US 20250115926A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2025/0115926 A1**

Maeda et al. (43) **Pub. Date: Apr. 10, 2025**

(54) **DEREGULATED DHS AND TYRA ENZYMES OF GRASSES ENABLE EFFICIENT PRODUCTION OF BOTH TYROSINE AND PHENYLALANINE**

(71) Applicant: **Wisconsin Alumni Research Foundation**, Madison, WI (US)

(72) Inventors: **Hiroshi Maeda**, Madison, WI (US); **Jorge El Azaz**, Madison, WI (US)

(21) Appl. No.: **18/906,694**

(22) Filed: **Oct. 4, 2024**

Related U.S. Application Data

(60) Provisional application No. 63/588,272, filed on Oct. 5, 2023.

Publication Classification

(51) **Int. Cl.**
C12N 15/82 (2006.01)

(52) **U.S. Cl.**
CPC ***C12N 15/8251*** (2013.01)

(57) **ABSTRACT**

The present invention provides engineered cells and plants that express deregulated aromatic amino acid synthesis pathway enzymes from grasses. Methods for increasing the production of aromatic amino acids and their derivatives in cells and plants by engineering them to express these enzymes and methods for producing aromatic amino acids or derivatives thereof and/or sequestering carbon dioxide by growing the plants are also provided.

Specification includes a Sequence Listing.

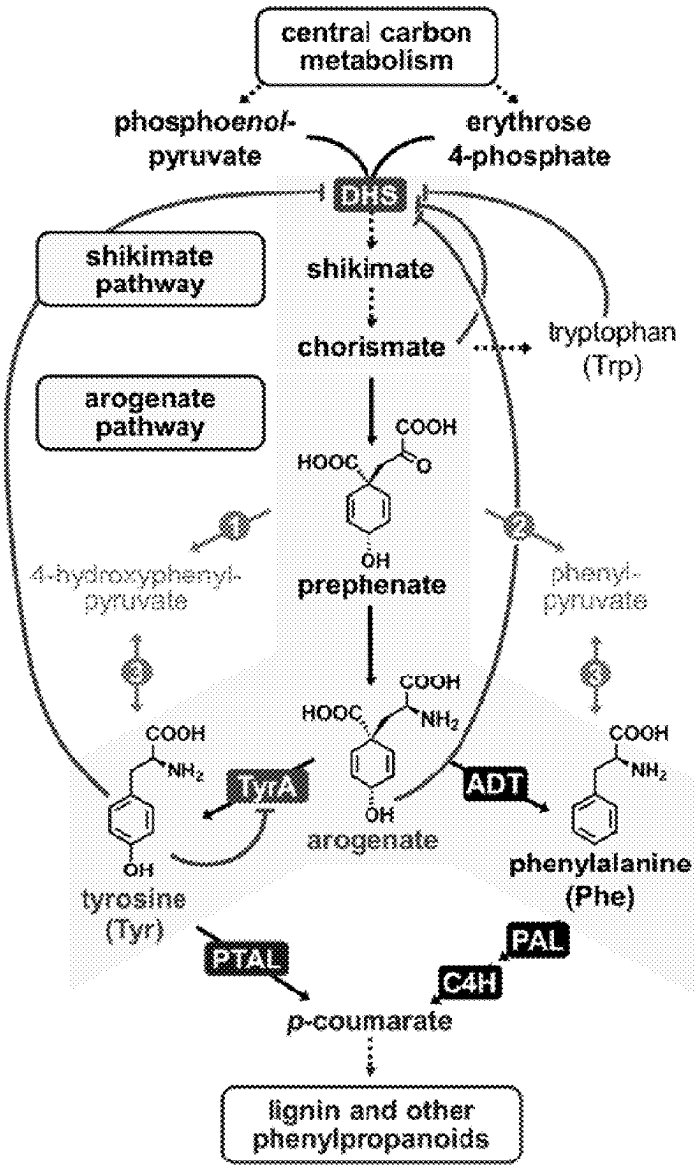


FIG. 1A

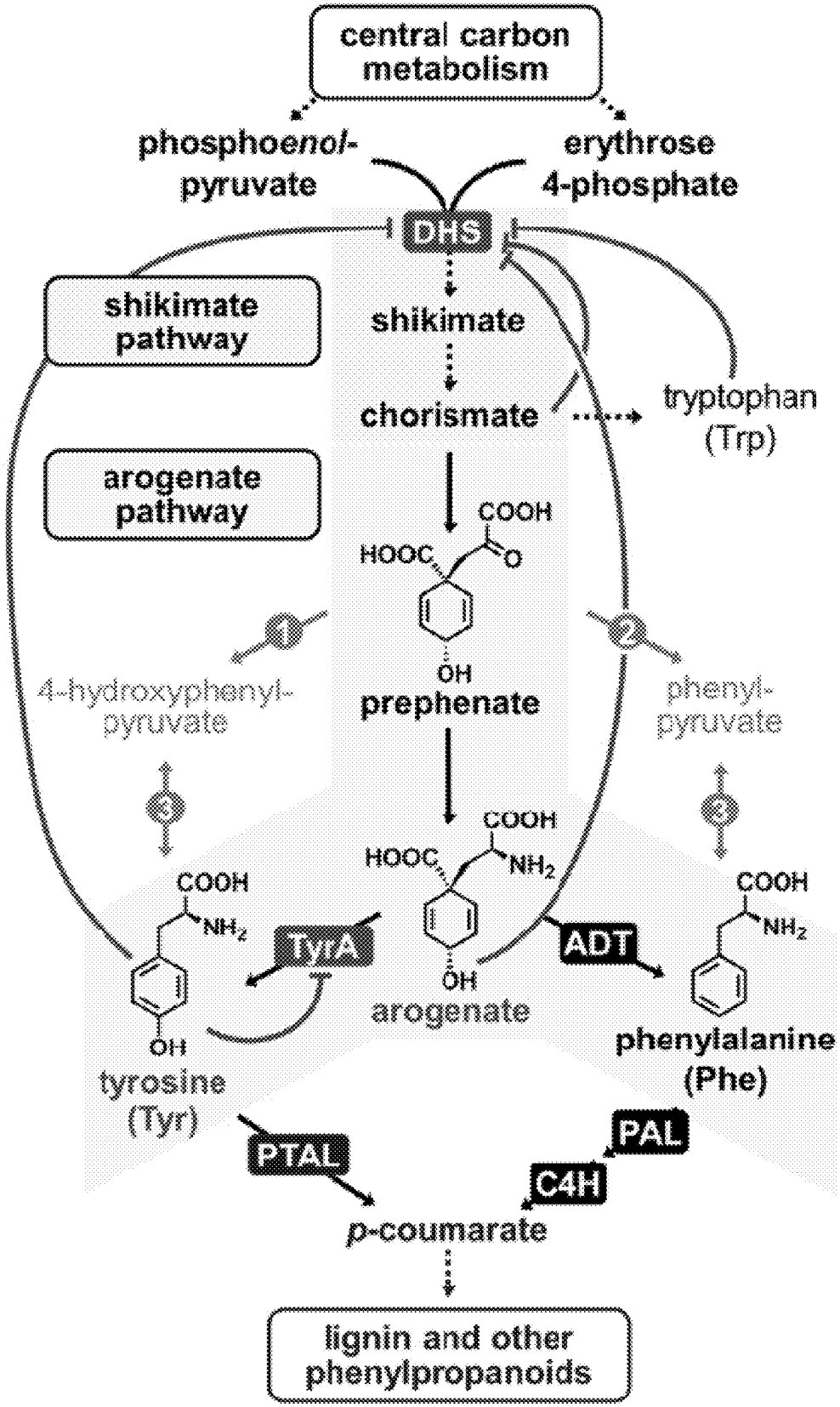
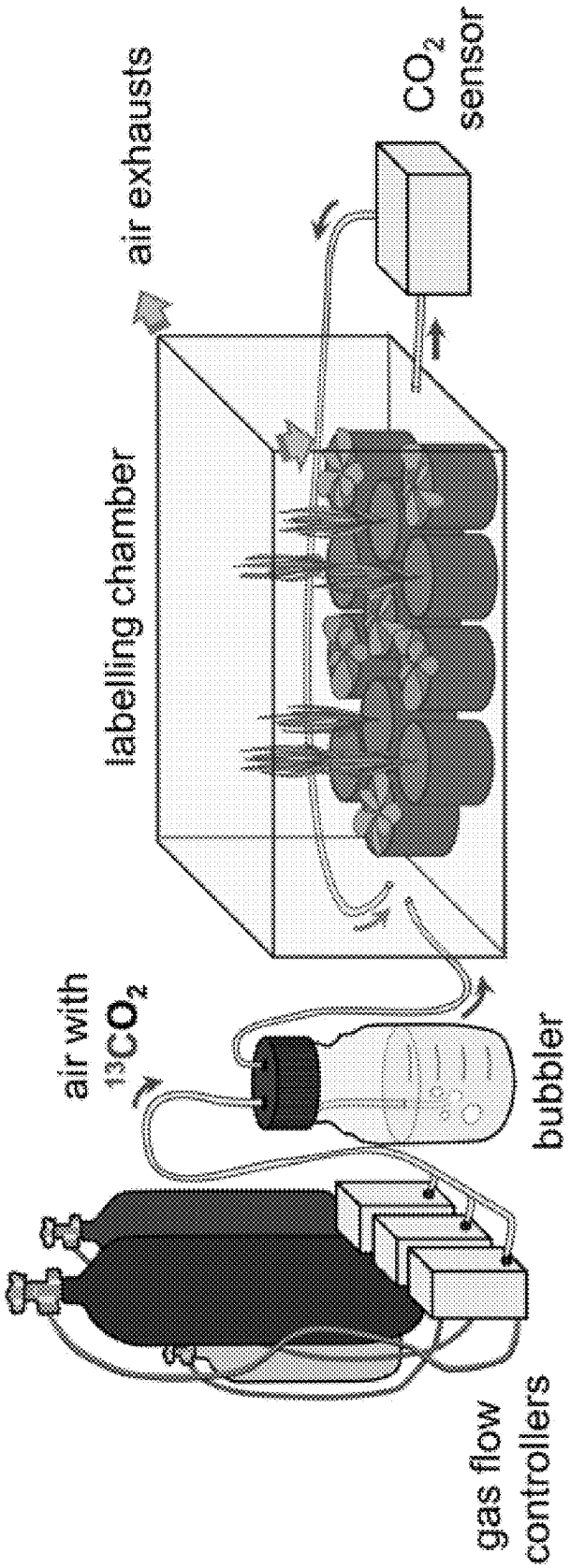

central carbon metabolism
phosphoenol-pyruvate
erythrose 4-phosphate
DHS
shikimate pathway
shikimate
chorismate
tryptophan (Trp)
arogenate pathway
COOH
HOOC
O
OH
prephenate
4-hydroxyphenyl-pyruvate
phenyl-pyruvate
COOH
NH2
HOOC
TyrA
ADT
OH
arogenate
tyrosine (Tyr)
phenylalanine (Phe)
PTAL
PAL
C4H
p-coumarate
lignin and other phenylpropanoids

FIG. 1B


air exhausts
CO_2 sensor
labelling chamber
air with $^{13}CO_2$
bubbler
gas flow controllers

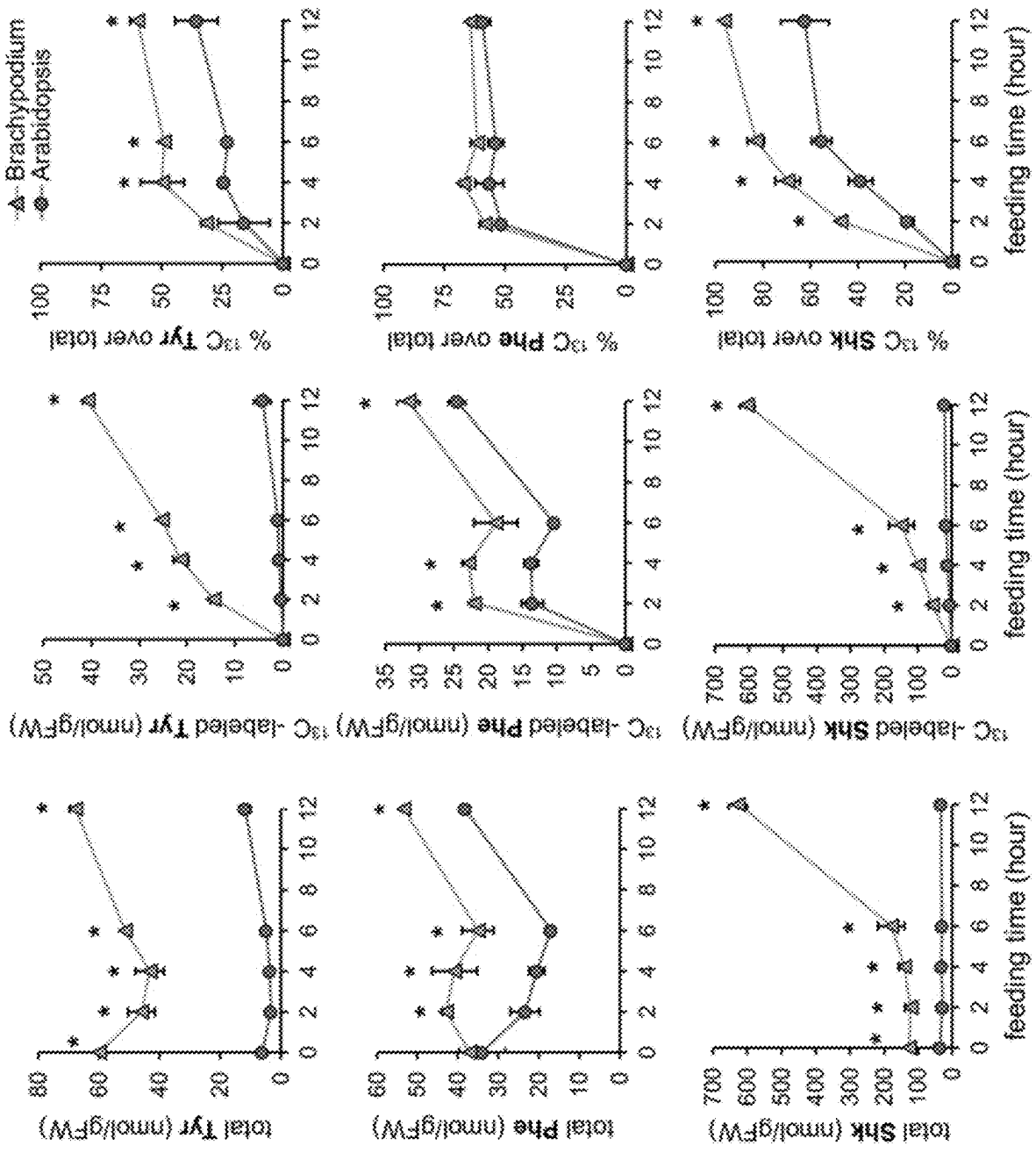


FIG. 1C

FIG. 1D

FIG. 1E

FIG. 2A

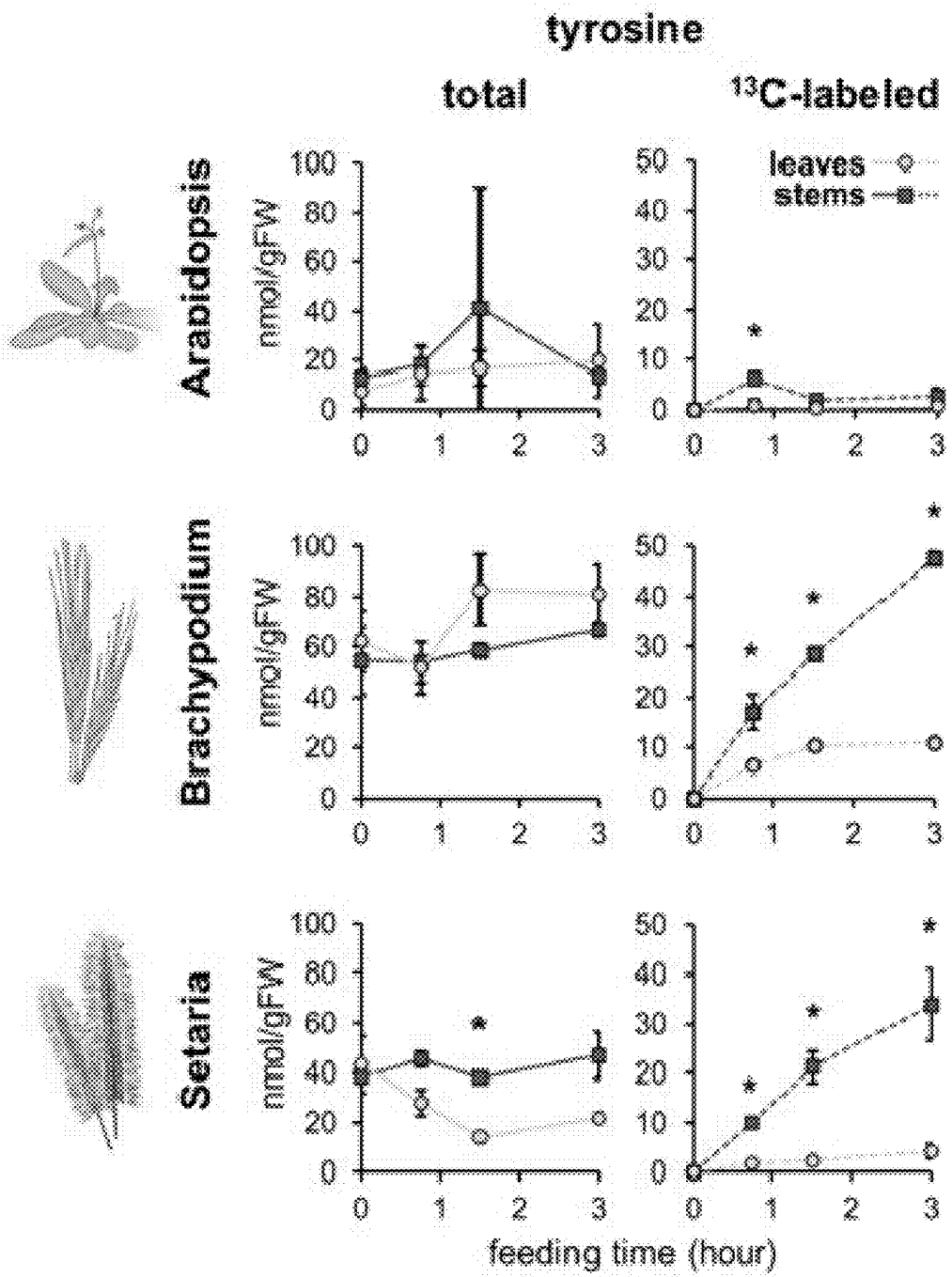

tyrosine
total
13C-labeled
leaves
stems
Arapidopsis
Brachypodium
Setaria
nmol/gFW
feeding time (hour)

FIG. 2B

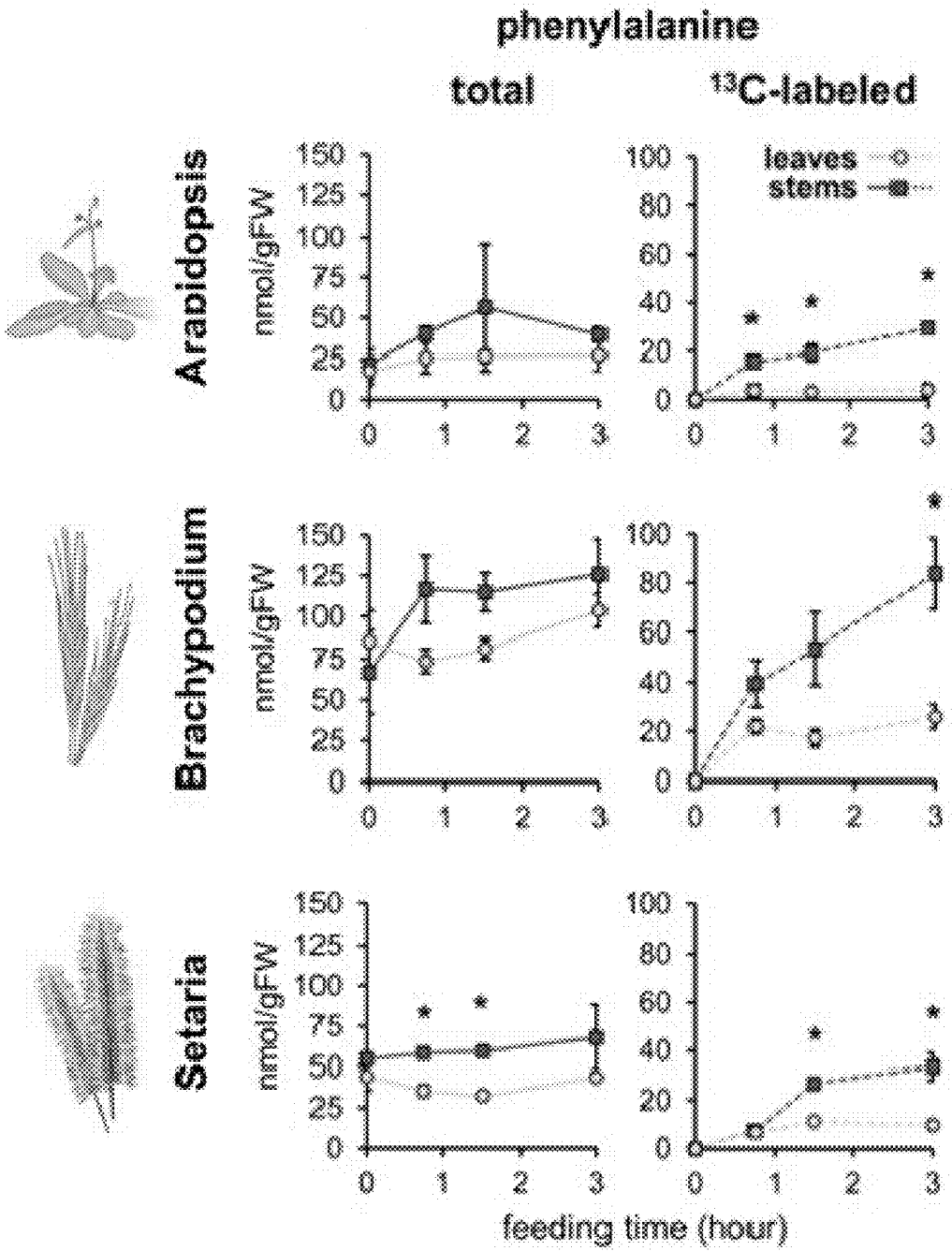

phenylalanine
total
13C-labeled
Arabidopsis
Brachypodium
Setaria
nmol/gFW
leaves
stems
feeding time (hour)

FIG. 2C

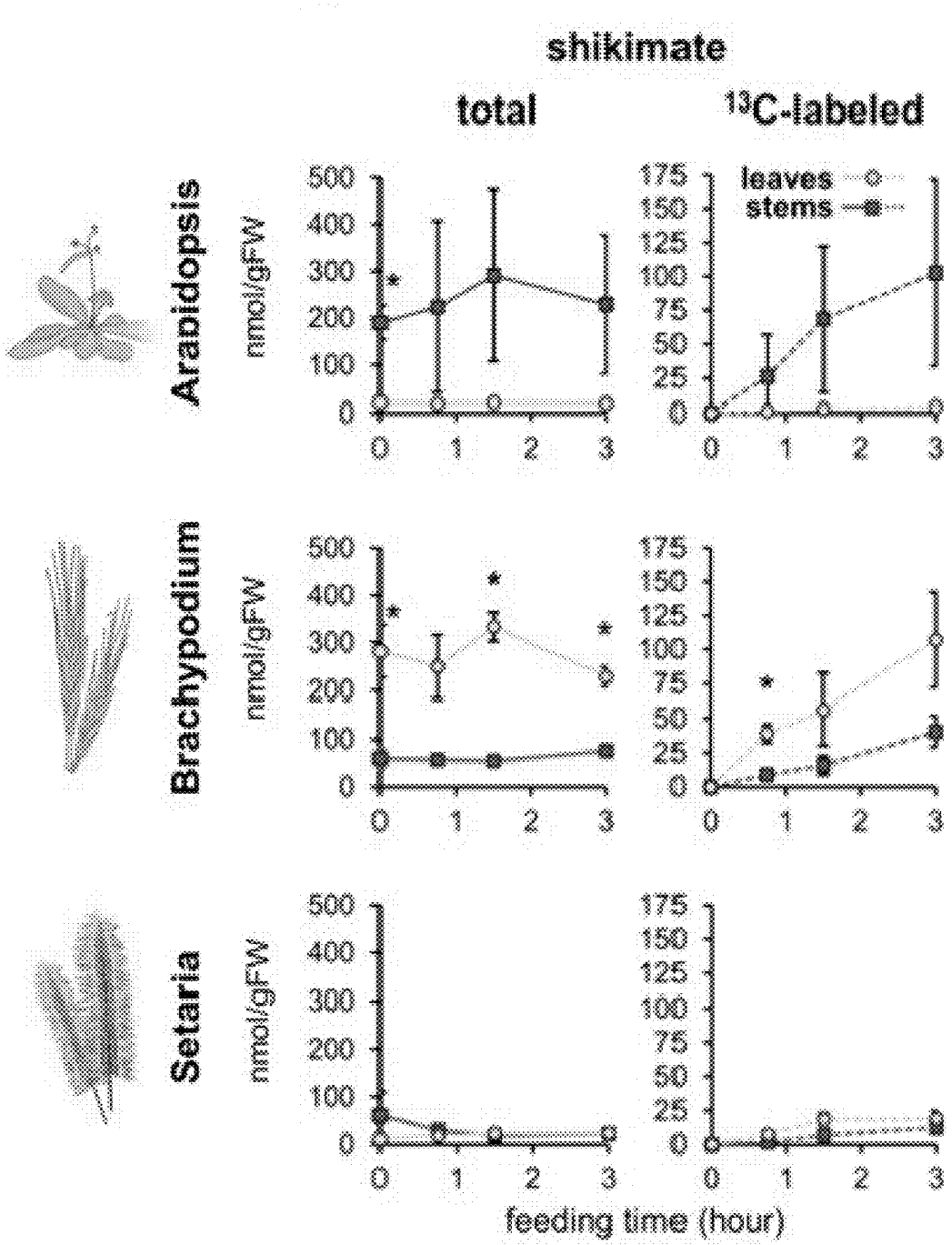

shikimate
total
13C-labeled
leaves
stems
Arabidopsis
Brachypodium
Setaria
nmol/gFW
feeding time (hour)

FIG. 3A

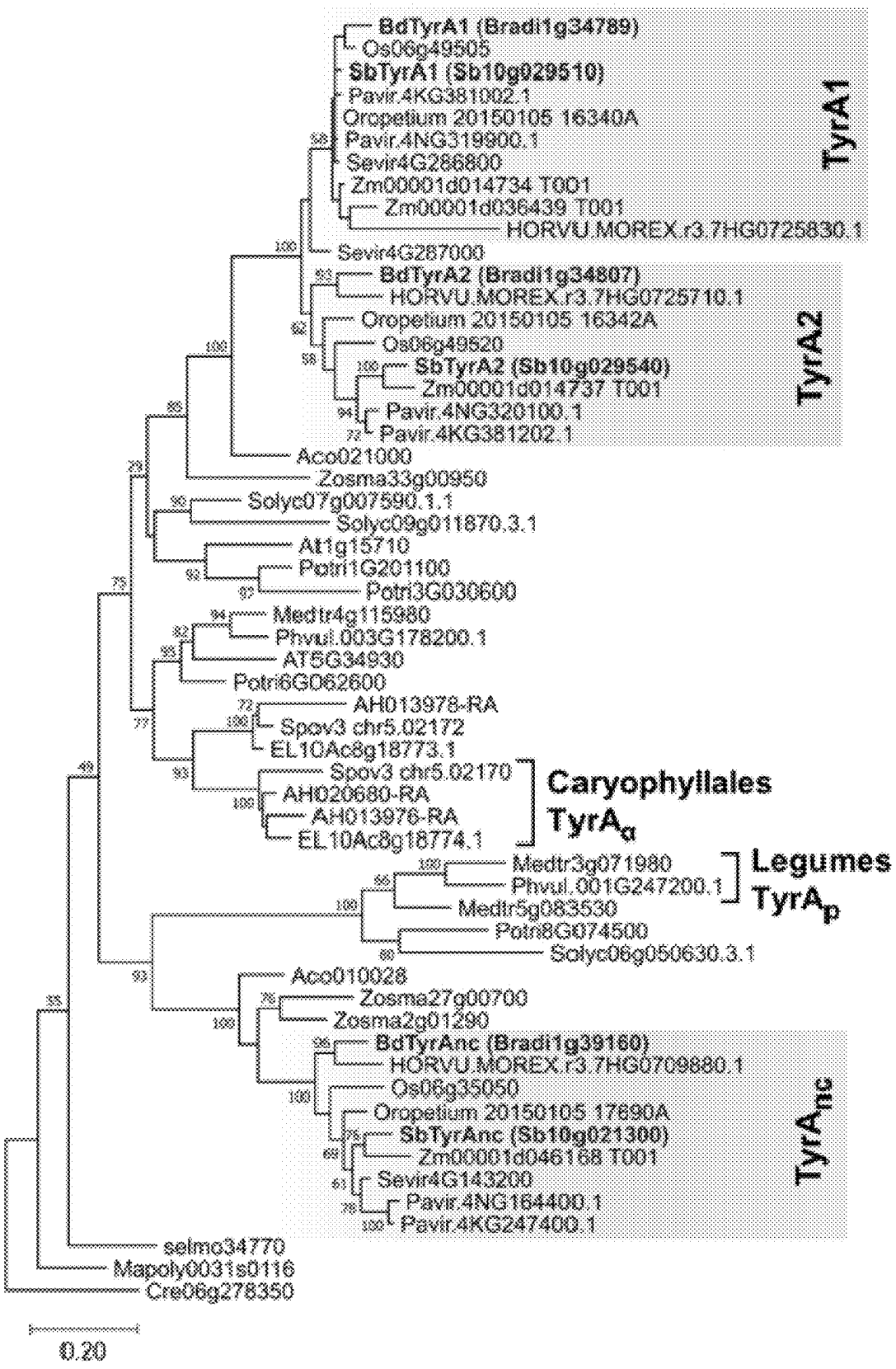

BdTyrA1 (Bradi1g34789)
Os06g49505
SbTyrA1 (Sb10g029510)
Pavir.4KG381002.1
Oropetium 20150105 16340A
Pavir.4NG319800.1
Sevir4G286800
Zm00001d014734 T001
Zm00001d036439 T001
HORVU.MOREX.r3.7HG0725830.1
TyrA1
Sevir4G287000
BdTyrA2 (Bradi1g34807)
HORVU.MOREX.r3.7HG0725710.1
Oropetium 20150105 16342A
Os06g49520
SbTyrA2 (Sb10g029540)
Zm00001d014737 T001
Pavir.4NG320100.1
Pavir.4KG381202.1
TyrA2
Aco021000
Zosma33g00950
Solyc07g007590.1.1
Solyc09g011870.3.1
At1g15710
Potri1G201100
Potri3G030600
Medtr4g115980
Phvul.003G178200.1
AT5G34930
Potri6G062600
AH013978-RA
Spov3 chr5.02172
EL10Ac8g18773.1
Spov3 chr5.02170
AH020680-RA
AH013976-RA
EL10Ac8g18774.1
Caryophyllales
TyrAα
Medtr3g071980
Phvul.001G247200.1
Medtr5g083530
Potri8G074500
Solyc06g050630.3.1
Legumes
TyrAp
Aco010028
Zosma27g00700
Zosma2g01290
BdTyrAnc (Bradi1g39160)
HORVU.MOREX.r3.7HG0709880.1
Os06g35050
Oropetium 20150105 17690A
SbTyrAnc (Sb10g021300)
Zm00001d046168 T001
Sevir4G143200
Pavir.4NG164400.1
Pavir.4KG247400.1
TyrAnc
selmo34770
Mapoly0031s0116
Cre06g278350
0.20

FIG. 3B

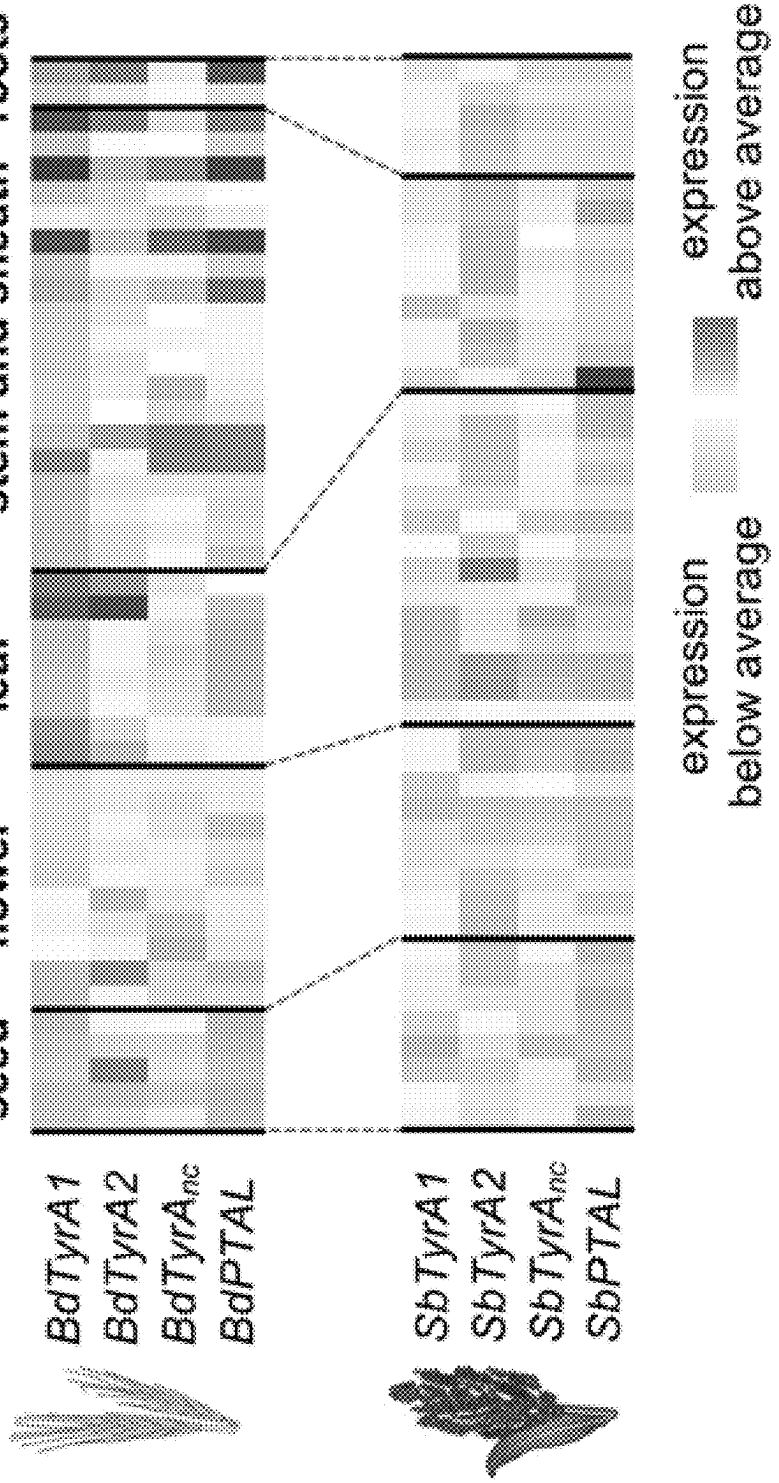

seed
flower
leaf
stem and sheath
roots
BdTyrA1
BdTyrA2
BdTyrA$_{nc}$
BdPTAL
SbTyrA1
SbTyrA2
SbTyrA$_{nc}$
SbPTAL
expression below average
expression above average

FIG. 3C

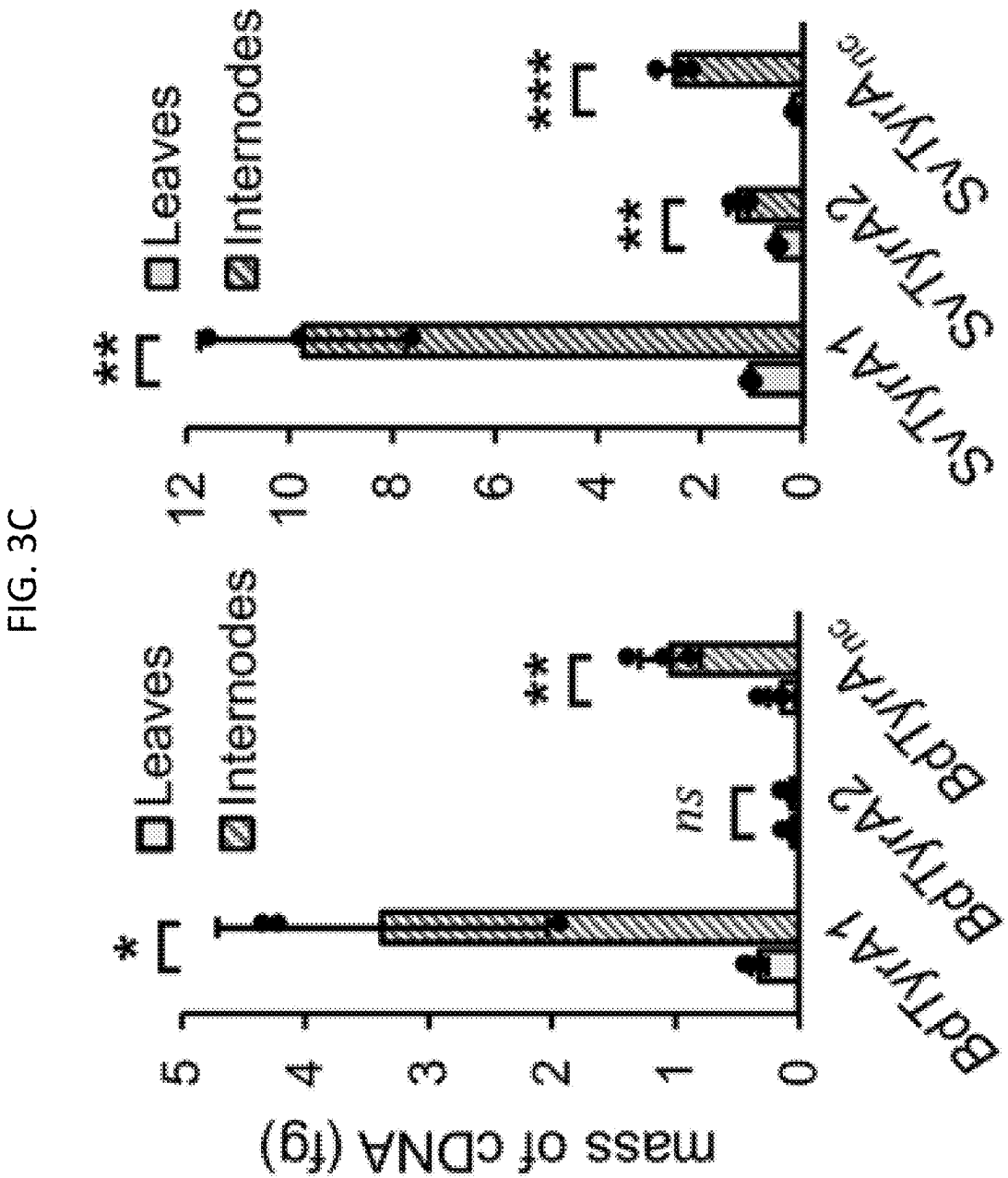

mass of cDNA (fg)
0
1
2
3
4
5
Leaves
Internodes
*
**
ns
BdTyrA1
BdTyrA2
BdTyrA$_{mc}$
0
2
4
6
8
10
12
Leaves
Internodes
**
***
SvTyrA1
SvTyrA2
SvTyrA$_{mc}$

