The Genetic Map of Artemisia annua L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin

Ian A. Graham, *et al.*

Science **327**, 328 (2010);
DOI: 10.1126/science.1182612

The Genetic Map of *Artemisia annua* L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin

Ian A. Graham,[1]* Katrin Besser,[1] Susan Blumer,[1] Caroline A. Branigan,[1] Tomasz Czechowski,[1] Luisa Elias,[1] Inna Guterman,[1] David Harvey,[1] Peter G. Isaac,[2] Awais M. Khan,[1] Tony R. Larson,[1] Yi Li,[1] Tanya Pawson,[1] Teresa Penfield,[1] Anne M. Rae,[1] Deborah A. Rathbone,[1] Sonja Reid,[1] Joe Ross,[1] Margaret F. Smallwood,[1] Vincent Segura,[1] Theresa Townsend,[1] Darshna Vyas,[1] Thilo Winzer,[1] Dianna Bowles[1]*

Artemisinin is a plant natural product produced by *Artemisia annua* and the active ingredient in the most effective treatment for malaria. Efforts to eradicate malaria are increasing demand for an affordable, high-quality, robust supply of artemisinin. We performed deep sequencing on the transcriptome of *A. annua* to identify genes and markers for fast-track breeding. Extensive genetic variation enabled us to build a detailed genetic map with nine linkage groups. Replicated field trials resulted in a quantitative trait loci (QTL) map that accounts for a significant amount of the variation in key traits controlling artemisinin yield. Enrichment for positive QTLs in parents of new high-yielding hybrids confirms that the knowledge and tools to convert *A. annua* into a robust crop are now available.

Malaria is a global health problem with more than 1 billion people living in areas with a high risk of the disease. Artemisinin combination therapies (ACTs) are the recommended treatment for uncomplicated malaria caused by the *Plasmodium falciparum* parasite (*1*). Parasite resistance to artemisinin has recently been confirmed in western Cambodia (*2*). It has long been recognized that the problem of artemisinin resistance is best addressed by increasing access to ACTs and discouraging the use of artemisinin monotherapies (*3*). This approach has strong support from the global health community with both funding and demand for ACTs expected to increase massively in the short- to midterm (*3*). However, there is growing concern that the supply chain will be unable to consistently produce high-quality artemisinin in the quantities that will be required (*3*). Artemisinin is a sesquiterpenoid synthesized in the glandular trichomes of the Chinese medicinal plant *Artemisia annua* L. (*4*–*10*). For a pharmaceutical with annual sales exceeding 100 million treatments, ACT supply remains reliant on the agricultural production of artemisinin. Plant-based production of artemisinin is challenging because *A. annua* remains relatively undeveloped as a crop. An alternative microbial-based system that synthesizes an artemisinin precursor for chemical conversion is in development (*11*, *12*). This would supplement but not replace agricultural production, which will continue to be an essential source of supply (*3*). Improved varieties of *A. annua* for developing-world farmers would bring immediate benefits to the existing artemisinin supply chain by reducing production costs, stabilizing supplies, and improving grower confidence in the crop (*3*).

[1]Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5YW, UK. [2]IDna Genetics Ltd., Norwich Research Park, Norwich NR4 7UH, UK.

*To whom correspondence should be addressed. E-mail: iag1@york.ac.uk (I.A.G.); djb32@york.ac.uk (D.B.)

A. annua is a member of the *Asteraceae* family that favors outcrossing over selfing (*13*). The artemisinin content of plants from different origins varies considerably and is highly heritable (*14*). The market leader for artemisinin production, at present, is Artemis, an F_1 hybrid (population) variety developed by Mediplant (Conthey, Switzerland) (*14*). Artemis seed is produced from a cross between two heterozygous and genetically different parental genotypes, called C4 and C1, that are themselves maintained vegetatively. In this study, we have used the Artemis pedigree to establish genetic linkage and QTL maps for this species and independently validated positive QTL for artemisinin yield.

