

A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop

Paulina Fuentes, Fei Zhou[†], Alexander Erban, Daniel Karcher, Joachim Kopka, Ralph Bock*

Max-Planck-Institut für Molekulare Pflanzenphysiologie, Potsdam-Golm, Germany

Abstract Artemisinin-based therapies are the only effective treatment for malaria, the most devastating disease in human history. To meet the growing demand for artemisinin and make it accessible to the poorest, an inexpensive and rapidly scalable production platform is urgently needed. Here we have developed a new synthetic biology approach, combinatorial supertransformation of transplastomic recipient lines (COSTREL), and applied it to introduce the complete pathway for artemisinic acid, the precursor of artemisinin, into the high-biomass crop tobacco. We first introduced the core pathway of artemisinic acid biosynthesis into the chloroplast genome. The transplastomic plants were then combinatorially supertransformed with cassettes for all additional enzymes known to affect flux through the artemisinin pathway. By screening large populations of COSTREL lines, we isolated plants that produce more than 120 milligram artemisinic acid per kilogram biomass. Our work provides an efficient strategy for engineering complex biochemical pathways into plants and optimizing the metabolic output.

DOI: [10.7554/eLife.13664.001](https://doi.org/10.7554/eLife.13664.001)

*For correspondence: rbock@mpimp-golm.mpg.de

Present address: [†]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

Competing interests: The authors declare that no competing interests exist.

Funding: See page 24

Received: 09 December 2015

Accepted: 08 May 2016

Published: 14 June 2016

Reviewing editor: Joerg Bohlmann, University of British Columbia, Canada

© Copyright Fuentes et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Artemisinin, a C₁₅ isoprenoid (sesquiterpene) naturally produced in the wild plant *Artemisia annua* (sweet wormwood, native to temperate Asia), is the main ingredient of artemisinin combination therapies (ACTs), currently the only effective cure of malaria (Okell et al., 2014). As ACTs are the mainstay of malaria treatment and no alternative to artemisinin derivatives is expected to enter the market in the foreseeable future, there is a steadily increasing demand for ACTs which reached nearly 400 million treatment courses in 2013 (http://www.who.int/malaria/publications/world_malaria_report_2014). The mechanism of action of artemisinin on the malaria parasites *Plasmodium falciparum* and *P. vivax* is not entirely clear, but it is generally believed that the reactive endoperoxide bridge present in the molecule (Figure 1) is responsible for its medicinal properties. In addition to their antimalarial activity, artemisinin and its derivatives are currently also considered as promising anti-cancer, antiviral and anti-inflammatory agents (e.g., Willoughby et al., 2009). In *A. annua*, artemisinin is produced in the cytosol of the glandular trichomes of leaves and flowers (Tang et al., 2014). The biosynthesis initiates with the conversion of the isoprenoid building blocks IPP and DMAPP into farnesyl pyrophosphate (FPP) which is then converted into amorpha-4,11-diene by amorphaadiene synthase (ADS), the enzyme catalyzing the first committed step of the pathway (Figure 1). Amorpha-4,11-diene is a volatile compound that is oxidized to artemisinic alcohol and subsequently to artemisinic aldehyde by the cytochrome P450 monooxygenase CYP71AV1 (CYP) and its redox partner, the cytochrome P450 reductase (CPR). Artemisinic aldehyde is then further oxidized to artemisinic acid by the same enzyme pair, or alternatively, is reduced to dihydroartemisinic aldehyde by the double bond reductase 2 (DBR2; Figure 1; Zhang et al., 2008). Artemisinic acid can be efficiently and cheaply converted to artemisinin by chemical means (Paddon et al., 2013;

eLife digest Malaria is by far the most devastating tropical disease in the world. It affects hundreds of millions of people – mainly in Africa and Asia – with almost half a million deaths every year. The most effective therapies against malaria all include the drug artemisinin, which is naturally found in an Asian plant called *Artemisia annua*. Unfortunately, the artemisinin content of *A. annua* plants is relatively low and the demand for this drug outstrips the supply of the plant. The costly production process makes artemisinin-based treatments inaccessible to many of the people in the most badly affected regions, and so researchers have been trying to find new ways to produce this drug.

Genetically modifying crop plants, such as tobacco, to produce artemisinin or related compounds could potentially provide a more sustainable and cheaper source of the drug. Inside plant cells, a structure called the nucleus contains DNA that encodes most of a plant's genes, but compartments called mitochondria and chloroplasts also contain some DNA. Existing methods to genetically modify plants are able to insert a few genes into either the nucleus or the chloroplasts at a time. However, the production of artemisinin in *A. annua* involves many different genes that act at different stages of the process, and the precise roles played by many of them remain unclear.

Fuentes et al. developed a new approach to insert many of the *A. annua* genes involved in artemisinin production into tobacco plants at the same time, instead of one-by-one. The new method, referred to as COSTREL, takes advantage of the researchers' ability to insert new genes into both the nucleus and the chloroplast of the tobacco plants. In the first step, Fuentes et al. inserted a core set of genes that are essential to make artemisinin into the chloroplast. This enabled the plants to produce a molecule called artemisinic acid, which the researchers can extract from the plants and convert into artemisinin by simple chemical reactions.

After testing different arrangements of the genes in the chloroplast, the plant line that had the highest levels of artemisinic acid was used to introduce a set of "accessory" genes into the nuclear DNA. These accessory genes are not strictly required to make the drug, but they help to regulate the process in a largely unknown manner. The experiments generated hundreds of genetically modified plant lines that each have different combinations of the accessory genes. Fuentes et al. examined these lines and were able to identify plants that could produce large amounts of artemisinic acid. Therefore, these findings lay the foundations for a cheap way to produce this life-saving drug in tobacco. In the future, the COSTREL method developed by Fuentes et al. could also be used to genetically engineer other complex biochemical processes into plants.

DOI: [10.7554/eLife.13664.002](https://doi.org/10.7554/eLife.13664.002)

Kopetzki et al., 2013) and, therefore, represents a high-value precursor for the industrial production of artemisinin-based pharmaceuticals (**Paddon and Keasling, 2014**).

In view of the great medicinal value of artemisinic compounds, their low accumulation levels in *A. annua* and the unstable supply of the plant, enormous efforts have been undertaken to produce artemisinic compounds synthetically or in heterologous biological systems. Currently, the semisynthetic synthesis in yeast (**Paddon et al., 2013**) represents the most efficient heterologous production system for artemisinic acid, the immediate precursor of artemisinin (**Figure 1**). However, the production costs are still high and ACTs remain unaffordable to many people in the tropical and subtropical regions of Africa and Asia that are most severely afflicted with malaria. Since production in yeast requires large volumes of costly synthetic culture media and large-capacity bioreactors run under sterile conditions, production in plants can potentially provide a much cheaper, renewable and easily scalable source of artemisinic acid. Although the production of artemisinic compounds at low levels has been shown to be feasible in heterologous plant systems (**Wu et al., 2006; van Herpen et al., 2010; Zhang et al., 2011; Farhi et al., 2011**), the development of an efficient production system for the drug precursor artemisinic acid has not been achieved.

Here we have pursued a novel synthetic biology approach towards high-level production of artemisinic acid in chloroplasts of tobacco (*Nicotiana tabacum*), a fast-growing crop that produces high amounts of biomass at very low cost. We show that by implementing the core pathway in the chloroplast and subsequently selecting for optimum combinations and expression levels of additional

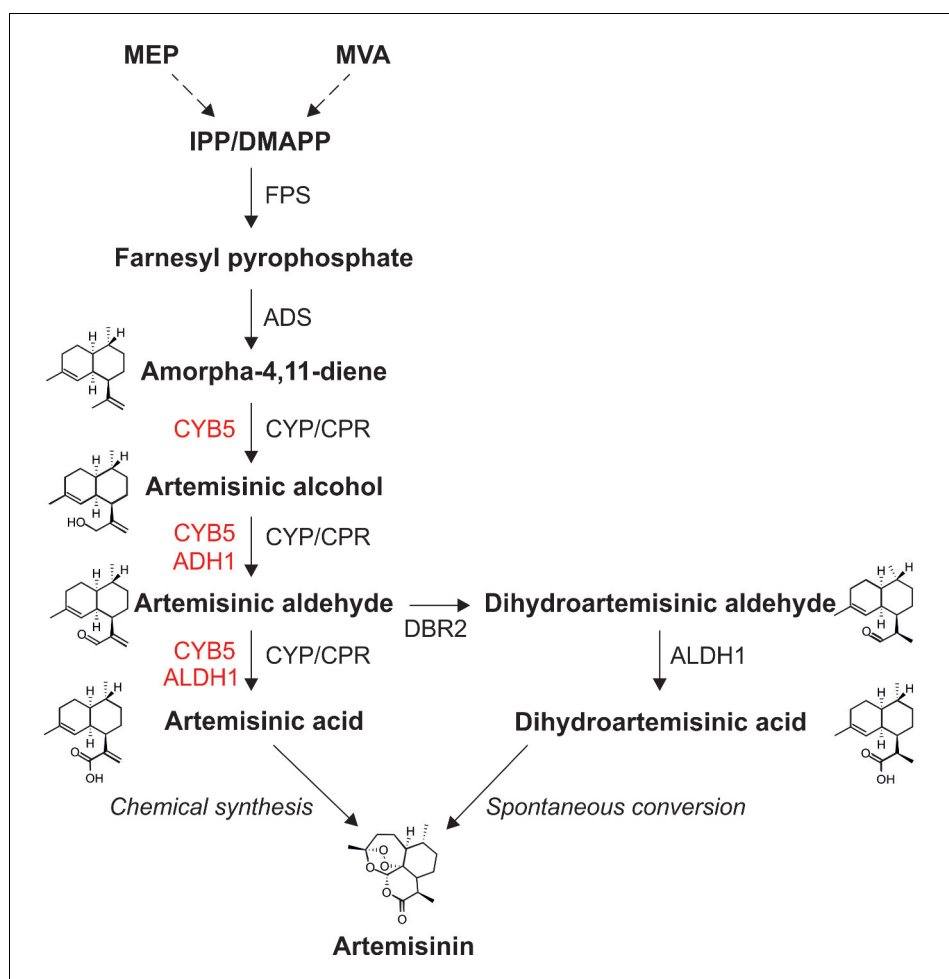


Figure 1. Metabolic pathway of artemisinin biosynthesis. The canonical pathway of artemisinin synthesis starts with the conversion of IPP/DMAPP (C_5 isoprenoids produced by the MVA pathway in the cytosol or the MEP pathway in the chloroplast) into farnesyl pyrophosphate (FPP), catalyzed by farnesyl pyrophosphate synthase (FPS). Amorpha-4,11-diene synthase (ADS) converts FPP into amorpha-4,11-diene in the first committed step of the pathway. Amorpha-4,11-diene is then successively oxidized to artemisinic alcohol, artemisinic aldehyde and artemisinic acid by the cytochrome P450 monooxygenase CYP71A1 (CYP) and its redox partner, the cytochrome P450 reductase (CPR). In *A. annua*, artemisinic aldehyde is converted to dihydroartemisinic aldehyde by DBR2, and then to dihydroartemisinic acid by ALDH1. Artemisinin is generated by the spontaneous oxidation of dihydroartemisinic acid *in planta*, and can be produced by chemical conversion of artemisinic acid *in vitro*. Enzymes depicted in red improve the efficiency of different oxidation steps in yeast (Paddon *et al.*, 2013; Paddon and Keasling, 2014). See text for details.

DOI: 10.7554/eLife.13664.003

pathway enzymes from a large population of combinatorially supertransformed transplastomic lines, artemisinic acid can be produced in tobacco leaves to levels of more than 120 mg/kg fresh weight.

Results

Expression of the core pathway for artemisinic acid synthesis from the plastid genome

The core enzymes to synthesize artemisinic acid are FPS, ADS, CYP and CPR (Figure 1). Accessory enzymes (indicated in red in Figure 1) and additional enzymes facilitating more efficient biosynthesis of artemisinin are CYB5, ADH1, ALDH1 and DBR2. We first implemented the canonical pathway from FPP to artemisinic acid in tobacco chloroplasts using stable plastid genome transformation

(Svab and Maliga, 1993; Bock, 2015). To this end, we designed a number of synthetic operons (Zhou et al., 2007; Lu et al., 2013) that combine the genes for the four core enzymes (FPS, ADS, CYP and CPR; Figure 1) in different arrangements and under the control of different expression signals (Figure 2A). Four synthetic artemisinic acid operon constructs (pAO1-4) were built and introduced into the chloroplast (plastid) genome of tobacco plants by particle gun-mediated transformation. Chloroplast-transformed (transplastomic) lines were selected on regeneration medium with spectinomycin and purified to homoplasmy by additional rounds of selection and regeneration (Svab and Maliga, 1993; Bock, 2015). Restriction fragment length polymorphism (RFLP) analysis verified integration of the synthetic operon constructs into the plastid genome by homologous recombination and successful elimination of all wild-type copies of the highly polyploid chloroplast genome (Figure 2B). Homoplasmy of the transplastomic lines was additionally verified by seed assays that confirmed lack of segregation of the spectinomycin resistance and uniparentally maternal inheritance (Figure 2C).

All homoplasmic transplastomic lines grew autotrophically under greenhouse conditions and produced viable seeds. However, transplastomic lines obtained with constructs pAO1 and pAO3 (Figure 2A) displayed a slightly pale-green phenotype and a subtle growth delay at the juvenile stage (Figure 3A; Figure 3—figure supplement 1). This phenotype could be due to toxicity of artemisinic metabolites produced in these plants (Bharati et al., 2012) or, alternatively, depletion of isoprenoid precursors from other metabolic pathways in the cell, such as carotenoid and chlorophyll biosyntheses. Measurement of chlorophylls and carotenoids confirmed that, indeed, both pigment classes are significantly reduced in plants exhibiting the mild phenotype (Figure 3—figure supplements 1 and 2). Metabolite profiling (see Materials and Methods) of the transplastomic lines revealed that all lines accumulated the volatile artemisinic acid precursor amorpha-4,11-diene and its first oxidation product artemisinic alcohol (Figure 1; Figure 3). Interestingly, amorpha-4,11-diene accumulated to lower levels in the lines displaying the subtle phenotype, whereas artemisinic alcohol was detected in similar amounts in all transplastomic plants. Accumulation of artemisinic acid correlated with the altered phenotype in *Nt-AO1-1* and *Nt-AO3-1*, suggesting that a more efficient conversion of amorpha-4,11-diene to downstream metabolites could be the cause of the phenotype. This hypothesis gained support from the analysis of a series of developmental stages and leaf ages which revealed that, while in *Nt-AO2* plants, artemisinic acid accumulates only in mature leaves of young and flowering plants, it accumulates throughout development in *Nt-AO3* plants. These analyses also confirmed the inverse relationship between artemisinic acid and amorpha-4,11-diene accumulation (Figure 4A–C).

