidates were prioritised to select a total of ~303 Kb of sequence based on preferential expression in developing xylem (among seven tissues; //eucgenie.org/), association with *Eucalyptus* biomass and bioprocessing traits (Dataset 1 from Mizrachi *et al.*[9]) and differential expression during tension wood formation[10]. The prioritised candidates thus had at least one evidence line linking them to wood formation in *Eucalyptus*. CDSs of 261 prioritised TFs from the *E. grandis* v2.0 genome (//phytozome.jgi.doe.gov/) were then "domesticated", entailing the synonymous substitution of the sequences to enhance synthetic gene synthesis and remove internal BsaI, BpaI and BsmBI sites required for Type IIS assembly. Two TF panels were successfully synthesized, with 40 TF candidates failing gene synthesis. The first comprises a Phytobrick parts panel of 173 constructs compatible with all Type IIS and GATEWAY approaches, containing standard syntax sequences[2] flanking each CDS followed by a convergent BsaI restriction site for Golden Gate assembly, Gibson-cloned into pCR8™/GW/TOPO® (Invitrogen™) via a "chew back" linker (Fig. 1A). We included bordering *att*L1 and *att*L2 sites to enable GATEWAY recombination to *att*R-enabled destination vectors, with spacers between the BsaI and *att*L sequences such that destination vector-encoded N-terminal tags remain in frame. The second panel, intended for DAP-seq analysis primarily, comprises 48 CDSs cloned into pIX-HALO in-frame with an N-terminal HALO (haloalkane dehalogenase) tag for protein purification using Promega Magne® HaloTag® Beads during the DAP-seq assay[4] (Fig. 1B). While standard syntax sequences and BsaI assembly sites are not included in this panel due to the presence of the tag, the constructs are otherwise domesticated and can easily be converted into Phytobricks by PCR amplification of the CDS with primers containing BpiI recognition sites and compatible overhangs, followed by one-step digestion-ligation into a Universal Acceptor Plasmid[2].

