LETTER

High-level semi-synthetic production of the potent antimalarial artemisinin

C. J. Paddon[1], P. J. Westfall[1]†, D. J. Pitera[1], K. Benjamin[1], K. Fisher[1], D. McPhee[1], M. D. Leavell[1], A. Tai[1], A. Main[1]†, D. Eng[1], D. R. Polichuk[2], K. H. Teoh[2]†, D. W. Reed[2], T. Treynor[1], J. Lenihan[1]†, M. Fleck[1], S. Bajad[1]†, G. Dang[1]†, D. Dengrove[1], D. Diola[1], G. Dorin[1], K. W. Ellens[2]†, S. Fickes[1], J. Galazzo[1], S. P. Gaucher[1], T. Geistlinger[1], R. Henry[1], M. Hepp[2]†, T. Horning[1], T. Iqbal[1], H. Jiang[1], L. Kizer[1], B. Lieu[1], D. Melis[1], N. Moss[1], R. Regentin[1]†, S. Secrest[1], H. Tsuruta[1], R. Vazquez[1], L. F. Westblade[1], L. Xu[1], M. Yu[1], Y. Zhang[2]†, L. Zhao[1], J. Lievense[1]†, P. S. Covello[2], J. D. Keasling[3,4,5,6], K. K. Reiling[1], N. S. Renninger[1] & J. D. Newman[1]

In 2010 there were more than 200 million cases of malaria, and at least 655,000 deaths[1]. The World Health Organization has recommended artemisinin-based combination therapies (ACTs) for the treatment of uncomplicated malaria caused by the parasite *Plasmodium falciparum*. Artemisinin is a sesquiterpene endoperoxide with potent antimalarial properties, produced by the plant *Artemisia annua*. However, the supply of plant-derived artemisinin is unstable, resulting in shortages and price fluctuations, complicating production planning by ACT manufacturers[2]. A stable source of affordable artemisinin is required. Here we use synthetic biology to develop strains of *Saccharomyces cerevisiae* (baker's yeast) for high-yielding biological production of artemisinic acid, a precursor of artemisinin. Previous attempts to produce commercially relevant concentrations of artemisinic acid were unsuccessful, allowing production of only 1.6 grams per litre of artemisinic acid[3]. Here we demonstrate the complete biosynthetic pathway, including the discovery of a plant dehydrogenase and a second cytochrome that provide an efficient biosynthetic route to artemisinic acid, with fermentation titres of 25 grams per litre of artemisinic acid. Furthermore, we have developed a practical, efficient and scalable chemical process for the conversion of artemisinic acid to artemisinin using a chemical source of singlet oxygen, thus avoiding the need for specialized photochemical equipment. The strains and processes described here form the basis of a viable industrial process for the production of semi-synthetic artemisinin to stabilize the supply of artemisinin for derivatization into active pharmaceutical ingredients (for example, artesunate) for incorporation into ACTs. Because all intellectual property rights have been provided free of charge, this technology has the potential to increase provision of first-line antimalarial treatments to the developing world at a reduced average annual price.

Before the discovery of the enzymes that complete the biosynthetic pathway of artemisinin production (see Supplementary Fig. 1 for a complete overview), several improvements were made to the original amorphadiene-producing strain Y337 (ref. 3). We replaced the *MET3* promoter with the copper-regulated *CTR3* promoter (Fig. 1a), enabling restriction of *ERG9* expression (*ERG9* encodes squalene synthase, which catalyses the competing reaction of joining two farnesyl diphosphate moieties to form squalene) by addition of the inexpensive repressor $CuSO_4$ to the medium rather than the more expensive methionine[4–6]. Strains Y1516 (P_{CTR3}-*ERG9*) and Y337 (P_{MET3}-*ERG9*)

(Supplementary Table 1) both produced similar amounts of amorphadiene (Supplementary Fig. 2), demonstrating the equivalence of the *MET3* and *CTR3* promoters for repression of *ERG9* expression. We compared the production of amorphadiene from Y337 with the production of artemisinic acid from Y285, a variant of Y337 that also expressed the amorphadiene oxidase *CYP71AV1* (a cytochrome P450) and *A. annua CPR1* (its cognate reductase) from a high-copy plasmid (pAM322)[3]. Both strains were grown in a fed-batch fermentor with mixed glucose and ethanol feed. Whereas Y337 produced more than $12\,\text{g}\,\text{l}^{-1}$ of amorphadiene, Y285 produced significantly less sesquiterpene: $3.3\,\text{g}\,\text{l}^{-1}$ of artemisinic acid (Fig. 2a and Supplementary Table 2) plus $0.3\,\text{g}\,\text{l}^{-1}$ amorphadiene, $0.18\,\text{g}\,\text{l}^{-1}$ artemisinic alcohol and no detectable artemisinic aldehyde (Supplementary Table 3). The viability of the Y285 culture also decreased markedly after *CYP71AV1* and *CPR1* expression (Fig. 2a). We surmised that the decreased viability and reduced production of sesquiterpene products in Y285 might be caused by the cytochrome P450 responsible for oxidizing amorphadiene, or by the rapid accumulation of artemisinic acid.

Poor coupling between P450 cytochromes and their reductases can result in the release of reactive oxygen species[7]. In liver microsomes, the P450 enzyme is generally present in excess over its reductase[8], whereas in Y285 and Y301 both enzymes are expressed from strong galactose-regulated promoters on a high-copy plasmid, and are presumably present at similar levels. We reduced expression of *CPR1* by expressing it from a weaker promoter (*GAL3* promoter) and integrating a single copy into genomic DNA, generating strain Y657 (Supplementary Table 1). Y657 had an increase in cell growth (Fig. 2b) and viability (Supplementary Fig. 3) compared to either Y285 or Y301 (isogenic to Y285, but with P_{CTR3}-*ERG9* replacing P_{MET3}-*ERG9*), but showed lower artemisinic acid production in shake-flask cultures (Fig. 2c) and mixed-feed fed-batch fermentors (Supplementary Tables 2 and 3). Comparison of all amorphadiene-derived sesquiterpenes showed that although reducing *CPR1* expression decreased artemisinic acid production, total sesquiterpene production remained relatively high, indicating that low *CPR1* levels increase cell health, but decrease the total rate of amorphadiene oxidations (Fig. 3a; compare Y301 and Y657). The reaction rate of some cytochromes P450 is enhanced by their interaction with cytochrome b_5 as explained by several possible mechanisms[9,10]. We identified a cytochrome b_5 complementary DNA from *A. annua* (*CYB5*; Supplementary Fig. 4) and expressed a chromosomally integrated copy from a strong promoter

[1]Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, California 94608, USA. [2]National Research Council of Canada, 110 Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada. [3]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, USA. [4]Department of Bioengineering, University of California, Berkeley, California 94720, USA. [5]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [6]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, California 94608, USA. †Present addresses: Dow AgroSciences, 9330 Zionsville Road, Indianapolis, Indiana 46268, USA (P.J.W.); Radiant Genomics, Inc., 2430 Fifth Street, Suite D, Berkeley, California 94710, USA (A.M.); Arkansas Biosciences Institute, Arkansas State University, Jonesboro, Arkansas 72401, USA (K.H.T.); Solazyme, Inc., 225 Gateway Boulevard, South San Francisco, California 94080, USA (J. Lenihan); Sutro Biopharma Inc., 310 Utah Avenue, South San Francisco, California 94080, USA (S.B.); Abbvie, 1500 Seaport Boulevard, Redwood City, California 94063, USA (G. Dang); Horticultural Sciences Department, University of Florida, PO Box 110690, Gainesville, Florida 32611-0690, USA (K.W.E.); Department of Plant Sciences, University of Saskatchewan, 51 Campus Drive, Saskatoon, Saskatchewan S7N 5A8, Canada (M.H.); Regentin Consulting, 25409 Modoc Court, Hayward, California 94542, USA (R.R.); AS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, The Chinese Academy of Science, Wuhan 430074, China (Y.Z.); Genomatica, Inc., 10520 Wateridge Circle, San Diego, California 92121, USA (J. Lievense).
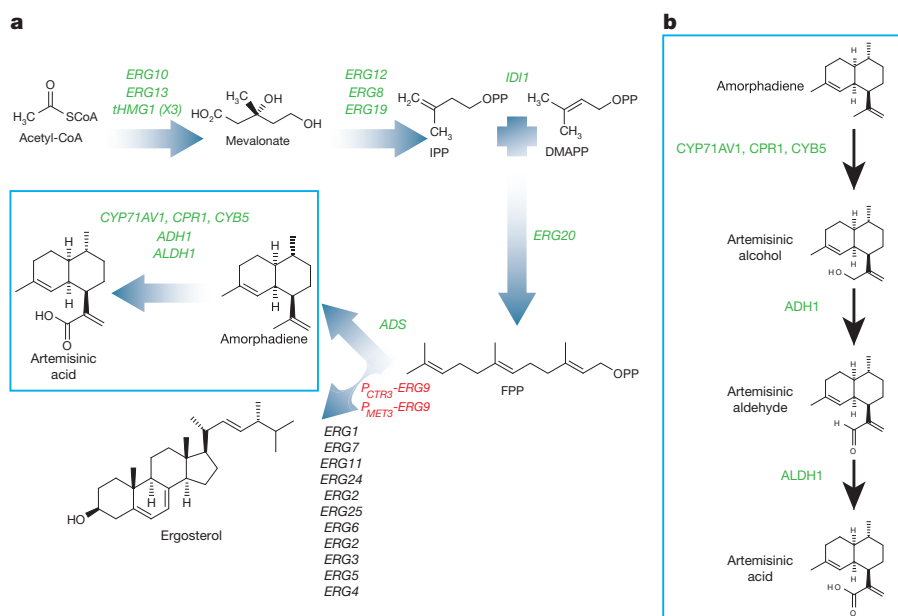
Figure 1 | Artemisinic acid production pathway in *S. cerevisiae* and summary of strains described. **a**, Overview of artemisinic acid production pathway. Overexpressed genes controlled by the GAL induction system are shown in green. Copper- or methionine-repressed squalene synthase (*ERG9*) is shown in red. DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; IPP, isopentenyl diphosphate. *tHMG1* encodes truncated HMG-CoA reductase. **b**, The full three-step oxidation of amorphadiene to artemisinic acid from *A. annua* expressed in *S. cerevisiae*. CYP71AV1, CPR1 and CYB5 oxidize amorphadiene to artemisinic alcohol; ADH1 oxidizes artemisinic alcohol to artemisinic aldehyde; ALDH1 oxidizes artemisinic aldehyde to artemisinic acid. Strains containing these genes are described in Supplementary Table 1.