We used the Roche 454 pyrosequencing platform to produce expressed sequence tag (EST) databases from cDNA libraries derived from enriched glandular trichome preparations of young leaves, mature leaves, and flower buds from the Artemis hybrid (*15*). cDNA libraries and EST databases were also prepared from meristem tissue (including very young leaf tissue) and cotyledons. A selection of key genes associated with metabolic pathways and phenotypic traits such as trichome development and plant architecture that could affect artemisinin yield are illustrated in fig. S1, together with their relative abundance in the different libraries (SOM Text and table S1). The EST sequences were also used for in silico identification of single-nucleotide polymorphisms (SNPs), short sequence repeats (SSRs), and insertions/deletions (InDels), which can be used as molecular markers for mapping and breeding (*15*). We identified 34,419 SNPs from DNA sequences contained in the five EST databases derived from the Artemis F_1 hybrid material, representing a mean SNP frequency of 1 in 104 base pairs

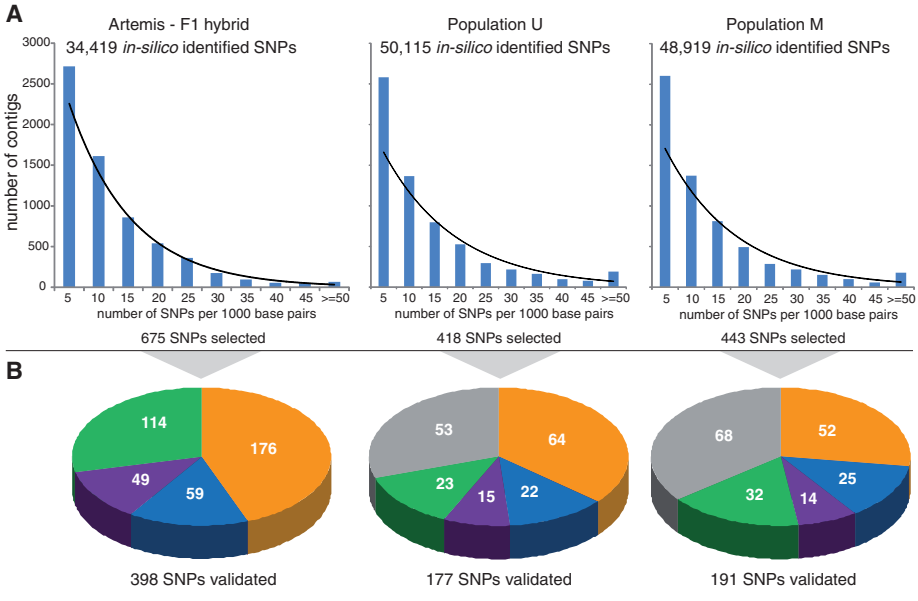

Fig. 1. High-throughput identification and validation of SNP markers in three *A. annua* populations. (**A**) Frequency distribution of potential SNPs identified in silico from EST databases produced by pyrosequencing cDNA libraries from the Artemis F_1 hybrid, Population U (commercially grown in Uganda) and Population M (commercially grown in Madagascar). The observed distribution of SNP frequency correlates closely with an exponential distribution for each data set as indicated by the curved black lines, which trace the expected distribution and R^2 values that are greater than 0.85 in all cases. Stringent selection criteria resulted in approximately only 10% of contigs being used for SNP identification (*15*). (**B**) Genotyping the Artemis pedigree. Subsets of in silico–identified SNPs from each population were selected for hybridization-based detection on the Illumina Goldengate Genotyping platform. The three pie charts show the genotyping of the Artemis pedigree with SNPs from Artemis, and Populations U and M. Color coding illustrates the proportion of SNPs that were polymorphic in the C4 parent (orange), polymorphic in the C1 parent (blue), polymorphic in both parents (purple), monomorphic in both parents for opposite alleles (green), and monomorphic in both parents for the same alleles (gray). This latter class is due to alleles being polymorphic in Population M or Population U but not in Artemis.

