A procedure to introduce point mutations into the Rubisco large subunit gene in wild-type plants

Myat T. Lin, Douglas J. Orr, Dawn Worrall, Martin A. J. Parry, Elizabete Carmo-Silva, Maureen R. Hanson

1 Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14850, USA.
2 Lancaster Environment Centre, Lancaster University, Library Avenue, Lancaster, LA1 4YQ, UK.

Running title: Introducing point mutations into chloroplast genes

Keywords: chloroplast transformation, food security, homologous recombination, Rubisco, photosynthesis, site-directed mutagenesis, Nicotiana tabacum

+ M.T. Lin and D.J. Orr should be considered joint first authors.

*Correspondence: m.parry@lancaster.ac.uk; tel +44 1524 595084
SUMMARY

Photosynthetic inefficiencies limit the productivity and sustainability of crop production, and the resilience of agriculture to future societal and environmental challenges. Rubisco is a key target for improvement as it plays a central role in carbon fixation during photosynthesis and is remarkably inefficient. Introduction of mutations to the chloroplast-encoded Rubisco large subunit rbcL is of particular interest to improve the catalytic activity and efficiency of the enzyme. However, manipulation of rbcL is hampered by its location in the plastome, with many species recalcitrant to plastome transformation, and by the plastid’s efficient repair system, which can prevent effective maintenance of mutations introduced with homologous recombination. Here we present a system where the introduction of a number of silent mutations into rbcL within the model plant *Nicotiana tabacum* facilitates simplified screening via additional restriction enzyme sites. This system was used to successfully generate a range of transplastomic lines from wild-type *N. tabacum* with stable point mutations within rbcL in 40% of the transformants, allowing assessment of the effect of these mutations on Rubisco assembly and activity. With further optimization, the approach offers a viable way forward for mutagenic testing of Rubisco function in planta within tobacco and modifying rbcL in other crops where chloroplast transformation is feasible. The transformation strategy could also be applied to introduce point mutations in other chloroplast-encoded genes.

Significance Statement

A simplified transformation strategy was developed to perform site-directed mutagenesis in the chloroplast-encoded Rubisco large subunit of tobacco. This approach reduces unwanted mismatch repair events, enables rapid screening of transformed lines that possess desired mutations and can be applied to modify any chloroplast-encoded gene in plant species amenable to plastid genome transformation.
INTRODUCTION

The transformation of chloroplast or plastid genomes in higher plants represents a promising technology in multiple biotechnological applications such as introducing agronomically important traits, metabolic engineering, recombinant protein expression and production of high-value therapeutic compounds (Maliga and Bock 2011). Chloroplast transformation offers several benefits: precise manipulation of its genome via homologous recombination, no transgene silencing and better control over the transgene escape into the environment. In addition to the introduction and expression of foreign genes, chloroplast transformation can also be used to delete or mutate plastid-encoded protein subunits for functional studies. Since many of these proteins are involved in photosynthesis, the technology has great potential to improve photosynthesis and productivity in crops (Hanson et al. 2013, Bock 2015, Martin-Avila et al. 2020).

One of the most commonly studied plastid genes is rbcL, which encodes the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). In plants, Rubisco catalyses two competing reactions in the stroma of chloroplasts: carboxylation or oxygenation of ribulose-1,5-bisphosphate (RuBP) (Ogren and Bowes 1971, Tcherkez et al. 2006, Andersson and Backlund 2008). The carboxylation of RuBP is an essential step for photosynthesis in plants, while its oxygenation generates 2-phosphoglycolate, which is recycled through the photorespiratory pathway, spanning multiple organelles (Ogren and Bowes 1971, Ogren 1984, Keys 1986, Busch 2020). Rubisco in C3 plants is a relatively slow enzyme with $k_{\text{cat}}$ about 3-4 CO$_2$ s$^{-1}$ (Orr et al. 2016, Flamholz et al. 2019). As a result, C3 plants generally express a large amount of Rubisco within the chloroplast stroma in order to achieve sufficient carbon fixation. Since the reactions of Rubisco control major metabolic fluxes, manipulating its kinetics has been an important target to improve photosynthesis in plants (Sharwood 2017, Zhu et al. 2020). Two general goals of engineering Rubisco are to improve its carboxylation efficiency under ambient O$_2$, which is defined as $k_{\text{cat}}^C / K_{M,\text{air}}^C$, and to increase its CO$_2$/O$_2$ specificity factor, which is the ratio of its carboxylation efficiency to oxygenation efficiency or $k_{\text{cat}}^C K_{M}^O / k_{\text{cat}}^O K_{M}^C$, where $k_{\text{cat}}$ and $K_M$ represent the catalytic turnover number and Michaelis-Menten constant, respectively (Whitney et al. 2011a). Increasing Rubisco content in maize and rice led to higher plant biomass,
supporting the hypothesis that Rubisco represents a bottleneck in photosynthesis, and crop yields can be improved with a more efficient Rubisco (Salesse-Smith et al. 2018, Yoon et al. 2020). However, to date attempts to engineer vascular plants with such a Rubisco enzyme have been unsuccessful.

Rubisco enzymes with different kinetic properties exist in nature despite its well-characterized catalytic constraints (Galmés et al. 2015, Galmés et al. 2016, Orr et al. 2016, Sharwood et al. 2016a, Flamholz et al. 2019). As organisms adapt to different environments, they have evolved Rubisco enzymes that are optimized to their immediate surroundings (Tcherkez et al. 2006, Savir et al. 2010). For example, Rubisco in C4 plants is generally associated with a lower affinity for CO2 or higher $K^C_M$ and a higher $k^C_{cat}$ compared to those in C3 plants (Whitney et al. 2011a). Modeling studies indicated that a typical C3 Rubisco is optimized for 220 ppm of atmospheric CO2 and Rubisco from several C4 plants would improve carbon fixation in C3 plants (Zhu et al. 2004, Sharwood et al. 2016b). Rubisco from red algae such as Griffithsia monilis were shown to have the highest known specificity factors and generally assumed to be promising candidates for improving photosynthesis in C3 plants (Whitney et al. 2011a).