(*GAL7* promoter) in a strain with low CPR1 expression. The resulting strain, Y692 (Supplementary Table 1), produced higher concentrations of artemisinic acid than strains without CYB5 (Fig. 2c and Supplementary Tables 2 and 3; compare Y657 and Y692). Expression of CYB5 also increased the production of artemisinic aldehyde, leading to a 40% increase in total sesquiterpene production in shake-flask cultures (Fig. 3a; compare Y657 and Y692), and almost doubled the production of artemisinic aldehyde in fermentors (Supplementary Table 3). In view of the reactivity and presumed toxicity of artemisinic aldehyde, we expressed a recently isolated cDNA encoding *A. annua* artemisinic aldehyde dehydrogenase (*ALDH1*)[11] in Y692 to produce Y973 and Y1368 (also expressing increased levels of cytosolic catalase to reduce oxidative stress; Supplementary Table 1). Expression of *ALDH1* markedly increased the production of artemisinic acid in both flask (Fig. 2c; Y1368) and fermentor cultures (Supplementary Tables 2 and 3; Y973). Artemisinic aldehyde was undetectable in flask cultures (Fig. 3a; Y1368), and barely detectable in fermentors (Supplementary Table 3; Y973). Furthermore, the expression of *ALDH1* in Y973 allowed early induction of fermentor cultures immediately after inoculation (previous attempts at early induction with Y285 and Y301 had resulted in rapid loss of viability), further increasing production to $7.7\,\mathrm{g\,l^{-1}}$ artemisinic acid (Supplementary Tables 2 and 4). The yield (Cmol% of substrate carbon incorporated into artemisinic acid) was more than doubled compared to the initial Y285 cultures (Supplementary Tables 3 and 4).

In the course of investigating the biosynthesis of artemisinin in *A. annua* glandular trichomes, a gene encoding a putative alcohol dehydrogenase (*ADH1*) was examined. The gene is represented by a contiguous set of glandular trichome-derived expressed sequence tags (ESTs)[12] corresponding to 1.3% of the EST collection. The *A. annua* *ADH1* open reading frame (ORF) was expressed as a fusion protein and purified from *Escherichia coli*. Sequence analysis and *in vitro* characterization revealed that ADH1 is an NAD-dependent alcohol dehydrogenase of the medium chain dehydrogenase/reductase super-family, with specificity towards artemisinic alcohol (Michaelis constant $(K_{\mathrm{m}}) = 11 \pm 3\,\mu\mathrm{M}$, $k_{\mathrm{cat}} = 41 \pm 5\,\mathrm{s^{-1}}$ (mean $\pm$ s.e.m.); Supplementary Fig. 5). This specificity and the evidence for strong glandular trichome expression indicate a role for ADH1 in the formation of artemisinic aldehyde in the artemisinin pathway of *A. annua*. Therefore, we propose that all five enzymes (CYP71AV1, CPR1, CYB5, ADH1 and ALDH1) are involved in the oxidation of amorphadiene to artemisinic acid in *A. annua* plants, and set out to reconstitute the entire heterologous biosynthetic pathway in yeast (Fig. 1b).

Observing the accumulation of artemisinic alcohol in strain Y1368 (Fig. 3a), we completed the biosynthetic pathway by expressing *ADH1* in conjunction with *ALDH1*, *CYP71AV1*, *CYB5* and *CPR1*. The resulting strain, Y1283, produced no detectable artemisinic alcohol in flask cultures, and increased artemisinic acid production by 18% (Figs 2c and 3a; compare Y1368 and Y1283). In fermentors, *ADH1* increased production to $8.1\,\mathrm{g\,l^{-1}}$ in an early induction, mixed-feed process, while
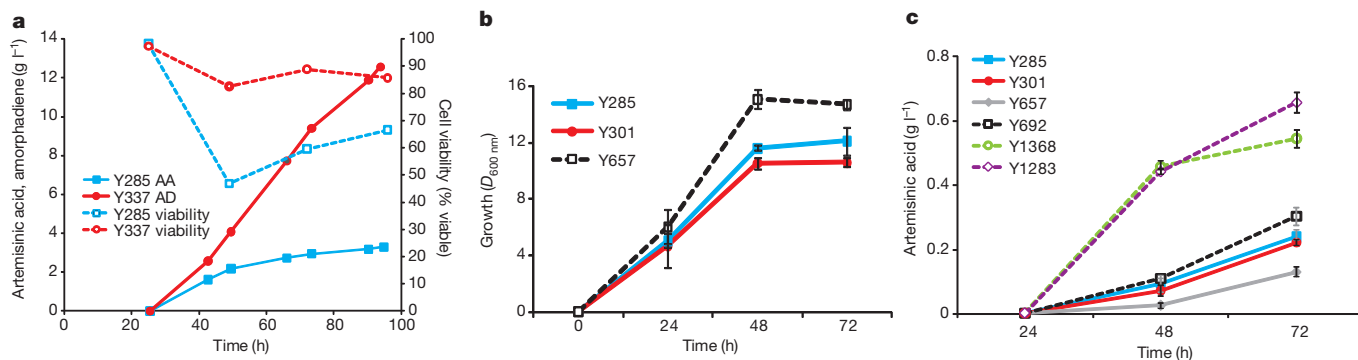

Figure 2 | Growth, viability and production by *S. cerevisiae* strains. **a**, Production and cell viability of artemisinic acid production strain Y285 and amorphadiene production strain Y337 in the glucose and ethanol mixed-feed, fed-batch fermentation process. AA, artemisinic acid; AD, amorphadiene. **b**, Growth of artemisinic-acid-producing strains in shake-flasks. **c**, Production of artemisinic acid in shake-flasks by different strains. Error bars denote standard deviation of triplicate shake-flask cultures.
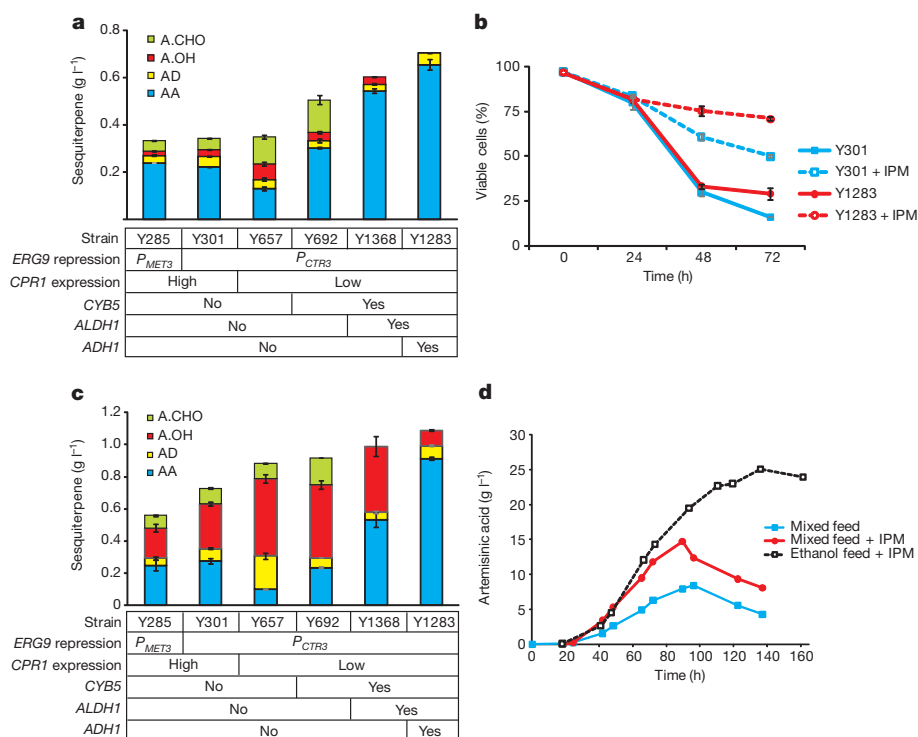
Figure 3 | Increasing production of artemisinic acid by strain engineering and addition of IPM to cultures. a, Artemisinic acid production in flasks (no IPM). A.CHO, artemisinic aldehyde; A.OH, artemisinic alcohol. **b**, Effect of IPM addition on the viability of strains Y301 and Y1283 in shake-flask cultures. **c**, Production of artemisinic acid in flasks containing IPM. **d**, Artemisinic acid production of strain Y1284 in fed-batch fermentation processes. All fermentations were run with early repression (150 μM CuSO$_4$ was added before inoculation). Because Y1284 has the *gal80*Δ genotype, no galactose was added to the fermentations. Error bars denote standard deviation of triplicate shake-flask cultures.

reducing production of artemisinic alcohol (Supplementary Table 4). The improved viability (Supplementary Fig. 6) of Y1283, which also contained the added catalase activity, allowed us to express all heterologous galactose-regulated enzymes constitutively by deletion of *GAL80*, a strategy used previously to increase production in amorphadiene-production strains[3]. This strain, Y1284, does not require the inducer galactose, yet produces higher concentrations of artemisinic acid compared to its parent (Supplementary Tables 2–4).

