Catalysis of 5-methyltetrahydrofolate to MeFox facilitates folate biofortification in crops

Chunyi Zhang (zhangchunyi@caas.cn)
Biotechnology Research Institute, Chinese Academy of Agricultural Sciences https://orcid.org/0000-0001-5685-6412

Ling Jiang
Biotechnology Research Institute, Chinese Academy of Agricultural Sciences https://orcid.org/0000-0002-4658-6989

Wenzhu Guo
Chinese Academy of Sciences

Yanjing Wang
Huazhong Agricultural University

Weiwei Wen
Huazhong Agricultural University

Weixuan Wang
Biotechnology Research Institute, Chinese Academy of Agricultural Sciences

Tong Lian
Biotechnology Research Institute, Chinese Academy of Agricultural Sciences; University of Liège

Qiuju Liang
Chinese Academy of Agricultural Sciences

Ji'an Liu
Biotechnology Research Institute, Chinese Academy of Agricultural Sciences

Haijun Liu
Huazhong Agricultural University https://orcid.org/0000-0001-7717-893X

Yuan Xue
Huazhong Agricultural University

Lixu Pan
Yangzhou University

Qiaoquan Liu
Yangzhou University https://orcid.org/0000-0001-5543-5798

Ping Yin
National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China https://orcid.org/0000-0001-8001-221X

Delin Zhang
Huazhong Agricultural University

Jianbing Yan
Huazhong Agricultural University https://orcid.org/0000-0001-8650-7811

Article

Keywords: CTM, 5-methyl-tetrahydrofolate, MeFox, natural variation, biofortification

Posted Date: March 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-308995/v1

Catalysis of 5-methyltetrahydrofolate to MeFox facilitates folate biofortification in crops

Ling Jiang[1#], Wenzhu Guo[2,3#], Yanjing Wang[2#], Weiwei Wen[2#], Weixuan Wang[1], Tong Lian[1,4], Qiuju Liang[1], Ji'an Liu[1], Haijun Liu[2], Yuan Xue[2], Lixu Pan[5], Qiaoquan Liu[5], Ping Yin[2], Delin Zhang[2]*, Jianbing Yan[2]*, Chunyi Zhang[1]*

*Correspondence should be addressed to Delin Zhang (zdl@mail.hzau.edu.cn), Jianbing Yan (yjianbing@mail.hzau.edu.cn), and Chunyi Zhang (zhangchunyi@caas.cn)

[#]These authors contributed equally to this work

Author Information

[1]Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Ling Jiang, jiangling@caas.cn

Weixuan Wang, wangweixuan@caas.cn

Tong Lian, liantong9111@163.com

Qiuju Liang, liangqiuju@caas.cn

Ji'an Liu, liujian1232021@163.com

Chunyi Zhang, zhangchunyi@caas.cn

[2]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

Wenzhu Guo, guowenzhu0534@163.com

Yanjing Wang, yjwang@webmail.hzau.edu.cn

Weiwei Wen, wwwen@mail.hzau.edu.cn

Haijun Liu, heroalone@qq.com

Yuan Xue, xueyuan@webmail.hzau.edu.cn

Ping Yin, yinping@mail.hzau.edu.cn

Delin Zhang, zdl@mail.hzau.edu.cn

Jianbing Yan, yjianbing@mail.hzau.edu.cn

[3]Tianjin Institute of Industrial Biotechnology, Chinese Academy of Science, Tianjin, 300308, China

Wenzhu Guo, guowenzhu0534@163.com

[4]Plant Genetics, Gembloux Agro-Bio Tech, University of Liège, Gembloux 5030, Belgium.

Tong Lian, liantong9111@163.com

[5]Key Laboratory of Crop Genomics and Molecular Breeding of Jiangsu Province, College of Agriculture, Yangzhou University, 225009 Yangzhou, China

Lixu Pan, panlixu288@163.com

Qiaoquan Liu, qqliu@yzu.edu.cn

Contributions:

C.Z., J.Y., D.Z., and P.Y. designed and supervised this study. L.J., W.G., T.L., Q.L., and J.L. performed the folate measurement. W.G., W.W., and H.L. performed GWAS data analysis. L.J., T.L., Q.L., J.L., Q.Q.L., and L.P. performed the analysis on transgenic plants. D.Z., Y.W., P.Y. and Y.X. performed the protein crystallization and enzymatic analysis. W.X.W. performed molecular simulation analysis. L.J., G.W., D.Z., W.W., J.Y., and C.Z. prepared the manuscript with inputs from other authors. L.J, W.G, Y.W., W.W. contributed equally to this work

Corresponding authors

Corresponding to Delin Zhang (0000000252990718), Jianbing Yan (0000000186507811), and Chunyi Zhang (0000000156856412)

Running title: CTM for folate biofortification

Abstract

Folate deficiency is a global health problem. Biofortification has been considered a cost-effective means to tackle this problem. Here, we describe the genetic cloning and functional identification of a previously uncharacterised plant protein, designated as CTM, which functions as an enzyme in folate metabolism. Plant CTMs are capable of catalysing 5-methyl-tetrahydrofolate to MeFox, a pyrazino-s-triazine derivative of 4α-hydroxy-5-methyl-tetrahydrofolate. The natural asparagine-to-glycine substitution caused by an A-to-G single nucleotide variation in maize CTM enhances its enzymatic activity, as demonstrated by *in vitro* enzymatic assays and *in silico* analyses using a maize CTM structure model based on a monomeric sorghum CTM crystal. Loss of the *CTM* function led to accumulation of 5-methyl-tetrahydrofolate, and overexpression of the maize CTM carrying the G-allele boosted the metabolic flow towards MeFox, and showed no negative impacts on plant growth. Our results suggest that CTM, which has evolved 5-methyl-tetrahydrofolate-to-MeFox converting activity in plants, could be valuable for developing folate-biofortified crops to provide an alternative to the challenge presented by the global folate deficiency.

Key words: CTM, 5-methyl-tetrahydrofolate, MeFox, natural variation, biofortification

Word count: 2780

Introduction

Folates, including tetrahydrofolate (THF) and its derivatives, are essential water-soluble B-vitamins for all living organisms. The human body cannot synthesize folates *de novo* and must capture folates from dietary foods[1]. Unfortunately, the dietary intake of folates is inadequate in both developing and developed countries[2-8]. The worldwide prevalence of neural tube defects, a severe disorder caused by folate deficiency, remains between 0.3 and 124.1 per 10,000 births[9, 10]. Accordingly, folate deficiency is considered a global public health problem and the prevention of folate deficiency remains a global priority.

To alleviate folate deficiency, food fortification and medical supplementation have been implemented in many countries. For example, folic acid supplementation proved effective for reducing the prevalence of neural tube and foetal abdominal wall defects[11, 12], but led to concerns regarding potential adverse effects of elevated folate status[13]. In contrast, biofortification is cost-effective, sustainable, and easily accessible (reviewed in ref.[14-17]), making it more acceptable for humans, and especially for poorer populations. Efforts thus far have focused on maize (*Zea mays*), potato (*Solanum tuberosum*), rice (*Oraza sativa*), tomato (*Solanum lycopersicum*), and wheat (*Triticum aestivum*) through the overexpression of the folate biosynthetic genes in combination with enhanced vitamin stability, leading to various degrees of folate accumulation[18-22]. Furthermore, a marker-assisted breeding approach has been attempted, but only with some genetically mapped quantitative trait loci/markers. Thus, there is an urgent need to identify new target genes and elite allelic variations for folate biofortification in crops. In this investigation, a genome-wide association study (GWAS) was conducted using a maize association panel containing hundreds of maize inbred lines genotyped by more than 1 million genome-wide single nucleotide polymorphisms (SNPs) [26,27]. We showed that *ZmCTM* (catalysis of 5-M-THF to MeFox) maps to a previously described quantitative trait locus[28] and encodes a protein that catalyses 5-M-THF to MeFox. The natural A-to-G variation explained 27.6% of the phenotypic variation, and led to a natural asparagine (Asn, N)-to-glycine (Gly, G) substitution in ZmCTM, which enhances its enzymatic activity. We also demonstrated that this natural variation is valuable for folate biofortification by the conserved function of CTMs in other plants and the verified functional natural variation in sweetcorn.

Results

***ZmCTM* is a major gene locus associated with folate derivative accumulation in corn seeds.** In a maize association panel[26], 5-F-THF-M, representing a mix of 5-F-THF and MeFox, comprised more than 50% of the total folates in most mature seeds (Supplementary Data 1). 5-F-THF-M levels ranged from 0.58 to 8.18 nmol g^{-1} and demonstrated high repeatability of 0.76 (Extended Data Table 1). We identified 36 genome-wide significant SNPs associated with the content of 5-F-THF-M at the level of $P < 1.81 \times 10^{-6}$, corresponding to nine loci from five chromosomes (Fig. 1b; Extended Data Table 2). The strongest association signal ($P = 1.46 \times 10^{-21}$) was mapped to a locus on chromosome 5 (GRMZM2G124863), encoding a protein that catalyses 5-M-THF to MeFox (described in detail below) and was thus designated *ZmCTM*. Two continuous SNPs, Chr5.S_19676906 and Chr5.S_19676907, which were in complete linkage disequilibrium with the highest –log*P* value, led to an Asn-to-Gly substitution and explained 27.6% of the phenotypic variation (Fig. 1c; Extended Data Table 2). Resequencing of *ZmCTM* in 134 inbred lines and subsequent candidate gene association analysis revealed 56 polymorphic sites that were significantly associated with levels of 5-F-THF-M ($P \leqslant 0.05$, MLM; Fig. 1c; Supplementary Data 2). Notably, the two SNPs at S2069 (Chr5.S_19676906) and S2070 (Chr5.S_19676907) exhibited a persistently strong association ($P = 1.51 \times 10^{-6}$, MLM; Fig. 1c; Supplementary Data 2). Allelic comparison based on S2069 indicated higher 5-F-THF-M levels and a greater MeFox/5-M-THF ratio among maize inbred lines carrying the G-allele than among maize inbred lines carrying the A-allele (Supplementary Results 1). To investigate whether *ZmCTM* is responsible for folate accumulation in corn seeds, *ZmCTM*-edited maize mutant plants carrying stop-gain mutation were generated using the CRISPR-Cas9 system (Supplementary Results 2). Loss of *ZmCTM* function led to a 3.3-fold increase of 5-M-THF, the most strongly affected folate derivative, and a decrease of MeFox (83% of the level in the wild-type) was observed in mature seeds (Fig. 1d). To investigate the potential effect of the A/G allele on folate accumulation, we introduced two genes into maize: A-allele carrying *ZmCTM-B73* and G-allele carrying *ZmCTM-Qi319*, under the control of constitutive promoters (Supplementary Results 2). MeFox was accumulated in mature seeds of both transgenic plants, compared with the wild-type, such that the G-allele demonstrated a significant

increase (1.6-fold; Fig. 1e). Thus, *ZmCTM* was considered as a major gene locus responsible for folate derivative accumulation in corn seeds, and the natural A-to-G variation may favour this accumulation.