(Fig. 1A). This polymorphism was confirmed experimentally with 19 amplified fragment length polymorphism (AFLP) primer combinations that revealed 322 polymorphic markers (table S2). The in silico approach also identified 49 SSR markers that segregated in the Artemis F_1 population (table S3). We extended the in silico approach to two other *A. annua* populations, commercially grown in Uganda (Population U) and Madagascar (Population M), and found that the mean SNP frequencies of 1 in 88 and 1 in 91 base pairs, respectively, are only slightly higher than that of the Artemis F_1 hybrid (Fig. 1A).

We used the Illumina GoldenGate Genotyping platform to exploit this genetic resource, employing stringent criteria for selection of 1536 SNPs from the pool of 133,000 in silico–identified SNPs from Artemis and Populations U and M (*15*). The subset of SNPs represented candidate genes and their homologs, as well as others chosen randomly with the aim of having well-spaced markers for the genetic linkage map. We developed size-based markers in addition for 104 of the 1536 SNPs that allowed the two alleles in each case to be distinguished by capillary electrophoresis, and these further confirmed the segregation data derived from the Illumina platform (table S3). Genotyping the Artemis pedigree confirmed that extensive heterozygosity exists in the Artemis parents (Fig. 1B). Of SNPs derived from Populations U and M, 70% and 64%, respectively, were also found to segregate in the Artemis pedigree. The heterozygosity in C4 is roughly double that of C1, reflecting differences in the history of these genotypes. A number of markers are monomorphic in both parents for opposite alleles. These fixed differences between parents will segregate in generations beyond the F_1, thereby offering additional segregation of alleles not revealed in Artemis.

Phenotypic variation can be seen in the Artemis pedigree, consistent with the high level of genetic variation. This is shown in Fig. 2 for our mapping population of the Artemis F_1, grown in UK field trials during 2007 (UK07). Metabolite profiling revealed concentrations of artemisinin that ranged from 0.93 to 20.65 μg/mg dry weight, with associated metabolites also showing variation (Fig. 2A). Leaf area ranged between 508.76 and 4696.08 mm^2 (Fig. 2B), glandular trichome density between 4.89 and 19.11 mm^{-2} (Fig. 2C), and plant fresh weight between 160 and 4440 g (Fig. 2D). These traits are targets for increasing artemisinin yield, which is a product of both artemisinin concentration and plant fresh weight.

The fact that the Artemis parents are heterozygous enabled us to produce genetic linkage maps for each parent based on data derived from an F_1 mapping population of 242 individuals (fig. S2) (*15*). Using a minimum LOD (logarithm of the odds ratio for linkage) score of 4.0, we defined nine linkage groups for the C4 parent and seven linkage groups for C1 (fig. S2). We hypothesized that the C1 map is missing two linkage groups, designated LG8 and LG9, because two chromosomes in the C1 parent are either homozygous or have a very low level of heterozygosity and therefore do not segregate for markers from this parent in the F_1, so cannot be mapped in this generation. To test this hypothesis, an individual F_1 plant showing high heterozygosity in molecular analysis was self-pollinated to produce an F_2 generation. Markers seen to be homozygous for opposite alleles in the parents, and therefore heterozygous in all F_1 progeny, were genotyped in the F_2 generation