We observed that strains expressing ALDH1 produced artemisinic acid as a crystalline extracellular precipitate (Supplementary Figs 7 and 8). Precipitation was also observed in early induction fed-batch mixed-feed fermentors, complicating accurate measurement of product from the heterogeneous fermentor samples. To overcome the sampling difficulties associated with artemisinic acid precipitation, we investigated the effect of solubilizing the precipitate by extractive fermentation[13], growing the cultures in the presence of isopropyl myristate (IPM) oil. Addition of 10% (v/v) IPM to flask cultures resulted in a marked increase in the viability of all strains, shown for Y301 and Y1283 in Fig. 3b. Whereas the addition of IPM to earlier strains (Y285, Y301, Y657 and Y692) lacking ALDH1 and ADH1 resulted in extraction of intermediates (amorphadiene, artemisinic alcohol and aldehyde; Fig. 3c), IPM addition to strains containing ALDH1 and ADH1 produced artemisinic acid to >14 g l^{-1} in a mixed-feed, early induction fermentation process (Supplementary Table 2), with a six-fold increase in yield compared to Y285 (Supplementary Table 5). The IPM extractive fermentation of Y1284 allowed us to increase production further by developing a feedback-controlled ethanol pulse-feed process; in this process Y1284 produced 25 g l^{-1} of artemisinic acid, with a sevenfold higher yield than Y285 (Fig. 3d and Supplementary Tables 2 and 5). Artemisinic acid production in the mixed-feed process was seen to reach a maximum concentration due to precipitation from the culture broth mixture and a cessation in production while the fed-batch fermentation continued. Additional IPM was added to the ethanol-feed fermentation to avoid precipitation. To take advantage of this high-titre process we developed a method for extracting artemisinic acid from IPM with high yield and purity (see Methods).

Numerous total and partial chemical syntheses of artemisinin exist[14–16]. Here we report one starting from our purified artemisinic acid (Fig. 4a). For this route to be scalable and practical, several modifications of the previously published syntheses are required. The first step is the reduction of the Δ11(13) double bond (numbering system of Sy and Brown[17]) to give dihydroartemisinic acid, which has two epimers, of which only the (*R*)-11 one has the correct stereochemistry found in artemisinin. This had typically been carried out with 'nickel boride' (NaBH$_4$ or LiBH$_4$ plus NiCl$_2$), which gives a ~3:1 to 85:15 ratio of the 11-epimers, favouring the desired one, but the use of a stoichiometric excess of the reducing agent and the poor isomer ratio are unacceptable for a cost-effective scalable synthesis. We found that catalytic hydrogenation using several different noble metal catalysts affords nearly quantitative yields of the reduced acid in (*R*)-*11*:(*S*)-*11* epimer ratios as high as 94:6, without any significant overreduction to tetrahydroartemisinic acid (Supplementary Fig. 9 and Supplementary Table 6).

The next step is the esterification of the carboxylic acid. The subsequent reactions can be performed using the acid (R = H), but this results in considerable yield losses owing to the formation of the five-membered lactone-containing compound dihydroepideoxyarteannuin B, a side-reaction blocked by the presence of the ester. Large-scale esterification was readily accomplished by carboxyl activation by acid chloride formation, followed by an alcohol quench (Supplementary Fig. 10).

The third step is an 'ene-type' reaction of the C4–C5 double bond with singlet oxygen (^{1}O$_2$) to give an allylic 3-hydroperoxide. In previous syntheses, the ^{1}O$_2$ was invariably generated by photosensitized energy transfer from dye molecules (such as rose bengal, methylene blue and porphyrins)[18–20], but because photosynthetic steps are rarely found in manufacturing facilities, we sought another source of ^{1}O$_2$. A practical alternative was found in the group VI metal salt-induced disproportionation of concentrated H$_2$O$_2$ (ref. 21). With this technique the hydroperoxide could be formed cleanly with no evidence of isomers or rearrangement products.

In the final step the allylic hydroperoxide undergoes an acid-catalysed Hock fragmentation and rearrangement to afford a ring-opened keto-aldehyde enol. Trapping of this enol with ^{3}O$_2$ produces

a vicinal hydroperoxide aldehyde, which in a cascade of acid-catalysed cyclizations forms an endoperoxide bridge, a seven-membered cyclic ether, and finally a six-membered lactone, thus producing good yields of artemisinin with the correct stereochemistry (Supplementary Figs 11–13). Improvements of this step involved substitution of expensive copper triflate[16] with benzenesulphonic acid/sulphonate copper(II) DOWEX resin and replacement (for safety reasons) of pure O_2 by air, in addition to reaction condition optimization to maximize the yield of artemisinin. These changes produced a scalable four-step synthesis that gave artemisinin in 40–45% overall yield, a marked
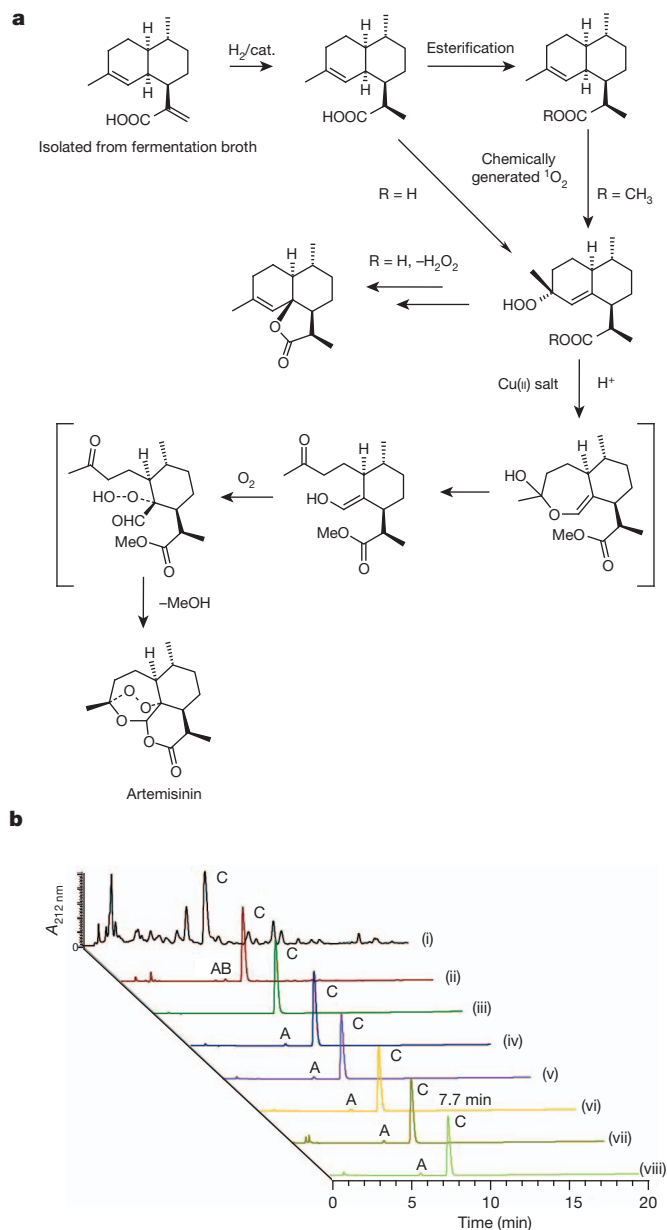

Figure 4 | Chemical conversion of artemisinic acid to artemisinin. a, Semi-synthesis of artemisinin from microbially produced artemisinic acid. Compounds in square brackets denote intermediates detected but not isolated. **b**, HPLC ultraviolet traces of crude, partially purified and purified semi-synthetic artemisinin and commercial samples of plant-derived artemisinin. For HPLC conditions see Supplementary Information. Major peak identification (retention time in min): artemisitene (5.93–5.96, peak A), 9-*epi*-artemisinin (6.54, peak B); artemisinin (7.72–7.74, peak C). Traces: (i) crude reaction mixture: <33% pure; (ii) crude mixture filtered through silica gel: 90.5% pure; (iii) recrystallized: 99.6% pure; (iv–viii) samples of plant artemisinin from different Chinese commercial suppliers, all quality control approved as raw materials suitable for ACT manufacture (97.2–97.8% pure).

improvement over the typical yields reported in the literature[18–20] (Supplementary Table 7). The product is purer than the plant-sourced artemisinin used in ACT production, as shown by high-performance liquid chromatography (HPLC) comparison of our material with a selection of commercial plant samples (Fig. 4b and Supplementary Table 8).

Our results describe for the first time, to our knowledge, the expression of the complete pathway for artemisinic acid production, which resulted in a greater than tenfold increase in artemisinic acid titres. In addition, we demonstrated a significant increase in the efficiency of artemisinic acid conversion to artemisinin compared with earlier work[18–20]. We show that expression of CYP71AV1 and its cognate reductase is not sufficient for high-level production of artemisinic acid, instead requiring three additional plant enzymes (CYB5, ADH1 and ALDH1 (ref. 11); Fig. 1b). Optimization of the CYP71AV1:CPR1 expression ratio, combined with CYB5 expression, overcame initial viability problems, but high titre and increased yield were only achieved when *A. annua* artemisinic alcohol and aldehyde dehydrogenases were co-expressed. These observations in yeast lend strong support to the importance of the dehydrogenases in artemisinin biosynthesis in the native plant system[22]. Extractive fermentation with a bio-compatible solvent (IPM) increased the production of artemisinic acid through an uncharacterized effect, perhaps by effectively removing the acid from the aqueous phase. The development of a facile procedure for purification of artemisinic acid from IPM in high yield and purity allows a ready supply for chemical conversion to artemisinin. The chemical conversion procedure is notable for its simplicity, scalability, economy of reagents, and the high yield obtained. Nevertheless, given investment in suitable large-scale photoreactors, improvements of the classical photochemistry syntheses have the potential to increase the overall process yields further (see Supplementary Information for details and comparison to photochemical syntheses[18–20]).

