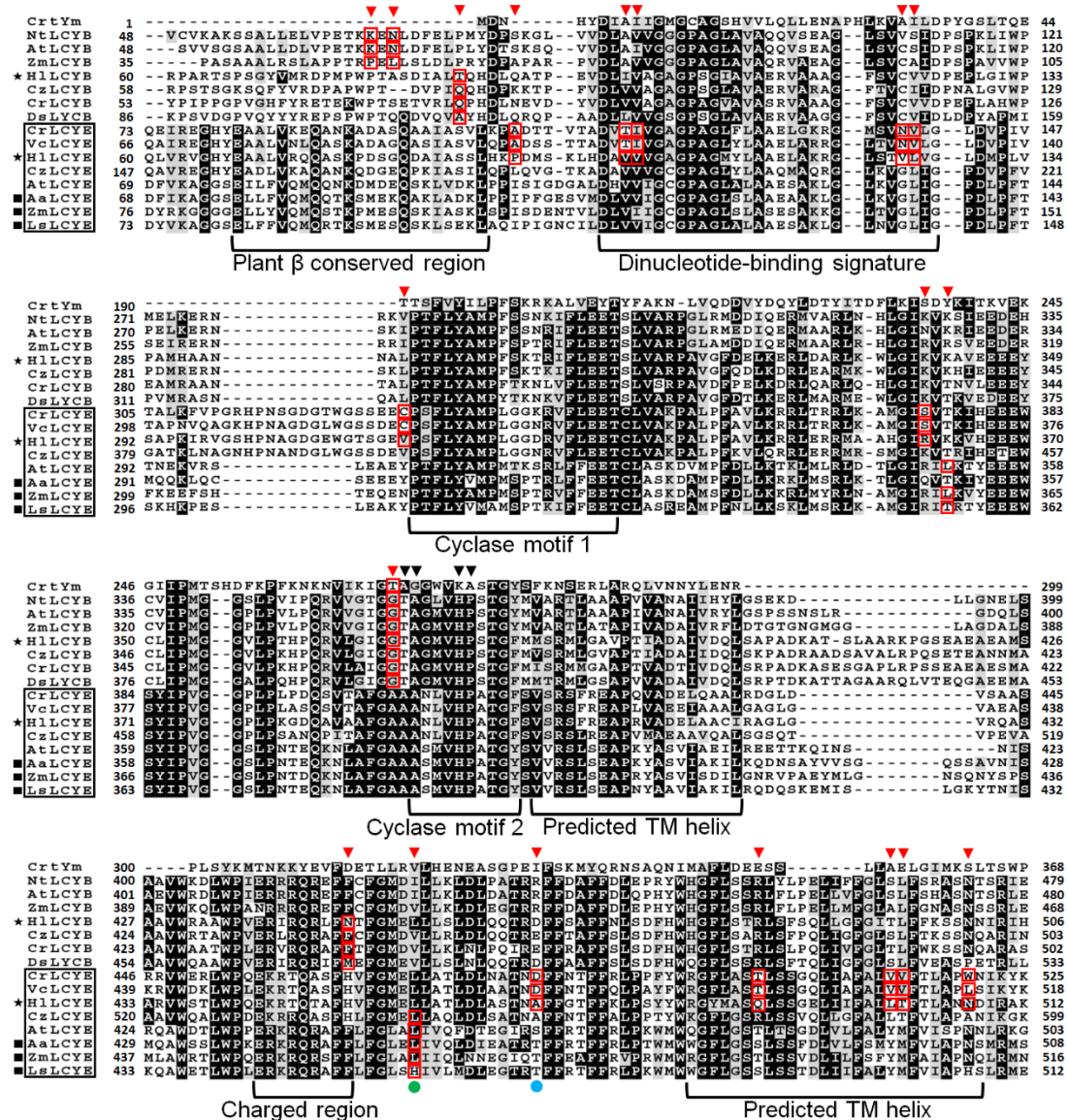
