Supporting Information

Daimon et al. 10.1073/pnas.1000479107

SI Materials and Methods

Insects and Cell Lines. Silkworms were reared on an artificial diet or mulberry leaves at 25 °C under standard conditions. Silkworm strains p50T and c108T, and the related species, *B. mandarina*, were maintained in the University of Tokyo, and the other strains were maintained in the National Institute of Agrobiological Sciences Genebank. The *Spodoptera frugiperda* Sf9 cell line was maintained as described previously (1).

HPLC Analysis of Flavonoids in Tissues and Cocoons. Silkworm larvae were reared on a diet containing mulberry leaf powder from hatching to the fourth ecdysis. To investigate the effect of *Gb* on the metabolism of a flavonoid quercetin, the newly molted larvae were fed a semisynthetic diet (2) supplemented with 0.5 g quercetin/100 g (dry weight) for 6 d. On the last day of the feeding experiment, the hemolymph, midgut, and silk glands were collected. Some larvae were transferred to a cage for cocooning. Cocoons were harvested after the larvae pupated. Flavonoids in the tissues and cocoons were identified and quantified as previously described (2) using a Shimadzu LC10 HPLC system equipped with an SPD-7AV UV-Vis detector set at 365 nm. Samples were loaded onto a Nova-Pak C18 column (150 × 3.9-mm i.d., Waters). Elution was performed using H_2O:HCO_2H (99.8:0.2 V/V) as solvent A and CH_3CN as solvent B at a flow rate of 1.0 mL/min at 40 °C. Separation was achieved with a linear gradient from 7% B to 40% B for 20 min and then to 100% for 10 min. Quercetin metabolites were identified based on their coelution with either commercially available standards or authentic reference samples purified from the cocoon shell of Pure-Mysore. Flavonoids were quantified using quercetin 5-*O*-glucoside as a reference compound.

Determination of UGT Activity. Silkworm larvae were reared on a diet containing mulberry leaf powder from hatching to the fourth ecdysis. To investigate the tissue distribution of UDP-glucosyltransferase, which catalyzes the glucosylation of quercetin, the newly molted larvae were fed a semisynthetic diet (2) for 5 d, after which the larvae were dissected and the midgut, anterior, middle, and posterior silk glands, fat body, hemocytes, Malpighian tubules, trachea, testis, and ovary were collected. The collected tissues were homogenized in 4 volumes (V/W) 0.1 M potassium phosphate buffer (pH 7.0). The precipitates were rehomogenized in 0.1 M potassium phosphate buffer (pH 7.0) and centrifuged at 23,000 × *g* for 20 min at 4 °C. The precipitates were rehomogenized in 0.1 M potassium phosphate buffer (pH 7.0), and the homogenate was used as the enzyme solution. The reaction mixtures consisted of 0.1 M potassium phosphate (pH 7.0), 5 mM UDP-glucose, 25 mM $MgCl_2$, 5 mM D-glucuronolactone (as an inhibitor of β-glucosidase), 0.1 mM quercetin, and enzyme preparation in a final incubation volume of 0.1 mL The enzyme reactions were started by addition of UDP-glucose, and then incubated for 30 min at 37 °C. The reaction was stopped by the addition of 0.2 mL methanol. Precipitated protein was removed by centrifugation at 20,000 × *g* for 10 min, and the supernatant was used for the determination of glucosides by HPLC. Control incubations were run, in which the enzyme source, substrates, or UDP-glucose was omitted. The glucosyl metabolites formed by tissue extracts were measured on the basis of the peak area obtained in the HPLC chromatogram at 365 nm, using commercially available standards or authentic reference samples purified from the cocoon shell, as described above. The protein concentration was measured with a commercial assay kit (Coomassie Plus; Pierce) using BSA as a standard.

