Escherichia coli expressing chloroplast chaperones as a proxy to test heterologous Rubisco production in leaves.

Sally Buck[1], Tim Rhodes[1], Matteo Gionfriddo[1], Tanya Skinner[1], Ding Yuan[1], Rosemary Birch[1], Maxim V. Kapralov[2], Spencer M. Whitney[1]*

[1]ARC Centre of Excellence in Translational Photosynthesis, Australian National University, Canberra, Australia 2000

[2]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

*Corresponding author: Spencer.whitney@anu.edu.au

ORCID; SB, 0000-0001-9569-0160: TR, 0000-0002-4146-7913: MG, 0000-0003-2364-0812: TS, 0000-0002-2661-6415: DY, 0000-0001-7895-0944: RB, 0000-0002-3050-7406: MK, 0000-0001-7966-0295: SMW, 0000-0003-2954-2359

Short title: *Escherichia coli* as a surrogate for Rubisco expression in plastids

Abstract

Rubisco is a fundamental enzyme in photosynthesis and therefore for life. Efforts to improve plant Rubisco performance have been hindered by the enzymes' complex chloroplast biogenesis requirements. New Synbio approaches however now allow the production of some plant Rubisco isoforms in *E. coli*. While this enhances opportunities for catalytic improvement, there remain limitations in the utility of the expression system. Here we generate, optimise and test a robust Golden Gate cloning *E. coli* expression system incorporating the protein folding machinery of tobacco chloroplasts. By comparing the expression of different plant Rubiscos in both *E. coli* and plastome transformed tobacco we show the *E. coli* expression system can accurately predict high level Rubisco production in chloroplasts but poorly forecasts the biogenesis potential of isoforms with impaired production *in planta*. We reveal heterologous Rubisco production in *E. coli* and tobacco plastids poorly correlates with Rubisco large subunit phylogeny. Our findings highlight the need to fully understand the factors governing Rubisco biogenesis if we are to deliver an efficient, low cost screening tool that can accurately emulate chloroplast expression.

Highlight:

Development of a modular Golden-Gate cloning system for expressing plant Rubisco and tobacco chloroplast chaperones in *E. coli* shows limited resolution in its capacity to forecast the level of heterologous Rubisco biogenesis in tobacco.

Keywords

CO_2-fixation, photosynthesis, chloroplast transformation, synthetic biology, Golden Gate cloning.

Introduction

Ribulose-1,5-bisphosphate (RuBP) carboxylase/oxygenase (Rubisco), a critical enzyme in the Calvin cycle of photosynthesis, often poses the cycles' rate-limiting step in plants due to functional inefficiencies. Rubisco catalysis involves 5 partial reactions spanning the binding of its 5-carbon substrate RuBP, facilitating its reaction with CO_2, and cleaving this 6-carbon intermediate into two molecules of 3-phosphoglycerate (3PGA). The complexity of this catalytic chemistry, combined with the complicated biogenesis requirements of plant Rubisco and phylogenetic constraints on its evolution, have imposed limitations on the enzymes' capacity to increase CO_2-fixation rates above 5 cycles per second in plants (Bouvier *et al.*, 2021; Bracher *et al.*, 2017; Cummins, 2021; Sharwood, 2017). Rubisco is also non-specific and can fix O_2 to RuBP to produce 3PGA and 2-phosphoglycolate (2PG). Recycling toxic 2PG into 3PGA employs the CO_2-releasing and energetically costly process of photorespiration (Busch, 2020). Improving the CO_2-affinity, fixation rate and specificity for CO_2 over O_2 has thus been a long-standing target to enhance plant photosynthesis, growth and yield (Evans and Lawson, 2020; Iñiguez *et al.*, 2021). Impeding success has been the inability of prokaryotic expression systems to meet the biogenesis requirements of either plant or algae Rubisco (Mueller-Cajar and Whitney, 2008; Sharwood, 2017), in addition to plant chloroplasts not being able to meet the assembly needs of more efficient Rubiscos found naturally in some red algae (Lin and Hanson, 2018; Whitney *et al.*, 2001).

Underpinning the challenge of bioengineering plant Rubisco is its complex folding and assembly requirements. Folding and assembly of the chloroplast made 50 kDa Rubisco large catalytic subunit (RbcL) into a hexamer $RbcL_8$ core comprising four $RbcL_2$ dimers requires the sequential action of at least seven auxiliary proteins (Aigner *et al.*, 2017; Bracher *et al.*, 2017; Conlan and Whitney, 2018; Vitlin Gruber and Feiz, 2018). Folding of the nascent RbcL peptide is facilitated by the ~840 kDa chloroplast chaperonin protein folding cylinders. In plants these hetero-oligomer, double heptameric ring complexes comprise 60 kDa CPN60α and CPN60β subunits (Wilson and Hayer-Hartl, 2018). A co-chaperonin lid comprising CPN20 and CPN10 subunits closes over one end of the cylinder upon encasement of an unfolded RbcL to facilitate its folding (Bracher *et al.*, 2017; Tsai *et al.*, 2012). The four other auxiliary proteins include the 45 kDa Rubisco Accumulation Factor 1 (RAF1, (Feiz *et al.*, 2012; Hauser *et al.*, 2015)), 18 kDa Rubisco Accumulation Factor 2 (RAF2, (Feiz *et al.*, 2014)), 10 kDa Bundle Sheath Deficient 2 (BSD2, (Aigner *et al.*, 2017; Brutnell *et al.*, 1999)) and 15 kDa RbcX (Bracher *et al.*, 2011; Liu *et al.*, 2010; Saschenbrecker *et al.*, 2007)

chaperones. Although the mechanism of the plant Rubisco biogenesis pathway is not fully resolved, the chaperones appear mostly specific to sequentially assembling the CPN60-folded RbcL through $RbcL_2$ to $RbcL_8$-chaperone bound intermediary complexes onto which two tetrads of 15 kDa Rubisco small subunits (RbcS) bind to form a functionally active, structurally stable, ~520 kDa L_8S_8 holoenzyme (Wilson and Hayer-Hartl, 2018).

Traditional plant Rubisco bioengineering strategies have relied on genetic transformation of the nucleus or chloroplast genome (plastome), or sometimes both (Martin-Avila *et al.*, 2020; Sharwood, 2017). In most instances the target has been transforming the plastome *rbcL* gene in tobacco, the model plant for chloroplast transformation (Whitney and Sharwood, 2021). The genetic precision of homologous recombination for directing transgene integration into the plastome has been used extensively to replace the native L_8S_8 tobacco enzyme by inserting *rbcL* (±*rbcS*) genes coding a range of phylogenetically divergent bacterial and Archaea Rubisco isoforms ((Gunn *et al.*, 2020) and reviewed in (Sharwood, 2017)). Replacing the tobacco *rbcL* with other plant *rbcL* genes has also been used to examine the kinetics of chimeric enzymes comprising the foreign plant RbcL and the endogenous tobacco RbcS. While this has proved useful in identifying catalytic switches (Whitney *et al.*, 2011b), these studies have revealed extensive variation in the biogenesis compatibility among different plant RbcL's despite sharing over 95% pairwise identity. For example, the biogenesis of chimeric Rubisco comprising Arabidopsis, sunflower or one of three different *Flaveria* RbcL's was impaired 50 to 75% in tobacco chloroplasts (Sharwood *et al.*, 2008; Whitney *et al.*, 2015; Whitney *et al.*, 2011b) while the assembly requirements of RbcL from two Panicum grasses were not met at all (Sharwood *et al.*, 2016). By contrast, the assembly needs of tomato and potato RbcL were fully met in tobacco, presumably as both Solanaceae RbcL differ by only 4 and 11 amino acids, respectively, from tobacco RbcL (Martin-Avila *et al.*, 2020; Zhang *et al.*, 2011). What remains unknown are the structural features in RbcL that influence structural compatibility with one or more chaperones in the chloroplast protein folding and assembly machinery. As a result, we remain unable to predict the potential for chimeric plant Rubisco biogenesis in tobacco or any other plant species amenable to chloroplast transformation. This is particularly important given the rate of photosynthesis under light saturation is primarily limited by the content and kinetics of Rubisco (Sharwood, 2017; von Caemmerer, 2020). As such, enhancing plant growth through Rubisco is not only limited by the identification of improved Rubiscos, but also by the ability

of these Rubiscos to be functionally expressed, and in sufficient quantities, to ensure above wild-type rates of photosynthetic carbon assimilation.

Knowledge of the accessory proteins involved in plant Rubisco biogenesis was first utilised to successfully produce Arabidopsis and tobacco Rubisco in *E. coli* (Aigner *et al.*, 2017). This fundamental breakthrough promised to broaden, and accelerate, the slow and costly traditional plant transformation bioengineering approaches (Conlan and Whitney, 2018; Wilson *et al.*, 2019). To date the technology has been used to study how the yield and reliability of Rubisco expression in *E. coli* can be improved; demonstrate the production of RbcL-protein intermediary complexes; show the RbcS produced in trichomes, but not the six RbcS produced in tobacco mesophyll cells, differentially impacts Rubisco catalysis; and test mutations identified by phylogenetic reconstruction of Solanaceae RbcL sequences that show promising impacts on tobacco Rubisco kinetics (Aigner *et al.*, 2017; Lin *et al.*, 2022; Lin *et al.*, 2020; Wilson *et al.*, 2019). Overlooked so far has been the suggestion to adapt plant Rubisco expression in *E. coli* into a Rubisco directed evolution system to facilitate identifying catalysis enhancing mutations (Conlan and Whitney, 2018; Gionfriddo *et al.*, 2019). Such directed evolution approaches have to date been the only means to simultaneously improve the carboxylation rate, efficiency and specificity of Rubisco, albeit only with non-plant Rubisco isoforms (Wilson *et al.*, 2016; Wilson *et al.*, 2018; Zhou and Whitney, 2019).