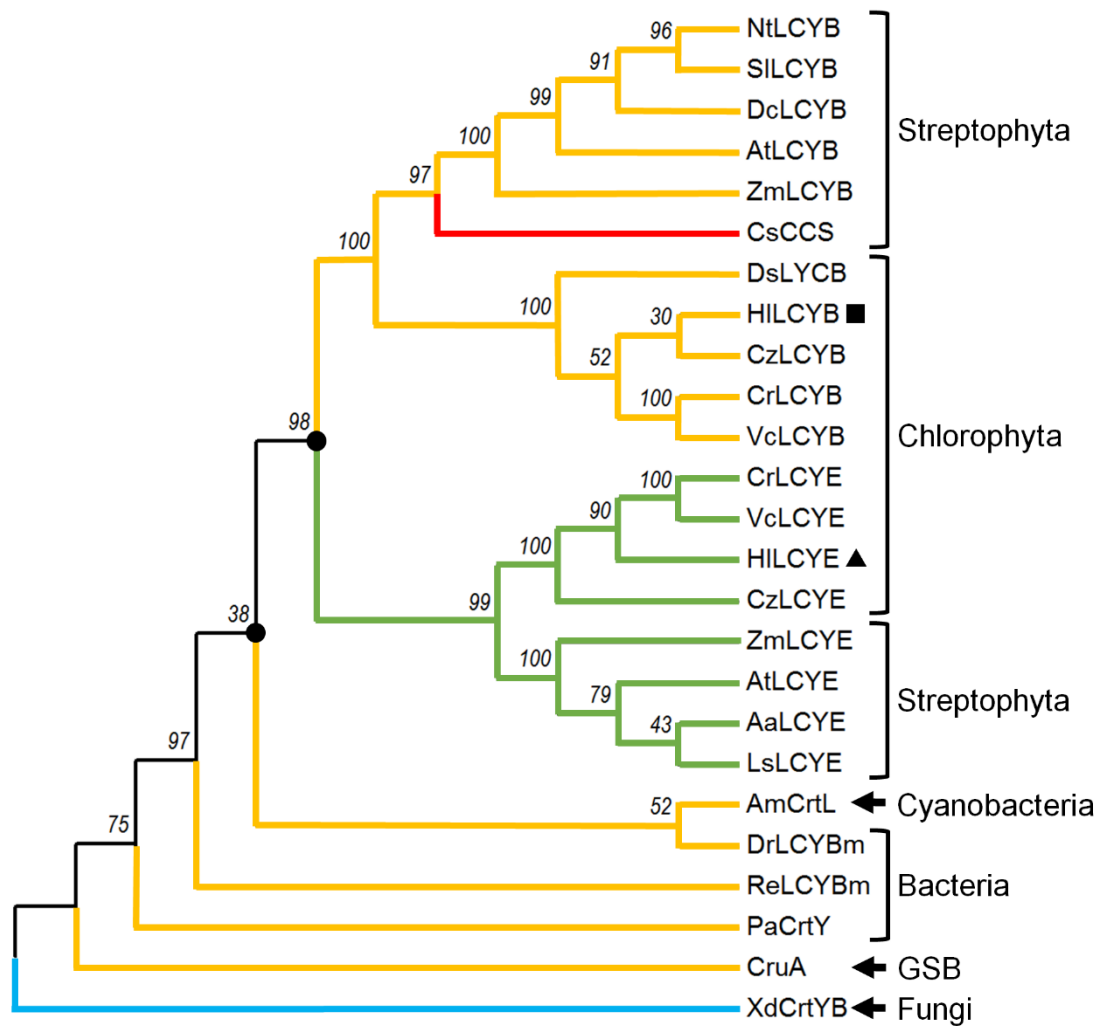