ZmCTM catalyses 5-M-THF to MeFox *in vitro*. Initially, *ZmCTM* was predicted to encode a glutamate formiminotransferase (GFT) because the presumed protein showed 20.7% amino acid sequence identity to the formiminotransferase (FT) domain of the porcine (*Sus scrofa*) formiminotransferase cyclodeaminase (SsFTCD; Supplementary Results 3). Algal and land plant CTMs showed a well-conserved motif pattern[29]. Therefore, we investigated whether ZmCTM had GFT activity. As expected, both recombinant SsFT and SsFTCD converted THF to 5,10-CH=THF in the presence of *N*-formimino-L-glutamate[30]. However, ZmCTMs from the inbred lines B73 (ZmCTM-B73) and Qi319 (ZmCTM-Qi319) did not display GFT activity, nor did orthologues from sorghum (*Sorghum bicolor*), wheat, sweet cherry (*Prunus avium*), and potato (Fig. 2a). These observations indicated that plant CTMs might participate in folate metabolism in a manner different from that of mammals.

In seeds of *ZmCTM*-edited maize plants, 5-M-THF and MeFox were inversely associated (Fig. 1d; Supplementary Results 2). Because MeFox was previously identified as a pyrazino-s-triazine derivative of 4α-hydroxy-5-M-THF[31], we speculated that ZmCTM might participate in the 5-M-THF-to-MeFox conversion. To investigate this possibility, ZmCTM-B73 and a mutant ZmCTM-B73 bearing an alanine substitution of the conserved histidine at the position 117 (i.e., ZmCTM^{H117A}) were expressed in insect cells and the purified proteins were subjected to an *in vitro* enzymatic assay. Incubation of 5-M-THF with the recombinant ZmCTM-B73 protein resulted in robust production of MeFox, such that approximately 40% of the 5-M-THF was converted (Fig. 2b, c). In contrast, incubation of 5-M-THF with the mutant protein ZmCTM^{H117A} produced only trace amounts of MeFox (Fig. 2c). Other plant CTMs (e.g., SbCTM, TaCTM, and PaCTM) were also capable of catalysing 5-M-THF to MeFox *in vitro* (Fig. 2d). Therefore, we concluded that ZmCTM, a previously uncharacterised plant protein, mainly contributed to the production of MeFox and used 5-M-THF as its substrate.

To unravel the structural characteristics of the ZmCTM protein at the atomic level, we constructed a three-dimensional (3D) model using the SbCTM crystal (Protein Data Bank [PDB]: 7DYH; 1.75 Å resolution) generated in this study as the main frame, and using the previously published SsFT (PDB: 1QD1) as a reference for the binding pocket[30] (Fig. 2e–h; Extended Data Table 3; Extended Data Fig. 1; Supplementary Results 3). Homology modelling revealed that ZmCTM and SbCTM shared an almost identical conformation, with a root-mean-square deviation between the two structures over 1665 Cα atoms of 0.162 Å. SsFT, SbCTM, and ZmCTM shared a similar global structure, with the N- and C-terminal subdomains comprising two α/β units (Fig. 2g). The binding pocket in SsFT formed a channel for transportation of the intermediate from the FT domain to the CD domain[30]. Compared with SsFT, both SbCTM and ZmCTM-B73 had a shorter pocket (Extended Data Fig. 1b, d, f). We investigated the optimal substrate of ZmCTM-B73 using molecular virtual screen. Notably, neither THF (the substrate for the mammalian FT domain) nor 5-F-THF (the ligand in the SsFT crystal complex) had the highest score; 5-M-THF displayed the highest binding affinity (Extended Data Table 4). Although most of the derivatives assessed were situated similarly in the binding pocket of ZmCTM-B73, 5-M-THF was in the most favorable position, such that N228 (corresponding to the A-allele at S2069) was located near the benzene ring of the aminobenzoate residue in the binding pocket (Fig. 2h). G57, R65, and T307 formed six hydrogen-bonds with the aminobenzoate head of 5-M-THF, while K25, D125, S127, and Y161 formed another four hydrogen-bonds with the glutamate tail. F56, G57, Y27, and L275 had pi-pi, amide-pi or alkyl-pi interactions with the aromatic ring of 5-M-THF. Other amino acids (e.g., D58, R63, H117, R175, and N228) determined the main shape of the binding pocket to adjust the ligand position. Hydrophobic bonds could be formed between the methyl group of 5-M-THF and the adjacent F56, G57, R63, and R65 residues, leading to a higher binding affinity to 5-M-THF than to THF or 5-F-THF (Extended Data Fig. 1g, h). These observations demonstrated that 5-M-THF is the most favorable substrate for ZmCTM.

The Asn-to-Gly substitution in ZmCTM causes changes in enzymatic activity. Because overexpression of the G-allele carrying *ZmCTM* resulted in a folate-accumulating pattern that differed from the pattern of the A-allele carrying *ZmCTM* (Fig. 1e), we investigated whether this variation affected ZmCTM activity *in vitro*. We introduced point mutations at S2069 into *ZmCTM-B73* and *ZmCTM-Qi319*, respectively. Specifically, we modified the favorable G-allele in ZmCTM-Qi319 to

the unfavorable A-allele, resulting in the ZmCTM-Qi319^{G232N} mutant protein. We also modified ZmCTM-B73 with the unfavorable A-allele to the favorable G-allele, resulting in the ZmCTM-B73^{N228G} mutant protein. As expected, ZmCTM-B73^{N228G} showed enhanced activity for the 5-M-THF-to-MeFox conversion (Fig. 3a), whereas ZmCTM-Qi319^{G232N} displayed reduced activity (Fig. 3b). An enzymatic kinetic assay showed that the k_{cat}/K_m value for ZmCTM-Qi319 was 1.24×10^4 $mol^{-1} \cdot min^{-1} \cdot L$. In contrast, the activity of the mutant ZmCTM-Qi319^{G232N} was reduced approximately six-fold, such that its k_{cat}/K_m value was 2.04×10^3 $mol^{-1} \cdot min^{-1} \cdot L$. Furthermore, the k_{cat}/K_m value for ZmCTM-B73 was 1.51×10^3 $mol^{-1} \cdot min^{-1} \cdot L$ and the Asn-to-Gly mutation enhanced its activity, leading to a four-fold enhancement of k_{cat}/K_m value (6.05×10^3 $mol^{-1} \cdot min^{-1} \cdot L$; Fig. 3b). Similar to ZmCTM-Qi319, SbCTM activity also dramatically decreased following the introduction of the G219N mutation (Fig. 3c). These observations demonstrated that the natural variation at S2069 strongly affected ZmCTM activity, such that greater activity was favoured by the G-allele than by the A-allele.

Consistent with the enzymatic activity findings, *in silico* molecular docking suggested that ZmCTM-Qi319 had a greater binding affinity to 5-M-THF than ZmCTM-B73 (Extended Data Table 5). The 3D structural models of ZmCTM-Qi319 and ZmCTM-B73 were almost identical, such that they differed by one amino acid (Gly versus Asn) caused by one SNP (S2069) near the binding pocket. G232 in ZmCTM-Qi319 occupied less space than did N228 in ZmCTM-B73, resulting that 5-M-THF positioned in a different subsite which was more suitable for the ligand binding. (Fig. 3e, f). These observations were presumed to explain how the natural variation at S2069 affects the enzymatic activity of the ZmCTM protein.

CTMs may be useful for folate biofortification in crops. *In vitro* enzymatic assays demonstrated that the CTMs from plants catalysed 5-M-THF to MeFox (Fig. 2b–d), indicating that CTM functions are conserved in plants. To further confirm this conservation in cereal, the expression of rice *CTM* (Os03g38540) was knocked down by RNA interference (RNAi). The results showed a 3.1-fold enhancement of 5-M-THF and a significant reduction of MeFox (81%, compared with wild-type) in mature transgenic seeds (Extended Data Fig. 2a; Supplementary Results 4). Notably, the RNAi rice plants had normal growth similar to the growth of *CTM*-edited maize plants (Supplementary Results 2 and 4). To investigate whether CTM could also catalyse 5-M-THF to MeFox in a heterologous

manner in other plants (e.g., *Arabidopsis* [*Arabidopsis thaliana*], where MeFox was almost undetectable in rosette leaves [Extended Data Fig. 2b]), we introduced ZmCTM-B73 into *Arabidopsis* under the control of the CaMV 35S promoter. This overexpression of ZmCTM-B73 resulted in an accumulation of MeFox in the rosette leaves of the transgenic plants (0.08 ± 0.03 nmol g^{-1} FW, P = 0.074; Extended Data Fig. 2b).

Importantly, CTM functioned more effectively in young seeds than in mature seeds, as demonstrated by the significant MeFox reduction in *ZmCTM*-edited plants (8% for young seeds versus 83% for mature seeds, compared with the wild-type; Fig. 1d and Fig. 4a) and significant MeFox elevation in *ZmCTM-Qi319*-overexpressing plants (2.6-fold for young seeds versus 1.6-fold for mature seeds, compared with the wild-type; Fig. 1e and Fig. 4b). Regarding 5-M-THF, we observed a similar variation for the increase (3.3-fold in mature seeds versus 3.0-fold in young seeds, compared with the wild-type; Fig. 1d and Fig. 4a), which resulted in significant accumulation of 5-M-THF in young mature seeds of *ZmCTM*-edited plants (1.95 ± 0.18 nmol g^{-1} FW versus 0.23± 0.01 nmol g^{-1} FW; Fig. 4a). Furthermore, the G-allele was preferable to the A-allele for the accumulation of folate derivatives in young seeds. Allele G-carrying *ZmCTM-Qi319* overexpression in maize led to significantly higher levels of both MeFox and 5-M-THF than those in allele A-carrying *ZmCTM-B73* (2.50 ± 0.08 nmol g^{-1} FW versus 1.88 ± 0.02 nmol g^{-1} FW for MeFox, and 1.85 ± 0.14 nmol g^{-1} FW versus 1.30 ± 0.20 nmol g^{-1} FW for 5-M-THF; Fig. 4b), and showed no negative impacts on plant growth. Additionally, only marginal increases in plant height and grain protein content were observed in *ZmCTM-Qi319* overexpressors, compared with the wild-type (Supplementary Results 2).