In summary, we have determined the full artemisinic acid biosynthetic pathway and developed a process for the production of the antimalarial drug artemisinin by fermentation of simple inexpensive carbon substrates using engineered *S. cerevisiae* to produce artemisinic acid, followed by extraction and chemical conversion to artemisinin. These key developments in yeast strain engineering, fermentation, and artemisinin synthetic chemistry pave the way for an industrial process capable of supplementing the world supply of artemisinin from a second source independent of the uncertainties associated with botanical production.

METHODS SUMMARY

The *A. annua* cytochrome b_5 cDNA sequence was identified from a trichome expressed sequence tag library (NCBI accession 35608) by searching for sequence similarity to *Crepis alpina* cytochrome b_5 type 11 (ref. 23). A cDNA encoding *A. annua* ADH1 was identified and the encoded protein characterized essentially as described for *ALDH1* (ref. 11). Growth of strains, general genetic methodology and construction of synthetic genes were as described[3]. Yeast strains were derived from Y337 (ref. 3). Salient features of DNA constructs are as follows: (1) for expression of *ERG9* from the *CTR3* promoter the *MET3* promoter was replaced with nucleotides −1 to −734 of the *CTR3* promoter, integration being selected by D-serine[24] (*erg9Δ::dsdA_P_{CTR3}-ERG9*) or nourseothricin[25] (*erg9Δ::natA_P_{CTR3}-ERG9*) resistance; (2) for reduced expression of cytochrome P450 reductase (*CPR1*), *CPR1* was removed from plasmid pAM322 (ref. 3) by digestion and recircularization to generate pAM552, which expresses only ADS and CYP71AV1. A single copy of *CPR1* was expressed from the *GAL3* promoter (P_{GAL3}) (nucleotides −1 to −660) integrated between *GAL1* and *GAL7* (*gal1/10/7Δ::natA_P_{GAL3}-CPR1-T_{CYC1}*, in which T_{CYC1} denotes the *CYC1* terminator); (3) a single integrated copy of *A. annua* cytochrome b_5 was expressed from the *GAL7* promoter (nucleotides −1 to −725; *leu2::hisMXΔ::kanA_P_{GAL7}-CYB5-T_{CYC1}*); (4) a single integrated copy of *A. annua* aldehyde dehydrogenase (*ALDH1*) was expressed from the *GAL7* promoter, selecting for hygromycin B (ref. 25) (*ndt80::hphA_P_{GAL7}-ALDH1-T_{TDH1}* and *his3::hphA_P_{GAL7}-ALDH1-T_{TDH1}*); (5) a single integrated copy of *A. annua* alcohol dehydrogenase (*ADH1*) was expressed from the *GAL7* promoter, selecting for uracil prototrophy (*natAΔ::URA3_P_{GAL7}-ADH1-T_{TDH1}* and *gal80Δ::URA3_P_{GAL7}-ADH1-T_{GAL80}*).

Flask and fermentor culture conditions were essentially as described[3]. Fermentations requiring IPM contained 400 ml IPM added to 800 ml fermentor volume before inoculation. Artemisinic acid was purified from IPM by aqueous extraction at pH 10.7, followed by precipitation at pH 5.0. Assays for amorph-4,11-diene and artemisinic acid are essentially as described[3]. Artemisinic alcohol and artemisinic aldehyde were monitored by gas chromatography with flame-ionization detection.

Full Methods and any associated references are available in the online version of the paper.

1. World Health Organization. *World Malaria Report 2011* http://www.who.int/malaria/publications/atoz/9789241564403/en/index.html (World Health Organization, 2011).
2. Hale, V., Keasling, J. D., Renninger, N. & Diagana, T. T. Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs. *Am. J. Trop. Med. Hyg.* **77,** 198–202 (2007).
3. Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl Acad. Sci. USA* **109,** E111–E118 (2012).
4. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440,** 940–943 (2006).
5. Paradise, E. M., Kirby, J., Chan, R. & Keasling, J. D. Redirection of flux through the FPP branch-point in *Saccharomyces cerevisiae* by down-regulating squalene synthase. *Biotechnol. Bioeng.* **100,** 371–378 (2008).
6. Asadollahi, M. A. *et al.* Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnol. Bioeng.* **99,** 666–677 (2008).
7. Zangar, R. C., Davydov, D. R. & Verma, S. Mechanisms that regulate production of reactive oxygen species by cytochrome P450. *Toxicol. Appl. Pharmacol.* **199,** 316–331 (2004).
8. Peterson, J. A., Ebel, R. E., O'Keeffe, D. H., Matsubara, T. & Estabrook, R. W. Temperature dependence of cytochrome P-450 reduction. A model for NADPH-cytochrome P-450 reductase:cytochrome P-450 interaction. *J. Biol. Chem.* **251,** 4010–4016 (1976).
9. Schenkman, J. B. & Jansson, I. The many roles of cytochrome b_5. *Pharmacol. Ther.* **97,** 139–152 (2003).
10. Zhang, H., Im, S. C. & Waskell, L. Cytochrome b_5 increases the rate of product formation by cytochrome P450 2B4 and competes with cytochrome P450 reductase for a binding site on cytochrome P450 2B4. *J. Biol. Chem.* **282,** 29766–29776 (2007).
11. Teoh, K. H., Polichuk, D. R., Reed, D. W. & Covello, P. S. Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua. Botany* **87,** 635–642 (2009).
12. Teoh, K. H., Polichuk, D. R., Reed, D. W., Nowak, G. & Covello, P. S. *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett.* **580,** 1411–1416 (2006).
13. Daugulis, A. J. Partitioning bioreactors. *Curr. Opin. Biotechnol.* **8,** 169–174 (1997).
14. Li, Y., Huang, H. & Wu, Y.-L. in *Medicinal Chemistry of Bioactive Natural Products* (eds Liang, X.-T. & Fang, W.-S.) 183–256 (John Wiley and Sons, 2006).
15. Kim, B. & Sasaki, T. Recent progress in the synthesis of artemisinin and its derivatives. *Org. Prep. Proced. Int.* **38,** 1–80 (2006).
16. Haynes, R. K. From artemisinin to new artemisinin antimalarials: biosynthesis, extraction, old and new derivatives, stereochemistry and medicinal chemistry requirements. *Curr. Top. Med. Chem.* **6,** 509–537 (2006).
17. Sy, L.-K. & Brown, G. D. The role of the 12-carboxylic acid group in the spontaneous autoxidation of dihydroartemisinic acid. *Tetrahedron* **58,** 909–923 (2002).
18. Haynes, R. K. & Vonwiller, S. C. Catalysed oxygenatation of allylic hydroperoxides derived from qinghao (artemisinic) acid. Conversion of qinghao acid into dehydroqinghaosu (artemisitene) and qinghaosu (artemisinin). *J. Chem. Soc. Chem. Commun.* **6,** 451–453 (1990).
19. Roth, R. J. & Roth, N. A. Simple conversion of artemisinic acid into artemisinin. US patent 4,992 561 (1991).
20. Lévesque, F. & Seeberger, P. H. Continuous-flow synthesis of the anti-malaria drug artemisinin. *Angew. Chem. Int. Edn Engl.* **51,** 1706–1709 (2012).
21. Boehme, K. & Brauer, H.-D. Generation of singlet oxygen from hydrogen peroxide disproportionation catalyzed by molybdate ions. *Inorg. Chem.* **31,** 3468–3471 (1992).
22. Bertea, C. M. *et al.* Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in *Artemisia annua. Planta Med.* **71,** 40–47 (2005).
23. Nam, J. W. & Kappock, T. J. Cloning and transcriptional analysis of *Crepis alpina* fatty acid desaturases affecting the biosynthesis of crepenynic acid. *J. Exp. Bot.* **58,** 1421–1432 (2007).
24. Vorachek-Warren, M. K. & McCusker, J. H. DsdA (D-serine deaminase): a new heterologous MX cassette for gene disruption and selection in *Saccharomyces cerevisiae. Yeast* **21,** 163–171 (2004).
25. Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae. Yeast* **15,** 1541–1553 (1999).

Author Information Novel DNA sequences have been deposited in GenBank with the following accession numbers: native *A. annua* ADH1, JF910157; codon-optimized *A. annua* ADH1, JQ582842; codon-optimized *A. annua* CYB5, JQ582841; codon-optimized *A. annua* ALDH1, JQ609276. Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.J.P. (paddon@amyris.com) or J.D.N. (newman@amyris.com).

METHODS

The *A. annua* cytochrome b_5 cDNA sequence was identified from a trichome expressed sequence tag library (NCBI accession 35608) by searching for sequence similarity to *Crepis alpina* cytochrome b_5 type 11 (ref. 23). Dominant selection markers for yeast strain engineering were D-serine[24], nourseothricin[25] and hygromycin B[25].

Yeast strain engineering. *S. cerevisiae* codon-optimized synthetic genes of *A. annua ADS*, *CYP71AV1* and *CPR1* have been previously described[3]. Codon-optimized synthetic genes for *A. annua CYB5* (GenBank accession JQ582841), *A. annua ALDH1* (JQ609276) and *A. annua ADH1* (JQ582842) were synthesized by DNA 2.0 (https://www.dna20.com/) or Biosearch Technologies.