To identify the molecular basis of the striking metabolic differences between the different operon constructs, a series of northern blot experiments was conducted. In view of the commonalities of the *Nt-AO2* and *Nt-AO4* plants versus the *Nt-AO1* and *Nt-AO3* plants, it seemed reasonable to assume that the relative orientation of the two operons (Figure 2A) is causally responsible for the different visual and metabolic phenotypes. When the expression of the four transgenes was assayed, a striking difference was observed in the *CYP/CPR* expression ratio in that high artemisinic acid accumulation correlated with a high *CYP/CPR* expression ratio in *Nt-AO3* plants (Figure 4D–G; Figure 4—figure supplement 1). In nature, CYPs are often found in excess to their CPR counterparts, with ratios of 10–100:1 or higher (reviewed, e.g., in Guengerich, 2002). Even though there is currently no consensus explanation for this observation, it is known that CPRs can activate molecular oxygen, thereby producing superoxide radicals and wasting redox capacity of the cell (Manoj et al., 2010). A high *CYP/CPR* ratio would prevent this CPR-mediated toxicity and result in a more efficient use of the redox power of the cell for artemisinic acid synthesis, as observed in our *Nt-AO1* and *Nt-AO3* transplastomic lines. This explanation is also in agreement with published data on transcript accumulation and protein abundance for these two enzymes in *A. annua*. While the levels of the *CPR* transcript and the *CPR* protein remain constant during development of the plant and in different organs, the transcript and protein levels of *CYP* increase in the developmental stages and organs where artemisinin synthesis is induced (Olofsson et al., 2011; Zeng et al., 2008). Especially the final oxidation step from artemisinic aldehyde to artemisinic acid appears to require an effective monooxygenase (Ting et al., 2013), suggesting that the higher levels of artemisinic acid in our transplastomic *Nt-AO1* and *Nt-AO3* lines are most likely related to their higher *CYP/CPR* expression ratio. However, determination of the *CYP* and *CPR* protein accumulation levels (and enzyme activities) would be

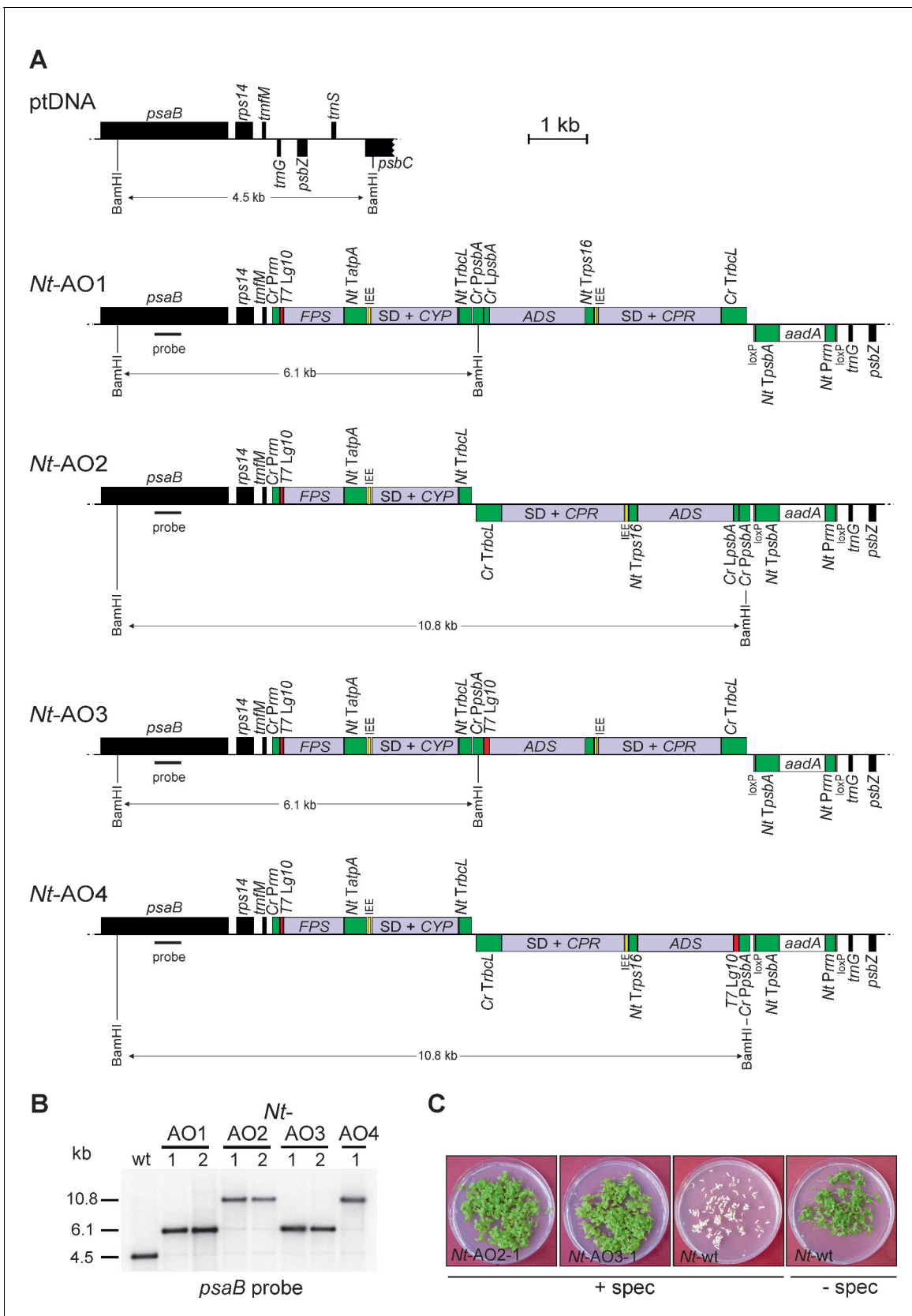


Figure 2. Implementation of the canonical pathway of artemisinin acid biosynthesis in chloroplasts. Synthetic codon-optimized genes for the four enzymes required to produce artemisinin acid (Figure 1) were introduced into the tobacco plastid genome by stable genetic transformation with four Figure 2 continued on next page

Figure 2 continued

different synthetic operon constructs (pAO1-4). The constructs differ in gene arrangement and in the translation signals that drive synthesis of the key pathway enzyme (ADS) catalyzing the first committed step. (A) Physical map of the plastid genome region (ptDNA) used for integration of the synthetic artemisinic acid operons and maps of the transgenic loci in the generated transplastomic tobacco lines (*Nt*-AO1-4). The artemisinic acid operon genes are depicted as light blue boxes. Chloroplast promoters and terminators are shown in green, the *aadA* selectable marker gene for chloroplast transformation is represented as a white box, and genes in flanking plastid sequences used for transgene targeting via homologous recombination are in black. Genes above the line are transcribed from left to right, genes below the line are transcribed in the opposite direction. The four transgenes are arranged in two dicistronic operons. *FPS* and *CYP* are driven by the *Chlamydomonas reinhardtii* plastid ribosomal RNA operon promoter (*Cr Prrn*) and the *g10* leader sequence from phage T7 (*T7 Lg10*). The second operon containing *ADS* and *CPR* is driven by the *C. reinhardtii psbA* promoter (*Cr PpsbA*) and either the *T7 Lg10* or the *psbA* leader sequence from *C. reinhardtii* (*Cr LpsbA*). This operon is arranged either in sense and downstream of the first operon (AO1, 3) or in antisense, downstream of the *aadA* cassette (AO2, 4). The genes in each operon are separated by an intercistronic expression element (IEE) conferring intercistronic RNA processing and, in this way, enhancing expression of downstream cistrons of the operon (Zhou et al., 2007; Drechsel and Bock, 2010). The BamHI restriction sites used in RFLP analyses and the expected fragment sizes are indicated. The location of the hybridization probe is shown as a black bar. *Cr*: *C. reinhardtii*; *Nt*: *N. tabacum*; *T7*: bacteriophage T7; *P*: promoter; *L*: leader sequence; *T*: terminator; *SD*: Shine-Dalgarno sequence. (B) RFLP analysis of transplastomic plants. Two independently isolated transplastomic lines are shown for constructs pAO1-3 and one for pAO4. (C) Seed assays confirming the homoplasmic state of the transplastomic plants. Seeds were germinated on medium containing 500 mg/L spectinomycin (*Nt*-AO2-1, *Nt*-AO3-1, *Nt*-wt) or antibiotic-free medium (*Nt*-wt).

DOI: 10.7554/eLife.13664.004

necessary to precisely assess these ratios and ultimately confirm their impact on metabolite conversion in the pathway.

Pathway optimization by combinatorial supertransformation

Having successfully implemented the canonical pathway of artemisinic acid synthesis into the chloroplast of tobacco plants, we next sought to maximize artemisinic acid production. In our best-performing transplastomic plants (*Nt*-AO3), artemisinic acid accumulation reached a maximum of 2–4 mg/kg fresh weight (FW), equivalent to approximately 20–40 mg/kg dry weight (DW; or 0.002–0.004% DW), a level significantly lower than artemisinin accumulation in *A. annua* (varying between 0.01 and 1% DW; Liu et al., 2011; Bryant et al., 2015). Recently, a number of accessory *A. annua* enzymes have been identified that enhance the flux through the pathway, including a cytochrome *b₅* (CYB5) that promotes electron transfer to P450 monooxygenases (Schenkman and Jansson, 2003), a new alcohol dehydrogenase (ADH1) that improves the oxidation of artemisinic alcohol to artemisinic aldehyde (Paddon et al., 2013) and an aldehyde dehydrogenase (ALDH1) that catalyzes the conversion of dihydroartemisinic aldehyde into dihydroartemisinic acid and, in yeast, also enhances the conversion of artemisinic aldehyde into artemisinic acid (Paddon et al., 2013; Figure 1). We also considered two additional enzymes: The double bond reductase 2 (DBR2) from *A. annua* introduces a branch point into the pathway by reducing artemisinic aldehyde to dihydroartemisinic aldehyde (Zhang et al., 2008) and, therefore, potentially can lead to the synthesis of artemisinin (Figure 1). Finally, the 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) from the cyanobacterium *Synechocystis*, a key regulatory enzyme in the MEP pathway of isoprenoid biosynthesis, was selected because its expression may improve precursor availability (Figure 1). Since the quantitative contributions of these enzyme activities to artemisinic acid biosynthesis are not well understood and, moreover, the optimum enzyme activities required to mediate maximum flux through the pathway are unknown, we decided to pursue a combinatorial supertransformation approach. Combinatorial transformation involves the mixing of multiple single-gene transformation constructs and their biolistic co-transformation followed by large-scale screening of many transgenic lines by their metabolic (or other) phenotypes (Zhu et al., 2008; Naqvi et al., 2009). Individual transgenic lines generated by this approach differ in the transgene combination they harbor as well as in transgene copy numbers and expression levels, thus facilitating selection of optimized genotypes that condition the desired metabolic output (Naqvi et al., 2009). We applied combinatorial nuclear transformation to our transplastomic *Nt*-AO2-1 (high accumulation of amorpha-4,11-diene but low levels of artemisinic acid) and *Nt*-AO3-1 (low accumulation of amorpha-4,11-diene and high accumulation of artemisinic acid) lines, assuming that artemisinic acid production can be substantially increased by identifying the optimum combination and expression levels of the additional pathway enzymes. Combinatorial supertransformation of transplastomic lines encoding a canonical metabolic pathway with a plasmid cocktail containing additional and/or accessory pathway enzymes represents a new approach in

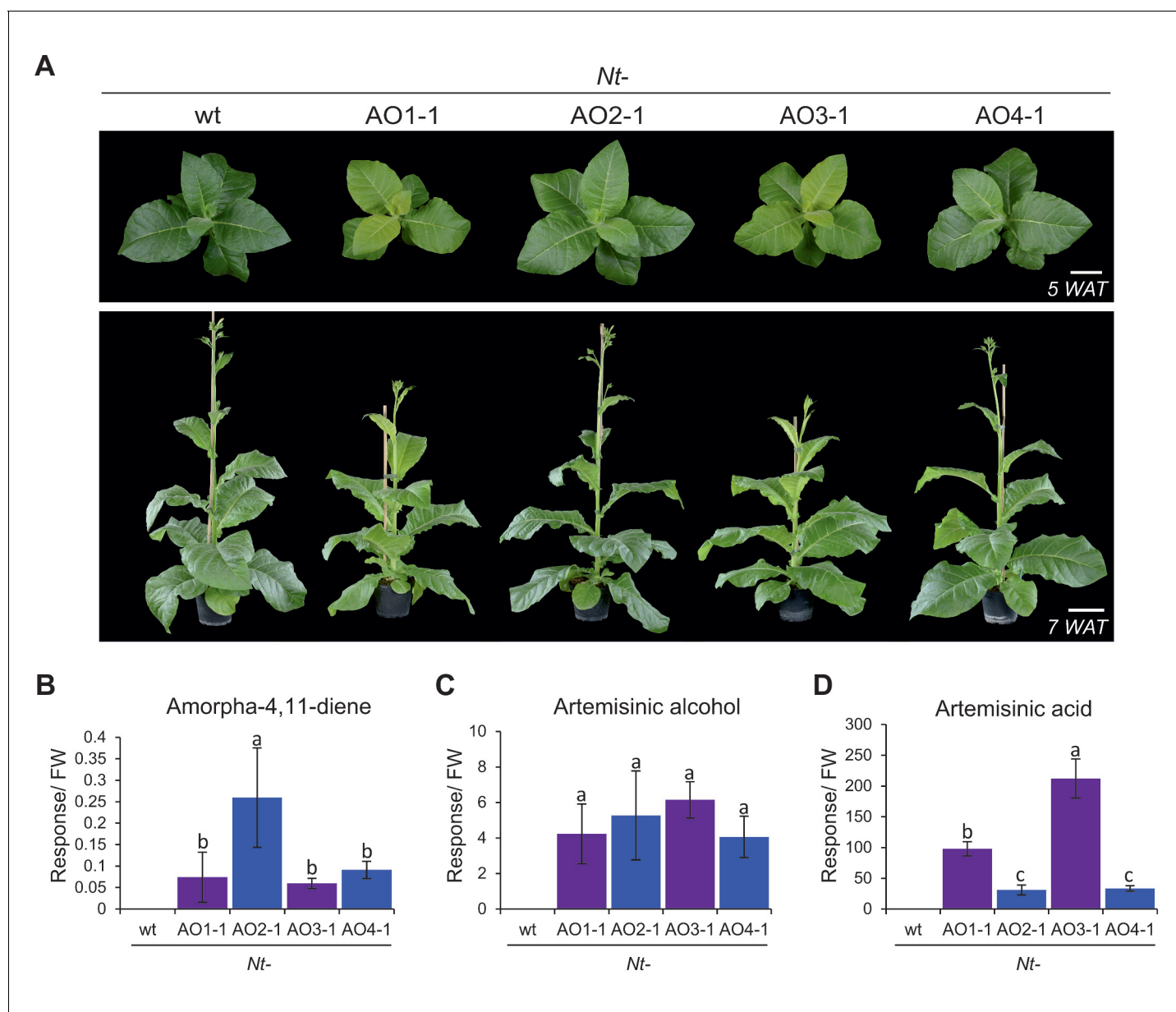


Figure 3. Phenotype of transplastomic tobacco plants and accumulation of artemisinic compounds. (A) Transplastomic lines *Nt*-AO1-1 and *Nt*-AO3-1 display a slightly pale and growth-delayed phenotype at the juvenile stage. WAT: weeks after transfer from tissue culture to soil; scale bars: 10 cm. (B) Amorpha-4,11-diene is synthesized in all transplastomic lines, but accumulates to lower levels in the lines displaying an altered phenotype (purple bars). (C) Artemisinic alcohol is detected in similar amounts in all transplastomic plants. (D) Accumulation of artemisinic acid correlates with the altered phenotype of *Nt*-AO1-1 and *Nt*-AO3-1. Relative accumulation of amorpha-4,11-diene was profiled by GC-MS analysis of volatile organic compounds (VOCs). Relative accumulation of the sum of free and conjugated artemisinic alcohol and artemisinic acid were determined by GC-MS analysis of the soluble metabolite fraction after saponification (see Materials and methods; **Figures 6** and **7**). In agreement with previous reports (*van Herpen et al., 2010*), these compounds were found to be present mainly as conjugates. Expanding leaves of 5–6 plants per line were used for each measurement. Error bars represent the SD. Different letters above the bars indicate significant differences as determined by One-way ANOVA ($p < 0.001$) and the Holm-Sidak post-hoc test.

DOI: [10.7554/eLife.13664.005](https://doi.org/10.7554/eLife.13664.005)

The following figure supplements are available for figure 3:

Figure supplement 1. Phenotypes of *Nt*-AO2-1 and *Nt*-AO3-1 plants throughout development.