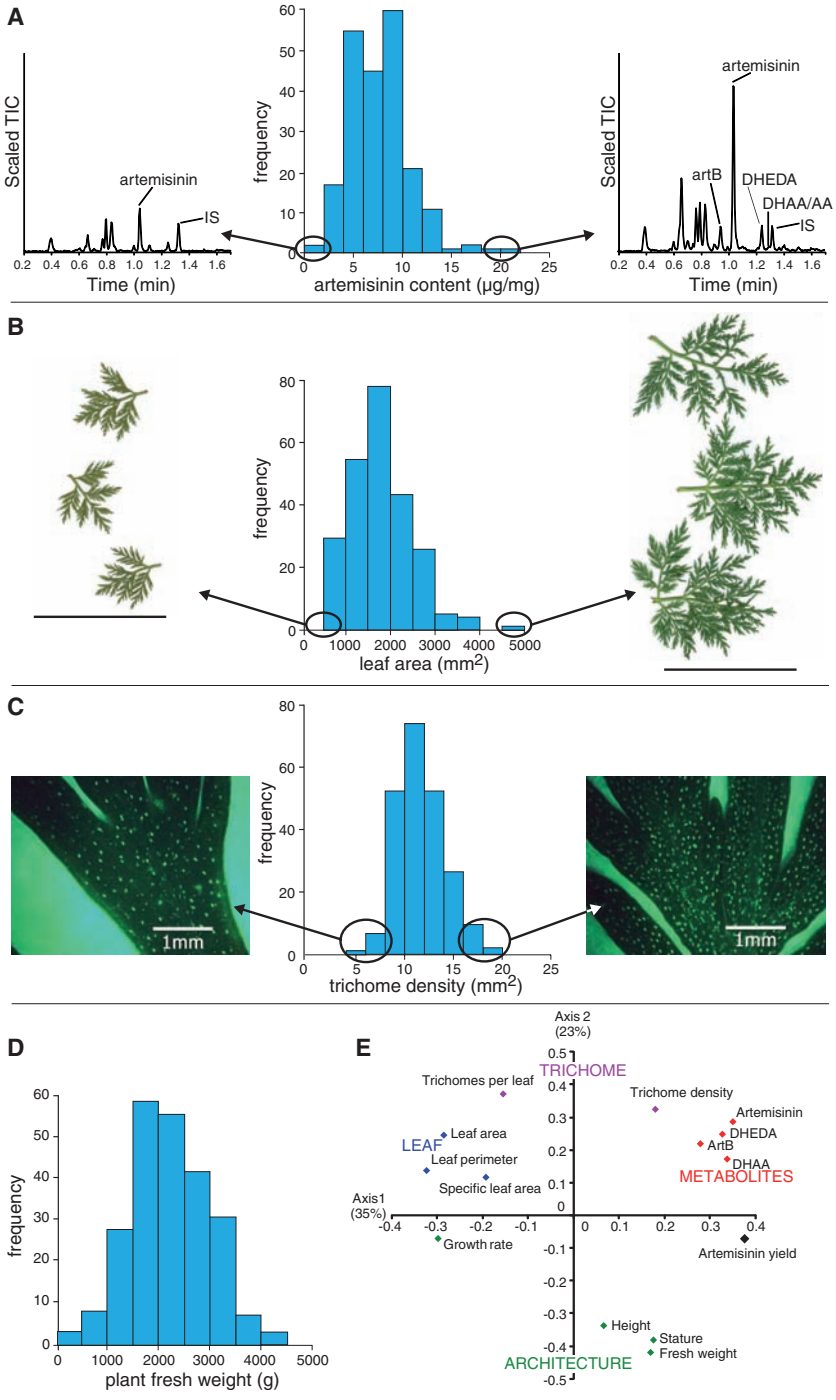

Fig. 2. Phenotypic variation in the Artemis F_1 grown in UK field trials in 2007. The distribution of four traits related to artemisinin yield is illustrated. (**A**) Artemisinin concentration at harvest (7 months after sowing). Metabolite profiles showing artemisinin and related metabolites from the lowest- and highest-yielding plants relative to an internal standard (IS) are presented. artB, arteannuin B; DHEDA, dihydro-epi-deoxyarteannuin B; DHAA, dihydroartemisinic acid; AA, artemisinic acid. (**B**) Leaf area 5 months after sowing. Images show leaves from positions 20, 21, and 22 from the apical meristem. (**C**) Trichome density 5 months after sowing. The abaxial surface of leaves 15, 16, and 17 from the apical meristem was visualized by fluorescent microscopy. Glandular trichomes appear as bright green spots. (**D**) Fresh weight of aboveground plant material at harvest (7 months after sowing). (**E**) Principal component analysis of traits related to artemisinin yield. Architecture, leaf, and metabolite traits additional to those detailed in (A) to (D) are also included in the analysis as shown.

together with markers known to map to LG8 and LG9 in the C4 parent. In support of our hypothesis, a number of these markers were found to segregate in the F_2 generation. These data allowed F_2 linkage groups for LG8 and LG9 to be defined and anchored to the corresponding C4 linkage groups (fig. S2). The identification of nine LGs is consistent with cytological studies reporting the diploid number of chromosomes to be 18 in *A. annua* (*16*). The marker positions shown on the map were validated by three independent approaches: coalignment on the C4 and C1 maps, common location of multiple markers from single candidate genes, and robustness of marker order after reconstruction of the map with a subset of markers (SOM Text).