The promising possibility plant Rubisco may be amenable to catalytic improvement by directed evolution in *E. coli* begs the question as to what extent plant Rubisco production in *E. coli* producing both its native and chloroplast protein folding and assembly machinery can serve as a reliable proxy for Rubisco biogenesis in a plant chloroplast? To address this question, here we compare the production of tobacco and chimeric plant Rubiscos in both tobacco chloroplasts and in *E. coli* expressing the RAF1, RAF2, RbcX, BSD2, CPN60α, CPN60β & CPN20 chaperones from tobacco. This was achieved by firstly generating *rbcL* transplastomic tobacco lines coding the RbcL from 15 phylogenetically distant, some commercially important, plant species. Secondly a new, modular, Golden Gate cloning toolkit was developed to generate a tobacco chloroplast chaperone expressing BL21 Star *E. coli* line (tcc*E. coli*) to compare tobacco and chimeric Rubisco biogenesis. Described here is the development of a two-plasmid expression system, one plasmid coding all 7 tobacco chloroplast chaperones, the other coding interchangeable plant RbcL and RbcS isoforms. Rubisco expression optimisation and quantification analyses show that while there are

parallels between Rubisco biogenesis in tobacco leaves and tcc*E. coli*, the resolution by which L_8S_8 expression in tcc*E. coli* can forecast the level of heterologous Rubisco production in tobacco chloroplasts is relatively limited.

Materials and Methods

Gene synthesis, cloning and transformation

All genes were supplied as gBlocks from Twist Biosciences. All gene codon use was modified to match the tobacco *rbcL* gene (gene sequences supplied in Supplemental data file S1). The ten new plant *rbcL* genes (corresponding RbcL accessions in Figure 1A) were cloned as 1439-bp *Nhe*I-*Sal*I fragments into pLEV4 and each pLEV-*rbcL* plasmid transformed into the tob*Rr* plastome as described ((Whitney *et al.*, 2015); Figure 1A). Of the 3 to 10 spectinomycin-resistant plants obtained from 5 leaf bombardments for each genotype, at least two independent lines were maintained through to homoplasmicity as determined by native PAGE as described (Whitney and Sharwood, 2008). The tobacco genotypes expressing the *rbcL* gene from dicot species were grown in soil to maturity in air supplemented with 0.5% (v/v) CO_2 and fertilized with wild-type pollen, as described previously (Whitney and Sharwood, 2008). The genotypes expressing monocot *rbcL* were unable to grow in soil and had to be maintained in tissue culture as described (Sharwood *et al.*, 2016) without success in generating seeds.

A tobacco Rubisco and chaperone cloning kit for *E. coli* expression was assembled by Golden Gate assembly. The kits incorporated modified components of the EcoFlex Kit (Moore *et al.*, 2016) into 9 new RFP/GFP coding Level 0 (pBP), Level 1 (pTU1), Level 2 (pTU2), pET28a(+) and pCDFDuet-1 based acceptor vectors incorporating differing *BsaI* and *BsmbI* (*Esp3I*) cloning sites (Table S1). Cloning sites not included in the gBlocks were introduced by PCR using Phusion polymerase (ThermoFisher) and gene specific primers (see Table S2) that introduced gene flanking *BsaI* sites with appropriate 4-base overhangs to the Level 0 acceptor vector required (Figure S1). Golden Gate cloning was performed according to Engler et al. (2008) with all Level 1 plasmids (coding a T7 promoter, a gene and a Bba_0015 terminator) fully sequenced using BigDye terminator sequencing (Macrogen). The genetic composition of Level 2 and 3 plasmids was confirmed by restriction enzyme digests and PCR mapping using combinations of gene specific primers.

Tobacco growth and sampling.

At least two plants of wildtype and each T_1 transplastomic tobacco line were germinated in tissue culture pots and grown to ~5 cm in height (~6 leaf stage juvenile plants) before transferring to 3 L pots of soil in a temperature-controlled 25°C glasshouse under natural summer illumination (Canberra, Australia). Plants were watered daily and fertilised twice weekly with Hoaglands nutrients. When 35 to 40 cm in height, replica 0.5 cm^2 samples from an upper canopy leaf of comparable physiological age (5^{th} leaf from apical meristem, 13 ± 1 cm in diameter) were rapidly frozen in liquid N_2 and stored at -80°C.

Tobacco and chimeric plant Rubisco expression in tcc*E. coli* BL21 Star.

Electrocompetent tcc*E. coli* cells were generated by transforming plasmid pCDF*Nt*Assembl into BL21 Star *E. coli* before transforming in a pET-*rbc*LS construct coding either tobacco Rubisco, chimeric plant Rubisco (coding a heterologous plant RbcL and the highest expressed tobacco RbcS (Donovan *et al.*, 2020)) or wheat, maize or rice Rubisco (coding the cognate RbcL and RbcS of each monocot). Overnight cultures (5 mL) grown at 37°C of independent tcc*E. coli* colonies (± a co-transformed pET-*rbc*L(±S) plasmid) were used to inoculate 100 mL LB-media and grown at 23°C to an OD_{600} of 0.5 before inducing with 1 mM IPTG. After 12, 24 and 48 hours at 23, 30 or 37°C (or 24 h at 23°C for cells expressing the chimeric and monocot Rubiscos) aliquots of the cells were harvested by centrifugation (8,000*g*, 5 min, 4°C), the cell pellets frozen in liquid N_2 and stored at -80°C.

Rubisco and protein quantification, PAGE and immunoblot analysis.

Total leaf protein was isolated in ice-cold extraction buffer (50 mM EPPES-NaOH, pH 8.0, 15 mM $MgCl_2$, 0.5 mM EDTA, 20 mM $NaHCO_3$, 5 mM DTT, 0.1% [v/v] plant protease inhibitor; Sigma-Aldrich) using 2 mL glass tissue homogenizers (Wheaton) and the soluble protein retained after centrifugation (1 min, 2°C, 16,000*g*). Samples of *E. coli* total protein were collected following cell disruption with a Q125 Sonicator (Qsonica, 30 sec pulse at 60% amplitude) in 1 mL ice cold extraction buffer and the soluble protein collected following centrifugation (1 min, 2°C, 16,000*g*). The Rubisco content in the leaf and *E. coli* soluble protein fractions was quantified by [^{14}C]-CABP binding (Whitney and Sharwood, 2021) and the protein concentration determined using the Coomassie Plus Protein Assay Reagent (Pierce) against BSA. See supplemental data file 1 for raw data. Samples of the total and soluble proteins were analysed by SDS PAGE (4 to 12% Bis-Tris gel; Invitrogen) and the soluble protein by native PAGE (4 to 12% Tris-Gly gels; Invitrogen) as described previously

(Whitney and Sharwood, 2007). The separated protein bands were either stained with GelCode Blue Stain Reagent (Invitrogen) or blotted onto Hybond-C Extra nitrocellulose membrane using an XCell II Blot Module (Invitrogen), and the immuno-reactive proteins recognised by primary antibodies raised in rabbits against purified tobacco Rubisco, RAF1 (*Nt*RAF1, (Whitney *et al.*, 2015)) and BSD2 (*Nt*BSD2, (Conlan *et al.*, 2019)), followed by the αRabbit-HRP conjugated secondary antibody were visualized using a Pharos FX Plus Molecular imager (Bio-Rad) as described (Martin-Avila *et al.*, 2020).

Results

Plastome transformation of heterologous plant *rbcL* genes into tobacco.

Synthesized *rbcL* genes coding the RbcL from 7 dicotyledonous (dicot) and 3 monocotyledonous (monocot) angiosperms were cloned into pLEV4, a plastome transforming plasmid that directs transgene integration at the *rbcL* locus in the tobacco plastome (Whitney *et al.*, 2015) (Figure 1A). As shown previously, the genetic design of the *rbcL-aadA* operon in pLEV4 produces wild-type levels of tobacco *rbcL* mRNA, an 80-90% less abundant *rbcL-aadA* transcript, and wild-type levels of tobacco Rubisco (Martin-Avila *et al.*, 2020; Whitney *et al.*, 2015). In the 10 pLEV-*rbcL* plasmids produced, each *rbcL* transgene shared the native tobacco *rbcL* promoter, 5′-UTR and the first 59 nucleotides of 5′-coding sequence (*i.e.* RbcL codon 19, supplemental data file 1), thus avoiding the impact that nucleotide substitutions in this region might have on RbcL translation (Kuroda and Maliga, 2001). Transformation of each pLEV-*rbcL* construct into the L_2-Rubisco producing tob*Rr* 'master-line' (Martin-Avila *et al.*, 2020; Whitney and Sharwood, 2008) generated spectinomycin resistant genotypes where those expressing monocot RbcL produced no Rubisco while those expressing the dicot RbcL's produced chimeric L_8S_8 Rubisco comprising endogenous tobacco RbcS (Figure 1B). At least three independent lines for each genotype were generated. For each chimeric Rubisco-producing genotype the flowers of two independent lines were fertilised with wild-type pollen and the seed collected for T_1 progeny study. As shown previously for tobacco expressing RbcL from the monocot grasses *P. bisulcatum* and *P. deustum* (Sharwood *et al.*, 2016), the homoplasmic tobacco lines producing the RbcL from maize, rice or wheat produced no Rubisco (including no *R. rubrum* L_2 Rubisco) and could only be grown in tissue culture (Figure 1C). Attempts to produce flowers and seed for the monocot RbcL coding lines growing in tissue culture were unsuccessful.

The dicot RbcL's show differing chimeric Rubisco biogenesis compatibilities in tobacco.

The folding and assembly compatibility of each heterologous dicot RbcL with the tobacco chloroplast chaperones and endogenous RbcS was examined by comparing the chimeric Rubisco content in the leaves of the T_1 progeny. This included analysing the T_1 or T_2 progeny of previously made tobacco genotypes producing the RbcL from sunflower (Sharwood *et al.*, 2008), Arabidopsis (Whitney *et al.*, 2015), potato (Martin-Avila *et al.*, 2020), *Flaveria bidentis* and *F. pringlei* (Whitney *et al.*, 2015). As shown previously by ^{35}S-Methionine pulse-chase in these lines, the structural stability of each chimeric Rubisco matches native tobacco Rubisco with limitations in RbcL folding and the assembly with tobacco RbcS impacting the amount and production rate of the chimeric L_8S_8 complexes (Martin-Avila *et al.*, 2020; Whitney *et al.*, 2015). Indeed, in the upper canopy leaves from glass house grown plants (37±3 cm in height) there was significant variation in the chimeric Rubisco content between the 12 tobacco genotypes examined (Figure 1D). As demonstrated in Figure 1E, the differing levels of chimeric Rubisco production resulted in corresponding impacts on plant growth rate with, for example, those producing Arabidopsis RbcL taking 27 days longer than wild-type tobacco to reach 37 cm.

Developing the tcc*E. coli* system; a modular cloning tool for plant Rubisco production in *E. coli* expressing tobacco chloroplast chaperones.

