Engineering Marchantia polymorpha chloroplasts for the production of high-value specialized terpenes

Summary:
Originally, three independent operon-like synthetic constructs should be built to achieve de novo synthesis of mono-, sesqui- and triterpenes in M. polymorpha chloroplasts. GoldenGate modules of coding sequences to be expressed in M. polymorpha were synthesized. However, two major issues were encountered during the project, including problems with transforming M. polymorpha chloroplasts with large constructs, and an assembly defect of the 2A peptide system used for generating the clusters. To circumvent these obstacles, constructs allowing nuclear transformation of M. polymorpha and subsequent chloroplast targeting of the proteins were designed and a new 2A peptide system has been created and is currently being evaluated.

Report and Outcomes:
Originally, three independent operon-like synthetic constructs were designed for the *de novo* synthesis of mono-, sesqui- and triterpenes in *M. polymorpha* chloroplasts (Genbank files pChloro). The coding sequence of each heterologous gene to be expressed in *M. polymorpha* was subjected to the *Marchantia* chloroplastic codon optimization tool previously available on the Integrated DNA Technologies (IDT) website (https://eu.idtdna.com/CodonOpt, “Chloroplast *Marchantia polymorpha* (liverwort)” tool). The codon-optimized sequences were used to design and synthesize Golden Gate modules (Genbank files EC81672 to EC81687, EC81704 and EC81705) compatible with the 2A peptide building blocks available at the JIC at the start of the project. From an early stage in the project it became evident that *Marchantia* chloroplast transformation using such large constructs could be an issue. As our objective was to synthesize terpenes in chloroplasts, three additional constructs allowing nuclear transformation of *M. polymorpha* and subsequent chloroplast targeting of the proteins were designed (Genbank files pNucl). For chloroplast targeting we used the *M. polymorpha* chloroplast-transit peptide, previously characterised at the University of Cambridge (module EC81783). A set of six constructs was built to test the ability of the designed vectors to produce mono-, sesqui- and triterpenes in *Nicotiana benthamiana* agroinfiltration-based transient expression assay, prior to stable transformation of *M. polymorpha*. The construction of these vectors revealed a major flaw in the 2A peptide system available at that time in JIC. The 4bp linkers that define the ordered assembly of the 2A-flanked sequences into the Golden Gate binary expression vectors were very similar to each other (only one base pair difference between each linker), which resulted in the creation of random incorrect assemblies. To overcome this, a new set of 2A peptide modules and terpene biosynthetic modules, fully compatible with the plant synthetic biology common syntax (Patron et al., 2015) was created (EC81823 to EC81828, EC81831, EC81832 and EC81846 to EC81855). The efficiency of the 2A modules to mediate production of independent fluorescent proteins was successfully assessed using an agrobacterium-mediated *N. benthamiana* transient expression assay (Figure 1, construct EC80046). Also the assembly of the six large operon-like constructs using the newly created modules was efficiently achieved (Genbank files EC81833 to EC81835 and EC81866 to EC81868). Nevertheless, none of these

six constructs were functional when transiently expressed in *N. benthamiana*. Confocal microscopy analysis of the infiltrated leaves revealed the absence of accumulation of mTurquoise fluorescent reporter present at the end of each construct. Furthermore, Gas Chromatography - Mass Spectrometry (GC-MS) analysis of the leaf extracts confirmed the absence of terpene production in the transiently transformed *N. benthamiana* leaves. Currently we are performing a series of tests in order to troubleshoot the experiment and identify the origin of the problem. Proper functioning of the 2A peptide system implies the production of a single polycistronic mRNA, in which the coding sequences of all the genes of interest and the 2A peptide sequences form a single open reading frame (ORF). It is thus critical that no modification of the DNA sequence (nucleotide insertion/deletion or nonsense mutation) disrupts that ORF. To try to rule out the possibility that codon optimizing all the sequences and assembling them did not create any cryptic splicing sites, all the expression vector sequences that were designed, were subjected to the intron/exon pattern prediction tool GENSCAN (http://genes.mit.edu/GENSCAN.html). This analysis confirmed the absence of theoretical cryptic splicing sites in the polycistronic mRNA encoded by our expression vectors. It is as well possible that the use of the 2A peptide system is responsible for the production of non-functional proteins, as the system relies on a co-translational mechanism where ribosomes skip the synthesis of the glycyl-prolyl peptide bond at the C-terminus of a 2A peptide, leading to the cleavage between a 2A peptide and its immediate downstream peptide. As a result, the cleaved-off downstream peptide has proline at its N-terminus and the cleaved-off upstream peptide harbours the entire 2A peptide sequence minus the last proline. The additional 23 amino acids fused to the C-terminus of the terpene biosynthetic enzymes and/or the proline residue remaining on their N-terminus could thus be responsible for their loss of function. To ensure this is not the case, the constructs are being built without the 2A peptide system but with independent expression cassettes for each gene. These will be transiently expressed in *N. benthamiana* for functional validation. Failure of that experiment would lead to the conclusion that the *Marchantia* chloroplast codon optimisation could be responsible for the non-functional constructs. Introduction of rare codons into these sequences could prevent the plant from efficiently translating the recombinant mRNAs into the proteins of interest.

Expenditure:

Item	**Cost (£)**	**Balance**
DNA synthesis	2173.23	+ 1826.77
Marchantia transformation consumables	585	+ 1241.77
GC-MS	50	+ 1191.77
Cloning consumables	120	+ 1071.77

Are you claiming the additional £1000 follow-on funding?: No

Follow-On Plans:

We still have 1071.77£ out of the original 4000£ (as described in the table above). That money should be sufficient to cover the future cost of the Marchantia transformation as well as the cost of the GC-MS analysis of the N. benthamiana and Marchantia material that will be generated. We then do not need the additional 1000£ but we would like to keep the 1071.77£ that were not spent during the 6 first months.