DOI: [10.7554/eLife.13664.006](https://doi.org/10.7554/eLife.13664.006)

Figure supplement 2. Isoprenoids levels throughout development in wild-type *Nicotiana tabacum* plants (*Nt*-wt) and the transplastomic lines *Nt*-AO2-1 and *Nt*-AO3-1.

DOI: [10.7554/eLife.13664.007](https://doi.org/10.7554/eLife.13664.007)

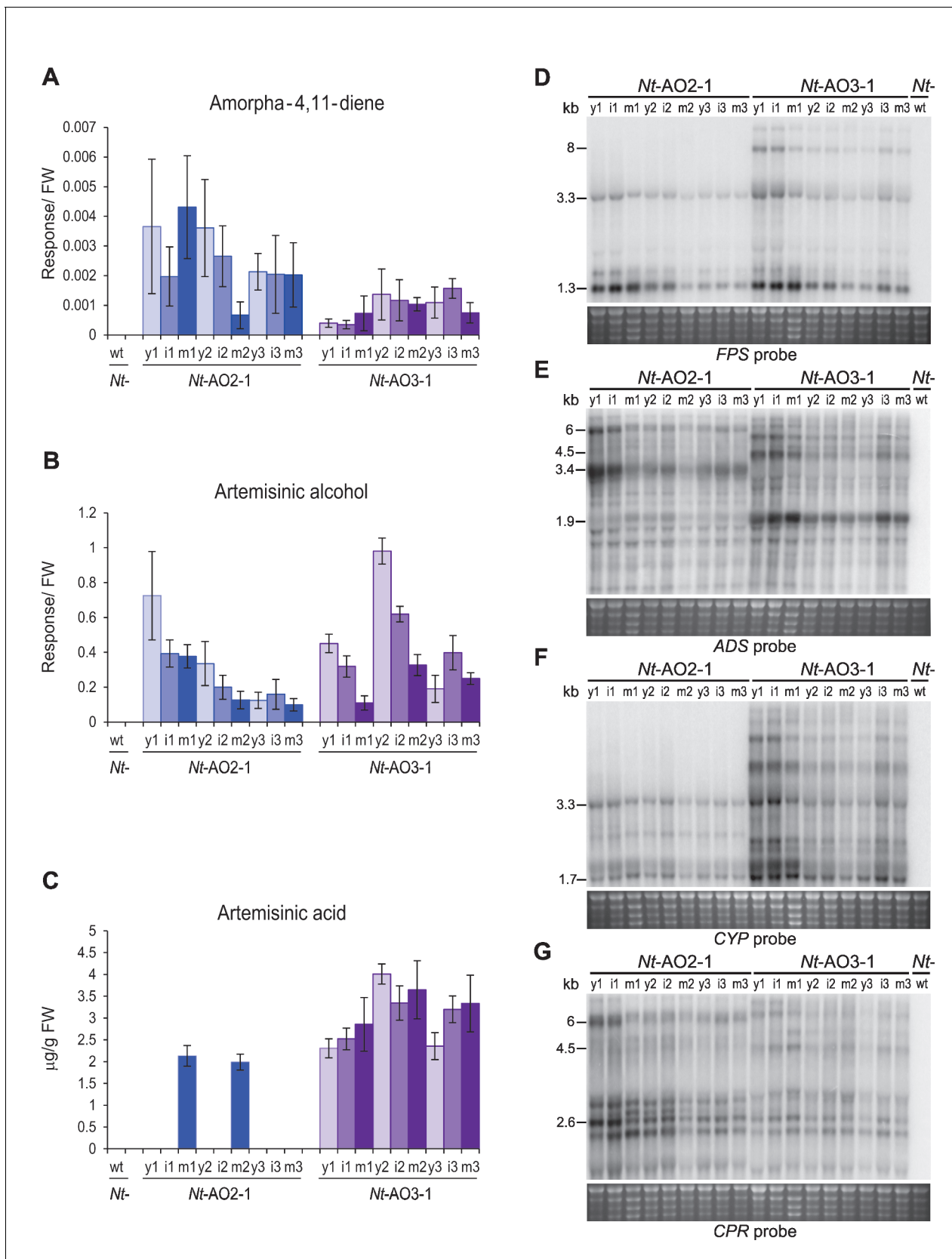


Figure 4. Production of artemisinic acid is maintained throughout plant development and correlates with a high CYP/CPR expression ratio. Artemisinic compounds and expression levels of the transgenes were measured in young (stage 1), flowering (stage 2) and old plants (stage 3; see **Figure 3**—**Figure 4** continued on next page

Figure 4 continued

figure supplement 1. (A) Amorpha-4,11-diene accumulates to higher levels in line *Nt-AO2-1* than in *Nt-AO3-1*. (B) Artemisinic alcohol is present at similar levels in early (1) and late stages (3) of development in lines *Nt-AO2-1* and *Nt-AO3-1*, but it is slightly higher in the flowering stage (2) of line *Nt-AO3-1*. (C) Artemisinic acid accumulates to high levels during all developmental stages of line *Nt-AO3-1*, whereas in line *Nt-AO2-1*, it is detectable only in mature leaves of young and flowering plants. Relative accumulation of amorpha-4,11-diene and artemisinic alcohol was profiled, the tissue content of artemisinic acid was quantified using an authenticated reference standard ($n = 5-6$ plants per line; **Figures 6** and **7**). The sum of free and conjugated artemisinic alcohol and artemisinic acid were determined. y: young leaf; i: expanding (intermediate) leaf; m: fully expanded (mature) leaf. Error bars represent the SD. (D–G) Northern blot analysis of the expression of the four transgenes. Total RNA samples from *N. tabacum* wild-type (*Nt-wt*) plants and the transplastomic lines *Nt-AO2-1* and *Nt-AO3-1* (at the developmental stages 1–3) were separated in denaturing 1.5% agarose gels, blotted and hybridized to strand-specific RNA probes. Below each blot, the rRNA-containing region of the ethidium bromide-stained gel prior to blotting is shown as a control for RNA integrity and equal loading. The *Nt-wt* sample corresponds to RNA extracted from a fully expanded leaf of a *N. tabacum* wild-type plant at developmental stage 2. The smallest labeled band in each blot corresponds to the monocistronic mRNA. Larger bands represent unprocessed polycistronic precursor transcripts and read-through transcripts (which are common in plastids; e.g., **Elghabi et al., 2011**; **Lu et al., 2013**). *CYP* transcripts accumulate to higher levels in line *Nt-AO3-1*, while *CPR* transcripts accumulate to higher levels in line *Nt-AO2-1*, resulting in a higher *CYP/CPR* expression ratio in line *Nt-AO3-1*.

DOI: [10.7554/eLife.13664.008](https://doi.org/10.7554/eLife.13664.008)

The following figure supplement is available for figure 4:

Figure supplement 1. Quantitation of the expression of *CYP* and *CPR* by qRT-PCR analysis.

DOI: [10.7554/eLife.13664.009](https://doi.org/10.7554/eLife.13664.009)

synthetic biology that we refer to as COSTREL (for COmbinatorial Supertransformation of Transplastomic REcipient Lines).

Genes for the five candidate enzymes (CYB5, ADH1, ALDH1, DBR2, DXR) were cloned into individual expression cassettes, the resulting plasmids were mixed and co-bombarded with a kanamycin resistance gene into the nuclear genomes of transplastomic *Nt-AO2-1* and *Nt-AO3-1* plants. 612 kanamycin-resistant shoots (*Nt-AO-CS* lines) were generated by supertransformation of the transplastomic recipient lines *Nt-AO2-1* and *Nt-AO3-1*. After rooting in kanamycin-containing medium, 512 plantlets were transferred to soil and grown to maturity under standard greenhouse conditions. At the onset of flowering, a fully expanded leaf was harvested for preliminary profiling of artemisinic acid and its precursors by GC-MS (see Materials and Methods). Based on growth, phenotype and fertility of the plants, 199 COSTREL lines were selected for metabolic screening of artemisinic compounds: 79 *Nt-AO2-CS* and 120 *Nt-AO3-CS* lines (**Figure 5—source data 1**). The various lines displayed great variation with respect to the accumulation levels of the compounds assayed (amorpha-4,11-diene, artemisinic alcohol, dihydroartemisinic alcohol, dihydroartemisinic acid and artemisinic acid). Importantly, COSTREL lines could be identified that contained strongly increased levels of the drug precursor artemisinic acid (**Figure 5—source data 1**).

In combinatorial transformation, all transgenes that simultaneously enter the nucleus of the recipient cell usually integrate into the same genomic locus (most likely into a transient DNA double-strand break), and therefore co-segregate into the next generation (**Naqvi et al., 2009**). This feature allowed us to raise a T1 generation of supertransformed lines from seeds and repeat the metabolite profiling with T1 leaf material grown under highly standardized conditions. These analyses confirmed the results obtained with the T0 plants and revealed that, in the case of the *Nt-AO2-CS* lines, the highest increase in artemisinic acid content occurred in line 132 showing a 33-fold increase compared to its transplastomic recipient line *Nt-AO2-1*, whereas among the *Nt-AO3-CS* lines, line 180 reached an even 77-fold increase compared to transplastomic line *Nt-AO3-1* (**Figure 5—source data 1**; **Figure 5A**). The trait artemisinic acid content was stable across generations, and the highest producer, line *Nt-AO3-CS180*, reached levels of 120.4 ± 42 mg per kg FW in the T1 generation.

Identification of limiting steps in artemisinic acid biosynthesis

To obtain insights into pathway regulation and identify bottlenecks in artemisinic acid synthesis, we investigated correlations between pathway metabolites and between artemisinic acid accumulation and the set of transgenes expressed in the nucleus of *Nt-AO2-CS* and *Nt-AO3-CS* COSTREL lines.

Increased amounts of artemisinic acid in the *Nt-AO-CS* lines were negatively correlated with the accumulation of artemisinic alcohol (**Figure 5A,B**), indicating that the efficiency of oxidation of the alcohol represents a key bottleneck in the pathway that we alleviated by supertransformation with the additional pathway genes. Importantly, artemisinic alcohol was reduced to nearly undetectable

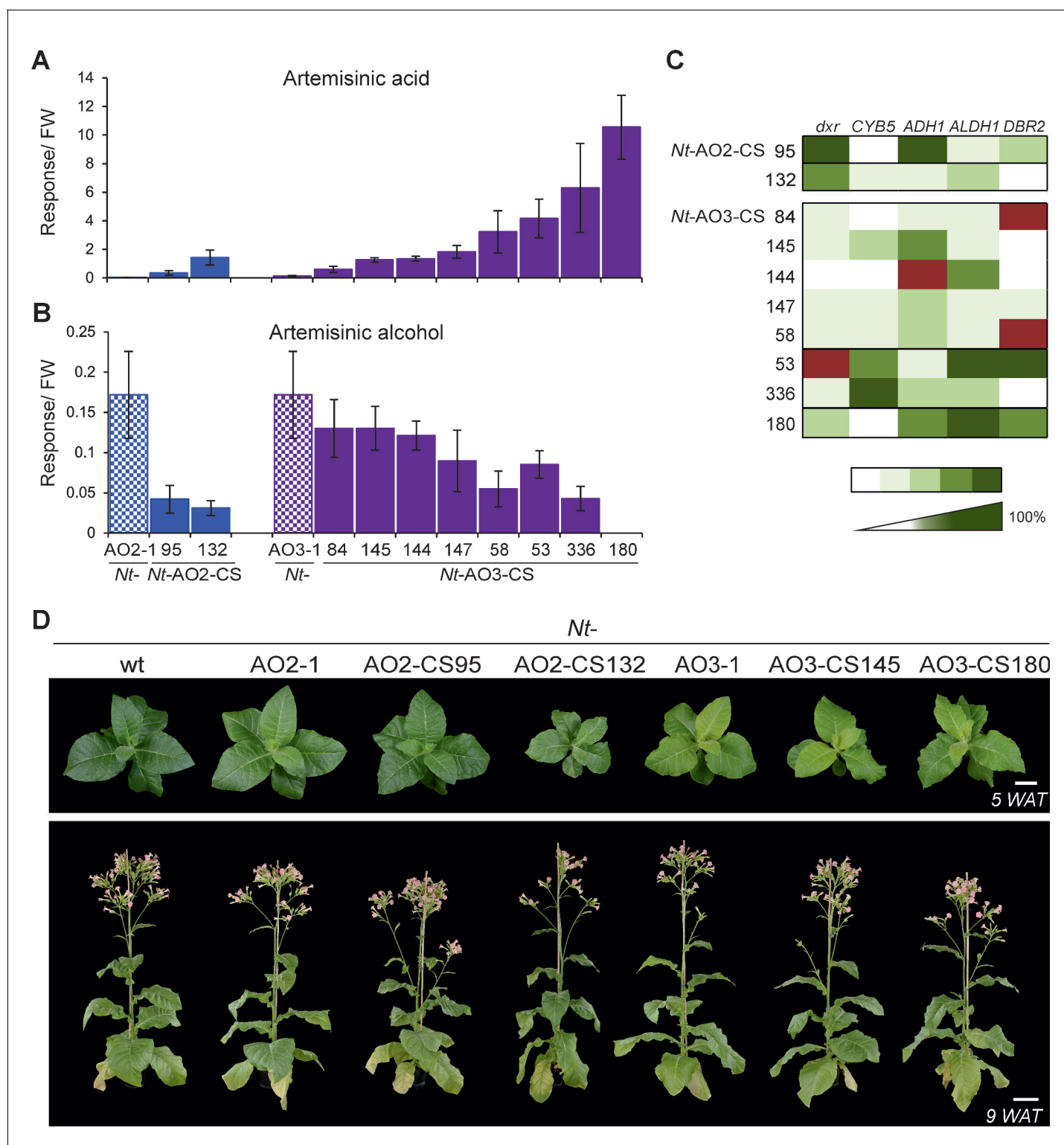


Figure 5. Isolation of combinatorially supertransformed transplastomic lines with a strong increase in artemisinic acid accumulation. Transplastomic lines *Nt*-AO2-1 and *Nt*-AO3-1 were combinatorially supertransformed with genes for additional enzymes of the pathway (Figure 1) to facilitate large-scale screening for increased artemisinic acid production. (A) Relative artemisinic acid levels (given in response/FW; see Materials and methods) in the T1 generation of two combinatorially supertransformed lines obtained with transplastomic recipient line *Nt*-AO2-1 (*Nt*-AO2-CS) and eight lines obtained with transplastomic recipient line *Nt*-AO3-1 (*Nt*-AO3-CS). An up to 77-fold increase in artemisinic acid was achieved (line *Nt*-AO3-CS180) in comparison to recipient line *Nt*-AO3-1. For a complete list of screened supertransformed lines, see Figure 5—source data 1. (B) Inverse relationship

Figure 5 continued on next page

Figure 5 continued

between artemisinic acid accumulation and artemisinic alcohol accumulation in supertransformed lines. Fully expanded leaves of 5–6 plants per line (at the flowering stage) were used for metabolite profiling. The sum of free and conjugated artemisinic alcohol and artemisinic acid were determined. (C) qRT-PCR analysis of transgene expression suggests a predominant role of *ALDH1* in boosting artemisinic acid synthesis. 2–3 plants per line were measured, and the expression levels were ranked after One-way ANOVA comparison ($p < 0.05$). Brown color indicates absence of gene expression. (D) Combinatorially supertransformed lines with a high increase in artemisinic acid (*Nt-AO2-CS132* and *Nt-AO3-CS180*) display a similar phenotype as the corresponding transplastomic recipient line. WAT: weeks after transfer from tissue culture to soil; scale bars: 10 cm.