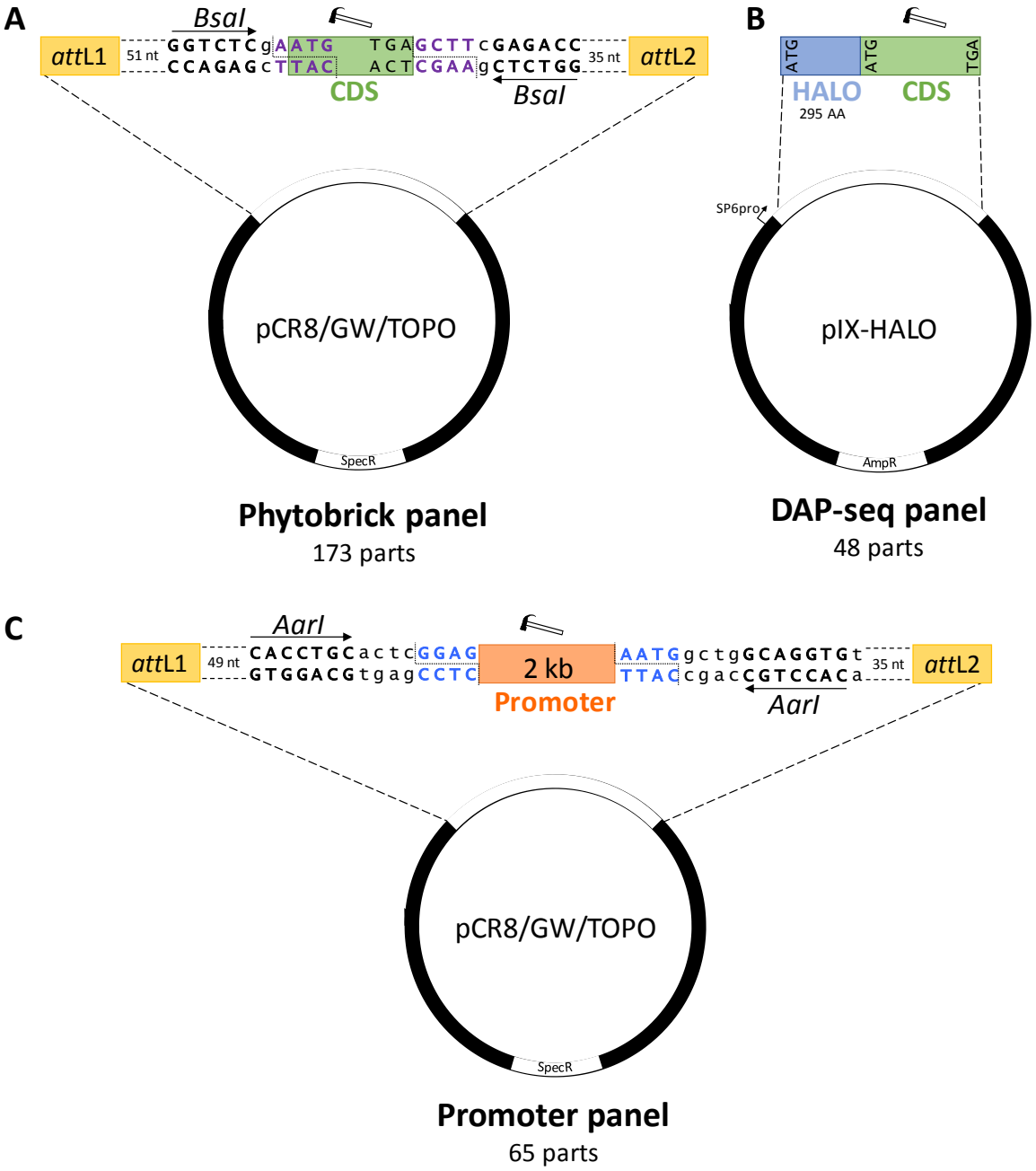


Figure 1. Design of standardized synthetic *E. grandis* SCW-related transcription factors and promoters. Hammer symbols represent domesticated sequences, dotted lines indicate restriction cleavage sites, coloured bases indicate standard syntax sequences and lowercase sequences indicate spacer nucleotides. **(A)** Transcription factor Phytobricks contain *att*L GATEWAY recombination sites (orange), chew back linkers, BsaI Type IIS recognition sites and standard syntax sequences (purple text). The start codon of the domesticated coding sequence (green) remains in frame with N-terminal tags in GATEWAY destination vectors, while the panel is primarily intended for Golden Gate, MoClo and GoldenBraid assembly. **(B)** The DNA Affinity Purification Sequencing (DAP-seq) panel of transcription factors is available as a C-terminal fusion to the HALO purification tag, intended for *in vitro* transcription and translation via the SP6 phage promoter in pIX-HALO. While not standardized by a universal syntax, the coding sequences are sequence-

domesticated and can thus be subcloned as standard parts into a universal acceptor plasmid. **(C)** The secondary cell wall promoter panel features 2 kb promoter sequences (including 5' UTRs) compatible with GATEWAY and AarI-mediated Golden Gate cloning. Standard prefix and suffix syntax sequences (blue text) allow for two-step Type IIS assembly to any Phytobrick panel CDS in (A). AmpR, ampicillin resistance gene; CDS, coding sequence; SpecR, spectinomycin resistance gene.

The promoter panel features 65 SCW-related structural genes (n = 50) and TFs (n = 15) in pCR8™/GW/TOPO® (Fig. 1C), successfully produced from 90 prioritized candidates (~180 Kb) selected using similar evidence lines as for the TF panels, as well as a set of what are likely bona fide cellulose, hemicellulose and lignin biosynthetic genes in *Eucalyptus*[3]. Since the sequences include the 5' UTR, we implemented the GGAG prefix and AATG suffix[2], allowing the promoter standards to be appended to any Phytobrick CDS. Due to the frequent occurrence of BsaI and other illegal sites in these sequences, sequence substitution of which could affect promoter function, the AarI heptanucleotide recognition sequence was adopted due to its considerably lower frequency relative to the hexanucleotide signatures of BsaI, BpaI and BsmBI recognition sequences. AarI cleavage yields tetranucleotide overhangs similar to BsaI and is hence Type IIS assembly-compatible. A limitation of this strategy is that BsaI-digested Phytobrick CDS fragments may need to be gel-purified prior to digestion-ligation with a particular promoter. The few illegal AarI sites found in the wild-type promoter sequences were eliminated, for all but seven constructs, by substitution with naturally occurring *E. grandis* SNPs (Myburg *et al*., unpublished). To enable compatibility of the promoter panel with GATEWAY cloning, in which case the AATG suffix opens the reading frame upstream of destination vector-encoded reporters, spacers were added such that the AarI recognition sequence, chew back linker and resulting *att*B2 site following LR recombination collectively encode a 44-residue peptide in-frame with the reporter. We also avoided repetitive microsatellites flanking the target sequences. This approach yielded construct lengths of 1,995±72 bp.