## Supplementary Figures



**Figure S1.** Multiple secondary domains in cyclases. The regulatory motif conserved in all  $\beta$ -bicyclases are highlighted by black arrowheads. Some positively selected sites are red-boxed and indicated by red arrowheads. The two residues varying between  $\epsilon$ -mono- and bicyclases are highlighted by filled circles. Filled rectangle indicated three  $\epsilon$ -bicyclases. Stars indicate the HILCYE and HILCYB characterized in this study.



**Figure S2.** Molecular evolutionary relationships of the *H. lacustris* cyclases by the Neighbor-Joining method. The orange, green, red, and blue lines highlight LCYBs, LCYEs, CCS, and fungal bi-functional cyclase, respectively. The nodes of gene duplication during evolution are dotted. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown above the branches.

## Supplementary Tables

**Table S1.** Primers used in this study.

Manipulation	Primer	Sequence (5' to 3')
<b>Expression plasmid construction</b>		
pACCRT-	EIBYb F	<u>TACCGCATTA<b>AAGCTT</b></u> aggaggattacaaa <b>ATG</b> TTGAGCAAACAACACGATT
EIB-B	EIBYb R	<u>TATCATCGATA<b>AAGCTT</b></u> CTACTTGATCATGGCTGGCG
pACCRT-	EIBYe F	<u>TACCGCATTA<b>AAGCTT</b></u> aggaggattacaaa <b>ATG</b> CAACCACTGCGTGTAGCAAC
EIB-E	EIBYe R	<u>TATCATCGATA<b>AAGCTT</b></u> TTAGGCCTGGTTCGGCTCCA
<b>Site-directed mutagenesis</b>		
LCYE mutants	SP <sup>b</sup>	CCACTGGCAG <b>ACTAGT</b> GTAACAGGATTAGC
	L457H	ATCCAGAGTGCCAG <b>ATG</b> CTCCATGCCAAATAC
	A498S	AGTGAAAGTCAGCAG <b>GCT</b> GAAGATGATCAGCTC
	L457H/A498S <sup>c</sup>	AGTGAAAGTCAGCAG <b>GCT</b> GAAGATGATCAGCTC
	TAGX3KA	CATGAACCCTGTTGA <b>GGCCTT</b> CACCATGCC <b>GCCGCGCGT</b> GCCAATGCCCAGCAC
	G370T	CACCATGCCCCGCCGT <b>GCT</b> GCCAATGCCCAGCAC
LCYB mutants	T371A	ATGCACCATGCCCGC <b>CGC</b> GCCGCCAATGCCCAG
	A372G	GGGATGCACCATGCC <b>GCC</b> CGTGCCGCCAATGCC
	T371A/A372G <sup>c</sup>	GGGATGCACCATGCC <b>GCC</b> CGTGCCGCCAATGCC
	H376K	GAACCCTGTTGAGGG <b>CTT</b> CACCATGCCCGCCGT
	P377A	CATGAACCCTGTTGA <b>GGC</b> ATGCACCATGCCCGC
	H376K/P377A <sup>3</sup>	CATGAACCCTGTTGA <b>GGC</b> ATGCACCATGCCCGC

Sequences overlapping those of vectors are underlined; restriction sites are shown in bold italics; initiation codon (ATG) artificially added is shown in bold; mutant sites are shown in red. The selection primer (SP) was designed according to the user protocol.

**Table S2.** Sequences used for sequence alignment and molecular evolution analyses.

Type	Species	Accession number
Streptophyta LCYB	<i>Nicotiana tabacum</i>	NP_001311716
	<i>Arabidopsis thaliana</i>	AAB53337
	<i>Zea mays</i>	AAO18661
	<i>Solanum lycopersicum</i>	ABR57232
	<i>Daucus carota</i>	NP_001316089
Chlorophyta LCYB	<i>Haematococcus lacustris</i>	KX424526 (this study)
	<i>Chlamydomonas reinhardtii</i>	AAX54906
	<i>Chromochloris zofingiensis</i>	CBH31263
	<i>Dunaliella salina</i>	ACA34344
	<i>Volvox carteri f. nagariensis</i>	EFJ41647
Streptophyta LCYE	<i>Arabidopsis thaliana</i>	AAB53336
	<i>Adonis aestivalis</i>	AAK07431
	<i>Zea mays</i>	NP_001146840
	<i>Lactuca sativa</i>	AAK07434
Chlorophyta LCYE	<i>Chlamydomonas reinhardtii</i>	XP_001696529
	<i>Volvox carteri f. nagariensis</i>	XP_002945935
	<i>Haematococcus lacustris</i>	AKT95178 (this study)
	<i>Chromochloris zofingiensis</i>	CCG06343
Bacterial $\beta$ -monocyclase	Unclassified <i>Flavobacterium</i>	BAC77673
	<i>Rhodococcus erythropolis</i>	AAR98749
	<i>Deinococcus radiodurans</i>	AAF10377
Cyanobacterial CrtL	<i>Acaryochloris marina</i>	WP_012166204
Eubacterial CrtY	<i>Pantoea ananatis</i>	ADD79327
Fungal CrtB	<i>Xanthophyllomyces dendrorhous</i>	AAY33923
Streptophyta CCS	<i>Citrus sinensis</i>	AAF18389
GSB CruA	<i>Chlorobium tepidum</i>	NP_661357.1