Current understanding of the chloroplast protein folding machinery required for plant Rubisco biogenesis has facilitated the expression of tobacco and Arabidopsis Rubiscos in BL21 Star *E. coli* (Aigner *et al.*, 2017; Lin *et al.*, 2020). The machinery comprises 7 ancillary proteins; the CPN60α, CPN60β and CPN20 subunits of the chaperonin CPN60 folding cage, 3 assembly chaperones with known mechanistic roles (RAF1, RbcX, BSD2) and RAF2 whose function remains unclear (Figure 2A). Homologues for each protein in tobacco were identified and their mature N-terminus (*i.e.* without their chloroplast transit peptide) derived from homology mapping to prior studies (Aigner *et al.*, 2017; Whitney *et al.*, 2015) or from available N-terminal sequence information for the chaperone (Conlan *et al.*, 2019) (Figure 2B). Gene blocks for each protein were synthesised using the codon use of tobacco *rbcL* (whose translation properties align with *E. coli*, (Whitney and Andrews, 2001)) and cloned into appropriate Golden Gate compatible Level 0 constructs (Figure S1). Two expression plasmids coding both tobacco Rubisco genes (Level 2 construct pET28-*Nt*LS) or all 7 tobacco chloroplast chaperones (Level 3 construct pCDF-*Nt*Asmbl) were assembled by Golden Gate cloning (Table S2). Each gene in both plasmids was regulated by independent

T7 promoter/terminator elements to facilitate their unified induction by IPTG (Figure 2C). Transformation of pCDF-*Nt*Asmbl into BL21 Star produced the tcc*E. coli* system. Preliminary expression trials in tcc*E. coli* showed abundant L_8S_8 Rubisco expression in cells expressing both RbcL and RbcS (Figure 2D). RbcL-bound CPN60 complexes (CPN60-L) were also apparent in tcc*E. coli* expressing RbcL alone or with RbcS. No RbcL complexes were detected when expressed alone in BL21 Star *E. coli* (*i.e.* without pCDF-*Nt*Asmbl) indicating the RbcL specifically formed stable complexes with the plant Cpn60 complexes and not the endogenous *E. coli* GroEL (Figure 2D, *Nt*Rubisco Ab blot).

Rubisco production in tcc*E. coli* is impacted by temperature and induction duration.

The influence of growth temperature and culturing duration on the efficacy of tobacco Rubisco production in tcc*E. coli* was examined. Cultures incubated at 23°C showed Rubisco production on a cell soluble protein (CSP) basis was sustained over 24 to 48 h (~4.3 ± 0.8% w/w CSP) while at 30°C and 37°C the Rubisco contents peaked at 24 h and 12 h (4.8 ± 1.1% and 1.8 ± 0.3% w/w CSP) respectively (Figure 3A). These findings correlated with the amount of L_8S_8 Rubisco identified following native-PAGE separation of the soluble protein from each culture (Figure 3B). In addition to the CPN60-L complex detected by the *Nt*Rubisco antibody in each sample, unidentified RbcL-containing complex(es) that separated slower than L_8S_8 Rubisco and faster than the CPN60-RbcL complex were produced at all temperatures, being more abundant in those grown at 30°C and 37°C (Figure 3B). Preliminary immunoblot analyses of these complexes following excision and electroelution indicate they comprise RbcL, RAF1 and BSD2 subunits in undetermined stoichiometries (Figure S2).

SDS-PAGE of the total and soluble protein from the same cells analysed by native-PAGE showed the total amount of RbcL and RbcS produced at each timepoint and growth temperature was relatively constant, but varied in the amount detected in the soluble fraction (Figure 3C, D). Even in the 23°C grown cultures only 8-9% of the RbcL and 12-16% of the RbcS produced was soluble (Figure 3D and S3). Noticeably the relative amount of soluble RbcS in the cells grown at 30°C and 37°C were considerably lower than in the 23°C grown cultures, suggesting L_8S_8 biogenesis at these temperatures may be more impacted by soluble RbcS availability or/and their accessibility to 'binding-ready' $RbcL_8$-$(BSD2)_8$ complexes (Figure 2A), consistent with that proposed previously (Wilson *et al.*, 2019).

Taken together these data indicate growing tcc*E.coli* at 23°C and inducing for 24 h afforded the maximal level of L_8S_8 Rubisco production and low amounts of intermediary RbcL-BSD2/RAF1 complexes. Such complexes are not expected to comprise functionally active RbcL catalytic pockets and are thus unlikely to bind the Rubisco inhibitor CABP.

Chimeric Rubisco biogenesis in tcc*E. coli* differs depending on heterologous plant RbcL origin.

The tobacco *rbcL* gene in pET28-*Nt*LS was replaced with those transformed into tobacco coding wheat, maize and rice RbcL and 12 different dicot RbcLs (Table S1). Each pET-LS construct was expressed in tcc*E. coli* and the level of chimeric L_8S_8 Rubisco production visualised by native-PAGE (Figure 4A) and quantified by ^{14}CABP binding (Figure 4B). Both analyses showed cells expressing potato, carrot and strawberry RbcL yielded high, more variable, amounts of chimeric Rubisco that often-exceeded tobacco Rubisco production. Rubisco chimers comprising *Flaveria* RbcL and tobacco RbcS were produced in moderate abundance (~50% of tobacco Rubisco) and the remaining 7 dicot RbcL supported little (olive, Arabidopsis, cassava, Eucalyptus) or no (sweet potato, sunflower, citrus) chimeric Rubisco biogenesis (Figure 4B, Table S2). SDS-PAGE analyses of the total and soluble cell protein fractions confirmed there was little variation in the total pool of dicot RbcL and tobacco RbcS produced in the tcc*E. coli*, only in the soluble fraction that successfully assembled into L_8S_8 holoenzyme (Figure 4C).

The folding and assembly requirements of monocot Rubisco are not met in tcc*E. coli.*

In addition to making the three monocot RbcL expressing pET-LS constructs, the tobacco *rbcS* gene in each was replaced with a synthetic *rbcS* coding a cognate RbcS from wheat (BAB19810.1), rice (NP_001391658.1) or maize (NP_001338725.1) Rubisco (Table S1). Native-PAGE analysis detected no L_8S_8 biogenesis in tcc*E. coli* expressing either wild-type or chimeric monocot Rubiscos (Figure 5A). SDS-PAGE analysis of the total and soluble cell protein fractions indicated each RbcL and RbcS was produced in equivalent abundance, but entirely insoluble (Figure 5B). Taken together these findings indicate that while structural incompatibilities between tobacco RbcS and monocot RbcL may impact chimeric monocot Rubisco biogenesis, the underpinning impediment precluding monocot Rubisco production is that the folding and assembly requirements of the monocot RbcL are not met in tcc*E. coli*, as observed in tobacco chloroplasts (Figure 1C).

Discussion

The expression of higher-plant Rubisco in *E. coli* represents an important milestone in Rubisco bioengineering in terms of opening a new avenue to 'tune up' its catalytic properties and better understand the chaperone compatibility requirements that impact the feasibility and rate of L_8S_8 holoenzyme biogenesis (Conlan and Whitney, 2018; Wilson and Hayer-Hartl, 2018; Wilson *et al.*, 2019). Both these applications depend on the reliability of the *E. coli* expression system to forecast mutated or heterologous Rubisco production and performance in chloroplasts. To assist such efforts, here we have designed a Golden Gate compatible system whose cloning efficiency and modularity provides a first-generation toolkit of genetic parts to facilitate the rapid, combinatorial, testing of transgenes and regulatory elements to optimise plant Rubisco biogenesis in *E. coli* (Table S1). A caveat identified in this study is that while the *E. coli* expression system presents a cost and time effective option for plant Rubisco bioengineering, further optimisation is needed to improve its accuracy to reflect the biogenesis properties of Rubisco in chloroplasts. A similar lack of translational correlation has already been observed for cyanobacteria Rubisco whose biogenesis potential in *E. coli* and tobacco chloroplasts are disproportionate (Wilson *et al.*, 2018).

Limitations in the translational potential of tcc*E. coli* to tobacco chloroplasts.

As summarised in Figure 6, the level of chimeric plant Rubisco expression in both tobacco chloroplasts and tcc*E. coli* appears largely independent of RbcL phylogeny. The 98% RbcL sequence similarity shared by tobacco and potato Rubisco (differing by 11 amino acids, Table S2) unsurprisingly facilitates tobacco-like levels of chimeric potato Rubisco production in both tob^{pL} leaves (Figure 1D, see also (Martin-Avila *et al.*, 2020)) and in the tcc*E. coli* (Figure 4B). Moderately high levels of chimeric carrot and strawberry Rubisco were also produced in both tobacco and tcc*E. coli*, their RbcL differing by 11 amino acids relative to each other, and by 23 amino acids relative to tobacco Rubisco. Somewhat surprisingly however, chimeric olive Rubisco production in both tcc*E. coli* and tobacco chloroplasts were substantially more impaired despite sharing greater homology with tobacco RbcL (differing by 18 amino acids). Differences in chimeric Rubisco production between tcc*E. coli* and tobacco leaves were most apparent for enzymes comprising sweet potato, sunflower and citrus RbcL whose near-fully impaired assembly in tcc*E. coli* contrasted with production levels in tobacco chloroplasts that were sufficient to support glasshouse plant growth (Figure 1E).

The need to better understand the chaperone requirements of plant Rubisco.

Understanding what factors underpin variation in the Rubisco biogenesis properties between tcc*E. coli* and tobacco chloroplasts is critical to the successful use of *E. coli* as a cheap, high throughput, Rubisco testing screen with reliant translational potential in plants. Prior studies have demonstrated differences in the structural compatibility requirements between RbcL and its folding chaperones. For example, RAF1 and RbcL in plants have co-evolved to share structural complementation requirements that significantly impact Rubisco biogenesis in chloroplasts (Whitney *et al.*, 2015). In contrast, the $RbcL_8$ binding and stabilising function of BSD2 appears more promiscuous in its structural compatibility with RbcL (Conlan *et al.*, 2019) while the structural specificity of RbcX on plant Rubisco biogenesis is less well understood. For plant Rubisco production in *E. coli* there is no absolute requirement for RbcX, although it does enhance the efficiency of Rubisco biogenesis (Aigner *et al.*, 2017). Clearly a systematic, combinatorial heterologous gene cloning approach is required to fully understand the structural compatibility requirements between RbcL and each of its folding and assembly chaperones. For example, there is evidence the efficiency of tobacco L_8S_8 biogenesis in *E. coli* is enhanced using the RAF2 and CPN60 from Arabidopsis (Lin *et al.*, 2020). Untested in this study was whether differences in the levels of the tobacco and Arabidopsis chaperone isoforms produced in *E. coli* influenced Rubisco biogenesis yields, an aspect of experimental rigor for consideration in follow up comparative studies examining Rubisco chaperone compatibility.