Next, we assessed the potential for the natural variation to facilitate folate biofortification in sweetcorn due to the human preference for sweetcorn at a young stage (days after pollination [DAP] 20 to 24) and the close kinship of sweetcorn with field corn[32] (Supplementary Data 3). Among all 47 commercial sweetcorn hybrids analysed, the MeFox/5-M-THF ratio was significantly greater in homozygous G-carrying corn than in either homozygous A-carrying or heterozygous G/A corn (Fig. 4c), consistent with the observations made using inbred lines (Supplementary Results 1). Notably, higher levels of folates were present in the G-allele carrying sweetcorn than in the A-allele carrying corn (Fig. 4c, d).

In total, 53.2% (25 of 47) of the hybrids were homozygous for the A-allele, 25.5% (12 of 47) were homozygous for the G-allele, and 21.3% (10 of 47) were heterozygous. Among the ten heterozygous corn hybrids, four contained total folate levels > 400 μg 100 g^{-1} FW folates (Supplementary Data 3) and readily satisfied the dietary allowance recommended by the World Health Organization[15] (400 μg per day) if an adult consumes 100 g corn kernels a day.

Discussion

Thus far, metabolic engineering of para-aminobenzoic acid and the pterin branch with the trunk of folate biosynthesis (but not the genes involved in one-carbon metabolism) has been the main strategy to enhance levels of plant folates (reviewed in references[16-17]). This study revealed a 5-M-THF-to-MeFox conversion catalysed by a previously uncharacterised plant protein in folate metabolism, and subsequently developed a new strategy to use a natural variation for folate biofortification in crops.

To our knowledge, this is the first report that CTM functions as an enzyme to catalyse 5-M-THF to MeFox. It was previously reported that two enzymes directly participate in 5-M-THF metabolism. 5,10-Methylene-THF is reduced by methylenetetrahydrofolate reductase (MTHFR) to form 5-M-THF, which subsequently enters the methionine cycle by methionine synthase as a one-carbon unit donor (reviewed in reference[33]; Fig. 1a). However, no other enzymes have been reported since MTHFR and methionine synthase were first reported in *Escherichia coli*, 60 years ago[34, 35]. Therefore, this study sheds a new light on folate metabolism and fills the knowledge gap regarding production of MeFox. Notably, plant CTMs have evolved a new enzymatic activity that differs from FTCD activity in mammals[30]. However, the molecular basis underlying the evolution of this new enzymatic activity remains unknown and further studies are needed to explore the catalytic mechanism by which CTMs convert 5-M-THF to MeFox. Nevertheless, irrespective of knowledge gaps regarding the protein structure and biochemical evidence, we speculate that the CTM-mediated 5-M-THF-to-MeFox conversion might involve two steps. First, a hydroxyl group is added to the 18th carbon of 5-M-THF to transform 5-M-THF into hmTHF (4α-hydroxy-5-methyltetrahydrofolate)[36]. Second, hmTHF undergoes a spontaneous intramolecular rearrangement and is subsequently transformed into MeFox.

There was a dramatic increase in 5-M-THF in mature seeds of *ZmCTM*-edited maize plants without disruption of growth performance. This indicates that the obstruction of metabolic flow towards MeFox by introducing a loss-of-function mutation into *CTM* can substantially improve 5-M-THF accumulation. This represents a new strategy for folate biofortification to benefit human health, especially for people with only partial MTHFR activity[37]. Furthermore, boosting the metabolic flow towards MeFox through expression of the allele G-carrying *CTM* gene increased the accumulation of both 5-M-THF and MeFox, especially in young seeds. This demonstrates that the G-allele is favorable for folate accumulation in sweetcorn. In many populations, dietary folate intake is considerably lower than recommended level[2-8]. Because high levels of folates were detected in the heterozygous G/A corn, breeders may consider using the G-allele to breed high-folate corn varieties by means of the marker-assisted approach. We anticipate that the *CTM* gene and its natural variation identified in this study will be valuable for developing folate-fortified crops to benefit human health.

Main references

1. Hanson, A.D. and Gregory, J.F. Folate biosynthesis, turnover, and transport in plants. *Annu. Rev. Plant Biol.* **62**, 105–125 (2011).
2. Akter, R., et al. Micronutrient adequacy in the diet of reproductive-aged adolescent girls and adult women in rural Bangladesh. *Nutrients* **13**, 337 (2021).
3. Ayensu, J., et al. Prevalence of anaemia and low intake of dietary nutrients in pregnant women living in rural and urban areas in the Ashanti region of Ghana. *PLoS ONE* **15**, e0226026(2020).
4. Barth-Jaeggi, T., et al. Nutritional status of Tajik children and women: Transition towards a double burden of malnutrition. *Matern. Child. Nutr.* **16**, e12886 (2020).
5. Chong, M.F., et al., A landscape of micronutrient status in women through the reproductive years: Insights from seven regions in Asia. *Women's Health* **16**, 1 –11(2020).
6. Herter-Aeberli, I., et al. Inadequate status and low awareness of folate in Switzerland— A call to strengthen public health measures to ensure sufficient intakes. *Nutrients* **12**, 3729 (2020).

7. Janowska-Miasik, E., et al. Diet quality in the population of Norway and Poland: Differences in the availability and consumption of food considering national nutrition guide lines and food market. *BMC Public Health* **21**, 319 (2021).
8. Madhari, R., et al., High dietary micronutrient inadequacy in peri-urban school children from a district in South India: Potential for staple food fortification and nutrient supplementation. *Matern. Child. Nutr.***16**, e13065 (2020).
9. Zaganjor, I., et al. Describing the prevalence of neural tube defects worldwide: A systematic literature review. *PloS One.* **11**, e0151586 (2016).
10. Van Der Straeten, D. et al. Multiplying the efficiency and impact of biofortification through metabolic engineering. *Nat. Commun.* **11**, 5203 (2020).
11. Ren, A. Prevention of neural tube defects with folic acid: The Chinese experience. *World J. Clin. Pediatr.* **4**, 41–44(2015).
12. Liu, J., et al. Folic acid supplementation and risk for foetal abdominal wall defects in China: Results from a large population-based intervention cohort study. *Br. J. Nutr.* **26**, 1–20 (2021).
13. Maruvada, P., et al. Knowledge gaps in understanding the metabolic and clinical effects of excess folates/folic acid: A summary, and perspectives, from an NIH workshop. *Am. J. Clin. Nutr.* **112**, 1390–1403 (2020).
14. Bouis, H.E., et al. Biofortification: A new tool to reduce micronutrient malnutrition. *Food Nutr. Bull.* **32**, S31–40 (2011).
15. Blancquaert, D., et al. Present and future of folate biofortification of crop plants. *J. Exp. Bot.* **65**, 895–906 (2014).
16. Strobbe, S., and Van Der Straeten, D. Folate biofortification in food crops. *Curr. Opin. Biotech*. **44**, 202–211(2017).
17. Jiang, L., et al. Regulation of plant vitamin metabolism: Backbone of biofortification for the alleviation of hidden hunger. *Mol. Plant* **14**, 40–60 (2021).
18. Liang, Q., et al. Improved folate accumulation in genetically modified maize and wheat. *J. Exp. Bot.* **70**, 1539–1551 (2019).
19. de Lepeleire, J., et al. Folate biofortification of potato by tuber-specific expression of four folate biosynthesis genes. *Mol. Plant* **11**, 175–188 (2018).
20. Storozhenko, S., et al. Folate fortification of rice by metabolic engineering. *Nat. Biotechnol.* **25**,

337 1277–1279 (2007).
338 21. Blancquaert, D., et al. Improving folate (vitamin B9) stability in biofortified rice through meta-
339 bolic engineering. *Nat. Biotechnol.* **33**, 1076–1078 (2015).
340 22. de la Garza, R.I.D., Gregory, J.F., and Hanson, A.D. Folate biofortification of tomato fruit. *Proc.*
341 *Nat. Acad. Sci. U. S. A*. **104**, 4218–4222 (2007).
342 23. Bali, S., et al. Single nucleotide polymorphism (SNP) markers associated with high folate content
343 in wild potato species. *PLoS ONE* **13**, e0193415 (2018).
344 24. Guo, W., et al. Genetic mapping of folate QTLs using a segregated population in maize. *J. Integr.*
345 *Plant Biol.* **61**, 675–690 (2019).
346 25. Jha, A.B., et al. Folate profile diversity and associated SNPs using genome wide association study
347 in pea. *Euphytica*, **216**, 18 (2020).
348 26. Yang, X.H., et al. Characterization of a global germplasm collection and its potential utilization for
349 analysis of complex quantitative traits in maize. *Mol Breeding*, **28**, 511–526 (2011).
350 27. Liu, H., et al. Distant eQTLs and non-coding sequences play critical roles in regulating gene
351 expression and quantitative trait variation in maize. *Mol. Plant* **10**, 414–426 (2017).
352 28. Guo, W., et al. Genetic mapping of folate QTLs using a segregated population in maize. *J. Integr.*
353 *Plant Biol.* **61**, 675–690 (2019).
354 29. Gorelova,V., et al. Evolution of folate biosynthesis and metabolism across algae and land plant
355 lineages. *Sci. Report* **9**, 5731 (2019).
356 30. Kohls, D., Sulea, T., Purisima, E.O., MacKenzie, R.E. & Vrielink, A. The crystal structure of the
357 formiminotransferase domain of formiminotransferase-cyclodeaminase: Implications for substrate
358 channeling in a bifunctional enzyme. *Structure* **8**, 35–46 (2000).
359 31. Fazili, Z., et al. Demographic, physiologic, and lifestyle characteristics observed with serum total
360 folate differ among folate forms: Cross-sectional data from fasting samples in the National Health
361 and Nutrition Examination Survey 2011–2016. *J. Nutr.* **150**, 851–860 (2020).
362 32. Olmstead, D.L., Nault, B.A. and Shelton, A.M. Biology, ecology, and evolving management of
363 *Helicoverpa zea* (Lepidoptera: Noctuidae) in sweet corn in the United States. *J. Econ. Entomol.*
364 **109**, 1667–1676 (2016).
365 33. Cossins, E. & Chen, L. Folates and one-carbon metabolism in plants and fungi. *Phytochemistry* 45,
366 437–452 (1997).