FIG. 3D

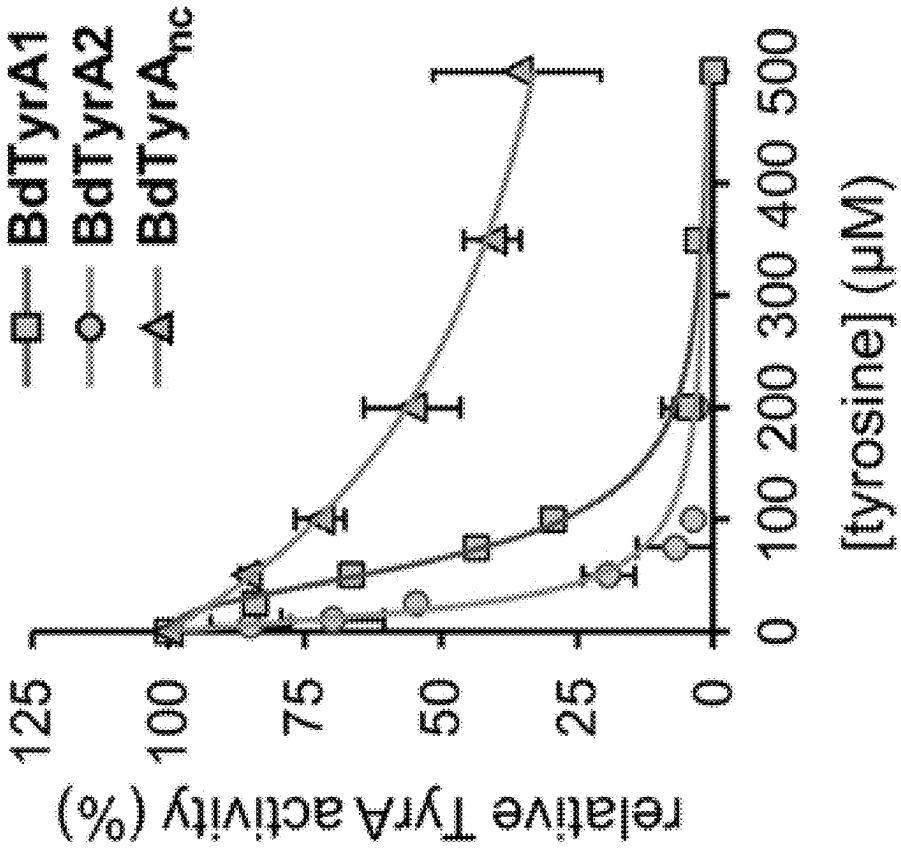

BdTyrA1
BdTyrA2
BdTyrA$_{nc}$
relative TyrA activity (%)
125
100
75
50
25
0
0 100 200 300 400 500
[tyrosine] (μM)

FIG. 3E

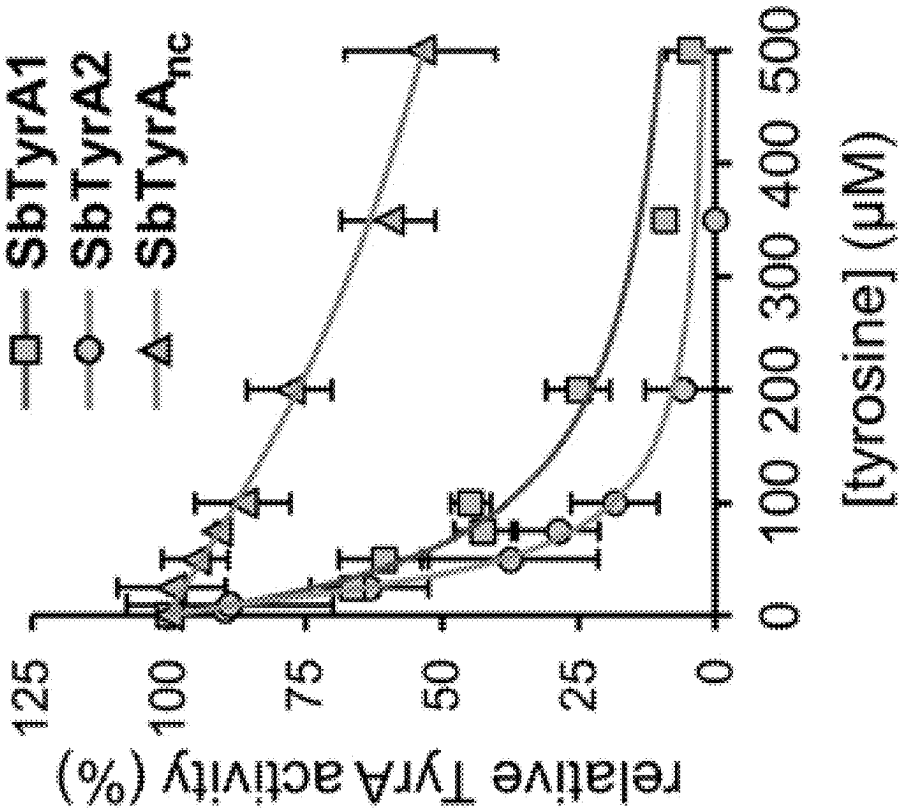

SbTyrA1
SbTyrA2
SbTyrA$_{nc}$
relative TyrA activity (%)
125
100
75
50
25
0
0 100 200 300 400 500
[tyrosine] (μM)

FIG. 3F

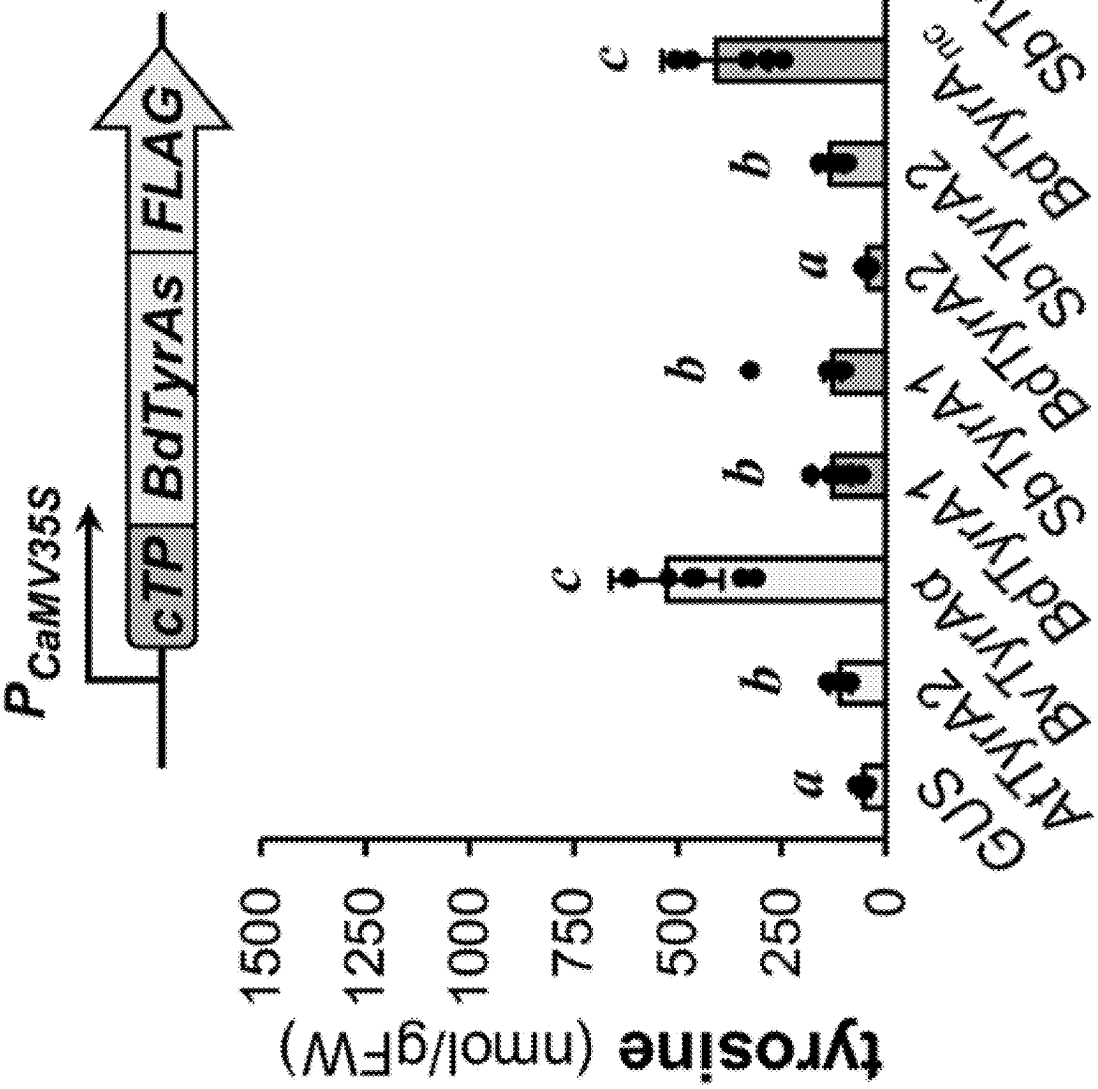

$P_{CaMV35S}$
cTP
BdTyrAs
FLAG
tyrosine (nmol/gFW)
1500
1250
1000
750
500
250
0
a
b
b
c
b
b
a
c
c
GUS
AtTyrA2
BvTyrAd
BdTyrA1
SbTyrA1
BdTyrA2
SbTyrA2
BdTyrA2mc
SbTyrA2mc

FIG. 4A

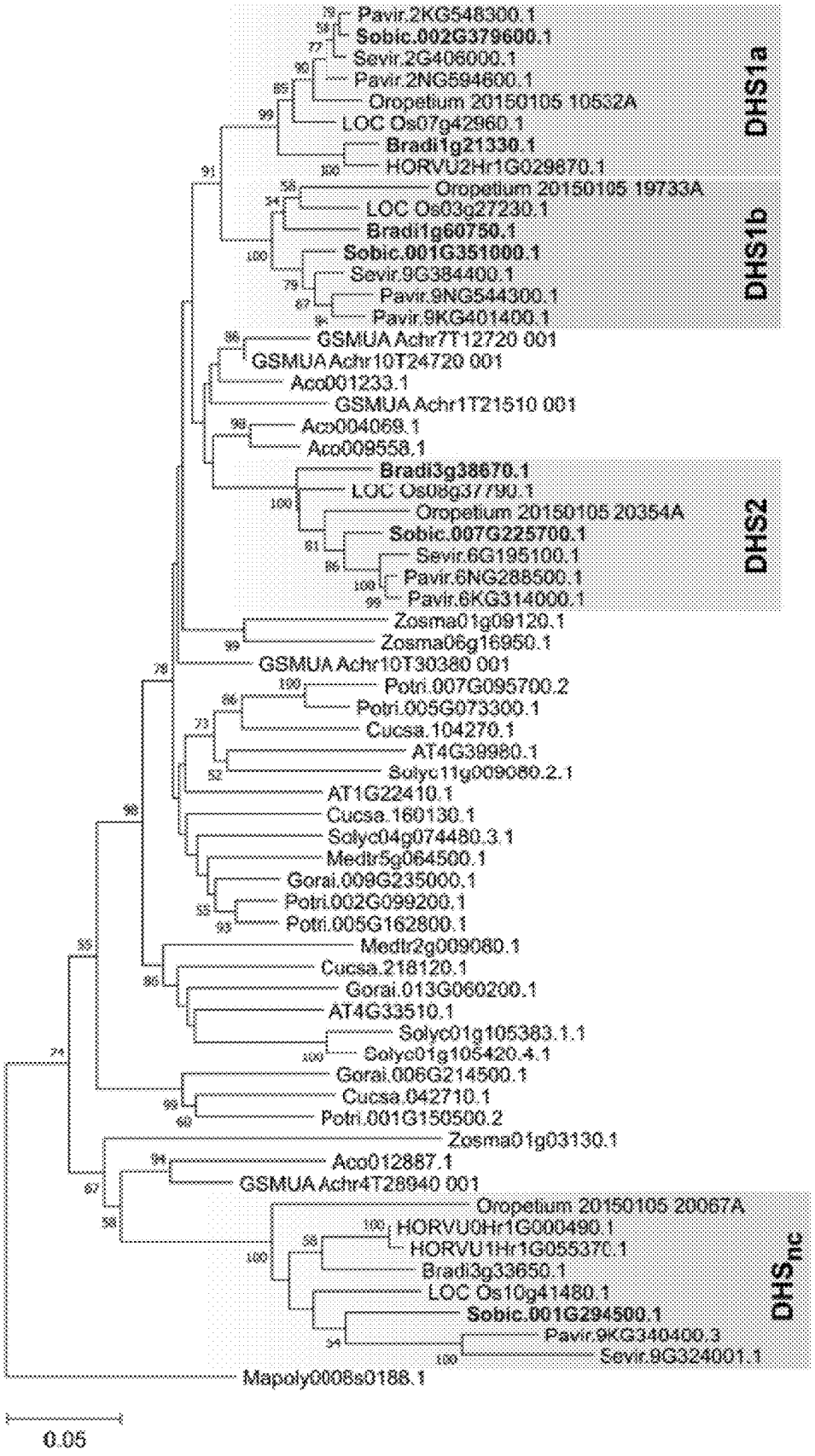

Pavir.2KG546300.1
Sobic.002G379600.1
Sevir.2G406000.1
Pavir.2NG594600.1
Oropetium 20150105 10532A
LOC Os07g42960.1
Bradi1g21330.1
HORVU2Hr1G029870.1
DHS1a
Oropetium 20150105 19733A
LOC Os03g27230.1
Bradi1g60750.1
Sobic.001G351000.1
Sevir.9G384400.1
Pavir.9NG544300.1
Pavir.9KG401400.1
DHS1b
GSMUA Achr7T12720 001
GSMUA Achr10T24720 001
Aco001233.1
GSMUA Achr1T21510 001
Aco004069.1
Aco009558.1
Bradi3g38670.1
LOC Os08g37790.1
Oropetium 20150105 20354A
Sobic.007G225700.1
Sevir.6G195100.1
Pavir.6NG288500.1
Pavir.6KG314000.1
DHS2
Zosma01g09120.1
Zosma06g16950.1
GSMUA Achr10T30380 001
Potri.007G095700.2
Potri.005G073300.1
Cucsa.104270.1
AT4G39980.1
Solyc11g009080.2.1
AT1G22410.1
Cucsa.160130.1
Solyc04g074480.3.1
Medtr5g064500.1
Gorai.009G235000.1
Potri.002G099200.1
Potri.005G162800.1
Medtr2g009080.1
Cucsa.218120.1
Gorai.013G060200.1
AT4G33510.1
Solyc01g105383.1.1
Solyc01g105420.4.1
Gorai.006G214500.1
Cucsa.042710.1
Potri.001G150500.2
Zosma01g03130.1
Aco012887.1
GSMUA Achr4T28940 001
Oropetium 20150105 20067A
HORVU0Hr1G000490.1
HORVU1Hr1G055370.1
Bradi3g33650.1
LOC Os10g41480.1
Sobic.001G294500.1
Pavir.9KG340400.3
Sevir.9G324001.1
DHSnc
Mapoly0008s0188.1
0.05

FIG. 4B

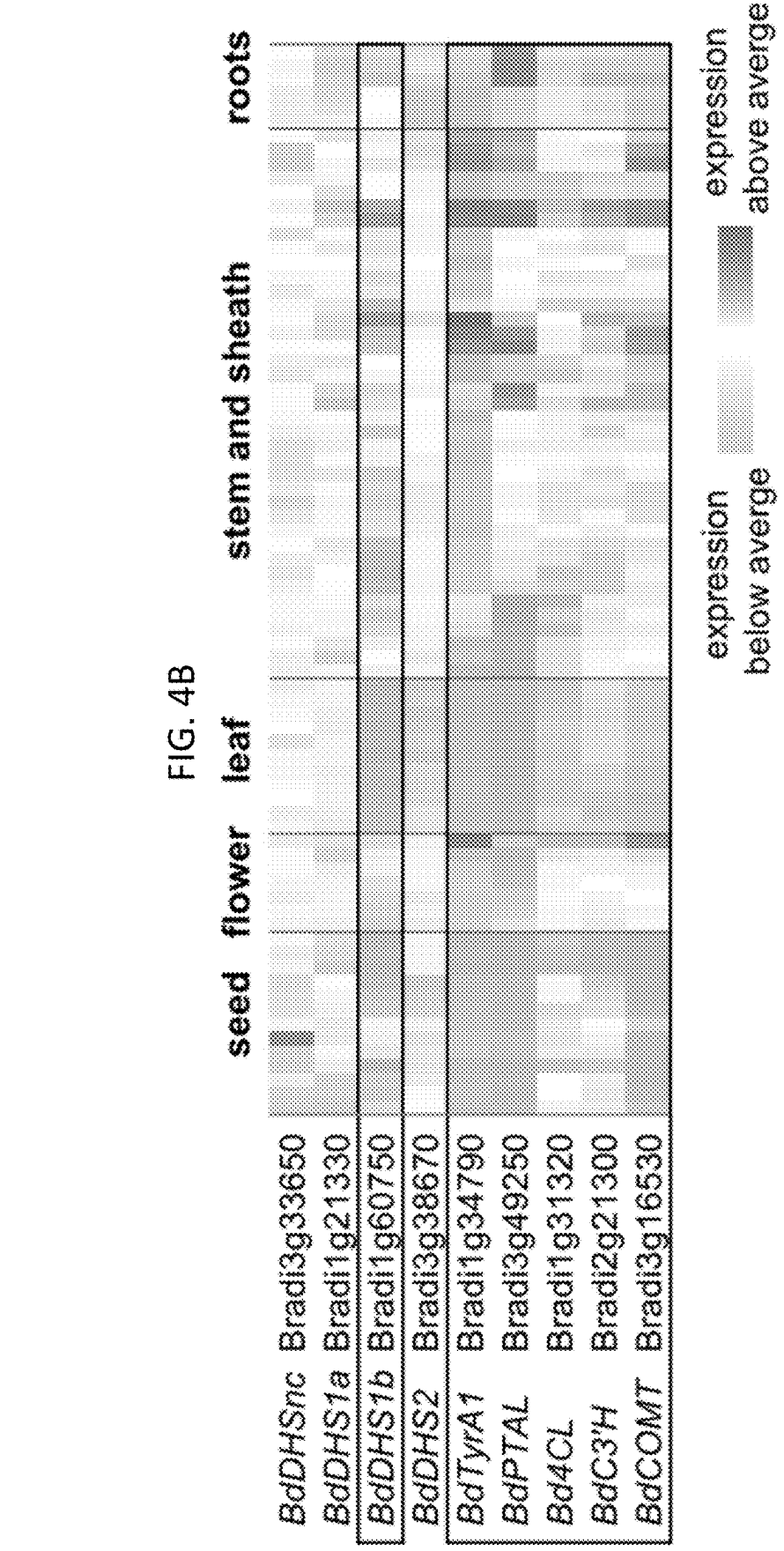

seed
flower
leaf
stem and sheath
roots
BdDHSnc Bradi3g33650
BdDHS1a Bradi1g21330
BdDHS1b Bradi1g60750
BdDHS2 Bradi3g38670
BdTyrA1 Bradi1g34790
BdPTAL Bradi3g49250
Bd4CL Bradi1g31320
BdC3'H Bradi2g21300
BdCOMT Bradi3g16530
expression below averge
expression above averge

FIG. 4C

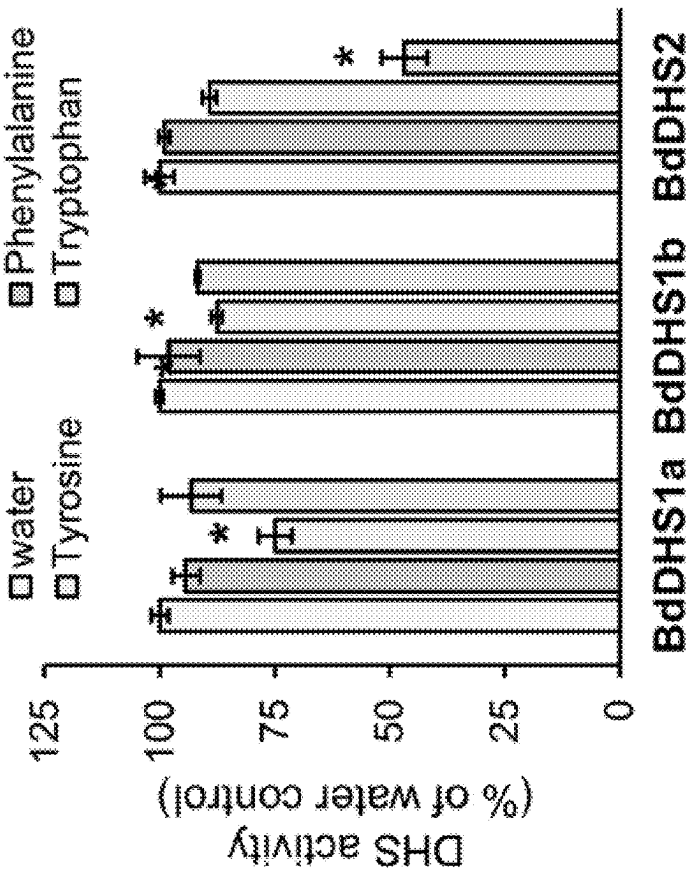

DHS activity
(% of water control)
125
100
75
50
25
0
water
Tyrosine
Phenylalanine
Tryptophan
*
*
*
BdDHS1a
BdDHS1b
BdDHS2

FIG. 4D

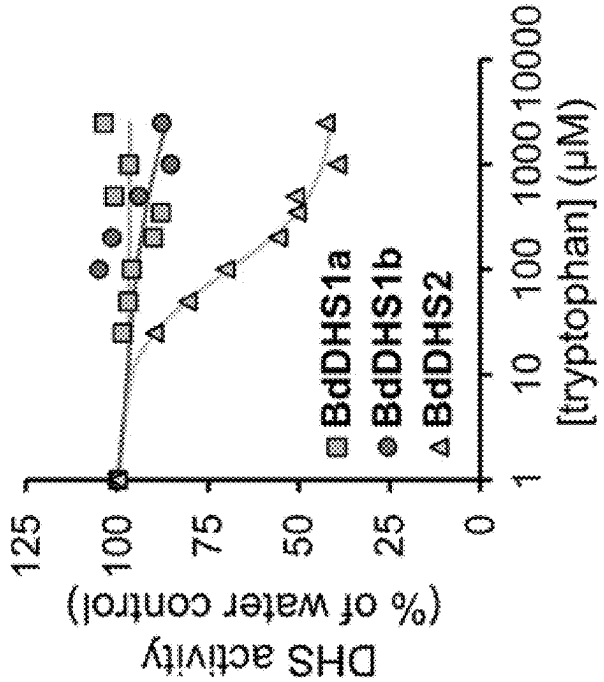

DHS activity
(% of water control)
125
100
75
50
25
0
1
10
100
1000
10000
[tryptophan] (μM)
BdDHS1a
BdDHS1b
BdDHS2

FIG. 4E

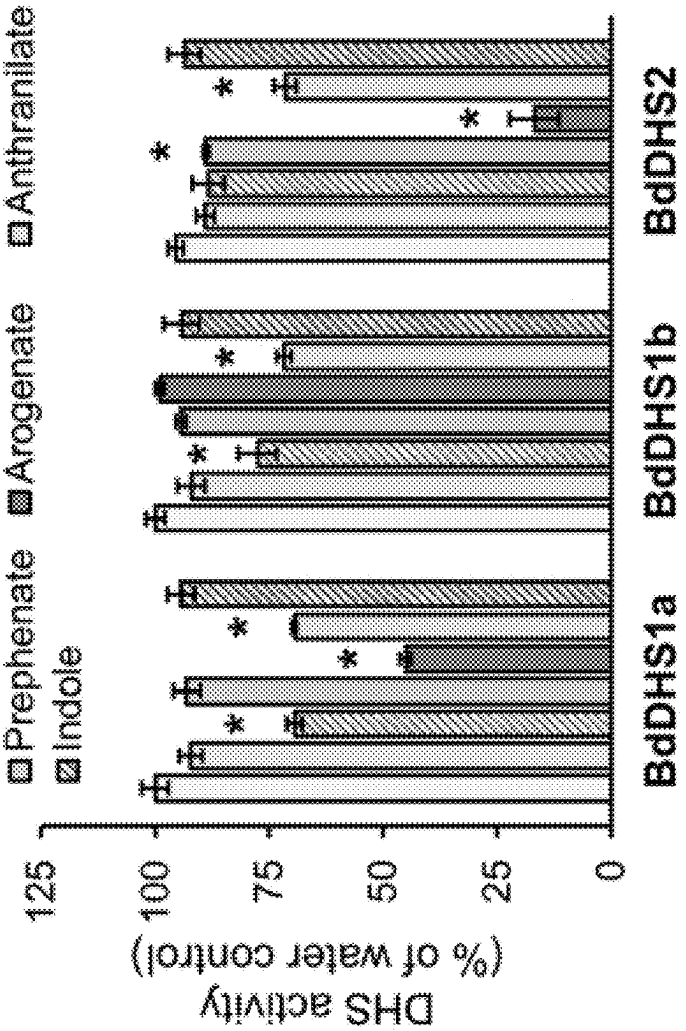

water
Prephenate
Indole
Shikimate
Arogenate
Chorismate
Anthranilate
DHS activity
(% of water control)
125
100
75
50
25
0
BdDHS1a
BdDHS1b
BdDHS2

FIG. 4F

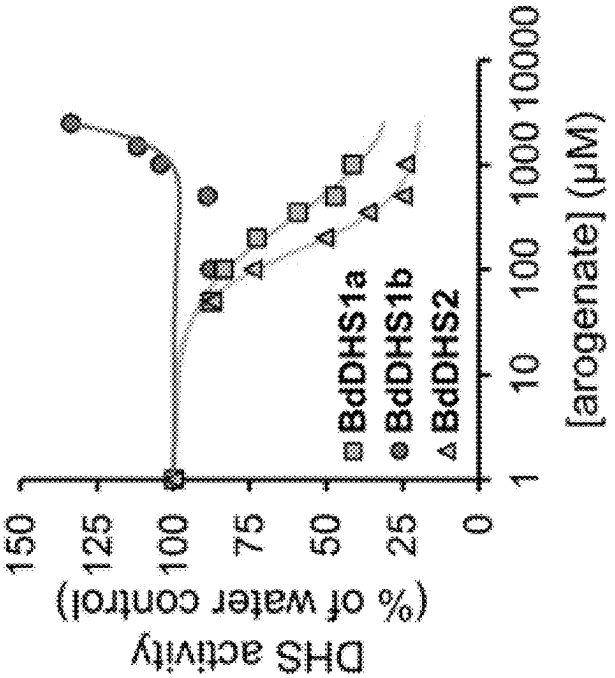

DHS activity
(% of water control)
150
125
100
75
50
25
0
1
10
100
1000
10000
[arogenate] (μM)
BdDHS1a
BdDHS1b
BdDHS2

FIG. 5A

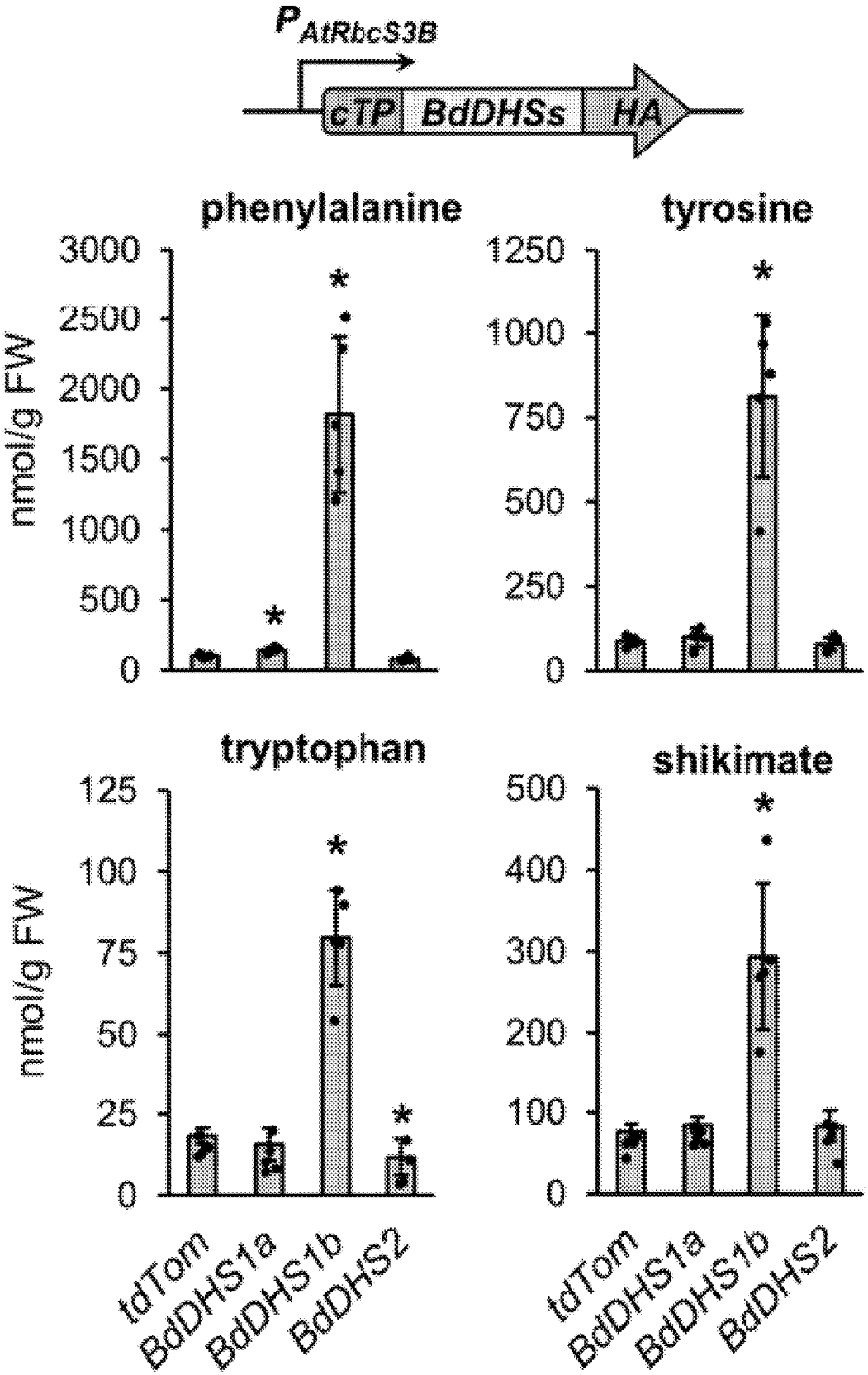

P AtRbcS3B
cTP
BdDHSs
HA
phenylalanine
tyrosine
tryptophan
shikimate
nmol/g FW
3000
2500
2000
1500
1000
500
0
1250
1000
750
500
250
125
100
75
50
25
400
300
200
100
*
tdTom
BdDHS1a
BdDHS1b
BdDHS2

FIG. 5B

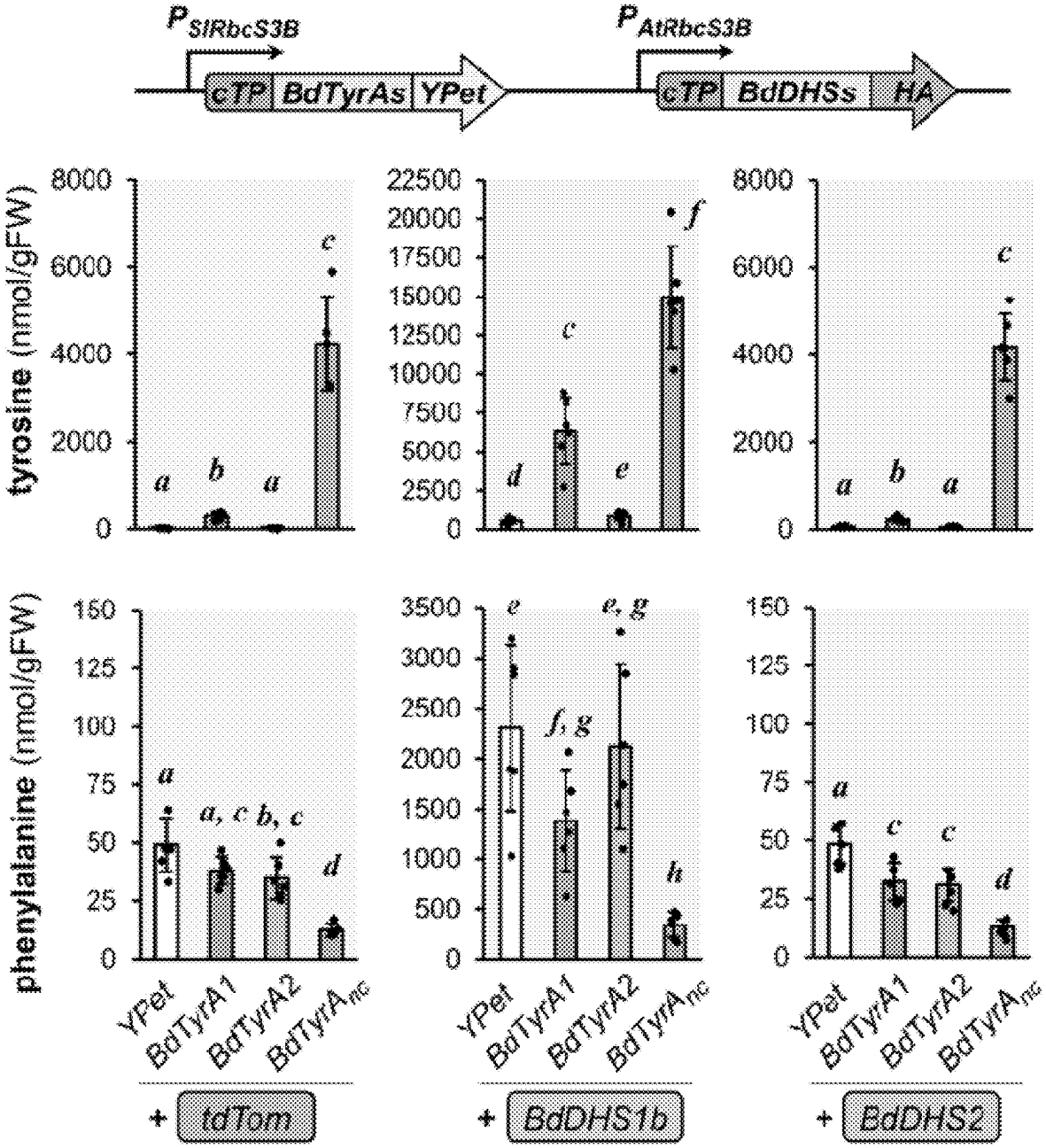

P SlRbcS3B
cTP
BdTyrAs
YPet
P AtRbcS3B
cTP
BdDHSs
HA
tyrosine (nmol/gFW)
phenylalanine (nmol/gFW)
YPet
BdTyrA1
BdTyrA2
BdTyrA nc
+ tdTom
+ BdDHS1b
+ BdDHS2

FIG. 6A

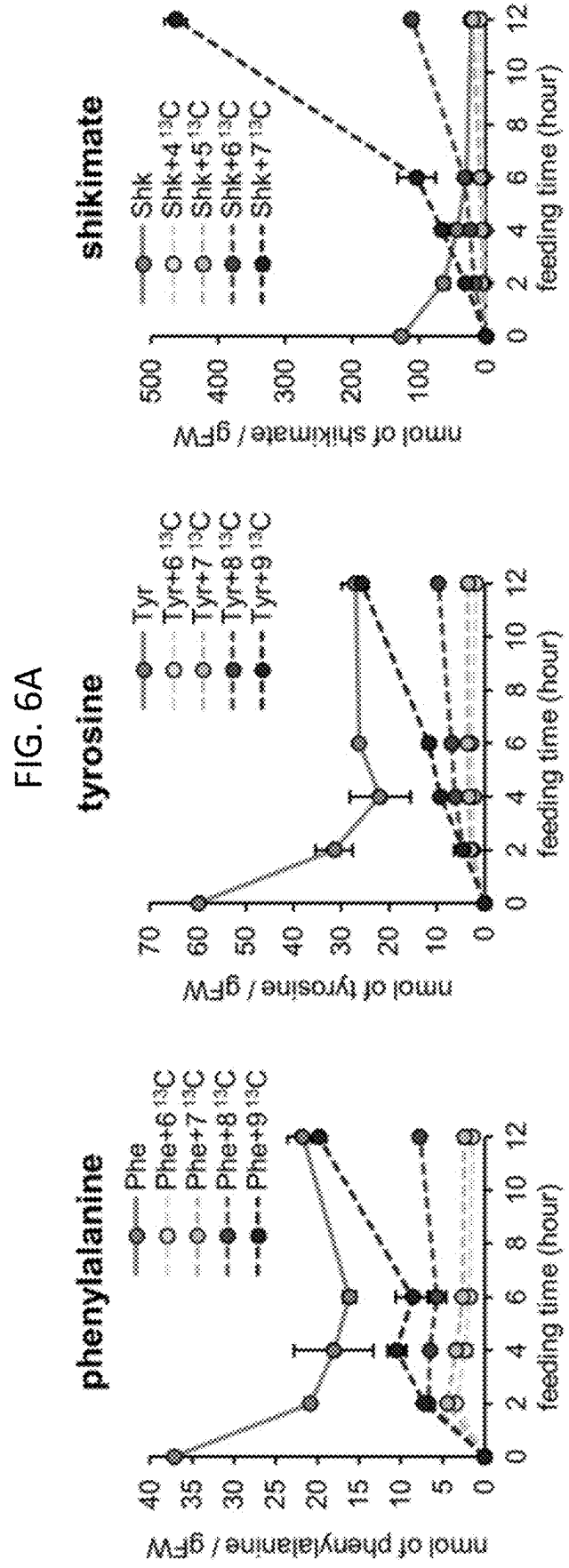

shikimate
Shk
Shk+4 13C
Shk+5 13C
Shk+6 13C
Shk+7 13C
nmol of shikimate / gFW
feeding time (hour)
tyrosine
Tyr
Tyr+6 13C
Tyr+7 13C
Tyr+8 13C
Tyr+9 13C
nmol of tyrosine / gFW
feeding time (hour)
phenylalanine
Phe
Phe+6 13C
Phe+7 13C
Phe+8 13C
Phe+9 13C
nmol of phenylalanine / gFW
feeding time (hour)