A consistent finding of plant RbcL and RbcS expression studies in *E. coli* is they are produced in abundance but mostly form mis-folded, insoluble protein (Figures 3C, 3D, 4C, (Aigner *et al.*, 2017; Wilson *et al.*, 2019)). This suggests Rubisco biogenesis in tcc*E. coli* remains impaired by post-translational limitations associated with impediments to the accessibility or compatibility of plant RbcL peptides to the tobacco CPN60 protein-folding cylinders or, as discussed above, one or more of the assembly chaperones. The limited capacity of the *E. coli* GroEL chaperonin to bind plant RbcL, and inability of GroEl to form heterologous complexes with the chloroplast CPN60 subunits (Figure 2D, (Aigner *et al.*, 2017)), suggests any impediment to RbcL folding is not associated with the formation of chimeric *E. coli*-chloroplast chaperonin folding cylinders. It is also feasible that one or more of the 7 chloroplast auxiliary proteins are produced in adverse stoichiometries ill-suited to optimal Rubisco holoenzyme production in tcc*E. coli*. This appears particularly problematic in faster growing cells (*i.e.* when cultured at 30 or 37°C) where the production of stable

377 BSD2 and or RAF1 bound RbcL-chaperone intermediary complexes (designated $L_8(B2/R1)_n$
378 in Figures 3B and S2) appear to exceed L_8S_8-Rubisco levels more than 5-fold (Figure 3B).
379 Keeping in mind what is currently known about the plant Rubisco biogenesis pathways
380 (Figure 2A), and our findings these complexes comprise BSD2 and RAF1 (Supplemental
381 Figure S2), it would appear either BSD2 and/or soluble RbcS availability is limiting L_8S_8
382 biogenesis in tcc*E. coli* grown at the elevated temperatures. The pervasive role of RbcS
383 availability on plant Rubisco production in *E. coli* has been proposed previously (Wilson *et*
384 *al.*, 2019) and supported by our observations of diminishing RbcS solubility in tcc*E. coli* with
385 increasing growth temperature (Figure 3D).

386 Clearly there is a need to more closely examine the impact that growth temperature
387 (e.g. below 23°C) and alternating the stoichiometric expression of each tobacco chaperone
388 has on both tobacco and heterologous plant Rubisco production in tcc*E. coli*. For example,
389 what toll the excessive levels of CPN20 produced in tcc*E. coli* (Figure 3C and Supplemental
390 Figure S2) might have on cell viability and/or Rubisco biogenesis remains unknown. Current
391 models of chaperonin protein folding cage function show the co-chaperonin lid comprises
392 heptamers of GroES subunits, or in chloroplasts equivalent sized hetero-oligomer lid
393 complexes comprising CPN20 alone (forming an asymmetrical octameric cap) or with
394 CPN10 subunits in a heptameric configuration (Tsai *et al.*, 2012; Zhao and Liu, 2017). The
395 use of the endogenous GroES, in place of chloroplast CPN10, with CPN20 however
396 enhanced plant Rubisco biogenesis in *E. coli* suggesting the possible formation of functional
397 CPN20-GroES cap complexes suited to facilitating RbcL folding by CPN60 chaperonin
398 cylinders (Aigner *et al.*, 2017). However, the structurally stable oligomeric CPN20 assembly
399 produced in tcc*E. coli* does not contain any detectable *E. coli* GroES subunits (Supplemental
400 Figure S2). This poses questions as to subunit composition of the CPN20-oligomeric
401 complexes produced in tcc*E. coli* and whether there are corresponding perturbations in its
402 function that impair the RbcL folding capacity by the plant CPN60 cages.

403 A critical question is whether the utility of tcc*E. coli* for plant Rubisco production can
404 also be improved by co-expressing additional chloroplast chaperone components. For
405 example, expressing components of the stromal HSP70 system involved in RbcS folding (Su
406 and Li, 2010) might help bolster L_8S_8 production. Likewise, co-expressing one or more of the
407 post-translational modifying (PTM) acetylation, methylation, and N-terminal proteolytic
408 RbcL processing enzymes (Houtz *et al.*, 2008) may benefit the assembly, function and
409 stability of Rubisco production in *E. coli*. For example, truncating the RbcL N-terminus by 2

amino acids to mimic the natural 2-amino acid processing in chloroplasts appears to be kinetically important (Ng *et al.*, 2020), a might explain the lower CO_2-fixation rates (k_{cat}^c, s^{-1}) reported for Rubisco produced in *E. coli* relative to that made in chloroplasts (Lin *et al.*, 2022; Lin *et al.*, 2020). Also in need of closer examination is the extent to which chaperone bound $RbcL_8$-intermediary complexes might actually bind CABP and thus impair accurate measurement of k_{cat}^c for *E. coli*-made Rubisco.

What is impeding monocot Rubisco biogenesis by chaperones from dicot chloroplasts?

The inability of tobacco chloroplasts and the tcc*E. coli* to meet the folding and assembly requirements of rice, wheat and maize Rubisco remains enigmatic. As shown previously, fully silencing Rubisco activity in tobacco through *rbcL* deletion (Kanevski and Maliga, 1994), mutagenic silencing of catalytic function (Whitney and Sharwood, 2008), or replacement with monocot *rbcL* genes (Sharwood *et al.*, 2016) produces chlorotic plant progeny that can only be propagated in tissue culture or by grafting onto wildtype scion tissue. Building on this we show the inability to produce chimeric monocot Rubisco in tobacco (Figure 1) or homogeneous monocot Rubisco in tcc*E. coli* expressing cognate RbcL and RbcS (Figure 5) arises from structural incompatibility between monocot RbcL and one or more components of the tobacco chloroplast folding machinery and not the tobacco RbcS. Producing monocot Rubisco in *E. coli* will therefore likely require the production of pCDF*Nt*Asmbl plasmid derivatives (Figure 2C) using Golden Gate cloning to allow for the easy exchange of genetic parts coding equivalent Rubisco chaperones from rice, maize or wheat chloroplasts. Such an advance would be of particular importance given the current tools for Rubisco engineering in these global staple grain crops remain limited to modulating Rubisco production through nuclear transgenesis (Matsumura *et al.*, 2020; Salesse-Smith *et al.*, 2018; Yoon *et al.*, 2020) as stably transforming their chloroplast genomes remains untenable (Bock, 2015).

Utilising tcc*E. coli* to evolve a 'better' plant Rubisco?

There is considerable apprehension as to the feasibility of improving Rubisco catalysis. Enzyme kinetic surveys have revealed some degree of natural catalytic diversity among terrestrial plant Rubiscos (Iñiguez *et al.*, 2021; Orr *et al.*, 2016) though likely not at a scale to have a discernible impact on crop yield (Wu *et al.*, 2019). By contrast, the kinetics of Rubisco from some red algae have the potential to improve crop photosynthesis and productivity by up to 30% (Zhu *et al.*, 2004), a discovery supporting the assertion that improving the

carboxylation properties of plant Rubisco is not an immutable challenge. Directed evolution studies in *E. coli* have since confirmed the feasibility of improving the carboxylation rate, efficiency and specificity of Rubisco, though to date only with non-plant enzyme isoforms (Wilson *et al.*, 2016; Wilson *et al.*, 2018; Zhou and Whitney, 2019). The *E. coli* screen used in these directed evolution studies involve the expression of phosphoribulokinase (PRK) to produce RuBP whose toxicity to *E. coli* growth can be alleviated by Rubisco catalysis. This allows for a life-death Rubisco Dependent *E. coli* (RDE) screen where higher activity Rubisco mutants can be selected as those that support cell growth under increased PRK expression.

The tcc*E. coli*-plant Rubisco two plasmid expression system developed in this study provides the genetic resource needed to develop an RDE screen for the directed evolution of plant Rubisco. As previously proposed (Conlan and Whitney, 2018; Gionfriddo *et al.*, 2019), establishing the RDE screen would require the co-transformation of a 3rd pACYC-PRKNPTII plasmid, where the expression of the PRK-NPTII fusion protein is regulated by an arabinose inducible promoter (Wilson *et al.*, 2018) and thus independent of the IPTG induced expression of all T7 promoter regulated 7 chaperone genes in pCDF*Nt*Asmbl and both Rubisco genes in pET28-*rbc*LS (Figure 2C). Of advantage, such a tcc*E. coli* based RDE screen incorporates the endogenous tobacco chloroplast chaperones and thus engenders hope that the biogenesis properties of any tobacco Rubisco catalytic mutants selected would also be met in tobacco chloroplasts. Such a consideration is critical to facilitating the translational testing of how a 'better' Rubisco impacts plant photosynthesis and growth.

Supplementary Data

Table S1. The tobacco Rubisco and chaperone Golden Gate compatible cloning kit and acceptor vectors for cloning and expression.

Table S2. Primer list.

Supplementary Figure S1. An overview of the modular Golden Gate cloning process used to generate the tcc*E. coli* plant Rubisco expression system.

Supplementary Figure S2. Preliminary compositional analysis of 3 protein complexes produced in 30°C grown tcc*E. coli* producing tobacco Rubisco.

Supplementary Figure S3. Densitometry analysis of relative soluble tobacco RbcL and RbcS produced in pET-*Nt*LS-tcc*E. Coli*.

Author contributions

S.B, T.R and S.M.W conceived the project; D.Y, M.K. and R.B generated the transgenic tobacco lines; S.B, T.R, and M.G developed the *E. coli* expression system and with T.S and S.M.W performed the molecular and protein analyses. All authors contributed to writing the manuscript.

Conflict of interest

The authors declare no conflict of interest

Funding

This work was supported by the Australian Government through the Australian Research Council Centre of Excellence for Translational Photosynthesis CE140100015 and the plant transformation work funded by the Bill and Melinda Gates Foundation grant OPP1060461 titled “RIPE - Realizing increased photosynthetic efficiency for sustainable increases in crop yield”.

Data availability

All data supporting the findings of this study are available within the paper and within its supplementary materials published online.

References

Aigner H, Wilson RH, Bracher A, Calisse L, Bhat JY, Hartl FU, Hayer-Hartl M. 2017. Plant Rubisco assembly in *E. coli* with five chloroplast chaperones including BSD2. Science **358**, 1272-1278.