34. Cathou, R.E. & Buchanan, J.M. Enzymatic synthesis of the methyl group of methionine. V. Studies with 5, 10-methylenetetrahydrofolate reductase from *Escherichia coli*. *J. Biol. Chem.* **238**, 1746–1751 (1963).
35. Cohn, M., Cohen, G.N. & Monod, J. The specific inhibiting effect of methionine in the formation of methionine-synthase by *Escherichia coli*. *C. R. Hebd. Seances. Acad. Sci.* **236**, 746-748 (1953).
36. Blair, J.A., Pearson, A.J. and Robb, A.J. Autoxidation of 5-Methyl-5,6,7,8-tetrahydrofolic Acid. *J. Chem. Soc. Perkin Transactions 2,* **1**, 18–21(1975).
37. Rommer, P.S., et al. Manifestations of neurological symptoms and thromboembolism in adults with MTHFR-deficiency. *J. Neurol. Sci.* **383**, 123–127 (2017).

Methods

Plant materials and field trials

A diverse maize association panel consisting of 531 maize inbred lines was used for folate profiling[37]. The panel was planted in three field trials including Hainan in 2009 (09ZHN), Yunnan in 2010 (10WY), and Hubei (10AMH) in 2010. Samples were harvested from 501, 406 and 464 lines in these environments, respectively (Supplementary Data 1). Field experiments followed the procedures described previously, where one row was planted for each inbred line and at least six ears in each row were self-pollinated for all 531 lines[39].

A collection of 143 lines from the panel containing 368 maize inbred lines, which were part of the 531-association panel were grown in Hebei (China) in 2019. A commercial collection of 47 fresh corn hybrids were collected from Shanghai, Guangdong, Beijing, and Sichuan (all in China) and grown in in Hebei (China) in 2020. Field experiments followed the procedures described previously[40]. For the 143 inbred lines, one row was planted for each inbred line and at least three ears in each row were self-pollinated and harvested for folate profiling of mature seeds (Supplementary Result 1). For the 47 commercial hybrid lines, one row was planted for each inbred line and at least three ears in each row were self-pollinated and harvested for folate profiling of the young seeds (DAP 20 to 24; Supplementary Data 3).

Extraction and profiling of folates

Preparation of mature seeds from the inbred line samples (09ZHN, 10WY, and 10AMH) for testing of folate composition, including 5-F-THF-M, was performed as described[41]. Detection of folate derivatives, including MeFox in fresh leaves, young seeds, and mature seeds from maize, rice, and *Arabidopsis* was carried out in accordance with the method described by Shahid et al[42].

Genome-wide association study

In total, 368 inbred lines were genotyped from the above the 531-association panel using RNA-Seq data[26, 27]. From among 1.03 million high-quality SNPs identified by RNA sequencing, approximately 560,000 polymorphisms with minor allele frequency > 0.05 were selected for the GWAS. Furthermore, 5-F-THF-M was measured in mature seeds of 513 maize inbred lines grown in the three environments mentioned above. In total 558,529 SNPs with a minor allele frequency > 0.0536 were used to investigate the association with 5-F-THF-M levels on the base of linear unbiased prediction data across these three environments by means of a mixed linear model that considered population structure and individual relatedness[43]. A strict threshold ($P = 1/n = 1.8 \times 10^{-6}$, where n = total markers used) was set to identify considerable genome-wide SNPs. To identify more SNPs or genes related to 5-F-THF-M, a conditional GWAS was conducted using the four most significant SNPs identified as covariates.

Re-sequencing of candidate gene *ZmCTM*

The sequence of the candidate gene *ZmCTM* was obtained from the B73 reference at the MaizeSequence database (http://ensembl.gramene.org/Zea_mays/Info/Index?db=core). Primers were designed using Primer Premier 5 software to cover the entire gene, including regions of the promoter, exons, introns, and 5' and 3' UTRs. Polymerase chain reaction products of 134 inbred lines from 155 temperate Chinese inbred lines, among which 89 lines were from the 531-association panel, were subsequently sequenced. The sequences were assembled using ContigExpress[44] and aligned using MUSCLE40[45], then refined manually in BioEdit[46]. Nucleotide polymorphisms including SNPs and InDels with a frequency of ≥ 0.05 were extracted.

Transgenic analysis

Zm*GFT1*-edited maize plants carrying stop-gain mutation were generated using the CRISPR-Cas9 system[47], with an SgRNA target sequence of CCAAGTTCATCTCCTGCAAG. The *ZmCTM* open reading frames of the B73 (S2069 = A) and Qi319 (S2069 = G) inbred lines driven by the *ubiquitin* promoter were introduced into the maize hybrid line C01 *via* Agrobacterium-mediated transformation to obtain transgenic lines[48]. The endogenous *CTM* gene carries G at S2069. After T1 seeds had been obtained, all transgenic plants were self-pollinated for two generations. Three individual lines of homozygous Zm*GFT1*-edited, *ZmCTM-B73* overexpressing and *ZmCTM-Qi319* overexpressing maize were characterised for 2 years, and representative data from 1 year were shown in this study because the lines exhibited similar patterns. Young seeds (DAP25) and mature seeds of third-generation-transgenic plants were collected for folate profiling. The ZmCTM open reading frames of the B73 (S2069 = A) and Qi319 (S2069 = G) inbred lines driven by the *CaMV 35S* promoter were introduced into Arabidopsis c*olumbia via* Agrobacterium-mediated transformation to obtain transgenic lines[49]. Rosette leaves of the transgenic plants were grown for 30 days and collected for folate profiling. Two individual lines of homozygous *ZmCTM-B73* overexpressors and *ZmCTM-Qi319* overexpressors were characterised twice, and represented data from one assay were shown in this study because they exhibited similar patterns. To knock down *OsCTM* (Os03g38540), an orthologue of ZmCTM in rice genome, an RNAi construct containing 360-bp *OsCTM* -specific fragment under the control of the ubiquitin promoter was also transformed into the *Oryza sativa* subsp. *japonica* rice cultivar Yandao 8 (Y8) via *Agrobacterium*[50]. Mature seeds of the transgenic plants were collected for folate profiling. Five individual lines of homozygous *OsCTM-RNAi* were characterised twice, and represented data from one assay were shown in this study because they exhibited similar patterns. Plant heights of the mature wild-type and transgenic maize and rice were measured by a ruler. Hundred-grain weight was measured using a weighing scale, and protein content and starch content were measured by near-infrared-transmission spectroscopy (Infratec^{TM}1241 Grain Analyzer; Foss, Denmark) when the mature seeds was harvested.

Real time-PCR

The B73 and C01 inbred lines were used to investigate potential patterns. Roots, stems, young leaves, tassels, and embryos/endosperm at 10, 15, 20, or 25 DAP were collected separately. Transcript levels

of *ZmGFT1* were analysed in the transgenic plants. Total RNA was extracted with TRIzol reagent[51]. cDNAs were constructed using the first-strand cDNA synthesis kit (Thermo). Quantitative real-time polymerase chain reactions were performed with an ABI7500. The primers used in this experiment are listed in Extended Data Table 6.

Sequence alignment

Phylogenetic evolutionary analyses were conducted using MEGA version 5[52]. The amino acid sequences of CTM in maize and other species were aligned using ClustalW (https://www.genome.jp/tools-bin/clustalw), and output by Espript3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi).

Purification of ZmCTMs, SbCTMs, PaCTM, TaCTM, StCTM, SsFTCD and SsFT proteins

ZmCTM genes were amplified from the cDNAs of Qi319 and B73 and subcloned into a modified pFastBac1 vector with a 10 × His affinity tag fused to the C-terminus, respectively. Bacmids were generated in DH10Bac cells in accordance with the instructions for the BAC-to-BAC baculovirus expression system (Invitrogen). Baculovirus was generated and amplified in Sf-9 insect cells. Variant ZmCTM proteins were expressed in Sf-9 insect cells at 27 °C for 60 h using individual viruses. Cells were harvested by centrifugation at 2,000 × g for 15 min and homogenised in ice-cold lysis buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 0.5 mM phenylmethanesulfonyl-fluoride. Cells were disrupted using a cell homogeniser (JuNeng Co., Ltd). The insoluble fraction was precipitated by ultracentrifugation (20,000 × g) for 1 h at 4°C. The supernatant was loaded onto a Ni-NTA superflow affinity column (Qiagen) and washed three times with lysis buffer plus 10 mM imidazole. Elution was performed in buffer containing 25 mM Tris-HCl (pH 8.0), and 250 mM imidazole. The protein was further purified using Source15Q (GE Healthcare) and concentrated to approximately 1 mg mL^{-1} (Amicon 30 kDa cutoff, Millipore), followed by size-exclusion chromatography (Superdex-200 Increase 10/300, GE Healthcare) and equilibrated with 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 5 mM dithiothreitol. Peak fractions were pooled for the enzymatic assay. Mutant *ZmCTM* genes were constructed by overlap polymerase chain reaction and the proteins were expressed and purified as description above.

The genes of *SbCTM*, *PaCTM*, *TaCTM*, *StCTM* and *SsFTCD* were synthesised by Genewiz (GENEWIZ, Inc., China) and optimised for expression in *E.coli*. The sequence of SsFT was amplified by polymerase chain reaction using the *SsFTCD* gene as the template. All genes were individually subcloned into a modified pET21b vector with a 6 × His tag fused at the C terminus (Novagen), and the plasmid was transformed into BL21 (DE3) cells. One litre of lysogeny broth medium supplemented with 100 μg mL^{-1} ampicillin was inoculated with a transformed bacterial pre-culture and shaken at 37 °C until the optical density reached 1.0 at 600 nm. After induction with 0.2-mM isopropyl-β-D-thiogalactoside and 16-h of growth at 16°C, the bacterial cells were collected and homogenised (JuNeng Co., Ltd) in a buffer containing 25 mM Tris-HCl (pH 8.0), and 150 mM NaCl, then centrifugated at 23,000 × g at 4°C. The supernatant was loaded onto a column equipped with Ni^{2+} affinity resin (Ni-NTA, Qiagen), washed with a buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 15 mM imidazole, and eluted with a buffer containing 25 mM Tris-HCl (pH 8.0) and 250 mM imidazole. The eluted protein was applied to Source15Q (GE Healthcare), then subjected to gradient NaCl elution (up to 1 M) in 25 mM Tris-HCl (pH 8.0). The elution peak was concentrated to 1 mL (approximately 10 mg mL^{-1}) and subjected to gel filtration chromatography (Superdex200 Increase 10/300, GE Healthcare) equilibrated with a buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM NaCl and 5 mM 1,4-dithiothreitol. The peak fractions were collected for enzymatic activity determination or crystallisation trials. Mutant *SbCTM* genes were constructed by overlap polymerase chain reaction and the proteins were expressed and purified as described above. The primers used in this experiment are listed in Extended Data Table 6.