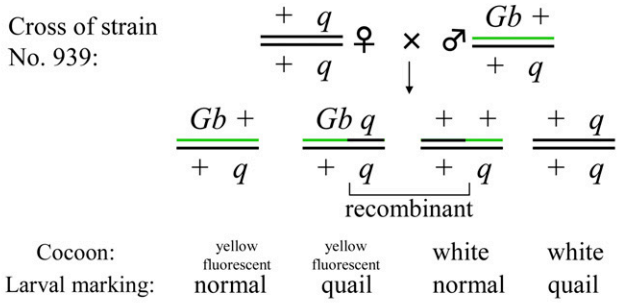
Cloning of *UGT* genes. Total RNAs extracted from midgut, fat bodies, and silk glands of fifth-instar day 3 larvae of strain p50T were mixed and used for 5′- and 3′-rapid amplification of cDNA ends (RACE) using GeneRacer Kit (Invitrogen). PCR was performed following the manufacturer's protocol using primers listed in Table S3. The PCR products were subcloned and sequenced as previously described (3).

RT-PCR. Total RNAs were isolated from various tissues of fifth-instar day 3 larvae of the p50T strain and reverse transcribed as previously described (4). Semiquantitative RT-PCR analysis for UGT expression was performed using primers listed in Table S3. PCR conditions were as follows: 30–35 cycles of 94 °C for 30 s, 54 °C for 30 s, and 72 °C for 1.5 min. The identity of each RT-PCR product was confirmed by DNA sequencing.

Expression and UGT Activity of Recombinant UGTs. The coding sequence of *UGT* genes was amplified by RT-PCR using primers listed in Table S3 and cloned into a pFastBac vector (Invitrogen). Recombinant *Autographa californica* nucleopolyhedrovirus (AcMNPV) expressing each UGT with a His6-tag at the C terminus was constructed using a Bac-to-Bac System (Invitrogen). Monolayers of Sf9 cells cultured in 150-mm dishes were infected with recombinant viruses at a multiplicity of infection of 10. After 72 h, cells were collected, washed twice with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, and 1.4 mM KH_2PO_4), and stored at −80 °C. The expression of UGT proteins was confirmed by immunoblot analysis using a Penta-His antibody (Qiagen) as previously described (4). The UGT activities in mock- or virus-infected cells were measured as described above. The expression level of each UGT was determined by immunoblot using a LAS-100 luminoimager (Fuji Film). We could estimate the relative expression level of UGT, as the reactivity of the antibody to each UGT appeared to be similar. The Q5GT activity of each recombinant UGT was normalized according to the relative expression level of UGTs, expressed as "relative Q5GT activity (%)," and the level of Q5GT activity in cells infected with the virus expressing *Bm-UGT10286* was set to 100%.

1. Daimon T, Katsuma S, Iwanaga M, Kang W, Shimada T (2005) The *BmChi-h* gene, a bacterial-type chitinase gene of *Bombyx mori*, encodes a functional exochitinase that plays a role in the chitin degradation during the molting process. *Insect Biochem Mol Biol* 35:1112–1123.
2. Hirayama C, Ono H, Tamura Y, Konno K, Nakamura M (2008) Regioselective formation of quercetin 5-*O*-glucoside from orally administered quercetin in the silkworm, *Bombyx mori*. *Phytochemistry* 69:1141–1149.
3. Daimon T, et al. (2003) A *Bombyx mori* gene, *BmChi-h*, encodes a protein homologous to bacterial and baculovirus chitinases. *Insect Biochem Mol Biol* 33:749–759.
4. Daimon T, et al. (2008) *β-Fructofuranosidase* genes of the silkworm, *Bombyx mori*: insights into enzymatic adaptation of *B. mori* to toxic alkaloids in mulberry latex. *J Biol Chem* 283:15271–15279.

A



B

Phenotype	individual #	AFLP marker 7–fr	C3–fr	D1–fr
Yellow fluorescent cocoon	2	GW	GW	W
	5	GW	W	W
	6	GW	Gw	W
	13	GW	W	W
	77	W	GW	GW
White cocoon	1	W	GW	GW
	3	W	W	GW
	4	W	W	GW
	6	W	W	GW
	7	W	GW	GW
	11	GW	W	W
	38	GW	W	W

C

Phenotype	individual #	AFLP marker 7–fr	10287–2fr	72–fr	10289–1fr	C3–fr	D1–fr
White cocoon & normal marking	5	W	W	W	W	GW	GW
	27	W	W	W	W	GW	GW
	31	W	W	W	W	GW	GW
	38	W	W	W	W	GW	GW
	51	W	W	W	W	GW	GW
	56	W	W	W	W	GW	GW
Yellow fluorescent cocoon & quail marking	68	GW	GW	GW	GW	W	W
	74	GW	GW	W	W	W	W
	83	GW	GW	GW	GW	W	W