Construction of genome integration cassettes. The oligonucleotide primers used in this study are listed in Supplementary Table 9.

dsdA-P_{CTR3}-*ERG9*. Replacement of the *MET3* promoter with the *CTR3* promoter in Y301 and Y592 was accomplished as follows. The *dsdA* gene (encoding D-serine deaminase) was amplified from pAM577 (containing the promoter and terminator of *Kluyveromyces lactis TEF1*) by PCR amplification with oligonucleotides PW91-031-CPK275-G and DE_PW91-027-CPK262-G. PCR amplification of the wild-type *CTR3* promoter from positions −1 to −734 was performed with oligonucleotides PW61-104-CPK116-G and DE_PW91-027-CPK263-G using CEN.PK2-1C (ref. 3) genomic DNA as the template. These two PCR products shared a 44-base-pair (bp) overlap at the 3′ end of the promoter and the 5′ end of the gene. For the secondary PCR, 25 ng each of the purified *CTR3* promoter (−1 to −734) and *dsdA* PCR were used as the DNA templates and PCR amplified with oligonucleotides PW91-031-CPK275-G and PW61-104-CPK116-G to give $P_{CTR3(−1\ to\ −734)}$-*dsdA*.

*gal1/10/7::natA*_P_{GAL3}-*CPR1*-T_{CYC1}. Targeted integration of the P_{GAL3}-*CPR1* expression cassette in Y657 at the *GAL7* locus was accomplished as follows. PCR amplification of the wild-type *GAL7* locus from positions 30 to 1021 was performed with oligonucleotides PW91-014-CPK236-G and PW-91-079-CPK384-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *CPR1* ORF and *CYC1* terminator was performed with oligonucleotides PW-91-079-CPK385-G and PW-91-079-CPK392-G using plasmid pAM322 (ref. 3) as the template. PCR of the wild-type *GAL3* promoter from positions −1 to −660 was performed with oligonucleotides PW-91-079-CPK393-G and PW-91-079-CPK394-G using CEN.PK2-1C genomic DNA as the template. PCR of the *natA* marker (nourseothricin resistance) was performed with oligonucleotides PW-91-079-CPK383-G and PW-91-079-CPK395-G. Each of the DNA elements from the first round of PCR was designed to share a 20–30-bp overlap with the adjacent element, by using non-templated tails on the oligonucleotides. For the secondary PCR amplification, 25 ng each of the purified *GAL7*, *CPR1*-*CYC1*, *GAL3* promoter and *natA* PCR products were used as the DNA template and PCR amplified with oligonucleotides PW91-014-CPK236-G and PW-91-079-CPK383-G to give *GAL7(30 to 1021)*_$P_{GAL3(−1\ to\ −660)}$-*CPR1*-T_{CYC1}_*natA*.

*leu2::hisMXΔ::kanA*_P_{GAL7}-*CYB5*-T_{CYC1}. Targeted replacement of the *leu2::hisMX* locus in Y657 was accomplished as follows. PCR amplification of the wild-type *ERG19* locus from positions 489 to 1341 was performed with oligonucleotides AM/PW-91-093-CPK461-G and AM/PW-91-093-CPK462-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *kanA* marker (G418 resistance) was performed with oligonucleotides AM/PW-91-093-CPK460-G and AM-125-50-CPK514-G using pAM575 (containing the promoter and terminator of *K. lactis TEF1*) as the template. PCR amplification of the *GAL7* promoter from positions −1 to −725 was performed with oligonucleotides AM-125-50-CPK513-G and AT-126-103-CPK593-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *S. cerevisiae* codon-optimized *A. annua CYB5* ORF was performed with oligonucleotides AT-126-103-CPK592-G and PW-91-093-CPK426-G. PCR amplification of the *CYC1* terminator (T_{CYC1}) from positions 331 to 830 was performed with oligonucleotides PW-91-093-CPK425-G and AT-126-103-CPK595-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *LEU2* locus from positions 1 to 450 was performed with oligonucleotides AT-126-103-CPK594-G and AM/PW-91-093-CPK457-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR, 25 ng each of the purified *ERG19(489 to 1341)*, *kanA*, $P_{GAL7(−1\ to\ −725)}$, *A. annua CYB5*, $T_{CYC1(331\ to\ 830)}$, and *LEU2(1 to 450)* PCR products were used as the DNA template and PCR amplified with oligonucleotides AM/PW-91-093-CPK462-G and AM/PW-91-093-CPK457-G to give *ERG19(489 to 1341)*_*kanA*_$P_{GAL7(−1\ to\ −725)}$-*CYB5*-$T_{CYC1(331\ to\ 830)}$_*LEU2(1 to 450)*.

ndt80Δ::P_{TDH1}-*HEM1*_*hphA*_P_{PGK1}-*CTT1*. Targeted replacement of the *NDT80* locus in Y1368 with constitutively expressed *CTT1* and *HEM1* was accomplished as follows. PCR amplification of the wild-type *NDT80* locus from positions −187 to −951 was performed with oligonucleotides PW-091-144-CPK640-G and PW-091-144-CPK654-G using CEN.PK2-1C genomic DNA as the template. PCR of the wild-type *HEM1* locus from positions 1 to 1947 was performed with oligonucleotides PW-091-144-CPK655-G and PW-091-144-CPK656-G using CEN.PK2-1C genomic DNA as the template. PCR of the *TDH1* promoter (P_{TDH1}) from positions −1 to −577 was performed with oligonucleotides PW-091-144-CPK657-G and PW-091-144-CPK658-G using CEN.PK2-1C genomic DNA as the template. PCR of the *hphA* marker was performed with oligonucleotides PW-091-144-CPK659-G and PW-091-144-CPK643-G using BY4710 (ref. 4) genomic DNA as the template. PCR of the *PGK1* promoter (P_{PGK1}) from positions −1 to −623 was performed with oligonucleotides PW-091-144-CPK644-G and PW-091-144-CPK645-G using CEN.PK2-1C genomic DNA as the template. PCR of the *CTT1* locus from positions 1 to 2000 was performed with oligonucleotides PW-091-144-CPK646-G and PW-091-144-CPK647-G using CEN.PK2-1C genomic DNA as the template. PCR of the wild-type *NDT80* locus from positions 1684 to 2470 was performed with oligonucleotides PW-091-144-CPK648-G and PW-091-144-CPK649-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR, 25 ng each of the purified *NDT80(−187 to −951)*, *HEM1(1 to 1947)*, $P_{TDH1(−1\ to\ −577)}$, *hphA*, $P_{PGK1(−1\ to\ −623)}$, *CTT1(1 to 2000)* and *NDT80(1684 to 2470)* PCR products were used as the DNA template and PCR amplified with oligonucleotides PW-091-144-CPK640-G and PW-091-144-CPK649-G to give *NDT80(−187 to −951)*_$P_{TDH1(−1\ to\ −577)}$-*HEM1(1 to 1947)*_*hphA*_$P_{PGK1(−1\ to\ −623)}$-*CTT1(1 to 2000)*_*NDT80(1684 to 2470)*.

*ndt80::hphA*_P_{GAL7}-*ALDH1*-T_{TDH1}. Targeted replacement of the *NDT80* locus with P_{GAL7}-*ALDH1* in Y973 was accomplished as follows. PCR amplification of the wild-type *NDT80* locus from positions −187 to −951 was performed with oligonucleotides PW-091-144-CPK640-G and PW-091-144-CPK641-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *hphA* marker was performed with oligonucleotides PW-091-144-CPK642-G and AM-125-50-CPK514-G using pAM578 pAM575 (containing the promoter and terminator of *K. lactis TEF1*) as the template. PCR amplification of the *GAL7* promoter (P_{GAL7}) from positions −1 to −725 was performed with oligonucleotides AM-125-50-CPK513-G and AM-125-107-CPK756-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *S. cerevisiae* codon-optimized *ALDH1* ORF was performed with oligonucleotides AM-125-107-CPK754-G and AM-125-107-CPK755-G. PCR amplification of the *TDH1* terminator (T_{TDH1}) from positions 1000 to 1997 was performed with oligonucleotides AM-125-107-752G and AM-125-107-CPK753-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the wild-type *NDT80* locus from positions 1684 to 2470 was performed with oligonucleotides AM-125-107-CPK751-G and PW-091-144-CPK649-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR amplification, 25 ng each of the purified *NDT80(−187 to −951)*, *hphA*, $P_{GAL7(−1\ to\ −725)}$, *A. annua ALDH1*, $T_{TDH1(1000\ to\ −1997)}$ and *NDT80(1684 to 2470)* PCR products were used as the DNA templates and PCR amplified with oligonucleotides PW-091-144-CPK640-G and PW-091-144-CPK649-G to give *NDT80(−187 to −951)*_*hphA*_$P_{GAL7(−1\ to\ −725)}$-*ALDH1*-$T_{TDH1(1000\ to\ −1997)}$_*NDT80(1684 to 2470)*.