FIG. 6B

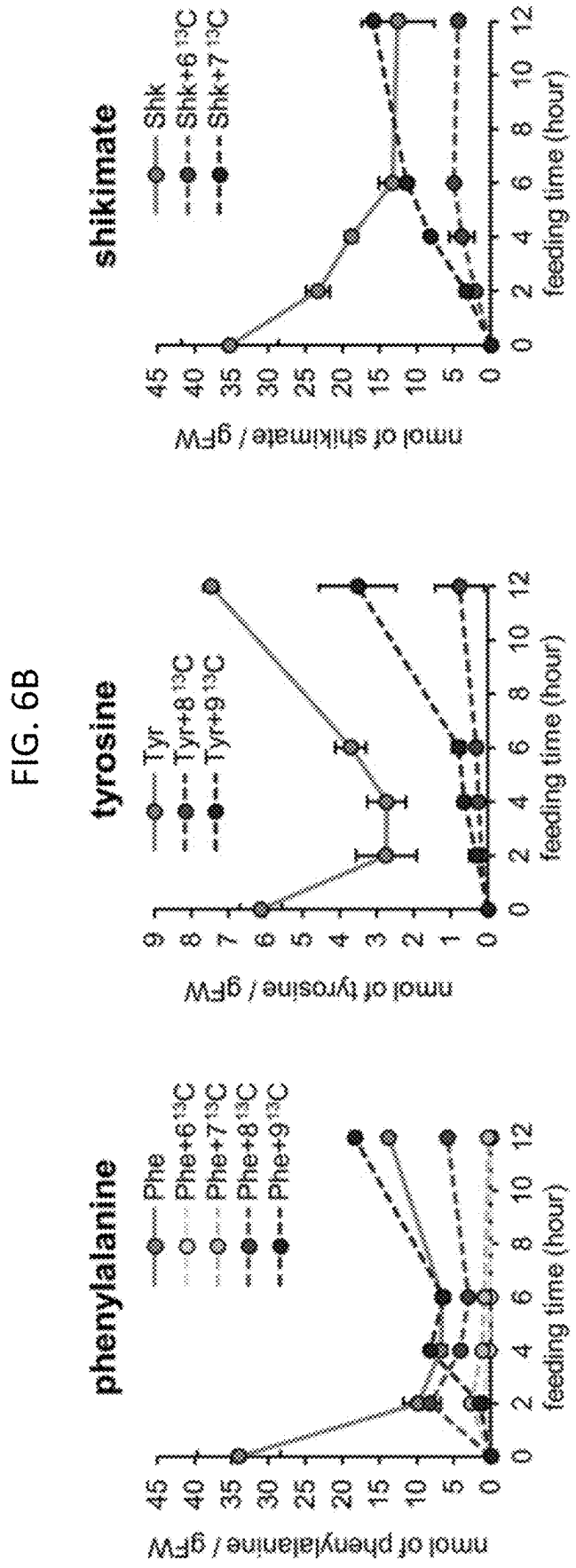

shikimate
Shk
Shk+6 13C
Shk+7 13C
nmol of shikimate / gFW
feeding time (hour)
tyrosine
Tyr
Tyr+8 13C
Tyr+9 13C
nmol of tyrosine / gFW
feeding time (hour)
phenylalanine
Phe
Phe+6 13C
Phe+7 13C
Phe+8 13C
Phe+9 13C
nmol of phenylalanine / gFW
feeding time (hour)

FIG. 7A

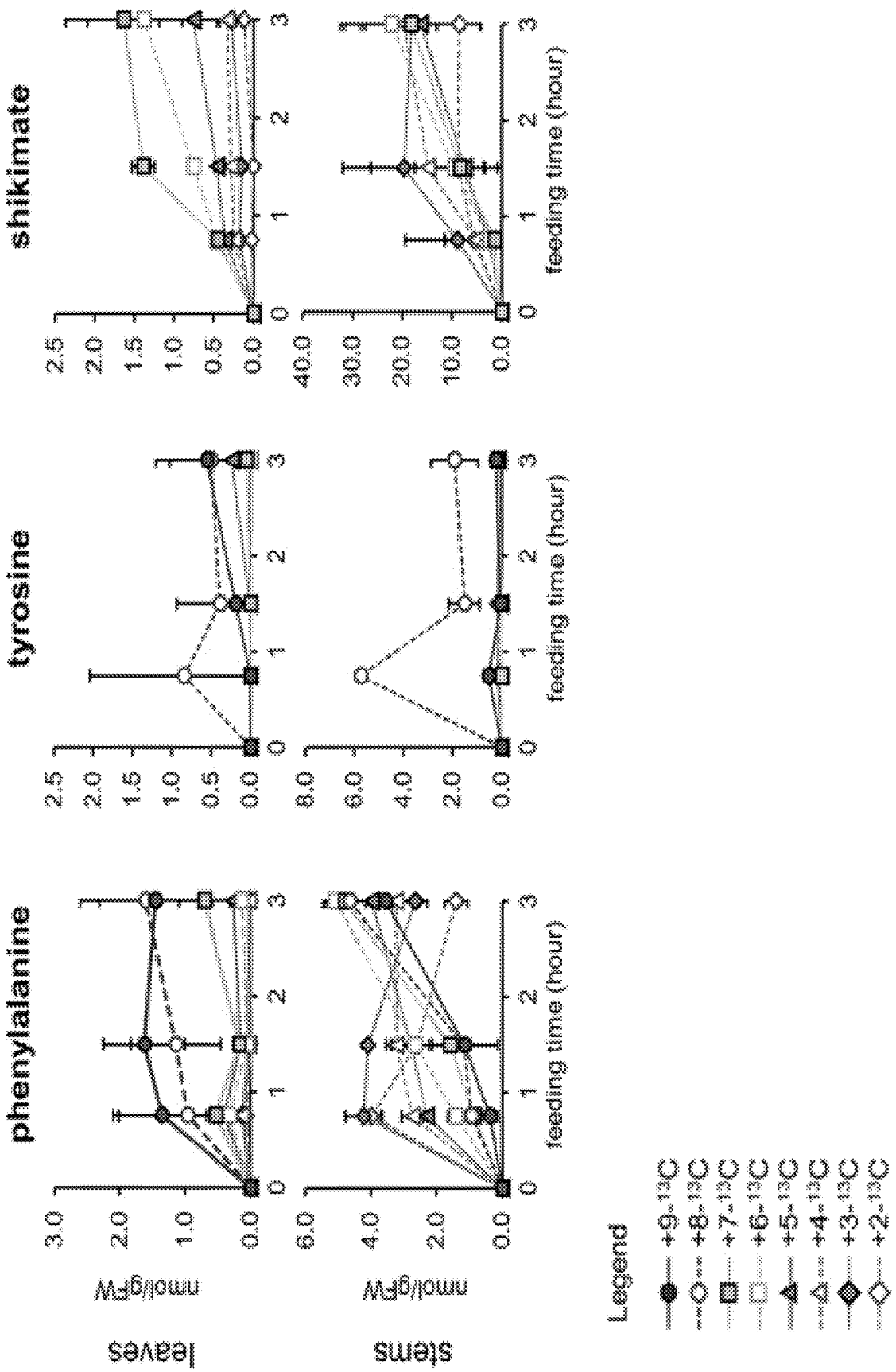

phenylalanine
tyrosine
shikimate
leaves
stems
nmol/gFW
feeding time (hour)
Legend
+9-13C
+8-13C
+7-13C
+6-13C
+5-13C
+4-13C
+3-13C
+2-13C

FIG. 7B

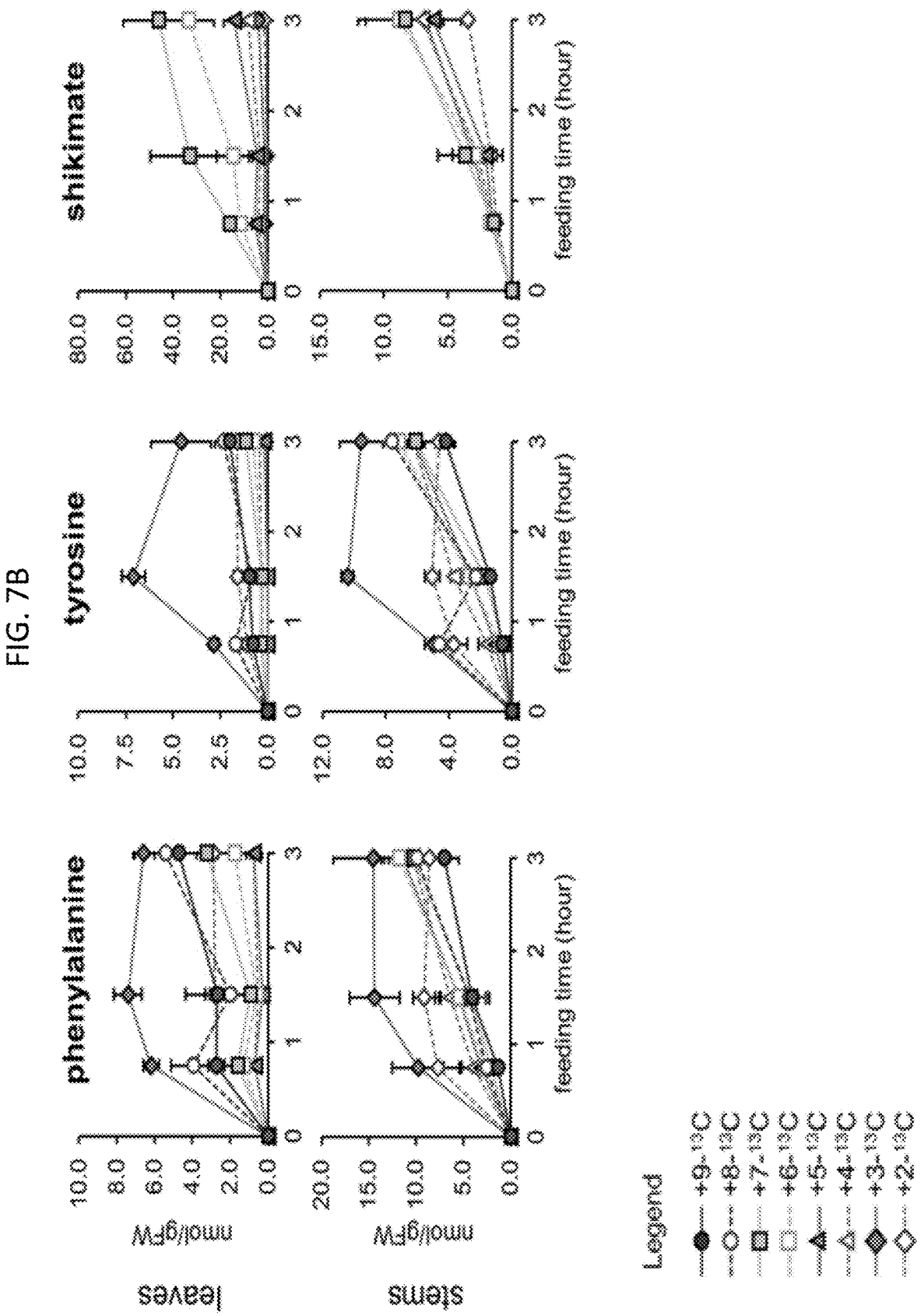

phenylalanine
tyrosine
shikimate
leaves
stems
nmol/gFW
feeding time (hour)
Legend
+9-13C
+8-13C
+7-13C
+6-13C
+5-13C
+4-13C
+3-13C
+2-13C

FIG. 7C

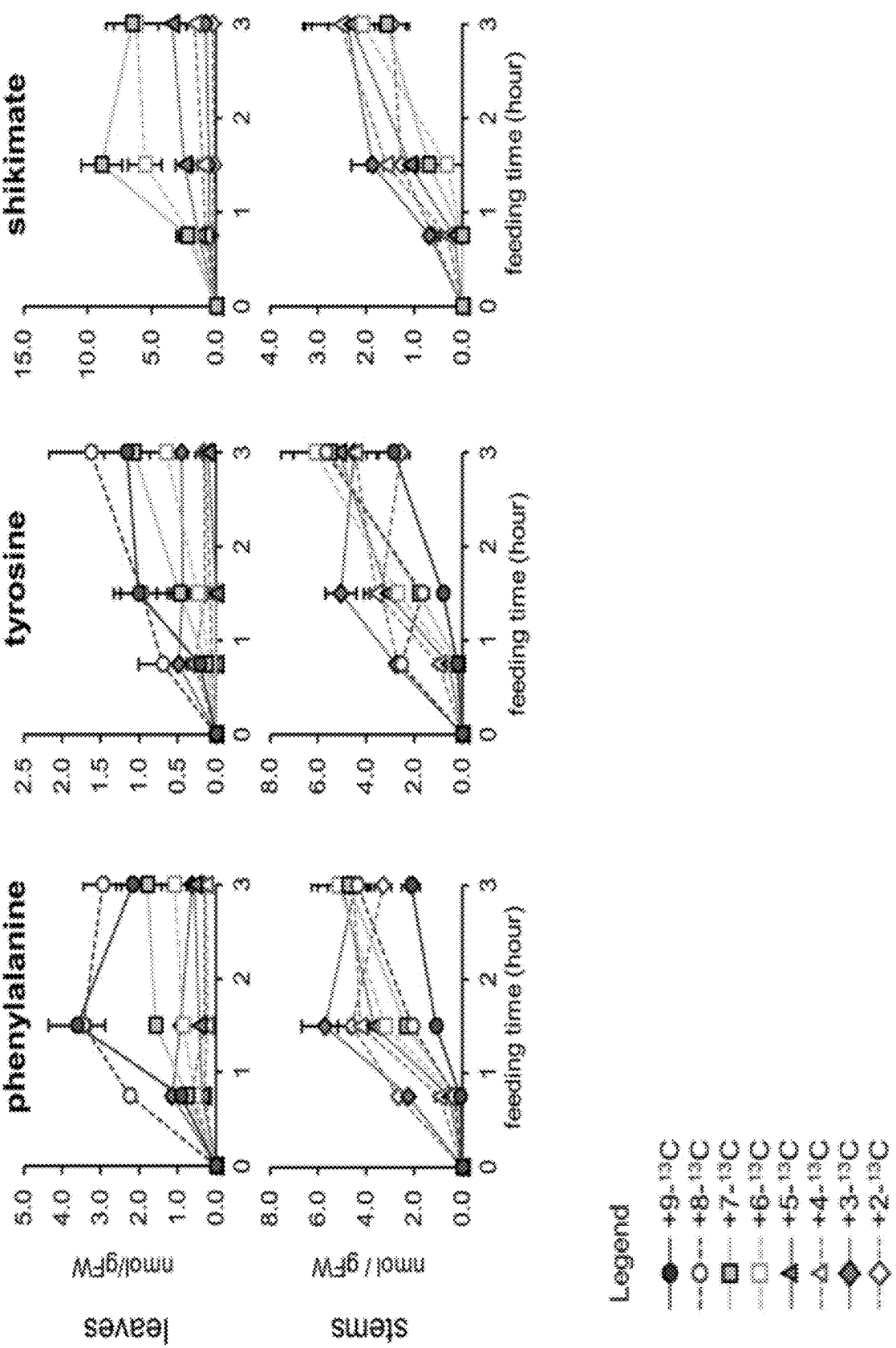

phenylalanine
tyrosine
shikimate
leaves
stems
nmol/gFW
feeding time (hour)
Legend
+9-13C
+8-13C
+7-13C
+6-13C
+5-13C
+4-13C
+3-13C
+2-13C

FIG. 8

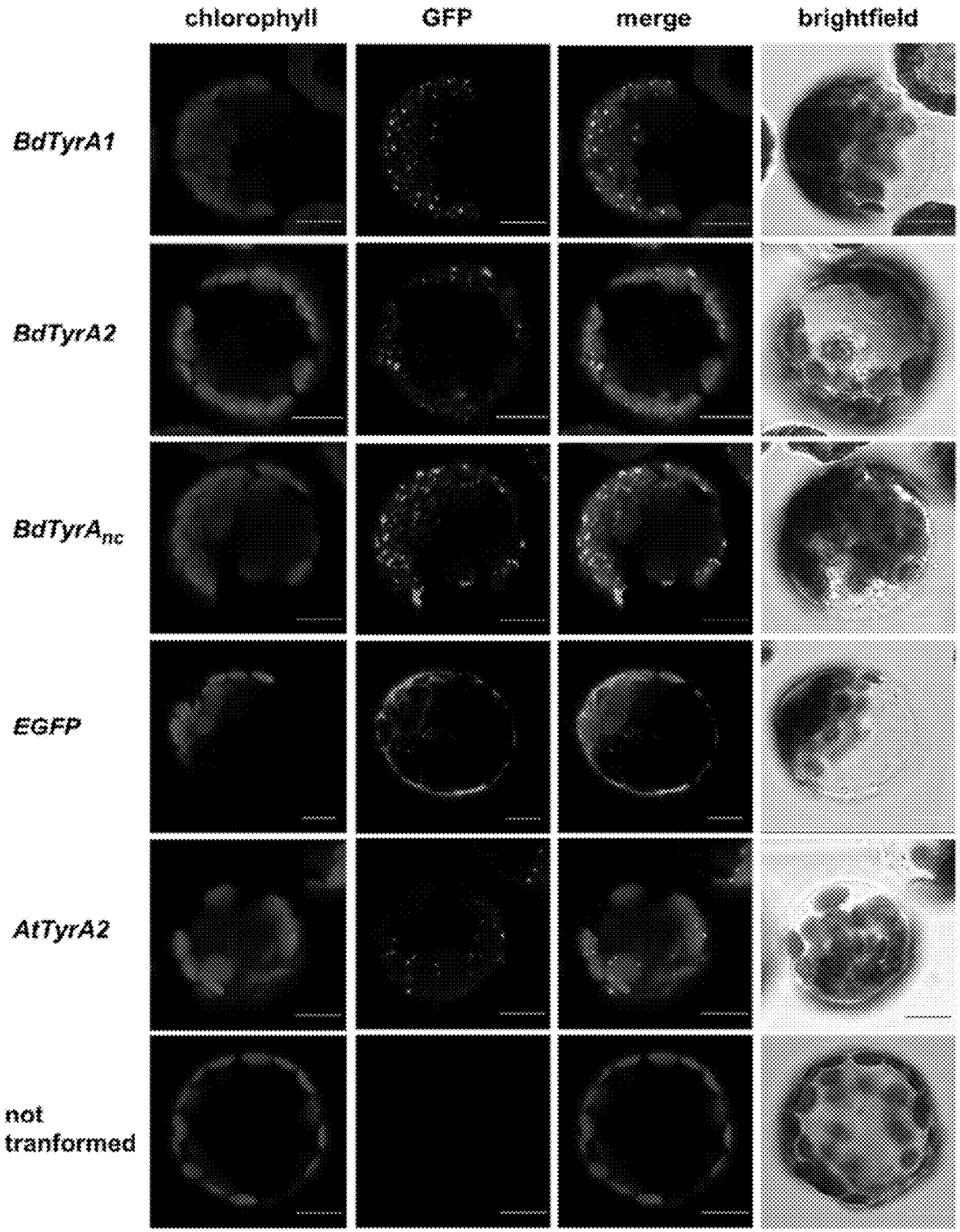

chlorophyll
GFP
merge
brightfield
BdTyrA1
BdTyrA2
BdTyrA$_{nc}$
EGFP
AtTyrA2
not tranformed

FIG. 9B

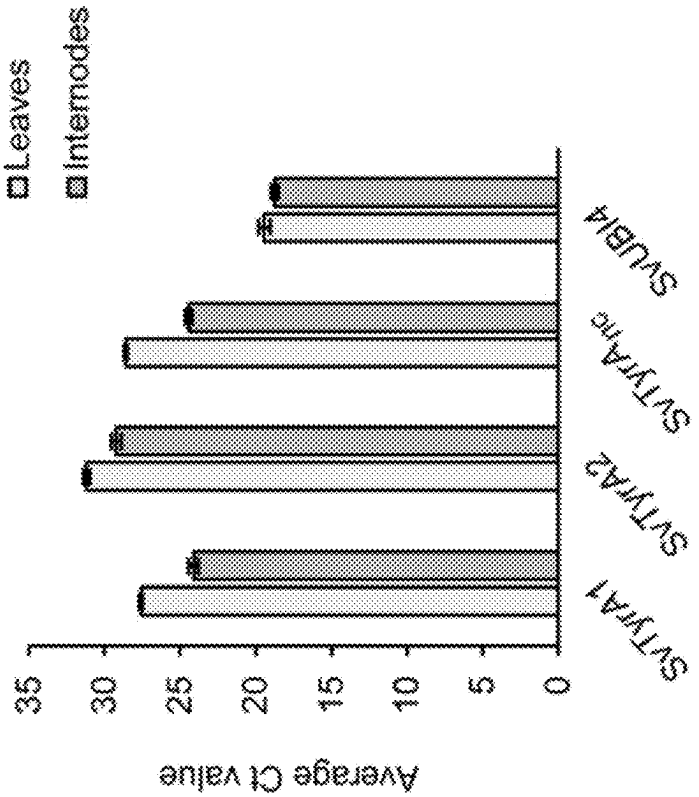

Leaves
Internodes
Average Ct value
35
30
25
20
15
10
5
0
SvTyrA1
SvTyrA2
SvTyrAnc
SvUbi4

FIG. 9A

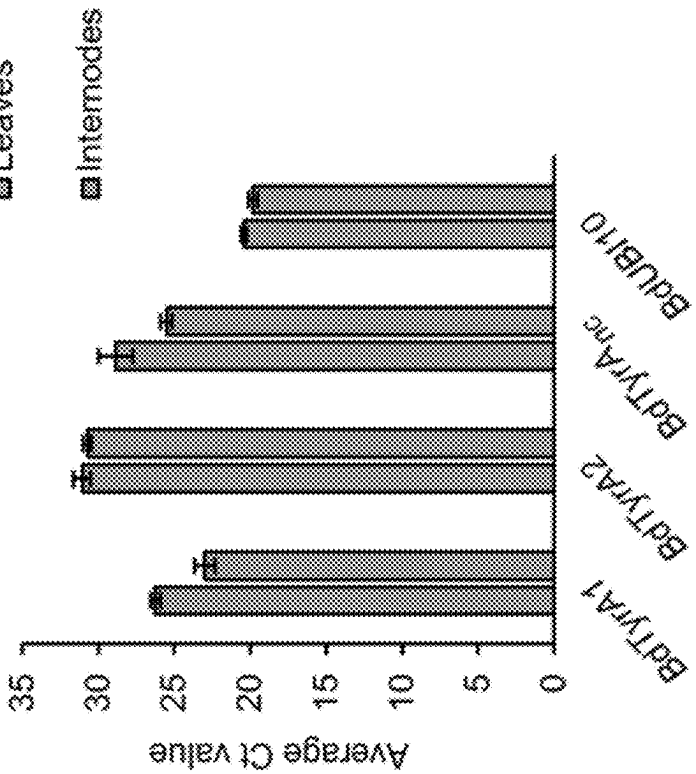

Leaves
Internodes
Average Ct value
35
30
25
20
15
10
5
0
BdTyrA1
BdTyrA2
BdTyrAnc
BdUbi10

FIG. 9C

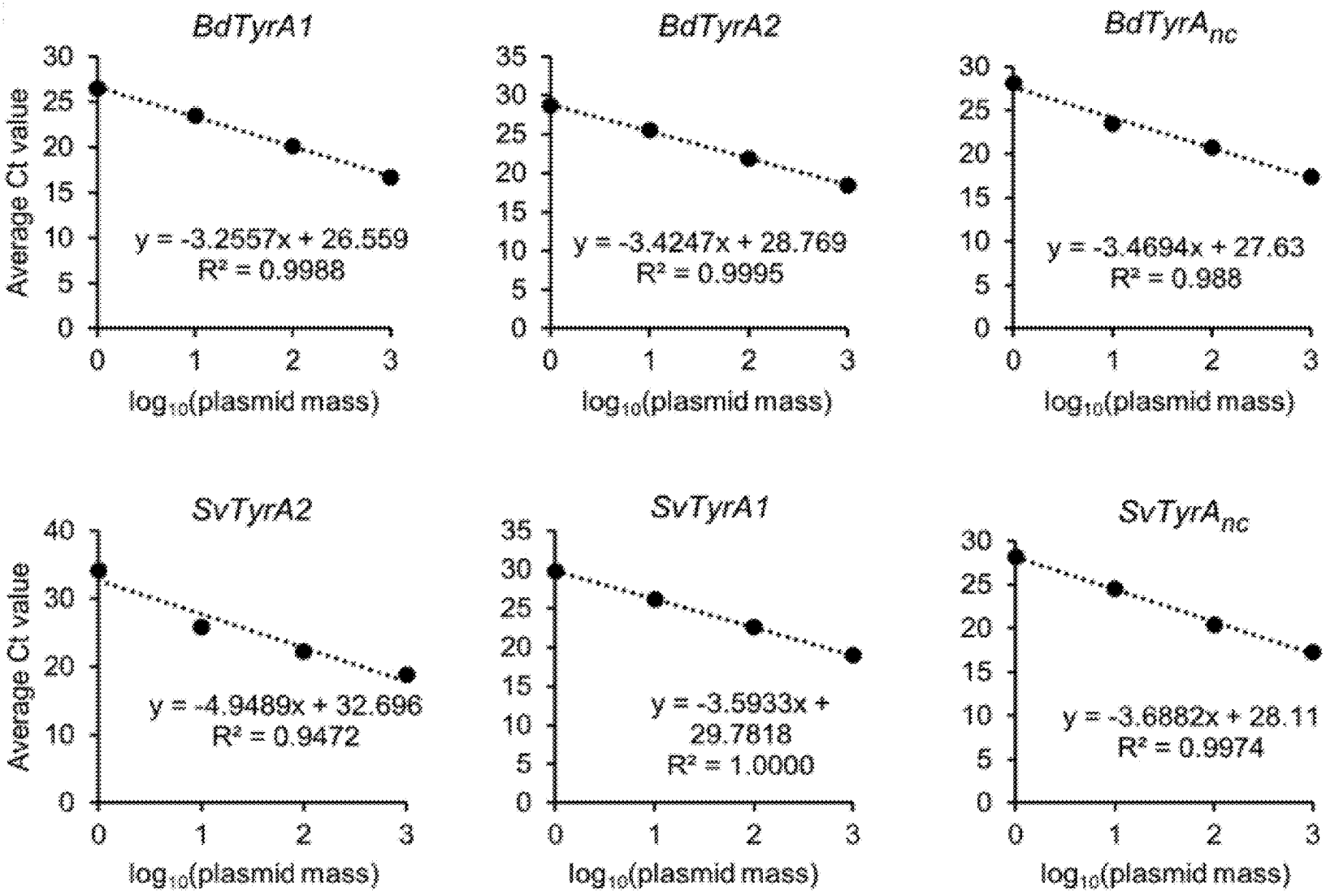

BdTyrA1
Average Ct value
y = -3.2557x + 26.559
R² = 0.9988
log10(plasmid mass)
BdTyrA2
y = -3.4247x + 28.769
R² = 0.9995
log10(plasmid mass)
BdTyrAnc
y = -3.4694x + 27.63
R² = 0.988
log10(plasmid mass)
SvTyrA2
Average Ct value
y = -4.9489x + 32.696
R² = 0.9472
log10(plasmid mass)
SvTyrA1
y = -3.5933x + 29.7818
R² = 1.0000
log10(plasmid mass)
SvTyrAnc
y = -3.6882x + 28.11
R² = 0.9974
log10(plasmid mass)

FIG. 10B

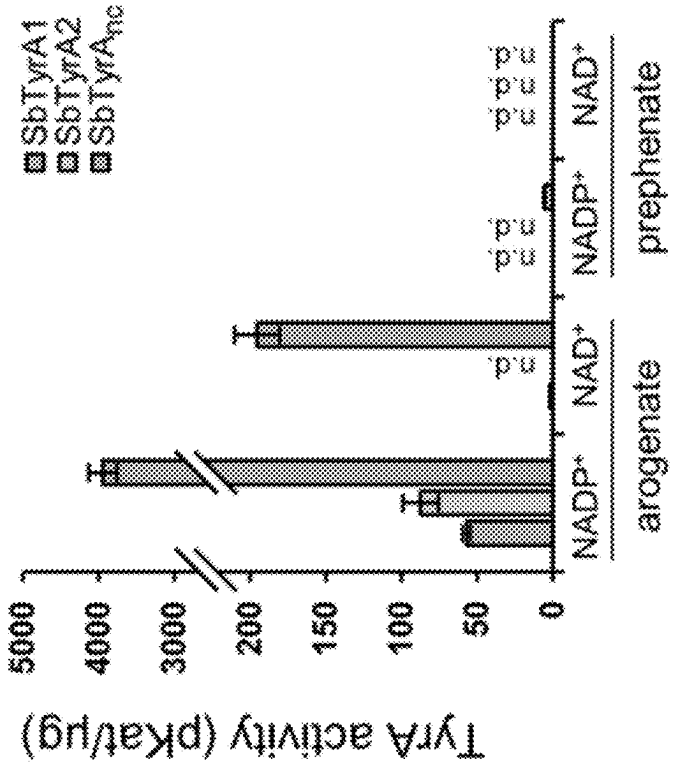


FIG. 10A

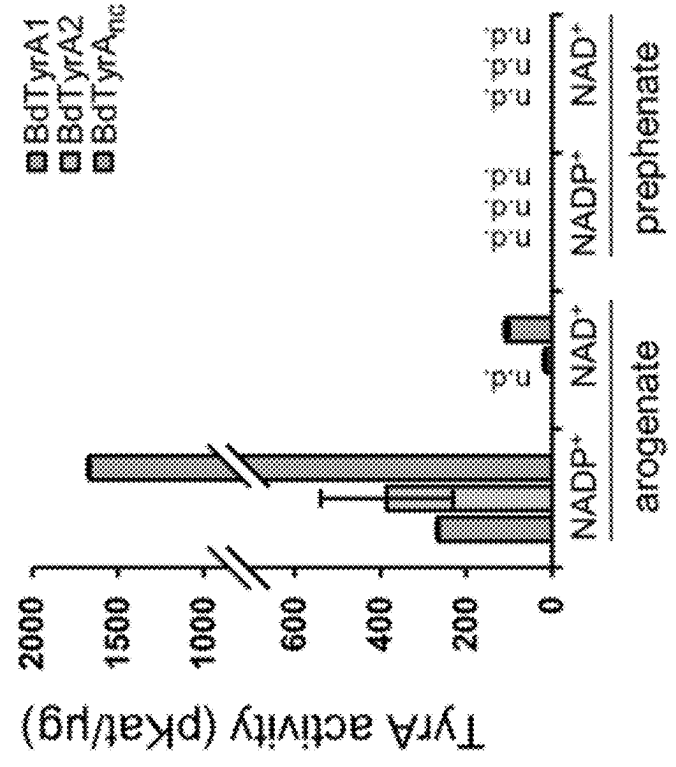


FIG. 11A

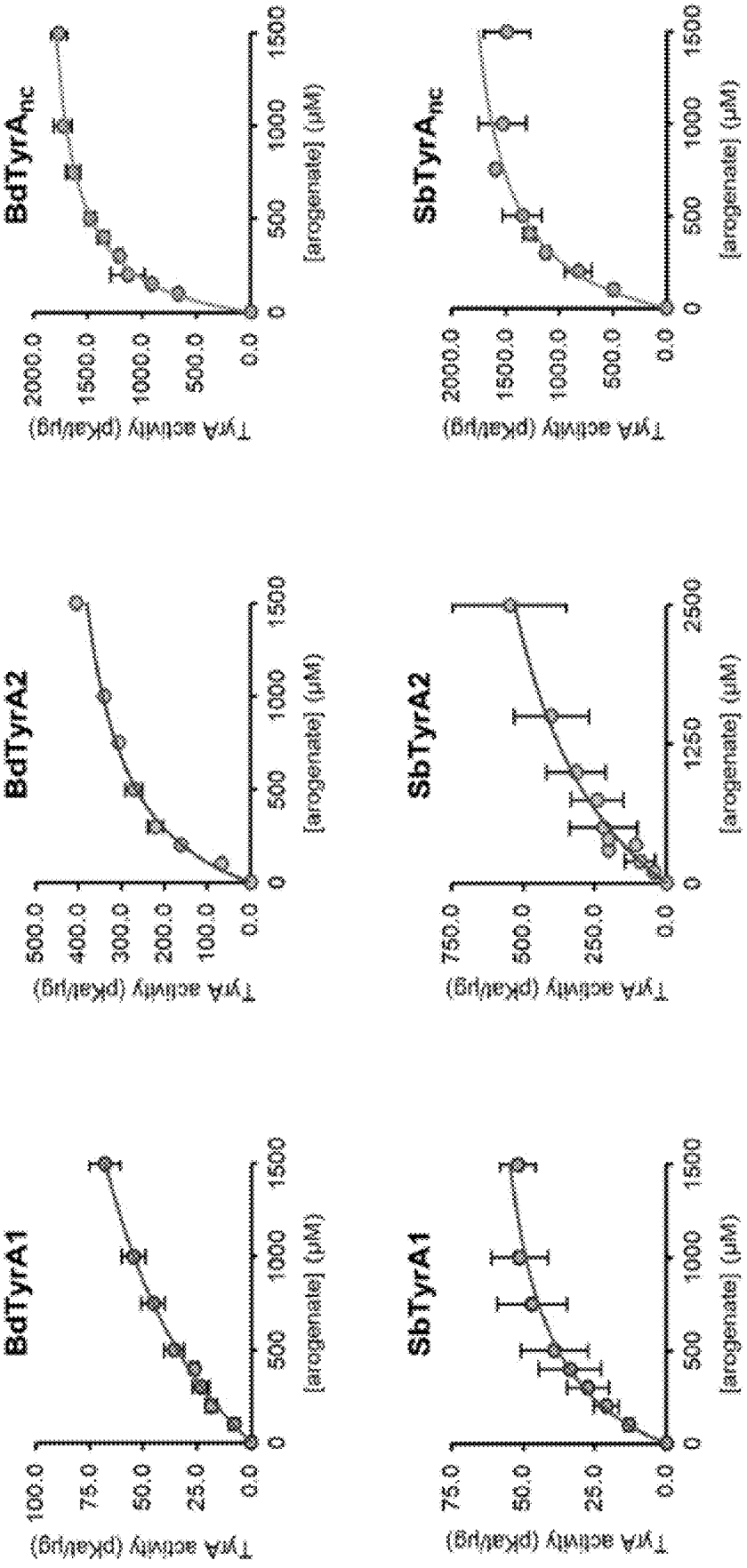

BdTyrA1
BdTyrA2
BdTyrA$_{mc}$
SbTyrA1
SbTyrA2
SbTyrA$_{mc}$
TyrA activity (pKat/µg)
[arogenate] (µM)

FIG. 11B

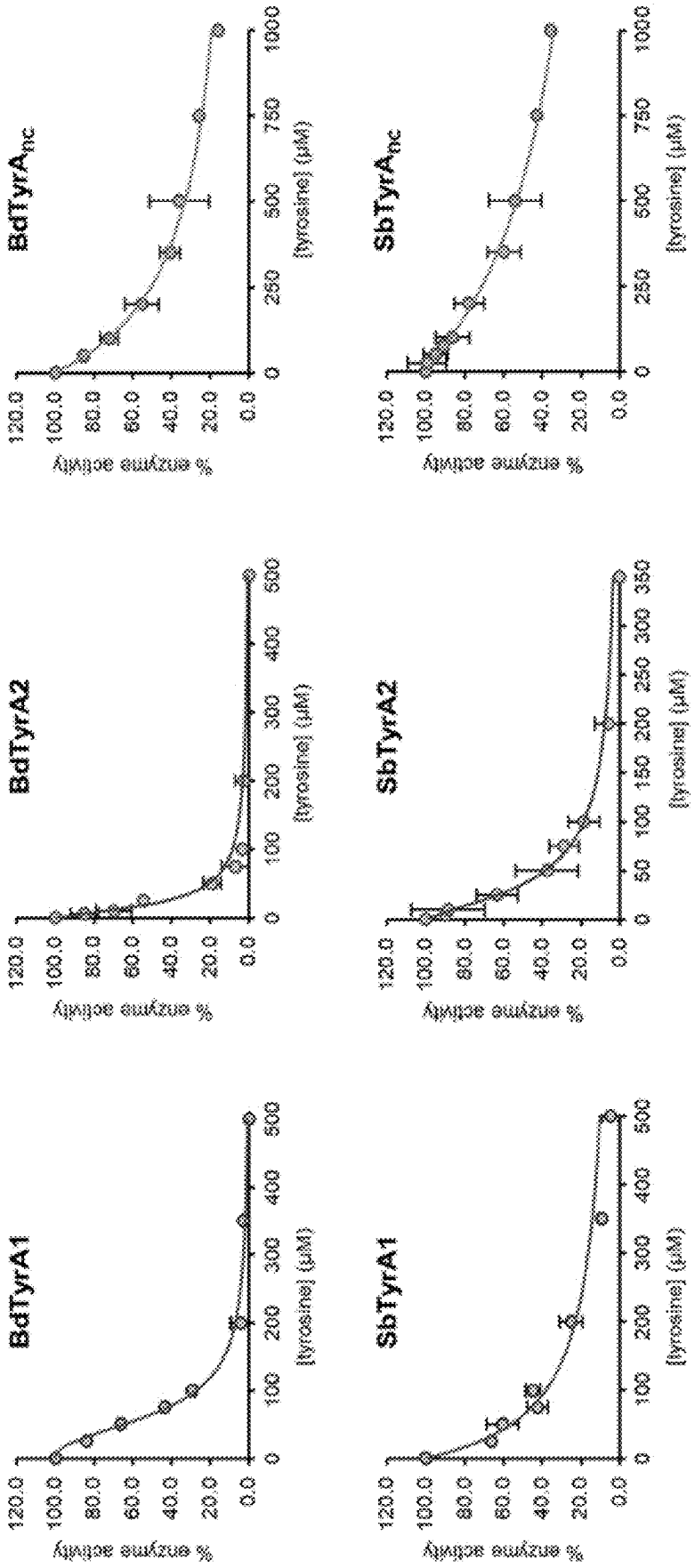

BdTyrA1
BdTyrA2
BdTyrA$_{mc}$
SbTyrA1
SbTyrA2
SbTyrA$_{mc}$
% enzyme activity
[tyrosine] (μM)