CCS: capsanthin-capsorubin synthase

**Table S3.** Product profiles of the *H. lacustris* cyclases and site-directed mutants expressed in *E. coli*.

Percentage of total carotenoids (n = 3)									
	lyc	δ-car	γ-car	ε-car	α-car	β-car	neur	α-zeacar	β-zeacar
EIB-E	67.89 ± 2	32.11 ± 3	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
EIB-B	1.08 ± 1	n.d.	n.d.	n.d.	n.d.	98.92 ± 4	n.d.	n.d.	n.d.
<b>Mutants of HILCYE</b>									
L457H	66.52 ± 0.5	n.d.	n.d.	33.48 ± 2	n.d.	n.d.	n.d.	n.d.	n.d.
A468D	87.7 ± 3	12.3 ± 1	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
V531T	99.83 ± 2	0.17 ± 0.1	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
<b>Mutants of HILCYB</b>									
G370T	2.22 ± 0.5	n.d.	n.d.	n.d.	n.d.	97.78 ± 4	n.d.	n.d.	n.d.
T371A	1.24 ± 0.5	n.d.	n.d.	n.d.	n.d.	98.76 ± 2	n.d.	n.d.	n.d.
A372G	1.27 ± 0.2	n.d.	n.d.	n.d.	n.d.	98.73 ± 2	n.d.	n.d.	n.d.
H376K	1.21 ± 0.3	n.d.	n.d.	n.d.	n.d.	98.79 ± 1	n.d.	n.d.	n.d.
P377A	1.19 ± 0.5	n.d.	n.d.	n.d.	n.d.	98.81 ± 3	n.d.	n.d.	n.d.
T371A/A372G	1.39 ± 1	n.d.	n.d.	n.d.	n.d.	98.61 ± 1	n.d.	n.d.	n.d.
H376K/P377A	2.87 ± 1	n.d.	n.d.	n.d.	n.d.	97.13 ± 1	n.d.	n.d.	n.d.
G370T/H376K/P377A	21.91 ± 1	n.d.	3.57 ± 1	n.d.	n.d.	74.52 ± 2	n.d.	n.d.	n.d.
GTAX <sub>3</sub> HP/TAGX <sub>3</sub> KA	53.57 ± 3	n.d.	14.94 ± 2	n.d.	n.d.	31.49 ± 2	n.d.	n.d.	n.d.

Data are expressed as dry cell weight (DCW) percentages of total carotenoids in the *E. coli* strains, and each value is the mean ± SD for three experiments (n = 3). Lyc: lycopene; δ-car: δ-carotene; γ-car: γ-carotene; ε-car: ε-carotene; α-car: α-carotene; β-car: β-carotene; neur: neurosporene; α-zeacar: α-zeacarotene; β-zeacar: β-zeacarotene; n.d.: not detected.

**Table S4.** Recombination analysis of cyclases

Recombinant	Breakpoint		Parent		Detection Methods								
	Begin	End	Minor	Major	RDP	GENECONV	Bootscan	Maxchi	Chimaera	SiScan	PhylPro	LARD	3Seq
ZmLCYB	1031	1365	AtLCYB	CrLCYB/VcLCYB	NS	0.0007797	NS	2.06E-02	1.65E-02	1.44E-09	NS	2.91E-02	NS

NS: No significant *P* value is recorded for this recombination event using this method.