Bock R. 2015. Engineering Plastid Genomes: Methods, Tools, and Applications in Basic Research and Biotechnology. Ann Rev Plant Biol **66**, 211-241.

Bouvier JW, Emms DM, Rhodes T, Bolton JS, Brasnett A, Eddershaw A, Nielsen JR, Unitt A, Whitney SM, Kelly S. 2021. Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off. Molecular Biology and Evolution.

Bracher A, Starling-Windhof A, Hartl FU, Hayer-Hartl M. 2011. Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. Nature structural & molecular biology **18**, 875-880.

Bracher A, Whitney SM, Hartl FU, Hayer-Hartl M. 2017. Biogenesis and metabolic maintenance of Rubisco. Ann Rev Plant Biol **68**, 29-60.

Brutnell TP, Sawers RJH, Mant A, Langdale JA. 1999. BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the rbcL gene of maize. The Plant Cell **11**, 849-864.

Busch FA. 2020. Photorespiration in the context of Rubisco biochemistry, CO2 diffusion and metabolism. Plant Journal **101**, 919-939.

Conlan B, Birch R, Kelso C, Holland S, De Souza AP, Long SP, Beck JL, Whitney SM. 2019. BSD2 is a Rubisco-specific assembly chaperone, forms intermediary hetero-oligomeric complexes, and is nonlimiting to growth in tobacco. Plant, Cell & Environment **42**, 1287-1301.

Conlan B, Whitney S. 2018. Preparing Rubisco for a tune up. Nature Plants **4**, 12-13.

Cummins PL. 2021. The coevolution of RuBisCO, photorespiration, and carbon concentrating mechanisms in higher plants. Frontiers in Plant Science **12**.

Donovan S, Mao Y, Orr DJ, Carmo-Silva E, McCormick AJ. 2020. CRISPR-Cas9-mediated mutagenesis of the Rubisco small subunit family in *Nicotiana tabacum*. Frontiers in Genome Editing **2**.

Evans JR, Lawson T. 2020. From green to gold: agricultural revolution for food security. Journal of Experimental Botany **71**, 2211-2215.

Feiz L, Williams-Carrier R, Belcher S, Montano M, Barkan A, Stern DB. 2014. A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. Plant J. **80**, 862-869.

Feiz L, Williams-Carrier R, Wostrikoff K, Belcher S, Barkan A, Stern DB. 2012. Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. Plant Cell **24**, 3435-3446.

Gionfriddo M, De Gara L, Loreto F. 2019. Directed Evolution of Plant Processes: Towards a green (r)evolution? Trends in Plant Science **24**, 999-1007.

Gunn LH, Martin Avila E, Birch R, Whitney SM. 2020. The dependency of red Rubisco on its cognate activase for enhancing plant photosynthesis and growth. Proc Nat Acad Sci **117**, 25890-25896.

Hauser T, Bhat JY, Milicic G, Wendler P, Hartl FU, Bracher A, Hayer-Hartl M. 2015. Structure and mechanism of the Rubisco-assembly chaperone Raf1. Nat. Struct. Mol. Biol. **22**, 720-728.

Houtz RL, Magnani R, Nayak NR, Dirk LMA. 2008. Co- and post-translational modifications in Rubisco: unanswered questions. J Exp Bot **59**, 1635-1645.

Iñiguez C, Aguiló-Nicolau P, Galmés J. 2021. Improving photosynthesis through the enhancement of Rubisco carboxylation capacity. Biochemical Society Transactions.

Kanevski I, Maliga P. 1994. Relocation of the plastid *rbcL* gene to the nucleus yields functional ribulose-1,5-bisphosphate carboxylase in tobacco chloroplasts. Proc Nat Acad Sci **91**, 1969-1973.
Kuroda H, Maliga P. 2001. Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. Plant Physiology **125**, 430-436.
Lin MT, Hanson MR. 2018. Red algal Rubisco fails to accumulate in transplastomic tobacco expressing Griffithsia monilis RbcL and RbcS genes. Plant Direct **2**, e00045.
Lin MT, Salihovic H, Clark FK, Hanson MR. 2022. Improving the efficiency of Rubisco by resurrecting its ancestors in the family Solanaceae. ScienceAdvances **8**, abm6871.
Lin MT, Stone WD, Chaudhari V, Hanson MR. 2020. Small subunits can determine enzyme kinetics of tobacco Rubisco expressed in Escherichia coli. Nature Plants **6**, 1289-1299.
Liu C, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, Rao KV, Berninghausen O, Mielke T, Hartl FU, Beckmann R, Hayer-Hartl M. 2010. Coupled chaperone action in folding and assembly of hexadecameric Rubisco. Nature **463**, 197-202.
Martin-Avila E, Lim Y-L, Birch R, Dirk LMA, Buck S, Rhodes T, Sharwood RE, Kapralov MV, Whitney SM. 2020. Modifying Plant Photosynthesis and Growth via Simultaneous Chloroplast Transformation of Rubisco Large and Small Subunits. The Plant Cell **32**, 2898-2916.
Matsumura H, Shiomi K, Yamamoto A, Taketani Y, Kobayashi N, Yoshizawa T, Tanaka S-i, Yoshikawa H, Endo M, Fukayama H. 2020. Hybrid Rubisco with Complete Replacement of Rice Rubisco Small Subunits by Sorghum Counterparts Confers C4 Plant-like High Catalytic Activity. Molecular Plant.
Moore SJ, Lai H-E, Kelwick RJR, Chee SM, Bell DJ, Polizzi KM, Freemont PS. 2016. EcoFlex: A multifunctional MoClo kit for *E. coli* synthetic biology. ACS Synthetic Biology **5**, 1059-1069.
Mueller-Cajar O, Whitney SM. 2008. Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. Photosynth Res **98**, 667-675.
Ng J, Guo Z, Mueller-Cajar O. 2020. Rubisco activase requires residues in the large subunit N terminus to remodel inhibited plant Rubisco. J Biol Chem **295**, 16427-16435.
Orr DJ, Alcântara A, Kapralov MV, Andralojc PJ, Carmo-Silva E, Parry MAJ. 2016. Surveying Rubisco diversity and temperature response to improve crop photosynthetic efficiency. Plant Physiol **172**, 707-717.
Salesse-Smith CE, Sharwood RE, Busch FA, Kromdijk J, Bardal V, Stern DB. 2018. Overexpression of Rubisco subunits with RAF1 increases Rubisco content in maize. Nature Plants **4**, 802-810.
Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU, Hayer-Hartl M. 2007. Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. Cell **129**, 1189-1200.
Sharwood R, von Caemmerer S, Maliga P, Whitney S. 2008. The catalytic properties of hybrid Rubisco comprising tobacco small and sunflower large subunits mirror the kinetically equivalent source Rubiscos and can support tobacco growth. Plant Physiol **146**, 83-96.
Sharwood RE. 2017. Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. New Phytologist **213**, 494-510.
Sharwood RE, Ghannoum O, Kapralov MV, Gunn LH, Whitney SM. 2016. Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C_3 photosynthesis. Nature Plants **2**, 16186.

Su PH, Li HM. 2010. Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts. Plant Cell **22**, 1516-1531.
Tsai Y-CC, Mueller-Cajar O, Saschenbrecker S, Hartl FU, Hayer-Hartl M. 2012. Chaperonin cofactors, Cpn10 and Cpn20, of green algae and plants function as hetero-oligomeric ring complexes. Journal of Biological Chemistry **287**, 20471-20481.
Vitlin Gruber A, Feiz L. 2018. Rubisco Assembly in the Chloroplast. Frontiers in Molecular Biosciences **5**.
von Caemmerer S. 2020. Rubisco carboxylase/oxygenase: From the enzyme to the globe: A gas exchange perspective. Journal of Plant Physiology **252**, 153240.
Whitney SM, Andrews TJ. 2001. Plastome-encoded bacterial Rubisco supports photosynthesis and growth in tobacco. Proc Natl Acad Sci **98**, 14738-14743.
Whitney SM, Baldet P, Hudson GS, Andrews TJ. 2001. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. Plant J **26**, 535-547.
Whitney SM, Birch R, Kelso C, Beck JL, Kapralov MV. 2015. Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. Proc Nat Acad Sci **112**, 3564-3569.
Whitney SM, Houtz RL, Alonso H. 2011a. Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. Plant Physiol **155**, 27-35.
Whitney SM, Sharwood RE. 2007. Linked Rubisco subunits can assemble into functional oligomers without impeding catalytic performance. J Biol Chem **282**, 3809-3818.
Whitney SM, Sharwood RE. 2008. Construction of a tobacco master line to improve Rubisco engineering in chloroplasts. J Exp Bot **59**, 1909-1921.
Whitney SM, Sharwood RE. 2021. Rubisco Engineering by Plastid Transformation and Protocols for Assessing Expression. Methods Mol Biol **2317**, 195-214.
Whitney SM, Sharwood RE, Orr D, White SJ, Alonso H, Galmés J. 2011b. Isoleucine 309 acts as a C_4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) carboxylation rate in *Flaveria*. Proc Nat Acad Sci **108**, 14688-14693.
Wilson RH, Alonso H, Whitney SM. 2016. Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. Sceintific Reports **6**, 22284.
Wilson RH, Hayer-Hartl M. 2018. Complex chaperone dependence of Rubisco biogenesis. Biochemistry **57**, 3210-3216.
Wilson RH, Martin-Avila E, Conlan C, Whitney SM. 2018. An improved *Escherichia coli* screen for Rubisco identifies a protein–protein interface that can enhance CO_2-fixation kinetics. J Biol Chem **293**, 18-27.
Wilson RH, Thieulin-Pardo G, Hartl F-U, Hayer-Hartl M. 2019. Improved recombinant expression and purification of functional plant Rubisco. FEBS Letters **593**, 611-621.
Wu A, Hammer GL, Doherty A, von Caemmerer S, Farquhar GD. 2019. Quantifying impacts of enhancing photosynthesis on crop yield. Nature plants **5**, 380-388.
Yoon D-K, Ishiyama K, Suganami M, Tazoe Y, Watanabe M, Imaruoka S, Ogura M, Ishida H, Suzuki Y, Obara M, Mae T, Makino A. 2020. Transgenic rice overproducing Rubisco exhibits increased yields with improved nitrogen-use efficiency in an experimental paddy field. Nature Food **1**, 134-139.
Zhang XH, Webb J, Huang YH, Lin L, Tang RS, Liu A. 2011. Hybrid Rubisco of tomato large subunits and tobacco small subunits is functional in tobacco plants. Plant Sci **180**, 480-488.
Zhao Q, Liu C. 2017. Chloroplast Chaperonin: An Intricate Protein Folding Machine for Photosynthesis. Front Mol Biosci **4**, 98.