D

Primers for PCR markers

marker	forward primer (5′ to 3′)	reverse primer (5′ to 3′)	position in nscaf2986	size (bp) of PCR product Gb/+	+/+
7–fr	GGGTCTTTAGGCCACTCAATGC	CATGTGTAAGCTGTGACGTCAG	5,067,488	800 + 600	600
10287–2fr	CTCAATGATCCAGAACAGAAGT	GAAGGTGGCACCAGACCTCTC	5,557,252	1300 + 1100	1100
72–fr	GCGACCTAGATCAATTCTGC	ACGCACTTTGAATGTGCTGC	5,572,042	1800 + 1300	1800
10289–1fr	GATCTACCGAAGGAGATTAAAG	CTGTTGTCGATAGAAGACCTCC	5,597,030	1500 + 1100	1500
C3–fr	GCTGTTTCACTTAACTTTTCCAC	CGCACTGGCAGTATACAATTACC	5,794,302	1000 + 700	700
D1–fr	GTGGACTATGAGACAAAGTCTGGG	GCTTTGCAATCAATGCTAAATTTGG	5,987,327	1000 + 770	770

Fig. S1. Detailed procedure for positional cloning of *Gb*. (*A*) Mating scheme for mapping *Gb*. Cross between homozygous quail female with white cocoon (+ *q*/+ *q*) and heterozygous WT male with yellow fluorescent cocoon (*Gb* +/+ *q*) of strain 939 produced F_1 offspring for mapping. (*B*) Results of rough mapping and (*C*) fine mapping of *Gb*. Because two chromosomes of this strain, *Gb* + (green line in *A*) and + *q* (black line in *A*), are of different origin, codominant PCR markers were designed that could distinguish each chromosome. GW, heterozygous genotype; W, + *q* homozygous genotype. (*D*) Codominant PCR markers used in this study.