We used vegetative propagation to replicate individuals from the mapping population, which enabled us to perform three independent field trials using the same genotypes. A single replicate of each genotype was tested in 2007 in the UK (UK07) and three replicates of each genotype were tested both in the UK (UK08) and Switzerland (SW08) in 2008. Fourteen traits were scored that could affect artemisinin yield (Fig. 2E). All these traits exhibited a moderate to high heritability ranging from 0.41 to 0.62, resulting in the discovery of multiple QTLs. Stable QTLs for artemisinin concentration were identified on C4 LG1, LG4, and LG9 (Fig. 3 and table S4), which describe 20% of the variation in UK07 and between 30 and 38% in

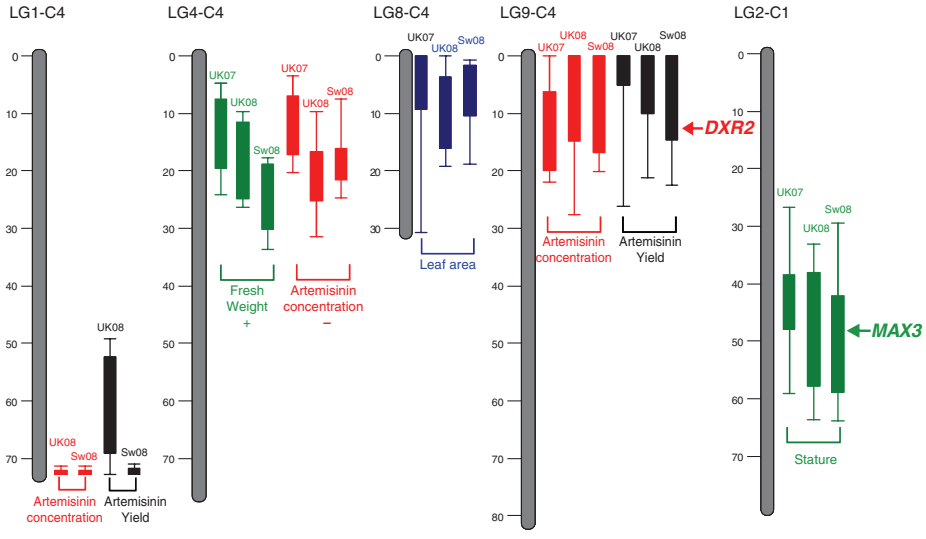

Fig. 3. A selection of QTLs for key traits identified across three field trials. QTLs are shown to the right and distances in centimorgans to the left of each linkage group. Thick and thin lines indicate the confidence intervals of the QTLs corresponding to 1 and 2 LOD units below the maximum LOD score, respectively. QTLs are shown for artemisinin concentration (in red), artemisinin yield (artemisinin concentration × fresh weight) (in black), architecture (fresh weight and stature) (in green), and leaf area (in blue). Trials in which QTLs were detected are denoted as UK07, UK08 and SW08. Candidate genes associated with QTL are *DXR2* (*1-Deoxy-D-Xylulose 5-phosphate Reductoisomerase 2*) and *MAX3* (*More Axillary Branching 3*).