Zhou Y, Whitney S. 2019. Directed evolution of an improved Rubisco; *In vitro* analyses to decipher fact from fiction. Int J Mol Sci **20**, 5019.
Zhu XG, Portis AR, Long SP. 2004. Would transformation of C_3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. Plant Cell Env **27**, 155-165.

Figures

Figure 1. Plastome transformation of heterologous plant *rbcL* genes into tobacco differentially impact chimeric Rubisco production and plant growth.

(**A**) Plastome pLEV4 derived transforming plasmids (Whitney *et al.*, 2015) containing synthetic *rbcL* genes coding the RbcL from 3 monocot and 7 dicot species (GenBank accessions shown) were transformed into the *R. rubrum* L_2-Rubisco producing master-line tob*Rr* (Whitney and Sharwood, 2008). Native-PAGE screening of spectinomycin resistant (coded by the *aadA* gene) plantlets identified homoplasmic, non-L_2 Rubisco producing, lines producing wither (**B**) chimeric L_8S_8 complexes (comprising foreign (orange) RbcL and tobacco RbcS) or no Rubisco (ΔRubisco). (**C**) Representative pale green phenotype of the ΔRubisco tob^{wheat}, tob^{maize} and tob^{rice} T_0 plants that could only be grown in tissue culture on sucrose. (**D**) Significant variation in chimeric Rubisco production among the T_1 tobacco genotypes quantified by ^{14}CABP binding and visualised by native-PAGE L_8S_8 Coomassie staining correlated with (**E**) differences in plant growth under ambient CO_2 25 days after transfer to soil from propagating in tissue culture (noting. Tobacco genotypes analysed in [1] (Martin-Avila *et al.*, 2020), [2](Whitney *et al.*, 2011a), [3] (Sharwood *et al.*, 2008) and [4] (Whitney *et al.*, 2015). Lower case letters indicate significant differences to $p<0.05$ using a Tukey multiple comparison test.

Figure 2. Building a multiplexed tobacco chloroplast chaperone (tcc)*E. coli* cloning system for plant Rubisco expression.

(**A**) Replicating the complex, multi-step biogenesis requirements of tobacco L_8S_8 Rubisco in *E. coli* requires the co-expression of RbcL, the 7 accessory proteins required to fold RbcL (CPN60α/β/20) and assemble them via dimer ($RbcL_2$RAF1/ $RbcL_2RbcX_2$) and octameric ($RbcL_8BSD2_8$) intermediary complexes to which RbcS binds to form a functional, stable L_8S_8 holoenzyme. (**B**) GenBank accessions and first 5 amino acids (AA, transit peptide sequence excluded) of the cloned tobacco chaperones. (**C**) The Golden Gate compatible tcc*E. coli* cloning system comprises plasmid pCDF-*Nt*Asmbl coding 7 tobacco chloroplast chaperones and the Rubisco expression plasmid pET-*Nt*LS. Each gene is regulated by independent, IPTG inducible, T7 promoter and 3'UTR (terminator) elements. RBS, ribosome binding sites. See supplementary Table S1 and Figure S1 for Golden Gate cloning and sequence detail. (**D**) Replica native-PAGE of 8 µg soluble protein from differentially transformed BL21 Star *E. coli*. The separated proteins were Coomassie stained or blotted onto nitrocellulose and probed

with antibodies to tobacco Rubisco (*Nt*Rubisco) or Cpn60α (*Nt*CPN60α). RbcL formed stable complexes with tobacco chaperonin (CPN60-L, lanes 4 and 6) but not the *E. coli* GroELS (lane 3) with tobacco Rubisco (L_8S_8) abundantly produced in cells expressing all 9 recombinant proteins (lane 6). CPN60, smaller (faster migrating) tobacco chaperonin complexes without RbcL bound (lanes 2 and 5).

Figure 3. Impact of temperature on tobacco Rubisco production in tcc*E. coli*

Cultures of three independent pET-*Nt*LS-tcc*E. coli* colonies were induced with 1mM IPTG at an OD_{600} of 0.5 and grown at 23, 30 or 37°C for 12h, 24h or 48h. (**A**) Rubisco content quantified by ^{14}CABP-binding (n = 4 ± SD, expressed as a percentage of cell soluble protein, CSP). Lower case letters indicate significant differences to $p<0.05$ using a Tukey multiple comparison test. (**B**) L_8S_8 Rubisco contents were confirmed by Coomassie staining and *Nt*Rubisco immuo-blotting of 6 µg soluble protein separated by native-PAGE. RbcL was detected associated with tobacco chaperonin (CPN60-L) and in $L_8(R1/B2)_n$ complexes comprising uncharacterised stoichiometries of RAF1 and BSD2 (see Figure S2). The relative amounts of total and soluble RbcL and RbcS produced in each culture was ascertained by (**C**) Coomassie staining and (**D**) *Nt*Rubisco immuo-blotting following SDS-PAGE separation of the soluble (S) and total (T) cellular protein. The percentage of soluble RbcL and RbcS (shown in brackets) was determined by immuno-blot densitometry (Figure S3).

Figure 4. Differential production of plant Rubisco chimers in tcc*E. coli*

Cultures of independent tcc*E. coli* colonies transformed with differing pET-LS constructs coding tobacco RbcS and a heterologous RbcL were induced with 1mM IPTG at an OD_{600} of 0.5 and grown at 23°C for 24h. Their soluble protein (8 µg) was separated by native-PAGE and L_8S_8 Rubisco production visualised by (**A**) Coomassie staining and (**B**) quantified as a percentage of cell soluble protein (CSP) by ^{14}CABP-binding (n = 3 to 12 ± SD, as indicated). Lower case letters indicate significant differences to $p<0.05$ using a Tukey multiple comparison test. (**C**) The relative level of soluble to total amount of RbcL and RbcS produced was ascertained by coomassie staining following SDS-PAGE separation of the soluble (S) and total (T) cellular protein.

Figure 5. Monocot Rubisco subunits are produced but their assembly not met in tcc*E. coli*

Cultures of tcc*E. coli* transformed with pET-LS constructs coding the RbcL from maize (*Zm*), wheat (*Ta*) or rice (*Os*) with either the tobacco RbcS (tob-RbcS) or their cognate RbcS

(*Zm*RbcS, *Ta*RbcS or *Os*RbcS) were induced with 1mM IPTG at an OD_{600} of 0.5 and grown at 23°C for 24h. (**A**) L_8S_8 Rubisco production in the soluble protein separated by native-PAGE was only detected in the tobacco Rubisco pET-*Nt*LS control. (**B**) Commassie staining and (**C**) *Nt*Rubisco immunoblotting following SDS-PAGE of the soluble (S) and total (T) cellular protein indicated the monocot RbcL and RbcS were produced, but insoluble. While the *Nt*Rubisco antisera recognises tob-RbcS, it does not recognise *Zm*RbcS, *Ta*RbcS or *Os*RbcS.

Figure 6. Correlative Rubisco biogenesis potential in tobacco chloroplasts and tcc*E. coli*

Mapping the comparative biogenesis potential of chimeric Rubiscos in tobacco chloroplasts and tcc*E. coli* against their RbcL phylogeny. The expression levels of each Rubisco were normalised relative to the tobacco Rubisco content in chloroplasts (Figure 1D) and tcc*E. coli* (Figure 4B). The RbcL phylogenetic tree was generated in Geneious using sequence identity and a global alignment, using the tree build model UPGMA and the genetic distance model Jukes-Candor. RbcL Genbank accessions were NP_051067.1 (Arabidopsis), SCM15159.1 (*P. deustum*), SCM15147.1 (*P. bisulcatum*), YP_588125.1 (*H. annus*), P19161.1 (*F. bidentis*), P19162.1 (*F. pringlei*) and those listed in Figure 1A.

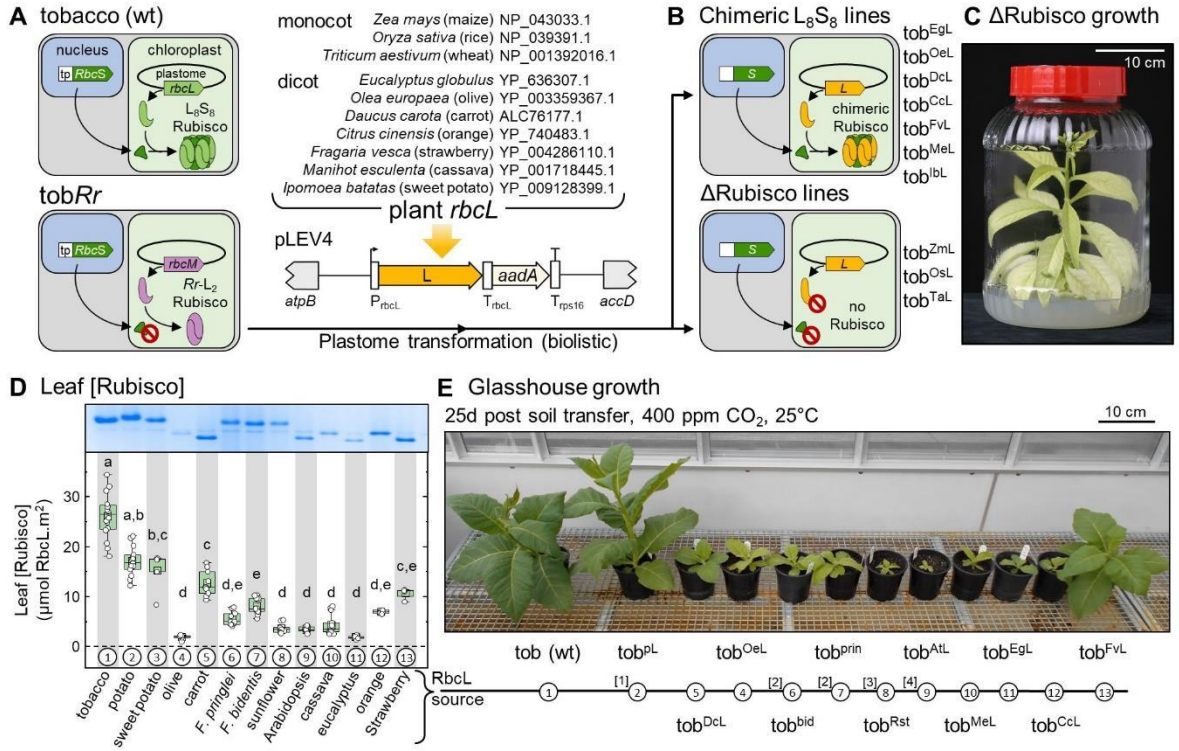

Figure 1. Plastome transformation of heterologous plant *rbcL* genes into tobacco differentially impact chimeric Rubisco production and plant growth.