Plants possess form I Rubisco, which is a hexadecameric complex made up of eight chloroplast-encoded large subunits (LSu) and eight small subunits (SSu) encoded by a family of RbcS nuclear genes and imported to the chloroplast stroma (Whitney et al. 2011a). Each holoenzyme complex consists of four LSu dimers with two active sites located inside each dimer at the interface between the two LSu (Andersson and Backlund 2008), and capped with 4 SSu monomers at each end. Plants produce specific chaperonins to prevent irreversible aggregation of LSu as well as multiple chaperones for step-by-step assembly of functional $L_8S_8$ complexes (Bracher et al. 2017, Wilson and Hayer-Hartl 2018). Thus, engineering plants with a more efficient Rubisco that can accumulate the enzyme at sufficiently high levels (Carmo-Silva et al. 2015) has been a major challenge.

Engineering Rubisco in higher plants has historically been carried out exclusively in tobacco (Nicotiana tabacum), where well-established procedures for chloroplast transformation allowed precise modification of its rbcl gene through homologous recombination and, recently, co-engineering of both Rubisco subunits (Sharwood 2017, Martin-Avila et al. 2020). Although
LSu from sunflower, Arabidopsis, and C₃ and C₄ Flaveria species were able to assemble with
native tobacco SSu to form functional enzymes in tobacco transformant plants, the
accumulation of such hybrid enzymes in the leaves was significantly lower than the normal
amount, likely due to suboptimal interactions between the foreign LSu and native chaperones
or SSu (Kanevski et al. 1999, Whitney et al. 2011b, Whitney et al. 2015). Attempts to replace the
*rbcL* gene in tobacco with red algal Rubisco genes produced transformants without functional
Rubisco due to incompatibility with the chaperonin machinery and chaperones in tobacco
chloroplast stroma (Whitney et al. 2001, Lin and Hanson 2018). A recent study demonstrated
that a red-type Rubisco from *Rhodobacter sphaeroides* was able to assemble and function in
tobacco chloroplasts, but its poor compatibility with native Rubisco activases led to low
activation levels in the absence of its cognate Rubisco activase from *Rhodobacter sphaeroides*
(Gunn et al. 2020). Thus, successfully expressing Rubisco sufficiently distant phylogenetically
from the host plant will likely require manipulation of ancillary proteins such as Rubisco activase
and/or assembly related factors.

Unless chaperones and SSu can also be optimized to work with a foreign LSu, an
alternative approach could be to introduce carefully selected site-directed mutations in the
native LSu via chloroplast transformation. However, generating such transformants from wild-
type tobacco has been inefficient because the desired mutations introduced by recombination
between the mutated *rbcL* and the native version of the gene can be removed by the plastid’s
repair system before transformants reached homoplasmy (Whitney et al. 1999, Kanevski et al.
1999). Thus, a tobacco master line has been created where the native *rbcL* gene was replaced
with the codon-modified *rbcL* gene from *Rhodospirillum rubrum* with low sequence homology
followed by removal of the selectable marker gene to facilitate the introduction of mutated and
foreign *rbcL* genes and polycistrons into tobacco (Whitney and Sharwood 2008). This tobacco
master line has been successfully used to study the residues critical for the catalytic properties
of C₄ Rubisco enzymes in Flaveria species as well as introduction of Arabidopsis LSu into tobacco
(Whitney et al. 2011b, Whitney et al. 2015). In a recent study, inhibiting the expression of native
SSu in the tobacco master line with RNA interference allowed the simultaneous transformation
of both Rubisco subunits and investigation of the effects of novel Rubisco complexes in tobacco
Martin-Avila et al. 2020). This tobacco master line has been a great resource to carry out Rubisco engineering in a model species, though the generation of such a master line in other plants is a lengthy process, and an efficient procedure to modify chloroplast genes including rbcL using a wild-type line would therefore be helpful.

In this study, we developed an approach to effectively introduce specific mutations into chloroplast-encoded genes. As a proof of concept, we synthesized a tobacco rbcL gene with synonymous or silent mutations resulting in unique restriction sites. Using this modified rbcL gene as a template, we introduced single or double residue substitutions that were predicted in previous phylogenetic and biochemical studies to be potentially important for the enzyme’s kinetic properties (Kapralov et al. 2012, Galmés et al. 2014, Studer et al. 2014, Orr et al. 2016). This approach allowed us to successfully replace the rbcL gene with the modified rbcL genes directly within wild-type tobacco plants with high retention of the mutations. A procedure to screen for transformants that possess the mutant rbcL genes is described, as well as preliminary analyses of Rubisco activities in ten different transplastomic lines.

RESULTS

Modified restriction sites in the rbcL gene allowed efficient screening of the transformants

We synthesized part of a modified rbcL gene (Nt-rbcLm) by introducing 26 silent mutations such that four restriction sites were removed while ten were added (Figure 1a, S1). The Nt-rbcLm gene was then seamlessly joined with the native rbcL promoter from tobacco and inserted into a chloroplast transformation plasmid, pCT-rbcL, described previously (Lin et al. 2014). In the resultant chloroplast transformation vector pCT-Nt-rbcLm, the Nt-rbcLm gene followed by the native rbcL terminator, Nt-TrbcL, and a selectable marker operon expressing the aadA gene driven by the tobacco psbA promoter are flanked between a 980 bp upstream homologous region or Flank 1, which contains the native rbcL promoter, and a 1kbp downstream homologous region or Flank 2 (Figure 1b). The homologous recombination between the plastid genome and pCT-Nt-rbcLm plasmid through the two flanking regions should introduce the aadA marker gene into the plastid genome and facilitate the selection of transformants with
spectinomycin on regeneration medium. However, it is expected that the intended mutations in
\( Nt-rbcl^m \) will be incorporated into only a portion of the transformants with the \( aadA \) gene. If the
cross-over site upstream of the \( aadA \) gene is located within \( Nt-rbcl^m \), the resulting
transformants will not possess those mutations upstream of the cross-over site (Figure 1c). Our
strategy is to use the unique restriction sites in \( Nt-rbcl^m \) to screen for the transformants with
the intended mutations.