Compared to microbial prokaryotic SynBio, plant SynBio is lagging with respect to the availability of open-source standard biological parts. Here we have described what is to the best of our knowledge the largest publically available standard biological parts collection of a forest tree yet produced. These synthetic DNA constructs are chiefly intended to provide forest biotechnologists with a molecular toolkit to modify and study wood formation in forest trees and enable the reconstruction of the SCW transcriptional network using high-throughput assays. However, the toolkit also has the potential to support SynBio innovation in other plant biomass crops. Moving forward, we propose that even basic cloning experiments should attempt to create Phytobrick

standards of DNA sequences wherever possible to advance the plant biotechnology community as a whole to a state of SynBio readiness.

Acknowledgements

SH, JP-G, EM and AM compiled and prioritized candidate genes, where SH, AM and EM are co-PIs and JP-G is a grant collaborator. SH drafted the manuscript and led the synthetic panel design in consultation with SD, YY and Nicola Patron (Earlham Institute, UK). SMB, a grant collaborator supported by an HHMI Faculty Scholar fellowship, provided expertise in eY1H. We thank Ronan O'Malley (Sequencing Technologies Group Leader, DOE Joint Genome Institute) for advice on DAP-seq. Gene synthesis was funded by the DNA Synthesis Science Programme of the DOE Joint Genome Institute, in partnership with Integrated DNA Technologies. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy (Contract No. DEAC02-05CHI/231). Profiling and selection of gene candidates was supported by the Department of Science and Technology (Strategic Grant for the *Eucalyptus* Genomics Platform), the National Research Foundation of South Africa (Bioinformatics and Functional Genomics Programme, Grants IUD 86936 and 97911), the Technology and Human Resources for Industry Programme (grant UID 80118) and by Sappi and Mondi through the Forest Molecular Genetics Programme (University of Pretoria). The authors declare no conflict of interest.

References

1. Hussey, S. G., Mizrachi, E., Creux, N. M., and Myburg, A. A. (2013) Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Front Plant Sci 4*, 325.

2. Patron, N. J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farré, G., Rogers, C., et al. (2015) Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol 208*, 13-19.

3. Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., et al. (2014) The genome of *Eucalyptus grandis*. *Nature 510*, 356-362.

4. Bartlett, A., O'Malley, R. C., Huang, S. C., Galli, M., Nery, J. R., Gallavotti, A., and Ecker, J. R. (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc 12*, 1659-1672.

5. Gaudinier, A., Zhang, L., Reece-Hoyes, J. S., Taylor-Teeples, M., Pu, L., Liu, Z., Breton, G., Pruneda-Paz, J. L., Kim, D., Kay, S. A., et al. (2011) Enhanced Y1H assays for Arabidopsis. *Nat Methods 8*, 1053-1055.

6. Hussey, S. G., Saidi, M. N., Hefer, C. A., Myburg, A. A., and Grima-Pettenati, J. (2015) Structural, evolutionary and functional analysis of the NAC domain protein family in *Eucalyptus*. *New Phytol 206*, 1337-1350.

7. Soler, M., Camargo, E. L. O., Carocha, V., Cassan-Wang, H., Clemente, H. S., Savelli, B., Hefer, C. A., Paiva, J. A. P., Myburg, A. A., and Grima-Pettenati, J. (2015) The *Eucalyptus grandis* R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. *New Phytologist 206*, 1364-1377.

8. Soler, M., Plasencia, A., Larbat, R., Pouzet, C., Jauneau, A., Rivas, S., Pesquet, E., Lapierre, C., Truchet, I., and Grima-Pettenati, J. (2017) The *Eucalyptus* linker histone variant EgH1.3 cooperates with the transcription factor EgMYB1 to control lignin biosynthesis during wood formation. *New Phytol 213*, 287-299.

9. Mizrachi, E., Verbeke, L., Christie, N., Fierro, A. C., Mansfield, S. D., Davis, M. F., Gjersing, E., Tuskan, G. A., Van Montagu, M., Van de Peer, Y., et al. (2017) Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing. *Proc Natl Acad Sci U S A 114*, 1195-1200.

10. Mizrachi, E., Maloney, V. J., Silberbauer, J., Hefer, C. A., Berger, D. K., Mansfield, S. D., and Myburg, A. A. (2015) Investigating the molecular underpinnings underlying morphology and changes in carbon partitioning during tension wood formation in *Eucalyptus*. *New Phytol 206*, 1351-1363.