DOI: [10.7554/eLife.13664.010](https://doi.org/10.7554/eLife.13664.010)

The following source data and figure supplements are available for figure 5:

Source data 1. Metabolic screening (phenotyping) and genotyping of the T0 generation of combinatorially supertransformed *Nt-AO-CS* lines.

DOI: [10.7554/eLife.13664.011](https://doi.org/10.7554/eLife.13664.011)

Figure supplement 1. Amorpha-4,11-diene, dihydroartemisinic alcohol and dihydroartemisinic acid accumulation in the T1 generation of combinatorially supertransformed plants.

DOI: [10.7554/eLife.13664.012](https://doi.org/10.7554/eLife.13664.012)

Figure supplement 2. Measurements of plant height and total leaf biomass of COSTREL line *Nt-AO3-CS180* (progeny of four different T1 lines), its transplastomic progenitor line *Nt-AO3-1* and the wild type (wt).

DOI: [10.7554/eLife.13664.013](https://doi.org/10.7554/eLife.13664.013)

levels in the best-performing line *Nt-AO3-CS180* suggesting that maximum conversion efficiency has been achieved (**Figure 5B**; **Figure 5—source data 1**). Another significant correlation at the metabolite level was a strong positive correlation between amorpha-4,11-diene and artemisinic alcohol (**Tables 1** and **2**) which may be a consequence of the enzymatic limitation downstream of artemisinic alcohol. By contrast, amorpha-4,11-diene was not significantly correlated with artemisinic acid accumulation. Dihydroartemisinic alcohol (presumably generated by an endogenous enzymatic activity in tobacco; **Ting et al., 2013**; **Zhang et al., 2011**), while accumulating in some lines, showed only a weak positive correlation with artemisinic acid in the *Nt-AO2-CS* but not in the *Nt-AO3-CS* lines (**Figure 5—figure supplement 1**; **Tables 1** and **2**; **Figures 6** and **7**). Dihydroartemisinic acid, the direct precursor of artemisinin, was detected only in line *Nt-AO3-CS180* in the T0 generation and in lines *Nt-AO3-CS53* and *Nt-AO3-CS180* in the T1 generation (**Figure 5—source data 1** and **Figure 5—figure supplement 1**; **Figure 7C**), and therefore had to be excluded from the correlation analysis.

To correlate the genotype of the COSTREL lines with their metabolic phenotypes, the transgene sets present in the nucleus of *Nt-AO2-CS* and *Nt-AO3-CS* lines were determined (**Figure 5—source data 1**). Statistical analysis revealed that elevated artemisinic acid levels in *Nt-AO3-CS* lines were most strongly correlated with the presence of the *ALDH1* transgene. Weaker (and statistically not significant) correlations were observed between high artemisinic acid levels and the presence of *dxr* and *ADH1*, and, in *Nt-AO2-CS* lines, also the *DBR2* and *ALDH1* transgenes (**Tables 1** and **2**). These results indicate that *ALDH1* and *dxr* are most likely the genes with the greatest impact on the increase in artemisinic acid content. As transgene presence is not necessarily indicative of transgene expression, we measured mRNA accumulation in the T1 generation of a selected set of COSTREL lines by qRT-PCR analyses. The results support the importance of *dxr*, *ADH1*, *ALDH1* and *DBR2* in boosting artemisinic acid synthesis and revealed that the best-performing line (*Nt-AO3-CS180*) expresses *dxr*, *ADH1*, *ALDH1* and *DBR2* to high levels (**Figure 5C**).

To test whether artemisinic acid accumulation is correlated with a visible phenotype of the plants, the best-performing COSTREL lines were compared with their transplastomic recipient lines. No significant phenotypic differences were observed and even line *Nt-AO3-CS180* (that showed the strongest increase in artemisinic acid accumulation; **Figure 5A**) was nearly indistinguishable from its transplastomic recipient *Nt-AO3-1* (**Figure 5D**). This suggests that artemisinic acid is not toxic to plant cells (and that a further increase in artemisinic acid might be achievable). Growth and biomass measurements confirmed that there are no significant differences between transplastomic line *Nt-AO3-1* and the best-performing COSTREL line *Nt-AO3-CS180*, and revealed only a small reduction in total leaf biomass (by on average 13%) of the COSTREL line compared to the wild type (**Figure 5—figure supplement 2**).

None of our transplastomic lines and none of the analyzed COSTREL lines accumulated detectable levels of artemisinin (see Materials and Methods). This could be because the set of transgenes introduced into tobacco was insufficient to obtain conversion of artemisinic acid into artemisinin. However, some of our best-performing COSTREL lines accumulated detectable amounts of

Table 1. Correlation analysis of artemisinic compounds and transgenes introduced into transplastomic line *Nt-AO2-1* by combinatorial supertransformation. The levels of the artemisinic compounds amorpho-4,11-diene, artemisinic alcohol, dihydroartemisinic alcohol and artemisinic acid, and the presence of the transgenes *dxr*, *CYB5*, *ADH1*, *ALDH1* and *DBR2* were correlated using Spearman's method in the 39 *Nt-AO2-CS* lines analyzed by genomic PCR in the T0 generation (see **Figure 5—source data 1**) using the SPSS software. Dihydroartemisinic acid was excluded from this analysis, because it was not detectable in any of the *Nt-AO2-CS* lines. CC: correlation coefficient. Positive values indicate positive correlations and negative values indicate negative correlations. *: $p < 0.05$; **: $p < 0.01$; N: number of samples where both variables are present.

<i>Nt-AO2-CS</i>	Amorpha-4,11-diene	Artemisinic alcohol	Dihydroartemisinic alcohol	Artemisinic acid	<i>dxr</i>	<i>CYB5</i>	<i>ADH1</i>	<i>ALDH1</i>	<i>DBR2</i>
Amorpha-4,11-diene	CC 1.000	0.602**	0.147	-0.112	0.067	-0.020	-0.009	-0.015	0.156
	N 35	29	13	24	35	35	35	35	35
Art. alcohol	CC 0.602**	1.000	-0.557*	-0.141	-0.159	-0.165	-0.170	0.013	-0.217
	N 29	29	13	24	29	29	29	29	29
Dihydroart. alcohol	CC 0.147	-0.557*	1.000	0.642*	0.462	0.248	0.383	0.180	0.496
	N 13	13	13	13	13	13	13	13	13
Art. acid	CC -0.112	-0.141	0.642*	1.000	0.404	0.159	0.179	0.207	0.317
	N 24	24	13	24	24	24	24	24	24
<i>dxr</i>	CC 0.067	-0.159	0.462	0.404	1.000	0.643**	0.402*	0.546**	0.578**
	N 35	29	13	24	39	39	39	39	39
<i>CYB5</i>	CC -0.020	-0.165	0.248	0.159	0.643**	1.000	0.192	0.507**	0.793**
	N 35	29	13	24	39	39	39	39	39
<i>ADH1</i>	CC -0.009	-0.170	0.383	0.179	0.402*	0.192	1.000	0.372*	0.270
	N 35	29	13	24	39	39	39	39	39
<i>ALDH1</i>	CC -0.015	0.013	0.180	0.207	0.546**	0.507**	0.372*	1.000	0.420**
	N 35	29	13	24	39	39	39	39	39
<i>DBR2</i>	CC 0.156	-0.217	0.496	0.317	0.578**	0.793**	0.270	0.420**	1.000
	N 35	29	13	24	39	39	39	39	39

DOI: 10.7554/eLife.13664.014

dihydroartemisinic acid (**Figure 5—source data 1** and **Figure 5—figure supplement 1**; **Figure 7**), the immediate precursor of artemisinin, indicating that *DBR2* and *ALDH1* can function in plastids. The chloroplast is likely to produce sufficient amounts of singlet oxygen (which is a regular by-product of photosynthetic electron transfer) to facilitate the spontaneous conversion of dihydroartemisinic acid into artemisinin (**Kopetzki et al., 2013**; **Figure 1**). An alternative explanation for the lack of artemisinin accumulation could be that *COSTREL* lines that produce artemisinin were not recovered, because artemisinin is highly toxic to photosynthetically active cells (**Bharati et al., 2012**). The fact that no artemisinin could be detected and only low amounts of dihydroartemisinic acid were obtained in a few lines, whereas artemisinic acid accumulated to high amounts, may indicate that future efforts should be focused on maximizing the production of artemisinic acid.

Discussion

In the course of this work, we have developed a new synthetic biology approach that combines chloroplast transformation with combinatorial nuclear transformation and large-scale metabolic screening of supertransformed plant lines. This strategy enabled the transfer of an entire biochemical pathway of secondary metabolism from a medicinal plant to a high-biomass crop.

For the foreseeable future, ACTs will remain the most powerful weapon in the world's battle against malaria (<http://www.who.int/malaria/areas/treatment/overview/en/>). When used as an oral monotherapy, artemisinin can promote the development of resistance in the parasite (**Mok et al., 2015**; **Straimer et al., 2015**; **Mbengue et al., 2015**) and, therefore, ACTs are based on fixed-dose co-formulations that combine two different active ingredients in one tablet. Development of an

Table 2. Correlation analysis of artemisinic compounds and transgenes introduced into transplastomic line *Nt-AO3-1* by combinatorial supertransformation. The levels of the artemisinic compounds amorpho-4,11-diene, artemisinic alcohol, dihydroartemisinic alcohol and artemisinic acid, and the presence of the transgenes *dxr*, *CYB5*, *ADH1*, *ALDH1* and *DBR2* were correlated using Spearman's method in the 61 *Nt-AO3-CS* lines analyzed by genomic PCR in the T0 generation (see **Figure 5—source data 1**) using the SPSS software. Dihydroartemisinic acid had to be excluded from this analysis, because it was detectable only in one of the *Nt-AO3-CS* lines in the T0 generation. Note that the negative correlation between artemisinic alcohol and artemisinic acid (**Figure 5A,B**) is restricted to those lines that display increased artemisinic acid contents, and therefore is not statistically significant over all COSTREL lines analyzed (cf. **Figure 5—source data 1**). CC: correlation coefficient. Positive values indicate positive correlations and negative values indicate negative correlations. *: $p < 0.05$; **: $p < 0.01$; N: number of samples where both variables are present.

<i>Nt-AO3-CS</i>	Amorpho-4,11-diene	Artemisinic alcohol	Dihydroartemisinic alcohol	Artemisinic acid	<i>dxr</i>	<i>CYB5</i>	<i>ADH1</i>	<i>ALDH1</i>	<i>DBR2</i>
Amorpho-4,11-diene	CC 1.000	0.363**	0.079	0.237	0.006	0.082	0.439**	-0.032	0.049
	N 60	58	39	59	60	60	60	60	60
Art. alcohol	CC 0.363**	1.000	0.124	-0.081	-0.088	-0.119	0.086	0.026	-0.009
	N 58	59	39	59	59	59	59	59	59
Dihydroart. alcohol	CC 0.079	0.124	1.000	0.029	-0.169	0.153	0.306	0.195	0.219
	N 39	39	39	39	39	39	39	39	39
Art. acid	CC 0.237	-0.081	0.029	1.000	0.241	0.183	0.217	0.386**	0.141
	N 59	59	39	60	60	60	60	60	60
<i>dxr</i>	CC 0.006	-0.088	-0.169	0.241	1.000	0.153	0.224	0.415**	0.170
	N 60	59	39	60	61	61	61	61	61
<i>CYB5</i>	CC 0.082	-0.119	0.153	0.183	0.153	1.000	0.185	0.322*	0.503**
	N 60	59	39	60	61	61	61	61	61
<i>ADH1</i>	CC 0.439**	0.086	0.306	0.217	0.224	0.185	1.000	0.339**	0.252*
	N 60	59	39	60	61	61	61	61	61
<i>ALDH1</i>	CC -0.032	0.026	0.195	0.386**	0.415**	0.322*	0.339**	1.000	0.444**
	N 60	59	39	60	61	61	61	61	61
<i>DBR2</i>	CC 0.049	-0.009	0.219	0.141	0.170	0.503**	0.252*	0.444**	1.000
	N 60	59	39	60	61	61	61	61	61

DOI: 10.7554/eLife.13664.015

inexpensive and sustainable production method that is suitable to meet the constantly growing demand for artemisinin and its derivatives has remained a grand challenge. Enormous breeding efforts are currently underway to produce new varieties of *A. annua* that accumulate higher and more consistent levels of the compound (Graham et al., 2010). However, as *A. annua* produces artemisinin only in a very small fraction of the leaf cells (the glandular trichomes) and its cultivation is inefficient, slow and vulnerable to adverse environmental conditions, the development of a production method that is independent of *A. annua* is highly desirable (Bryant et al., 2015). If accomplished in a high-biomass non-food/non-feed crop, this would provide a stable supply of the feedstock that can be scaled up at will and at short notice, and take full advantage of the existing agricultural infrastructure. In the course of this work, we have established tobacco as an efficient production factory for artemisinic acid. Tobacco is a high-biomass crop, grown in large acreages, for which alternative uses (that are unrelated to smoking) have long been sought. Since tobacco is well suited for cultivation at high cropping densities and multiple harvests (4–5) per season are possible, 40 t of biomass can be obtained from a single acre of tobacco field at a cost of only around \$100 per ton (<http://tobacco.ces.ncsu.edu/wp-content/uploads/2012/07/tobacco-production-cost-2011-1.pdf?fw=no>). Thus, with our best-performing COSTREL line, production levels of ~4.8 kg artemisinic acid per acre can be obtained, suggesting that the current world demand (of ~100 t artemisinin) can be met by cultivating tobacco on an area of ~200 km², which is less than the area of the city of Boston (assuming ~50% loss during extraction and conversion of artemisinic acid to artemisinin; Paddon et al., 2013; Kopetzki et al., 2013).