*his3::hphA*_P_{GAL7}-*ALDH1*-T_{TDH1}. Targeted replacement of the *his3::hisMX* locus with P_{GAL7}-*ALDH1* to create Y1368 was accomplished as follows. PCR amplification of the wild-type *HIS3* locus from positions −32 to −630 was performed with oligonucleotides PW-91-129-CPK543-G and PW-91-129-CPK544-G using BY4710 genomic DNA as the template. PCR of the *hphA* marker was performed with oligonucleotides PW-91-129-CPK545-G and AM-125-50-CPK514-G using pAM578 as the template. PCR amplification of the *GAL7* promoter (P_{GAL7}) from positions −1 to −725 was performed with oligonucleotides AM-125-50-CPK513-G and AM-125-107-CPK756-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *A. annua ALDH1* ORF was performed with oligonucleotides AM-125-107-CPK754-G and AM-125-107-CPK755-G using a synthetic, *S. cerevisiae* codon-optimized template (DNA2.0). PCR amplification of the *TDH1* terminator (T_{TDH1}) from positions 1000 to 1997 was performed with oligonucleotides PW-191-015-CPK859-G and AM-125-107-CPK753-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the wild-type *ERG12* locus from positions 883 to 1456 was performed with oligonucleotides PW-191-015-CPK860-G and PW-91-129-CPK550-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR amplification, 25 ng each of the purified *HIS3(−32 to −630)*, *hphA*, $P_{GAL7(−1\ to\ −725)}$, *A. annua ALDH1*, $T_{TDH1(1000\ to\ −1997)}$ and *ERG12(883 to 1456)* PCR products were used as the DNA templates and PCR amplified with oligonucleotides PW-91-129-CPK543-G and PW-91-129-CPK550-G to give *NDT80(−187 to −951)*_*hphA*_$P_{GAL7(−1\ to\ −725)}$-*ALDH1*-$T_{TDH1(1000\ to\ −1997)}$_*ERG12(883 to 1456)*.

*natAΔ::URA3*_P_{GAL7}-*ADH1*-T_{TDH1}. Targeted replacement of the *natA* locus with P_{GAL7}-*ADH1* for creation of Y1283 was accomplished as follows. PCR amplification of the wild-type *GAL3* promoter (P_{GAL3}) from positions −77 to −660 was performed with oligonucleotides PW-191-015-CPK866-G and

PW-191-015-CPK867-G using BY4710 genomic DNA as the template. PCR amplification of the *URA3* locus from position −226 to 884 was performed with oligonucleotides PW-191-015-CPK868-G and PW-191-015-CPK869-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *GAL7* promoter (P_{GAL7}) from positions −1 to −725 was performed with oligonucleotides PW-191-015-CPK870-G and PW-191-015-CPK871-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *S. cerevisiae* codon-optimized *A. annua ADH1* ORF was performed with oligonucleotides PW-191-015-CPK872-G and PW-191-015-CPK873-G. PCR amplification of the *TDH1* terminator (T_{TDH1}) from positions 1000 to 1750 was performed with oligonucleotides PW-191-015-CPK874-G and PW-191-015-CPK875-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the wild-type *GAL1* locus from positions 1637 to 2436 was performed with oligonucleotides PW-191-015-CPK876-G and PW-191-015-CPK877-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR amplification, 25 ng each of the purified $P_{GAL3(−77\ to\ −660)}$, $URA3(−226\ to\ 884)$, $P_{GAL7(−1\ to\ −725)}$, *A. annua ADH1*, $T_{TDH1(1000\ to\ −1750)}$ and *GAL1(1637 to 2436)* PCR products were used as the DNA templates and PCR amplified with oligonucleotides PW-191-015-CPK866-G and PW-191-015-CPK877-G to give $P_{GAL3(−77\ to\ −660)}$_$URA3(−226\ to\ 884)$_ $P_{GAL7(−1\ to\ −725)}$- ADH1-$T_{TDH1(1000\ to\ −1750)}$-*GAL1(1637 to 2436)*.

gal80Δ::URA3_P_{GAL7}-ADH1-T_{GAL80}. Targeted replacement of the *GAL80* locus with P_{GAL7}-ADH1 to create Y1284 was accomplished as follows. PCR amplification of the wild-type *GAL80* locus from positions −28 to −760 was performed with oligonucleotides PW-191-015-CPK882-G and PW-191-015-CPK883-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *URA3* locus from position −226 to 884 was performed with oligonucleotides PW-191-015-CPK884-G and PW-191-015-CPK869-G using BY4710 genomic DNA as the template. PCR of the *GAL7* promoter (P_{GAL7}) from positions −1 to −725 was performed with oligonucleotides PW-191-015-CPK870-G and PW-191-015-CPK871-G using CEN.PK2-1C genomic DNA as the template. PCR amplification of the *A. annua ADH1* ORF was performed with oligonucleotides PW-191-015-CPK872-G and PW-191-015-CPK873-G using a synthetic *S. cerevisiae* codon-optimized template (DNA2.0). PCR amplification of the wild-type *GAL80* locus from positions 1320 to 2117 was performed with oligonucleotides PW-191-015-CPK886-G and PW-191-015-CPK887-G using CEN.PK2-1C genomic DNA as the template. For the secondary PCR amplification, 25 ng each of the purified *GAL80(−28 to −760)*, *URA3(−226 to 884)*, $P_{GAL7(−1\ to\ −725)}$, *A. annua ADH1*, and *GAL80(1320 to 2117)* PCR products were used as the DNA templates and PCR amplified with oligonucleotides PW-191-015-CPK882-G and PW-191-015-CPK887-G to give *GAL80(−28 to −760)*_*URA3(−226 to 884)*_$P_{GAL7(−1\ to\ −725)}$-ADH1-*GAL80(1320 to 2117)*.

All strains were confirmed with diagnostic PCR to contain the expected integration constructs and, where appropriate, all integrations were verified by sequence analysis.

Cloning and characterization of *A. annua* ADH1. Analysis of a previously developed *A. annua* EST collection[12,26,27] identified a contig corresponding to an apparently full-length ORF encoding a putative trichome-expressed alcohol dehydrogenase. The corresponding gene, designated *A. annua ADH1*, was associated with 2.2%, 1.3% and 0.06% of ESTs in the 'trichome-minus-flower-bud' (designated GSTSUB in ref. 12), glandular trichome (designated AAGST[12]) and flower bud (designated AAFB[12]) collections, respectively. Similarly, based on the generation of expressed sequence tags by 454 sequencing of *A. annua*[28], the gene expression pattern of *ADH1* was found to be comparable to that of *CYP71AV1* in a range of tissues, with negligible expression in cotyledons and mature leaf trichomes and 0.21, 0.53 and 0.03% of sequences in each of the EST collections derived from young leaf trichomes, flower bud trichomes and meristem/young leaf, respectively. A full-length *ADH1* ORF was cloned by reverse transcriptase PCR (RT–PCR) (using oligonucleotide primers PSC1 and PSC2 and the vector pENTR/D TOPO (Invitrogen)). The *A. annua ADH1* gene has an ORF encoding a polypeptide of 378 amino acids with a relative molecular mass of 40,415 daltons. On the basis of sequence similarities, *A. annua* ADH1 is a member of the medium chain alcohol dehydrogenase/reductase superfamily that is related to predicted proteins of *Populus trichocarpa* (61% identity, GenBank accession XP_002324694) and *Cynara cardunculus* (72% identity over 214 amino acids, GenBank accession GE588275).

The *A. annua ADH1* ORF was subcloned into the pET15b vector modified to contain a PreScission protease cleavage site. The vector containing the *A. annua ADH1* ORF was used to transform the *E. coli* BLR (DE3). Protein expression was induced by adding 0.4 mM isopropyl-β-D-thiogalactoside (IPTG) and cultures totalling 5 l were incubated at 16 °C for 16 h. ADH1 was subjected to two rounds of Ni-column purification, analysis by SDS–PAGE and dialysis. The final fractions containing ADH1 were pooled and dialysed against protein storage buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 200 mM KCl, 10% glycerol and 1 mM dithiothreitol (DTT)). Protein concentration was determined by Bradford assay and aliquots were stored at −80 °C. ADH1 purity by SDS–PAGE was judged to be 95%.

Unless otherwise stated, ADH1 enzyme assays included 50 mM Tris buffer, pH 8.5, 250 mM NaCl, 0.4 mg ml⁻¹ BSA, 50 μM substrate, 1 mM NAD, 3 μg of octadecane (as internal standard; Sigma-Aldrich) and 80 ng of recombinant *A. annua* ADH1 in a total volume of 200 μl. Negative controls were carried out in the absence of NAD. Reactions were allowed to proceed for 4 min at 30 °C with shaking (500 r.p.m.), and immediately stopped by extraction with 500 μl pentane. All quantitative analyses were done with 3–6 technical replicates per treatment. Pentane extracts were concentrated to ∼30 μl under a stream of nitrogen and either 10 μl ethyl acetate or 10 μl of a mixture of 1:1 *N,O*-bis-(trimethylsilyl)acetamide (Sigma-Aldrich)/pyridine (Fluka) was added. The remainder of the pentane was carefully removed under a stream of nitrogen and the final 10-μl sample was analysed by gas chromatography–mass spectrometry (GC–MS)[26].

Substrate specificity was determined in 15-min, 600-μl assays using (+)-borneol (Fluka), (−)-borneol (Fluka) and artemisinic, dihydroartemisinic, artemisia[26], coniferyl (Sigma-Aldrich), and cinnamyl (Sigma-Aldrich) alcohols. ADH1 only showed considerable dehydrogenase activity with artemisinic alcohol and to a lesser extent with dihydroartemisinic alcohol (4.2% relative to artemisinic alcohol; Supplementary Fig. 5). The identity of the aldehyde products was confirmed by GC–MS in comparison to authentic standards. When assayed with artemisinic alcohol and NAD, recombinant *A. annua* ADH1 showed a pH optimum of 8.5. Using 1 mM NADP as the cofactor, oxidation of artemisinic alcohol by ADH1 was 30-fold lower (Supplementary Fig. 5). The linear range of the ADH1 assay with respect to time was tested by reactions under standard assay conditions except varying the time up to 30 min. The pH optimum of the purified ADH1 was determined to be 8.5 based on a series of 15-min assays with the pH range from 5.5 to 10 in intervals of 0.5-pH units using 50 mM citrate, phosphate, Tris, CHES and CAPS buffers. Kinetic parameters were determined by varying the concentrations of artemisinic alcohol (3.0–25 μM) in 600-μl assays using 240 ng ADH1. Substrate solubility prevented the use of higher concentrations. Octadecane was used as an internal standard to quantify the substrate and product from the reactions by gas chromatography using response factors determined by using known concentrations of standards. Kinetic constants were determined by fitting the data to the Michaelis–Menten equation using nonlinear regression and EnzFitter software (Biosoft).