**Table S5.** LRT of the models used in the PAML analysis

Model	np	Estimates of parameters	lnL	LRT pair	df	2ΔlnL	Positive selected sites
<b>Site models</b>							
M0	49	$\omega=0.07152$	- 33663.366682				
M3	53	$p_0=0.38470, p_1=0.44096,$ $(p_2=0.17434), \omega_0=0.01935,$ $\omega_1=0.09159, \omega_2=0.29940$	- 32903.382604	M0/M3	4	1519.9700**	
M1a	50	$p_0=0.86750, (p_1=0.13250),$ $\omega_0=0.07646, (\omega_1=1.00000)$	- 33424.440636				
M2a	52	$p_0=0.86750, p_1=0.01616,$ $(p_2=0.11635), \omega_0=0.07646,$ $(\omega_1=1.00000), \omega_2=1.00000$	- 33424.440636	M1a/M2a	2	0	
M7	50	$p=1.12383, q=9.31286$ $p_0=0.99027, p=1.15604,$ $q=9.95730, (p_1=0.00973),$ $\omega=1.00000$	- 32906.043654				
M8	52		- 32904.834280	M7/M8	2	2.4188	
<b>Branch models</b>							
Fr	95	See Figure 4	- 33550.427866	M0/Fr	46	225.8780**	
Ta	50	$\omega_0=0.07145, \omega_a=447.95515$	- 33663.294828	M0/Ta	1	0.1437	
Tb	50	$\omega_0=0.07145, \omega_b=3.41633$	- 33660.058438	M0/Tb	1	6.6165**	
Tc	50	$\omega_0=0.07152, \omega_c=287.71757$	- 33663.291103	M0/Tc	1	0.1512	
Td	50	$\omega_0=0.07161, \omega_d=999.00000$	- 33662.259393	M0/Td	1	2.2146	
Te	50	$\omega_0=0.07206, \omega_e=0.00184$	- 33662.761843	M0/Te	1	1.2097	
Tf	50	$\omega_0=0.07700, \omega_f=0.00587$	- 33653.210680	M0/Tf	1	20.3120**	
Tg	50	$\omega_0=0.07216, \omega_g=999.00000$	- 33660.221006	M0/Tg	1	6.2914**	
Th	50	$\omega_0=0.07112, \omega_h=999.00000$	- 33660.722873	M0/Th	1	5.2876*	
Ti	50	$\omega_0=0.06958, \omega_i=999.00000$	- 33657.758291	M0/Ti	1	11.2168**	
<b>Branch-site models</b>							
A1b	51	$p_0=0.68674, p_1=0.10482,$ $(p_{2a}=0.18083, p_{2b}=0.02760),$ $\omega_0=0.07585, (\omega_1=1.00000), b:$ $\omega_{2a}=0.07585, \omega_{2b}=1.00000, f:$ $\omega_{2a}=1.00000, \omega_{2b}=1.00000$	- 33418.029102				