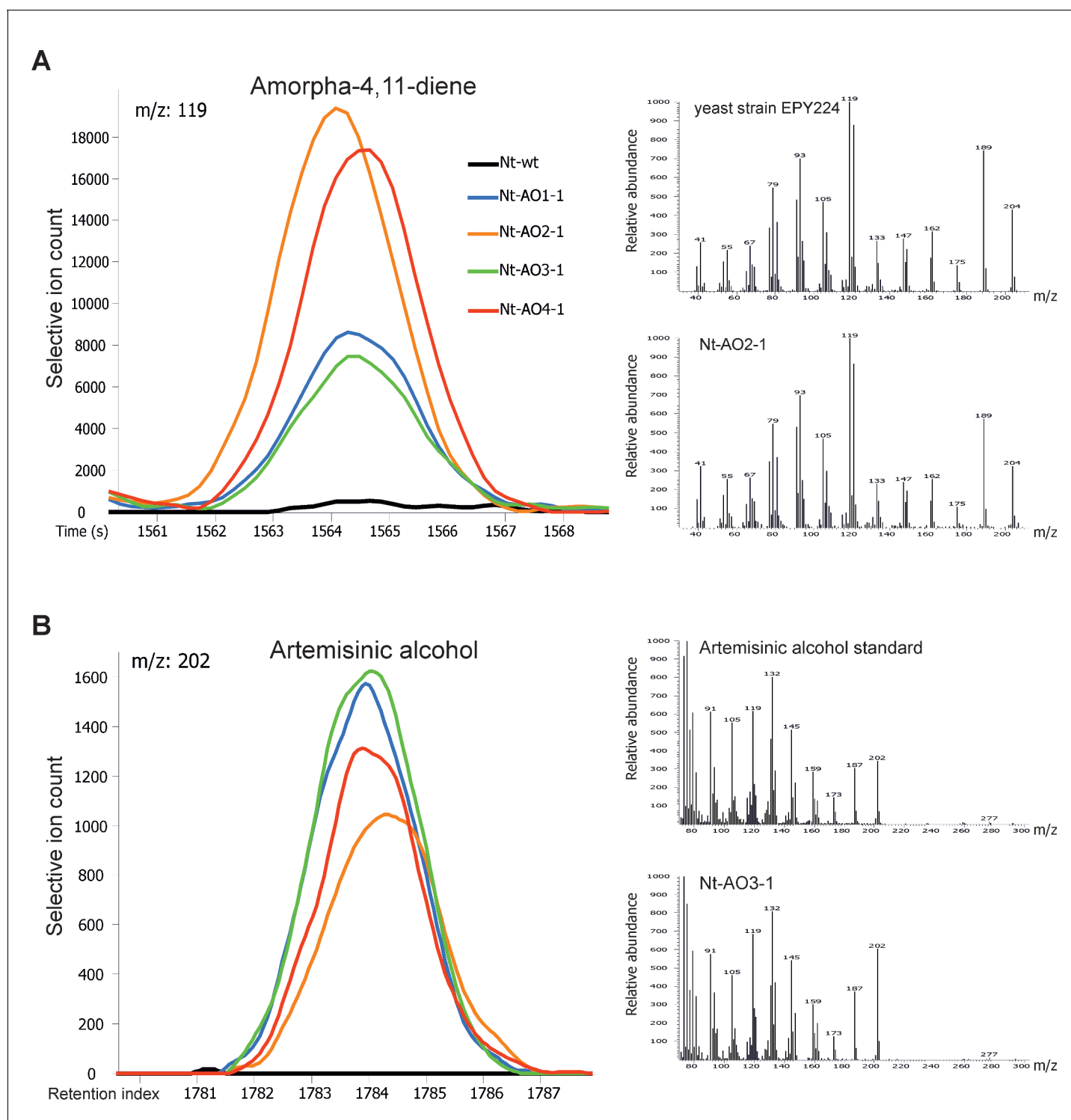


Figure 6. Chromatograms and mass spectra of amorpha-4,11-diene and artemisinic alcohol. Characteristic peaks for one specific fragment at the expected retention time or index are displayed for each compound. (A) Amorpha-4,11-diene-specific mass feature 119 at a retention time of 1564 s. This metabolite is present in all *Nt-AO* lines, and at slightly higher levels in lines *Nt-AO2-1* and *Nt-AO4-1*. (B) Artemisinic alcohol-specific mass feature 202 at a retention index of 1784. The compound is present at similar levels in all *Nt-AO* lines. Both compounds are absent from the wild-type sample. In addition to the chromatograms, the characteristic mass spectrum (m/z) of each compound is shown for the standard and for one of the artemisinic acid operon lines. EPY224: yeast strain that produces amorpha-4,11-diene (Ro et al., 2006). One representative plant per line is depicted. Mass spectra and mass features of trimethylsilylated artemisinic alcohol are shown.

DOI: 10.7554/eLife.13664.016

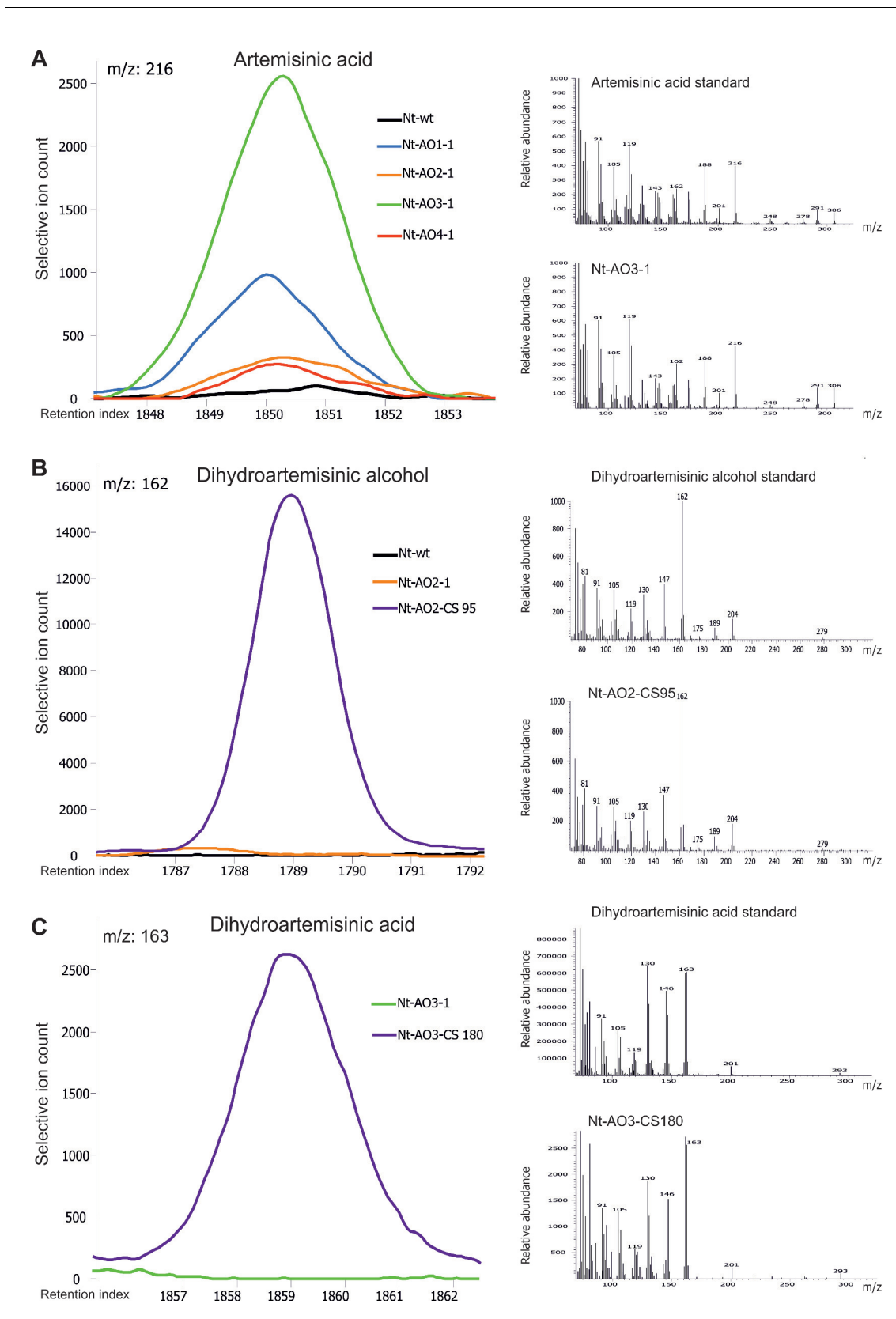


Figure 7. Chromatograms and mass spectra of the artemisinic compounds artemisinic acid, dihydroartemisinic alcohol and dihydroartemisinic acid. Characteristic peaks for one specific fragment at the expected retention index are shown for each compound. (A) Artemisinic acid-specific mass feature
Figure 7 continued on next page

Figure 7 continued

216 is shown at a retention index of 1850. This compound accumulates to higher levels in lines *Nt*-AO1-1 and *Nt*-AO3-1. (B) Dihydroartemisinic alcohol-specific mass feature 162 at a retention index of 1789. The compound is present at high levels in COSTREL line *Nt*-AO2-CS95, but is absent from transplastomic line *Nt*-AO2-1. (C) Dihydroartemisinic acid-specific mass feature 163 at a retention index of 1859. This compound accumulates in COSTREL line *Nt*-AO3-CS180, but is absent from transplastomic line *Nt*-AO3-1. All compounds are absent from the wild-type sample. In addition to the chromatograms, the characteristic mass spectrum of each compound is shown for the standard and for one of the artemisinic acid operon lines. Mass spectra and mass features of trimethylsilylated artemisinic compounds are shown.

DOI: 10.7554/eLife.13664.017

Whereas in *A. annua* the artemisinin biosynthetic pathway is confined to glandular trichomes, our COSTREL tobacco lines produce artemisinic acid in chloroplasts and, thus, in the whole leaf. Together with the absence of toxic effects of artemisinic acid on the chloroplast (**Figure 5D**; **Figure 5—figure supplement 2**), this offers great potential for further enhancement of the pathway by addressing the bottlenecks that limit flux in our current best-performing lines. Furthermore, previous transgenic work has shown that the redox environment in the cytosol of tobacco cells favors reduction of aldehydes to alcohols rather than their oxidation to acids, thus limiting the ability of the cytosolically located pathway to produce high quantities of artemisinic acid (**Zhang et al., 2011**). The high levels of artemisinic acid achieved in this work by implementing the pathway into plastids suggest that the chloroplast offers a more favorable redox milieu that allows the quantitative conversion of artemisinic alcohol into artemisinic acid (**Figure 5A,B**). Although tobacco leaves also possess glandular trichomes (where artemisinin is produced in *A. annua*), the trichomes in our COSTREL plants are unlikely to accumulate large amounts of artemisinic acid. This is because transgene expression from the plastid genome is generally very low in non-photosynthetic tissues and cell types. It can be significantly enhanced by designing specific (chimeric) expression signals that confer high transgene activity in non-green tissues (**Zhang et al., 2012**; **Caroca et al., 2013**), but the expression signals used to drive our synthetic artemisinic acid operons (**Figure 2**) are not suitable to trigger efficient gene expression in non-photosynthetic plastids.

The chloroplast represents an attractive site for engineering new metabolic pathways into plants. Being the biosynthetic center of the plant cell, the chloroplast contains large pools of diverse metabolites that can be tapped. Expression of genes for metabolic enzymes from the plastid genome has a number of attractions, including high expression levels, simple stacking of multiple transgenes in synthetic operons (**Lu et al., 2013**; **Gnanasekaran et al., 2016**) and high-precision engineering via homologous recombination (**Maliga, 2004**; **Bock, 2015**). Previously, plastid transformation was employed to enhance endogenous metabolic pathways (**Apel et al., 2009**; **Lu et al., 2013**) or to produce novel metabolites, such as ketocarotenoids and biopolymers (**Hasunuma et al., 2008**; **Bohmert-Tatarev et al., 2011**). Recently, two ER-resident cytochrome P450 enzymes of the dhurrin pathway (a cyanogenic glucoside from sorghum) were successfully expressed from a synthetic operon in tobacco chloroplasts (**Gnanasekaran et al., 2016**). Together with the third pathway enzyme, a glucosyltransferase, the two P450 enzymes catalyzed the formation of dhurrin from tyrosine. The activity of the P450 enzymes was strictly light-dependent, indicating that the electrons used come from the photosynthetic electron transport chain (**Gnanasekaran et al., 2016**). This suggests that, at least when P450 enzymes are anchored to the thylakoid membrane, reduced ferredoxin can replace the NADPH-dependent native reductase (**Gnanasekaran et al., 2016**), thus making the chloroplast a superb compartment for the implementation of secondary metabolic pathways that involve P450-catalyzed reactions.

By transplastomic introduction of the core pathway for artemisinic acid synthesis, our COSTREL approach takes advantage of the stability and high efficiency of transgene expression from the plastid genome (**Maliga, 2004**; **Bock, 2015**). Subsequent combinatorial supertransformation of the nuclear genome with genes for auxiliary and regulatory factors then allows fine-tuning of the pathway and optimization of metabolic flux by screening metabolic phenotypes of hundreds of transgenic lines that differ in the set of transgenes they harbor in the genome and the expression levels of the transgenes (**Zhu et al., 2008**; **Naqvi et al., 2009**; **2010**). Importantly, this approach requires no prior knowledge about the contributions of the individual factors to metabolic flux and the optimum expression strength of each transgene. Previous metabolic engineering work in microorganisms has demonstrated that the success is often more dependent on achieving the optimum balance

of enzyme activities than on the absolute levels of enzyme (over)expression (e.g., *Peralta-Yahya et al., 2012*). The use of combinatorial supertransformation, therefore, provides a significant advantage over the construction of large transformation vectors expressing multiple pathway genes, because the great variation between transgenic events in (i) the transgene combination present, (ii) the copy numbers of the individual transgenes and (iii) the absolute and relative expression strengths of the transgenes (depending, e.g., on the integration site in the genome and the structure of the transgenic locus) is likely to yield at least some events that harbor the optimum combination of transgenes and provide the right balance of enzyme activities. Moreover, the characterization of these elite events can provide valuable information about pathway regulation, limiting steps and bottlenecks that should be the target of future engineering and optimization efforts. In sum, our COSTREL strategy provides a new synthetic biology tool that facilitates the efficient transfer of complex metabolic pathways into new host organisms while, at the same time, maximizing the metabolic output.

Materials and methods

Plant material and growth conditions

Tobacco plants (*Nicotiana tabacum* cv. Petit Havana) were grown under sterile conditions on agar-solidified MS medium (*Murashige and Skoog, 1962*) supplemented with 30 g/L sucrose. Genetically modified plants were selected, propagated and rooted in the same medium containing additionally 500 mg/L spectinomycin (transplastomic plants) or 50 mg/L kanamycin (combinatorially supertransformed plants). For sampling and seed production, plants were transferred to soil and grown under standard greenhouse conditions.

Construction of transformation vectors

The synthetic operon constructs for chloroplast transformation (pAO1-4) are based on plastid transformation vector pKP9 (*Zhou et al., 2008*). They all contain the four genes required for the canonical artemisinic acid biosynthetic pathway in *Artemisia annua*: *FPS* (AF112881), *ADS* (AF138959), *CYP71AV1* (*CYP*, DQ268763) and *CPR* (DQ318192; **Figure 1**). The genes were codon optimized for expression in the chloroplast and chemically synthesized (GeneArt, Regensburg, Germany). The four genes were then assembled into synthetic operons as follows. The *CYP71AV1* (*CYP*) gene was synthesized with a Shine-Dalgarno (SD) sequence derived from the chloroplast *rbcl* gene and with the flanking restriction sites *NheI* (at the 5' end) and *XbaI* (at the 3' end). The gene was cloned into pZF1 replacing the *P24* gene (*Zhou et al., 2008*) and generating construct pZF83. pZF1 is an intermediate cloning construct that contains the promoter from the rRNA operon from tobacco (*Prrn*), the leader sequence from the *gene 10* of bacteriophage T7 (*T7 Lg10*), the *P24* capsid protein gene of HIV-1 and the terminator of the chloroplast *rbcl* gene (*TrbcL*; *Zhou et al., 2008*). A fragment containing the rRNA operon promoter from *Chlamydomonas reinhardtii* (*Cr Prrn*), the *T7 Lg10*, the *gfp* gene, the terminator of the *atpA* gene from the chloroplast genome of *C. reinhardtii* (*Cr TatpA*) and the intercistronic expression element (IEE; *Zhou et al., 2007*) was excised with *SacI* and *NheI* from a modified version of construct pDK139 in which the *Clal*, *Sall* and *XhoI* restriction sites between *Cr TatpA* and IEE were removed by *XhoI*/*HindIII* digestion and blunting of the overhanging ends by a fill-in reaction with Klenow enzyme. pDK139 is a chloroplast transformation construct based on vector pHK20 (*Kuroda and Maliga, 2001*). The excised fragment was cloned into pZF83, replacing the region spanning *Prrn* and *T7 Lg10* and generating construct pZF84. Next, the *FPS* gene was synthesized flanked by *NdeI* and *PacI* restriction sites at the 5' and 3' ends, respectively. The excised *NdeI*/*PacI* restriction fragment was cloned into the identically digested pZF84, replacing the *gfp* gene and giving rise to plasmid pZF85. The complete fragment from *Cr Prrn* to *TrbcL* was then cut out from pZF85 with *SacI* and *Clal* and ligated into chloroplast transformation vector pKP9 (*Zhou et al., 2008*), producing clone pZF90. The *ADS* gene was synthesized (flanked by *NcoI* and *EcoRV* restriction sites) and cloned into vector pKCZaphA-6, replacing the *aphA-6* gene and giving rise to plasmid pZF86. pKCZaphA-6 (*Fleischmann et al., 2011*) is an intermediate cloning construct that contains the *C. reinhardtii psbA* promoter (*Cr PpsbA*), the *C. reinhardtii psbA* leader (*Cr LpsbA*), the *aphA-6* gene for kanamycin resistance and the *C. reinhardtii rbcl* terminator (*Cr TrbcL*). Next, the terminator of the tobacco *rps16* gene (*Trps16*) was amplified by PCR with primers containing *EcoRV* and *PstI* restriction sites at the 5' and 3' ends, respectively, and cloned into pZF86 digested with the same

enzymes, generating vector pZF87. The *CPR* gene was synthesized as a PstI/SphI restriction fragment with the *rbcl* SD sequence and an IEE element at its 5' end. The fragment was cloned into pZF87 digested with the same enzymes, giving rise to pZF88. Artemisinic acid operon constructs pAO1 and pAO2 were generated by digesting pZF88 with Clal (releasing the cassette containing the *ADS-CPR* dicistron between *Cr PpsbA* and *Cr TrbcL*) and cloning this cassette into pZF90 digested with the same enzyme. In vector pAO1, the *ADS-CPR* cassette is integrated in sense orientation, downstream of the *FPS-CYP* cassette, whereas in construct pAO2, the fragment is integrated in antisense (**Figure 2A**). For generation of pAO3 and pAO4, the *Cr PpsbA* - *Cr LpsbA* fragment was eliminated from pZF88 by digestion with MluI and NcoI and subsequently replaced by a PCR-amplified *Cr PpsbA* - T7 *Lg10* fragment obtained by digestion with the same enzymes, thus generating plasmid clone pZF89. pZF89 was then digested with Clal and cloned into pZF90 in a similar way as for generation of pAO1 and pAO2. Construct pAO3 originates from integration of the *ADS-CPR* cassette in sense orientation, whereas pAO4 harbors the cassette in antisense orientation (**Figure 2A**).