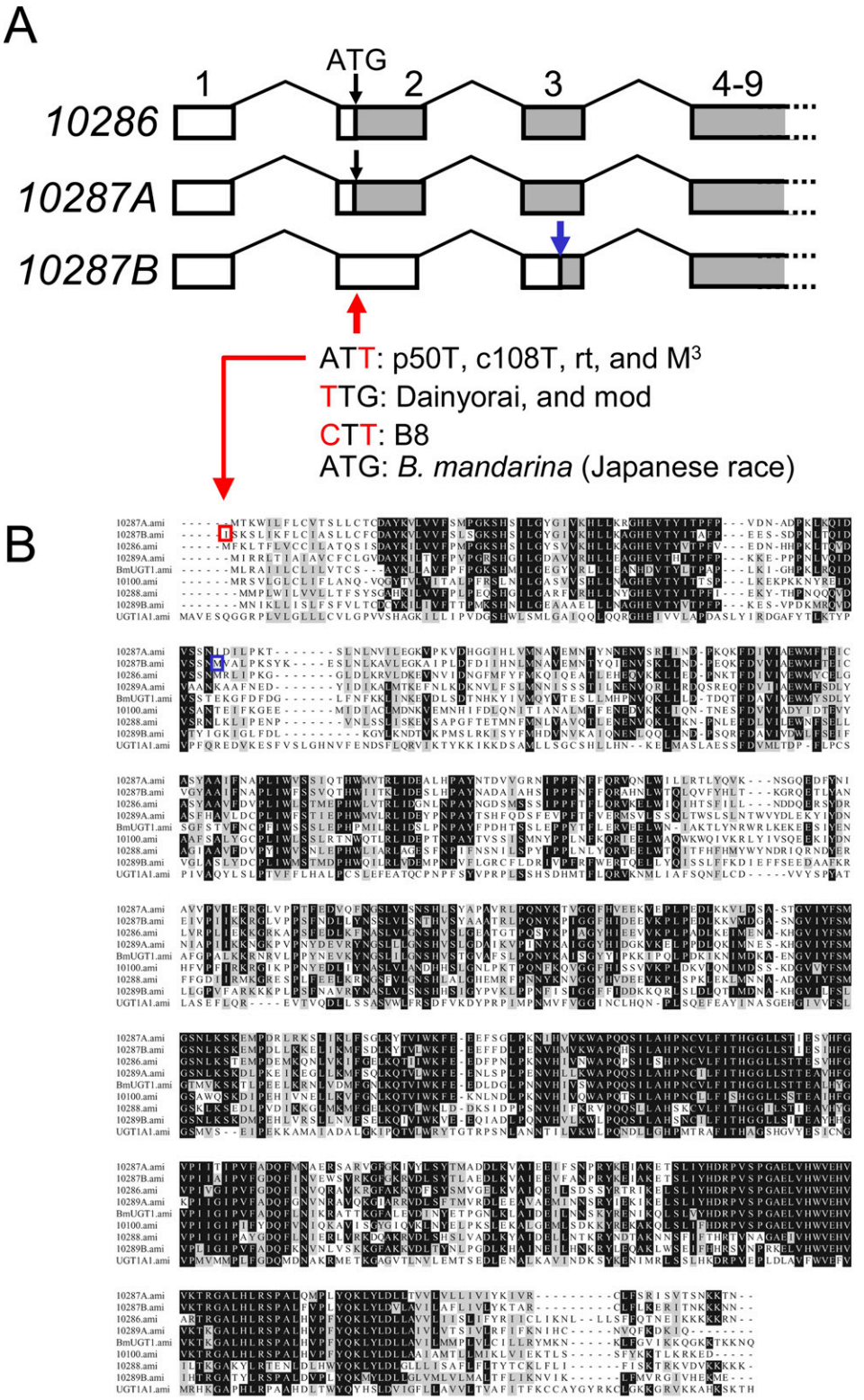


Fig. S2. Comparisons of genome structure and amino acid sequences of UGTs. (*A*) Putative point mutations at the translation start site of *Bm-UGT10287B*. Genome structures are highly conserved among *Bm-UGT10286*, *10287A,* and *10287B*. In *Bm-UGT10286* and *10287A*, translation starts from fourth nucleotide of second intron (black arrow). However, the nucleotide sequence of *Bm-UGT10287B* at this position (red arrow) is highly variable, and no silkworm strain had ATG codon here; instead, the first start codon was present at the third exon (blue arrow), causing truncation of 75 amino acid residues at the N terminus. (*B*) Alignment of amino acid sequences of UGTs. Amino acid sequences of 7 UGTs were aligned with those of BmUGT1 (AF324465) and human UGT1A1 (NP_000454) with Clustal X program (1). If *Bm-UGT10287B* is translated from same position as *Bm-UGT10286* and *10287A* (isoleucine, red box), the deduced amino acid sequence shows high homology to other UGTs, suggesting recent nucleotide substitutions at this site. It is also noteworthy that these substitutions might have occurred during the domestication process of the silkworm, as *B. mandarina*, an ancestral species of *B. mori*, retains the ATG codon here.

1. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.

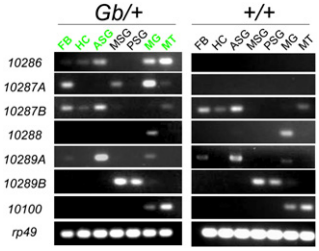


Fig. S3. RT-PCR analysis of *UGT*s in strain 939 (*Gb*/+ and +/+). Semiquantitative RT-PCR analysis of seven *UGT*s in each genotype of strain 939. Total RNAs from fifth instar day 3 larvae were analyzed. Tissues that showed Q5GT activity (Table 1) are highlighted in green. Abbreviations for tissues are the same as in Fig. 3. As a control, *rp49* was analyzed. Note that expression of *Bm-UGT10286* and *10287A* was not detected in +/+ individuals because of the genomic deletion, but that this deletion did not change the spatial expression patterns of other *UGT*s.