Determination of ZmCTMs and SbCTMs activity *in vitro*

MeFox, 5-M-THF, and THF powders were purchased from Toronto Research Chemicals (Canada, http://www.trc-canada.com) and Schircks Laboratories (Switzerland, http://www.schircks.ch), and resolved in a buffer containing 50 mM sodium phosphate (pH 7.0) and 10 mM β-mercaptoethanol as the standard sample for mass spectrometry determination or substrate for enzymatic activity measurement. *N*-formimino-L-glutamate powder was purchased from Sigma-Aldrich (China, https://www.sigmaaldrich.com/china-mainland.html).

For the formiminotransferase activity assay, a 0.1 mL mixture containing 0.1 M phosphate buffer (pH

7.4), 10 mM β-mercaptoethanol, 0.5 mM THF, 5 mM *N*-formimino-L-glutamate, and an appropriate amount of enzyme was incubated for 2 h at 30°C. The reaction was stopped by the addition of 0.1 mL of 0.36 M HCl. The tubes were then heated at 100°C for 55 s, cooled on ice, and centrifuged to remove the precipitant. The absorbance at 350 nm was determined by comparison with a blank which did not contain formiminoglutamate.

For mass spectrometry assays, Mefox and 5-M-THF were detected using multiple reaction monitoring in the ESI positive mode on a mass spectrometer. Multiple reaction monitoring (477.4/327.2 Da) was used to identify MeFox, and multiple reaction monitoring (460.2/313.1Da) was used to identify 5-M-THF[53]. Standard curves of MeFox or 5-M-THF were drawn by measuring the gradient concentration of the sample in 10 parts per billion (ppb: μg L^{-1}), 20 ppb, 40 ppb, and 80 ppb, respectively. One millimolar MeFox or 5-M-THF substrate was incubated with a gradient ZmCTM protein in 100 μL at 30°C for 12 h, and the reaction was then diluted by 10-fold and quenched with a buffer containing 50% methanol (v/v), 0.1% sodium ascorbic acid (w/v), 0.5% β-mercaptoethanol (v/v), and 20 mM ammonium acetate. A 2 μL reaction mixture was used for liquid chromatography (UFLC SHIMADZU CBM20A system) with a C18 column (VP-ODS, 150 L × 20) and mass spectrometry (Applied Biosystems 4000 Q TRAP) to detect MeFox production or 5-M-THF consumption. The amounts of the product generated (or substrate remaining) during the reaction were calculated based on the peak areas.

For ZmCTM or SbCTM kinetic assay, 0.5 μM protein was incubated at 30°C with 50–900 μM of 5-M-THF substrate in a 100 μL reaction buffer containing 50 mM sodium phosphate (pH 7.0) and 10 mM β-mercaptoethanol at 30°C. The reaction was quenched and subjected to liquid chromatography-mass spectrometry as described above. The initial velocity of the reaction was calculated by measuring MeFox generation at 3, 6, 9, or 12 min, respectively. The initial velocity in a gradient concentration of 5-M-THF was fitted by the Michaelis-Menten equation to obtain the k_{cat} and K_m values. k_{cat} and K_m values, as well as errors resulting from fitting, were calculated using GraphPad Prism software (GraphPad Software, Inc, USA, https://www.graphpad.com/)

Crystallisation of SbCTM

SbCTM (A6-D317, C260S/C314S, C-His) was crystallised using the hanging-drop vapour-diffusion method at 18°C by mixing 1 μL of the sample with an equal volume of reservoir solution. Crystal optimisation was carried out in the drop plate. After the pH buffer, salt concentration, and additive, high-resolution crystals were obtained in in the conditions containing 15% w/v polyethylene glycol 20,000, 0.1 M HEPES (pH7.0) and 3% w/v 1.6-hexanediol. The crystals were flash-frozen in liquid nitrogen using 20% (v/v) ethylene glycol as the cryoprotective buffer and diffracted to 1.75 Å at the Shanghai Synchrotron Radiation Facility beamline BL17U1[54]. The structure of SbCTM was resolved by the isomorphous replacement method using the selenomethionine method.

Data Collection and Structural Determination

All data sets were collected at the Shanghai Synchrotron Radiation Facility beamline BL17U1 or BL19U and processed with the HKL3000 or HKL2000 packages[55]. Further processing was performed with programs from the CCP4 suite[56]). Data collection and structure refinement statistics are summarised in Extended Data Table 3. The structure was manually iteratively refined with the PHENIX[57] and COOT[58] tools. All figures were generated using the PyMOL software (http://www.pymol.org).

Structure modeling

The X-ray crystal structure of SbCTM was used as the template to construct the 3D model of ZmCTM because they share the high amino acid sequence identity. In addition, a blast search[59] of the amino acid sequence of ZmCTM was conducted against the current PDB (http://www.rcsb.org) to obtain further information. The FT domain of SsFTCD (PDB: 1QD1) was chosen as the reference because the crystal structure of the SsFT complex contained an analogue molecule, 6R-5-F-THF. Several initial models were constructed using the Modeler module[60] in Discovery Studio 2.0 (Accelrys Software Inc.), and the model with the highest Profiles-3D[61] score was retained. Energy minimisation procedures were processed under the CHARMm[62] (Brooks et al., 1983) force field. The SHAKE algorithm[63] was applied to constrain covalent bonds to hydrogen atoms during the minimisation. Finally, the Profiles-3D method was used to evaluate the fitness between the sequence and the current 3D model. ZmCTM-B73 and ZmCTM-Q319 were modelled separately.

Molecular docking

Using the established homology model, AutoDock Vina[64] was employed to find the potential ligand of ZmCTM as well as the binding mode between ligand and protein. The 3D structures of folate derivatives (Extended Data Tables 4 and 5) were sketched and further refined with the steepest descent minimisation for 2000 steps, followed by gradient minimisation for another 2000 steps, using the CHARMm force field. The active site pocket of the receptor was identified by Discovery Studio 2.0, and the location of the ligand 6R-5-F-THF in 1DQ1 was referenced for additional information. A box size of 20 × 18 × 18 was set as the entire binding pocket using ADT software[65]. Other parameters were set as default. The top nine docking positions ranked by the binding affinity were preserved to find the most probable binding mode. To investigate the binding abilities of 5M-THF to ZmCTM-B73 and ZmCTM-Qi319, four docking programs (i.e., AutoDock Vina, AutoDock 4[65], CDOCKER[66], and LigandFit[67]) were employed to find the best binding mode. The binding pockets were defined as above. The top docking scores of each program were collected to generate a consensus score.

Methods references

38. Yang, X. *et al*. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breeding* **28**,511–526 (2011).
39. Li, Q., et al. Genome-wide association studies identified three independent polymorphisms associated with α-tocopherol content in maize kernels. *PLoS One* **7**, e36807 (2012).
40. Guo, W. *et al.* Genetic mapping of folate QTLs using a segregated population in maize. *J. Integr. Plant Biol.* **61**, 675–690 (2019).
41. Jiang, L., et al. The mitochondrial folylpolyglutamate synthetase gene is required for nitrogen utilization during early seedling development in *Arabidopsis*. *Plant Physiol.* **161**, 971–989 (2013).
42. Shahid, M., et al. Folate monoglumate in cereal grains: Evaluation of extraction techniques and determination by LC-MS/MS. *J. Food Comp. Analyt.* **91**, 103510 (2020).
43. Yu, J., et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208 (2006).
44. Lu, G. et al. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief Bioinform.* **5**, 378–88 (2004).

603 45. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl.*
604 *Acids Res.* **32**, 1792–1797 (2004).
605 46. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for
606 Windows 95/98/NT. *Nucl. Acid.s Symp. Ser.* **41**, 95–98 (1999).
607 47. Li, C., et al. RNA-guided Cas9 as an in vivo desired-target mutator in maize. *Plant Biotechnol. J.*
608 **15**, 1566–1576 (2017).
609 48. Ishida,Y., et al. High efficiency transformation of maize (*Zea mays* L.) mediated by Agrobacterium
610 tumefaciens. *Nat. Biotechnol.* **14**, 745–750 (1996).
611 49. Clough, S.J. and Bent, A.F. Floral dip: A simplified method for Agrobacterium-mediated
612 transformation of Arabidopsis thaliana. *Plant J.* **16**, 735–743(1998).
613 50. Zhao, D., et al. GS9 acts as a transcriptional activator to regulate rice grain shape and appearance
614 quality. *Nat. Communi.* **9**, 1240 (2018).
615 51. Chomczynski, P. A. reagent for the single-step simultaneous isolation of RNA, DNA and proteins
616 from cell and tissue samples. *Bio. Techniques.* **15**, 532–537 (1993).
617 52. Tamura, K., et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood,
618 evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731-2739 (2011).
619 53. Fazili, Z.& Pfeiffer, C.M. Accounting for an isobaric interference allows correct determination of
620 folate vitamers in serum by isotope dilution-liquid chromatography-tandem mass spectrometry. *J.*
621 *Nutr.* **143**, 108–113 (2013).
622 54. Wang,Q.S., et al, The macromolecular crystallography beamline of SSRF. *Nucl. Sci. Tech.* **26**, 12–
623 17 (2015).
624 55. Otwinowski, Z. & Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode.
625 *Methods Enzymology* **276**, 307–326 (1997).
626 56. Winn, M.D., et al. Overview of the CCP4 suite and current developments. *Acta Cryst.* **D67**, 235–
627 242 (2011).
628 57. Adams, P.D., et al. PHENIX: Building new software for automated crystallographic structure
629 determination. *Acta Cryst.* **D58**, 948–1954(2002).
630 58. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Cryst.* **D60**,
631 2126–12132 (2004).
632 59. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment search tool.