Phylogenetic studies of Rubisco had previously suggested that several amino acid
substitutions in LSu were positively selected during the \( C_3 \) to \( C_4 \) transition (Kapralov et al. 2011,
Kapralov et al. 2012, Studer et al. 2014). As detailed in Table 1, we selected five of those
mutations (V101I, V255A, L270I, A281S and H282N) and introduced each into the \( Nt-rbcl^m \) gene
as well as a double mutant with both A281S and H282N. We also included the \( Nt-rbcl^n \) gene
with a C449G mutation that was suggested to be associated with improved catalytic efficiency
by a previous wide survey of Rubisco kinetic properties from 75 plant species (Orr et al. 2016).
In addition, we introduced the K429Q mutation to obtain the enzyme from tobacco’s paternal
parent, \( N. tomentosiformis \) (accession YP_398871.1). We also created the \( Nt-rbcl^n \) gene that
encodes both K429Q and L225I mutations as the inverse l225L change was found to be selected
in evolution of \( C_3 \) branches, and ancestors of tobacco Rubisco potentially possessed an L225I
mutation (Studer et al. 2014).

We introduced each of the pCT-Nt-rbcl\( ^m \) vectors with these mutations into tobacco
seedlings of either Samsun or Petit Havana tobacco cultivar with biolistics and performed
restriction digestion of the PCR-amplified \( rbcl \) gene from the transformed shoots arising from
the first selection round (Figure 2a). We found that typically 40 percent of the shoots at this
stage possessed the restriction sites corresponding to the \( Nt-rbcl^m \) gene. Our result was
comparable to that in a previous study, which introduced a different set of silent mutations to
remove commonly used restriction sites in the tobacco \( rbcl \) gene and found that 6 out of 12 or
50% of the transformants had the silent mutations (Sinagawa et al. 2009). After the shoots we
produced with the modified restriction sites at the \( rbcl \) locus were subjected to a second round
of selection, we analyzed restriction fragment length polymorphisms (RFLPs) in the
transformants with DNA blotting (Figure 2b,c). Those transformants that had achieved
homoplasmy with the \textit{Nt-rbcL}^m gene were transferred to rooting medium and subsequently to soil until they set seeds. We obtained multiple independent transformants for six out of nine LSu mutations and one each for the remaining three mutations (Table 1). Sequencing of the \textit{rbcL} locus in the final transformants confirmed that all but one of the transformants possessed the entire set of silent mutations in the \textit{Nt-rbcL}^m gene along with the intended non-synonymous mutations. The NtLwt transformant inherited all the silent mutations in the \textit{Nt-rbcL}^m gene minus the single nucleotide change for the new HindIII site (Table 1). Our results are consistent with the previous study which found that the majority, five out of six transformants with silent mutations in the \textit{rbcL} gene, possessed the entire set of mutations and only one originated from a cross-over event within the \textit{rbcL} gene (Sinagawa \textit{et al.} 2009).

\textbf{Mutations in the Rubisco large subunit did not prevent assembly of Rubisco holoenzyme}

We further investigated one tobacco transformant for each Rubisco LSu mutant to determine the effects of the introduced mutations. DNA blotting with a probe hybridized to a region upstream of the \textit{rbcL} gene locus confirmed that all transformants were homoplasmic and possessed a restriction site from the \textit{Nt-rbcL}^m gene (Figure 3). Likewise, an RNA blot of the same samples with a probe to detect the \textit{Nt-rbcL} gene showed that all transformants had an extra dicistronic transcript with both the \textit{Nt-rbcL}^m and downstream \textit{aadA} genes, in addition to the monocistronic \textit{Nt-rbcL}^m transcript (Figure 4a). Both of the \textit{rbcL} transcripts present in the transformants were much less abundant than the single monocistronic transcript from the wild-type sample, however, the total \textit{rbcL} transcript levels in the transformants were comparable to that from the single transcript in the wild-type (Figure 4b). The proportion of discistronic mRNA amongst the transgenic lines did not significantly vary (ANOVA, \(P = 0.848\)). Soluble protein samples from the leaf tissues of the wild-type plant and all transformants displayed a similar band for \(L_8S_8\) holoenzyme on blue native PAGE, indicating these mutations did not prevent Rubisco assembly, though measurements indicate an effect on enzyme abundance in the leaf (Table 2).

\textbf{Introduced mutations often affected Rubisco carboxylation}
To compare the potential impact of these introduced mutations on Rubisco activity an analysis was conducted using glasshouse-grown plants of each transplastomic line and wild-type controls. Leaf discs were taken from 26 day old plants and analysed for maximum Rubisco carboxylation rate ($k_{\text{cat}}$) and other related parameters including total soluble protein and chlorophyll content (Table 2, Table S1). There was significant variation amongst genotypes in the cv. Samsun background (Table 2). Two mutations, V101I and V255A, showed significant negative impact on Rubisco maximum carboxylation rate. Amongst cv. Petit-Havana genotypes there was some evidence for the negative impacts of the C449G mutation, with variation among these genotypes approaching statistical significance ($P = 0.054$, ANOVA). In cv. Petit-Havana genotypes there was also some support for the effect of mutations on Rubisco content and reflected in total soluble protein content (Table 2). Plants of each transplastomic line grew to comparable size and displayed no obvious phenotype under the conditions used (Figure 5). There were no significant differences in chlorophyll content (Table S1). Growth under additional environment conditions will be needed to further explore the effect of the mutations on plant phenotype and enzymatic activity.
DISCUSSION