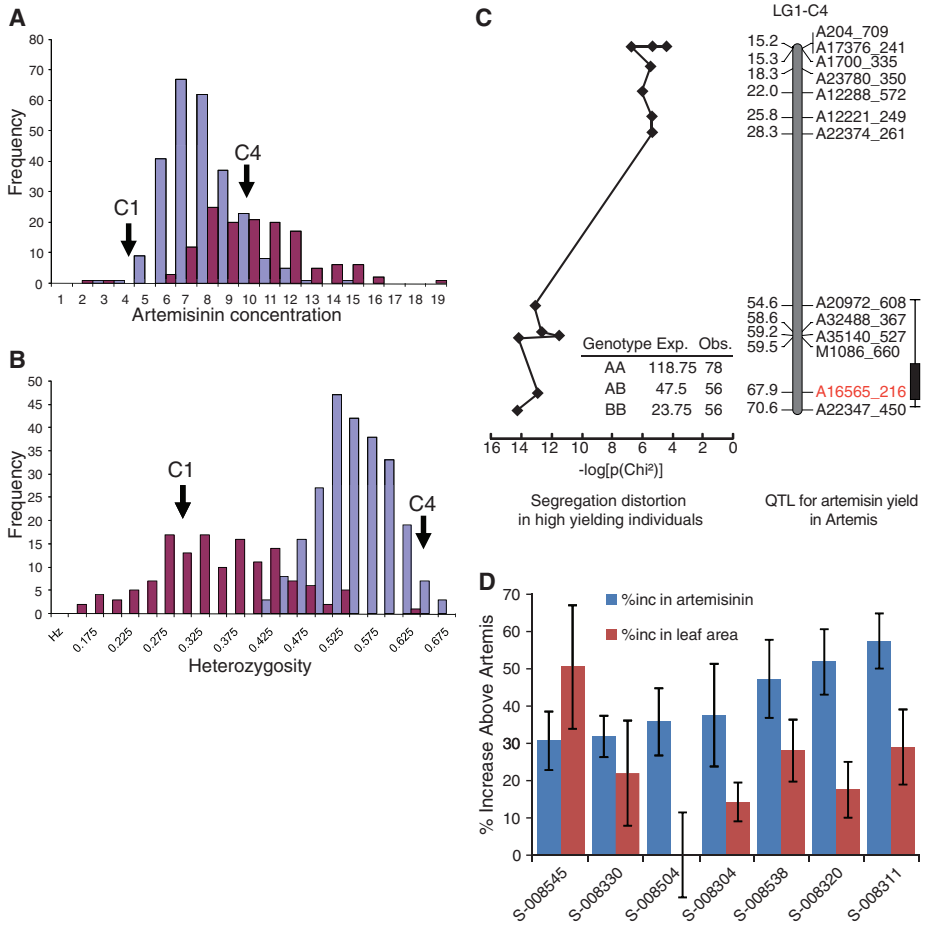

Fig. 4. Genetic analysis of high-yielding plants. (**A**) Distribution of artemisinin concentration (μg/mg dry weight) in the F_1 mapping population of 242 individuals is shown in blue and that in 130 selected high-yielding F_2 individuals grown in the same trial (UK08), is shown in red. The artemisinin concentrations of C4 and C1 grown in the same trial are indicated. (**B**) Distribution of heterozygosity scores for the same individuals as in (A). (**C**) The position of a major QTL for artemisinin yield on LG1 and markers in this region show high segregation distortion in favor of the increasing alleles in 190 F_2 high-yielding individuals. For the marker highlighted in red, the B allele has a positive effect on yield ($P = 4.4 \times 10^{-7}$) and is overrepresented in the high-yielding individuals summarized in the table inset. The plotted values for segregation distortion represent the –log [Chi-squared] based on the observed and expected values for genotype classes at a number of markers on linkage group 1. (**D**) The percentage increase in artemisinin concentration (in red) and leaf area (in blue), over Artemis F_1 for seven hybrids produced from crosses of selected high-yielding individuals. Values are the mean ± SE for a minimum of five individual replicates.

UK08 and SW08. Artemisinin yield is a product of both artemisinin concentration and fresh weight. QTLs for yield colocate to those for artemisinin concentration on LG1 and LG9, thus representing targets for a breeding program. The artemisinin concentration QTL on LG4 colocates with a QTL for fresh weight but with opposing effects on artemisinin yield (Fig. 3). Markers in candidate genes colocate with a number of the QTLs. For example, the precursor supply gene candidate *DXR2* colocates with the QTL for artemisinin yield on C4 LG9 and the architectural trait candidate gene *MAX3* colocates with the QTL for stature on C1 LG2.