633 *J. Mol. Biol.* **215**, 403–410 (1990).
634 60. Sali, A. & Blundell, T. L. Comparative protein modeling by satisfaction of spatial restraints. *Mol.*
635 *Biol.* **234**, 779–815 (1993).
636 61. Luthy, R., Bowie, J. U., and Eisenberg, D. Assessment of protein models with three-dimensional
637 profiles. *Nature* **356**, 83–85 (1992).
638 62. Brooks, B. R., et al. CHARMM: A program for macromolecular energy minimization and
639 dynamics calculations. *J. Comp. Chem.* **4**, 187–217 (1983).
640 63. Rychaert, J. P., Ciccotti, G., Berendsen, H. J. C. Numerical integration of the Cartesian equations
641 of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**,
642 327–341 (1977).
643 64. Trott, O., Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new
644 scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461(2010).
645 65. Morris, G. M., et al. Autodock4 and AutoDockTools4: Automated docking with selective receptor
646 flexibility. *J. Comput. Chem.* **16**, 2785–2791 (2009).
647 66. Wu, G., Robertson, D. H., Brooks, C. L. III., Vieth, M. Detailed analysis of grid-based molecular
648 docking: A case study of CDOCKER - A CHARMm-based md docking algorithm. *J. Comp. Chem.*
649 **24**, 1549–1562 (2003).
650 67. Venkatachalam, C. M., Jiang, X., Oldfield, T., Waldman, M. LigandFit: A novel method for the
651 shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* **21**, 289–
652 307(2003).
653

654 # Figure legends

655 **Fig. 1. GWAS and identification of the major gene locus that underlies folate variation.** (**a**)
656 Scheme of folate metabolism. Chemicals: 5-F-THF, 5-formyl-tetrahydrofolate; 10-F-THF, 10-formyl-
657 tetrahydrofolate; GTP, guanosine triphosphate; Hcy, homocysteine; MeFox, a pyrazino-s-triazine
658 derivative of 4α-hydroxy-5-methyl-tetrahydrofolate; 5-M-THF, 5-methyl-tetrahydrofolate; 5,10-
659 CH=THF, 5,10-methenyl-tetrahydrofolate; 5,10-CH_2-THF, 5,10-methylene-tetrahydrofolate; THF,
660 tetrahydrofolate. The underlined chemicals were described in this report. Enzymes: CTM, catalysis

from 5-M-THF to MeFox; DHFR, dihydrofolate reductase; DHFS, dihydrofolate synthase; DHC, 5,10-methylene-THF dehydrogenase/5,10-methenyl-THF cyclohydrolase; 5-FCL, 5-formyl THF cycloligase; GDC, glycine decarboxylase complex; FTCD, glutamate formiminotransferase-cyclodeaminase; MS, methionine synthase; MTHFR, methylenetetrahydrofolate reductase; SHMT, serine hydroxymethyltransferase. (**b**) Manhattan plot of GWAS for the content of 5-F-THF-M in mature corn seeds, with the position of CTM is indicated. (**c**) Structure of *ZmCTM* and association mapping with more genetic variants. Dots represent 56 variants identified by resequencing that were significantly associated with the content of 5-F-THF-M ($P < 0.05$). Different dot colours represent linkage disequilibrium with the peak SNP (S2069). X-axis, position relative to the starting site of resequencing (0 bp). (**d**) Folate profiles (mean ± standard deviation [SD] of three biological replates with n = 3 each) in the wild-type (WT) and *ZmCTM*-edited (CRISPR) maize mature seeds. *P*-value (Student's *t*-test, $P < 0.05$) represents *ZmCTM*-edited compared with the wild-type maize. (**e**) Folate profiles (mean ± SD of three biological replicates with n = 3 each) in mature seeds of the wild-type (WT), *ZmCTM-B73* (OE-B73) and *ZmCFTM-Qi319* (OE-Qi319) overexpressing maize lines. A in the brackets after OE-B73 represents A-allele at S2069, and G in the brackets after OE-Qi319 represents G-allele at S2069. The C01 inbred line (genetic transformation recipient) was used as the wild-type (WT), and the endogenous *CTM* gene carries G at S2069. *P*-values (Student's *t*-test, $P < 0.05$) represent comparisons among *ZmCTM-B73* and *ZmCFTM-Qi319* overexpressing transgenic, and wild-type maize. n.s., not significant. See Extended Data Tables 1 and 2, Supplementary Data 1 and 2, and Supplementary Results 1 and 2.

Fig. 2. ZmCTM directly converts 5-M-THF to MeFox. (**a**) The formiminotransferase (FT) activity of wild-type ZmCTMs (ZmCTM-B73 and ZmCTM-Qi319) and their homologues. SsFTCD and SsFT have three concentration gradients, while the other samples had fixed concentration. Enzymes: CTM, catalysis from 5-M-THF to MeFox; FT, formiminotransferase; FTCD, formiminotransferase-cyclodeaminase. Species: Pa, *Prunus avium*; Sb, *Sorghum bicolor*; Ss, *Sus scrofa*; Ta, *Triticum aestivum*; St, S*olanum tuberosum*; Zm, *Zea mays*. (**b**) Activity of ZmCTM during catalysis of 5-M-THF to MeFox. The standard 5-M-THF and MeFox, as well as the actual catalytic products of ZmCTM and ZmCTM^{H117A}, were detected by routine liquid chromatography-mass spectrometry as described in the Methods. 5-M-THF, 5-methyltetrahydrofolate; MeFox, pyrazino-s-

triazine derivative of 4a-hydroxy-5-M-THF. (**c**) Relative contents of 5-M-THF and MeFox in the enzymatic assay. Quantification of MeFox production or 5-M-THF consumption by the addition of ZmCTM or ZmCTM^{H117A}. The amounts of MeFox or 5-M-THF were normalized to the blank. Each column represents a mean value of three independent measurements and the error bar represents standard deviation. The *P*-value is based on a two-sided Student's *t*-test. n.s., not significant. (**d**) CTM catalytic activities of plant CTMs. The concentrations of SsFTCD and SsFT for CTM assay were fixed at 0.8 μM, while concentrations of other samples for assays were 0.8 μM, 0.4 μM, or 0.2 μM. Enzymes: CTM, catalysis from 5-M-THF to MeFox; FT, formiminotransferase; FTCD, formiminotransferase-cyclodeaminase. Species: Pa, *Prunus avium*; Sb, *Sorghum bicolor*; Ss, *Sus scrofa*; Ta, *Triticum aestivum*; St, S*olanum tuberosum*. (**e**) The overall structure of the apo state SbCTM. The N-terminal domain (NTD) and C-terminal domain (CTD) are shown in lime green and aquamarine. (**f**) The view of SbCTM is rotated 180 degrees; the colour representation is identical to that shown in (**e**). (**g**) Superimposed structure of SbCTM (PDB code 7DYH; lime green for the NTD and aquamarine for the CTD), SsFT (PDB code 1QD1, grey), and 3D model of ZmCTM-B73 (yellow). The 5-M-THF docked into ZmCTM-B73 is shown as grey sticks, and all atoms are coloured according to the element (carbon, grey; oxygen, red; nitrogen, blue). (**h**) Close-up view of the ligand binding pocket of ZmCTM-B73. The colour representations of ZmCTM-B73 and the 5-M-THF molecule are identical to those shown in (**g**). Cartoons of SbCTM and SsFT are included for clarity. The amino acid N228 in ZmCTM-B73 is shown as a stick, and the colour representation is identical to that of 5-M-THF. See Extended Data Fig. 1, Extended Data Tables 3 and 4, and Supplementary Results 3.

Fig. 3. Asn-to-Gly substitution of ZmCTM significantly affects its catalytic activity. Kinetics analysis of ZmCTM and SbCTM proteins. The velocity of the product MeFox generation is plotted against the concentration of substrate 5-M-THF, which ranges from 50 μM to 900 μM. The concentrations of ZmCTM and SbCTM are fixed at 0.5 μM, respectively. Black and white squares represent the velocity values of wild ZmCTM from B73 and its mutant ZmCTM-B73^{N228G}, respectively, in **(a)**. The black and white circles represent the values of wild-type ZmCTM from Qi319 and ZmCTM-Qi319^{G232N}, respectively, in **(b)**. Black and white diamonds represent the values of wild-type SbCTM and its mutant SbCTM^{G219N}, respectively, in **(c)**. Each point represents the mean value of two independent measurements, and the error bars represent standard deviation. Solid black lines fitting

curves were obtained using the Michaelis-Menten equation. **(a,b,c)**. **(d)** k_{cat} and K_m values, as well as errors resulting from fitting, were calculated using GraphPad Prism software. (**e**) Binding pocket and binding model of 5-M-THF with ZmCTM-B73. The relative position of N228 in ZmCTM-B73 is shown. N228 is shown as a yellow stick and 5-M-THF is shown as a grey stick. All atoms are coloured according to the element (carbon, grey; oxygen, red; nitrogen, blue). N, asparagine. (**f**) Binding pocket and binding model of 5-M-THF with ZmCTM-Qi319. The relative position of G232 in ZmCTM-Qi319 is shown. G232 is shown as a green stick and 5-M-THF is shown as a grey stick, and all atoms are coloured according to the element (carbon, gray; oxygen, red; nitrogen, blue). G, glycine. See Extended Data Table 5.

Fig. 4. Potentials of CTMs for folate biofortification in crops. (**a**) Folate profiles (mean ± SD of three biological replicates with n = 3 each) in young seeds (25 DAP) of *ZmCTM*-edited (CRISPR) and wild-type maize (WT). *P*-value (Student's *t*-test, $P < 0.05$) represents *ZmCTM*-edited compared with the wild-type maize. (**b**) Folate profiles (mean ± SD of three biological replicates with n = 3 each) in young seeds (25 DAP) of the wild-type (WT), *ZmCTM-B73* (OE-B73) and *ZmCFTM-Qi319* (OE-Qi319) overexpressing maize lines. A in the brackets after OE-B73 represents A-allele at S2069, and G in the brackets after OE- Qi319 represents G-allele at S2069. The C01 inbred line (the genetic transformation recipient) was used as the wild-type (WT), and the endogenous *CTM* gene carries G at S2069. *P*-values (Student's *t*-test, $P < 0.05$) represent comparisons among *ZmCTM-B73* and *ZmCFTM-Qi319* overexpressing transgenic, and wild-type maize. n.s., not significant. (**c**) Ratios of MeFox/5-M-THF in young seeds (20-24 DAP at harvest stages) of sweetcorn. (**d**) Total folates in young seeds of sweetcorn. (**e**) Folate profiles in young seeds of sweetcorn. *P*-values (Student's *t*-test, $P < 0.05$) represent comparisons among homozygous G, homozygous A, and heterozygous G/A. n.s., not significant. See Supplementary Data 3 and Extended Data Fig 2.