Engineering the chloroplast genome, or plastome, to utilise the unique characteristics of this organelle is of increasing interest for goals such as improving photosynthesis and the use of plants as bio-factories. The central CO$_2$ fixing enzyme Rubisco, due its inefficiencies, has long been considered a promising target to improve photosynthesis and increase biomass and yield in crops. Manipulation of Rubisco subunit genes presents contrasting complications. Chloroplast transformation allows precise site-directed mutagenesis of the large subunit (rbcL) gene in the chloroplast genome, and although this has long been successfully used to investigate Rubisco biogenesis and biochemistry, the technology is currently limited to a relatively small set of species. Understanding of the pervasive role, in some cases, of the small subunit (encoded by the nuclear RbcS family) on catalysis is rapidly increasing, and despite the complexity of manipulating highly similar gene families, often involving a large number of homologues, advances are being made in this area (e.g. Martin-Avila et al. 2020, Khumsupan et al. 2020). In this study, we introduced ten restriction sites with 26 silent mutations in a modified tobacco rbcL gene (Nt-rbcL\textsuperscript{m}), which allowed PCR-RFLP screening of transformants after the first round of selection. The elimination of null transformants without the intended mutations after the first selection round also improved the efficiency of workflow to obtain the final stable transformants. Compared to a previous report where only one in eight transformants had the desired mutation in the rbcL gene (Whitney et al. 1999), about 40 percent of our transformants obtained after the first selection had Nt-rbcL\textsuperscript{m}, with the majority of these giving rise to stable transplastomic plants with the entire set of mutations in the Nt-rbcL\textsuperscript{m} gene. Our results are generally consistent with a previous study to remove common restriction sites in the rbcL gene with silent mutations (Sinagawa et al. 2009).

It proved useful to generate multiple shoots or subclones from each transformant after the first selection round since more than half of the shoots tested after the second selection round had lost the Nt-rbcL\textsuperscript{m}. This is not surprising since Nt-rbcL\textsuperscript{m} is over 98% identical to the wild-type sequence and can be removed before the transformants reach homoplasmy (Whitney et al. 1999, Kanevski et al. 1999). There may be potential to further improve the transformation efficiency by increasing the number of silent mutations in Nt-rbcL\textsuperscript{m} such that there is no
sequence homology to the wild-type gene. However, having too many silent mutations could possibly interfere with translation efficiency and other underlying sequence-specific regulatory processes, which are not yet fully understood. One such regulatory element that is widely conserved among plants is a major translation pause site within \( rbcl \) transcripts caused by an internal ribosome-binding site and mRNA structure (Gawroński et al. 2018).

Our analyses indicate that many of the transformants with the \( Nt-rbcl^m \) gene had lower Rubisco content than the wild-type plants although in some cases the difference was not statistically significant. One reason for reduced Rubisco levels could be the 23 modified codons in the \( Nt-rbcl^m \) gene, with the majority of these changes resulting in the incorporation of less frequently used codons that could negatively impact the protein’s translation efficiency. Although translation efficiencies in chloroplasts cannot always be predicted from codon usage and chloroplasts do not possess rare codons similar to those found in \textit{Escherichia coli}, different codons have varying translation efficiencies so that it may still be desirable to avoid unnecessary codon changes (Nakamura and Sugiura 2011). For example, strategically targeting a single restriction site closest to each non-synonymous mutation instead of introducing the entire set of silent mutations in the \( Nt-rbcl^m \) gene should minimise unintentional influence on translation efficiencies.

The processing of the \( rbcl \) mRNA 3’ end in our transformants was not efficient, giving rise to a dicistronic transcript with \( rbcl \) and \( aadA \) genes. This is likely due to insufficient length of the \( rbcl \) 3’-UTR, which was 205 nucleotides in the transformants. In addition to the dicistronic transcript, we also observed two sizes for monocistronic mRNAs, indicating a second transcript processing site that is likely in the \( psbA \) promoter downstream of the \( rbcl \) 3’-UTR. In a previous study where an \( rbcl \) 3’-UTR that was 269 nucleotides in length was incorporated downstream of the \( rbcl \) gene, the dicistronic transcripts were much less abundant, probably due to more efficient processing of the transcripts (Whitney and Sharwood 2008). It was previously shown that 410 nucleotides following \( rbcl \) was necessary for proper maturation of the \( rbcl \) transcript (Sinagawa \textit{et al.} 2009). Thus, future work should consider incorporation of a complete \( rbcl \) 3’-UTR that is at least 410 nucleotides long so that the transformants can produce \( rbcl \) transcripts that are similar in size and abundance to those in the wild-type. In
addition, stem-loop structures at the *loxP* sites flanking the *aadA* operon were suggested to interfere with the processing of both *rbcL* and *aadA* transcripts (Sinagawa *et al.* 2009). Thus, it may be preferable to replace *loxP* sites with long direct repeats that can spontaneously trigger removal of the marker gene through homologous recombination (Iamtham and Day 2000). Alternatively, the marker operon can be flanked with *attB* and *attP* sequences and subsequently removed with the expression of PhiC31 phage integrase (Kittiwongwattana *et al.* 2007).

Previously, the development of a tobacco master line, where the *rbcL* gene had been replaced with a homolog from *Rhodospirillum rubrum* encoding a form II Rubisco, allowed modification of the *rbcL* gene and rapid characterization of the subsequent mutant Rubisco enzymes within 6-9 weeks of transformation, although this master line required a high CO₂ environment to grow in soil (Whitney and Sharwood 2008). Recently, functional Rubisco enzymes from Arabidopsis and tobacco were successfully assembled in *Escherichia coli* with the co-expression of at least five chaperones (Aigner *et al.* 2017, Wilson *et al.* 2019, Lin *et al.* 2020). Thus, modified Rubisco enzymes from plants can now be readily produced in *E. coli*, and once modifications that lead to superior carboxylation kinetics are identified, they can be introduced into host plants for further characterization of their effects on photosynthesis and plant growth. These new tools are complemented by directed evolution approaches with cyanobacterial Rubisco, which have shown promising improvements to catalytic efficiency and specificity (Wilson *et al.* 2018).