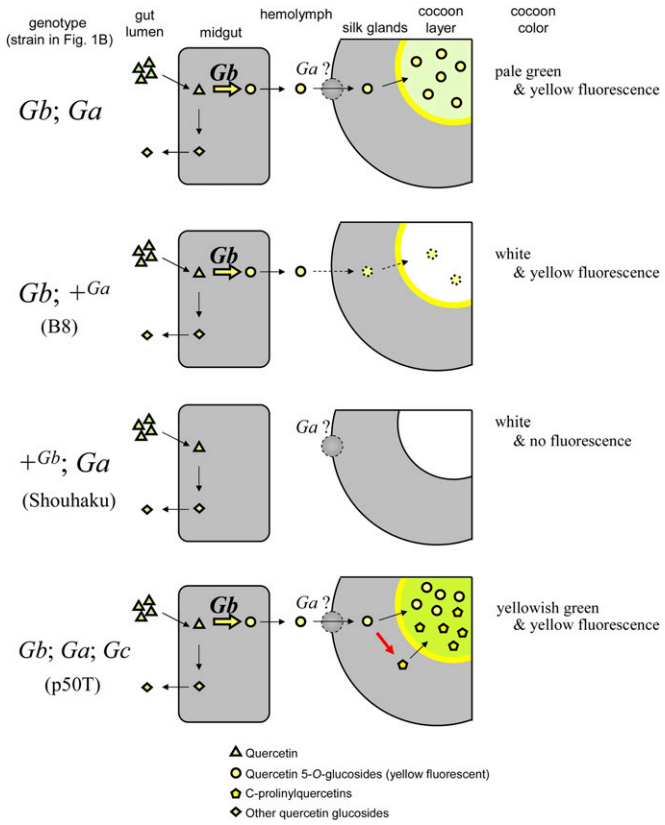


Fig. S4. Model of metabolism, uptake, and transport of dietary quercetin in the silkworm. In *Gb* larvae, dietary quercetin is glucosylated at the 5-*O* position in the midgut as the first-pass metabolite by the action of *Gb* (yellow arrow) (this study). Such 5-*O*-glucosylated quercetins are transported to the silk glands and incorporated into cocoon shells. If *Ga* (Green a, locus unknown) coexists with *Gb*, uptake of 5-*O*-glucosylated quercetins in silk glands is greatly facilitated (dotted gray circle), resulting in pale green coloration of cocoon shells with yellow fluorescence. In contrast, if *Gb* larvae do not carry *Ga* (*Gb*; $+^{Ga}$), the uptake of quercetins in the silk glands may be greatly reduced, which could impair green coloration of cocoon shells. However, yellow fluorescence may still be visible because the trace amount of quercetin 5-*O*-glucosides contained in the cocoons is sufficient for the production of detectable yellow fluorescence (strain B8 in Fig. 1*B*). In $+^{Gb}$ larvae, dietary quercetins are not glucosylated at the 5-*O*-position, and this may dramatically reduce the uptake of dietary quercetins, resulting in white cocoon shells with no yellow fluorescence (strain Shouhaku in Fig. 1*B*). In strains carrying another locus involving the green cocoon trait, *Gc* (Green c, chromosome 15, locus unknown), quercetins in the silk glands are further metabolized to C-prolinylquercetins by the action of *Gc* (red arrow), causing yellowish-green coloration of cocoon shells with yellow fluorescence (p50T in Fig. 1*B*). Our data demonstrate that *Gb* controls the first critical step in the biosynthesis of cocoon flavonoids in the silkworm.

Table S1. Quantification of quercetin metabolites in *Gb*/+ and +/+ individuals of strain 939