Constructs pCS1-5 for combinatorial supertransformation contain the genes *dxr* (BA000022) from *Synechocystis* sp. and *CYB5* (JQ582841.1), *ADH1* (JF910157.1), *ALDH1* (FJ809784.1) and *DBR2* (EU704257.1) from *A. annua*. The genes were codon optimized for expression in the nucleus and synthesized (Eurofins MWG Operon). The five constructs are derivatives of pUC18 and contain the terminator from the nopaline synthase gene (*Tnos*), the transit peptide from *RBCS* and either the 35S promoter from the cauliflower mosaic virus (CaMV), the mannopine synthase gene promoter from *Agrobacterium tumefaciens* (*Pmas*) or the ubiquitin-10 promoter from *Arabidopsis thaliana* (*PUBIQ10*). To generate these constructs, the *RBCS* transit peptide (TP) was amplified by PCR with primers introducing XbaI/XhoI, ApaI/XhoI or SpeI/XhoI restriction sites into the 5' and 3' ends of the amplification product, respectively. The TP was then digested with the corresponding restriction enzymes and cloned into a *P35S-Tnos* cassette (opened with XbaI/XhoI), a *Pmas-Tnos* cassette (opened with ApaI/XhoI) and a *PUBIQ10-Tnos* cassette (opened with SpeI/XhoI), producing constructs pPF28, pPF29 and pPF30, respectively. Constructs pCS1 and pCS2 are derivatives of pPF28 and were generated by cloning the synthetic genes *dxr* and *CYB5* into pPF28 as XhoI/SacI fragments. Constructs pCS4 and pCS5 are derivatives of pPF29 and were obtained by cloning the synthetic genes *ALDH1* and *DBR2* into pPF29 as XhoI/SacI fragments. Finally, construct pCS3 was obtained in a similar way, by cloning the synthetic gene *ADH1* into pPF30 as an XhoI/XmaI restriction fragment.

The plasmid cocktail for combinatorial transformation was produced by mixing equal quantities of constructs pCS1-5 (each at a concentration of 2 µg/µL) and plasmid pPH200 that contains the *nptII* gene for kanamycin resistance between the CaMV 35S promoter and terminator.

Plastid transformation and selection of transplastomic tobacco plants

For chloroplast transformation, young leaves harvested from aseptically grown wild-type tobacco plants were bombarded with gold particles covered with plasmid-DNA (pAO1-4) using the DuPont PDS1000He biolistic gun. Spectinomycin-resistant shoots were selected on plant regeneration medium with 500 mg/L spectinomycin (**Svab and Maliga, 1993**). Primary transformants were identified by Southern blot analysis and at least one additional regeneration round was performed to obtain homoplasmic plants. Independently generated transplastomic lines are designated by the construct number followed by the number of the individual line (e.g., *Nt-AO1-2* stands for *Nicotiana tabacum* plant obtained with construct pAO1, transplastomic line number 2). Homoplasmy was confirmed by Southern blot analyses and seed assays.

Combinatorial nuclear supertransformation and selection of transgenic tobacco plants

Young leaves from transplastomic plants *Nt-AO2-1* and *Nt-AO3-1* grown under aseptic conditions were harvested and bombarded with gold particles coated with a plasmid DNA mixture containing pCS1-5 and pPH200 using the DuPont PDS1000He biolistic gun. Kanamycin-resistant shoots were selected on plant regeneration medium containing 50 mg/L kanamycin. Resistant shoots were rooted in the same medium, then transferred to soil and grown to maturity under standard greenhouse conditions. Material from T0 plants was used for initial molecular analyses and preliminary

metabolite profiling experiments. To generate standardized material for metabolite measurements and molecular analysis of the T1 generation, seeds from candidate supertransformed lines were surface-sterilized and sown on MS medium with 200 mg/L kanamycin. After three weeks, six green (resistant) seedlings per line were transferred to soil and raised under standard greenhouse conditions.

Plant growth and biomass measurements

Plant height and total leaf biomass were determined for six plants each of *N. tabacum* wild type (wt), the transplastomic line *Nt-AO3-1* and the progeny of four *Nt-AO3-CS180* T1 lines. Measurements were performed at two different stages. The first measurement was done when the wild-type plants started to flower ('same age'). The second measurement was done when the *Nt-AO3-1* and *Nt-AO3-CS180* plants started to flower (typically five days after the first measurement), to compensate for the slightly delayed development of the transplastomic plants and the COSTREL plants. The height was measured from the top of the pot to the top of the inflorescence. The total leaf biomass (fresh weight, FW) was determined by weighing all leaves of an individual plant.

Isolation of nucleic acids

Total plant DNA was extracted from frozen leaf material by a CTAB-based protocol (Doyle and Doyle, 1990). For total RNA extraction, samples of 300–400 mg of frozen powdered plant material were extracted with the peqGOLD Trifast reagent (Peqlab GmbH, Erlangen, Germany), following the manufacturer's instructions. The RNA pellet was resuspended in 100 μ L of RNase-free water and

Table 3. List of oligonucleotides used in this study. The reverse primers (R) for amplification of the genes *FPS*, *ADS*, *CYP* and *CPR* contain the sequence of the T7 promoter (bold) to facilitate in vitro transcription.

Gene	Primer	Sequence (5'→3')	Purpose
<i>dxr</i>	DXR_F	CAACCTATGTACGTTGTTGGAGAAGAGGG	qPCR
	DXR_R	CTGGAGCACCAGCAATCAATGTCTC	
<i>CYB5</i>	CYB5_F	CCAGGAGGAGATGAAGTCTTTTGCTG	qPCR
	CYB5_R	GCTGGAGGAACGTAAGCTCTCTTTG	
<i>ADH1</i>	qADH1_F2	TCCAGGTCATGAAGGTGTTG	qPCR
	qADH1_R2	ATTGTCCACACTACCAAGG	
<i>ALDH1</i>	ALDH1_F	CCTGTTTCTTTGGAATTGGGTGGTAAGTC	qPCR
	ALDH1_R	CAGCAACACACATCTCACCTTTGTTAGTG	
<i>DBR2</i>	DBR2_F	GAGCAAGTTGAGGGTTGGAAGAAAGTTG	qPCR
	DBR2_R	TAGAAGAGATAGGAGCAGCTCCACCTG	
<i>CYP</i>	CYP_qF	CCTGAACCTTGAGATTACC	qPCR
	CYP_qR	GCCCATTTAGGAGAAGATACAAC	
<i>CPR</i>	CPR_qF	CCTGTTGGAATGGGTGATG	qPCR
	CPR_qR	CCTACAGCAGCAGTATAAGGAG	
<i>ACTIN</i>	qTac9actin f	CCTGAGGTCCTTTTCCAACCA	qPCR
	qTac9actin r	GGATTCCGGCAGCTTCCATT	
<i>FPS</i>	FPS_probe_F	CCTGCTTTTGAATTTGATGATG	RNA probe
	FPS_probe_R	TAATACGACTCACTATAGGGCGAAACCAACAAGGTTGTCC	
<i>ADS</i>	ADS_probe_F	CTGAAGCTGTTGAAAGATGGTC	RNA probe
	ADS_probe_R	TAATACGACTCACTATAGGGGGAGCAGATACAGCCCATT	
<i>CYP</i>	CYP_probe_F	TCCTCATCGAGGAGTACGAG	RNA probe
	CYP_probe_R	TAATACGACTCACTATAGGGTTACAGGTCGTCCAGATCCAG	
<i>CPR</i>	CPR_probe_F	ATGATTGGTCCTGGAAGTGG	RNA probe
	CPR_probe_R	TAATACGACTCACTATAGGGGCCATTCTTTAGCATCTCC	

DOI: 10.7554/eLife.13664.018

mixed with 250 μL buffer RA1 from the NucleoSpin RNA Plant kit (Macherey-Nagel, Düren, Germany). 350 μL of 70% EtOH were mixed with the RNA solution, passed through the RNA-binding column and purified following the protocol of the supplier. Finally, the RNA was eluted in 45 μL of RNase-free water and stored at -80°C until use.

cDNA synthesis

Prior to reverse transcription, isolated RNAs were tested for the presence of contaminating DNA by a standard PCR using 1 ng of RNA as template. If no DNA amplification was observed, cDNA was synthesized as follows. 1.5 μg of RNA were incubated with 1 μL of oligo(dT) primer (10 μM) and 1 μL of dNTPs (10 mM) for 5 min at 65°C . Then, 7 μL of a master mix were added (4 μL of 5x First Strand buffer, 1 μL 0.1 M DTT, 40 U RNaseOUT and 200 U SuperScript III Reverse Transcriptase; Invitrogen, Carlsbad, CA) and incubated for 1 hr at 50°C , followed by an inactivation step of 15 min at 70°C . Alternatively, cDNA was synthesized using the QuantiTect Reverse Transcription kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The quality of the cDNA was tested by standard PCR.

Quantitative real-time PCR (qRT-PCR)

Quantitative RT-PCR was performed in a LightCycler 480 (Roche, Mannheim, Germany) using cDNA as template in 5 μL reactions containing 1 μL of each gene-specific primer (1.25 μM ; **Table 3**), 2.5 μL of the LightCycler 480 SYBR green I Master mix and 0.5 μL of a 1:50 cDNA dilution. Three biological (independent plants) and three technical replicates per line were analyzed. The relative transcript levels were determined using the formula $(1+E)^{-\Delta\Delta\text{Ct}}$ where E is the binding efficiency of the primers (**Pfaffl, 2001**). E was calculated from the slope of the expression level of each gene in a dilution series of a given cDNA. Results were normalized to the mRNA levels of *ACTIN* as a housekeeping gene (**Table 3**), and relative mRNA accumulation levels were calculated according to the delta-delta Ct method. To identify the key genes involved in the increased levels of artemisinic acid in super-transformed plants, the expression levels of each transgene (in all lines where it was present) were compared by One-way ANOVA analysis ($p < 0.05$). The results were expressed as a heat map, where the darkest green color represents the highest expression level (brown: no expression).

Synthesis of hybridization probes and gel blot hybridizations

For Southern blot analysis, samples of 2–3 μg DNA were digested with BamHI, separated by electrophoresis in 0.8% agarose gels and transferred onto Hybond XL nylon membranes (GE Healthcare, Little Chalfont, UK) by capillary blotting. For northern blot analysis, samples of 4–5 μg total RNA were separated in denaturing formaldehyde-containing agarose gels (1.5%) and transferred onto nylon membranes. As RFLP probe, a 550-bp fragment of the *psaB* gene was amplified by PCR using primer pair P7247 / P7244 (**Wurbs et al., 2007**) and purified. The probe was labeled with [$\alpha^{32}\text{P}$] dCTP by random priming (Multiprime DNA labeling kit; GE Healthcare). Probes for *FPS*, *ADS*, *CYP* and *CPR* were generated by in vitro transcription and radioactive labeling with [$\alpha^{32}\text{P}$]UTP. PCR fragments of 200–300 bp were amplified for each gene using specific primers (**Table 3**) that contain the T7 promoter sequence in the reverse primer. Radiolabeled probes were generated by incubating 5 μL PCR product with 4 μL H_2O , 2 μL 10x buffer, 3 μL of an equimolar mixture of ATP, CTP and GTP, 2 μL T7 RNA polymerase (15 U/ μL) and 4 μL [$\alpha^{32}\text{P}$]UTP (40 μCi) for 30 min at 37°C . Hybridizations were performed overnight at 65°C . Following standard washing steps, autoradiographic screens were exposed to the membranes for 3–4 hr and then scanned in a Typhoon TRIO+ scanner (GE Healthcare).

Cultivation of yeast reference strains

The following genetically engineered strains of *Saccharomyces cerevisiae* were used as reference strains for artemisinic metabolites (**Ro et al., 2008**): (i) EPY300 (MAT $\alpha\alpha$ his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0 PGAL1-tHMGR PGAL1-upc2-1 erg9::PMET3-ERG9 PGAL1-tHMGR PGAL1-ERG20; **Ro et al., 2008**), a control strain that does not produce artemisinic compounds, (ii) EPY224 (EPY300 transformed with plasmid pRS425-Leu::ADS; **Ro et al., 2006**), a strain that produces amorpho-4,11-diene, and (iii) EPY302 (EPY300 transformed with plasmids pRS425-Leu::ADS and pESC-Ura::AMO/CPR; **Ro et al., 2006**), a strain that produces artemisinic acid and artemisinic acid pathway intermediates. Plasmid

pRS425-Leu::ADS complements the leucine auxotrophy and plasmid pESC-Ura::AMO/CPR complements the uracil auxotrophy of strain EPY300. The three yeast strains were kindly provided by Dr. Jay D. Keasling (UC Berkley, USA). Yeast strains were maintained on solid synthetically defined (SD) medium supplemented with 2% (w/v) sucrose and 0.002% (w/v) uracil for EPY224, and additionally 0.01% (w/v) leucine for EPY300. Induction of the synthesis of artemisinic compounds in strains EPY224 and EPY302 was done by adding 1.8% (w/v) galactose and 1 mM methionine to liquid SD medium and reducing the sucrose content to 0.2% (w/v). Yeast strains were incubated for 120 hr at 30°C and 160 rpm, until they reached an OD₆₀₀ of 1.4. Control strain EPY300 was incubated under the same conditions in liquid SD medium supplemented with 2% (w/v) sucrose, 0.002% (w/v) uracil and 0.01% (w/v) leucine.