FIG. 12A

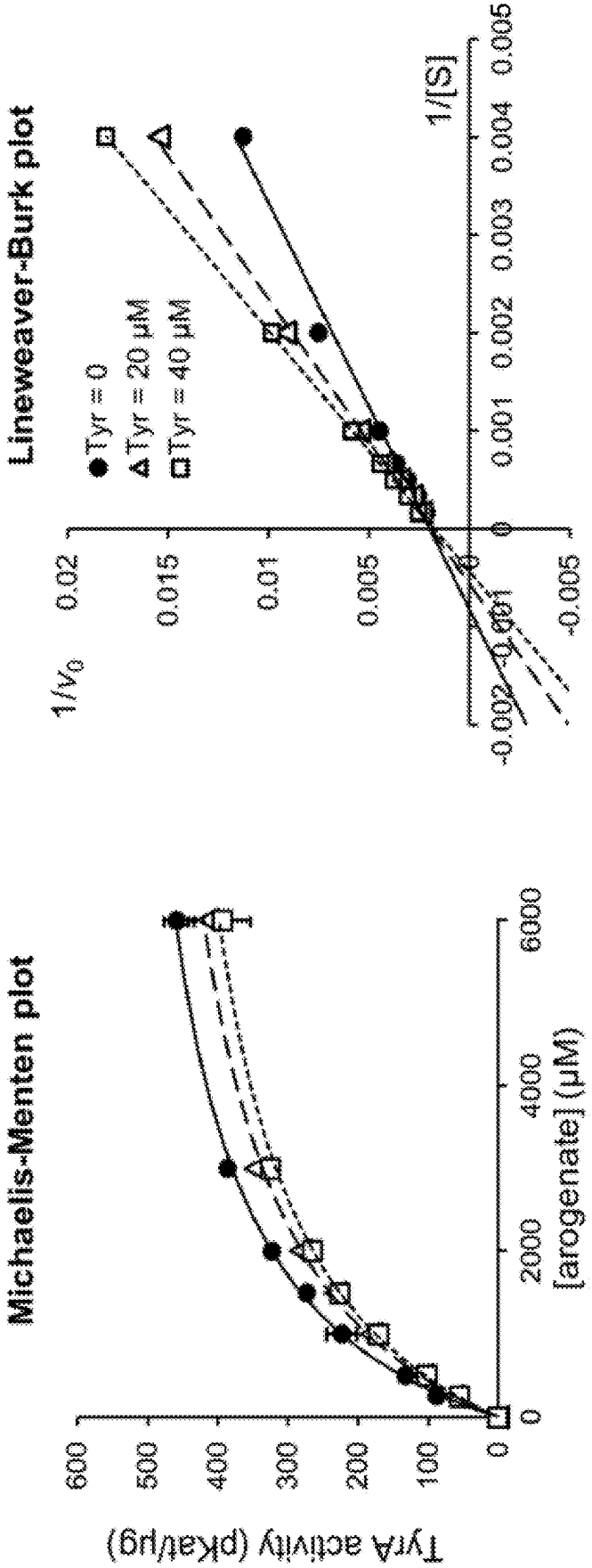

Lineweaver-Burk plot
Tyr = 0
Tyr = 20 µM
Tyr = 40 µM
1/v0
1/[S]
0.02
0.015
0.01
0.005
0
-0.005
-0.002
-0.001
0.001
0.002
0.003
0.004
0.005
Michaelis-Menten plot
TyrA activity (pKat/µg)
[arogenate] (µM)
600
500
400
300
200
100
0
2000
4000
6000

FIG. 12B

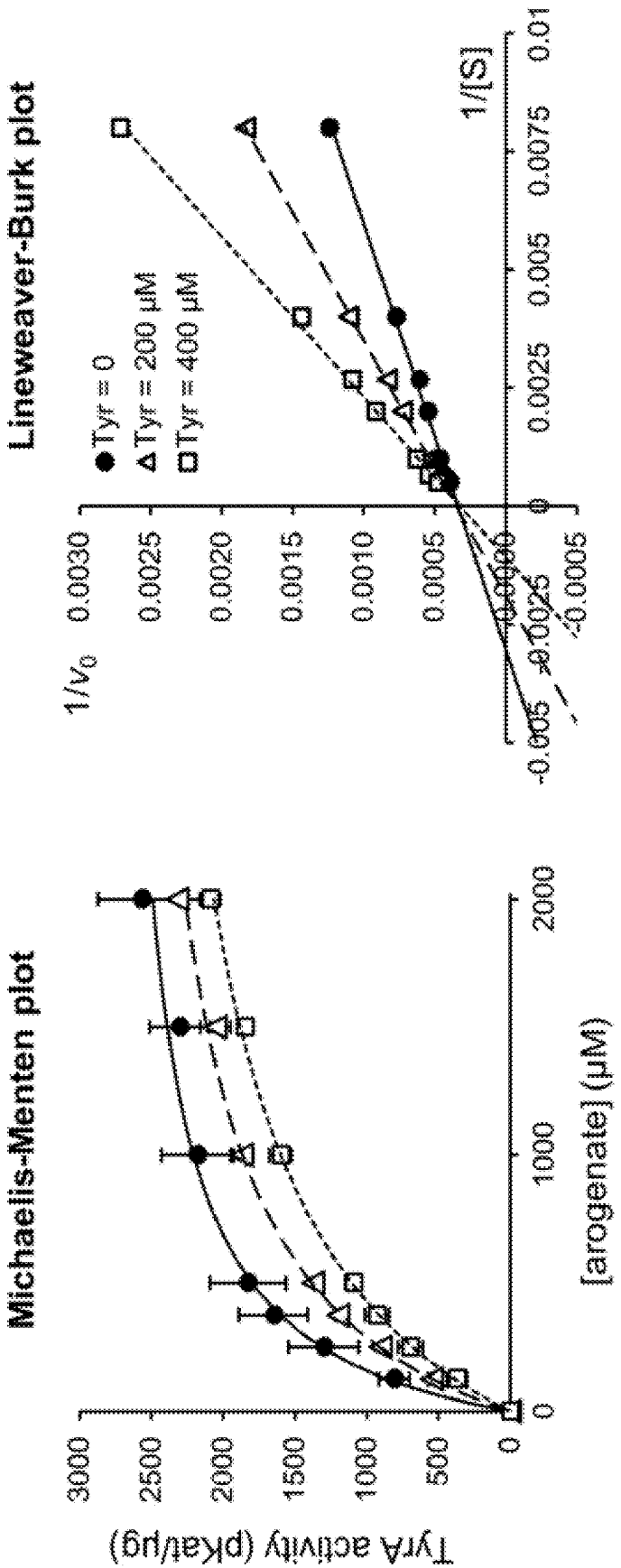

Lineweaver-Burk plot
Tyr = 0
Tyr = 200 μM
Tyr = 400 μM
1/v0
1/[S]
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0
-0.0005
-0.0010
-0.0015
-0.0020
-0.0025
-0.0030
-0.005
0
0.0025
0.005
0.0075
0.01
Michaelis-Menten plot
TyrA activity (pKat/μg)
[arogenate] (μM)
3000
2500
2000
1500
1000
500
0
1000
2000

FIG. 12C

		Tyr = 0 μM	Tyr = 20 μM	Tyr = 40 μM
BdTyrA1	**$K_{m\ app}$ (μM arogenate)**	1623 ± 269	1876 ± 194	2162 ± 493
	$V_{max\ app}$ (pKat/μg)	586 ± 38	563 ± 64	551 ± 69
		Tyr = 0 μM	**Tyr = 200 μM**	**Tyr = 400 μM**
$BdTyrA_{nc}$	**$K_{m\ app}$ (μM arogenate)**	299 ± 38	549 ± 131	849 ± 248
	$V_{max\ app}$ (pKat/μg)	2861 ± 257	2894 ± 380	2946 ± 302

FIG. 13A

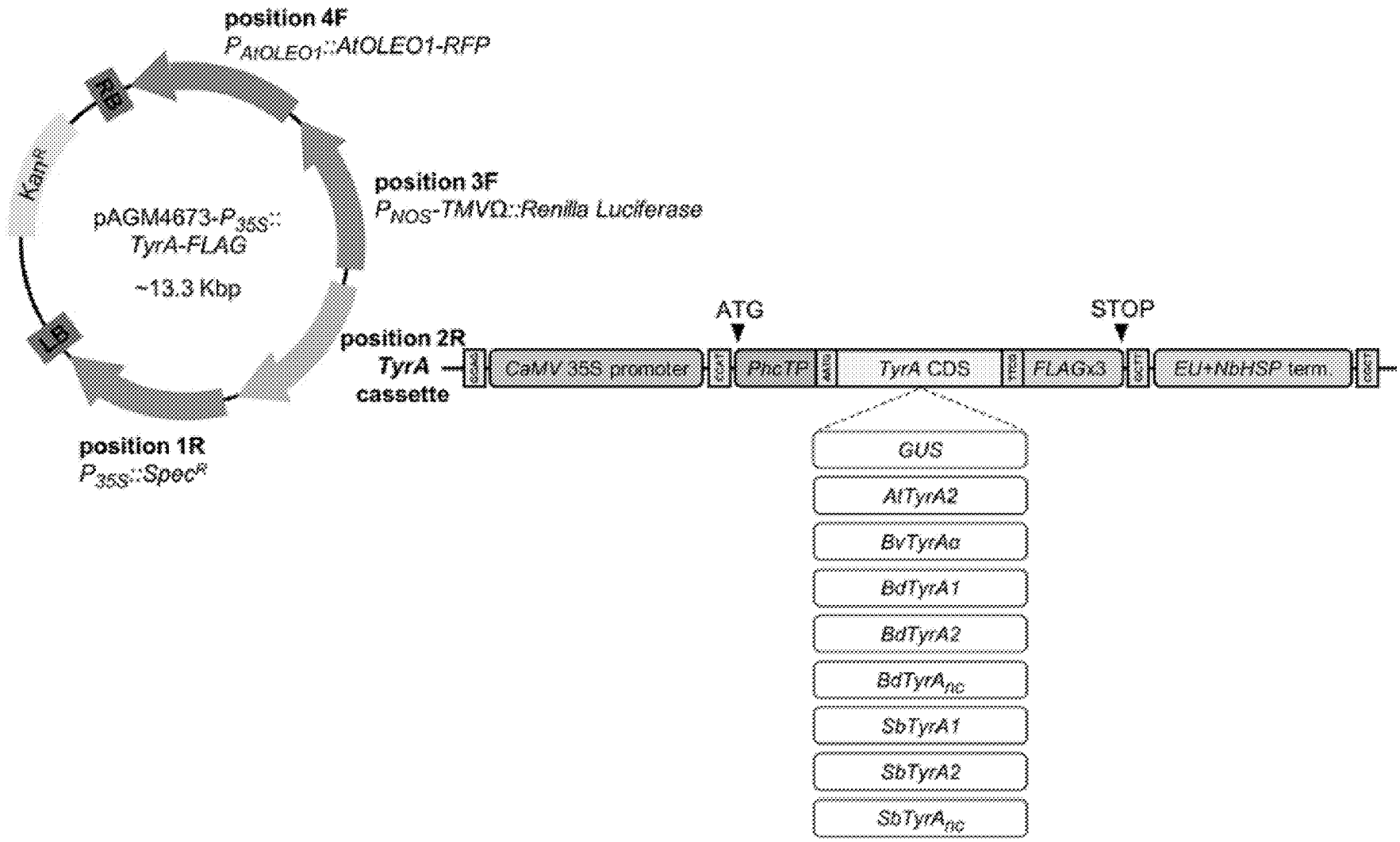

position 4F
$P_{AtOLEO1}$::AtOLEO1-RFP
RB
Kan^R
pAGM4673-P_{35S}::
TyrA-FLAG
~13.3 Kbp
LB
position 3F
P_{NOS}-TMVΩ::Renilla Luciferase
position 2R
TyrA
cassette
ATG
STOP
CaMV 35S promoter
PhcTP
TyrA CDS
FLAGx3
EU+NbHSP term.
GUS
AtTyrA2
BvTyrAα
BdTyrA1
BdTyrA2
BdTyrA$_{nc}$
SbTyrA1
SbTyrA2
SbTyrA$_{nc}$
position 1R
P_{35S}::Spec^R

FIG. 13B

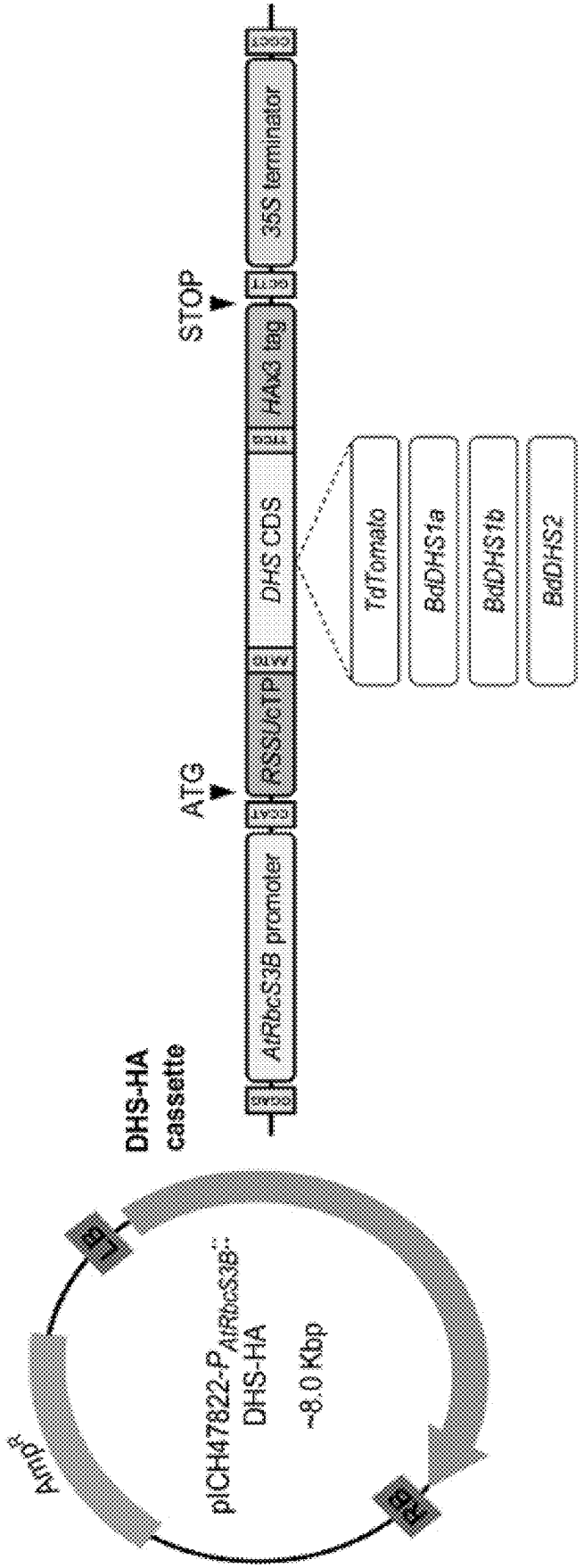
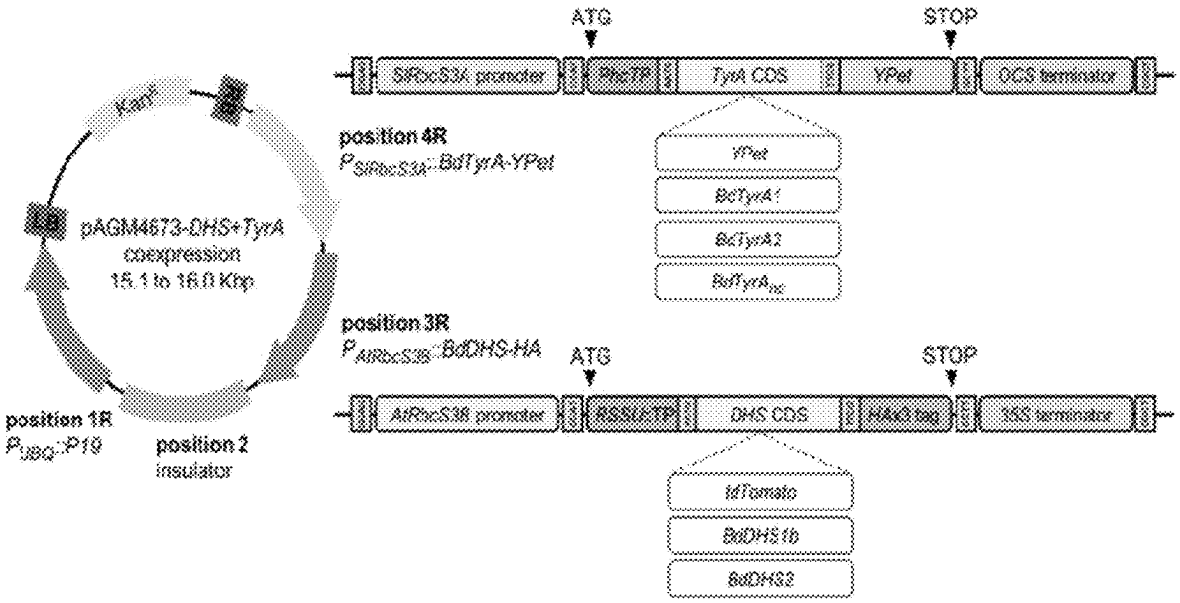
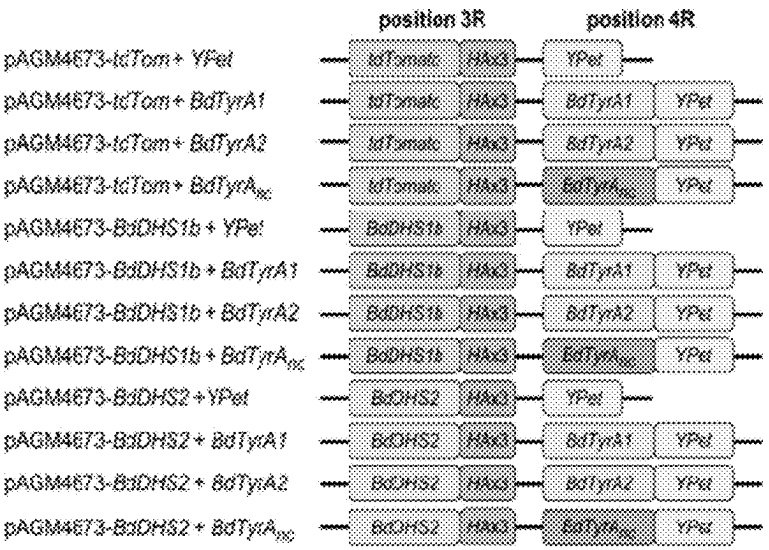

pICH47822-P AtRbcS3B::
DHS-HA
~8.0 Kbp
AmpR
DHS-HA cassette
AtRbcS3B promoter
ATG
RSSUcTP
DHS CDS
HA x3 tag
STOP
35S terminator
TdTomato
BdDHS1a
BdDHS1b
BdDHS2

FIG. 13C


ATG
STOP
SlRbcS3A promoter
TyrA CDS
YPet
OCS terminator
position 4R
$P_{SlRbcS3A}$::BdTyrA-YPet
YPet
BdTyrA1
BdTyrA2
$BdTyrA_{fbr}$
pAGM4673-DHS+TyrA
coexpression
15.1 to 16.0 Kbp
position 3R
$P_{AtRbcS3B}$::BdDHS-HA
ATG
STOP
AtRbcS3B promoter
DHS CDS
HA-tag
35S terminator
tdTomato
BdDHS1b
BdDHS2
position 1R
P_{UBQ}::P19
position 2
insulator


position 3R
position 4R
pAGM4673-tdTom + YPet
pAGM4673-tdTom + BdTyrA1
pAGM4673-tdTom + BdTyrA2
pAGM4673-tdTom + $BdTyrA_{fbr}$
pAGM4673-BdDHS1b + YPet
pAGM4673-BdDHS1b + BdTyrA1
pAGM4673-BdDHS1b + BdTyrA2
pAGM4673-BdDHS1b + $BdTyrA_{fbr}$
pAGM4673-BdDHS2 + YPet
pAGM4673-BdDHS2 + BdTyrA1
pAGM4673-BdDHS2 + BdTyrA2
pAGM4673-BdDHS2 + $BdTyrA_{fbr}$
tdTomato
BdDHS1b
BdDHS2
HA-tag
YPet
BdTyrA1
BdTyrA2
$BdTyrA_{fbr}$

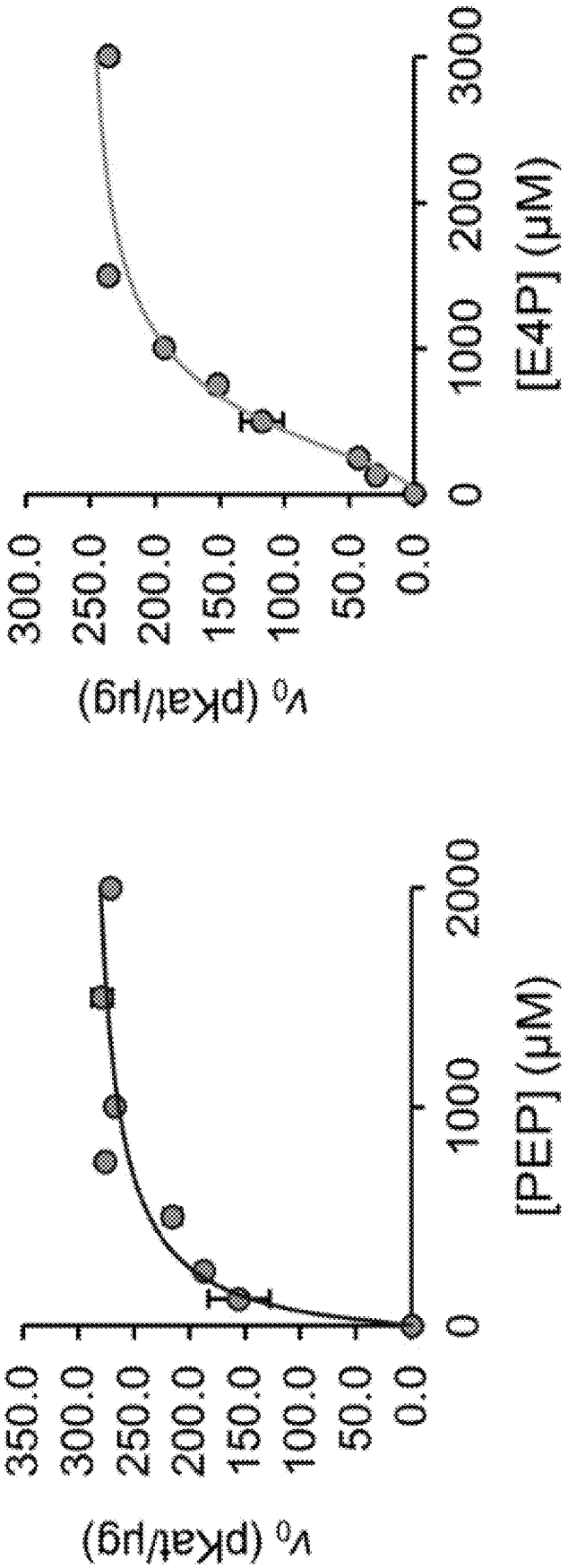

v_0 (pKat/µg)
300.0
250.0
200.0
150.0
100.0
50.0
0.0
0
1000
2000
3000
[E4P] (µM)
v_0 (pKat/µg)
350.0
300.0
250.0
200.0
150.0
100.0
50.0
0.0
0
1000
2000
[PEP] (µM)

FIG. 14A

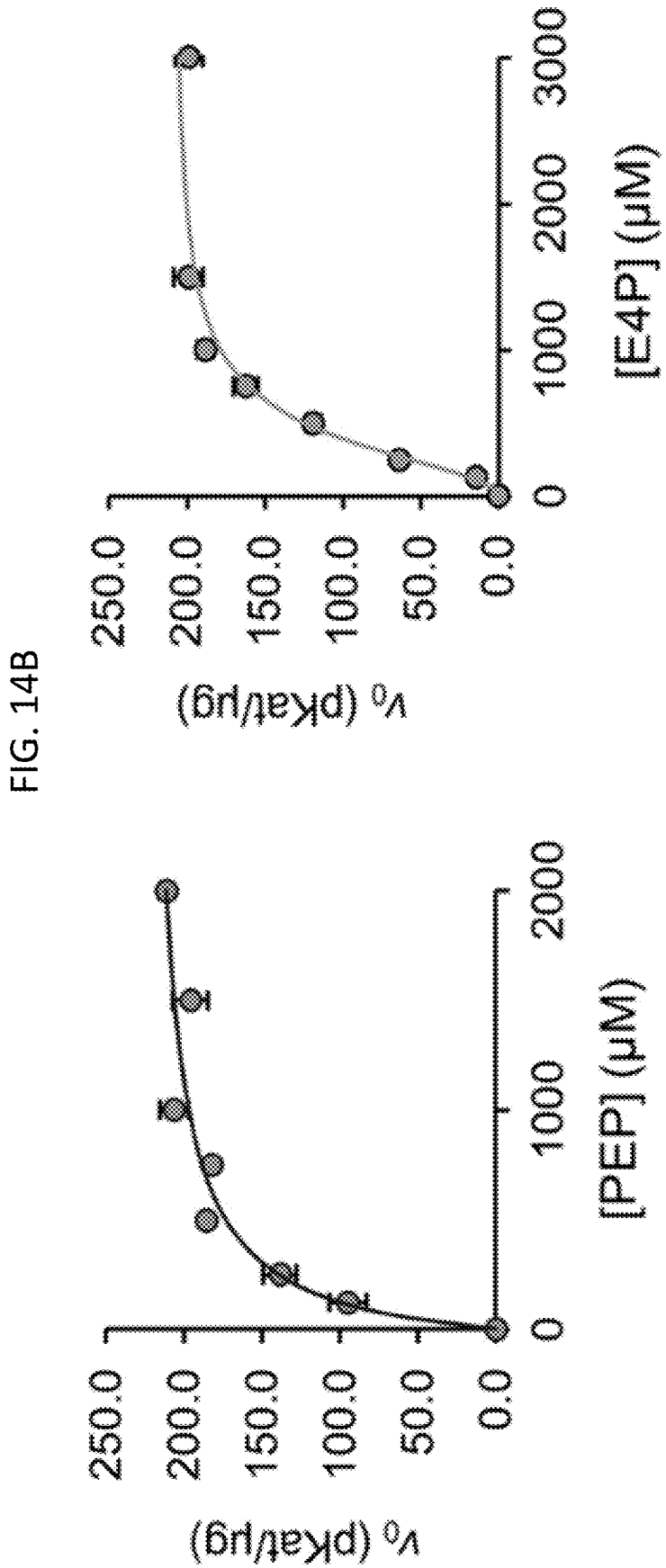

250.0
200.0
150.0
100.0
50.0
0.0
v_0 (pKat/µg)
0
1000
2000
3000
[E4P] (µM)
250.0
200.0
150.0
100.0
50.0
0.0
v_0 (pKat/µg)
0
1000
2000
[PEP] (µM)

FIG. 14B

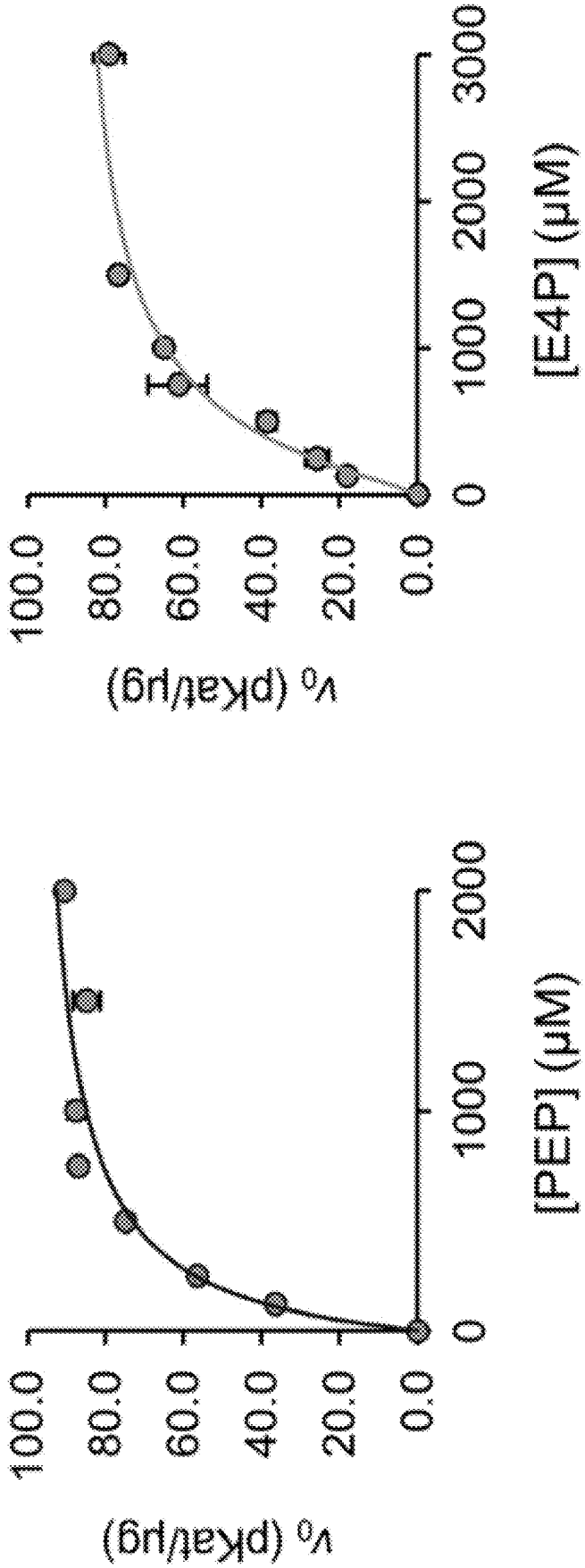

v0 (pKat/µg)
100.0
80.0
60.0
40.0
20.0
0.0
0
1000
2000
3000
[E4P] (µM)
v0 (pKat/µg)
100.0
80.0
60.0
40.0
20.0
0.0
0
1000
2000
[PEP] (µM)

FIG. 14C

FIG. 15A

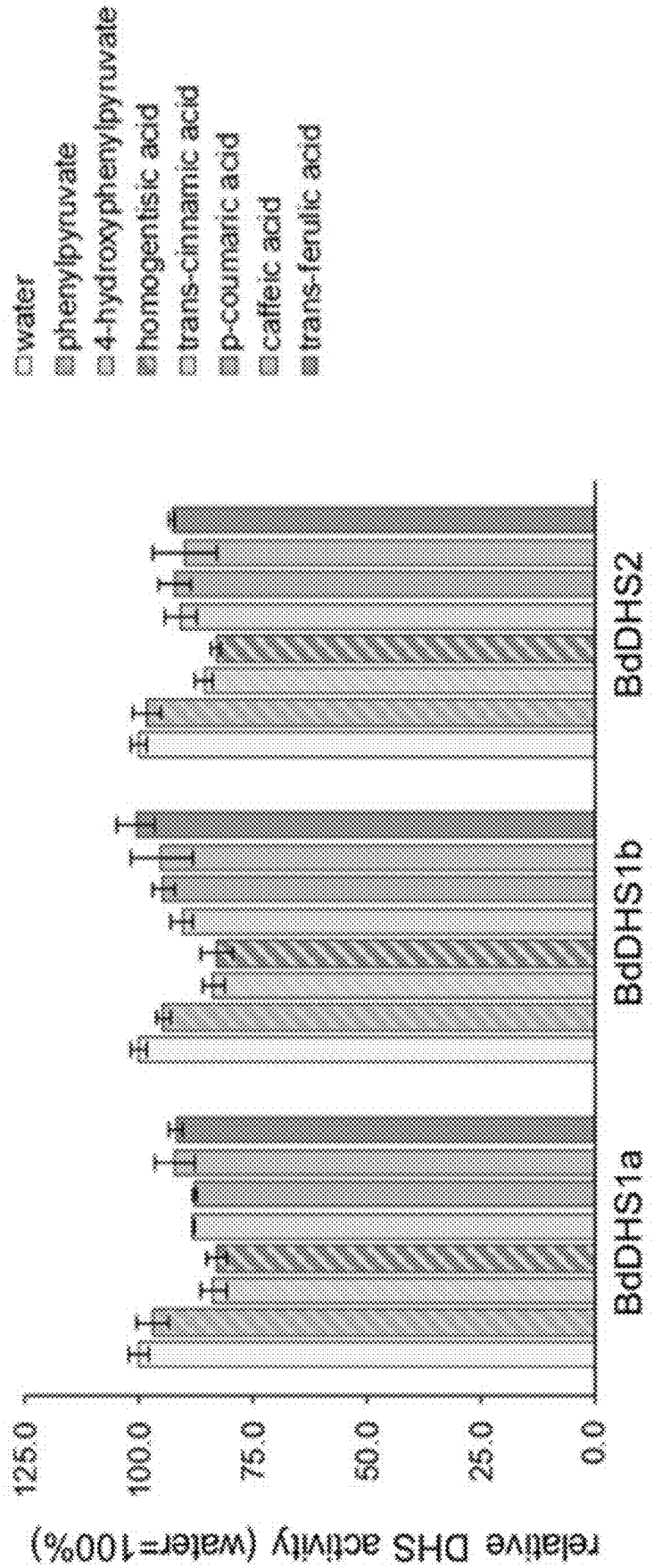

water
phenylpyruvate
4-hydroxyphenylpyruvate
homogentisic acid
trans-cinnamic acid
p-coumaric acid
caffeic acid
trans-ferulic acid
relative DHS activity (water=100%)
125.0
100.0
75.0
50.0
25.0
0.0
BdDHS1a
BdDHS1b
BdDHS2

FIG. 15B

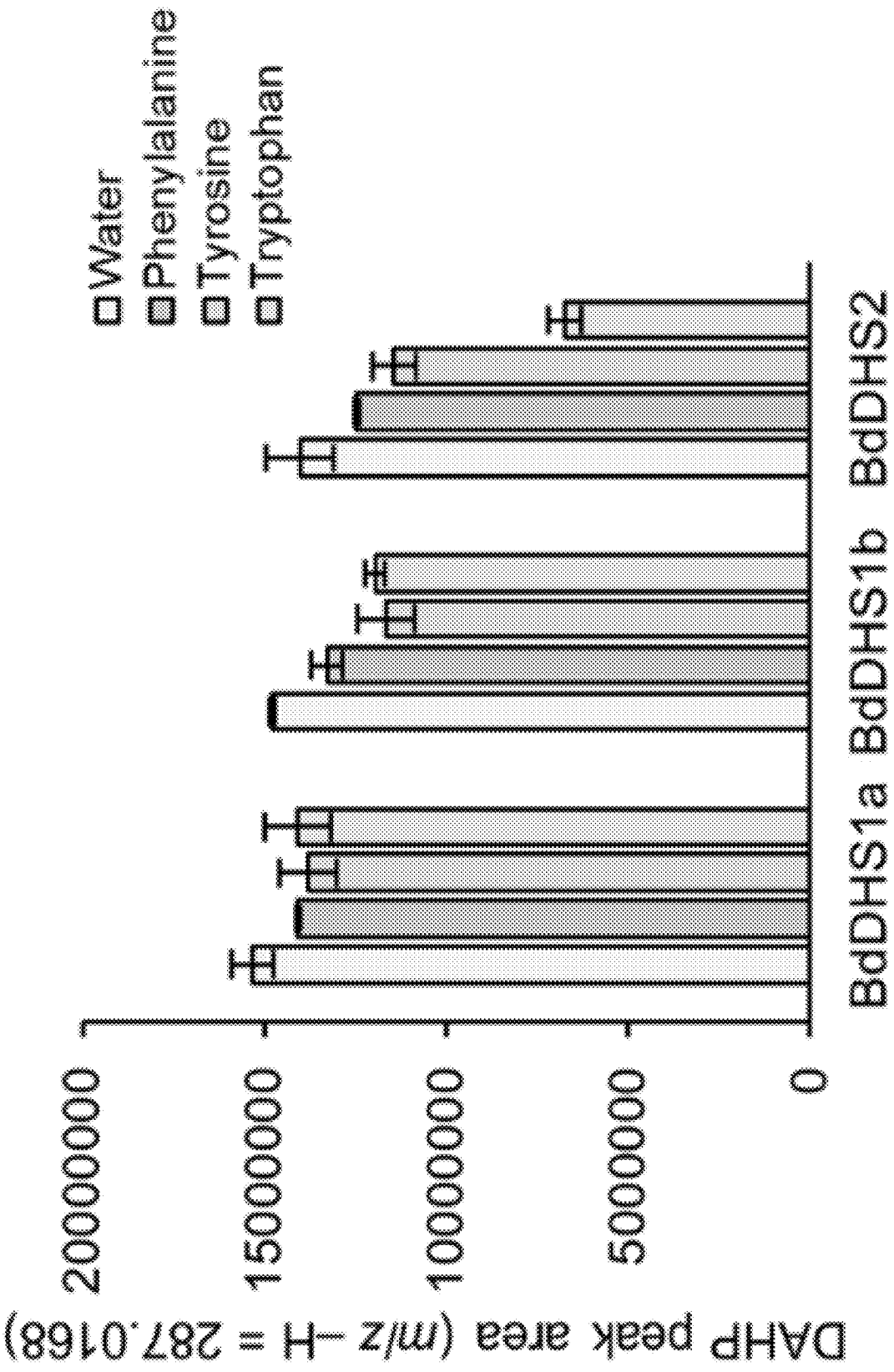

Water
Phenylalanine
Tyrosine
Tryptophan
DAHP peak area (m/z −H = 287.0168)
20000000
15000000
10000000
5000000
0
BdDHS1a
BdDHS1b
BdDHS2