		<p>p<sub>0</sub>=0.66662, p<sub>1</sub>=0.10212, (p<sub>2a</sub>=0.20054, p<sub>2b</sub>=0.03072), ω<sub>0</sub>=0.07672, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07672, ω<sub>2b</sub>=1.00000, f: <b>ω<sub>2a</sub>=921.69928, ω<sub>2b</sub>=921.69928</b></p> <p>p<sub>0</sub>=0.82918, p<sub>1</sub>=0.12680, (p<sub>2a</sub>=0.03818, p<sub>2b</sub>=0.00584), ω<sub>0</sub>=0.07576, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07576, ω<sub>2b</sub>=1.00000, f: ω<sub>2a</sub>=1.00000, ω<sub>2b</sub>=1.00000</p>	-	33412.910439	A1b/Ab	1	10.2373**	HILCYE S89*, Y112*, K119*, G128*, D130*, P132*, V134**, K145*, H151*, S180*, V209**, <b>A210*</b> , T216*, A228*, T232**, L233*, A235*, A237*, P250*, S254*, E261*, F271**, T286*, K339*, H366*, Y372*, V375*, G382*, D383**, A391*, S403*, <b>L457*</b> , A459*, S465*, <b>A468*</b> , A498*	HILCYB <b>T86*</b> , A109*, A116*, D125*, E127*, I131*, P133**, E144*, E191*, D181*, V208**, N209*, S215*, G227*, L231**, D232*, T234*, H236*, F247*, Y251*, V258*, F266**, T281*, R318*, E345*, L351*, M354*, H361*, P362**, <b>G370*</b> , M382*, L451*, L453*, Q459*, D462*, G492*	
Ab	52									
A1f	51		-	33422.509336						
		<p>p<sub>0</sub>=0.83668, p<sub>1</sub>=0.12892, (p<sub>2a</sub>=0.02981, p<sub>2b</sub>=0.00459), ω<sub>0</sub>=0.07596, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07596, ω<sub>2b</sub>=1.00000, f: <b>ω<sub>2a</sub>=21.31733, ω<sub>2b</sub>=21.31733</b></p> <p>p<sub>0</sub>=0.53134, p<sub>1</sub>=0.08186, (p<sub>2a</sub>=0.33516, p<sub>2b</sub>=0.05163), ω<sub>0</sub>=0.07470, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07470, ω<sub>2b</sub>=1.00000, f: ω<sub>2a</sub>=1.00000, ω<sub>2b</sub>=1.00000</p>	-	33417.738928	A1f/Af	1	9.5408*	AtLCYB K65**, N67*, N494*		
Af	52									
A1g	51		-	33416.077226						
		<p>p<sub>0</sub>=0.81800, p<sub>1</sub>=0.12905, (p<sub>2a</sub>=0.04574, p<sub>2b</sub>=0.00722), ω<sub>0</sub>=0.07527, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07527, ω<sub>2b</sub>=1.00000, f: <b>ω<sub>2a</sub>=999.00000, ω<sub>2b</sub>=999.00000</b></p> <p>p<sub>0</sub>=0.74005, p<sub>1</sub>=0.11651, (p<sub>2a</sub>=0.12393, p<sub>2b</sub>=0.01951), ω<sub>0</sub>=0.07480, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07480, ω<sub>2b</sub>=1.00000, f: ω<sub>2a</sub>=1.00000, ω<sub>2b</sub>=1.00000</p>	-	33406.684493	A1g/Ag	1	18.7855**	HILCYB <b>T86**</b> , K242*, Q245*, N445*	DsLCYB A112**, K268*, Q271*, M472*	
Ag	52									
A1h	51		-	33410.003873						
		<p>p<sub>0</sub>=0.79428, p<sub>1</sub>=0.13049, (p<sub>2a</sub>=0.06461, p<sub>2b</sub>=0.01062), ω<sub>0</sub>=0.07616, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07616, ω<sub>2b</sub>=1.00000, f: <b>ω<sub>2a</sub>=999.00000, ω<sub>2b</sub>=999.00000</b></p> <p>p<sub>0</sub>=0.57455, p<sub>1</sub>=0.08544, (p<sub>2a</sub>=0.29600, p<sub>2b</sub>=0.04401), ω<sub>0</sub>=0.07413, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07413, ω<sub>2b</sub>=1.00000, f: ω<sub>2a</sub>=1.00000, ω<sub>2b</sub>=1.00000</p>	-	33388.535879	A1h/Ah	1	42.9360*	HILCYE P93*, V102**, V103*, V125**, L126**, A198*, L200**, <b>A210*</b> , L226**, V231**, S251*, V315**, R361**, <b>A468**</b> , Q488*, L500*, T501**, N507**, V531**	CrLCYB A106*, T115**, I116*, N138**, V139**, R211*, S213**, V223*, Y239**, T244**, I264*, C328**, S374**, D481**, T501*, V513*, V514**, W520**, Q544**	VcLCYB A99*, T108**, I109*, N131**, V132**, R204*, S206**, V216*, Y232**, T237**, I257*, C321**, S467**, D464**, T494*, V506*, V507**, L513**, Q537**
Ah	52									
A1i	51		-	33419.336472						
		<p>p<sub>0</sub>=0.80198, p<sub>1</sub>=0.11853, (p<sub>2a</sub>=0.06926, p<sub>2b</sub>=0.01024), ω<sub>0</sub>=0.07368, (ω<sub>1</sub>=1.00000), b: ω<sub>2a</sub>=0.07368, ω<sub>2b</sub>=1.00000, f: <b>ω<sub>2a</sub>=999.00000, ω<sub>2b</sub>=999.00000</b></p>	-	33417.040223	A1i/Ai	1	4.5925*	AtLCYB L351*	ZmLCYB L358*	LsLCYB T355*
Ai	52									

Selection analysis by three kinds of models is performed using Codeml implemented in PAML. *np*: number of free parameters. *lnL*: log



likelihood. LRT: likelihood ratio test. *df*: degrees of freedom.  $2\Delta\ln L$ : twice the log-likelihood difference of the models compared. The significant tests at 5% cutoff are labeled with \* and at 1% cutoff are labeled with \*\*. Sites of HILCYE detected in both branches B and H are showed in bold. Sites of HILCYB detected in both branches B and G are also showed in bold. G<sub>370</sub> of HILCYB and L<sub>457</sub> of HILCYE detected in branch B is highlighted by italic bold.