(**A**) Plastome pLEV4 derived transforming plasmids (Whitney *et al.*, 2015) containing synthetic *rbcL* genes coding the RbcL from 3 monocot and 7 dicot species (GenBank accessions shown) were transformed into the *R. rubrum* L_2-Rubisco producing master-line tob*Rr* (Whitney and Sharwood, 2008). Native-PAGE screening of spectinomycin resistant (coded by the *aadA* gene) plantlets identified homoplasmic, non-L_2 Rubisco producing, lines producing wither (**B**) chimeric L_8S_8 complexes (comprising foreign (orange) RbcL and tobacco RbcS) or no Rubisco (ΔRubisco). (**C**) Representative pale green phenotype of the ΔRubisco tob^{wheat}, tob^{maize} and tob^{rice} T_0 plants that could only be grown in tissue culture on sucrose. (**D**) Significant variation in chimeric Rubisco production among the T_1 tobacco genotypes quantified by ^{14}CABP binding and visualised by native-PAGE L_8S_8 Coomassie staining correlated with (**E**) differences in plant growth under ambient CO_2 25 days after transfer to soil from propagating in tissue culture (noting. Tobacco genotypes analysed in [1] (Martin-Avila *et al.*, 2020), [2](Whitney *et al.*, 2011), [3] (Sharwood *et al.*, 2008) and [4] (Whitney *et al.*, 2015). Lower case letters indicate significant differences to $p<0.05$ using a Tukey multiple comparison test.

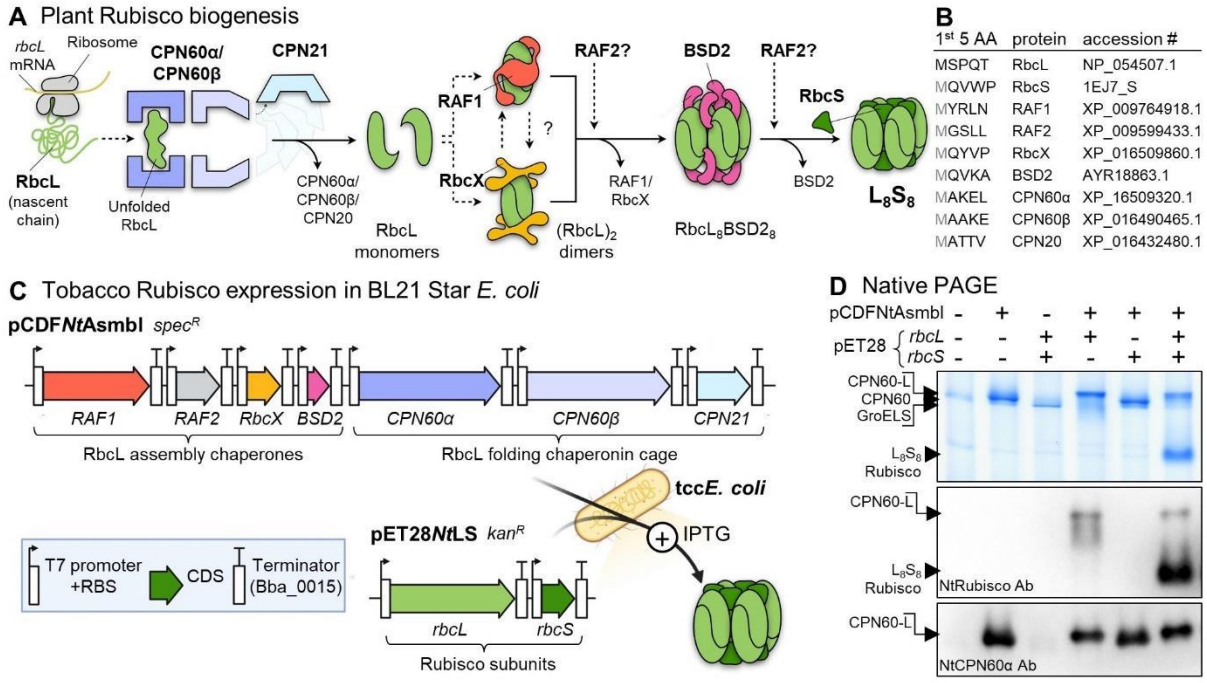

Figure 2. Building a multiplexed tobacco chloroplast chaperone (tcc)*E. coli* cloning system for plant Rubisco expression.

(**A**) Replicating the complex, multi-step biogenesis requirements of tobacco L_8S_8 Rubisco in *E. coli* requires the co-expression of RbcL, the 7 accessory proteins required to fold RbcL (CPN60α/β/20) and assemble them via dimer ($RbcL_2RAF1$/ $RbcL_2RbcX_2$) and octameric ($RbcL_8BSD2_8$) intermediary complexes to which RbcS binds to form a functional, stable L_8S_8 holoenzyme. (**B**) GenBank accessions and first 5 amino acids (AA, transit peptide sequence excluded) of the cloned tobacco chaperones. (**C**) The Golden Gate compatible tcc*E. coli* cloning system comprises plasmid pCDF-*Nt*Asmbl coding 7 tobacco chloroplast chaperones and the Rubisco expression plasmid pET-*Nt*LS. Each gene is regulated by independent, IPTG inducible, T7 promoter and 3'UTR (terminator) elements. RBS, ribosome binding sites. See supplementary Table S1 and Figure S1 for Golden Gate cloning and sequence detail. (**D**) Replica native-PAGE of 8 µg soluble protein from differentially transformed BL21 Star *E. coli*. The separated proteins were Coomassie stained or blotted onto nitrocellulose and probed with antibodies to tobacco Rubisco (*Nt*Rubisco) or Cpn60α (*Nt*CPN60α). RbcL formed stable complexes with tobacco chaperonin (CPN60-L, lanes 4 and 6) but not the *E. coli* GroELS (lane 3) with tobacco Rubisco (L_8S_8) abundantly produced in cells expressing all 9 recombinant proteins (lane 6). CPN60, smaller (faster migrating) tobacco chaperonin complexes without RbcL bound (lanes 2 and 5).

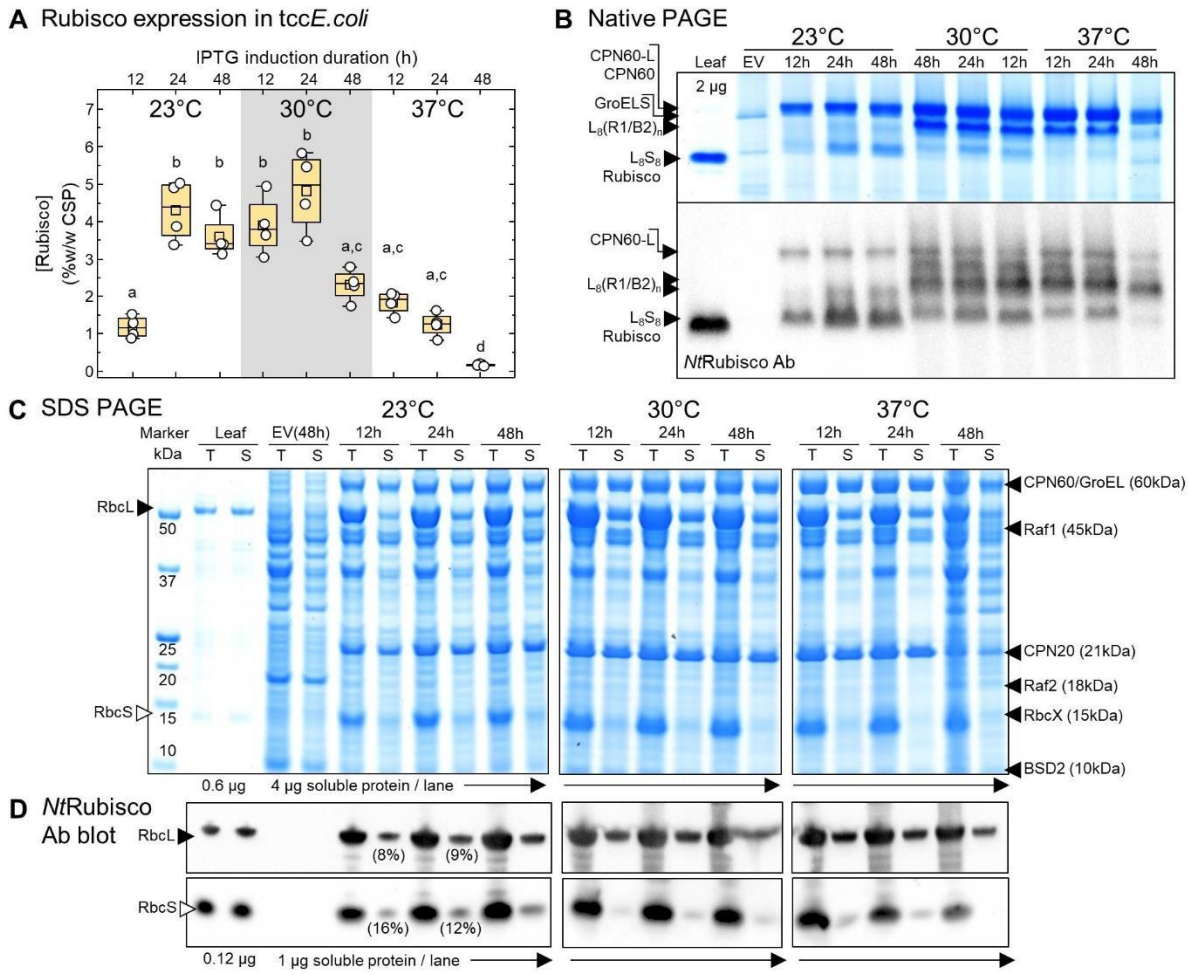

Figure 3. Impact of temperature on tobacco Rubisco production in tcc*E. coli*

Cultures of three independent pET-*Nt*LS-tcc*E. coli* colonies were induced with 1mM IPTG at an OD_{600} of 0.5 and grown at 23, 30 or 37°C for 12h, 24h or 48h. (**A**) Rubisco content quantified by ^{14}CABP-binding (n = 4 ± SD, expressed as a percentage of cell soluble protein, CSP). Lower case letters indicate significant differences to $p<0.05$ using a Tukey multiple comparison test. (**B**) L_8S_8 Rubisco contents were confirmed by Coomassie staining and *Nt*Rubisco immuo-blotting of 6 µg soluble protein separated by native-PAGE. RbcL was detected associated with tobacco chaperonin (CPN60-L) and in $L_8(R1/B2)_n$ complexes comprising uncharacterised stoichiometries of RAF1 and BSD2 (see Figure S2). The relative amounts of total and soluble RbcL and RbcS produced in each culture was ascertained by (**C**) Coomassie staining and (**D**) *Nt*Rubisco immuo-blotting following SDS-PAGE separation of the soluble (S) and total (T) cellular protein. The percentage of soluble RbcL and RbcS (shown in brackets) was determined by immuno-blot densitometry (Figure S3).