GC-MS analyses

For GC-MS profiling of volatile organic compounds (VOCs), leaves of *N. tabacum* plants were collected, immediately frozen in liquid nitrogen and processed in a cryogenic grinding robot (Labman, North Yorkshire, UK). Aliquots of 500 ± 10 mg of frozen powdered leaf tissue were weighed in frozen microcentrifuge tubes, and then transferred to frozen 20 mL head-space screw cap vials. The powdered plant material was kept at 15°C in the closed vials for at least 1 hr and then incubated for 10 min at 50°C prior to VOC analysis. VOCs were sampled in a replicated randomized block sequence design by solid phase micro extraction (SPME) using a StableFlex™ SPME fiber with 65 μm polydimethylsiloxane/divinylbenzene coating (Supelco, Bellefonte, USA), and profiled as described previously (Agudelo-Romero et al., 2015; 2013) using a DB-624 capillary column of 60 m length, 0.25 mm internal diameter and 1.40 μm film thickness (Agilent Technologies Deutschland GmbH, Waldbronn, Germany). VOCs were analyzed by gas chromatography coupled to electron impact ionization/quadrupole mass spectrometry (GC-El/QUAD-MS) using an Agilent 6890N24 gas chromatograph connected to an Agilent 5975B VL mass spectrometer (Agilent Technologies, Böblingen, Germany). Data files were visually controlled, exported in NetCDF file format and baseline-corrected using the Agilent ChemStation software and the MetAlign software (Lommen, 2009). Data processing into a standardized numerical data matrix and compound identification were performed using the TagFinder software (Luedemann et al., 2008). Criteria for manually supervised metabolite identification were the presence of at least three specific and selective mass fragments and a retention time deviation <1.0%. The relative accumulation of amorpha-4,11-diene in VOC profiles of leaf tissue was analyzed using the mass spectral intensity of specific and selective mass fragments (Response) after normalization to fresh weight (Response/FW).

Amorpha-4,11-diene was identified in VOC profiles with the help of the reference substance obtained from cultures of the genetically engineered yeast strain EPY224 (Ro et al., 2006). To this end, the strain was grown in 25 mL of inducing SD medium for 120 hr at 30°C under vigorous shaking (160 rpm), until an OD₆₀₀ of ~1.4 was reached. Additionally, control strain EPY300 was grown in 25 mL of SD medium supplemented with 2% (w/v) sucrose, 0.002% (w/v) uracil and 0.01% (w/v) leucine under the same conditions. Amorpha-4,11-diene was identified by differential display of 1 mL cell suspensions in 20 mL head-space screw cap vials comparing the VOC profiles of the compounds obtained from strain EPY224 with those from control strain EPY300, and analysis of the main differential VOC. VOC profiles obtained from leaf material of *A. annua* were used to further validate the identification of amorpha-4,11-diene. Amorpha-4,11-diene present in the VOCs of tobacco leaf material from transplastomic and combinatorially supertransformed lines was annotated by mass spectral (m/z) and retention time matching to the reference data in the Golm Metabolome Database (GMD, <http://gmd.mpimp-golm.mpg.de/>; Kopka et al., 2005). For compound information and reference data, see the GMD entry for amorpha-4,11-diene (GMD identifier: A149010; <http://gmd.mpimp-golm.mpg.de/search.aspx>). The retention time of amorpha-4,11-diene in the VOC analysis (Agudelo-Romero et al., 2015) of tobacco plant samples was on average 1563 s, with less than 1% deviation between independent experiments. The specific fragments used for verification of the identity of the compound in complex samples were m/z 93, 105, 119, 133, 189 and 204 (Figure 6A).

For preparation and GC-MS profiling of lipophilic saponification products from total leaf tissue containing artemisinic acid and/or intermediates of artemisinic acid biosynthesis, aliquots of 150 ± 5 mg of frozen powdered *N. tabacum* leaves were mixed with 500 μL of 2 N KOH/methanol, and incubated at 70°C for 1 hr with gentle shaking (at 800 rpm). After acidification of the saponified samples with 100 μL of 12 M HCl, 300 μL of hexane were added and the samples were vortexed for 1 min.

After centrifugation for 5 min at 14,000 rpm, 200 μ L of the hexane extract were transferred into a clean microcentrifuge tube and concentrated under a mild N_2 flow to near dryness. Samples were manually trimethylsilylated. Trimethylsilylation was performed by adding 50 μ L of a mixture of N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and an n-alkane standard in hexane (7:1, v/v) followed by incubation at 37°C for 30 min with gentle shaking (800 rpm). Metabolite profiling was performed as detailed previously (Erban et al., 2007) by gas chromatography coupled to electron impact ionization/time-of-flight mass spectrometry (GC-EI/TOF-MS) using an Agilent 6890N24 gas chromatograph (Agilent Technologies) connected to a Pegasus III time-of-flight mass spectrometer (LECO Instrumente GmbH, Mönchengladbach, Germany). Retention indices were calibrated in the range relevant for the intermediates of artemisinic acid biosynthesis by addition of a $C_{15}/C_{18}/C_{19}$ alkane reference mixture to each sample (Strehmel et al., 2008). Chromatograms were acquired, visually controlled, baseline-corrected and exported in NetCDF file format using the ChromaTOF software (Version 4.22; LECO, St. Joseph, USA). Data analysis of GC-EI/TOF-MS profiles of lipophilic saponification products was performed as described for the VOC analysis. Relative quantification of the intermediates of artemisinic acid biosynthesis was performed by calculating normalized responses/FW values using the response of the C_{18} n-alkane and the fresh weight of the sample.

Intermediates of artemisinic acid biosynthesis were initially identified by comparing GC-EI/TOF-MS profiles from yeast strain EPY224 (synthesizing amorpho-4,11-diene) to those of yeast strain EPY302 (synthesizing artemisinic acid and also accumulating all pathway intermediates) and control strain EPY300 (that does not express any of the pathway enzymes). Strain EPY302 was cultured in the same way as strain EPY224, but without addition of uracil. All extractions were performed in duplicate omitting saponification. One of the two sample sets was trimethylsilylated as described above, while the other sample set remained non-derivatized. Non-derivatized samples from yeast were compared to trimethylsilylated samples to unambiguously link the non-derivatized soluble metabolic intermediates of artemisinic acid (A188031) and artemisinic alcohol (A177023) to their respective trimethylsilylated analytes (artemisinic acid 1TMS, A185023; artemisinic alcohol 1TMS, A178029). Dihydroartemisinic alcohol 1TMS (A179026) and dihydroartemisinic acid 1TMS (A186033) were identified after trimethylsilylation. To further validate the identity of the artemisinic compounds, the trimethylsilylated and non-derivatized GC-EI/TOF-MS profiles from yeast were compared to equivalently processed leaf material of *A. annua*. Finally, GC-EI/TOF-MS profiles from authenticated reference compounds (kindly provided by Andreas Pallidis and Dr. Alexander R. van der Krol, Wageningen University, The Netherlands) were used to unambiguously confirm identification of artemisinic acid (AA), artemisinic aldehyde (AAA), artemisinic alcohol (AAOH), dihydroartemisinic acid (DHAA), dihydroartemisinic aldehyde (DHAAA) and dihydroartemisinic alcohol (DHAAOH). Artemisinic alcohol, dihydroartemisinic alcohol, dihydroartemisinic acid and artemisinic acid were identified as trimethylsilylated chemical derivatives in complex profiles according to their mass spectrum (m/z) and retention time index relative to the $C_{15}/C_{18}/C_{19}$ n-alkanes, using reference data from the Golm Metabolome Database. Guidelines for manually supervised metabolite identification were the presence of at least 3 specific mass fragments per compound and a retention index deviation <1.0% (Strehmel et al., 2008). The average retention index of artemisinic alcohol (1TMS) was 1785 and the specific fragments used for verification were m/z 91, 105, 119, 132, 162, 187 and 202. The average retention index of dihydroartemisinic alcohol (1TMS) was 1789 and the specific fragments used for verification were m/z 91, 105, 162, 189 and 204. The average retention index of dihydroartemisinic acid (1TMS) was 1859 and the specific fragments used for verification were m/z 91, 105, 119, 130, 162, 163, 293 and 308. The average retention index of artemisinic acid (1TMS) was 1851 and the specific fragments used for verification were m/z 91, 105, 119, 188, 216, 291 and 306 (Figures 6 and 7). Retention indices of each compound showed a deviation of less than 1% in all measurements performed. For quantification purposes, the most abundant and specific among the selective mass features of each artemisinic metabolite was chosen, i.e., m/z 162 or 202 for artemisinic alcohol (Figure 6B), m/z 162 or 204 for dihydroartemisinic alcohol (Figure 7B), m/z 163 for dihydroartemisinic acid (Figure 7C), and m/z 188 or 216 for artemisinic acid (Figure 7A).

For absolute quantification of artemisinic acid, we first determined the percentage of recovery of artemisinic acid spiked into wild-type tobacco leaf tissue samples in comparison to the recovery of pure artemisinic acid processed without saponification and in the absence of leaf material. To this end, 150 ± 5 mg of powdered frozen leaf material from *N. tabacum* was mixed with 10 μ L of an artemisinic acid standard of known concentration (2 mg/mL in methanol) and subjected to the

saponification protocol. The matrix-free artemisinic acid standard was prepared by dissolving 2 mg of artemisinic acid powder (Apin Chemicals, Oxon, UK) in 1 mL of methanol. All spiked samples were prepared and measured in six replicates and compared to the non-saponified matrix-free artemisinic acid reference samples. The average of the artemisinic acid response values obtained from the reference samples was set to 100%, and the percentage of recovery of artemisinic acid from the leaf tissue matrix after saponification was calculated to be $66 \pm 16\%$. This value was used to correct for the final amount of total artemisinic acid in saponified extracts from plant samples. The artemisinic acid concentration in transplastomic and combinatorially supertransformed plants was calibrated using a dilution series of the commercial non-saponified standard. GC-EI/TOF-MS analysis was as described above. The final quantification of artemisinic acid in line *Nt-AO3-CS180* was done as described above except that, due to the high amounts, only 1/10 of the standard extract volume was used.

For identification of artemisinin or degradation products of artemisinin, aliquots of 1.2 ± 0.01 g of frozen powdered leaf tissue were placed in 20 mL head-space screw cap vials, mixed with 3.6 mL hexane and incubated for 1 hr in a water bath at 69°C. The tubes were shortly vortexed and opened every 10 min to release the vapor pressure. Samples were then centrifuged for 5 min at 14,000 rpm. 300 μ L of the hexane extracts were transferred to 1.1 mL Chromacol vials and reduced to 50 μ L under a mild flow of N_2 . For identification of artemisinin or its degradation products, 500, 1000 or 2500 ng of an artemisinin standard (1 mg/mL; Sigma-Aldrich, Steinheim, Germany) were subjected to the same procedure. GC-EI/TOF-MS profiling was performed as described for soluble metabolites using the whole tissue saponification protocol. As reported previously, only the degradation products of artemisinin (peaks A and B; *Sipahimalani et al., 1991*), were detected, likely due to thermal instability of artemisinin. Peaks A and B were only detected in samples that contained the artemisinin reference compound, but not in any of the plant samples.

UPLC analysis of isoprenoids

For UPLC analysis of pigments, samples of 40 ± 2 mg of frozen powdered leaf tissue were extracted with 500 μ L HPLC grade acetone. A stainless steel ball was added to the mixture and the samples incubated for 20 min at 30°C and 1,400 rpm in the dark. After centrifugation for 5 min at 12,000 rpm and 4°C, the upper phase was collected in a new microcentrifuge tube and stored on ice in darkness. The acetone extraction was repeated two more times, using 250 μ L of acetone each time and combining the three upper phases. Following centrifugation for 5 min at 12,000 rpm and 4°C to precipitate any remaining insoluble material, 600 μ L of the acetone extracts were transferred to 9 mm glass vials. Samples were analyzed using a Waters UPLC Class H (Milford, USA) equipped with an autosampler, Quaternary Solvent Manager, and e λ PDA detector. Pigments were separated in a Waters ACQUITY UPLC BEH C_{18} 1.7 μ m C_{18} 2.1 \times 50 mm column at 28°C, using UPLC solutions A and B. Elution was carried out at a flow rate of 0.5 mL/min with the following gradient: 100–0% of solution A from 0 to 5 min, 100% solution B from 5 to 6 min, 0–100% solution A from 6 to 6.5 min, and 100% solution A from 6.5 to 7.5 min. Carotenoids were detected at 450 nm and chlorophylls at 640 nm. Three biological replicates (i.e., independent plants) per condition were measured and data were analyzed with the Empower 3 software.

Acknowledgements

We thank Dr. Stephanie Ruf and Pierre Endries for help with plant transformation, Ines Fehrle for help with mass spectrometry, Andreas Pallidis and Dr. Alexander R. van der Krol (Wageningen University, The Netherlands) for providing reference compounds for GC-MS, Dr. Jay D. Keasling (UC Berkley, USA) for making available yeast strains synthesizing artemisinic compounds, and the MPI-MP GreenTeam for plant cultivation. This research project received funding from the Max Planck Society and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-ADG-2014; grant agreement No 669982).

Additional information

Funding

Funder	Grant reference number	Author
European Research Council	ERC-ADG-2014 grant agreement No 669982	Ralph Bock
Max-Planck-Gesellschaft		Ralph Bock

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

PF, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; FZ, Conception and design, Acquisition of data, Analysis and interpretation of data; AE, Acquisition of data, Analysis and interpretation of data; DK, Conception and design, Analysis and interpretation of data; JK, RB, Conception and design, Analysis and interpretation of data, Drafting or revising the article

Author ORCIDs

Ralph Bock,  <http://orcid.org/0000-0001-7502-6940>

References

- Agudelo-Romero P**, Erban A, Sousa L, Pais MS, Kopka J, Fortes AM. 2013. Search for transcriptional and metabolic markers of grape pre-ripening and ripening and insights into specific aroma development in three Portuguese cultivars. *PLoS One* **8**:e60422. doi: [10.1371/journal.pone.0060422](https://doi.org/10.1371/journal.pone.0060422)
- Agudelo-Romero P**, Erban A, Rego C, Carbonell-Bejerano P, Nascimento T, Sousa L, Martínez-Zapater JM, Kopka J, Fortes AM. 2015. Transcriptome and metabolome reprogramming in *Vitis vinifera* cv. Trincadeira berries upon infection with *Botrytis cinerea*. *Journal of Experimental Botany* **66**:1769–1785. doi: [10.1093/jxb/eru517](https://doi.org/10.1093/jxb/eru517)
- Apel W**, Bock R. 2009. Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin A conversion. *Plant Physiology* **151**:59–66. doi: [10.1104/pp.109.140533](https://doi.org/10.1104/pp.109.140533)
- Bharati A**, Kar M, Sabat SC. 2012. Artemisinin inhibits chloroplast electron transport activity: mode of action. *PLoS One* **7**:e38942. doi: [10.1371/journal.pone.0038942](https://doi.org/10.1371/journal.pone.0038942)
- Bock R**. 2015. Engineering plastid genomes: Methods, tools, and applications in basic research and biotechnology. *Annual Review of Plant Biology* **66**:211–241. doi: [10.1146/annurev-arplant-050213-040212](https://doi.org/10.1146/annurev-arplant-050213-040212)
- Bohmert-Tatarev K**, McAvoy S, Daughtry S, Peoples OP, Snell KD. 2011. High levels of bioplastic are produced in fertile transplastomic tobacco plants engineered with a synthetic operon for the production of polyhydroxybutyrate. *Plant Physiology* **155**:1690–1708. doi: [10.1104/pp.110.169581](https://doi.org/10.1104/pp.110.169581)
- Bryant L**, Flatley B, Patole C, Brown GD, Cramer R. 2015. Proteomic analysis of *Artemisia annua*—towards elucidating the biosynthetic pathways of the antimalarial pro-drug artemisinin. *BMC Plant Biology* **15**:175. doi: [10.1186/s12870-015-0565-7](https://doi.org/10.1186/s12870-015-0565-7)
- Caroca R**, Howell KA, Hasse C, Ruf S, Bock R. 2013. Design of chimeric expression elements that confer high-level gene activity in chromoplasts. *The Plant Journal* **73**:368–379. doi: [10.1111/tpj.12031](https://doi.org/10.1111/tpj.12031)
- Doyle JJ**, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**:13–15.
- Drechsel O**, Bock R. 2011. Selection of shine-dalgarno sequences in plastids. *Nucleic Acids Research* **39**:1427–1438. doi: [10.1093/nar/gkq978](https://doi.org/10.1093/nar/gkq978)
- Elghabi Z**, Karcher D, Zhou F, Ruf S, Bock R. 2011. Optimization of the expression of the HIV fusion inhibitor cyanovirin-N from the tobacco plastid genome. *Plant Biotechnology Journal* **9**:599–608. doi: [10.1111/j.1467-7652.2011.00598.x](https://doi.org/10.1111/j.1467-7652.2011.00598.x)
- Erban A**, Schauer N, Fernie AR, Kopka J. 2007. Nonsupervised construction and application of mass spectral and retention time index libraries from time-of-flight gas chromatography-mass spectrometry metabolite profiles. *Methods in Molecular Biology* **358**:19–38. doi: [10.1007/978-1-59745-244-1_2](https://doi.org/10.1007/978-1-59745-244-1_2)
- Farhi M**, Marhevka E, Ben-Ari J, Algamas-Dimantov A, Liang Z, Zeevi V, Edelbaum O, Spitzer-Rimon B, Abeliovich H, Schwartz B, Tzfira T, Vainstein A. 2011. Generation of the potent anti-malarial drug artemisinin in tobacco. *Nature Biotechnology* **29**:1072–1074. doi: [10.1038/nbt.2054](https://doi.org/10.1038/nbt.2054)
- Fleischmann TT**, Scharff LB, Alkatib S, Hasdorf S, Schöttler MA, Bock R. 2011. Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *The Plant Cell* **23**:3137–3155. doi: [10.1105/tpc.111.088906](https://doi.org/10.1105/tpc.111.088906)
- Gnanasekaran T**, Karcher D, Nielsen AZ, Martens HJ, Ruf S, Kroop X, Olsen CE, Motawie MS, Pribil M, Møller BL, Bock R, Jensen PE. 2016. Transfer of the cytochrome P450-dependent dhurrin pathway from *Sorghum bicolor* into *Nicotiana tabacum* chloroplasts for light-driven synthesis. *Journal of Experimental Botany* **67**:2495–2506. doi: [10.1093/jxb/erw067](https://doi.org/10.1093/jxb/erw067)