Accession codes

Gene resequencing data are available under GenBank accession codes KT727273-KT727912. Isolated

coding sequences of ZmCTM in maize inbred lines B73 and Qi319 are available under GenBank accession codes KT727913 and KT727914, respectively. ZmCTM (ZmCTM-B73), protein sequence from *Zea mays* inbred line B73, NP_001130076.1; ZmCTM-Qi319, protein sequence from *Zea mays* inbred line Qi319, AMK92167.1; SbCTM, protein sequence from *Sorghum bicolor*, XP_002466878.1; The atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank (PDB, http://www.rcsb.org) with the accession codes 7DYH. TaCTM, protein sequence from *Triticum aestivum*, KAF6990789.1; PaCTM, protein sequence from *Prunus avium*, XP_021832372.1; StCTM, protein sequence from *Solanum tuberosum*, XP_006357514.1; SsFTCD, protein sequence from *Sus scrofa*, NP_999440.1; RnFTCD, protein sequence from *Rattus norvegicus*, NP_446019.1.

Acknowledgements

We thank the research associates at the Center for Protein Research (CPR), Huazhong Agricultural University and Metabolome Platform of Public Laboratory in Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, for their technical support. This research was financially supported by the National Key Research and Development Program of China, China (2016YFD0100503 to L. J., 2020YFE0202300 to J. Y.), National Natural Science Foundation of China, China (31870283 to L. J., 31770878 to D.Z.), Shanghai Agriculture Applied Technology Development Program, China (Z20180103), Agricultural Science and Technology Innovation Program, China (ZDRW202010 to C.Z.), National Program on Key Basic Research Project of China, China (2013CB127003 to C. Z. and J. Y.), Fundamental Research Funds for the Central Universities, China (Program 2662017PY031 to P. Y.), Agricultural Science and Technology Innovation Program, China (ASTIP; CAAS-ZDRW202010 to C.Z.).

Author contributions

C. Z., J. Y., D. Z., and P.Y. designed and supervised this study. W.G., W.W., and H.L. performed GWAS data analysis. L.J., T.L., Q.L., J.A.L., Q.Q.L., and L.P. performed the analysis on transgenic plants. L.J., T.L., Q.L., and J.A.L. performed the folate measurement. D.Z., Y.W., P.Y. and Y.X. performed the protein crystallization and enzymatic analysis. W.X.W. performed molecular simulation analysis. L. J., G. W., D. Z., W. W., J. Y., and C. Z. prepared the manuscript with inputs from other authors. L. J, W. G, Y.W., W. W contributed equally to this work. All authors read and commented on the manuscript.

Competing interest

The authors declare no competing interests.

Data availability

The data supporting the findings of this study are available within the paper, and its **related manuscript files** (**Supplementary Data** and **Supplementary Information)**. **Supplementary Data** includes Supplementary Data 1, folate profile for GWAS in relation to Fig. 1b; Supplementary Data 2, association analysis of folates and resequencing of ZmCTM in relation to Fig. 1c; and Supplementary Data 3, folate profile in sweetcorn in relation to Fig. 4c–e. **Supplementary Information** includes Supplementary Results 1 (allelic comparison of natural variation), Supplementary Results 2 (phenotypic analyses of *ZmCTM* transgenic maize plants), Supplementary Results 3 (construction of 3D model for ZmCTM protein), and Supplementary Results 4 (folate profile and phenotypic analyses of *OsCTM*-RNAi rice plants). All other data that support the findings of this study are available from the corresponding authors on request.

Additional information

Extended data figures and tables

Extended Data Fig. 1 Global structure and binding pocket of SbCTM (PDB: 7DYH), SsFT (PDB: 1QD1), and 3D model of ZmCTM-B73. Related to Fig. 2. (**a**) Global structure of SbCTM. (**b**) Binding pocket of SbCTM. (**c**) Global structure of SsFT. (**d**) Binding pocket of SsFT. (**e**) Global structure of ZmCTM-B73. (**f**) Binding pocket of ZmCTM-B73. (**g**) 3D binding model of 5M-THF with ZmCTM-B73. Residues in the binding pocket are shown as lines and 5-M-THF is shown as a grey stick. Hydrogen *bond*s are shown as dashed lines (**h**) Two-dimensional interaction map of 5M-THF with binding pocket residues.

Extended Data Fig. 2 Folate profiles of *OsCTM*-RNAi rice plants and overexpression of *ZmCTM-B73* in Arabidopsis. Related to Fig. 4. (a) Folate profile (mean ± SD of five biological replicates with n = 3 each) in mature seeds of wild-type (WT) and *OsCTM*-RNAi (*OsRNAi*) rice plants. *P*-value (Student's *t*-test, $P < 0.05$) represents *OsCTM*-RNAi compared with the wild-type rice. (**b**) Folate profile (mean ± SD of two biological replicates with n = 3 each) in 30-day rosette leaves of *ZmCTM-B73* transgenic (OE-B73) and wild-type *Arabidopsis* plants. A in the brackets after OE-B73 represents A-allele at S2069. *P*-value (Student's *t*-test, $P < 0.05$) represents *ZmCTM-B73* overexpressor compared with the wild-type *Arabidopsis*. n.s., not significant.

Extended Data Table 1. Phenotypic variation and heritability analysis in the association panel (n = 531). Related to Fig. 1.

Extended Data Table 2. Genes associated with the content of 5-F-THF-M identified in genome-wide association analysis. Related to Fig. 1.

Extended Data Table 3. Data collection and refinement statistics. Related to Fig. 2.

Extended Data Table 4. Virtual screen against ZmCTM-B73 protein. Related to Fig. 2.

Extended Data Table 5. Docking data of ZmCTM-Qi319 and ZmCTM-B73 with 5-M-THF. Related to Fig. 3.

Extended Data Table 6. Primers used in this study. Related to Methods

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see: http://www.textcheck.com/certificate/index/ZBn29p

Correspondence and requests for materials should be addressed to D.Z., J.Y., or C.Z.

Figures

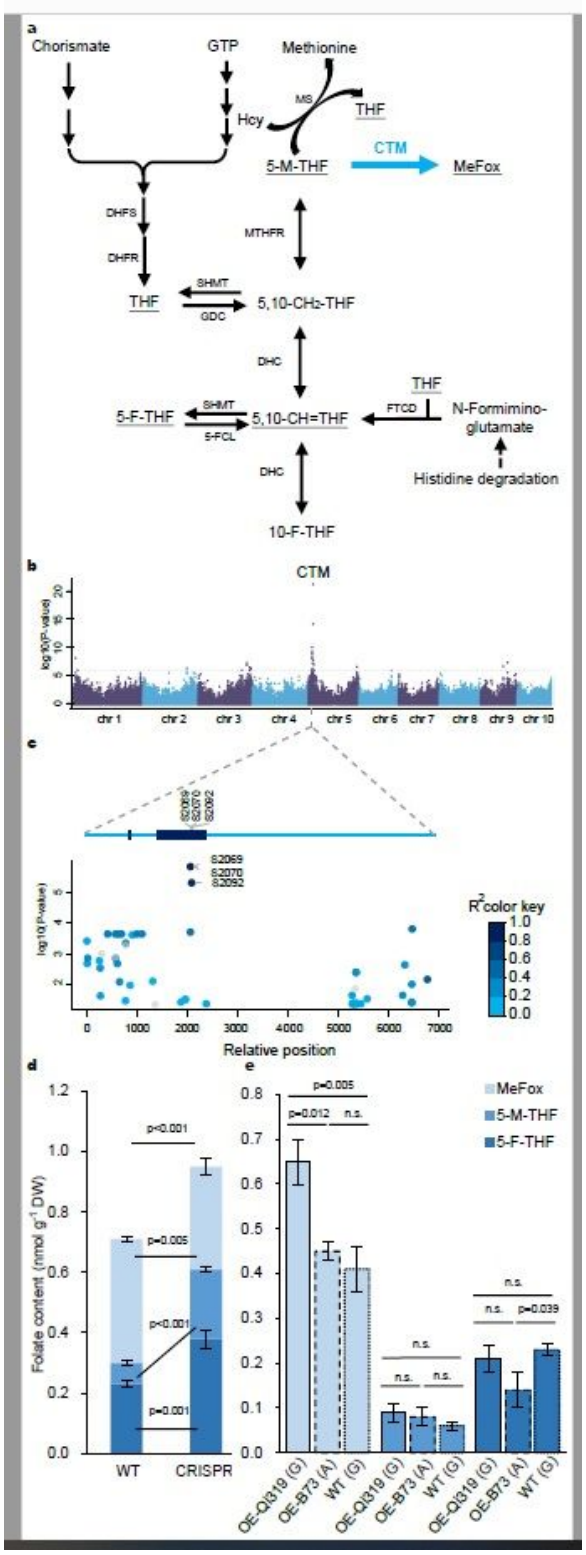


Figure 1

GWAS and identification of the major gene locus that underlies folate variation. (see Manuscript file for full figure legend)