FIG. 15C

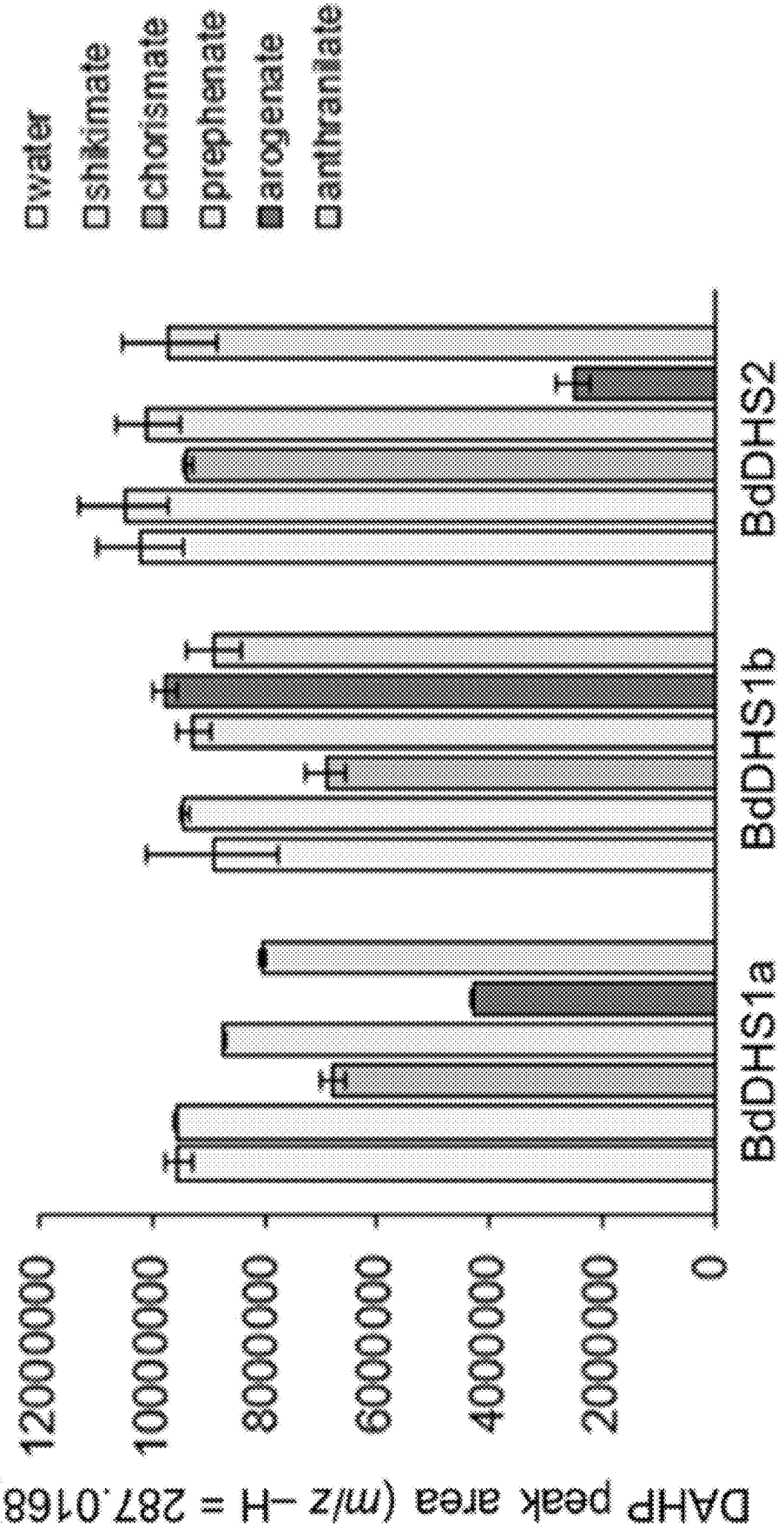

water
shikimate
chorismate
prephenate
arogenate
anthranilate
DAHP peak area (m/z −H = 287.0168)
12000000
10000000
8000000
6000000
4000000
2000000
0
BdDHS1a
BdDHS1b
BdDHS2

FIG. 16A

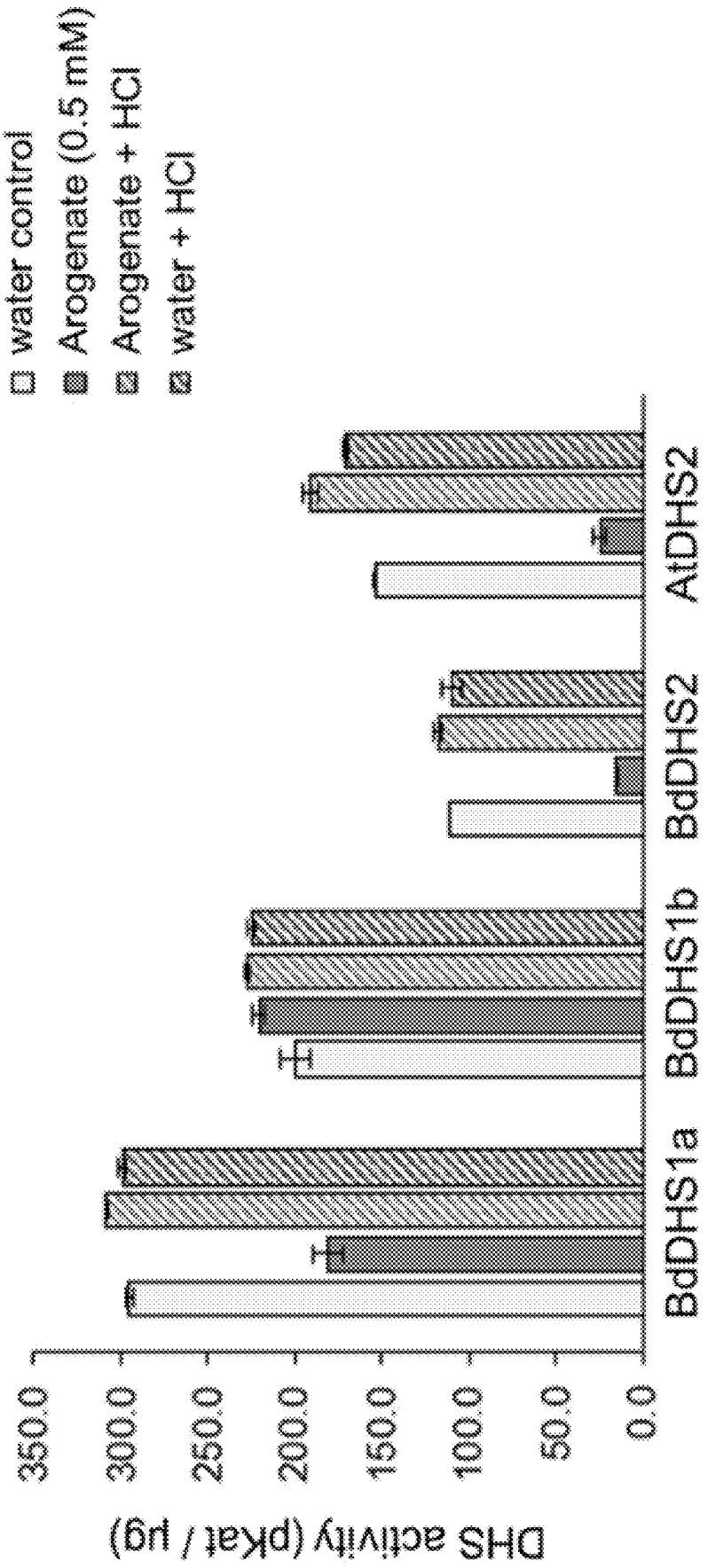

water control
Arogenate (0.5 mM)
Arogenate + HCl
water + HCl
DHS activity (pKat / µg)
350.0
300.0
250.0
200.0
150.0
100.0
50.0
0.0
BdDHS1a
BdDHS1b
BdDHS2
AtDHS2

FIG. 16B

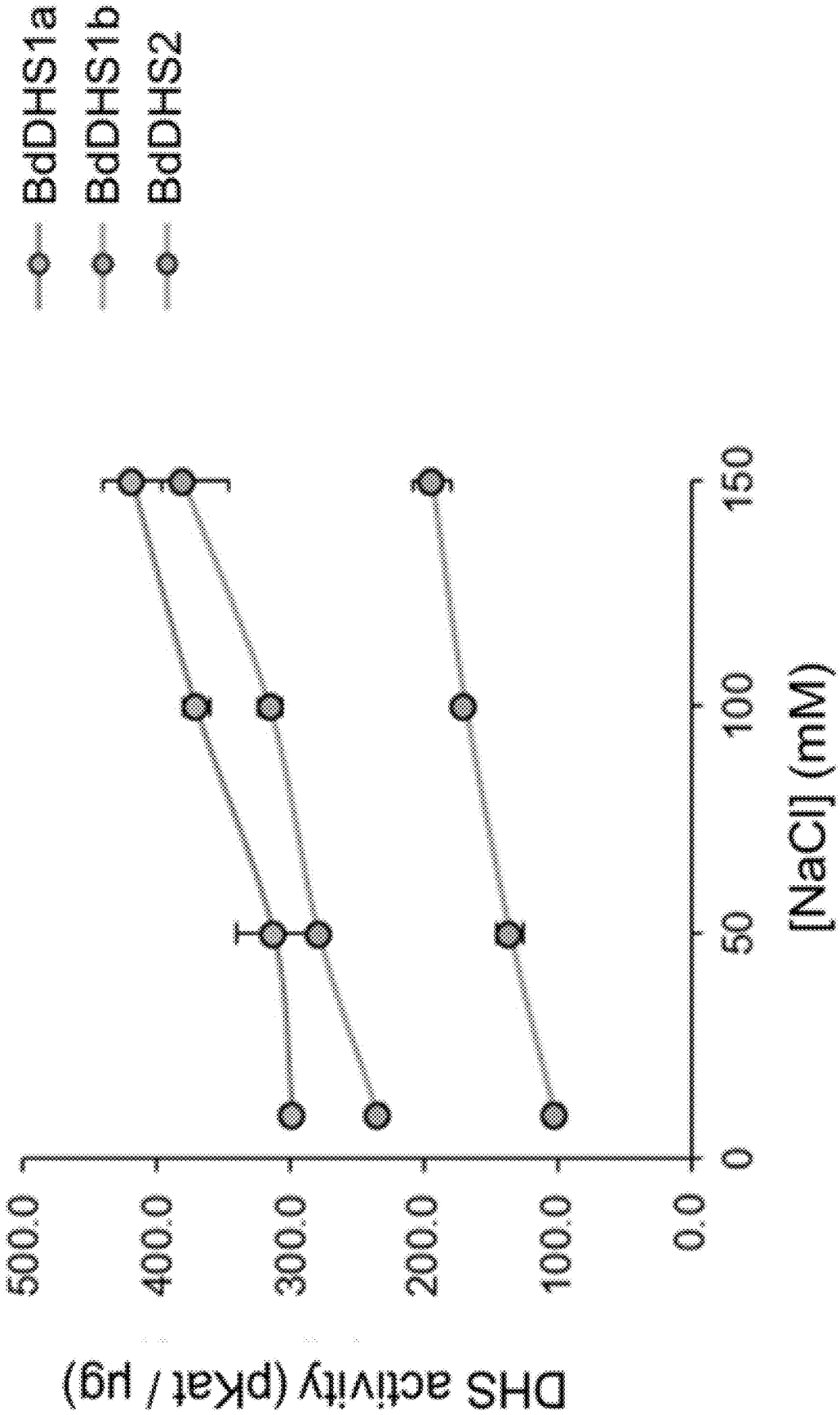

BdDHS1a
BdDHS1b
BdDHS2
DHS activity (pKat / µg)
500.0
400.0
300.0
200.0
100.0
0.0
0
50
100
150
[NaCl] (mM)

FIG. 17A

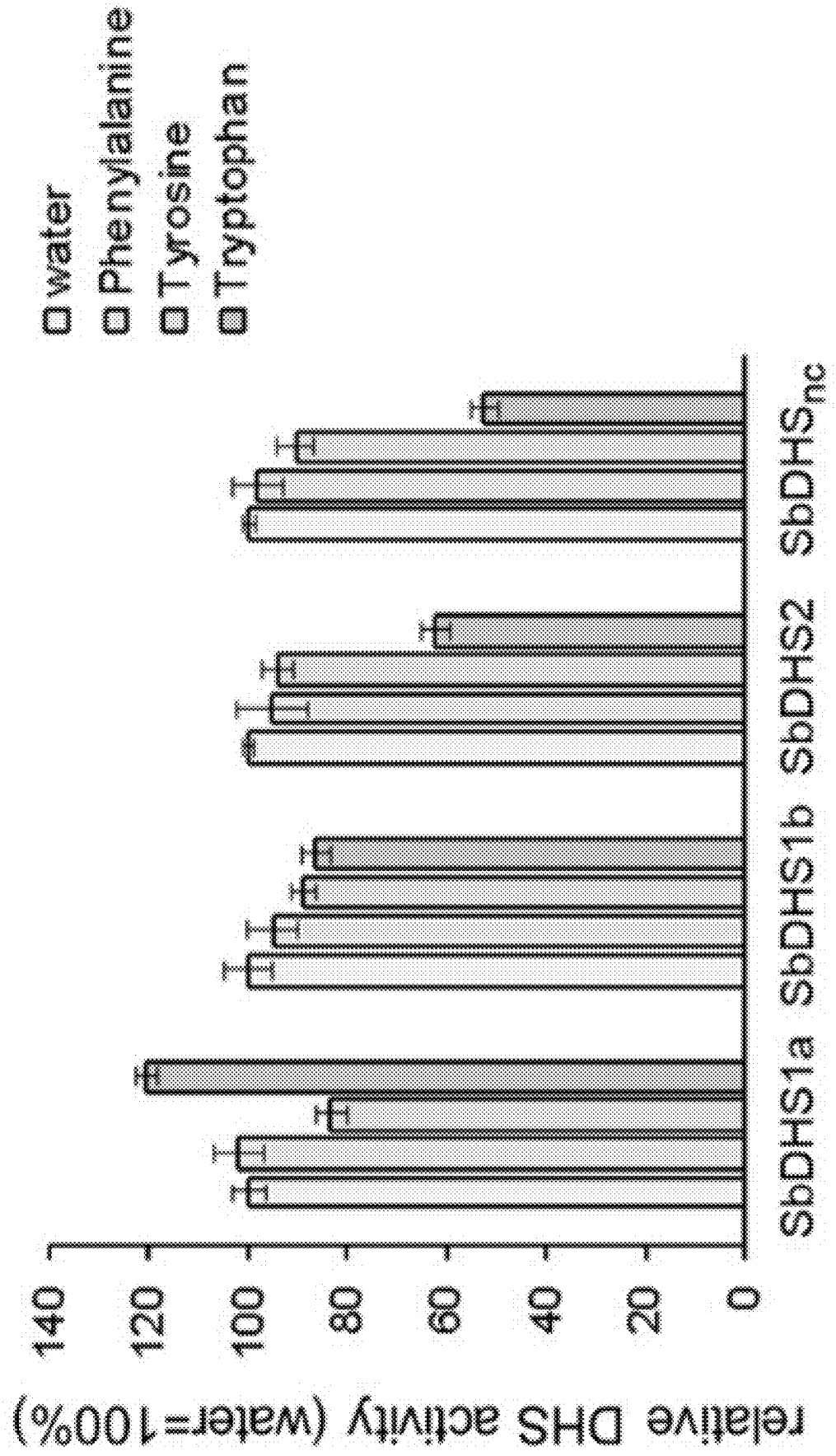

water
Phenylalanine
Tyrosine
Tryptophan
relative DHS activity (water=100%)
140
120
100
80
60
40
20
0
SbDHS1a
SbDHS1b
SbDHS2
SbDHSnc

FIG. 17B

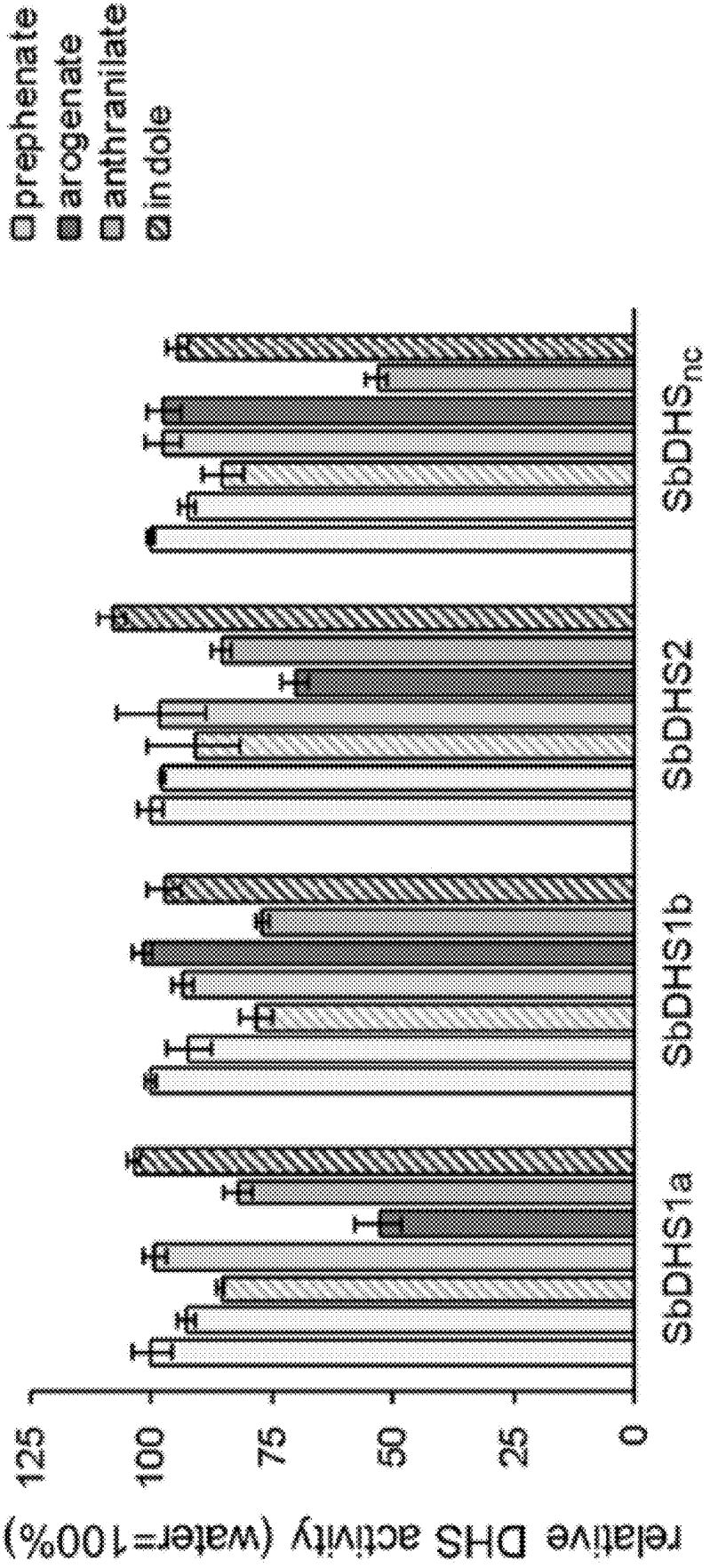

water
shikimate
chorismate
prephenate
arogenate
anthranilate
indole
relative DHS activity (water=100%)
125
100
75
50
25
0
SbDHS1a
SbDHS1b
SbDHS2
SbDHS$_{nc}$

FIG. 18B

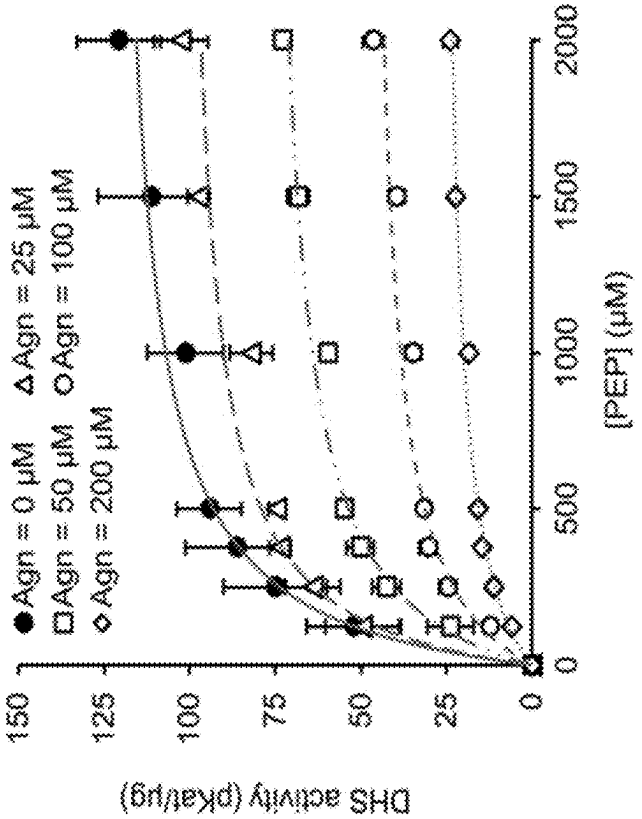


FIG. 18A

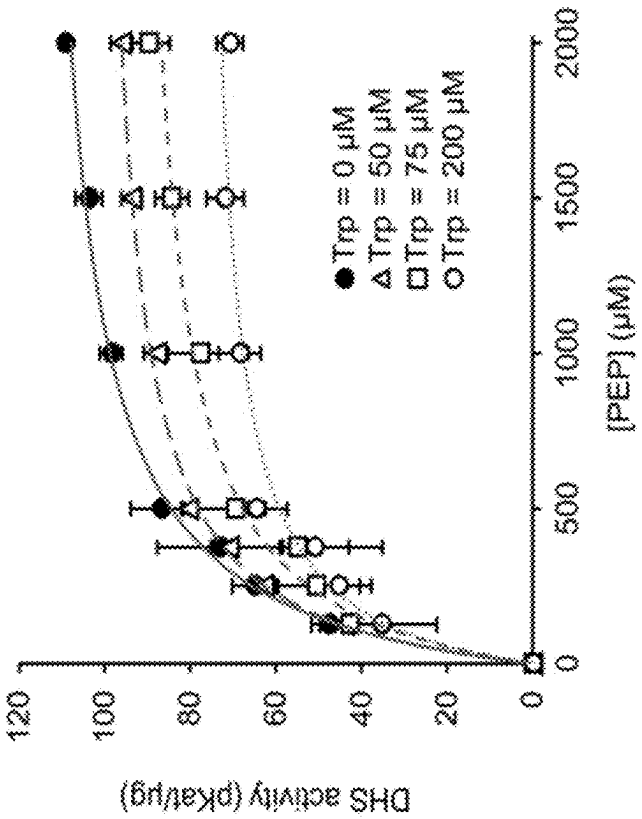


FIG. 18D

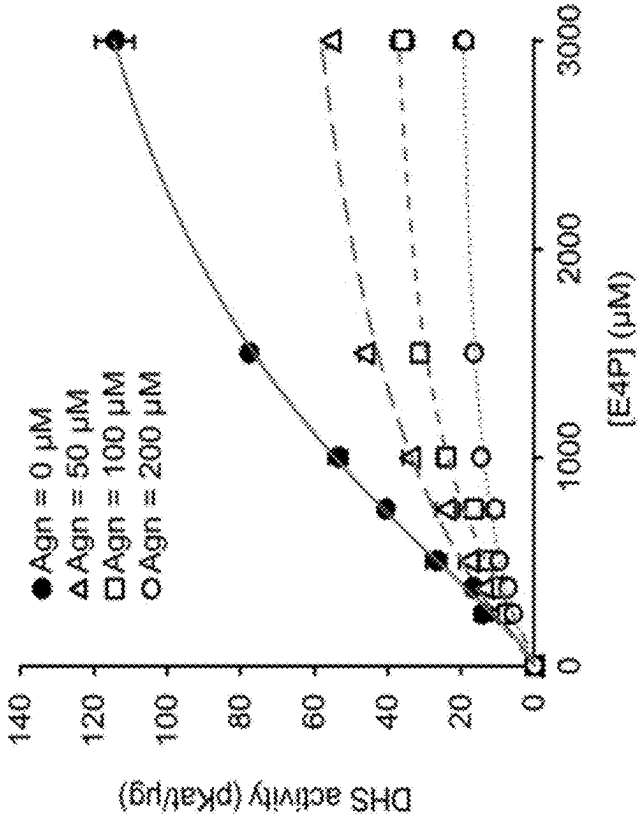


FIG. 18C

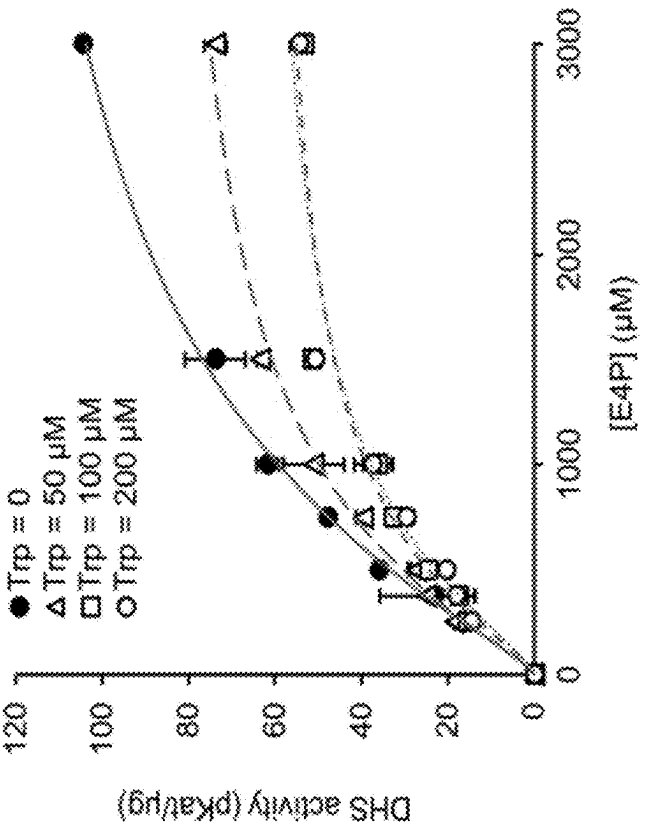


FIG. 18E

		Tryptophan (Trp)				Arogenate (Agn)				
		no Trp	50 μM	75 μM	200 μM	no Agn	25 μM	50 μM	100 μM	200 μM
Phosphoenol-pyruvate (PEP)	$V_{max\ app}$ (pKat/μg)	119 ± 2	103 ± 2	97 ± 1	80 ± 4	126 ± 11	104 ± 2	80 ± 1	49 ± 5	28 ± 3
	$K_{m\ app}$ (μM PEP)	210 ± 65	166 ± 7	236 ± 131	186 ± 56	185 ± 72	161 ± 37	245 ± 60	288 ± 55	390 ± 68
		no Trp	50 μM	100 μM	200 μM	no Agn	50 μM	100 μM	200 μM	
Erythrose 4-phosphate (E4P)	$V_{max\ app}$ (pKat/μg)	140 ± 9	90 ± 6	63 ± 6	66 ± 7	170 ± 30	69 ± 2	43 ± 5	22 ± 3	
	$K_{0.5\ app}$ (μM E4P)	1293 ± 155	840 ± 117	694 ± 139	791 ± 166	1755 ± 446	1012 ± 17	870 ± 191	589 ± 173	
	Hill Coefficient	1.2	1.3	1.3	1.3	1.4	1.4	1.5	1.3	

FIG. 19

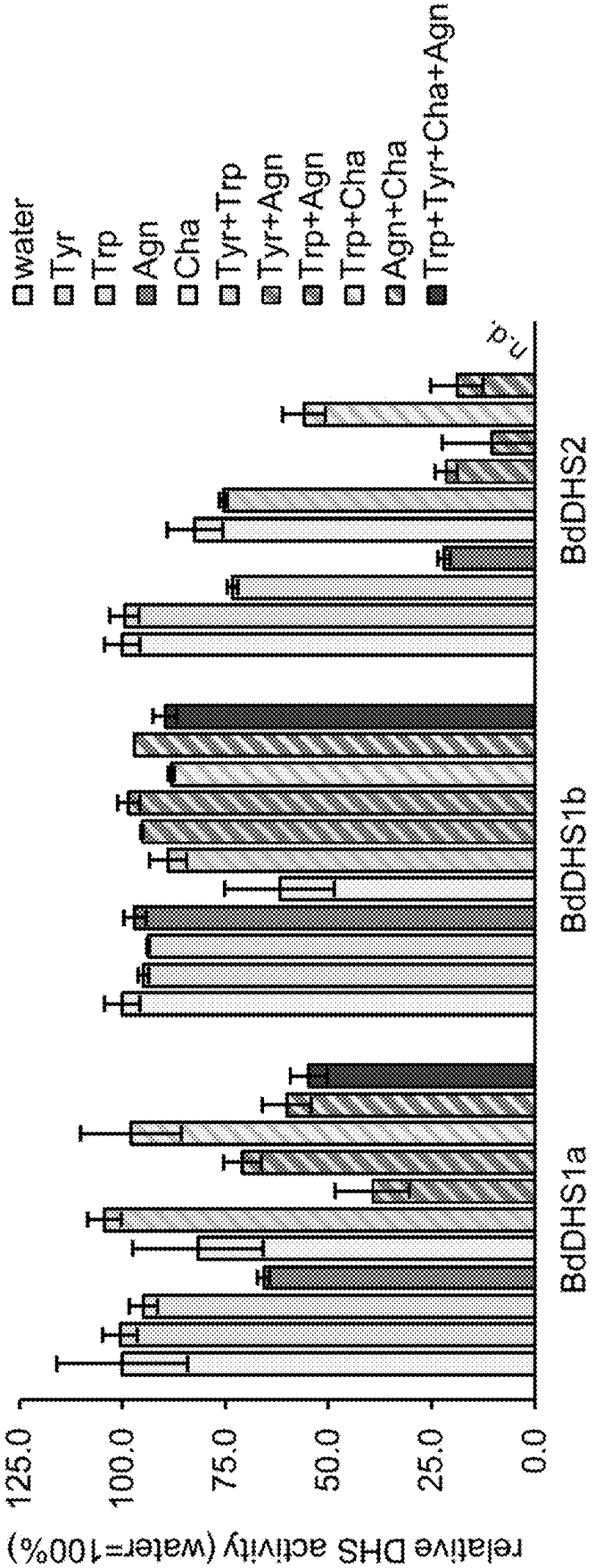
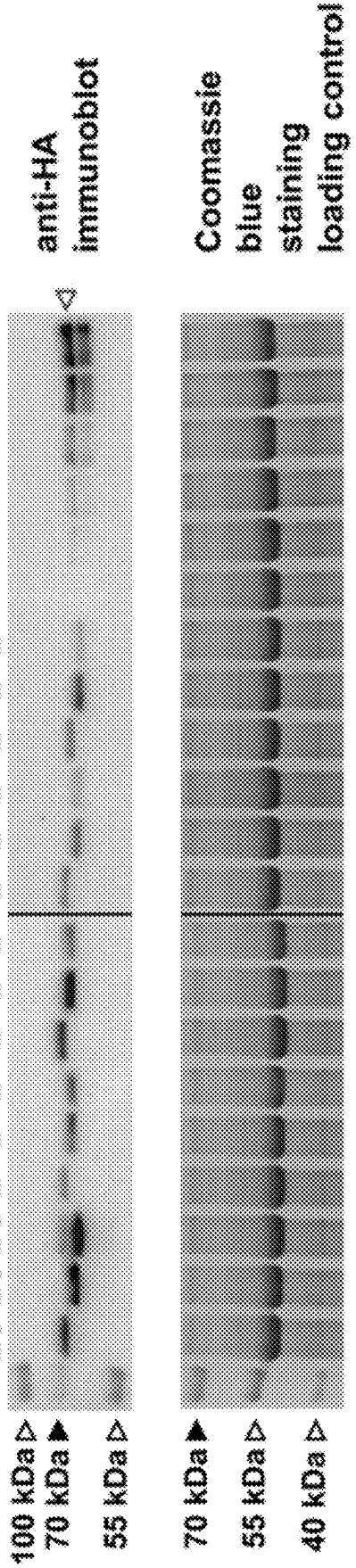

water
Tyr
Trp
Agn
Cha
Tyr+Trp
Tyr+Agn
Trp+Agn
Trp+Cha
Agn+Cha
Trp+Tyr+Cha+Agn
n.d.
BdDHS1a
BdDHS1b
BdDHS2
125.0
100.0
75.0
50.0
25.0
0.0
relative DHS activity (water=100%)

FIG. 20A


anti-HA immunoblot
Coomassie blue staining loading control
recombinant BdDHS1b-HA (ng/lane)
0
2
4
10
20
40
Plant 1
Plant 2
Plant 3
Plant 4
Plant 5
BdDHS1b (83.4 kDa)
BdDHS1a (83.4 kDa)
BdDHS2 (80.7 kDa)
BdDHS1b
BdDHS1a
BdDHS2
100 kDa
70 kDa
55 kDa
70 kDa
55 kDa
40 kDa

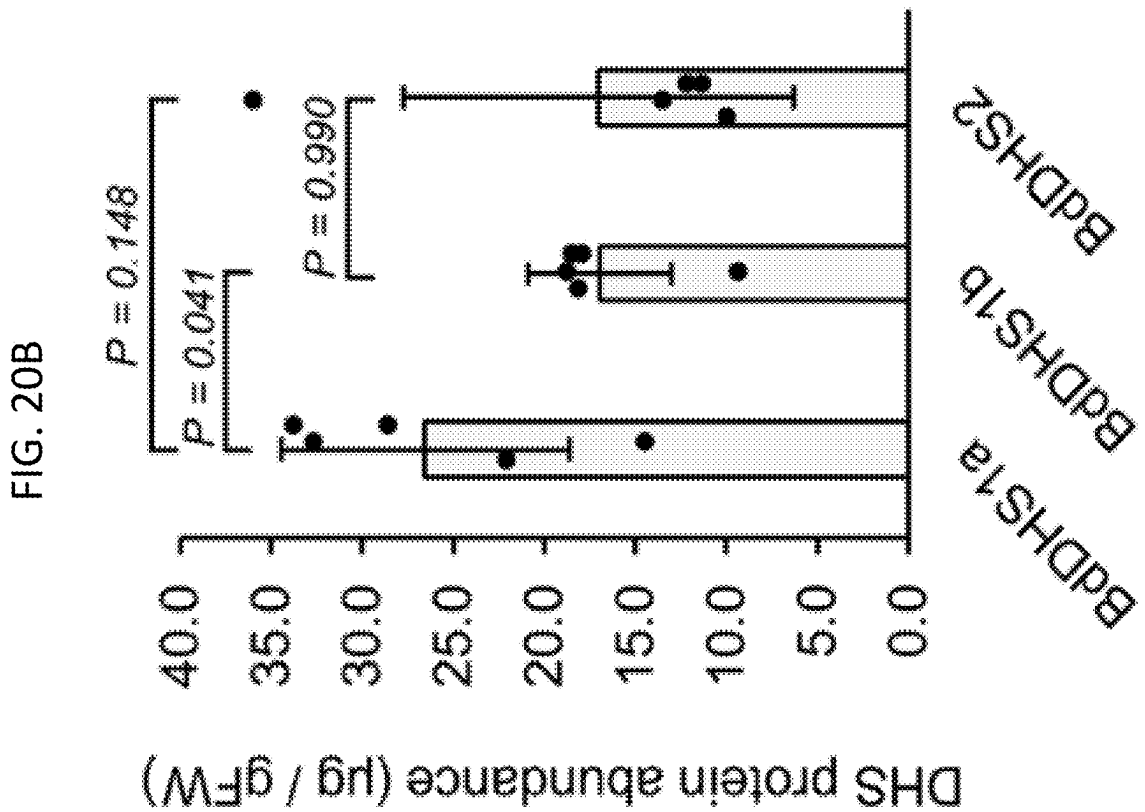

BdDHS2
BdDHS1b
BdDHS1a
P = 0.990
P = 0.148
P = 0.041
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
DHS protein abundance (µg / gFW)

FIG. 20B

FIG. 21

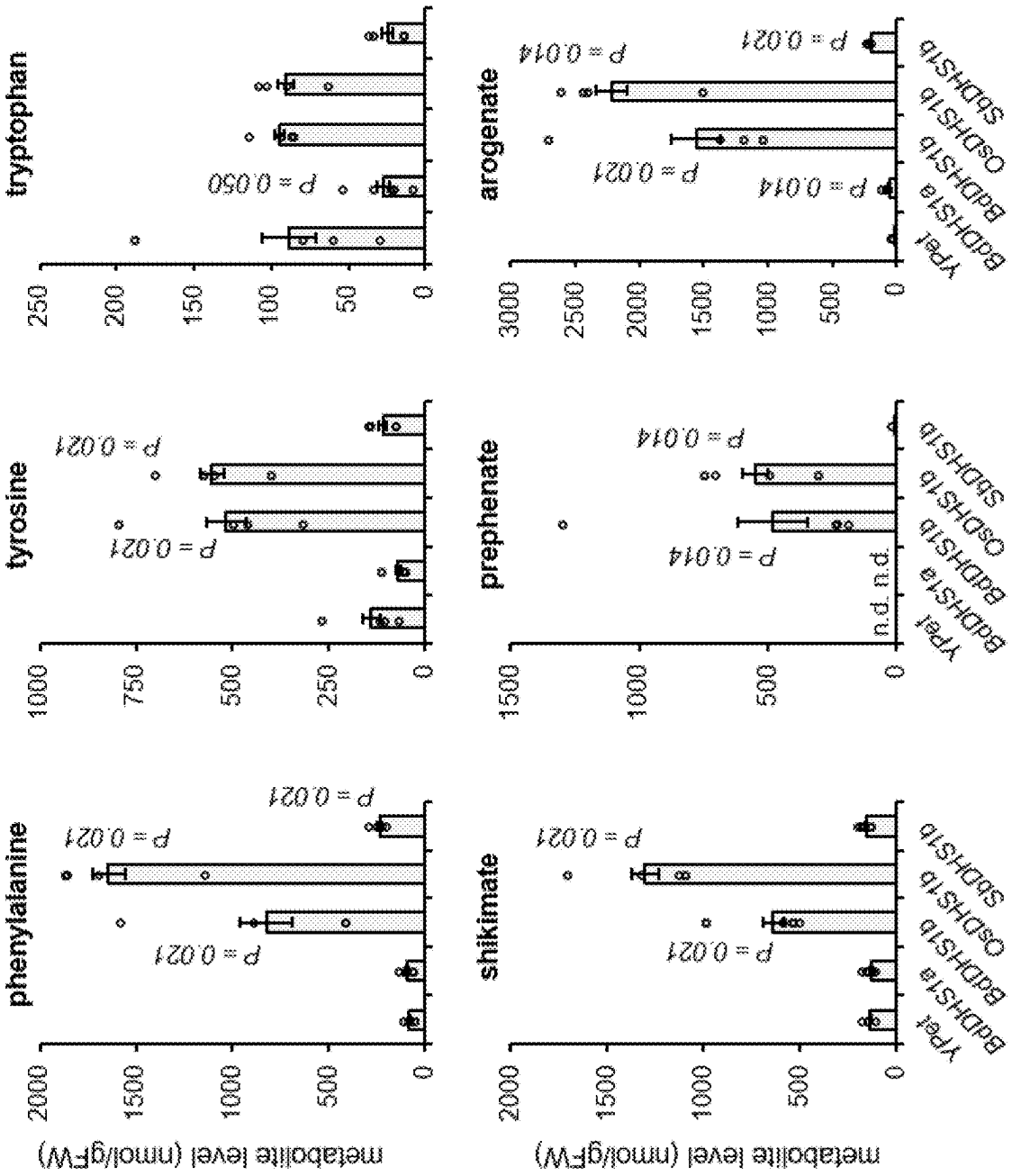
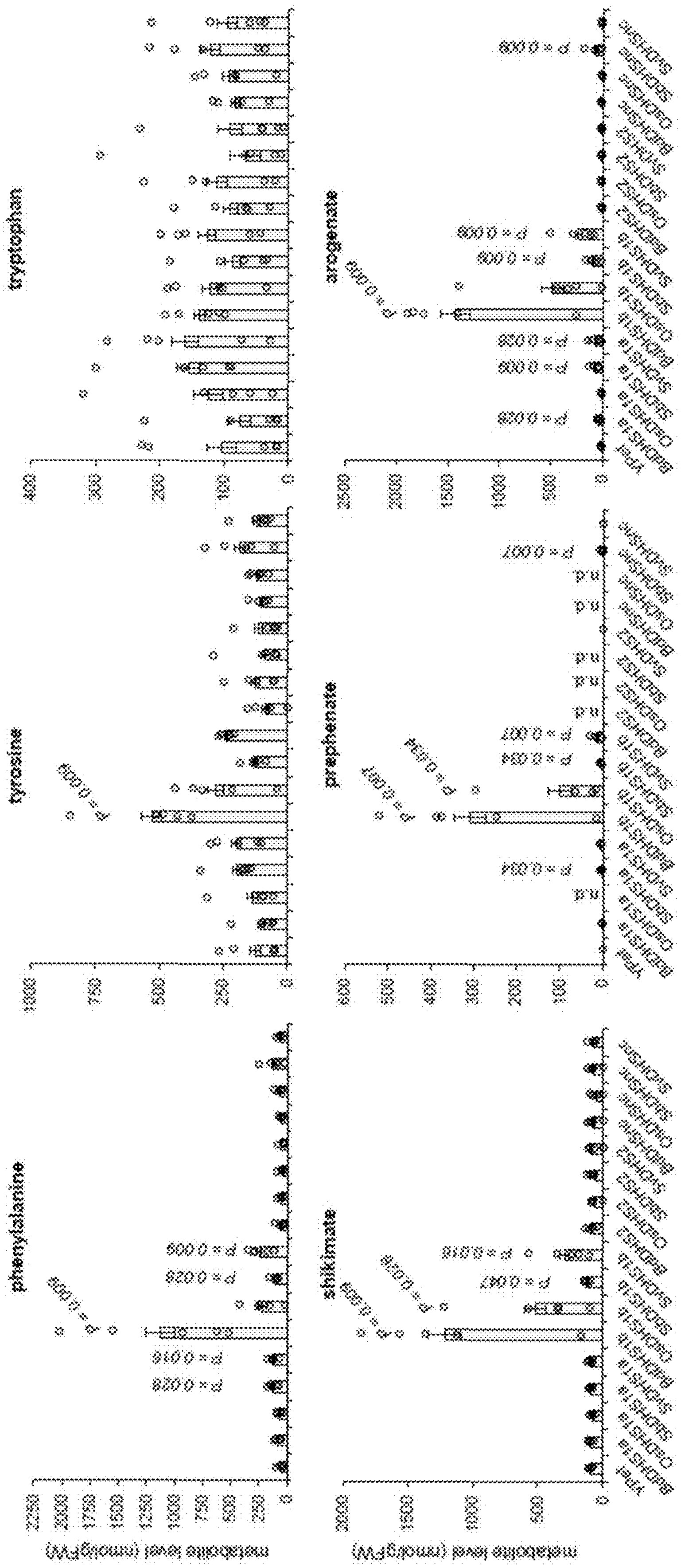

phenylalanine
tyrosine
prephenate
shikimate
arogenate
tryptophan
metabolite level (nmol/gFW)
P = 0.021
P = 0.050
P = 0.014
n.d.
VpeI
BdDHS1a
BdDHS1b
OsDHS1b
SbDHS1b

FIG. 22


phenylalanine
tyrosine
tryptophan
shikimate
prephenate
arogenate
metabolite level (nmol/gFW)
metabolite level (nmol/gFW)

DEREGULATED DHS AND TYRA ENZYMES OF GRASSES ENABLE EFFICIENT PRODUCTION OF BOTH TYROSINE AND PHENYLALANINE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 63/588,272 filed on Oct. 5, 2023, the contents of which are incorporated by reference in their entireties.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under 1836824 awarded by the National Science Foundation. The government has certain rights in the invention.