Media and growth conditions. Fermentation media: The media used for this work were based on media described previously[29]. The trace metal solution contained 5.75 g l⁻¹ ZnSO₄·7H₂O, 0.32 g l⁻¹ MnCl₂·4H₂O, 0.47 g l⁻¹ CoCl₂·6H₂O, 0.48 g l⁻¹ Na₂MoO₄·2H₂O, 2.9 g l⁻¹ CaCl₂·2H₂O, 2.8 g l⁻¹ FeSO₄·7H₂O and 80 ml l⁻¹ 0.5 M EDTA, pH 8.0. The vitamin solution contained 0.05 g l⁻¹ biotin, 1 g l⁻¹ calcium pantothenate, 1 g l⁻¹ nicotinic acid, 25 g l⁻¹ *myo*-inositol, 1 g l⁻¹ thiamine HCl, 1 g l⁻¹ pyridoxal HCl and 0.2 g l⁻¹ *p*-aminobenzoic acid. The batch medium for all fermentations contained 19.5 g l⁻¹ glucose, 15 g l⁻¹ (NH₄)₂SO₄, 8 g l⁻¹ KH₂PO₄, 6.2 g l⁻¹ MgSO₄·7H₂O, 12 ml l⁻¹ vitamin solution and 10 ml l⁻¹ trace metal solution.

The batch medium also contained additional components depending on the strain and the process being run. For the glucose and ethanol mixed-feed process, for all strains except Y285, CuSO₄ was added to the batch medium to a concentration of 0.25 μM CuSO₄. For Y285, the batch medium contained 20 μM CuSO₄.

The bioreactor feed media also varied for the different processes and for different strains. All mixed glucose/ethanol processes used bioreactor feed base that contained 386 g l⁻¹ glucose, 9 g l⁻¹ KH₂PO₄, 5.12 g l⁻¹ MgSO₄·7H₂O, 3.5 g l⁻¹ K₂SO₄, 0.28 g l⁻¹ Na₂SO₄ and 237 ml l⁻¹ ethanol (95% v/v).

For the glucose and ethanol mixed feed process, two different feed media were prepared for the fermentation: pre-induction-feed media and induction-feed media (includes small molecule inducers and repressors). For both feed media, stock solutions of vitamins and trace metals were added to the bioreactor feed base as follows: 12 ml vitamin solution per litre feed base, and 10 ml trace metals solution per litre feed base. The pre-induction-feed media contained 0.25 μM CuSO₄ and no other additions. For fermentation of Y285, the pre-induction-feed media contained 20 μM CuSO₄.

To the induction-feed medium, two different inducers/repressors were added to the medium dependent on the strain. For Y285, concentrated solutions of galactose and methionine were added to the induction-feed medium to bring the final concentrations to 10 g l⁻¹ galactose and 1 g l⁻¹ methionine. For all other strains, concentrated solutions of galactose and CuSO₄ were added to the feed medium to bring the final concentrations to 10 g l⁻¹ galactose and 150 μM CuSO₄. All additions to the medium were made in a sterile hood.

Shake-flask media: Seed medium for pre-culture was the batch fermentation medium modified with the addition of 100 ml l⁻¹ succinate buffer (0.5 M, pH 5.0). For strain Y285, the concentration of CuSO₄ in the seed medium was 20 μM. For

all other strains tested (which contain a copper-repressible promoter controlling ERG9 expression), low-copper seed medium, containing only 0.25 µM CuSO$_4$, was used.

Flask-production medium was modified seed media which contained 40 g l^{-1} glucose, 5 g l^{-1} galactose, 1.7 mM methionine and 150 µM CuSO$_4$.

Shake-flask methods. To prepare seed vials, single isolates of each strain from agar plates were grown for 18–24 h in 20-ml low-copper seed medium containing 0.25 µM CuSO$_4$. Cultures were then inoculated at an attenuance ($D_{600\,nm}$) of 0.05 into fresh low-copper seed medium and grown for a further 18–24 h to an $D_{600\,nm}$ of between 2 and 3 (measured using a Thermo Scientific Genesys 10 Vis spectrophotometer). Six-hundred microlitres of this culture was added to 400 µl of 50% glycerol and stored in 1-ml aliquots (20% glycerol (v/v) final) at −80 °C.

To acclimate cells before inoculation into production medium, frozen seed vials were thawed to room temperature and inoculated into 20-ml low-copper seed medium. Cultures were grown for 18–24 h at 30 °C with shaking at 200 r.p.m. The next day, the cultures were diluted to a $D_{600\,nm}$ of 0.05 in 20-ml low-copper seed medium and grown for ~18 h at 30 °C with shaking at 200 r.p.m.

Cells from the second overnight acclimation were diluted to a $D_{600\,nm}$ of 0.05 in 250 ml unbaffled flasks containing 25 ml flask-production medium. Flasks contained an additional 5 ml IPM where indicated. All cultures were inoculated in triplicate and incubated at 30 °C for 72 h with shaking at 200 r.p.m. in a humidified Innova incubator. Flasks were sampled periodically for growth ($D_{600\,nm}$), viability and product titres. Viability was measured using the LIVE/DEAD Funga Light yeast viability kit for flow cytometry (Invitrogen Corporation) and a Guava technologies EasyCyte Plus flow cytometer.

For the production of insoluble artemisinic acid in shake-flask cultures, 15.8 g l^{-1} of 95% ethanol was added to the production flask after 72, 96 and 120 h growth. The flask was inspected for formation of insoluble material after 144 h.

Glucose and ethanol mixed-feed process. Preparation of seed cultures and procedures for setting up and running glucose and ethanol mixed-feed fermentations have been described[3]. The production process was induced with the addition of 10 g l^{-1} galactose and 0.25 g l^{-1} methionine (for Y285), or 10 g l^{-1} galactose and 150 µM CuSO$_4$ (all other strains) to the bioreactor after the culture reached an $D_{600\,nm}$ value of approximately 50. At this time, the feed bottle containing pre-induction-feed medium was exchanged for a feed bottle containing induction-feed medium.

Mixed glucose/ethanol feed process with early induction/repression. The mixed glucose/ethanol feed process was modified by changing the time of induction (the time of addition of galactose and methionine, or galactose and high CuSO$_4$) from the time the culture reached a $D_{600\,nm}$ of 50 to the time of inoculation. The mixed glucose/ethanol feed process with early induction/repression was identical to the mixed feed process except that batch and feed media were modified. Immediately before inoculation, concentrated solutions of galactose, methionine and/or CuSO$_4$ were added to the batch medium to bring the final concentrations to 10 g l^{-1} galactose and 0.25 g l^{-1} methionine (for Y285), or 10 g l^{-1} galactose and 150 µM CuSO$_4$ (for all other strains). Only the induction-feed medium was used in the fed-batch phase of the process (no pre-induction medium). All other parameters were the same as the mixed glucose/ethanol feed process.

At later time points (42–96 h), in some fermentation runs of strains Y1283 and Y1284, artemisinic acid precipitated from the liquid fermentor broth. Solid precipitate was visible in the fermentor and adhered onto the side of the bioreactor and the head plate. During the runs, artemisinic acid concentration was still assayed from fermentor broth samples over the course of the fermentation as described below. However, for select fermentations of Y1284, artemisinic acid was also assayed at the end of the fermentation after complete solubilization of the precipitate by high pH treatment of the fermentor broth. At the end of the fermentation, the culture was adjusted to pH 8.1 with 10 M NH$_4$OH and allowed to stir at 1,500 r.p.m. (maximum rpm) at 30 °C for at least 1 h to dissolve the precipitated artemisinic acid. After the pH adjustment, the cell broth was collected and water was added the tank to wash any residual precipitate from the tank. The water was adjusted to pH 9.1 with 10 M NH$_4$OH and allowed to stir at 1,200 r.p.m. overnight. Together, this provided a more accurate measurement of artemisinic acid at the final time point.

Mixed glucose/ethanol feed process with induction/repression and IPM. The addition of an IPM phase to yeast cultures was tested at the fermentor scale using strains Y1283 and Y1284. The fermentation process used was the mixed glucose/ethanol feed process with early induction/repression. The only process change was the addition of 200 ml IPM to the fermentor before inoculation. The initial aqueous batch volume of the fermentations was 0.7 l.

The IPM phase and the aqueous cell broth formed a well-mixed emulsion in the reactor at later times in the fermentation (>24 h). To assay artemisinic acid titre in fermentations with IPM, samples of the combined IPM and cell broth mixture/emulsion were extracted with solvent. The mixture was first vortexed then added to the methanol/formic acid as described below. The concentrations measured by liquid chromatography with ultraviolet detection are reported in terms of grams per litre total volume (aqueous cell broth plus IPM). Using the ratio of aqueous cell broth volume to IPM volume at the time of sampling, the titres are converted to terms of grams per litre aqueous volume to allow for direct comparison with runs that do not use IPM.

At 30 °C, artemisinic acid has a solubility of approximately 100–115 g l^{-1} in IPM (empirically determined). At later times in the fed-batch fermentation, after a large volume of aqueous feed has been added to the fermentor, the ratio of IPM/aqueous volume is significantly lower and the solubility limit could restrict additional production.

Ethanol pulse-feed process with IPM. Yeast strain Y1284 was tested in an ethanol pulse-feed process with the addition of IPM to the culture medium. The temperature, pH and dissolved oxygen were controlled at the set points described above. The batch medium for this process was the same as for the glucose/ethanol mixed-feed process with early induction, described above, except that no galactose was added (Y1284 does not require galactose for induction). Four-hundred millilitres of IPM was added to a starting aqueous fermentor volume of 0.8 l, before inoculation.