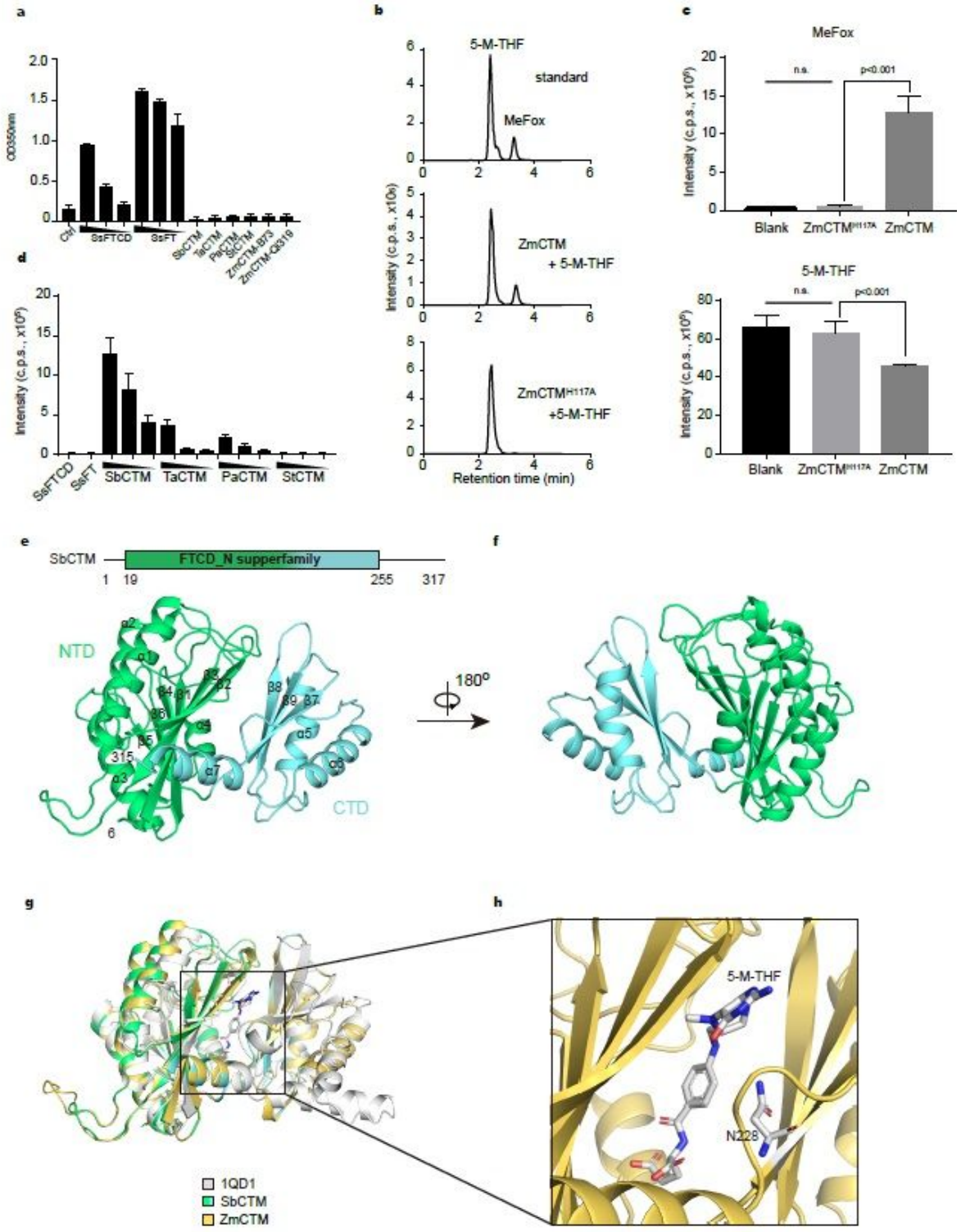
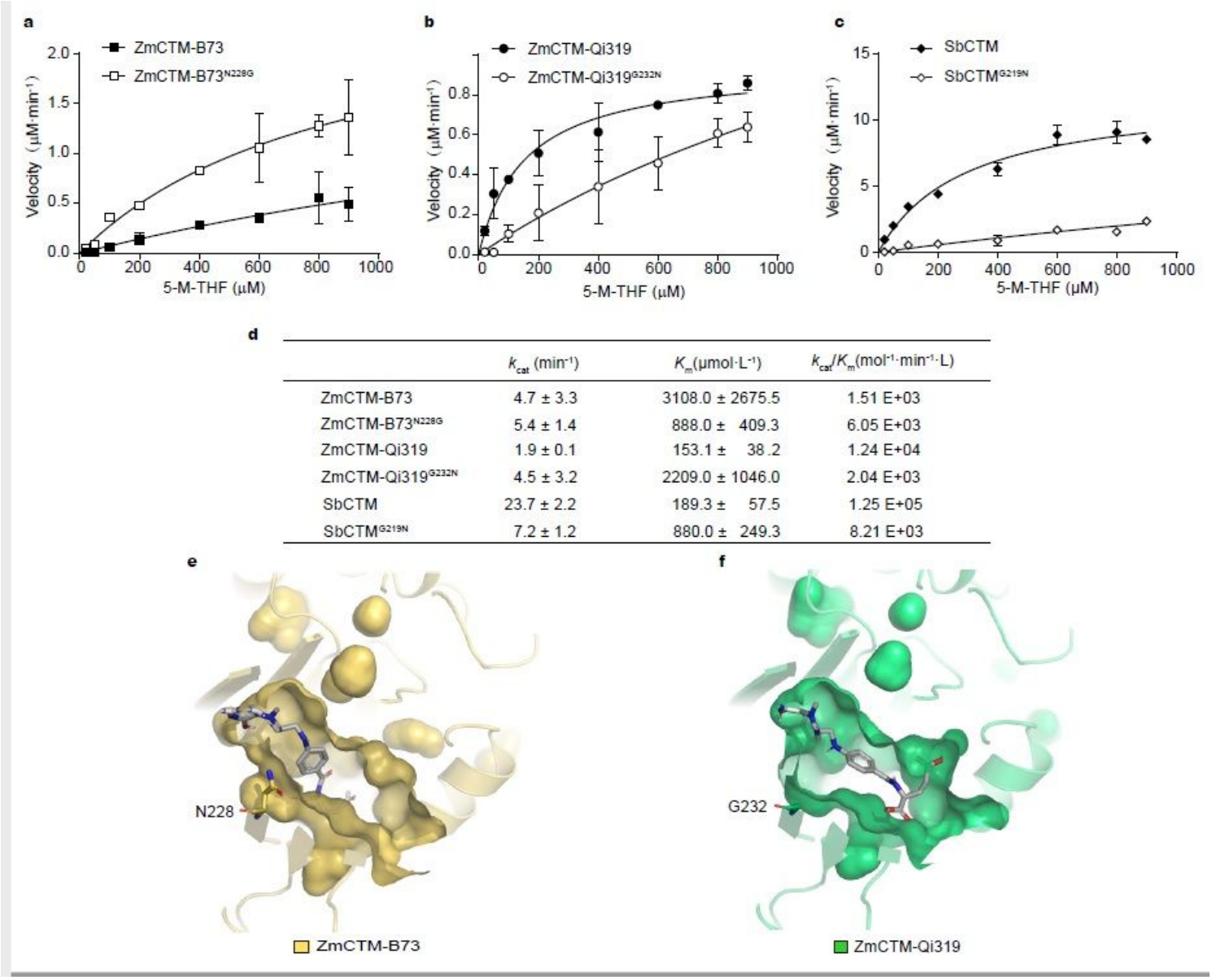


Figure 2

ZmCTM directly converts 5-M-THF to MeFox. (see Manuscript file for full figure legend)



d

	k_{cat} (min^{-1})	K_m(μmol·L^{-1})	k_{cat}/K_m(mol^{-1}·min^{-1}·L)
ZmCTM-B73	4.7 ± 3.3	3108.0 ± 2675.5	1.51 E+03
ZmCTM-B73^{N228G}	5.4 ± 1.4	888.0 ± 409.3	6.05 E+03
ZmCTM-Qi319	1.9 ± 0.1	153.1 ± 38.2	1.24 E+04
ZmCTM-Qi319^{G232N}	4.5 ± 3.2	2209.0 ± 1046.0	2.04 E+03
SbCTM	23.7 ± 2.2	189.3 ± 57.5	1.25 E+05
SbCTM^{G219N}	7.2 ± 1.2	880.0 ± 249.3	8.21 E+03

Figure 3

Asn-to-Gly substitution of ZmCTM significantly affects its catalytic activity. (see Manuscript file for full figure legend)

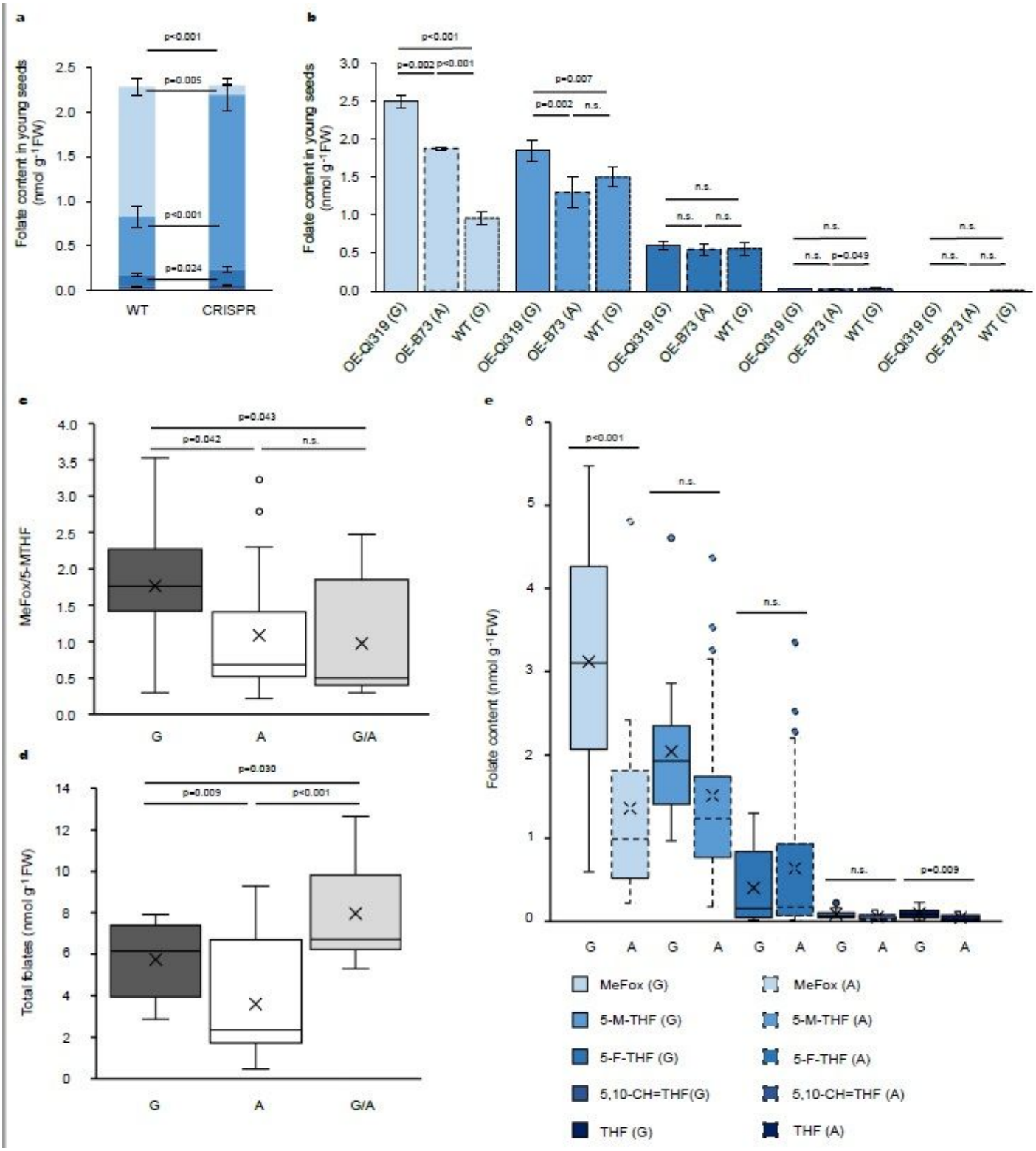


Figure 4

Potentials of CTMs for folate biofortification in crops. (see Manuscript file for full figure legend)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 20210309CTMsupplementalmaterial.pdf