In parallel with the development of the marker-assisted breeding program, we performed a high-throughput screen for artemisinin content in 23,000 12-week-old glasshouse-grown F_2 and F_3 plants derived from F_1 Artemis seed that had been mutagenized with ethylmethane sulfonate (*15*). The mutation frequency in this material was determined with the TILLING method and found to be approximately one EMS-induced mutation per 5.4 Mb (*15*). This is less than the SNP frequency determined for Artemis at one polymorphism per 104 base pairs. This screen should therefore identify individuals carrying beneficial mutations derived from the EMS treatment and also individuals carrying improved genetic backgrounds as a result of segregation of favorable alleles derived from natural variation. We found that the distribution of artemisinin content among selected high-yielding F_2 individuals is higher than in the UK08 Artemis F_1 mapping population (Fig. 4A) even though overall heterozygosity is lower (Fig. 4B). Next, we determined whether any of the QTLs we had identified for artemisinin yield on the basis of field trials are overrepresented in the high-yielding individuals that had been selected under glasshouse conditions. We found strong segregation distortion in favor of the advantageous alleles for an artemisinin yield QTL on C4 LG1 (Fig. 4C). These data validate this QTL and confirm that for artemisinin yield, the genotype has a strong influence on both glasshouse-grown and field-grown material.

An ongoing empirical hybridization program of high-yielding plants identified in the high-throughput phenotypic screen produced hybrid progeny that outperformed Artemis for artemisinin concentration and leaf area after 12 weeks' growth under glass (Fig. 4D). The choice of parents for this program preceded the availability of QTL data and was based on phenotypic characteristics (*15*). In terms of utility in a molecular breeding program, we found a significant association of positive artemisinin yield QTL in those parents that produced hybrids with increased artemisinin yield ($P < 0.001$).

Our study has established the molecular basis for marker-assisted breeding of this medicinal plant species and highlights the reduced timelines that are now feasible for developing this platform of knowledge and tools. The artemisinin from *A. annua* is the key component in the ACT treatment of malaria, and demand for ACTs is expected to increase in the immediate future. Development of new high-yielding varieties optimized for production in different geographic regions is now a realistic target.

References and Notes

1. World Malaria Report 2008, World Health Organisation; http://apps.who.int/malaria/wmr2008/malaria2008.pdf.
2. A. M. Dondorp *et al.*, *N. Engl. J. Med.* **361**, 455 (2009).
3. "*Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance*," National Academy of Sciences 2004, www.nap.edu/catalog/11017.html. Global Malaria Action Plan, Report of the 2008 Artemisinin Conference, 8 to 10 October, York, UK (www.york.ac.uk/org/cnap/artemisiaproject/pdfs/AEconference-report-web.pdf).
4. M. V. Duke, R. N. Paul, H. N. Elsohly, G. Sturtz, S. O. Duke, *Int. J. Plant Sci.* **155**, 365 (1994).
5. C. M. Bertea *et al.*, *Planta Med.* **71**, 40 (2005).
6. C. M. Bertea *et al.*, *Arch. Biochem. Biophys.* **448**, 3 (2006).
7. K. H. Teoh, D. R. Polichuk, D. W. Reed, G. Nowak, P. S. Covello, *FEBS Lett.* **580**, 1411 (2006).
8. Y. Zhang *et al.*, *J. Biol. Chem.* **283**, 21501 (2008).
9. K. H. Teoh, D. R. Polichuk, D. W. Reed, P. S. Corvello, *Can. J. Bot.* **87**, 635 (2009).
10. P. S. Covello, *Phytochemistry* **69**, 2881 (2008).
11. M. C. Chang, R. A. Eachus, W. Trieu, D. K. Ro, J. D. Keasling, *Nat. Chem. Biol.* **3**, 274 (2007).
12. D. K. Ro *et al.*, *Nature* **440**, 940 (2006).
13. J. F. S. Ferreira, J. Janick, *Int. J. Plant Sci.* **156**, 807 (1995).
14. N. Delabays, X. Simonnet, M. Gaudin, *Curr. Med. Chem.* **8**, 1795 (2001).
15. Information on materials and methods is available on *Science* Online.
16. M. Torrell, J. Vallés, *Genome* **44**, 231 (2001).