As a proof of concept for this approach to mutating chloroplast genes, we generated seven tobacco transformants each with one residue substitution and two transformants each with two residue substitutions in the Rubisco LSu. Most of these residue substitutions were predicted to be selected during the C₃ to C₄ transitions or associated with a higher catalytic efficiency, while L225I and K429Q were potentially present in ancestors of tobacco Rubisco (Studer *et al.* 2014). Mutations that are *bona fide* kinetic switches between C₃ and C₄ Rubisco should result in enzymes with higher *k*ₕ (Whitney *et al.* 2011b). However, our analyses indicated a lack of significant improvement in the carboxylation rate of Rubisco from these transformants, indeed a number of the mutant Rubiscos displayed lower maximum carboxylation activities than the control plants. Importantly, the transformants exhibited
Rubisco with similar mobility on native PAGE as the wild-type plants, suggesting no impairment of Rubisco biogenesis as a result of the mutations introduced in rbcL.

A previous study identified a residue substitution in the Rubisco LSu of Flaveria species as a C3-C4 kinetic switch, but the same residue substitution which is native to the tobacco enzyme did not convey the same kinetic effect (Whitney et al. 2011b). Thus, kinetic switches likely require accessorial components, which may be one or more additional complementary mutations in the large or even the small subunit of Rubisco. Indeed, evidence that the residues at the interface between the two subunits play important roles in determining the kinetic properties was previously reported by studies of the Rubisco from Chlamydomonas (Spreitzer et al. 2005, Genkov and Spreitzer 2009) and cyanobacteria (Wilson et al. 2018). Understanding the role of the small subunit in plant Rubisco catalysis is an expanding area, and will be an important complement to engineering the large subunit via chloroplast transformation (e.g. Ishikawa et al. 2011, Morita et al. 2014, Atkinson et al. 2017, Khumosupan et al. 2020), alongside newly developed tools for rbcL-rbcS co-engineering in the plastome (Martin-Avila et al. 2020).

Reliable plastid transformation procedures for many agriculturally important crops are not yet available. A recent study showed that spectinomycin-resistant Arabidopsis plastid transformants could be selected with reasonably high efficiency once the nuclear-encoded ACC2, an enzyme subunit involved in fatty acid biosynthesis pathway inside plastids, was knocked out, rendering the plants hypersensitive to spectinomycin (Yu et al. 2017). Fertile Arabidopsis plastid transformants were then successfully generated when root-derived microcalli of acc2 knockout lines were used for transformation and regeneration (Ruf et al. 2019). These latest developments have potential to inform future attempts to extend plastid transformation technology to other species. Once the plastid genome in crops can be readily transformed, an important goal would be to employ changes within the Rubisco large subunit to improve its properties in key crops such as rice and wheat. Since our approach does not require initial generation of master lines in target plant species, which can be time-intensive, it can greatly expedite future work in introducing targeted changes not only in the Rubisco large subunit, but also in other chloroplast-encoded protein subunits.
EXPERIMENTAL PROCEDURES

Construction of the chloroplast transformation vectors with modified *Nt-rbcL* genes

All primers used were obtained from Integrated DNA Technologies and listed in Table S2. Phusion™ high-fidelity DNA polymerase, FastDigest restriction enzymes and T4 DNA ligase from Thermo Scientific were used to generate amplicons, restriction digests and ligation products respectively. The partial *Nt-rbcL*™ gene (bp172-1362) with silent mutations was synthesized by GenScript and amplified with NtLm-154f and NtLm-1343r primers (Figure 1a, S1). The upstream and downstream fragments were amplified from tobacco DNA with LSU-FL1f+NtLm-171r and NtLm-1363f+NtLrev primers respectively. The three amplicons were joined with overlapping PCR with LSU-FL1f and NtLrev primers, digested with ClaI and MauBI and ligated into the similarly digested pCT-rbcL vector (described in Lin et al. 2014) to obtain pCT-Nt-rbcL™ vector (Figure 1b).

To introduce V101I, L225I, A281S, H282N or A281S/H282N amino-acid substitutions, amplicons obtained with the respective forward primers and NtLrev were digested with *Mlu*-*Age*, *Age*-*Hind* III or *Hind* III-*Not* I and ligated into the similarly digested pCT-Nt-rbcL™ vector. V255A, L270I and C449G mutations were introduced by ligating the amplicons from NtL-97SF and the respective reverse primers into the *Mlu*-*Sac*, *Mlu*-*Hind* III or *Hind* III-*Not* I sites of pCT-Nt-rbcL™ vector. To generate a *Nt-rbcL*™ gene with the K429Q mutation, the two amplicons obtained with NtL-783SF+K429Qr and K429Qf+NtLrev primers were joined with overlapping PCR, digested with *Kpn*-MauBI and ligated into the similarly digested pCT-Nt-rbcL™ vector. To obtain pCT-Nt-rbcL™ vector with L225I/K429Q mutations, *Nt-rbcL*™ with K429Q was digested with *Kpn*-NotI and ligated into the similarly digestd pCT-Nt-rbcL™-L225I vector. All modified *Nt-rbcL*™ genes were confirmed by Sanger sequencing with NtL-219SR, NtL-97SF and NtL-783SF primers.

Generation of tobacco transformants with modified *Nt-rbcL*™ genes

Each MS agar plate with 25 two-week-old seedlings of either the Samsun or the Petit Havana cultivars of *Nicotiana tabacum* occupying 1 inch² area was bombarded twice with 0.6 µm gold
nanoparticles coated with pCT-Nt-rbcLm vectors in a PDS-1000/He Particle Delivery System (Bio-Rad Laboratories) and the leaf tissues were placed on RMOP agar medium with 500 µg/mL spectinomycin as described previously (Maliga and Tungsuchat-Huang 2014). Five plates of seedlings were bombarded for each transformation vector. The rbcL locus was amplified from DNA extracted from the shoots arising from the first selection round with NtL-97SF and NtLrev primers. The amplicons were purified, digested with MluI, AgeI or HindIII and analyzed on an agarose gel. The shoots possessing the restriction sites in the rbcL gene were cut into 25 mm² pieces and placed on RMOP agar medium with 500 µg/mL spectinomycin. After shoots appeared 4-6 weeks later, DNA was extracted from each shoot with CTAB solution (Allen et al. 2006), 1 µg each was digested with Nhel and Xhol and the RFLP was analyzed on a DNA blot with a DIG-labeled probe as described previously (Orr et al. 2020). The shoots homoplasmic with Nt-rbcLm genes were then placed on MS agar medium with 500 µg/mL spectinomycin for rooting and subsequently transferred to soil. The Nt-rbcLm genes in those plants were confirmed by Sanger sequencing.