Genotype	Flavonoid	Midgut (nmol/g tissue)	Hemolymph (nmol/mL)	Silk glands (nmol/g tissue)	Cocoon (nmol/g)
Gb/+	Quercetin 5-*O*-glucoside	21.6 ± 7.6	11.7 ± 2.1	20.3 ± 4.9	67.9 ± 14.3
	Quercetin 5,4′-di-*O*-glucoside	4.5 ± 0.69	49.8 ± 3.9	165.1 ± 58.2	767.7 ± 188.3
	Quercetin 5,3′-di-*O*-glucoside	3.2 ± 0.6	4.3 ± 0.5	9.6 ± 2.2	36.7 ± 7.5
	Quercetin 7-*O*-glucoside	—	1.7 ± 0.3	46.8 ± 7.7	133.2 ± 53.2
	Quercetin 4′-*O*-glucoside	—	0.5 ± 0.1	27.2 ± 15.9	85.0 ± 17.2
	Quercetin 3-*O*-glucoside	6.6 ± 1.2	—	10.7 ± 6.1	37.8 ± 19.0
	Quercetin 3,7-di-*O*-glucoside	—	9.5 ± 1.1	22.5 ± 5.5	86.0 ± 19.5
	Quercetin 3,3′-di-*O*-glucoside	—	4.2 ± 0.6	30.3 ± 17.6	143.6 ± 32.7
	Unknown	—	—	—	—
	Quercetin	6.0 ± 1.1	—	4.5 ± 2.6	14.4 ± 8.3
	Total	**41.9 ± 9.6***	**81.8 ± 7.4***	**337.1 ± 82.7***	**1369.6 ± 242.1***
+/+	Quercetin 5-*O*-glucoside	—	—	—	—
	Quercetin 5,4′-di-*O*-glucoside	—	1.1 ± 0.4	—	—
	Quercetin 5,3′-di-*O*-glucoside	—	—	—	—
	Quercetin 7-*O*-glucoside	—	2.1 ± 0.7	2.9 ± 0.7	—
	Quercetin 4′-*O*-glucoside	—	0.3 ± 0.2	—	—
	Quercetin 3-*O*-glucoside	8.8 ± 3.0	0.9 ± 0.5	12.7 ± 2.4	49.9 ± 7.6
	Quercetin 3,7-di-*O*-glucoside	—	8.1 ± 2.8	11.5 ± 3.9	73.1 ± 20.7
	Quercetin 3,3′-di-*O*-glucoside	—	1.9 ± 0.5	2.4 ± 0.5	52.1 ± 10.7
	Unknown	2.7 ± 0.3	2.7 ± 0.4	—	—
	Quercetin	3.2 ± 1.9	—	9.2 ± 6.5	28.2 ± 23.4
	Total	**15.7 ± 4.9**	**17.3 ± 4.7**	**34.1 ± 11.6**	**196.3 ± 15.5**

Values are mean ± SD (n = 5). —, not detected.
*Total amount in each tissue was significantlly different between Gb/+ and +/+ ($P < 0.01$, t test).

Table S2. Glucosyltransferase activity toward quercetin in tissues from strain p50T

Tissue	UGT activity (nmol/min per mg protein) 3-*O*-glc	5-*O*-glc	7-*O*-glc	3′-*O*-glc	4′-*O*-glc
Fat body	2.24 ± 0.28	0.18 ± 0.16	0.85 ± 0.14	0.20 ± 0.07	4.73 ± 0.30
Trachea	—	—	—	—	—
Hemocyte	0.18 ± 0.05	1.76 ± 0.21	0.07 ± 0.02	0.12 ± 0.05	0.14 ± 0.06
Testis	1.09 ± 0.26	—	—	0.32 ± 0.04	0.90 ± 0.19
Ovary	2.37 ± 0.14	—	0.45 ± 0.11	—	1.64 ± 0.33
Anterior silk gland	15.93 ± 4.84	2.36 ± 0.82	—	—	0.69 ± 0.31
Middle silk gland	—	—	7.80 ± 0.73	—	0.71 ± 0.17
Posterior silk gland	—	—	10.67 ± 2.78	—	1.25 ± 0.48
Midgut	0.34 ± 0.09	1.72 ± 0.27	—	—	0.04 ± 0.02
Malpighian tubule	28.96 ± 2.61	7.29 ± 0.74	—	—	—

Values are mean ± SD (n = 5). —, not detected.