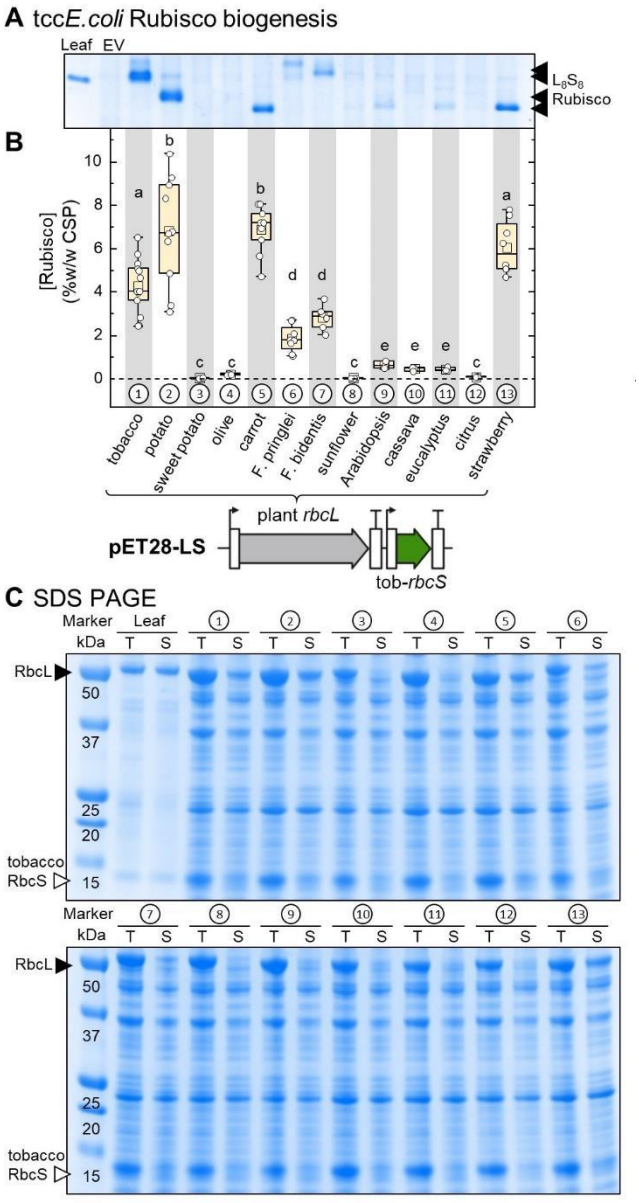

Figure 4. Differential production of plant Rubisco chimers in tcc*E. coli*

Cultures of independent tcc*E. coli* colonies transformed with differing pET-LS constructs coding tobacco RbcS and a heterologous RbcL were induced with 1mM IPTG at an OD_{600} of 0.5 and grown at 23°C for 24h. Their soluble protein (8 µg) was separated by native-PAGE and L_8S_8 Rubisco production visualised by (**A**) Coomassie staining and (**B**) quantified as a percentage of cell soluble protein (CSP) by ^{14}CABP-binding (n = 3 to 12 ± SD, as indicated). Lower case letters indicate significant differences to $p<0.05$ using a Tukey multiple comparison test. (**C**) The relative level of soluble to total amount of RbcL and RbcS produced was ascertained by coomassie staining following SDS-PAGE separation of the soluble (S) and total (T) cellular protein.

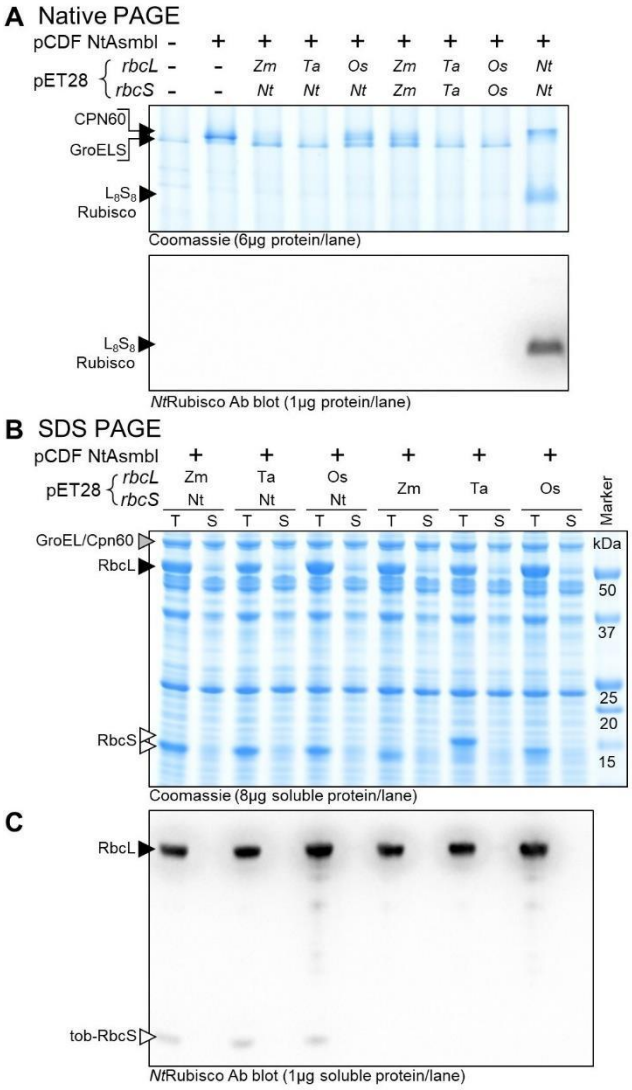

Figure 5. Monocot Rubisco subunits are produced but their assembly not met in tcc*E. coli*

Cultures of tcc*E. coli* transformed with pET-LS constructs coding the RbcL from maize (*Zm*), wheat (*Ta*) or rice (*Os*) with either the tobacco RbcS (tob-RbcS) or their cognate RbcS (*Zm*RbcS, *Ta*RbcS or *Os*RbcS) were induced with 1mM IPTG at an OD_{600} of 0.5 and grown at 23°C for 24h. (**A**) L_8S_8 Rubisco production in the soluble protein separated by native-PAGE was only detected in the tobacco Rubisco pET-*Nt*LS control. (**B**) Commassie staining and (**C**) *Nt*Rubisco immunoblotting following SDS-PAGE of the soluble (S) and total (T) cellular protein indicated the monocot RbcL and RbcS were produced, but insoluble. While the *Nt*Rubisco antisera recognises tob-RbcS, it does not recognise *Zm*RbcS, *Ta*RbcS or *Os*RbcS.

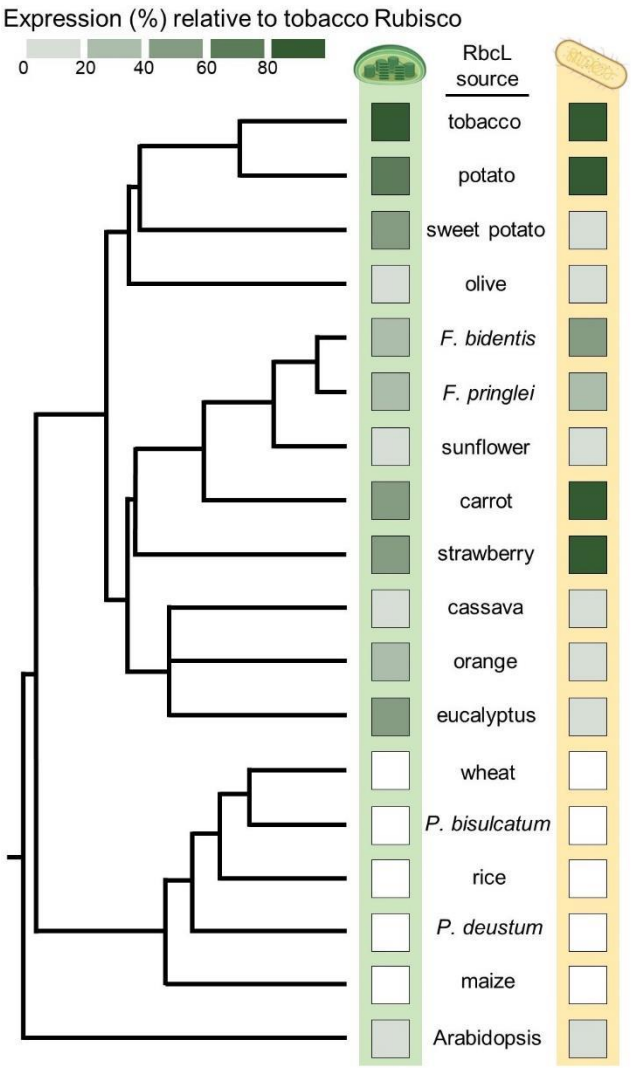

Figure 6. Correlative Rubisco biogenesis potential in tobacco chloroplasts and tcc*E. coli*

Mapping the comparative biogenesis potential of chimeric Rubiscos in tobacco chloroplasts and tcc*E. coli* against their RbcL phylogeny. The expression levels of each Rubisco were normalised relative to the tobacco Rubisco content in chloroplasts (Figure 1D) and tcc*E. coli* (Figure 4B). The RbcL phylogenetic tree was generated in Geneious using sequence identity and a global alignment, using the tree build model UPGMA and the genetic distance model Jukes-Candor. RbcL Genbank accessions were NP_051067.1 (Arabidopsis), SCM15159.1 (*P. deustum*), SCM15147.1 (*P. bisulcatum*), YP_588125.1 (*H. annus*), P19161.1 (*F. bidentis*), P19162.1 (*F. pringlei*) and those listed in Figure 1A.