The feed for the ethanol pulse-feed fed-batch phase of the process was 95% (v/v) ethanol. Because none of the salts, trace metals or vitamins was soluble in 95% (v/v) ethanol, concentrated feed components were combined into a concentrated post-sterile addition (PSA) solution. The concentrated PSA solution consisted of 72.9 g l^{-1} KH$_2$PO$_4$, 41.4 g l^{-1} MgSO$_4$·7H$_2$O, 28.3 g l^{-1} K$_2$SO$_4$, 2.3 g l^{-1} Na$_2$SO$_4$, 1.2 mM CuSO$_4$, 10 ml l^{-1} trace metals solution and 12 ml l^{-1} vitamin solution. The concentrated PSA solution was injected through a septum in the bioreactor head plate with a syringe once per day according to how much volume of 95% (v/v) ethanol volume had been delivered since the previous addition of feed components. One-hundred-and-twenty-four millilitres of concentrated PSA solution was added per litre of 95% (v/v) ethanol added.

After the batch carbon was consumed (detected as described above) the ethanol pulse-feed algorithm was initiated. As the culture grew and consumed O$_2$, dissolved O$_2$ was maintained at 40% by an agitation cascade followed by oxygen enrichment (as described above). In the first phase of the fed-batch fermentation, before the stir rate of the reactor reached the maximum allowed for the unit, the pulse feed algorithm used stir rate (Stir) measurements to control ethanol feed delivery (Supplementary Fig. 14a). The computer algorithm assigned a variable (Stir Max) that tracked the maximum stir rate obtained so far in the process. While growing on ethanol, O$_2$ demand increased and stir rate increased until the substrate was depleted from the fermentor medium. At that point, the dissolved O$_2$ increased and the controller decreased stir rate to maintain dissolved O$_2$ = 40%. When Stir decreased to less than 75% of the value of Stir Max, the ethanol feed pump was activated for the length of time necessary to add 10 g ethanol per litre fermentor volume to the reactor. The computer algorithm calculated the time necessary to add 10 g ethanol per litre fermentor volume to the reactor (Timer Max) after each cycle. The first phase of the algorithm iterated unit O$_2$ enrichment as required.

After the stir rate of the reactor reached the maximum allowed for the unit, oxygen enrichment was used to maintain dissolved O$_2$ = 40%. During this stage of the fed-batch fermentation, the second phase of the control algorithm was initiated. Dissolved O$_2$ measurements were used to control ethanol feed delivery (Supplementary Fig. 14b). When ethanol was depleted from the fermentor medium, the dissolved O$_2$ began to increase rapidly—faster than the dissolved O$_2$ controller could compensate. When dissolved O$_2$ > 50%, the ethanol feed pump was activated for the length of time necessary to add 10 g ethanol per litre fermentor volume to the reactor. After the addition of ethanol, the dissolved O$_2$ would rapidly decrease to <50%. The variable Timer Max was again calculated by the computer algorithm after each cycle. This algorithm iterated for the remainder of the fermentation.

Purification of artemisinic acid from IPM. IPM was isolated from artemisinic acid fermentations by centrifugation. IPM was mixed with 1% NaH$_2$PO$_4$·12H$_2$O and the pH was adjusted to 10.7 by the addition of 5 M NaOH. The solution was then stirred at ambient temperature for 60 min. After mixing, the solution was allowed to separate by gravity in a separatory funnel at ambient temperature. The bottom aqueous phase was drawn off from the upper IPM phase. The bottom aqueous phase was run through a liquid: liquid annular centrifugal contactor (CINC Industries) to ensure complete removal of any residual IPM. A 10% (w/v) SDS solution was added to the aqueous phase to bring the final SDS concentration to 0.03%. The solution was mixed and the pH adjusted to 5.0 with 2.5 M H$_2$SO$_4$. The acidification resulted in the formation of a fine white precipitate, which was captured on a 0.45-µm PTFE (polytetrafluoroethylene) filter, rinsed

with purified water and then dried. Analysis of the IPM before and after aqueous extraction showed that 5% of the artemisinic acid remained in the IPM after extraction (~95% step yield). Analysis of the filtrate after precipitation showed that 2% of the artemisinic acid present in the aqueous phase remained in the filtrate after acidification (~98% step yield). The overall purification yield obtained was ~93%. Additional aqueous extractions of the remaining IPM should increase the overall yield. Analysis of the dried precipitate by gas chromatography with flame-ionization detection (GC–FID) gave artemisinic acid purities of ~96% by area and ~98% by weight.

Broth extraction. Amorpha-4,11-diene, artemisinic alcohol and artemisinic aldehyde were extracted from cells and broth as follows. Cell lysis cocktail was prepared by combining two parts Novagen YeastBuster protein reagent (EMD Biosciences) and one part 2 M HCl. Samples were prepared by mixing 0.4 ml cell lysis cocktail with 0.1 ml whole broth and 1 ml ethyl acetate containing 10 mg l^{-1} *trans*-caryophyllene (internal standard, ≥98.5% purity; Sigma-Aldric) in a 2-ml glass vial. The sample was mixed for 30 min on a vortex mixer. After mixing, the vial was placed on the bench top to allow the phases to separate. If necessary, the vial was centrifuged at 1,000g to break any emulsion that had formed. Six-hundred microlitres of the ethyl acetate layer was transferred to a gas chromatography vial for analysis.

Gas chromatography. The production of amorpha-4,11-diene, artemisinic alcohol and artemisinic aldehyde was monitored by GC–FID. The ethyl acetate-extracted samples were analysed using on the GC–FID. Amorpha-4,11-diene, artemisinic alcohol and A.CHO peak areas were converted to concentration measurements from external standard calibrations using authentic compounds. To expedite run times, the temperature program and column were modified to achieve optimal resolution and the shortest overall run-time with minimal interferences. A 10-µl sample was split 1:20 and was separated using a DB-WAX column (50 m × 200 µm × 0.2 µm; Agilent), with hydrogen as the carrier gas at a flow rate of 1.57 ml min^{-1}. The temperature program for the analysis was as follows: the column was initially held at 150 °C for 3 min, followed by a temperature gradient of 5 °C min^{-1} to a temperature of 250 °C, and then the column was held at 250 °C for 5 min to elute all remaining components. Under these conditions, *trans*-caryophyllene, amorpha-4,11-diene, artemisinic aldehyde and artemisinic alcohol elute at 4.95, 5.77, 12.94 and 18.60 min, respectively.

Broth preparation with and without IPM. A 1-ml aliquot of well-mixed fermentation broth was diluted in 9 ml of methanol plus 0.1% formic acid (IPM formed an emulsion with the cell broth when it was used). The mixture was then mixed on a vortex mixer for 30 min and centrifuged at 16,000g for 5 min. One-hundred microlitres of the supernatant was diluted into 900 µl methanol plus 0.1% formic acid, and analysed by the HPLC method described below.

In-process assay for titre measurement. A screening method was developed to rank artemisinic-acid-producing strains. This method was used only to rank strains, and not determine final titre. A 20-µl aliquot was injected on an Agilent 1200 HPLC with ultraviolet detection at 212 nm. An Supelco Discovery C$_8$ column (4.6 mm × 100 mm × 5.0 µm; Supelco) equipped with the appropriate guard column (4.0 mm × 20.0 mm; Supelco) was used for separation, with the following gradient at a flow rate of 1 ml min^{-1} (channel A: water plus 0.1% formic acid; channel B: methanol plus 0.1% formic acid): 0–0.5 min 70% B, gradually increased to 97% B from 0.5 to 6.7 min, held at 97% B until 7 min, decreased to 70% B from 7 to 7.5 min, and re-equilibrated to 70% B from 7.5 to 9.5 min. The column was held at 25 °C during the separation. Under these conditions, artemisinic acid was found to elute at 6.3 min. AA peak areas were converted to concentrations from external standard calibrations of authentic compounds.

Final titre measurement. A 20-µl aliquot was injected on an Agilent 1200 HPLC with ultraviolet detection at 212 nm. An Agilent Eclipse XDB-C$_{18}$ column (4.6 mm × 100 mm × 3.5 µm, Agilent) equipped with the appropriate guard column (4.6 mm × 12.5 mm × 5 µm, Agilent) was used for separation, with the following gradient at a flow rate of 1 ml min^{-1} (channel A: water plus 0.1% formic acid; channel B: acetonitrile plus 0.1% formic acid): 0–10 min 50% B, gradually increased to 100% B from 10 to 21 min, held at 100% B until 26 min, decreased to 50% B from 26 to 26.1 min, and re-equilibrated to 50% B from 26.1 to 31 min. The column was held at 45 °C during the separation. Under these conditions, artemisinic acid was found to elute at 15.54 min. AA peak areas were converted to concentrations from external standard calibrations of authentic compounds.

26. Polichuk, D. R., Zhang, Y., Reed, D. W., Schmidt, J. F. & Covello, P. S. A glandular trichome-specific monoterpene alcohol dehydrogenase from *Artemisia annua*. *Phytochemistry* **71**, 1264–1269 (2010).
27. Zhang, Y. *et al.* The molecular cloning of artemisinic aldehyde Δ11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*. *J. Biol. Chem.* **283**, 21501–21508 (2008).
28. Graham, I. A. *et al.* The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* **327**, 328–331 (2010).
29. van Hoek, P., de Hulster, E., van Dijken, J. P. & Pronk, J. T. Fermentative capacity in high-cell-density fed-batch cultures of baker's yeast. *Biotechnol. Bioeng.* **68**, 517–523 (2000).