Analyses of Nt-rbcLm transcripts in the tobacco transformants

RNA was extracted and purified from young leaves of five-week-old plants with PureLink RNA mini kit from Life Technologies, and 200 ng of each sample was separated in a denaturing gel with 1.3% agarose gel and 2% formaldehyde, transferred to a Nylon membrane and detected with a DIG-labeled RNA probe as described previously (Occhialini et al. 2016). The chemiluminescence detection was performed with a ChemiDoc MP imaging system (Bio-Rad) and the band intensities were quantified with Gel Analyzer options in ImageJ software (https://imagej.nih.gov/ij/). The experiment was performed for two sets of plants. Five different RNA concentrations (25, 50, 100, 200 and 400 ng) from a wild-type cv. Samsun plant were applied along with each set of samples, and a quadratic function obtained from their band intensities was used to estimate the relative rbcL transcript quantity in each band. Each rbcL transcript quantity was then normalized with the corresponding 25S rRNA band intensity obtained from the UV exposure of the agarose gel before transfer.
Analyses of Rubisco holoenzyme on blue native PAGE

Leaf tissues were homogenized in an extraction buffer consisting of 100 mM Bicine-NaHO pH 7.9, 5 mM MgCl$_2$, 1 mM EDTA, 5 mM ε-aminocaproic acid, 50 mM 2-mercaptoethanol, 5% (w/v) PEG 4000, 10 mM NaHCO$_3$, 10 mM DTT and Pierce™ protease inhibitor mini tablets (Thermo Scientific part # A32955) and insoluble materials were removed with centrifugation at 16,000 x g at 4°C for 5 min. The protein concentrations were estimated with the Bradford method using a protein assay dye reagent concentrate (Bio-Rad part # 5000006), and 3 µg of each total soluble extract mixed with a sample buffer consisting of 50 mM BisTris, 50 mM NaCl, 10% w/v glycerol and 0.001% w/v Ponceau S pH 7.2 was loaded to a NativePAGE™ 3-12% Bis-Tris protein gel (Thermo Scientific part # BN1003BOX). The electrophoresis was carried out in an XCell™ SureLock™ Mini-Cell with an anode buffer consisting of 50 mM Bis-Tris and 50 mM Tricine pH 6.8 and a cathode buffer consisting of additional 0.002% (w/v) Coomassie G-250 dye at 4°C 150V for 30 min and 250V for ~ 60 min. The gel was fixed with 100 mL of 40% methanol and 10% acetic acid solution for 15-30 min, stained with 100 mL 0.02% Coomassie R-250 in 30% methanol and 10% acetic acid solution for 15-30 min and destained with 100 mL 8% acetic acid solution for 2-5 hours.

Plant material

To obtain plant material for protein analyses, seeds of wild type and transgenic *N. tabacum* cv. Petit Havana and cv. Samsun were sown into planter trays of a commercial potting mix (Petersfield Products, UK) with a slow-release fertiliser (Osmocote, Scotts UK Professional, UK). All lines germinated at similar time post sowing (6 d), and seedlings were thinned out after ca. two weeks, with individuals transferred to 1 L pots after ca. three weeks. Plants were grown inside a heated glasshouse at Lancaster University, UK, during June-July with minimum day/night temperatures of 25/18 ± 2°C and a 16 h photoperiod. Supplemental lighting was supplied by sodium lamps when light levels fell below 200 µmol m$^{-2}$ s$^{-1}$. Plants were kept well-watered.

Protein extraction and analysis
Leaf samples for Rubisco analyses were collected from the youngest fully expanded leaf 32 days after sowing. Samples were collected 3 h after the beginning of the photoperiod, and plants positioned to avoid shading for at least 1 h prior to sampling. Three 0.5 cm² leaf discs were rapidly collected using a cork borer and immediately snap frozen, then stored at -80 °C prior to analysis.

Fozen leaf samples were homogenised in an ice-cold mortar and pestle in 0.6 mL of extraction buffer, the soluble proteins collected after centrifugation for 1 min at 4 °C and 14,700 × g, followed immediately by assays of initial and total Rubisco activity, as described by Carmo-Silva et al. (2017). Rubisco activation state was calculated as the ratio of initial/total activity. An aliquot (100 µL) of the same soluble protein extract was incubated at RT for 30 mins with 100 µL of CABP binding buffer (Carmo-Silva et al. 2017) including [14C]CABP (carboxyarabintol-1,5-bisphosphate), for determination of Rubisco content via [14C]CABP binding (Sharwood et al. 2016c).

The same soluble protein extract was used to determine total soluble protein (TSP) via Bradford assay (Bradford 1976). The method of Wintermans and de Mots (1965) was used to determine chlorophyll content, using 20 µL of the homogenate taken in duplicate prior to centrifugation. This was added to 480 µL ethanol, inverted to mix, and kept in the dark for 2-3 h. Absorbance was measured using a SPECTROstar Nano (BMG LabTech, UK) to determine concentrations of chlorophyll a and b.

Statistical analysis
Statistical differences between biochemical trait means were assessed via ANOVA. Where a genotype effect was observed (P < 0.05), a post-hoc Tukey’s honest significance difference test was used to conduct multiple pairwise comparisons. Analyses were performed using RStudio (version 1.2.5033 (R Studio Team 2019)) and R (version 3.6.2, (R Core Development Team 2013)). Plots were prepared with ggplot2 (Wickham 2016). Outliers were detected using the Tukey’s fences method, where outliers are defined as extreme values that are 1.5 times the interquartile range (1.5 IQR) below the first quartile or 1.5 IQR above the third quartile.
Data availability

The data generated in this study can be obtained from the corresponding author upon request.