SEQUENCE LISTING

[0003] This application includes a sequence listing in XML format titled "960296_04546_ST26.xml", which is 129,604 bytes in size and was created on Sep. 16, 2024. The sequence listing is electronically submitted with this application via Patent Center and is incorporated herein by reference in its entirety.

BACKGROUND

[0004] Plants can directly convert atmospheric carbon dioxide (CO_2) into diverse aromatic natural products, which are primarily derived from the aromatic amino acids tyrosine, phenylalanine, and tryptophan. Aromatic compounds have unusual stability due to their aromaticity (i.e., electron delocalization). As a result, aromatic compounds have potential to be used as a carbon sink for reducing atmospheric CO_2. Aromatic compounds are also key precursors for pharmaceuticals, commodity chemicals, and industrial materials, for which there is rapidly growing global demand. However, the chemical conversion of CO_2 into aromatic compounds remains challenging, and fossil fuels remain the primary source of aromatic compounds. Thus, there remains a need in the art for improved methods for harvesting aromatic compounds from renewable sources, such as plants.

SUMMARY

[0005] In a first aspect, the present invention provides cells engineered to express or overexpress a deregulated enzyme selected from: (a) a DHS1b enzyme comprising SEQ ID NO: 3 (BdDHS1b), SEQ ID NO: 19 (OsDHS1b), SEQ ID NO: 27 (SvDHS1b) or a DHS1b enzyme having at least 95% identity to one of SEQ ID NOs: 3, 19 or 27; or (b) a noncanonical TyrA enzyme comprising SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 43 (SbTyrAnc) or a TyrA enzyme having at least 95% identity to SEQ ID NO: 37 or 43. In some embodiments, the cells are plant cells.

[0006] In a second aspect, the present invention provides plants comprising the engineered cells described herein.

[0007] In a third aspect, the present invention provides methods for increasing production of one or more aromatic amino acids in a cell. The methods comprise engineering the cell to express or overexpress an enzyme selected from: (a) a DHS1b enzyme comprising SEQ ID NO: 3 (BdDHS1b), SEQ ID NO: 19 (OsDHS1b), SEQ ID NO: 27 (SvDHS1b) or a DHS1b enzyme having at least 95% identity to one of SEQ ID NOs: 3, 19 or 27; or (b) a noncanonical TyrA enzyme comprising SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 43 (SbTyrAnc) or a TyrA enzyme having at least 95% identity to SEQ ID NO: 37 or 43. In some embodiments, the cell is engineered to express an enzyme that is not native to the cell. In other embodiments, the cell is engineered to overexpress an enzyme that is native to the cell as compared to a control cell. In some embodiments, the cell is a plant cell.

[0008] In a fourth aspect, the present invention provides methods for using the plants described herein to (1) produce aromatic amino acids or derivatives thereof, or (2) sequester CO_2. Both sets of methods comprise growing the plants described herein. The methods for producing aromatic amino acids or derivatives thereof further comprise purifying the aromatic amino acids or derivatives thereof produced by the plant.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIGS. **1**A-**1**E show results of a $^{13}CO_2$ feeding experiment that demonstrate that tyrosine is synthesized at high levels in *Brachypodium distachyon*. FIG. **1**A: Schematic representation of plant aromatic amino acid biosynthesis pathways and feedback regulation (red lines). The grass-specific, bifunctional enzyme phenylalanine tyrosine ammonia-lyase (PTAL, blue) constitutes a shortcut in the phenylpropanoid pathway, transforming tyrosine into p-coumaric acid in a single step. Dashed arrows indicate multiple enzymatic steps. Enzyme abbreviations: TyrA, arogenate dehydrogenase; ADT, arogenate dehydratase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 1, prephenate dehydrogenase $TyrA_p$ (EC 4.2.1.91), only found in legumes; 2, prephenate dehydratase, a side activity of plant ADT enzymes; 3, aromatic amino acid aminotransferase. FIG. **1**B: Schematic representation of the $^{13}CO_2$ feeding circuit. FIGS. **1**C-**1**E: Graphs showing the total content per gram of fresh weight (left panel), content of ^{13}C-labeled metabolite (central panel), and relative ^{13}C-labeled metabolite over the total content (right panel) for tyrosine (FIG. **1**C), phenylalanine (FIG. **1**D), and shikimate (FIG. **1**E). The levels in 4-week-old *Brachypodium distachyon* (turquoise, triangles) and *Arabidopsis thaliana* (magenta, circles) plants are compared. All data points are means±SD of n=2. *P<0.05, according to Student's t-test (two-sided test for two samples with equal variance).

[0010] FIGS. **2**A-**2**C show graphs comparing tyrosine (FIG. **2**A), phenylalanine (FIG. **2**B), and shikimate (FIG. **2**C) production in ^{13}C-labeled leaves and stems, and reveals that grass stems, unlike *Arabidopsis* stems, maintain a high rate of both tyrosine and phenylalanine production. Total and ^{13}C-labeled phenylalanine, tyrosine, and shikimate detected in leaves (pale green circles) and developing stems (dark blue squares) of 6-week-old *Brachypodium distachyon, Arabidopsis thaliana*, and *Setaria viridis* plants. All data points are means±SD of n=2. *P<0.05, according to Student's t-test (two-sided test for two samples with equal variance).

[0011] FIGS. **3**A-**3**F demonstrate that grass TyrA enzymes differ in their transcriptional and biochemical regulation. FIG. **3**A: Phylogenetic tree of plant TyrA enzymes (outgroup: *Chlamydomonas reinhardtii*) highlighting three clades of grass TyrA enzymes: TyrA1, TyrA2, and non-

canonical TyrA (TyrAnc). The enzymes highlighted in blue text were characterized in this study. Bootstrap test values (based on 1000 replications) below 50 have been omitted. The scale bar indicates the number of amino acid substitutions per site. Species abbreviations: AH, *Amaranthus hypochondriacus*; Aco, *Ananas comosus* v3; At, *Arabidopsis thaliana*; EL, *Beta vulgaris* EL10_1.0; Cre, *Chlamydomonas reinhardtii*; Bd, *Brachypodium distachyon*; HORVU, *Hordeum vulgare*; Mapoly, Marchantia *polymorpha*; Medtr, *Medicago truncatula; Oropetium, Oropetium thomaeum*, LOC_Os, *Oryza sativa*; Pavir, *Panicum virgatum*; Phvul, *Phaseolus vulgaris*; Potri, *Populus trichocarpa*; selmo, *Selaginella moellendorffii*; Sevir, *Setaria viridis; Solyc, Solanum lycopersicum*; Spov, *Spinacia oleracea*; Sb, *Sorghum bicolor*; Zm, *Zea mays*; Zosma, *Zostera marina*. FIG. **3**B: Heatmaps showing expression patterns of TyrA and PTAL genes in *Brachypodium distachyon* (top) and *Sorghum bicolor* (bottom) across different organs and developmental stages. The levels of expression are relative to the average abundance for each individual gene. Data were obtained from the databases PlaNet[41] and MOROKOSHI[42] for Brachypodium and Sorghum, respectively. FIG. **3**C: RT-qPCR results from an analysis of TyrA transcript abundance in young leaves and internodes from *Brachypodium distachyon* (left panel) and *Setaria viridis* (right panel). Error bars=SD; n=3; *P<0.05, **P<0.01, ***P<0.001, ns=not significant (P>0.05), according to Student's t-test (two-sided test for two samples with equal variance). FIGS. **3**D-**3**E: Graphs showing the in vitro sensitivity to feedback-inhibition by tyrosine of the recombinant TyrA enzymes of Brachypodium (FIG. **3**D) and Sorghum (FIG. **3**E) (error bars=SD; n=4 from two independent experiments). FIG. **3**F: Graph showing tyrosine content per gram of fresh weight at 72 hours following *Agrobacterium*-mediated transient expression of different TyrA genes in the leaves of *Nicotiana benthamiana. Arabidopsis* AtTyrA2, beet BvTyrAαi, and GUS were expressed as side-by-side controls. Letters indicate significant differences between treatments according to Student's t-test (P<0.05; two-sided test for two samples with equal variance). Error bars=SD; n=6.

[0012] FIGS. **4**A-**4**F demonstrate that grasses have a feedback-insensitive DHS1b enzyme that is expressed in internodes. FIG. **4**A: Phylogenetic tree of plant DHS enzymes (outgroup: Marchantia *polymorpha*). Enzymes highlighted in blue text were characterized in this study. Bootstrap test values (based on 1000 replications) below 50 have been omitted. Scale bar indicates number of amino acid substitutions per site. Additional species abbreviations that were not included in FIG. **3**A are: Cucsa, *Cucumis sativus*; Gorai, *Gossypium raimondii*, and GSMUA, *Musa acuminata*. FIG. **4**B: Heatmap showing expression patterns of Brachypodium DHS genes in different organs and developmental stages compared to BdTyrA1, BdPTAL, and other genes of the lignin pathway. Enzyme abbreviations: 4CL, 4-coumarate: CoA ligase; C3'H, 4-Coumarate 3-hydroxylase; and COMT, caffeic acid/5-hydroxyferulic acid O-methyltransferase. FIG. **4**C: Graph showing the inhibition of recombinant Brachypodium BdDHS1a, BdDHS1b, and BdDHS2 by 0.5 mM of aromatic amino acids (error bars=SD; n=3). FIG. **4**D: Graph showing the IC_{50} determination curve of tryptophan inhibition on the Brachypodium DHS enzymes. FIG. **4**E: Graph showing the effect of 0.5 mM of AAA pathway intermediates on Brachypodium DHS enzymes (error bars=SD; n=3). FIG. **4**F: Graph showing the IC_{50} determination curve of arogenate inhibition on the Brachypodium DHS enzymes. Asterisks indicate significant differences compared to the water control according to Student's t-test (P<0.01; two-sided test for two samples with equal variance).

[0013] FIGS. **5**A-**5**B show that co-expression of DHS1b and TyrA1 in *Nicotiana benthamiana* has a synergistic impact on tyrosine production while maintaining high phenylalanine accumulation. FIG. **5**A: Graphs showing the levels of phenylalanine, tyrosine, tryptophan, and their common intermediate shikimate, three days after the transient expression of Brachypodium DHS1a, DHS1b, and DHS2 in the leaves of *Nicotiana benthamiana* under control of a RuBisCO promoter. Asterisks indicate significant differences compared to the tdTomato (tdTom) negative control according to Student's t-test (P<0.05; two-sided test for two samples with equal variance). Error bars=SD; n=5. FIG. **5**B: Impact of the expression of BdDHS1b or BdDHS2 together with BdTyrA1, BdTyrA2, or BdTyrAnc on tyrosine (top) and phenylalanine (bottom) accumulation in *Nicotiana benthamiana* leaves. Level 2 Golden Gate vectors were assembled to ensure the simultaneous expression of the different DHS and TyrA genes. Letters indicate significant differences between treatments according to Student's t-test (P<0.05; two-sided test for two samples with equal variance). Error bars=SD; n=6.

[0014] FIGS. **6**A-**6**B show the ^{13}C-isotopologue abundance of phenylalanine (left), tyrosine (center) and shikimate (right) in *Arabidopsis thaliana* (FIG. **6**A) and *Brachypodium distachyon* (FIG. **6**B) as detected in the $^{13}CO_2$ feeding experiment described in FIG. **1**. Solid blue lines represent the concentration of unlabeled metabolite, dashed red-pink lines correspond to different ^{13}C isotopologues. The other possible isotopologues were undetectable or only detectable as a trace, so were not represented in these graphs. $1\times^{13}C$ isotopologues, which are naturally abundant, were not quantified for the study. Error bars=SD; n=2 for each individual time point.

[0015] FIGS. **7**A-**7**C show the ^{13}C-isotopologue abundance of phenylalanine (left), tyrosine (center) and shikimate (right) in leaves and stems of *Arabidopsis thaliana* (FIG. **7**A), *Brachypodium distachyon* (FIG. **7**B), and *Setaria viridis* (FIG. **7**C) as detected in the $^{13}CO_2$ feeding experiment described in FIG. **2**. Isotopologues with only one ^{13}C atom, which are naturally present at high abundance, were not represented. Error bars=SD; n=2 for each individual time point.

[0016] FIG. **8** shows transient expression of TyrA-EGFP fusion proteins in *Arabidopsis* protoplasts. Laser scan confocal microscopy of *Arabidopsis* protoplasts expressing the full-length CDS of *Brachypodium distachyon* BdTyrA enzymes fused to EGFP. AtTyrA2-EGFP was used as positive control for plastidial localization. Free EGFP was used as positive control for cytosolic localization.

[0017] FIGS. **9**A-**9**C are graphs showing the original Ct values and calibration curves for the measurement of TyrA expression in the leaves and developing internodes of the grasses *Brachypodium distachyon* and *Setaria viridis* by RT-qPCR. Ct values in *Brachypodium distachyon* (FIG. **9**A) and *Setaria viridis* (FIG. **9**B) (Error bars=SD; n=3). BdUBI10 (UBIQUITIN LIGASE 10) and SvUBI4 (UBIQUITIN LIGASE 4) were used as reference genes. Standard curves were generated for each individual TyrA amplicon (FIG. **9**C) (n=2).

[0018] FIGS. **10**A-**10**B are graphs showing the activity of pure recombinant TyrA enzymes from Brachypodium (FIG. **10**A) and Sorghum (FIG. **10**B) in the presence of different substrates (arogenate or prephenate) and electron acceptors (NAD^+ or $NADP^+$). Substrates and cofactors were tested at a concentration of 1 mM. Enzyme concentration was increased up to 10-times when using prephenate as a substrate or NAD^+ as an acceptor to increase the assay's sensitivity. Error bars=SD; n=3; n.d.=not detected.

[0019] FIGS. **11**A-**11**B show Michaelis-Menten and tyrosine inhibition plots for grass TyrA enzymes. FIG. **11**A: Michaelis-Menten plots corresponding to the kinetic parameters (K_m and V_{max}) shown in Table 1. FIG. **11**B: Tyrosine-inhibition plots used to calculate the IC_{50} values shown in Table 1. Individual points represent the average of 4 to 6 datapoints coming from at least two independent experiments conducted on different days using different batches of purified recombinant enzyme. Error bars=SD.

[0020] FIGS. **12**A-**12**C show inhibition kinetics for BdTyrA1 and BdTyrAnc. Michaelis-Menten (left panel) and Lineweaver-Burk (right panel) plots for BdTyrA1 (FIG. **12**A) and BdTyrAnc (FIG. **12**B), assayed at two alternative concentrations of tyrosine (legends are shown in the right panels). FIG. **12**C: Apparent K_m and V_{max} values, calculated from the data shown in FIG. **12**A and FIG. **12**B. Individual points represent the average of two technical replicates from the same experiment. Error bars=SD.

[0021] FIGS. **13**A-**13**C show schematic representations of the Golden-Gate plant expression constructs used in this study for transient expression of TyrA and DHS enzymes in *Nicotiana benthamiana*. FIG. **13**A: Level 2 vector for the expression of grass TyrA enzymes under the control of the CaMV 35S promoter. AtTyrA2, GUS, and BvTyrAα were used as controls. FIG. **13**B: Level 1 vector for the expression of Brachypodium DHS enzymes as HA-tagged proteins under the control of the *Arabidopsis* AtRbcS3B promoter. The fluorescent protein tdTomato was used as negative control. FIG. **13**C: Level 2 assemblies of TyrA-YPet and DHS-HA level 1 vectors, with their corresponding controls. The P19 repressor of the RNA silencing machinery was cloned in position 1 under control of the *Arabidopsis* ubiquitin ligase promoter, as it was found to be important for enhancing the expression level of BdDHS1b and BdDHS2. Yellow boxes represent Golden Gate overhangs used to assemble the constructs. Abbreviations: PhcTP, plastid transit peptide of *Petunia×hybrida* 5-enol-pyruvyl-shikimate-3-phosphate synthase (Della-Cioppa et al., 1986); RSSUcTP, plastid transit peptide of *Arabidopsis* RuBisCO Small Subunit; EU+NbHSP terminator was based on Diamos A G, and Mason H S (2018); NOS, *Agrobacterium tumefaciens* nopaline synthase; and OCS, *Agrobacterium tumefaciens* octopine synthase.

[0022] FIGS. **14**A-**14**C show Michaelis-Menten plots for Brachypodium DHS enzymes. Plots are shown for the substrates phosphoenolpyruvate (PEP, left) and erythrose 4-phosphate (E4P, right) for the enzymes BdDHS1a (FIG. **14**A), BdDHS1b (FIG. **14**B), and BdDHS2 (FIG. **14**C). Data points represent the average of at least two replicates from independent experiments using different preparations of purified recombinant enzyme. Error bars=SD.

[0023] FIGS. **15**A-**15**C are graphs showing the effect of AAAs and related compounds on the activity of recombinant Brachypodium DHS enzymes (validation of the results shown in FIGS. **4**C and **4**E). FIG. **15**A: Modulation of Brachypodium DHS enzyme activity by products of AAA catabolism and lignin biosynthesis intermediates, determined by measuring phosphoenolpyruvate consumption at 232 nm. FIGS. **15**B-**15**C: Effect of AAAs (FIG. **15**B) and AAA biosynthesis intermediates (FIG. **15**C) on Brachypodium DHS enzyme DAHP production, determined by UHPLC-MS. All compounds were tested at a fixed concentration of 0.5 mM. Error bars=SD; n=3.

[0024] FIGS. **16**A-**16**B show the results of DHS inhibition assays using hydrolyzed arogenate as control and the effect of NaCl concentration on DHS activity. FIG. **16**A: Arogenate that is hydrolyzed to phenylalanine via treatment with 1N HCl for 30 minutes (Zamir et al., 1980) does not inhibit Brachypodium DHS enzymes or *Arabidopsis thaliana* AtDHS2, which was also previously reported to be inhibited by arogenate (Yokoyama et al., 2021). FIG. **16**B: Increasing concentrations of NaCl result in increased activity of recombinant Brachypodium DHS enzymes. This is likely the cause of the increase in BdDHS1b activity observed when using high concentrations of arogenate, such as in FIG. **4**F. Error bars=SD; n=3.

[0025] FIGS. **17**A-**17**B show the effects of AAAs and related compounds on the activity of recombinant Sorghum DHS enzymes. FIG. **17**A: Effect of AAAs determined based on phosphoenolpyruvate consumption. FIG. **17**B: Effect of AAA intermediates determined based on phosphoenolpyruvate consumption. All compounds were tested at a fixed concentration of 0.5 mM. Error bars=SD; n=3.

[0026] FIGS. **18**A-**18**E show inhibition kinetics of BdDHS2 in the presence of tryptophan and arogenate. FIGS. **18**A-**18**D: Michaelis-Menten plots showing the activity of BdDHS2 in the presence of various concentrations of phosphoenolpyruvate (PEP) and tryptophan (FIG. **18**A), PEP and arogenate (FIG. **18**B), erythrose 4-phosphate (E4P) and tryptophan (FIG. **18**C), and E4P and arogenate (FIG. **18**D). FIG. **18**E: Apparent kinetic parameters calculated from the data shown in FIGS. **18**A-**18**D. Hill equation was used to fit E4P kinetic data. Individual points represent the average of at least two to three technical replicates from the same experiment. Error bars=SD.

[0027] FIG. **19** shows the effect of various combinations of the effectors tryptophan (Trp), tyrosine (Tyr), chorismate (Cha), and arogenate (Agn) on the activity of recombinant Brachypodium DHS enzymes. All effectors were tested at a concentration of 0.15 mM each. Error bars=SD; n=3; n.d. =not detected.

[0028] FIGS. **20**A-**20**B show an immunoblot quantification of the abundance of recombinant Brachypodium DHS enzymes (which comprised a C-terminal HA tag) in *Nicotiana benthamiana*. FIG. **20**A: Anti-HA immunoblot (top) and Coomassie blue staining of total protein as loading control (bottom) in samples from five independent plants. The expected sizes of the mature full-length proteins were: BdDHS1a, 62.4 kDa; BdDHS1b, 63.4 kDa; and BdDHS2, 60.7 kDa. A calibration curve for the HA tag quantification was generated using pure recombinant BdDHS1b-HA protein mixed with total protein extract from a non-infiltrated leaf of *Nicotiana benthamiana*. The arrow on the right indicates the expected size of the full-length BdDHS1b-HA protein. Bands of lower molecular weight in the BdDHS1b-HA standard were presumed to be degradation products and were not considered for quantitative purposes. All immunoblot images were non-saturated, and all shown membranes were exposed simultaneously. Vertical lines separate inde-

pendent membranes/gels. FIG. **20**B: Graph showing DHS protein abundance by gram of fresh weight of plant sample, based on anti-HA immunoblot signal. P values according to Student's t-test for two samples with equal variance. Error bars=SD; n=5.

[0029] FIG. **21** shows graphs that demonstrate that transient expression of BdDHS1b and OsDHS1b in planta results in increased accumulation of aromatic amino acids and their precursors. The levels of phenylalanine, tyrosine, tryptophan, their common intermediate shikimate, and the phenylalanine and tyrosine precursors prephenate and arogenate, were determined by LCMS in the leaves of *Nicotiana benthamiana* at three days post-infiltration with constructs encoding the *Brachypodium distachyon* enzymes BdDHS1a and BdDHS1b, the *Oryza sativa* enzyme OsDHS1b, and the *Sorghum bicolor* enzyme SbDHS1b under control of the RuBisCO S3B promoter from *Arabidopsis thaliana*. Note that different scales have been used for each individual graph. Data are presented as the average of n=4 individual plants; error bars=SE. Significant P values (<0.05) with respect to the YPet negative control, calculated using the Mann-Whitney U test, are shown next to the corresponding bar. n.d.=not detected.

[0030] FIG. **22** shows graphs that demonstrate that grass DHS enzymes other than BdDHS1b and OsDHS1b do not trigger in planta accumulation of aromatic amino acids. The levels of phenylalanine, tyrosine, tryptophan, shikimate, prephenate, and arogenate, were determined by LCMS in the leaves of *Nicotiana benthamiana* at three days post-infiltration with constructs encoding the four DHS isoforms (i.e., DHS1a, DHS1b, DHS2, and DHSnc) of *Brachypodium* dystachion (Bd), *Oryza sativa* (Os), *Sorghum bicolor* (Sb), and *Setaria viridis* (Sv) under control of the RuBisCO S3B promoter from *Arabidopsis thaliana*.[52] Note that different scales have been used for each individual graph. Data are presented as the average of n=5 individual plants; error bars=SE. Significant P values (<0.05) with respect to the YPet negative control, calculated using the Mann-Whitney U test, are shown above the corresponding bar. n.d.=not detected.

DETAILED DESCRIPTION

[0031] The present invention provides engineered cells and plants that express deregulated aromatic amino acid synthesis pathway enzymes from grasses. Methods for increasing the production of aromatic amino acids and their derivatives in cells and plants by engineering them to express these enzymes and methods for producing aromatic amino acids or derivatives thereof and/or sequestering carbon dioxide by growing the plants are also provided.

[0032] As is described in the Examples, the present inventors have identified 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DHS) and TyrA arogenate dehydrogenase (TyrA) enzymes in grasses that are naturally "deregulated," i.e., exhibit low sensitivity to feedback inhibition. They demonstrate that transiently expressing these deregulated enzymes in a non-grass plant (i.e., *Nicotiana benthamiana*) results in increased production of aromatic amino acids. Specifically, they show that expression of a DHS1b enzyme from *Brachypodium distachyon* (BdDHS1b; SEQ ID NO: 3), *Oryza sativa* (OsDHS1b; SEQ ID NO: 19), or *Setaria viridis* (SvDHS1b; SEQ ID NO: 27) increases the production of aromatic amino acids to varying degrees in *Nicotiana benthamiana* leaves (FIG. **5**A, FIG. **21**, FIG. **22**), and that expression of a nonconical TyrA enzyme from *Brachypodium distachyon* (BdTyrAnc; SEQ ID NO: 37) or *Sorghum bicolor* (SbTyrAnc; SEQ ID NO: 43) increases the production of tyrosine in *Nicotiana benthamiana* leaves (FIG. **3**F).

[0033] Expression of the deregulated enzymes described herein can be used to increase the production and accumulation of aromatic amino acids and their derivatives in plants. These products are valuable, and many are difficult to synthesize. Thus, increasing the levels of these products in plants being grown for pulp, paper, or biofuel production increases the value of the waste left following biomass extraction. Moreover, expression of the deregulated enzymes may enhance the ability of plants to pull carbon from the atmosphere to feed into the aromatic amino acid biosynthesis pathway and downstream pathways, resulting in increased carbon flow from carbon dioxide (CO_2) into diverse plant products, including phenylpropanoid compounds.

Engineered Cells:

[0034] In a first aspect, the present invention provides cells engineered to express or overexpress a deregulated enzyme selected from: (a) a DHS1b enzyme comprising SEQ ID NO: 3 (BdDHS1b), SEQ ID NO: 19 (OsDHS1b), SEQ ID NO: 27 (SvDHS1b) or a DHS1b enzyme having at least 95% identity to one of SEQ ID NOs: 3, 19 or 27; or (b) a noncanonical TyrA enzyme comprising SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 43 (SbTyrAnc) or a TyrA enzyme having at least 95% identity to SEQ ID NO: 37 or 43. Cells that express or overexpress combinations of two or more of these enzymes are also provided and may also be used herein.

[0035] A "cell" is a mass of cytoplasm that is bound externally by a cell membrane. This term encompasses both isolated single cells and cells that exists in cellular aggregates. The cells of the present invention may be from any organism and may be of any cell type. For example, the cells may be bacterial cells, fungal cells, archaeal cells, animal calls, or plant cells. However, in preferred embodiments, the cells are plant cells. Examples of suitable plant cells for use with the present invention include, without limitation, tomato plant cells, tobacco plant cells, soybean plant cells, cotton plant cells, poplar plant cells, sorghum plant cells, rice plant cells, corn plant cells, beet plant cells, mung bean plant cells, opium poppy plant cells, alfalfa plant cells, wheat plant cells, barley plant cells, millet plant cells, oat plant cells, rye plant cells, rapeseed plant cells, miscanthus plant cells, and grass plant cells.

[0036] The term "enzyme" is used to describe a biological catalyst (i.e., a substance that speeds up a chemical reaction). The enzymes of the present invention are proteins. The terms "protein," "polypeptide," and "peptide" are used interchangeably herein to refer to a series of amino acid residues connected by peptide bonds between the alpha-amino and carboxy groups of adjacent residues. Proteins may include modified amino acids and amino acid analogs. "Percentage of sequence similarity" or "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two

sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Protein and nucleic acid sequence identities are evaluated using the Basic Local Alignment Search Tool ("BLAST"), which is well known in the art (Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA* 87:2267-2268; Altschul et al., 1997, *Nucl. Acids Res.* 25:3389-3402). The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. The BLAST programs can be used with the default parameters or with modified parameters provided by the user. The term "substantial identity" of amino acid sequences for purposes of this invention normally means polypeptide sequence identity of at least 80%. Preferred percent identity of polypeptides can be any integer from 80% to 100%. Allelic differences in proteins are encompassed herein and thus the sequences provided include sequences with at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to those sequences provided herein.

[0037] The grass enzymes described herein are "deregulated" in that they exhibit reduced sensitivity to feedback inhibition by one or more effectors as compared to other isoforms of the enzyme found in the same grass plant and/or as compared to homologs of the enzyme found in non-grass plants. As used herein, the term "effector" refers to an organic molecule, other than the substrate of the reaction catalyzed by an enzyme, that can physically interact with the enzyme and interfere with its activity. For example, the inventors have demonstrated that the noncanonical TyrA enzymes from *Brachypodium distachyon* (BdTyrAnc; SEQ ID NO: 37) and *Sorghum bicolor* (SbTyrAnc; SEQ ID NO: 43) exhibit a lower sensitivity to inhibition by tyrosine as compared to the TyrA1 and TyrA2 isoforms found in the same plants, such that they remain active at concentrations of tyrosine that completely inactivate these other isoforms (FIG. **3**D and FIG. **3**E).

[0038] The cells of the present invention are "engineered," meaning that they have been altered by the hand of man. Specifically, the cells have been engineered to either express or overexpress a deregulated DHS or TyrA enzyme. In some embodiments, the cell is engineered to express an enzyme that is not native to the cell. In other embodiments, the cell is engineered to overexpress an enzyme that is native to the cell as compared to a control cell. As used herein, the term "native" is used to describe an enzyme that is naturally expressed by a cell. Conversely, an enzyme that is "not native" to a cell is an enzyme that is not naturally expressed by the cell.

[0039] A cell "overexpresses" an enzyme if it is artificially forced to express the enzyme at a higher level than the enzyme is expressed in a control cell or at a higher level than the enzyme would be expressed naturally in the absence of the genetic engineering or recombinant expression of the enzyme. As used herein, a "control cell" is a comparable cell (e.g., of the same species, cell type, and age) that developed under the same or substantially similar conditions but that was not engineered to express or overexpress a deregulated enzyme described herein.

[0040] In some embodiments, the cell is engineered to express or overexpress the deregulated enzyme via introduction of an exogenous nucleic acid (i.e., a nucleic acid that is not native to the cell) encoding the enzyme. The exogenous nucleic acid may either be inserted in the genome of the cell or may be present extrachromosomally (i.e., outside of the cell's chromosomes).

[0041] The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably to refer a polymer of DNA or RNA. A nucleic acid may be single-stranded or double-stranded and may represent the sense or the antisense strand. A nucleic acid may be synthesized or obtained from a natural source. A nucleic acid may contain natural, non-natural, or altered nucleotides, as well as natural, non-natural, or altered internucleotide linkages (e.g., phosphoroamidate linkages, phosphorothioate linkages). The term nucleic acid encompasses any form of DNA or RNA including, without limitation, constructs, vectors, plasmids, messenger RNA (mRNA), and viral RNA. Those of skill in the art understand the degeneracy of the genetic code and that a variety of nucleic acids can encode the same polypeptide. Examples of suitable nucleic acid sequences encoding deregulated enzymes described herein include SEQ ID NO: 4, which encodes the BdDHS1b enzyme of SEQ ID NO: 3; SEQ ID NO: 20, which encodes the OsDHS1b enzyme of SEQ ID NO: 19; SEQ ID NO: 28, which encodes the SvDHS1b enzyme of SEQ ID NO: 27; SEQ ID NO: 38, which encodes the BdTyrAnc enzyme of SEQ ID NO: 37; and SEQ ID NO: 44, which encodes the SbTyrAnc enzyme of SEQ ID NO: 43.

[0042] In some embodiments, the exogenous nucleic acid further comprises a promoter operably linked to the nucleic acid encoding the enzyme. As used herein, the term "promoter" refers to a DNA sequence that defines where transcription of a nucleic acid begins. RNA polymerase and the necessary transcription factors bind to the promoter to initiate transcription. Promoters are typically located directly upstream (i.e., at the 5' end) of the transcription start site. However, a promoter may also be located at the 3' end, within a coding region, or within an intron of a gene that it regulates. Promoters may be derived in their entirety from a native or heterologous gene, may be composed of elements derived from multiple regulatory sequences found in nature, or may comprise synthetic DNA. A promoter is "operably linked" to a nucleic acid if the promoter is positioned such that it can affect transcription of the nucleic acid.

[0043] The promoter used in the nucleic acids described herein may be a heterologous promoter (i.e., a promoter that is not naturally associated with the native gene encoding deregulated enzyme), an endogenous promoter (i.e., a promoter that is naturally associated with the native gene encoding deregulated enzyme), or a synthetic promoter that is designed to function in a desired manner in a particular host cell. Suitable promoters for use with the present invention include, but are not limited to, constitutive, inducible, temporally regulated, developmentally regulated, chemically regulated, tissue-preferred, and tissue-specific promoters. In some cases, it may be advantageous to use a tissue-specific promoter or a developmental stage-specific promoter to drive expression of the deregulated enzyme in a particular tissue of an organism or during a particular

developmental stage. For example, one may wish to drive expression of the deregulated enzyme in a plant tissue in which lignin deposition takes place, such as in growing stems.

[0044] In preferred embodiments, the cell is a plant cell and the promoter is a "plant promoter," i.e., a promoter that is active in plant cells. Suitable plant promoters include, without limitation, the 35S promoter of the cauliflower mosaic virus, the tCUP cryptic constitutive promoter, the Rsyn7 promoter, the maize In2-2 promoter, the maize ubiquitin promoter, the tobacco PR-1a promoter, the *Arabidopsis* RuBisCO S3B promoter, and the *Arabidopsis* ubiquitin ligase promoter.

[0045] In the Examples, the inventors demonstrate that expressing different combinations of grass DHS and TyrA enzymes in *Nicotiana benthamiana* leaves results in production of different levels of the aromatic amino acids tyrosine and phenylalanine. Specifically, they demonstrate that co-expressing the DHS1b enzyme from *Brachypodium distachyon* (BdDHS1b; SEQ ID NO: 3) with the nonconical TyrA enzymes from *Brachypodium distachyon* (BdTyrAnc; SEQ ID NO: 37) increases tyrosine production in *Nicotiana benthamiana* leaves by 500-fold and phenylalanine production by 8-fold compared to control leaves (i.e., *Nicotiana benthamiana* leaves that do not express non-native DHS and TyrA enzymes), whereas co-expressing BdDHS1b with the feedback-regulated TyrA1 enzyme from *Brachypodium distachyon* (BdTyrA1; SEQ ID NO: 33) increases tyrosine production by 230-fold and phenylalanine production by 32-fold compared to control leaves (FIG. **5**B). These data demonstrate that expression of different combinations of grass DHS and TyrA enzymes can be utilized to shift the flow through the aromatic amino acid synthesis pathway in different ways. Thus, in some embodiments, the cell is engineered to express or overexpress both a deregulated DHS1b enzyme and a specific TyrA enzyme. The TyrA enzyme may be any one of the three TyrA isoforms found in grasses, which are referred to herein as TyrA1, TyrA2, and non-canonical TyrA (TyrAnc). Specifically, the TyrA enzyme may be selected from any one of the TyrA enzymes tested in the Examples, namely SEQ ID NO: 33 (BdTyrA1), SEQ ID NO: 35 (BdTyrA2), SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 39 (SbTyrA1), SEQ ID NO: 41 (SbTyrA2), and SEQ ID NO: 43 (SbTyrAnc). Notably, TyrA enzymes having at least 95% or more sequence identity to those sequences provided herein may also be used.