Table S3. Primers used in this study

Name	Sequence (5′-3′ end)	Object
10286-RACE-F1	GGAGAAGCAACCGGAACTCCGCAGAGCT	RACE for *Bm-UGT10286*
10286-RACE-R1	CCGCAGCATAACTAGCACCCAATTCGCA	
10287A-RACE-F1	GTACCGAAAGTTGATCACGGTGGGATTCAT	RACE for *Bm-UGT10287A*
10287A-RACE-R1	CCTAGTCTTGACGACGTGCTCCACCCAGT	
10287B-RACE-F1	GAGAGGTCTGGTGCCACCTTCATTCAAT	RACE for *Bm-UGT10287B*
10287B-RACE-R1	CCGCCAACGTGTAAGAAAGATCGACCCTT	
10288-RACE-F1	CAACGCGTAGAGGAGCTTTGGACGCAA	RACE for *Bm-UGT10288*
10288-RACE-R1	CGATCCGTTGCGTTTGAGTTCCTCAAACA	
10289A-RACE-F1	CCTCACACTTTCAGGACAGCTTCGAAGTA	RACE for *Bm-UGT10289A*
10289A-RACE-R1	CGGAGACACCGGACGATCGTGATATATCA	
10289B-RACE-F1	CTCGGTCCCGTATTCGCTAGAAAGAA	RACE for *Bm-UGT10289B*
10289B-RACE-R1	CTGCTGCGGCAACCATTTCAAAACATG	
10100-RACE-F1	GCTTCGCTGGTGCTAGCAAACGATCACCA	RACE for *Bm-UGT10100*
10100-RACE-R1	CACGGCTTTCTGGATGTTGACAAACTGGT	
RT10286-F1	GCTCACATTCCTGGTCTGCT	RT-PCR for *Bm-UGT10286*
RT10286-R1	ACTGGAAGACATAGAATCCC	
RT10287A-F1	CCATTTCCGGTGGATAACGCG	RT-PCR for *Bm-UGT10287A*
RT10287A-R1	TTAACTTGATACAGTGTTCGT	
RT10287B-F1	GCGTTTCCCGAGGAAAGTTCA	RT-PCR for *Bm-UGT10287B*
RT10287B-R1	CTATTTCATTAGCGTACAATGT	
RT10288-F1	ATGCCGTTGTGGATCCTAGT	RT-PCR for *Bm-UGT10288*
RT10288-R1	CGTTATACCAATACATGTGGA	
RT10289A-F1	GATACGAAGACTCACAATTGC	RT-PCR for *Bm-UGT10289A*
RT10289A-R1	GGTACTTCGAAGCTGTCCTGA	
RT10289B-F1	GAATCTGTTCCCGATAAGAT	RT-PCR for *Bm-UGT10289B*
RT10289B-R1	CAACGATGAGATCTGATAAAGT	
RT10100-F1	GCGAATCAAGTTCAAGGCTAT	RT-PCR for *Bm-UGT10100*
RT10100-R1	CGGATAGTTCATTGATATCGA	
pFB10286-F1	CTAGGATCCAACATGTTCAAGCTCACATTC	Construction of expression vector for *Bm-UGT10286*
pFB10286-R1	GCTAGGCCTCAATGGTGATGGTGATGATGATTCCTCTTCTTCTTCTTCTTGATTTC	
pFB10287A-F1	GCAAGGCCTGAAAATGACCAAGTGGATATTGTT	Construction of expression vector for *Bm-UGT10287A*
pFB10287A-R1	CACTGCAGTCAATGGTGATGGTGATGATGATTCGTTTTCTTATTACTAGTTACGCT	
pFB10288-F1	GTCACTAGTAAAATGATGCCGTTGTGGATC	Construction of expression vector for *Bm-UGT10288*
pFB10288-R1	GCACTCGAGTCAATGGTGATGGTGATGATGCTTCTTCTTTTTCACATCCACTTTCCG	
pFB10289A-F1	CATGGATCCGAAAAATGATGATACGAAGACTCACAATTGC	Construction of expression vector for *Bm-UGT10289A*
pFB10289A-R1	GGTGAATTCAATGGTGATGGTGATGATGCTGAATTTTATCTTTAAATTGTACGTTACAGTG	
pFB10289B-F1	GCAACTAGTCGTACGGACGCAGACAGTAA	Construction of expression vector for *Bm-UGT10289B*
pFB10289B-R1	CGTCTCGAGTCAATGGTGATGGTGATGATGTTCCATCTTCTCGTGTACTATACC	
pFB10100-F1	GTAGGATCCGAATGAGGTCGGTTTTAGGATTG	Construction of expression vector for *Bm-UGT10100*
pFB10100-R1	GGTGAATTCAATGGTGATGGTGATGATGATCTTCTTTCCGTTTTAATGTCTTCTT	
del-F1	GTTCAAGCTCACATTCCTGGT	Genotyping of *Gb* locus
del-F2	GACAGTAACTGGCCGTAGCT	
del-R1	GGTTCGGGTCTGAACTTTCCT	
del-R2	TTAGAGCGATCTCCTTGAGAC	