Accession numbers

Tobacco RbcL sequence data can be found in the GenBank data library under accession number NP_054507.1.

ACKNOWLEDGMENTS

This work was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) under grant number BB/I024488/1 to M.A.J.P., the Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy under awards numbers DE-SC0014339 and DE-SC0020142 to M.R.H. and the US National Science Foundation (NSF) under grant number MCB-1642386 to M.R.H.

List of author contributions:

MTL, MRH, DJO, ECS & MAJP conceived research. MTL, MRH, DJO & ECS designed experiments. MTL, DJO & DW performed the experiments and analyzed data. All authors contributed to writing the manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.
SUPPORTING INFORMATION

Supplemental Figure S1. Sequence alignment of native \textit{rbcl} and modified \textit{rbcl} (\textit{Nt-rbcl}\textsuperscript{m})

Supplemental Figure S2. Blue native PAGE analyses of soluble proteins in the transformants

Supplemental Table S1. Chlorophyll data from analysis of tobacco transplastomic lines.

Supplemental Table S2. Oligonucleotide sequences used in the construction of transformation vectors

REFERENCES


Table 1. Summary on the generation of tobacco chloroplast transformants with point mutations in the Rubisco large subunit and the number of independent shoots obtained at each step of selection.

<table>
<thead>
<tr>
<th>Cultivar</th>
<th>Mutations</th>
<th>Predicted roles and references</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; round</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; round with expected restriction fragments</th>
<th>homoplasmic shoots after the 2&lt;sup&gt;nd&lt;/sup&gt; selection round</th>
<th>Final independent transformants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samsun</td>
<td>None (NtLwt)</td>
<td>Unmodified Rubisco</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1*</td>
</tr>
<tr>
<td>Samsun</td>
<td>V101I</td>
<td>C&lt;sub&gt;3&lt;/sub&gt; to C&lt;sub&gt;4&lt;/sub&gt; transition (Christin et al. 2008, Studer et al. 2014)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Samsun</td>
<td>V255A</td>
<td>C&lt;sub&gt;3&lt;/sub&gt; to C&lt;sub&gt;4&lt;/sub&gt; transition (Studer et al. 2014, Christin et al. 2008, Kapralov and Filatov 2007)</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Samsun</td>
<td>A281S</td>
<td>C&lt;sub&gt;3&lt;/sub&gt; to C&lt;sub&gt;4&lt;/sub&gt; transition (Christin et al. 2008, Studer et al. 2014)</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Samsun</td>
<td>H282N</td>
<td>C&lt;sub&gt;3&lt;/sub&gt; to C&lt;sub&gt;4&lt;/sub&gt; transition (Studer et al. 2014)</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Samsun</td>
<td>A281S,H282N</td>
<td>See A281S and H282N above</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Petit Havana</td>
<td>L270I</td>
<td>C&lt;sub&gt;3&lt;/sub&gt; to C&lt;sub&gt;4&lt;/sub&gt; transition (Christin et al. 2008, Studer et al. 2014)</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Petit Havana</td>
<td>L225I, K429Q</td>
<td>C&lt;sub&gt;3&lt;/sub&gt; branch for L225I (Studer et al. 2014, Kapralov and Filatov 2007)</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Petit Havana</td>
<td>K429Q</td>
<td>Found in paternal parent (N. tomentosiformis)</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Petit Havana</td>
<td>C449G</td>
<td>Improved catalytic efficiency (Orr et al. 2016)</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

* The NtLwt transformant possesses all the silent mutations in the \textit{Nt-rbcL} gene except for a single mutation necessary for the new HindIII site. All the other transformants have the entire set of silent mutations in the \textit{Nt-rbcL} gene as well as the intended non-synonymous mutations.
Table 2. Rubisco carboxylation rate ($k_{cat}$), content and total soluble protein (TSP) of transplastomic tobacco producing Rubisco with mutations in the large subunit. Transplastomic lines were compared statistically with the respective wild-type background. Values represent mean ± S.E. (n = 3-6 replicate plants), letters denote significant differences (p < 0.05) as determined by Tukey’s pairwise comparisons following ANOVA.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Rubisco $k_{cat}$</th>
<th>Rubisco (µmol sites m$^{-2}$)</th>
<th>TSP (g m$^{-2}$)</th>
<th>TSP (% Rubisco)</th>
<th>Activation state (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-SS</td>
<td>2.6$^a$ ± 0.2</td>
<td>24.6 ± 2.8</td>
<td>4.8 ± 0.5</td>
<td>36.2 ± 2.7</td>
<td>82.4 ± 2.6</td>
</tr>
<tr>
<td>NtLwt</td>
<td>2.4$^{ab}$ ± 0.2</td>
<td>18.6 ± 2.5</td>
<td>4.0 ± 0.7</td>
<td>36.0 ± 4.6</td>
<td>78.0 ± 2.9</td>
</tr>
<tr>
<td>V101I</td>
<td>1.6$^b$ ± 0.1</td>
<td>19.5 ± 1.1</td>
<td>4.7 ± 0.1</td>
<td>29.1 ± 1.3</td>
<td>71.2 ± 6.2</td>
</tr>
<tr>
<td>V255A</td>
<td>1.6$^b$ ± 0.0</td>
<td>18.2 ± 2.9</td>
<td>3.7 ± 0.2</td>
<td>34.1 ± 3.1</td>
<td>69.4 ± 5.9</td>
</tr>
<tr>
<td>A281S</td>
<td>1.8$^{ab}$ ± 0.2</td>
<td>17.8 ± 2.7</td>
<td>3.5 ± 0.3</td>
<td>35.4 ± 4.3</td>
<td>74.4 ± 5.1</td>
</tr>
<tr>
<td>H282N</td>
<td>2.2$^{ab}$ ± 0.3</td>
<td>15.3 ± 2.8</td>
<td>3.2 ± 0.2</td>
<td>33.6 ± 4.4</td>
<td>71.1 ± 4.5</td>
</tr>
<tr>
<td>A2815/H282N</td>
<td>2.1$^{ab}$ ± 0.2</td>
<td>18.9 ± 3.3</td>
<td>3.5 ± 0.3</td>
<td>37.2 ± 3.8</td>
<td>72.4 ± 2.1</td>
</tr>
</tbody>
</table>