- Graham IA**, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, Guterman I, Harvey D, Isaac PG, Khan AM, Larson TR, Li Y, Pawson T, Penfield T, Rae AM, Rathbone DA, Reid S, Ross J, Smallwood MF, Segura V, et al. 2010. The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* **327**:328–331. doi: [10.1126/science.1182612](https://doi.org/10.1126/science.1182612)
- Guengerich FP**. 2002. Cytochrome P450 enzymes in the generation of commercial products. *Nature Reviews Drug Discovery* **1**:359–366. doi: [10.1038/nrd792](https://doi.org/10.1038/nrd792)
- Hasunuma T**, Miyazawa S, Yoshimura S, Shinzaki Y, Tomizawa K, Shindo K, Choi SK, Misawa N, Miyake C. 2008. Biosynthesis of astaxanthin in tobacco leaves by transplastomic engineering. *The Plant Journal* **55**:857–868. doi: [10.1111/j.1365-313X.2008.03559.x](https://doi.org/10.1111/j.1365-313X.2008.03559.x)
- Kopetzki D**, Lévesque F, Seeberger PH. 2013. A continuous-flow process for the synthesis of artemisinin. *Chemistry - a European Journal* **19**:5450–5456. doi: [10.1002/chem.201204558](https://doi.org/10.1002/chem.201204558)
- Kopka J**, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, Dörmann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D. 2005. GMD@CSB.DB: The golm metabolome database. *Bioinformatics* **21**:1635–1638. doi: [10.1093/bioinformatics/bti236](https://doi.org/10.1093/bioinformatics/bti236)
- Kuroda H**, Maliga P. 2001. Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. *Nucleic Acids Research* **29**:970–975. doi: [10.1093/nar/29.4.970](https://doi.org/10.1093/nar/29.4.970)
- Liu B**, Wang H, Du Z, Li G, Ye H, Du Z, Li G. 2011. Metabolic engineering of artemisinin biosynthesis in *Artemisia annua* L. *Plant Cell Reports* **30**:689–694. doi: [10.1007/s00299-010-0967-9](https://doi.org/10.1007/s00299-010-0967-9)
- Lommen A**. 2009. Metalign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry* **81**:3079–3086. doi: [10.1021/ac900036d](https://doi.org/10.1021/ac900036d)
- Lu Y**, Rijzaani H, Karcher D, Ruf S, Bock R. 2013. Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proceedings of the National Academy of Sciences of the United States of America* **110**:E623–E632. doi: [10.1073/pnas.1216898110](https://doi.org/10.1073/pnas.1216898110)
- Luedemann A**, Strassburg K, Erban A, Kopka J. 2008. Tagfinder for the quantitative analysis of gas chromatography–mass spectrometry (gc-ms)-based metabolite profiling experiments. *Bioinformatics* **24**:732–737. doi: [10.1093/bioinformatics/btn023](https://doi.org/10.1093/bioinformatics/btn023)
- Maliga P**. 2004. Plastid transformation in higher plants. *Annual Review of Plant Biology* **55**:289–313. doi: [10.1146/annurev.arplant.55.031903.141633](https://doi.org/10.1146/annurev.arplant.55.031903.141633)
- Manoj KM**, Gade SK, Mathew L. 2010. Cytochrome P450 reductase: A harbinger of diffusible reduced oxygen species. *PLoS One* **5**:e13272. doi: [10.1371/journal.pone.0013272](https://doi.org/10.1371/journal.pone.0013272)
- Mbengue A**, Bhattacharjee S, Pandharkar T, Liu H, Estiu G, Stahelin RV, Rizk SS, Njimoh DL, Ryan Y, Chotivanich K, Nguon C, Ghorbal M, Lopez-Rubio J-J, Pfreder M, Emrich S, Mohandas N, Dondorp AM, Wiest O, Haldar K. 2015. A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature* **520**:683–687. doi: [10.1038/nature14412](https://doi.org/10.1038/nature14412)
- Mok S**, Ashley EA, Ferreira PE, Zhu L, Lin Z, Yeo T, Chotivanich K, Imwong M, Pukrittayakamee S, Dhorda M, Nguon C, Lim P, Amaratunga C, Suon S, Hien TT, Htut Y, Faiz MA, Onyamboko MA, Mayxay M, Newton PN, et al. 2015. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science* **347**:431–435. doi: [10.1126/science.1260403](https://doi.org/10.1126/science.1260403)
- Murashige T**, Skoog F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum* **15**:473–497. doi: [10.1111/j.1399-3054.1962.tb08052.x](https://doi.org/10.1111/j.1399-3054.1962.tb08052.x)
- Naqvi S**, Zhu C, Farre G, Ramessar K, Bassie L, Breitenbach J, Perez Conesa D, Ros G, Sandmann G, Capell T, Christou P, Conesa DP. 2009. Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proceedings of the National Academy of Sciences of the United States of America* **106**:7762–7767. doi: [10.1073/pnas.0901412106](https://doi.org/10.1073/pnas.0901412106)
- Naqvi S**, Farré G, Sanahuja G, Capell T, Zhu C, Christou P, Farré G, Christou P. 2010. When more is better: Multigene engineering in plants. *Trends in Plant Science* **15**:48–56. doi: [10.1016/j.tplants.2009.09.010](https://doi.org/10.1016/j.tplants.2009.09.010)
- Okell LC**, Cairns M, Griffin JT, Ferguson NM, Tarning J, Jagoe G, Hugo P, Baker M, D’Alessandro U, Bousema T, Ubben D, Ghani AC. 2014. Contrasting benefits of different artemisinin combination therapies as first-line malaria treatments using model-based cost-effectiveness analysis. *Nature Communications* **5**:5606. doi: [10.1038/ncomms6606](https://doi.org/10.1038/ncomms6606)
- Olofsson L**, Engström A, Lundgren A, Brodelius PE. 2011. Relative expression of genes of terpene metabolism in different tissues of *Artemisia annua* L. *BMC Plant Biology* **11**:45. doi: [10.1186/1471-2229-11-45](https://doi.org/10.1186/1471-2229-11-45)
- Paddon CJ**, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, Polichuk DR, Teoh KH, Reed DW, Treynor T, Lenihan J, Jiang H, Fleck M, Bajad S, Dang G, Dengrove D, et al. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**:528–532. doi: [10.1038/nature12051](https://doi.org/10.1038/nature12051)
- Paddon CJ**, Keasling JD. 2014. Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. *Nature Reviews Microbiology* **12**:355–367. doi: [10.1038/nrmicro3240](https://doi.org/10.1038/nrmicro3240)
- Peralta-Yahya PP**, Zhang F, del Cardayre SB, Keasling JD. 2012. Microbial engineering for the production of advanced biofuels. *Nature* **488**:320–328. doi: [10.1038/nature11478](https://doi.org/10.1038/nature11478)
- Pfaffl MW**. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**:e45. doi: [10.1093/nar/29.9.e45](https://doi.org/10.1093/nar/29.9.e45)
- Ro D-K**, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MCY, Withers ST, Shiba Y, Sarpong R, Keasling JD. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**:940–943. doi: [10.1038/nature04640](https://doi.org/10.1038/nature04640)

- Ro DK**, Ouellet M, Paradise EM, Burd H, Eng D, Paddon CJ, Newman JD, Keasling JD. 2008. Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. *BMC Biotechnology* **8**:83. doi: [10.1186/1472-6750-8-83](https://doi.org/10.1186/1472-6750-8-83)
- Schenkman JB**, Jansson I. 2003. The many roles of cytochrome b5. *Pharmacology & Therapeutics* **97**:139–152. doi: [10.1016/S0163-7258\(02\)00327-3](https://doi.org/10.1016/S0163-7258(02)00327-3)
- Sipahimalani AT**, Fulzele DP, Heble MR. 1991. Rapid method for the detection and determination of artemisinin by gas chromatography. *Journal of Chromatography A* **538**:452–455. doi: [10.1016/S0021-9673\(01\)88868-3](https://doi.org/10.1016/S0021-9673(01)88868-3)
- Straimer J**, Gnadig NF, Witkowski B, Amaratunga C, Duru V, Ramadani AP, Dacheux M, Khim N, Zhang L, Lam S, Gregory PD, Urnov FD, Mercereau-Puijalon O, Benoit-Vical F, Fairhurst RM, Menard D, Fidock DA. 2015. K13-propeller mutations confer artemisinin resistance in plasmodium falciparum clinical isolates. *Science* **347**:428–431. doi: [10.1126/science.1260867](https://doi.org/10.1126/science.1260867)
- Strehmel N**, Hummel J, Erban A, Strassburg K, Kopka J. 2008. Retention index thresholds for compound matching in GC-MS metabolite profiling. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **871**:182–190. doi: [10.1016/j.jchromb.2008.04.042](https://doi.org/10.1016/j.jchromb.2008.04.042)
- Svab Z**, Maliga P. 1993. High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proceedings of the National Academy of Sciences of the United States of America* **90**:913–917. doi: [10.1073/pnas.90.3.913](https://doi.org/10.1073/pnas.90.3.913)
- Tang K**, Shen Q, Yan T, Fu X. 2014. Transgenic approach to increase artemisinin content in artemisia annua L. *Plant Cell Reports* **33**:605–615. doi: [10.1007/s00299-014-1566-y](https://doi.org/10.1007/s00299-014-1566-y)
- Ting HM**, Wang B, Rydén AM, Woittiez L, van Herpen T, Verstappen FW, Ruyter-Spira C, Beekwilder J, Bouwmeester HJ, van der Krol A, Rydén A-M. 2013. The metabolite chemotype of nicotiana benthamiana transiently expressing artemisinin biosynthetic pathway genes is a function of CYP71AV1 type and relative gene dosage. *The New Phytologist* **199**:352–366. doi: [10.1111/nph.12274](https://doi.org/10.1111/nph.12274)
- van Herpen TWJM**, Cankar K, Nogueira M, Bosch D, Bouwmeester HJ, Beekwilder J. 2010. Nicotiana benthamiana as a production platform for artemisinin precursors. *PLoS ONE* **5**:e14222–11. doi: [10.1371/journal.pone.0014222](https://doi.org/10.1371/journal.pone.0014222)
- Willoughby JA**, Sundar SN, Cheung M, Tin AS, Modiano J, Firestone GL, Sundar SN, Cheung M. 2009. Artemisinin blocks prostate cancer growth and cell cycle progression by disrupting sp1 interactions with the cyclin-dependent kinase-4 (CDK4) promoter and inhibiting CDK4 gene expression. *The Journal of Biological Chemistry* **284**:2203–2213. doi: [10.1074/jbc.M804491200](https://doi.org/10.1074/jbc.M804491200)
- Wu S**, Schalk M, Clark A, Miles RB, Coates R, Chappell J. 2006. Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nature Biotechnology* **24**:1441–1447. doi: [10.1038/nbt1251](https://doi.org/10.1038/nbt1251)
- Wurbs D**, Ruf S, Bock R. 2007. Contained metabolic engineering in tomatoes by expression of carotenoid biosynthesis genes from the plastid genome. *The Plant Journal* **49**:276–288. doi: [10.1111/j.1365-313X.2006.02960.x](https://doi.org/10.1111/j.1365-313X.2006.02960.x)
- Zeng Q-P**, Zeng X-M, Yin L-L, Yang R-Y, Feng L-L, Yang X-Q. 2009. Quantification of three key enzymes involved in artemisinin biogenesis in artemisia annua by polyclonal antisera-based ELISA. *Plant Molecular Biology Reporter* **27**:50–57. doi: [10.1007/s11105-008-0056-1](https://doi.org/10.1007/s11105-008-0056-1)
- Zhang J**, Ruf S, Hasse C, Childs L, Scharff LB, Bock R. 2012. Identification of cis-elements conferring high levels of gene expression in non-green plastids. *The Plant Journal* **72**:115–128. doi: [10.1111/j.1365-313X.2012.05065.x](https://doi.org/10.1111/j.1365-313X.2012.05065.x)
- Zhang Y**, Teoh KH, Reed DW, Maes L, Goossens A, Olson DJ, Ross AR, Covello PS. 2008. The molecular cloning of artemisinic aldehyde delta11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in artemisia annua. *The Journal of Biological Chemistry* **283**:21501–21508. doi: [10.1074/jbc.M803090200](https://doi.org/10.1074/jbc.M803090200)
- Zhang Y**, Nowak G, Reed DW, Covello PS. 2011. The production of artemisinin precursors in tobacco. *Plant Biotechnology Journal* **9**:445–454. doi: [10.1111/j.1467-7652.2010.00556.x](https://doi.org/10.1111/j.1467-7652.2010.00556.x)
- Zhou F**, Karcher D, Bock R. 2007. Identification of a plastid intercistronic expression element (IEE) facilitating the expression of stable translatable monocistronic mRNAs from operons. *The Plant Journal* **52**:961–972. doi: [10.1111/j.1365-313X.2007.03261.x](https://doi.org/10.1111/j.1365-313X.2007.03261.x)
- Zhou F**, Badillo-Corona JA, Karcher D, Gonzalez-Rabade N, Piepenburg K, Borchers AM, Maloney AP, Kavanagh TA, Gray JC, Bock R. 2008. High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. *Plant Biotechnology Journal* **6**:897–913. doi: [10.1111/j.1467-7652.2008.00356.x](https://doi.org/10.1111/j.1467-7652.2008.00356.x)
- Zhu C**, Naqvi S, Breitenbach J, Sandmann G, Christou P, Capell T. 2008. Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proceedings of the National Academy of Sciences of the United States of America* **105**:18232–18237. doi: [10.1073/pnas.0809737105](https://doi.org/10.1073/pnas.0809737105)