[0046] The cells of the present invention may be further engineered to express one or more additional aromatic amino acid/phenylpropanoid biosynthesis pathway enzymes. Examples of such enzymes include the prephenate and arogenate TyrA dehydrogenases and engineered PAL enzymes described in U.S. Patent Publication US2015/0150157, U.S. Patent Publication US2018/0265880, U.S. Patent Publication US2018/0216083, and U.S. patent application Ser. No. 18/611,181, and the engineered DHS enzymes described in International Patent Application Publication No. WO2023108018, the contents of which are each incorporated by reference in their entireties.

Plants:

[0047] In a second aspect, the present invention provides plants comprising the engineered cells described herein. The term "plant" is used broadly herein to refer to a plant at any stage of development or to part of a plant, including a plant cutting, a plant cell culture, a plant organ, a plant tissue, a plant seed, or a plantlet. Particularly useful parts of a plant include harvestable parts and parts that can be used for propagation of progeny plants. A harvestable part of a plant can be any useful part of a plant, for example, flowers, pollen, seedlings, tubers, leaves, stems, fruit, seeds, roots, and the like. A part of a plant useful for propagation includes, for example, seeds, fruits, cuttings, seedlings, tubers, rootstocks, cells, callus and the like.

[0048] The plants of the present invention may comprise a single engineered plant cell, comprise a plurality of engineered plant cells, or consist entirely of engineered plant cells. For example, the plants may comprise a specific organ or tissue that comprises engineered plant cells.

[0049] In some embodiments, the plant is a tomato plant, tobacco plant, soybean plant, cotton plant, poplar plant, sorghum plant, corn plant, beet plant, mung bean plant, opium poppy plant, alfalfa plant, wheat plant, barley plant, millet plant, oat plant, rye plant, rapeseed plant, miscanthus plant, or grass plant. In some embodiments, the deregulated enzyme expressed by the plant cell is native to the plant. In these embodiments, the plant may be a grass plant selected from a *Sorghum bicolor* plant, an *Oryza sativa* plant, and a *Brachypodium distachyon* plant.

[0050] In the Examples, the inventors demonstrate that engineering plants to express the deregulated DHS and TyrA enzymes described herein increases their production of one or more aromatic amino acids. Thus, in some embodiments, the quantity of aromatic amino acids (i.e., tyrosine, phenylalanine, and/or tryptophan) or derivatives thereof produced by the plant is greater than the quantity produced by a control plant. Increased plant production of aromatic amino acids results in increased accumulation of aromatic amino acids in plant tissues, and the terms "increased production" and "increased accumulation" are therefore used interchangeably herein. Thus, in some embodiments, the quantity of aromatic amino acids or derivatives thereof accumulated in the plant is greater than the quantity accumulated in a control plant. Suitably, the plant produces/accumulates at least one aromatic amino acid in a quantity that is at least 1.5-, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11-, 12-, 13-, 14-, 15-, 16-, 17-, 18-, 19-, or 20-fold higher as compared to the quantity produced/accumulated by a control plant. The plant may exhibit increased accumulation of these products in any tissue. Namely, the plant may exhibit increased accumulation of these products in either a tissue that comprises one or more engineered cells described herein or a tissue that does not comprise engineered cells (i.e., via export of products from the engineered cells). Production/accumulation of aromatic amino acids may be measured, for example, using stable isotope tracing (e.g., $^{13}CO_2$ labeling) followed by quantification via gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), or nuclear magnetic resonance (NMR).

[0051] As used herein, the term "control plant" refers to a comparable plant (e.g., of the same species, cultivar, and age) that was raised under the same or substantially similar conditions but that does not comprise an engineered cell described herein (i.e., a cell that expresses or overexpresses a deregulated enzyme selected from BdDHS1b, OsDHS1b, SvDHS1b, BdTyrAnc, and SbTyrAnc). Plants that are grown in "substantially similar conditions" are grown in similar locations and soil conditions, are planted with similar timing, are subjected to similar abiotic stresses, and the like.

[0052] In some embodiments, the plant assimilates a greater quantity of CO_2 or assimilates CO_2 at a greater rate as compared to a control plant. Suitably, the phenylpropanoid compound or CO_2 assimilation of the plant is at least 2%, 5%, 10%, 20%, 30%, 40%, 50%, or 60% greater than that of a control plant. CO_2 assimilation may be quantified by measuring the gas exchange activity of the plant. For example, CO_2 assimilation may be measured using an LI-6400XT photosynthesis system equipped with the 6400-40 leaf chamber (LI-COR). Alternatively, labeled $^{13}CO_2$ can be fed to plants and the rate of ^{13}C incorporation into plants can be measured over time.

Methods for Increasing Production of Aromatic Amino Acids and Derivatives Thereof:

[0053] In a third aspect, the present invention provides methods for increasing production of one or more aromatic amino acid or derivative thereof in a cell. The methods comprise engineering the cell to express or overexpress an enzyme selected from: (a) a DHS1b enzyme comprising SEQ ID NO: 3 (BdDHS1b), SEQ ID NO: 19 (OsDHS1b), SEQ ID NO: 27 (SvDHS1b) or sequences having at least 95% identity thereto; or (b) a noncanonical TyrA enzyme comprising SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 43 (SbTyrAnc) or sequences having at least 95% sequence identity thereto. In some embodiments, the cell is engineered to express an enzyme that is not native to the cell. In other embodiments, the cell is engineered to overexpress an enzyme that is native to the cell as compared to a control cell. In preferred embodiments, the cell is a plant cell.

[0054] In some embodiments, the cell is engineered by introducing an exogenous nucleic acid encoding the enzyme into the cell. Suitable methods for introducing nucleic acids into cells include, without limitation, *Agrobacterium*-mediated transformation, the floral dip method, bacteriophage or viral infection, electroporation, heat shock, lipofection, microinjection, and particle bombardment. In these embodiments, the exogenous nucleic acid may either be inserted into the genome of the cell or remain extrachromosomal after it has been introduced into the cell. Insertion into the genome may be random or targeted to a specific locus (e.g., via homologous recombination). Further, in these embodiments, the exogenous nucleic acid may comprise a promoter operably linked to the nucleic acid encoding the enzyme, as described above.

[0055] In some embodiments, the cell is engineered via genome editing. In cells in which the deregulated enzyme is not native, the genome may be engineered to insert a copy of a DNA sequence encoding the deregulated enzyme or to replace a sequence encoding a regulated homolog of the deregulated enzyme with a sequence encoding the deregulated enzyme. In cells in which the deregulated enzyme is native, the genome may be engineered to introduce one or more additional copies of a DNA sequence encoding the deregulated enzyme or to modify a gene regulatory element (e.g., a promoter, an enhancer) associated with the native gene encoding the deregulated enzyme to increase its expression. Genome editing may involve use of an engineered nuclease (e.g., a meganuclease, zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), or CRISPR-Cas nuclease).

[0056] As is described above, the inventors have demonstrated that expressing different combinations of grass DHS and TyrA enzymes in *Nicotiana benthamiana* leaves results in production of different levels of the aromatic amino acids tyrosine and phenylalanine. Thus, in some embodiments, the cell is engineered to express or overexpress a specific combination of a deregulated DHS1b enzyme and a TyrA enzyme. The TyrA enzyme may be any one of the three TyrA isoforms found in grasses (i.e., TyrA1, TyrA2, and TyrAnc).

[0057] In some embodiments, the cell is part of a plant, and the methods comprise engineering one or more cells of the plant to express or overexpress the deregulated enzyme. In these embodiments, a single cell, a plurality of cells (e.g., a specific organ or tissue), or all the cells of the plant may be engineered.

[0058] In some embodiments, the methods further comprise purifying aromatic amino acids or derivatives thereof from the cell. As used herein, the term "purifying" refers to the process of separating a desired product from other cellular components and impurities. Suitable methods for purifying aromatic amino acids and derivatives thereof include, without limitation, high performance liquid chromatography (HPLC) and other chromatographic techniques, such as affinity chromatography. A "purified" product may be at least 85% pure, at least 95% pure, or at least 99% pure.

Methods for Using Plants:

[0059] In a fourth aspect, the present invention provides methods for using the plants described herein to (1) produce aromatic amino acids or derivatives thereof, or (2) sequester CO_2. Both sets of methods comprise growing the plants described herein. The methods for producing aromatic amino acids or derivatives thereof further comprise purifying the aromatic amino acids or derivatives thereof produced by the plant.

[0060] As used herein, a "derivative" of an aromatic amino acid is any cellular product that is produced using an aromatic amino acid or that incorporates an aromatic amino acid. Examples of aromatic amino acid derivatives that could be produced using the methods of the present invention include the tyrosine derivatives homogentisate (HGA), α-tocopherols, and γ-tocopherols. Suitable products made from overexpression of the enzymes as described herein may include, without limitation, vitamin E, plastoquinone, a cyanogenic glycoside, a benzylisoquinoline alkaloid, rosmarinic acid, betalains, suberin, mescaline, morphine, salidroside, a phenylpropanoid compound, dhurrin, a tocochromanol, ubiquinone, lignin, a catecholamine such as epinephrine (adrenaline) or dopamine (i.e., L-dihydroxyphenylalanine (L-DOPA)), melanin, an isoquinoline alkaloid, hydroxycinnamic acid amide (HCAA), an amaryllidaceae alkaloid, hordenine, hydroxycinnamate, hydroxylstyrene, phenylethanol, phenyllactate, phenylacetic acid, mandelic acid, or tyrosine. Phenylpropanoid compounds (e.g., lignin, tannins, flavonoids, stilbene, resveratrol, lignans) may be produced from tyrosine.

[0061] "Carbon sequestration" is a process in which atmospheric CO_2 is captured and stored. It is one method for reducing the amount of CO_2 in the atmosphere (i.e., to reduce global climate change). In some embodiments, the methods further comprise harvesting part of the plant while leaving the roots of the plant in the soil such that the carbon contained in the roots is sequestered therein. Harvestable parts of plants include, without limitation, flowers, pollen, seedlings, tubers, leaves, stems, fruit, seeds, roots, cuttings, and the like. Above ground tissues that are enriched for aromatic compounds will be decomposed slowly by soil

microbes, which also enhances carbon sequestration. The harvested plant materials can be also converted via pyrolysis to biochar, which can substantially extend the retention of organic molecules and carbon sequestration.

Enzyme Sequences:

TABLE 4

Grass DHS enzyme sequences

Enzyme	Organism	Protein sequence	DNA (CDS) sequence
BdDHS1a	*Brachypodium distachyon*	SEQ ID NO: 1	SEQ ID NO: 2
BdDHS1b	*Brachypodium distachyon*	SEQ ID NO: 3	SEQ ID NO: 4
BdDHS2	*Brachypodium distachyon*	SEQ ID NO: 5	SEQ ID NO: 6
BdDHSnc	*Brachypodium distachyon*	SEQ ID NO: 7	SEQ ID NO: 8
SbDHS1a	*Sorghum bicolor*	SEQ ID NO: 9	SEQ ID NO: 10
SbDHS1b	*Sorghum bicolor*	SEQ ID NO: 11	SEQ ID NO: 12
SbDHS2	*Sorghum bicolor*	SEQ ID NO: 13	SEQ ID NO: 14
SbDHSnc	*Sorghum bicolor*	SEQ ID NO: 15	SEQ ID NO: 16
OsDHS1a	*Oryza sativa*	SEQ ID NO: 17	SEQ ID NO: 18
OsDHS1b	*Oryza sativa*	SEQ ID NO: 19	SEQ ID NO: 20
OsDHS2	*Oryza sativa*	SEQ ID NO: 21	SEQ ID NO: 22
OsDHSnc	*Oryza sativa*	SEQ ID NO: 23	SEQ ID NO: 24
SvDHS1a	*Setaria viridis*	SEQ ID NO: 25	SEQ ID NO: 26
SvDHS1b	*Setaria viridis*	SEQ ID NO: 27	SEQ ID NO: 28
SvDHS2	*Setaria viridis*	SEQ ID NO: 29	SEQ ID NO: 30
SvDHSnc	*Setaria viridis*	SEQ ID NO: 31	SEQ ID NO: 32

TABLE 5

Grass TyrA enzyme sequences

Enzyme	Organism	Protein sequence	DNA (CDS) sequence
BdTyrA1	*Brachypodium distachyon*	SEQ ID NO: 33	SEQ ID NO: 34
BdTyrA2	*Brachypodium distachyon*	SEQ ID NO: 35	SEQ ID NO: 36
BdTyrAnc	*Brachypodium distachyon*	SEQ ID NO: 37	SEQ ID NO: 38
SbTyrA1	*Sorghum bicolor*	SEQ ID NO: 39	SEQ ID NO: 40
SbTyrA2	*Sorghum bicolor*	SEQ ID NO: 41	SEQ ID NO: 42
SbTyrAnc	*Sorghum bicolor*	SEQ ID NO: 43	SEQ ID NO: 44

[0062] The present disclosure is not limited to the specific details of construction, arrangement of components, or method steps set forth herein. The compositions and methods disclosed herein are capable of being made, practiced, used, carried out and/or formed in various ways that will be apparent to one of skill in the art in light of the disclosure that follows. The phraseology and terminology used herein is for the purpose of description only and should not be regarded as limiting to the scope of the claims. Ordinal indicators, such as first, second, and third, as used in the description and the claims to refer to various structures or method steps, are not meant to be construed to indicate any specific structures or steps, or any particular order or configuration to such structures or steps. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples or exemplary language (e.g., "such as") provided herein, is intended merely to facilitate the disclosure and does not imply any limitation on the scope of the disclosure unless otherwise claimed. No language in the specification, and no structures shown in the drawings, should be construed as indicating that any non-claimed element is essential to the practice of the disclosed subject matter. The use herein of the terms "including," "comprising," or "having," and variations thereof, is meant to encompass the elements listed thereafter and equivalents thereof, as well as additional elements. Embodiments recited as "including," "comprising," or "having" certain elements are also contemplated as "consisting essentially of" and "consisting of" those certain elements.

[0063] Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. For example, if a concentration range is stated as 1% to 50%, it is intended that values such as 2% to 40%, 10% to 30%, or 1% to 3%, etc., are expressly enumerated in this specification. These are only examples of what is specifically intended, and all possible combinations of numerical values between and including the lowest value and the highest value enumerated are to be considered to be expressly stated in this disclosure. Use of the word "about" to describe a particular recited amount or range of amounts is meant to indicate that values very near to the recited amount are included in that amount, such as values that could or naturally would be accounted for due to manufacturing tolerances, instrument and human error in forming measurements, and the like. All percentages referring to amounts are by weight unless indicated otherwise.

[0064] No admission is made that any reference, including any non-patent or patent document cited in this specification, constitutes prior art. In particular, it will be understood that, unless otherwise stated, reference to any document

herein does not constitute an admission that any of these documents forms part of the common general knowledge in the art in the United States or in any other country. Any discussion of the references states what their authors assert, and the applicant reserves the right to challenge the accuracy and pertinence of any of the documents cited herein. All references cited herein are fully incorporated by reference unless explicitly indicated otherwise. The present disclosure shall control in the event there are any disparities between any definitions and/or descriptions found in the cited references.

[0065] The following examples are meant only to be illustrative and are not meant as limitations on the scope of the invention or of the appended claims.

EXAMPLES

Example 1

[0066] In the following example, the inventors describe the identification of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DHS) and TyrA enzymes in grasses that are naturally deregulated. They demonstrate that transiently expressing these deregulated enzymes in a non-grass plant (i.e., *Nicotiana benthamiana*) results in increased synthesis of aromatic amino acids. Specifically, they show that expression of a 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1b (DHS1b) enzyme from *Brachypodium distachyon* (BdDHS1b; SEQ ID NO: 3) increases the production of phenylalanine, tyrosine, tryptophan, and shikimate in *Nicotiana benthamiana* leaves (FIG. **5**A), and that expression of a nonconical TyrA enzyme from *Brachypodium distachyon* (BdTyrAnc; SEQ ID NO: 37) or *Sorghum bicolor* (SbTyrAnc; SEQ ID NO: 43) increases the production of tyrosine in *Nicotiana benthamiana* leaves (FIG. **3**F).

BACKGROUND

[0067] The biosynthesis of aromatic amino acids (AAAs)—phenylalanine, tyrosine, and tryptophan—represents one of the major routes of plant metabolism that supplies essential building blocks for the production of proteins and a myriad of plant natural products[1, 2]. Yet, it remains poorly understood how the AAA biosynthetic pathway is regulated to meet various demands for AAA precursors in different species. The most abundant of these AAA derived compounds is lignin, which accounts for up to 30% of plant dry weight and plays a critical role in strengthening and waterproofing secondary cell walls. In most plant species, lignin and the other phenylpropanoids are synthesized exclusively from phenylalanine by the enzyme phenylalanine ammonia lyase (PAL, FIG. **1**A)[4, 5]. In contrast, grasses (family Poaceae), arguably one of the most important plant lineages from both an ecological and economic perspective, can produce lignin from both tyrosine and phenylalanine due to the presence of bifunctional phenylalanine/tyrosine ammonia lyase (PTAL) enzymes[6-11]. Multiple lines of evidence support that a significant proportion of grass lignin is synthesized from tyrosine[10, 12, 13]. However, it remains unknown how grasses regulate the upstream AAA biosynthetic pathways to provide high amount of both tyrosine and phenylalanine precursors to support the unique dual lignin pathway.

[0068] Our current knowledge on the regulation of plant AAA biosynthesis, mostly derived from dicot models, indicates that plants balance AAA production by targeting activities of key enzymes of the AAA pathway(s) through a combination of transcriptional and feedback regulation[1,2,14]. For instance, the first enzyme in AAA biosynthesis, 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase (DHS; EC:2.5.1.54) (FIG. **1**A), is feedback regulated by AAAs and multiple downstream metabolites[15, 16]. This feedback regulation at the entry point of the shikimate pathway controls the biosynthesis of AAAs and, when released by point mutations targeting the regulatory domain of DHS, can largely increase AAA production and CO_2 fixation in *Arabidopsis thaliana*[16]. Similarly, the enzymes controlling tyrosine and phenylalanine biosynthesis from arogenate, arogenate dehydrogenase (TyrA; EC 1.3.1.78) and arogenate dehydratase (ADT; EC 4.2.1.91), are subjected to feedback inhibition by their corresponding reaction products (FIG. **1**A)[2, 14]. However, while TyrA enzymes are sensitive to feedback inhibition even at low tyrosine levels[17-20], vascular plants have specialized ADT isoforms that maintain their activity at high concentrations of phenylalanine[21, 22]. Furthermore, unlike bacterial DHS enzymes[23], none of the plant DHS enzymes characterized so far are inhibited by phenylalanine[15, 24], likely providing abundant phenylalanine precursor for phenylpropanoid production. Besides the feedback regulation at the enzyme level, DHS and ADT genes are often strongly co-expressed with PAL and other lignin and phenylpropanoid-related genes across different plants[15, 25-29]. Through this conjunction of transcriptional and feedback regulation, many plant species prioritize the production of phenylalanine for phenylpropanoid biosynthesis, to the detriment of tyrosine and tryptophan levels[26, 30-33]. However, given the presence of the unique dual lignin pathway, we hypothesized that grasses may regulate AAA production differently from other plants.

[0069] Here, we combine stable-isotope labeling, phylogenetic and expression analyses, detailed enzyme characterization, and combinatorial in planta expression analysis to demonstrate that the coordinated regulation of the entry and final steps of AAA biosynthesis allows grasses to efficiently provide both tyrosine and phenylalanine precursors to meet the unique demand of the dual tyrosine/phenylalanine lignin pathway. This study highlights the importance of transcriptional and biochemical regulation at key metabolic branching points in fine-tuning the supply of AAA precursors for the downstream lignin and phenylpropanoid pathway. These basic findings and the novel enzymes identified in grasses can be utilized to engineer plants to efficiently produce natural and bio-based aromatic products.

Results:

Grasses Synthesize Tyrosine at a Much Higher Rate than *Arabidopsis* without Compromising Phenylalanine Production

[0070] While prior studies reported that grasses accumulate high levels of tyrosine[35-37], an elevated steady-state level of a metabolite does not necessarily imply a high synthesis and usage rate[38]. Therefore, we performed $^{13}CO_2$ feeding experiments to compare the turnover rates of AAAs between the grass *Brachypodium distachyon* Bd21-3 and the dicot *Arabidopsis thaliana* Col-0. Four-weeks-old *Brachypodium* and *Arabidopsis* plants, before bolting, were fed side by side (FIG. **1**B) with an air mixture containing ~400 ppm of $^{13}CO_2$. Then, samples were collected at regular intervals for determination of ^{13}C labeled tyrosine, phenylalanine,

and shikimate by ultra-high performance liquid chromatography coupled to electrospray ionization mass spectrometry (UHPLC-MS).

[0071] Total tyrosine content (with either ^{12}C or ^{13}C) ranged between ~3 to 12 nmol per gram of fresh weight (nmol/gFW) in *Arabidopsis* but was much higher in *Brachypodium* (FIG. **1**C), where it reached up to 70 nmol/gFW. Moreover, ^{13}C-labeled tyrosine (mostly eight or nine ^{13}C-isotopologues, FIGS. **6**A-**6**B) was 10-times more abundant in *Brachypodium* (~40 nmol/gFW) than in *Arabidopsis* (~4 nmol/gFW) after 12h (FIG. **1**C). In contrast, total phenylalanine levels were comparable between the two species, in the range of 35-55 nmol/gFW in *Brachypodium*, and 20-35 nmol/gFW, in *Arabidopsis* (FIG. **1**D). Labeled ^{13}C-phenylalanine over time was also similar between the two species (FIG. **1**D). As observed for tyrosine, most ^{13}C-phenylalanine was fully labeled, containing eight or nine ^{13}C atoms (FIG. **6**). In addition, we detected striking differences in the dynamics of the shikimate pool, with up to 20-times more total shikimate accumulating in *Brachypodium* than in *Arabidopsis* by the end of the day, and a higher incorporation rate of ^{13}C (FIG. **1**E).

[0072] We next performed additional $^{13}CO_2$ labeling experiments using older six-weeks-old plants of *Arabidopsis, Brachypodium*, and *Setaria viridis* A10.1 (hereafter, *Setaria*), comparing young leaves with elongating stems, where lignin is actively formed. These experiments further confirmed that grass species accumulate more tyrosine than *Arabidopsis* (FIG. **2**A) and showed that incorporation of ^{13}C into tyrosine was particularly rapid in grass stems, which accumulated up to ~50 nmol/gFW of ^{13}C-tyrosine after 3 hours of $^{13}CO_2$ feeding (FIG. **2**A), more than 10-times faster than *Arabidopsis* (<2 nmol/gFW; FIG. **2**A). On the contrary, the three species exhibited comparable labeling kinetics for phenylalanine, with faster ^{13}C-phenylalanine accumulation in the stems than in the leaves (FIG. **2**B). In the case of shikimate, *Arabidopsis* stems showed 10 to 20-times more total shikimate and higher rate of ^{13}C-labeling than the leaves (FIG. **2**C), despite high biological variation. In contrast, grass species showed a faster shikimate labeling in the leaves (FIG. **2**C). The time-course labeling of different ^{13}C-isotopologues of phenylalanine, tyrosine, and shikimate differed between leaf and stem tissues and among species (FIG. **7**). These results showed that, while the three species have a high rate of phenylalanine biosynthesis in the stems, only grasses exhibit high tyrosine turnover in this organ. Furthermore, the high rate of tyrosine biosynthesis in grasses did not seem to compromise phenylalanine biosynthesis.

Grass TyrA1 and TyrAnc Isoforms are Highly Expressed in Growing Stems

[0073] To understand the mechanism behind the increased production of tyrosine in grasses, we next examined the family of grass TyrA enzymes, which catalyze the final and key regulatory step in tyrosine biosynthesis[2, 14, 39]. Reconstruction of the plant TyrA protein phylogeny showed that grass genomes have at least three TyrA isoforms (FIG. **3**A) corresponding with the *Brachypodium distachyon* v3.2 loci Bradi1g34789, Bradi1g34807, and Bradi1g39160, which we named TyrA1, TyrA2, and non-canonical TyrA (TyrAnc), respectively. Whereas grass TyrA1 and TyrA2 are closely related to each other and to TyrA enzymes from most dicot plants, grass TyrAnc cluster in a more distant group and is a sister to cytosolic TyrAnc enzymes from legumes and other dicots (FIG. **3**A)[39, 40]. All three BdTyrA proteins have predicted plastid transit peptides in their N-terminus (TargetP-2.0, DTU Health Tech), similarly to most plant TyrA enzymes[40], and were targeted to the plastids when expressed in *Arabidopsis* protoplast fused to enhanced green fluorescent protein (EGFP) in their C-terminus (FIG. **8**).

[0074] To examine the potential involvement of TyrA in the tyrosine-lignin pathway of grasses, we compared the expression of TyrA genes with PTAL using publicly available expression datasets from *Brachypodium*[41] and *Sorghum*[42]. Interestingly, the expression profile of TyrA enzymes, in particular TyrA1 and TyrAnc, resembled that of PTAL, showing higher expression in stem internodes and roots, and low in seeds, flowers, and leaves (FIG. **3**B), which correlates with the elevated rate of tyrosine production observed in grass internodes in the $^{13}CO2$ feeding experiments (FIG. **2**). Furthermore, gene co-expression networks in *Brachypodium*[41], showed that BdTyrA1 and BdPTAL (Bradi3g49250) expression correlate with many other genes of the lignin pathway (data not shown). Absolute real-time quantitative PCR (RT-qPCR) comparing TyrA expression in young leaves and developing internodes of *Brachypodium* and *Setaria* found that TyrA1 had the highest expression among TyrA genes, followed by TyrAnc and TyrA2 (FIG. **3**C, FIG. **9**). Importantly, TyrA1 and TyrAnc transcripts were up to 10-times more abundant in young developing internodes than in leaves (FIG. **3**C). This was not clearly observed for TyrA2 genes, which showed a fold change comparable to ubiquitin ligase reference genes (~1.5-times; FIG. **9**). Altogether, these results support that the expression of TyrA1 and TyrAnc, but not TyrA2, is strongly induced in developing stems, where active PTAL expression and lignin deposition take place.

TyrAnc Enzymes, but not TyrA1 Enzymes, Exhibit Low Sensitivity to Feedback Inhibition by Tyrosine

[0075] Next, to study the biochemical properties of grass TyrA enzymes, we generated and characterized recombinant purified TyrA proteins from *Brachypodium* (BdTyrA) and *Sorghum* (SbTyrA), two distantly related grass species[43, 44]. Whereas TyrA enzymes from dicot plants are generally $NADP^+$-dependent arogenate dehydrogenases[20, 39, 45], except for $NADP^+$-prephenate dehydrogenases in the legume family[40, 46], previous reports on the substrate preference of grass TyrA enzymes are inconclusive[17, 19]. Initial biochemical assays to test the substrate and cofactor preference revealed that grass TyrA enzymes are most active with arogenate as substrate, rather than prephenate, and $NADP^+$ as cofactor, exhibiting only minor NAD^+-arogenate dehydrogenase activity (up to 5% of the main activity; FIG. **10**). NAD^+-prephenate dehydrogenase activity was absent in all cases, ruling out a significant contamination of the enzyme preparations with the TyrA enzyme from *E. coli*[47, 48]. Detailed kinetic analyses of the activity with arogenate and $NADP^+$ showed that the six TyrA enzymes obey Michaelis-Menten kinetics, with the TyrAnc isoforms having, by a wide margin, the highest turnover number (k_{cat}) and lowest Michaelis-Menten constant (K_m) (Table 1, FIG. **11**). Consequently, the catalytic efficiencies (k_{cat}/K_m) of BdTyrAnc (~580 s^{-1} mM^{-1}) and SbTyrAnc (~337 s^{-1} mM^{-1}) were the highest amidst the three isoforms of each species (Table 1).

TABLE 1

Kinetic parameters of TyrA enzymes. K_m and k_{cat} were calculated from Michaelis-Menten plots shown in FIG. 11. k_{cat}/K_m was calculated based on K_m, k_{cat}, and the molecular weight of each recombinant enzyme (including the mass of the poly-histidine tag). The half-inhibitory concentration of tyrosine (IC_{50}) was calculated at 0.5 mM of arogenate and 1 mM of $NADP^+$ from the data shown in FIG. 11. The inhibition constant for tyrosine (K_i) was calculated from K_m and IC_{50} values under a competitive inhibition model[19, 20]. All data are means ± SD of n = 4-6 derived from at least two independent experiments conducted on different days using different batches of purified recombinant enzyme.

	K_m (mM arogenate)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1} \cdot mM^{-1}$)	IC_{50} (μM tyrosine)	K_i (μM tyrosine)
BdTyrA1	1.41 ± 0.24	5.2 ± 0.8	3.6 ± 0.8	66 ± 2	46 ± 3
SbTyrA1	0.57 ± 0.15	2.9 ± 0.6	5.3 ± 0.8	71 ± 24	53 ± 10
BdTyrA2	0.45 ± 0.04	18.5 ± 0.6	41.2 ± 1.7	20 ± 6	8 ± 2
SbTyrA2	2.13 ± 0.53	38.5 ± 8.8	18.1 ± 9.6	47 ± 23	36 ± 13
BdTyrAnc	0.13 ± 0.04	76.5 ± 5.5	579.6 ± 24.3	242 ± 45	64 ± 25
SbTyrAnc	0.22 ± 0.06	73.5 ± 18.2	337.5 ± 19.5	406 ± 85	137 ± 44

[0076] The activity of most plant TyrA enzymes is inhibited competitively at low concentration of tyrosine, generally in the half maximum inhibition (IC_{50}) range of 10 to 50 μM when assayed in vitro[17-20]. Like other plant TyrA enzymes, BdTyrA1 showed an IC_{50} for tyrosine at ~65 μM, and BdTyrA2 had even lower IC_{50} of ~20 μM (FIG. **3**D, Table 1, FIG. **11**). In contrast, BdTyrAnc exhibited a low sensitivity to inhibition by tyrosine, with an estimated IC_{50} of ~240 μM. Hence, BdTyrAnc retains >50% of its activity at 200 μM of tyrosine, where BdTyrA1 and BdTyrA2 are fully inactive (FIG. **3**D). Despite this marked difference in sensitivity, the inhibition of BdTyrAnc by tyrosine is competitive with arogenate (FIG. **12**), as reported for other TyrA enzymes[19, 20]. Like BdTyrAnc, SbTyrAnc also showed low sensitivity to feedback-inhibition, having a high IC_{50} for tyrosine of ~475 μM, whereas SbTyrA1 and SbTyrA2 did not (IC_{50} at 71 and 44 μM, respectively) (FIG. **3**E, Table 1).

[0077] To investigate if the difference in sensitivity to feedback inhibition impacts the activity of the TyrA isoforms in planta, we transiently expressed *Brachypodium* and *Sorghum* TyrA genes in *Nicotiana benthamiana* through *Agrobacterium* leaf infiltration (FIG. **3**F). As controls, β-glucuronidase (GUS), the tyrosine-inhibited AtTyrA2 from *Arabidopsis*[20], and the deregulated BvTyrAα from *Beta vulgaris*[45] were also expressed, all under control of the CaMV 35S promoter ($P_{CaMV35S}$) (FIG. **13**). Tyrosine content in the transfected leaves was 2.5 to 3 times higher in BdTyrA1, SbTyrA1, and SbTyrA2 infiltrated leaves compared to the GUS control (FIG. **3**F). Similar tyrosine levels were observed in the leaves infiltrated with AtTyrA2. Overexpression of BdTyrA2, which encodes a strongly feedback inhibited enzyme with the lowest IC_{50} among grass TyrA enzymes (Table 1), did not significantly increase tyrosine levels. In contrast, infiltration with the BdTyrAnc and SbTyrAnc constructs increased tyrosine content by 8 and 14-times relative to the GUS control, respectively, causing an effect similar to that of the deregulated BvTyrAα (FIG. **3**F). The in planta accumulation of tyrosine correlated better with the sensitivity of the different TyrA enzymes to feedback inhibition (IC_{50}), rather than the other kinetical parameters (k_{cat}, K_m, kcal/K_m; Table 1). These results support that, in agreement with their sensitivity to feedback inhibition in vitro, grass TyrAnc, but not TyrA1 or TyrA2, can greatly increase tyrosine production when expressed in planta.

Grasses have a Feedback Insensitive DHS1b Enzyme

[0078] Feeding experiments using $^{13}CO_2$ revealed that, beyond high tyrosine production, grass species also synthesize shikimate and phenylalanine at a higher rate than *Arabidopsis* (FIG. **1** and FIG. **2**). These findings suggest that the regulation of the upstream shikimate pathway may be different in grass species. To test this hypothesis, we characterized the DHS enzymes from *Brachypodium* and *Sorghum*, which catalyze a key regulatory step at the entry point of the shikimate pathway (FIG. **1**A)[1, 14, 24].

[0079] The phylogeny of plant DHS enzymes shows that grasses generally have four DHS isoforms (FIG. **4**A), which correspond with the *Brachypodium* loci Bradi1g21330 (namely BdDHS1a), Bradi1g60750 (BdDHS1b), Bradi3g38670 (BdDHS2), and Bradi3g33650 (BdDHSnc, from non-canonical). Whereas DHS2 and DHSnc are conserved in other monocots, DHS1a and DHS1b (which share ~90% of protein sequence identity) are likely derived from a gene duplication event within the grass family. The four *Brachypodium* DHS genes differ in their spatio-temporal expression profile (FIG. **4**B). BdDHS2 is dominant in photosynthetic organs, BdDHS1a is expressed across different organs and stages, and BdDHSnc is mostly expressed in seeds. Notably, BdDHS1b expression is induced in the internodes (FIG. **4**B), and is co-expressed with BdPTAL, BdTyrA1 and other lignin pathway genes (data not shown)[41].

[0080] To examine their functional properties, the recombinant DHS enzymes of *Brachypodium* were produced and characterized in vitro. Though BdDHSnc was also produced, it was not soluble in bacteria and could not be studied. Enzyme assays showed Michaelis-Menten kinetics for phosphoenolpyruvate, with K_m values in the range of 135 to 200 μM (Table 2, FIG. **14**), but weak to moderate positive cooperativity for erythrose 4-phosphate, with a $K_{0.5}$—the analogous parameter to K_m in cooperative kinetics—in between 400 and 550 μM (Table 2, FIG. **14**). k_{cat} values were in the same order of magnitude for the three isoforms (Table 2, FIG. **14**). Thus, the three *Brachypodium* DHS enzymes seem to have similar kinetical parameters.

TABLE 2

Kinetic parameters of DHS enzymes. The data used to determine the kinetic parameters $K_m/K_{0.5}$ and k_{cat} are shown in FIG. 14. k_{cat}/K_m was calculated as described for TyrA enzymes in the legend of Table 1. IC_{50} for tryptophan and arogenate were determined based on the original data shown in FIG. 4D and FIG. 4F, respectively. n.i. = not inhibited. All data are means ± SD of n = 4-6 derived from at least two independent experiments conducted on different days using different batches of purified recombinant enzyme.

	K_m	EC_{50}	H		k_{cat}/K_m (PEP)	k_{cat}/EC_{50} (E4P)	IC_{50}	
	PEP (mM)	E4P (mM)	Coefficient (only E4P)	k_{cat} (s^{-1})	(s^{-1} · mM^{-1})	(s^{-1} · mM^{-1})	(μM Agn)	(μM Trp)
BdDHS1a	0.13 ± 0.02	0.55 ± 0.04	1.8 ± 0.1	17.0 ± 0.2	126.0 ± 1.7	30.7 ± 0.4	285 ± 24	n.i.
BdDHS1b	0.20 ± 0.07	0.40 ± 0.02	2.1 ± 0.1	11.3 ± 0.1	57.2 ± 0.3	28.3 ± 0.1	n.i.	n.i.
BdDHS2	0.19 ± 0.01	0.51 ± 0.10	1.5 ± 0.6	5.6 ± 0.01	29.0 ± 0.1	10.8 ± 0.01	91 ± 11	120 ± 4

[0081] As recent studies have shown that plant DHS enzymes are feedback-inhibited by multiple effector molecules[15, 16], we tested the effect of AAAs and another 14 related metabolites, including various intermediates of the shikimate pathway and the pathways downstream of AAAs, on *Brachypodium* DHS enzymes. The effect of these compounds was determined at a concentration of 0.5 mM with two alternative methods: real-time spectrophotometric quantification of phosphoenolpyruvate consumption[49], and final-point quantification of the reaction product, DAHP, by UHPLC-MS.

[0082] Among the three AAAs, phenylalanine did not cause significant effects on grass DHS activities, which seems to be a common feature in plant DHS enzymes[24]. Tyrosine, which strongly inhibits *Arabidopsis* DHS enzymes[15], only caused ~25% inhibition in BdDHS1a and ~10% in BdDHS1b and BdDHS2 (FIG. **4**C). Conversely, tryptophan strongly inhibited BdDHS2 at an IC_{50} of ~120 μM but had no effect on BdDHS1a or BdDHS1b (FIG. **4**C, FIG. **4**D, Table 2). We did not observe any remarkable inhibitory effect caused by intermediates of AAA catabolism or lignin biosynthesis (FIG. **15**). From the different AAA pathway(s) intermediates tested, we observed that arogenate caused the most dramatic effect, causing ~50% inhibition of BdDHS1a, and >75% in BdDHS2 at 0.5 mM (FIG. **4**E, FIG. **15**), with a calculated IC_{50} of ~285 and ~91 μM, respectively (FIG. **4**F, Table 2). In contrast, BdDHS1b was not inhibited by arogenate even up to 2.5 mM. In fact, DHS1b activity increased in the presence of high concentrations of arogenate (FIG. **4**F), likely due to high concentrations of contaminant NaCl present in the arogenate preparation (FIG. **16**). Under acidic conditions, arogenate is known to undergo a spontaneous dehydration and decarboxylation into phenylalanine[50]. Arogenate incubated with HCl, which is therefore fully converted into phenylalanine, did not inhibit the DHS enzymes, supporting that the inhibitory compound was arogenate instead of other possible contaminants (FIG. **16**). The characterization of the recombinant *Sorghum* DHS enzymes confirmed that no strong inhibition was caused by tyrosine or phenylalanine, whereas 0.5 mM of tryptophan caused 40 to 50% inhibition of SbDHS2 and SbDHSnc (FIG. **17**). Like in *Brachypodium* DHS enzymes, 0.5 mM arogenate inhibited SbDHS1a and SbDHS2 at ~50% and ~40%, respectively, but had no effect on SbDHS1b.