P = 0.007

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Rubisco $k_{cat}$</th>
<th>Rubisco (µmol sites m$^{-2}$)</th>
<th>TSP (g m$^{-2}$)</th>
<th>TSP (% Rubisco)</th>
<th>Activation state (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-PH</td>
<td>2.5 ± 0.2</td>
<td>26.6$^a$ ± 1.3</td>
<td>5.6$^a$ ± 0.4</td>
<td>33.9 ± 2.5</td>
<td>71.1 ± 2.5</td>
</tr>
<tr>
<td>C449G</td>
<td>1.5 ± 0.2</td>
<td>22.1$^{ab}$ ± 4.2</td>
<td>5.2$^{ab}$ ± 0.5</td>
<td>29.3 ± 3.2</td>
<td>68.3 ± 1.9</td>
</tr>
<tr>
<td>K429Q</td>
<td>1.9 ± 0.1</td>
<td>19.7$^{ab}$ ± 2.2</td>
<td>5.0$^{ab}$ ± 0.6</td>
<td>28.0 ± 2.6</td>
<td>68.4 ± 2.8</td>
</tr>
<tr>
<td>L225I/K429Q</td>
<td>2.4 ± 0.3</td>
<td>15.1$^b$ ± 0.7</td>
<td>3.2$^b$ ± 0.6</td>
<td>35.6 ± 6.7</td>
<td>83.6 ± 4.5</td>
</tr>
<tr>
<td>L270I</td>
<td>1.9 ± 0.2</td>
<td>19.3$^{ab}$ ± 2.4</td>
<td>4.4$^{ab}$ ± 0.3</td>
<td>30.9 ± 3.3</td>
<td>74.7 ± 3.9</td>
</tr>
</tbody>
</table>

P = 0.054

FIGURE LEGENDS

Figure 1. Modification of the tobacco rbcL gene for chloroplast transformation. (a) Schematic diagrams of the synthesized Nt-rbcL$^m$ gene with modified restriction sites. (b) The chloroplast transformation vector pCT-Nt-rbcL$^m$ for replacing the tobacco rbcL gene with Nt-rbcL$^m$. (c) Comparison of two hypothetical scenarios of cross-over events between the rbcL locus of the plastid genome and pCT-Nt-rbcL$^m$ plasmid upstream of the selectable marker aadA gene. The location of a hypothetical point mutation in Nt-rbcL$^m$ gene on pCT-Nt-rbcL$^m$ plasmid is indicated
with an ‘X’. The two cross-over events are indicated with dashed lines between the plastid genome and pCT-Nt-rbcl\(^m\) plasmid. If the cross-over takes place in Flank 1 or inside Nt-rbcl upstream of ‘X’ as in event 1, the point mutation will be introduced into the transformant. On the other hand, the cross-over taking place downstream of ‘X’ will fail to introduce the point mutation into the transformant. Note that although the segment upstream of Nt-rbcl has been marked as “Flank 1”, Nt-rbcl should be considered as part of Flank 1 due to high homology between Nt-rbcl and Nt-rbcl\(^m\).

**Figure 2.** Selection of tobacco transformants with the Nt-rbcl\(^m\) gene. (a) Restriction analyses of the amplified rbcl locus from tobacco transformants after the first selection round. The examples shown include MluI, Agel and HindIII digests of transformants with V101I, H282N and A281S/H282N mutations respectively. The samples with digestion fragments expected for Nt-rbcl\(^m\) are indicated with asterisks (*). (b) Schematics of the rbcl loci of the wild-type and Nt-rbcl\(^m\) transformants with the restriction sites used in DNA blot analyses. The green bars represent the region where the probe binds. (c) DNA blot analyses of tobacco transformants with Nt-rbcl\(^m\). The samples with successful replacement of the native rbcl gene with Nt-rbcl\(^m\) are indicated with asterisks (*).

**Figure 3.** DNA blot analyses of the tobacco chloroplast transformants with Nt-rbcl\(^m\) gene encoding Rubisco large subunit mutants. (a) Schematics of the the rbcl loci of the wild-type (wt) and Nt-rbcl\(^m\) transformants with the restriction sites used in DNA blot analyses. The green bars represent the region where the probe binds. (b) DNA blot indicating successful replacement of the native rbcl gene with Nt-rbcl\(^m\) in the transformants.

**Figure 4.** RNA blot analyses of the rbcl transcripts in tobacco transformants with Nt-rbcl\(^m\) encoding Rubisco large subunit mutants. (a) The RNA blot for rbcl transcripts in the wt (left six lanes and transformants (right ten lanes). In addition to the monocistronic rbcl transcript, all transformants display an additional dicistronic transcript with rbcl and aadA genes. The bottom panel shows ethidium bromide staining of the agarose gel. 25S rRNA band was used to estimate
the relative amounts of total RNA loaded for each sample. The transformants are in the same order as in (b). (b) The relative quantities of \( rbcL \) transcripts in the transformants compared to that in the wild-type plants obtained from two sets of plants. \( P \) values were determined by Tukey's pairwise comparisons following ANOVA.

**Figure 5.** Visual comparison of plant development. (a) wild-type tobacco cv. Samsun (WT-SS) and \( rbcL \) transplastomic lines, and (b) wild-type tobacco cv. Petite Havana (WT-PH) and \( rbcL \) transplastomic lines. Photos were taken of plants at the same age (35 days after sowing) and growth stage.