17. We thank L. Doucet, H. Martin, N. Nattriss, M. Segura, and A. Czechowska for horticulture assistance; G. Chigeza for horticulture management; S. Graham, S. Heywood, B. Kowalik, S. Pandey, R. Simister, and C. Whitehead for laboratory assistance; C. Calvert, P. Dickes, W. Lawley, and D. Rotherham for project management; E. Bartlet for communications advice; and P. Roberts for graphic design. We thank T. Brewer, H. Klee, and K. Stuart for insightful advice on this project. We thank X. Simonnet and Médiplant for access to the Artemis pedigree. We acknowledge financial support for this project from The Bill and Melinda Gates Foundation and Medicines for Malaria Venture, as well as from The Garfield Weston Foundation for the Centre for Novel Agricultural Products.

Supporting Online Material

www.sciencemag.org/cgi/content/full/327/5963/328/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
Tables S1 to S4
References

29 September 2009; accepted 20 November 2009
10.1126/science.1182612

Tetrathiomolybdate Inhibits Copper Trafficking Proteins Through Metal Cluster Formation

Hamsell M. Alvarez,[1]* Yi Xue,[1]* Chandler D. Robinson,[1] Mónica A. Canalizo-Hernández,[1] Rebecca G. Marvin,[1] Rebekah A. Kelly,[3] Alfonso Mondragón,[2] James E. Penner-Hahn,[3] Thomas V. O'Halloran[1,2]†

Tetrathiomolybdate (TM) is an orally active agent for treatment of disorders of copper metabolism. Here we describe how TM inhibits proteins that regulate copper physiology. Crystallographic results reveal that the surprising stability of the drug complex with the metallochaperone Atx1 arises from formation of a sulfur-bridged copper-molybdenum cluster reminiscent of those found in molybdenum and iron sulfur proteins. Spectroscopic studies indicate that this cluster is stable in solution and corresponds to physiological clusters isolated from TM-treated Wilson's disease animal models. Finally, mechanistic studies show that the drug-metallochaperone inhibits metal transfer functions between copper-trafficking proteins. The results are consistent with a model wherein TM can directly and reversibly down-regulate copper delivery to secreted metalloenzymes and suggest that proteins involved in metal regulation might be fruitful drug targets.

Excess dietary molybdate (MoO_4^{2-}) uptake was first linked to a fatal disorder in cattle known as "teart" pastures syndrome (*1*) and later to a neurological disorder in sheep known as "swayback" (*2*). Both disorders arise from Mo-induced copper deficiency, and the symptoms are readily reversed with copper supplementation. Although molybdate itself has little or no affinity for copper ions, the active copper-depleting agent, TM (MoS_4^{2-}), is formed in the ruminants' digestive track and readily reacts with Cu^{I} or Cu^{II} to form insoluble compounds. These zoogenic studies inspired the development of molybdenum compounds to treat copper-dependent diseases in humans (*3*). The potent chelating and antiangiogenic activities of orally active formulations of TM, such as the ammonium salt [$(NH_4)_2(MoS_4)$] (*4*–*6*) and the choline salt (ATN-224) (*7*, *8*), have been used in treatment of Wilson's disease, where copper accumulation leads to hepatic and neurological disorders, as well as in the inhibition of metastatic cancer progression in a number of clinical trials (*9*–*11*). TM inhibits several copper enzymes, including ceruloplasmin (Cp), ascorbate oxidase, cytochrome oxidase, superoxide dismutase (SOD1), tyrosinase, and the *Enterococcus*

[1]The Chemistry of Life Processes Institute, Northwestern University, Evanston, IL 60208, USA. [2]Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, IL 60208, USA. [3]Department of Chemistry, The University of Michigan, Ann Arbor, MI 48109, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: t-ohalloran@northwestern.edu