[0083] Determination of the kinetic parameters of BdDHS2 at different concentrations of tryptophan and arogenate showed that both effectors decrease V_{max} but had distinct impacts on K_m or EC_{50}. For PEP, tryptophan did not cause a significant change in the K_m, which is indicative of non-competitive inhibition, but arogenate caused the K_m to increase, indicating a mixed inhibition mechanism (FIG. **18**). In respect to E4P, both tryptophan and arogenate decreased $K_{0.5}$, suggesting uncompetitive inhibition kinetics (FIG. **18**). These findings resemble previous studies from bacterial type-II DHS enzymes[51] and support that type-II DHS enzymes, which include plant DHS enzymes, are allosteric enzymes with a complex response to the binding of their substrates and effectors.

[0084] DHS effector molecules can have synergistic effects when combined in vitro[51]. To explore this possibility, we tested the impact of different combinations of tryptophan, tyrosine, arogenate and chorismate, at 0.15 mM each, on the activity of *Brachypodium* DHS enzymes. Although most combinations did not exhibit strong additive effects, some of the combinations, such as tryptophan plus arogenate for BdDHS2 and tyrosine plus arogenate for BdDHS1a, caused additional inhibition (FIG. **19**). None of the combinations tested had a significant impact on BdDHS1b (FIG. **19**). Hence, although grass DHS1a is inhibited by arogenate and DHS2 by both arogenate and tryptophan, DHS1b seems largely insensitive to feedback inhibition in vitro.

Co-Expression of BdDHS1b and BdTyrA1 Synergistically Enhances Tyrosine Production while Maintaining High Phenylalanine Production

[0085] To evaluate in planta how DHS biochemical regulation may impact the production of AAAs, we expressed BdDHS1a, BdDHS1b, and BdDHS2 in *Nicotiana benthamiana* leaves under control of the *Arabidopsis* RuBisCO S3B promoter ($P_{AtRbcS3B}$), which provides 15-20% of the expression level of CaMV 35S promoter[52] (FIG. **5**A). Transient expression of BdDHS1a and BdDHS2, both sensitive to feedback inhibition in vitro, did not significantly alter the content of phenylalanine, tyrosine, tryptophan, nor their common precursor shikimate, compared to the control expressing tdTomato (tdTom) (FIG. **5**A). To the contrary, the expression of BdDHS1b triggered the accumulation of 10-times more tyrosine, 3-times more shikimate and tryptophan, and 18-times more phenylalanine, which was the

most abundant AAA (~2000 nmol/gFW; FIG. **5**A). Quantification of the BdDHS-HA tagged proteins by immunoblotting showed that, despite the marked differences in metabolite levels, the protein levels of these three DHS isoforms were comparable (FIG. **20**). Taken together, these results support in vitro results showing that grasses possess a naturally deregulated DHS1b that can boost AAA production, mostly phenylalanine, when expressed heterologously in planta.

[0086] Gene expression data in *Brachypodium* (FIG. **4**B) indicate that BdTyrA1 is co-expressed with BdDHS1b and BdPTAL in the internodes, where we detected a high rate of tyrosine production (FIG. **2**). Nevertheless, BdTyrA1 is strongly inhibited by tyrosine in vitro (FIG. **3**D) and its expression alone in *Nicotiana* leaves had little impact on tyrosine levels (FIG. **3**F). We therefore hypothesized that BdTyrA1 and BdDHS1b may co-operate in the high production of tyrosine and phenylalanine observed in grass tissues. To test this possibility, we co-expressed in planta different combinations of BdTyrA enzymes and BdDHS enzymes and measured the impact on phenylalanine and tyrosine levels. To this end, we took advantage of the Golden Gate modular cloning system[52] and assembled the BdTyrA and BdDHS expression cassettes into the same vector backbone (FIG. **13**). To avoid causing a strong over-expression, BdTyrA and BdDHS were driven by RuBisCO small subunit promoters from *Arabidopsis* ($P_{AtRbcS3B}$) and tomato ($P_{SlRbcS3A}$), respectively (FIG. **13**)[51]. Consistent with the results shown in FIG. **3**F, the expression of BdTyrAnc, but not BdTyrA1 or BdTyrA2, together with the tdTom control, led to a strong increase in tyrosine levels (~100-times) (FIG. **5**B). Notably, BdTyrAnc expression decreased phenylalanine levels by 4-times compared to the negative control co-expressing YPet with tdTom (FIG. **5**B). Co-expression of BdTyrAnc with BdDHS1b showed additive effects and further increased tyrosine to a dramatic ~400-times the tdTom+YPet control, while still negatively impacting phenylalanine level (FIG. **5**B). Interestingly, co-expression of feedback-regulated BdTyrA1 together with BdDHS1b boosted tyrosine content to a level similar to BdTyrAnc, while still maintaining a high production of phenylalanine (FIG. **5**B). This synergistic effect was not observed upon co-expression of BdDHS1b with BdTyrA2, possibly due to the tight feedback regulation of BdTyrA2 (Table 1). Similarly, no significant additive effects were observed upon co-expression of the BdTyrA enzymes with feedback-regulated BdDHS2 (FIG. **5**B). These results show that simultaneous expression of deregulated BdDHS1b with the feedback-regulated enzyme BdTyrA1 can render high levels of both tyrosine and phenylalanine in planta.

DISCUSSION

[0087] The emergence of phenylpropanoid metabolism is a key adaptation during the transition of plants from water to land, conferring plants with enhanced mechanical strength, and protection against UV radiation and desiccation. Phenylpropanoids are remarkably diverse across the plant phylogeny and are synthesized exclusively from phenylalanine and by the PAL enzyme in almost all plant groups[5, 53, 54]. Grasses are an exception, as they use the PTAL reaction to synthesize phenylpropanoids from tyrosine, which constitutes a shortcut in the "canonical" phenylpropanoid pathway (FIG. **1**)[5, 55]. In this study, we used this unique grass feature to investigate how enzyme evolution re-shapes metabolic flows and regulation at the interface of primary metabolism and the downstream natural product pathways.

[0088] Labeling experiments using $^{13}CO_2$ provided well-grounded evidence that supports high tyrosine production in grasses, especially in the internodes (FIG. **1** and FIG. **2**). Consistently, the expression of TyrA genes was induced in lignifying tissues of grasses (FIG. **3**), unlike in dicots where ADT but not TyrA genes are co-expressed with lignin pathway genes[29, 56, 57]. One of the grass TyrA enzymes, TyrAnc, also showed low sensitivity to inhibition by tyrosine in vitro, and its expression in *Nicotiana benthamiana* boosted tyrosine levels (FIG. **3**F and FIG. **5**B), though at the expense of phenylalanine production (FIG. **5**B). These findings clearly support that, unlike most plant TyrA enzymes[18-20], grass TyrAnc has low sensitivity to feedback inhibition. Partially feedback insensitive TyrA enzymes have been previously described in Caryophyllales, with IC_{50} values between ~400 and 700 μM of tyrosine[45, 58], and in legumes, where the prephenate dehydrogenase TyrAp is completely feedback insensitive[40, 46]. However, phylogenetic evidence shows that grass TyrAnc enzymes are not related to the Caryophyllales TyrAα enzyme and legume TyrAp enzyme (FIG. **3**A), indicating that feedback deregulated TyrA enzymes likely evolved independently in these three plant groups.

[0089] Despite their highly active tyrosine biosynthesis, grasses still maintain a high rate of phenylalanine production, particularly in stems (FIG. **1** and FIG. **2**). Detailed characterization of grass DHS isoforms further revealed the presence of DHS1b which, unlike DHS1a and DHS2, was insensitive to feedback inhibition (FIG. **4**). The expression of DHS1b, but not DHS1a or DHS2, in *Nicotiana* leaves also boosted the level of shikimate, tryptophan, tyrosine and, especially phenylalanine (FIG. **5**A). Although feedback-insensitive bacterial DHS enzymes have been introduced in plants to increase the production of AAAs and/or phenylpropanoids[21, 32, 59-61] DHS1b constitutes, to our knowledge, the first report of a naturally occurring deregulated plant DHS.

[0090] Previous studies on biochemical characterization of DHS activity from plant extracts-most of them dicots-reported varying observations about the sensitivity of DHS to feedback regulation[24], likely due to the presence of multiple DHS isoforms in plants[24, 62]. In monocots, a single study found that DHS activity in crude extracts from 9-day-old maize plantlets is inhibited by tryptophan, but not by phenylalanine or tyrosine[63], consistent with a predominant role of the tryptophan-inhibited DHS2 in green tissues (FIG. **4**). Based on the differences in feedback regulation and expression patterns of individual grass DHS isoforms (FIG. **4**), DHS1a and DHS2 may display more general or housekeeping roles, whereas DHS1b may have organ-specific (e.g., internode) functions that demand high AAA production. Recent studies in *Arabidopsis* also showed a complex isoform-dependent feedback regulation where different DHS isoforms are expressed and regulated differently[15, 16]. Moreover, *Arabidopsis* DHS enzymes were inhibited by chorismate, caffeic acid, tyrosine, and its metabolites 4-hydroxyphenylpyruvate and homogentisic acid[39], which all had little effect on grass DHS enzymes (FIG. **4**, FIG. **15**, and FIG. **17**). In the case of tyrosine and its derivatives 4-hydroxyphenylpyruvate and homogentisic acid, the insensitivity of grass DHS enzymes might be linked to the high

tyrosine levels present in grass tissues (FIG. **1** and FIG. **2**). Therefore, it seems that the feedback regulation of plant DHS enzymes is not only isoform-dependent, but also species-dependent. These findings indicate that both biochemical and transcriptional regulatory mechanisms targeting DHS enzymes give plant species a precise yet adaptable tool to modulate AAA production.

[0091] Although the feedback insensitive TyrAnc is likely contributing to the high rate of tyrosine production in grass internodes (FIG. **2**), it is somewhat incongruent that TyrA1, expression of which is strongly induced in the internodes (FIG. **3**C), encodes a tyrosine-inhibited enzyme (Table 1). Interestingly, the combinatorial expression of the feedback inhibited TyrA1 with the deregulated DHS1b led to high production of both tyrosine and phenylalanine in planta (FIG. **5**). This synergistic effect might be a consequence of "pushing" the carbon flow into the shikimate pathway, causing an accumulation of arogenate that would alleviate the competitive feedback inhibition of TyrA1 by tyrosine (FIG. **12**)[19, 20] and support high rates of both tyrosine and phenylalanine production. The relatively high K_m values of grass TyrA1 enzymes for arogenate (Table 1) may reflect their specialization for working at the high substrate levels provided by DHS1b, while TyrAnc is possibly better suited for producing tyrosine at low arogenate concentrations thanks to its lower K_m value (Table 1).

[0092] Overall, the current findings highlight that the interplay between feedback-regulated (TyrA1) and deregulated (DHS1b, TyrAnc) enzymes at the entry and exit steps of AAA biosynthesis can maintain the high production of both tyrosine and phenylalanine. This fine-tuning of the upstream AAA pathway likely supports the unique dual lignin pathway found in grasses. Future studies of these key enzymes from different monocot species will address the evolutionary history of the coordinated regulation of the grass AAA and lignin pathways. This fundamental knowledge also provides useful genetic tools for the rationale engineering of plant primary metabolism to support the production of aromatic products.

Materials and Methods

Plant Materials and Growth Conditions

[0093] The following grass cultivars were used in this study: *Brachypodium distachyon* 21-3, *Sorghum bicolor* RTx430, and *Setaria viridis* A10.1.

[0094] *Arabidopsis, Brachypodium*, and *Setaria* plants used for $^{13}CO_2$ feeding and RT-qPCR analysis were kept in a growth chamber at 22° C., 12h-photoperiod under ~100 μE of light intensity, 60% humidity, and watered with a 1:10 dilution of Hoagland's solution.

[0095] *Nicotiana benthamiana* plants used for transient expression experiments were grown at 22° C. in a 12h-photoperiod under ~200 μE of light intensity, 60% humidity, and watered with a 12:4:8 (N:P:K) plant nutritive solution (Miracle-Gro) at a 1:1000 dilution.

Gene Expression Analysis and RT-qPCR

[0096] Spatiotemporal gene expression data from *Brachypodium distachyon* and *Sorghum bicolor* were retrieved from the databases PlaNet[41] and MOROKOSHI[42], respectively.

[0097] Total RNA was isolated from young leaves and developing internodes of 1.5-month-old *Brachypodium* and *Setaria* plants using RNeasy Plant Mini Kit (Qiagen), following the manufacturer's instructions. RNA was treated with RQ1 RNase-free DNase (Promega) and was reverse transcribed with M-MLV Reverse Transcriptase (Promega) using random hexamer primers. Quantitative PCR analysis was carried out in a Stratagene Mx3000P (Agilent Technologies) thermocycler using GoTaq qPCR Master Mix (Promega). Ct values were determined using LinRegPCR[64] version 2018.0. Primers used are listed in Table 3. Ct values were converted into mass of template by using a calibration curve made of the corresponding RT-qPCR amplicon, cloned into the EcoRV site of pML94 vector using conventional blunt-end ligation protocols. Ubiquitin ligase genes of *Brachypodium* and *Setaria* were chosen as reference genes based on previous publications[65].

TABLE 3

Primers

Primer Name	Sequence (from 5'-)	Target gene	Purpose
pHM532Bd3 47pML94F	GATCTAGACTCGAGGGT ACCATGCTGTCGTCTTCC (SEQ ID NO: 45)	BdTyrA1	In-Fusion cloning into pML94 for transient expression in *Arabidopsis* protoplasts
pHM533Bd3 47pML94R	CTAGTGCATGCGGCCGC ATTCCGGACGGTGG (SEQ ID NO: 46)	BdTyrA1	In-Fusion cloning into pML94 for transient expression in *Arabidopsis* protoplasts
pHM534Bd3 47pET28aF	CGCGCGGCAGCCATATG CGCGCCATCGACGC (SEQ ID NO: 47)	BdTyrA1	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM535Bd3 47pET28aR	GCTCGAATTCGGATCCC TAATTCCGGACGGTGG (SEQ ID NO: 48)	BdTyrA1	In-Fusion cloning into NdeI/BamHI sites of pET28a

TABLE 3-continued

Primers

Primer Name	Sequence (from 5'-)	Target gene	Purpose
pHM536Bd3 48pML94F	GATCTAGACTCGAGGGT ACCATGTCTTCCTCCGGT C (SEQ ID NO: 49)	BdTyrA2	In-Fusion cloning into pML94 for transient expression in *Arabidopsis* protoplasts
pHM537Bd3 48pML94R	CTAGTGCATGCGGCCGC CCGTCCGTCATCCAA (SEQ ID NO: 50)	BdTyrA2	In-Fusion cloning into pML94 for transient expression in *Arabidopsis* protoplasts
pHM538Bd3 48pET28aF	CGCGCGGCAGCCATATG CGTGCCACGGACGC (SEQ ID NO: 51)	BdTyrA2	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM539Bd3 48pET28aR	GCTCGAATTCGGATCCT CACCGTCCGTCATCC (SEQ ID NO: 52)	BdTyrA2	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM540Bd3 9pML94F	GATCTAGACTCGAGGGT ACCATGGCTTCCTCCCTT G (SEQ ID NO: 53)	BdTyrAnc	In-Fusion cloning into pML94 for transient expression in *Arabidopsis* protoplasts
pHM541Bd3 9pML94R	CTAGTGCATGCGGCCGC AATGGGGACCCTCTCT (SEQ ID NO: 54)	BdTyrAnc	In-Fusion cloning into pML94 for transient expression in *Arabidopsis* protoplasts
pHM542Bd3 9pET28aF	CGCGCGGCAGCCATATG GCCGAGCAGGAGCAA (SEQ ID NO: 55)	BdTyrAnc	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM543Bd3 9pET28aR	GCTCGAATTCGGATCCTT AAATGGGGACCCTCTC (SEQ ID NO: 56)	BdTyrAnc	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM1751-SbTyrA1GGF	TCACTCTGTGGTCTCAAA TGCGCGCGCTGGACGCC GCCC (SEQ ID NO: 57)	SbTyrA1	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM1752-SbTyrA1GGR	CCACTTCGTGGTCTCACG AACTCTTGCGGACGTTG GAGGAG (SEQ ID NO: 58)	SbTyrA1	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM1753-SbTyrA2GGF	TCACTCTGTGGTCTCAAA TGCGCGCCACGGGTGCC TCGC (SEQ ID NO: 59)	SbTyrA2	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM1754-SbTyrA2GGR	CCACTTCGTGGTCTCACG AACTATTATTCTCCTCTC CCGACTTG (SEQ ID NO: 60)	SbTyrA2	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM1755-SbTyrA3GGF	TCACTCTGTGGTCTCAAA TGAGCCCGCCCGCCGCC ACCGC (SEQ ID NO: 61)	SbTyrAnc	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM1756-SbTyrA3GGR	CCACTTCGTGGTCTCACG AACTTAAGAGGCGAGCT GCAGGAG (SEQ ID NO: 62)	SbTyrAnc	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM1757-SbTyrA3nestF	ATGGCCTCCTCGCTCCGC C (SEQ ID NO: 63)	SbTyrAnc	Nested PCR for Sorghum TyrAnc
pHM1758-SbTyrA3nestR	GTATCCGGTTGAAGTGT AGG (SEQ ID NO: 64)	SbTyrAnc	Nested PCR for Sorghum TyrAnc
pHM2274_BdTyrA1_qF	CACCACCGTCCGGAATT AGC (SEQ ID NO: 65)	BdTyrA1	RT-qPCR

TABLE 3-continued

Primers

Primer Name	Sequence (from 5'-)	Target gene	Purpose
pHM2275_ BdTyrA1_qR	GCACCAGTTTCTCCCCA AAG (SEQ ID NO: 66)	BdTyrA1	RT-qPCR
pHM2316 BdTyrA2 qF2	GATGACGGACGGTGATC TCG (SEQ ID NO: 67)	BdTyrA2	RT-qPCR
pHM2317 BdTyrA2 qR2	TTCGTACCGCTTGTTGGT CG (SEQ ID NO: 68)	BdTyrA2	RT-qPCR
pHM2278_ BdTyrA3_qF	TGCTGTGTCCCCTCTCCT C (SEQ ID NO: 69)	BdTyrAnc	RT-qPCR
pHM2279_ BdTyrA3_qR	AGGGCTGAAAGACACTG GGC (SEQ ID NO: 70)	BdTyrAnc	RT-qPCR
pHM2318 SvTyrA2 qF2	CAGACAATGCGGAGATG ATCG (SEQ ID NO: 71)	SvTyrA2 (Sevir.4G287000.1)	RT-qPCR
pHM2319 SvTyrA2 qR2	TTTGCTTCAGAAACCAT GTCAC (SEQ ID NO: 72)	SvTyrA2 (Sevir.4G287000.1)	RT-qPCR
pHM2282_ SvTyrA1_qF	GGTAGTAATTCAGTGCG TCGG (SEQ ID NO: 73)	SvTyrA1 (Sevir.4G286800.1)	RT-qPCR
pHM2283_ SvTyrA1_qR	GGTGTTCTTCTTCCAGAG AGG (SEQ ID NO: 74)	SvTyrA1 (Sevir.4G286800.1)	RT-qPCR
pHM2284_ SvTyrA3_qF	GATCGCTTCCATCCCAA GGC (SEQ ID NO: 75)	SvTyrAnc (Sevir.4G143200.1)	RT-qPCR
pHM2285_ SvTyrA3_qR	CAGGCGGTCTGAAAGGA AGG (SEQ ID NO: 76)	SvTyrAnc (Sevir.4G143200.1)	RT-qPCR
pHM2290_ BdUBI10_qF	AGTTGTCGCGTGTCTGA GTC (SEQ ID NO: 77)	Bradilg32860.3	RT-qPCR, reference gene Polyubiquitin 10
pHM2291_ BdUBI10_qR	ACACGGGCTCACTTATT CATC (SEQ ID NO: 78)	Bradilg32860.3	RT-qPCR, reference gene Polyubiquitin 10
pHM2294_ SvUBI4_qF	GGGCTCATTGTGCTGCT GTC (SEQ ID NO: 79)	Sevir.5G079801.1	RT-qPCR, reference gene Polyubiquitin 4
pHM2295_ SvUBI4_qR	CCGGAGGACATAGGACT TGC (SEQ ID NO: 80)	Sevir.5G079801.1	RT-qPCR, reference gene Polyubiquitin 4
pHM2178 BdDHS1b Fwd	GGTGCCGCGCGGCAGCC ATATGGCCGTCCACGCC GCGGAGCC (SEQ ID NO: 81)	BdDHS1b	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2179 BdDHS1b Rvs	CGGAGCTCGAATTCGGA TCCTCAGAAACCATAGG TTGGCAATG (SEQ ID NO: 82)	BdDHS1b	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2180 BdDHS1a Fwd	GGTGCCGCGCGGCAGCC ATATGGCCGTGCACGCC GCCGACCC (SEQ ID NO: 83)	BdDHS1a	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2181 BdDHS1a Rvs	CGGAGCTCGAATTCGGA TCCTTAGAAGGCCAATG GCGGCAGTG (SEQ ID NO: 84)	BdDHS1a	In-Fusion cloning into NdeI/BamHI sites of pET28a

TABLE 3-continued

Primers

Primer Name	Sequence (from 5'-)	Target gene	Purpose
pHM2184 BdDHS2 Fwd	GGTGCCGCGCGGCAGCC ATATGATCCGCGCGCAC GCGGTGCG (SEQ ID NO: 85)	BdDHS2	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2185 BdDHS2 Rvs	CGGAGCTCGAATTCGGA TCCTCAGAGTCCCATGG ATGATGG (SEQ ID NO: 86)	BdDHS2	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2182 BdDHSnc Fwd	GGTGCCGCGCGGCAGCC ATATGCGCGCGACGTCG GTCGCGGC (SEQ ID NO: 87)	BdDHSnc	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2183 BdDHSnc Rvs	CGGAGCTCGAATTCGGA TCCTTAAGCTTCTACTCT AGATATCAAGC (SEQ ID NO: 88)	BdDHSnc	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2342 SbDHS1a Fwd	GGTGCCGCGCGGCAGCC ATATGGCCATCCACGCC GCCGACCC (SEQ ID NO: 89)	SbDHS1a	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2343 SbDHS1a Rvs	CGGAGCTCGAATTCGGA TCCTCAGAAAGCCAGTG GTGGCAGC (SEQ ID NO: 90)	SbDHS1a	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2344 SbDHS2 Fwd	GGTGCCGCGCGGCAGCC ATATGCTCCGCGCCCGC GCCGTCC (SEQ ID NO: 91)	SbDHS2	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2345 SbDHS2 Rvs	CGGAGCTCGAATTCGGA TCCTCACAGACGAATGG AACCCAGC (SEQ ID NO: 92)	SbDHS2	In-Fusion cloning into NdeI/BamHI sites of pET28a
pHM2675 BdDHS3a GGF	TCACTCTGTGGTCTCAAA TGGCCGTGCACGCCGCC G (SEQ ID NO: 93)	BdDHS1a	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM2676 BdDHS3a GGR	CCACTTCGTGGTCTCACG AACTGAAGGCCAATGGC GGC (SEQ ID NO: 94)	BdDHS1a	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM2546 BdDHS3b GGF	TCACTCTGTGGTCTCAAA TGGCCGTCCACGCCGCG GAGCC (SEQ ID NO: 95)	BdDHS1b	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM2547 BdDHS3b GGF	CCACTTCGTGGTCTCACG AACTGAAACCATAGGTT GGCAATG (SEQ ID NO: 96)	BdDHS1b	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM2677 BdDHS3c GGF	TCACTCTGTGGTCTCAAA TGATCCGCGCGCACGCG G (SEQ ID NO: 97)	BdDHS2	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part
pHM2678 BdDHS3c GGR	CCACTTCGTGGTCTCACG AACTGAGTCCCATGGAT GATG (SEQ ID NO: 98)	BdDHS2	In-Fusion cloning into pAGM1287 BsaI sites as Golden Gate level 0 part

Enzyme Assays

[0098] TyrA assays were conducted in a plate reader at 37° C. (Tecan Infinite M Plex, Tecan) using half-area plates (Greiner Bio-One) by tracking the conversion of NAD (P)$^+$ into NAD (P) H as the increment of absorbance at 340 nm. TyrA reactions consisted of a final volume of 50 μL of 50 mM HEPES buffer pH 7.5, 50 mM KCl, 1 mM NADP$^+$ (NAD$^+$), and the enzyme (variable concentration, see details below). For IC_{50} assays, tyrosine from 10×-stocks adjusted to PH~10 with NaOH was included in the reaction mixture,

as tyrosine solubility is low at neutral pH. Enzyme concentration was adjusted using TyrA desalting buffer supplemented with bovine serum albumin (BSA, protease-free powder purified by heat shock process; Fisher bioreagents), to ensure at least 3 minutes of linear reaction. For arogenate-$NADP^+$ activity, the mass of enzyme was adjusted to 10 to 200 ng per reaction, depending on the specific activity of the TyrA isoform being tested. For assays using NAD^+ and/or prephenate, the enzyme mass per reaction was scaled up to 200-1,000 ng to increase sensitivity. The reactions mixtures with the enzyme and without substrate (arogenate or prephenate) were incubated at 37° C. for 3 minutes upon the addition of the substrate. The final concentration of substrate varied depending on the experiment. For determination of the enzyme substrate, 1 mM of prephenate or arogenate was used. For K_m and k_{cat} determination, variable concentrations of up to 2.5 mM of arogenate were used. For IC_{50} determination, 0.5 mM of arogenate was used.

[0099] Except when specified (FIG. **14**), DHS activity was measured using a real-time method by tracking the consumption of PEP at 232 nm[49] at 37° C. in a plate reader (Tecan Infinite M Plex, Tecan) in half-area UV-transparent 96-well plates (UV-Star®, Greiner Bio-One). DHS reactions consisted of a final volume of 50 μL of 25 mM HEPES buffer pH 7.5, 2 mM $MgCl_2$, 3 mM DTT, the enzyme (variable mass, see details below), the effector (if tested) and the substrates (PEP, E4P). All DHS effectors tested were included in the initial reaction mixtures at a concentration of 0.5 mM, except for IC_{50} determination for arogenate and tryptophan, in which variable concentrations were used. To ensure at least 10 minutes of linear reaction, enzyme mass per reaction was carefully adjusted to between 100 to 300 ng (depending on the specific activity of each specific isoform) using DHS storage buffer supplemented with BSA. The reaction mixtures, having the enzyme and all the other components except PEP nor E4P, were incubated for 5 minutes at room temperature to allow for DTT-mediated activation of DHS. After this, PEP (variable concentration, see below) was mixed into the reaction, and a second incubation step of 5 minutes at 37° C. was performed. The enzymatic reaction was started with the addition of E4P (variable concentration, see below). The concentrations of the substrates, PEP and E4P, were varied depending on the specific experiment. For testing potential feedback inhibitors and determination of IC_{50}, 1.5 mM PEP and 2 mM E4P was used. For calculating K_m and k_{cat} for PEP, a fixed concentration of 2 mM E4P and variable concentrations of PEP (up to 2 mM) were used. For calculating EC_{50} and k_{cat} for E4P, a fixed concentration of 1.5 mM PEP and variable concentrations of E4P (up to 3 mM) were used.

[0100] For DHS effector molecules overlapping with PEP absorbance in the UV range, PEP-based quantification of DHS activity was contrasted by a final-point quantification of the reaction product DAHP by UHPLC-MS. The DHS assay for UHPLC-MS quantification was set up using the same settings as described in the previous paragraph for the UV-based DHS assay, which guaranteed at least 10 minutes of reaction linearity. After a 10-minute incubation, 20 μL of the reactions (out of a total volume of 50 μL) were mixed into 80 μL of methanol, vortexed and transferred to vials for injection. Analysis of DAHP by UHPLC-MS was conducted using the same chromatographic settings as described for the UHPLC-MS analysis of soluble metabolites and compared with an authentic DAHP standard (Sta. Cruz biotechnology, cat. no. sc-216432).

[0101] Kinetic parameters of both TyrA and DHS enzymes were determined in MS-Excel using the Solver add-in function. Arogenate was prepared by enzymatic conversion from prephenate (Prephenate Barium salt, Sigma-Aldrich) as previously described[66].

Transient Expression Experiments in *Nicotiana benthamiana*

[0102] *Agrobacterium tumefaciens* strain GV3101 transformed with the plant expression constructs were grown at 28° C. for 24 to 36 hours in 10 mL of LB liquid media containing the corresponding antibiotics. The saturated cultures were spun down at 3,000 g for 5 minutes at room temperature and washed twice with 3 mL of induction media (IM; 10 mM MES [2-(N-morpholino) ethanesulfonic acid] buffer pH 5.6, 0.5% glucose, 2 mM NaH_2PO_4, 20 mM NH_4Cl, 1 mM $MgSO_4$, 2 mM KCl, 0.1 mM $CaCl_2$), 0.01 mM $FeSO_4$, and 0.2 mM acetosyringone). After washing, bacteria cultures were incubated in IM for 2 to 3 hours at room temperature in the dark, pelleted at 3000 g for 5 minutes, and resuspended into 3 mL of 10 mM MES buffer pH 5.6 with 0.2 mM acetosyringone. OD_{600nm} was adjusted to a final density of 0.25 units for pAGM4673::TyrA (FIG. **3**F) and pICH47822::DHS (FIG. **5**A) infiltration, or 0.5 units for pAGM4673::TyrA-DHS co-expression constructs (FIG. **5**B) using 10 mM MES buffer pH 5.6 with 0.2 mM acetosyringone. For infiltration of pICH47822::DHS constructs, the *Agrobacterium* suspensions were adjusted to OD_{600nm}=0.5 and mixed with an equal volume of a suspension of an *Agrobacterium* clone transformed with pICH4780::p19 under control of the *Arabidopsis* ubiquitin ligase promoter adjusted to OD_{600nm}=0.5, resulting in a final mixture of 0.25 OD_{600nm} units for each construct. The inclusion of the p19 gene silencing suppressor was found to improve transient expression of grass DHS genes in *Nicotiana benthamiana. Nicotiana benthamiana* plants of around 4-weeks-old were infiltrated close to the end of the light period into four different spots per plant, distributed into two leaves at two infiltrations per leaf, with each individual spot corresponding with a different construct/treatment. In total, each construct was infiltrated as 5 or 6 independent replicates into different plants following a randomized pattern. Samples consisting of the infiltrated leaf limbs, without the main veins, were harvested at ~72 hours post infiltration and subjected to HPLC or UHPLC-MS analysis (see details below).

$^{13}CO_2$ Feeding

[0103] *Brachypodium, Setaria*, and *Arabidopsis* plants were grown in 2.5×2.5-inch pots and randomly distributed into a plexiglass labeling chamber of approximately 32 liters of total volume. The artificial air mixture containing 79% N_2, 21% O_2, and 0.040% (400 ppm) $^{13}CO_2$ was pumped at a normal flow rate of 2 liters per minute. The air flow was connected 15 minutes before the beginning of the light period. For sampling, the air flow was interrupted, and the plant samples (entire plants for the experiment represented in FIG. **1**; fully expanded leaves and stem tissue for the experiment represented in FIG. **2**) were harvested and frozen immediately in liquid nitrogen. Feeding was resumed by reconnecting the air flow at 10 liters per minute with no $^{13}CO_2$ included in the mixture to quickly purge atmospheric $^{12}CO_2$. After 5 minutes, the flow rate was resumed at 2 liters

per minute, ~400 ppm $^{13}CO_2$, and kept constant until the next sampling time. Light intensity and temperature during the experiment were ~ 100 μE and 22° C., respectively.

UHPLC-MS/MS Analysis of Metabolites

[0104] Around 30-40 mg of pulverized frozen plant tissue were resuspended into 400 μL of chloroform:methanol (1:2) for ~1 hour with regular vortexing, followed by centrifugation at 20,000 g for 5 minutes at room temperature. The whole supernatant was transferred to a fresh tube, mixed with 125 μL of chloroform, 300 μL of water, and spun down at 15,000 g for 5 minutes for phase separation. The upper, aqueous phase was recovered and dried down for 4 hours to overnight in a speed-vac at 40° C. The dried pellets were resuspended into 100 μL of methanol 80%, spun down at 20,000 g for 5 minutes, and the supernatant transferred to vials for injection. All reagents used for the extraction were UHPLC-MS grade.

[0105] Aromatic amino acids and shikimate were detected using a Vanquish Horizon Binary UHPLC (Thermo Scientific) coupled to a Q Exactive mass spectrometer (Thermo Scientific). Two microliters of the sample were analyzed using a InfinityLab Poroshell 120 HILIC-Z column (150× 2.1 mm, 2.7-μm particle size; Agilent) in a gradient of 5 mM ammonium acetate/0.2% acetic acid buffer in water (solvent A) and 5 mM ammonium acetate/0.2% acetic acid buffer in 95% acetonitrile (solvent B) at a flow rate of 0.45 mL/min and column temperature of 40° C. The phase B gradient was: 0-2 min, 94%; 2-9 min, 94-88%; 9-19 min, 88-71%; 19-20 min, 71-20%, 20-21.5 min, 20%; 21.5-22 min, 20-94%; 22-25 min, 94%. All chemicals used to prepare the mobile phases were LC-MS grade. Full MS spectra were recorded between 2 and 19 minutes using full scan in negative mode, under the following parameters: sheath gas flow rate, 55; auxiliary gas flow rate, 20; sweep gas flow rate, 2; spray voltage, 3 kV; capillary temperature, 400° C.; S-lens RF level, 50; resolution, 70,000; AGC target 3×10^6, maximum scan time 100 ms; scan range 70-1050 m/z. Spectral data were integrated manually using Xcalibur 3.0. For ^{13}C labeled plant samples, ^{13}C-isotopologues were detected based on a mass increase of 1.00335 atomic mass units for each ^{13}C atom. Compound abundance was calculated based on high purity standards: Amino Acid Standard H for tyrosine and phenylalanine (Thermo Scientific, cat. no. PI20088), and shikimic acid ≥99% (Millipore Sigma, cat. no. S5375).

Determination of Tyrosine Content by HPLC in *Nicotiana benthamiana* Extracts

[0106] The infiltrated leaf areas, excluding the midrib and major veins, were harvested at 72 hours after infiltration and frozen immediately in liquid nitrogen. Then, 15 to 25 mg of pulverized frozen plant tissue were extracted into 400 μL of 0.5% 2-amino-2-methyl-1-propanol buffer pH 10.0 in 75% ethanol, as described before[67]. Plant extracts were analyzed by HPLC (Infinity 1260, Agilent, Santa Clara, CA) equipped with a Water's Atlantis T3 C18 column (3μ, 2.1×150 mm) using mobile phases of A (water with 0.1% formic acid) and B (acetonitrile with 0.1% formic acid) in a 20 min gradient of the mobile phase B: 0 to 5 min, 1% isocratic; 5 to 10 min, linear increase from 1% to 76%; 10 to 12 min, linear decrease from 76% to 1%; 12 to 20 min, 1% isocratic. A tyrosine peak was detected at the retention time of ~3.5 minutes using fluorescence detection mode (excitation wavelength 274 nm, emission wavelength 303 nm) and quantified with an authentic tyrosine standard (Alfa Acsar, catalog number AAA1114118).

TyrA Sequence Identification and Phylogenetic Analysis

[0107] TyrA and DHS protein sequences were downloaded from Phytozome[68] v13 using pBLAST search in the following genomes (species abbreviations between parenthesis): *Amaranthus hypochondriacus* v2.1 (AH), *Ananas comosus* v3 (Aco), *Arabidopsis thaliana* TAIR10 (At), *Beta vulgaris* EL10_1.0 (EL), *Chlamydomonas reinhardtii* v5.6 (Cre), *Cucumis sativus* v1.0 (Cucsa), *Brachypodium distachyon* v3.2 (Bd), *Gossypium raimondii* v2.1 (Gorai), *Hordeum vulgare* Morex v3 (HORVU), *Marchantia polymorpha* v3.1 (Mapoly), *Medicago truncatula* Mt4.0v1 (Medtr), *Musa acuminata* v1 (GSMUA), *Oropetium thomaeum* v1.0 (*Oropetium*), *Oryza sativa* v7.0 (LOC_Os), *Panicum virgatum* v5.1 (Pavir), *Phaseolus vulgaris* v2.1 (Phvul), *Populus trichocarpa* v4.1 (Potri), *Selaginella moellendorffii* v1.0 (Selmo), *Setaria viridis* v2.1 (Sevir), *Solanum lycopersicum* ITAG4.0 (Solyc), *Spinacia oleracea* Spov3 (Spov), *Sorghum bicolor* v3.1.1 (Sb), *Zea mays* RefGen_V4 (Zm), *Zostera marina* v3.1 (Zosma). Protein sequences without the putative plastid transit peptide were aligned using MUSCLE in MEGA-1169. Phylogenies were reconstructed in MEGA-11 using the Neighbor-Joining method and a site coverage cutoff was set at 90%. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. Bootstrap values were calculated based on 1000 replications.

Cloning of TyrA and DHS Genes into pET28a

[0108] Plant total RNA used for cloning was extracted from young leaf tissue using the CTAB/LiCl method[71] with modifications[72]. cDNA was synthesized with SuperScript IV VILO Master Mix (Thermo Scientific) following the manufacturer's instructions.

[0109] All genes were cloned without the predicted plastid transit peptide (TargetP v2.0 server, DTU Health Tech) using specific primers listed in Table 3. TyrA1 and TyrA2 genes were directly cloned from genomic DNA, as these genes lack introns. Grass TyrAnc genes were cloned from cDNA. DHS genes from *Brachypodium* were cloned from cDNA. SbDHS1a and SbDHS2 were cloned from *Sorghum bicolor* cDNA. SbDHS1b and SbDHSnc were synthesized into the pET28a vector (GeneArt, Thermo-Fisher). All cloning PCRs were conducted using high fidelity DNA polymerase (PrimeSTAR Max DNA polymerase, Takara Bio). PCR amplicons were purified from gel using QIAquick gel extraction kit (QIAGEN) and cloned into the pET28a vector between the NdeI and BamHI sites by In-Fusion cloning (Clontech). All cloned genes were confirmed by Sanger sequencing.

Plant Expression Constructs

[0110] For transient expression of TyrA genes of *Brachypodium* in *Arabidopsis* protoplasts, the full-length CDSs, without stop codon, were amplified by PCR from cDNA (BdTyrAnc) or genomic DNA (BdTryA1 and BdTyrA2, which lack introns) using corresponding gene-specific primers (Table 3). cDNA was prepared as described for pET28a constructs. The PCR fragments were purified from gel and inserted into the vector backbone pML94 at KpnI and NotI

sites, using the In-Fusion cloning (Clontech). The constructs were confirmed by restriction digestion and Sanger sequencing.

[0111] For TyrA expression in *Nicotiana benthamiana* under control of CaMV 35S promoter, the TyrA genes were amplified from pET28a constructs and assembled into a modified version of the binary vector pAGM4673 (Addgene plasmid #48014, courtesy of Sylvestre Marillonnet[70] (FIG. **13**) using BsaI sites introduced downstream of the CaMV 35S promoter. The plastid transit peptide from the enzyme 3-enol-pyruvyl-shikimate-3-phosphate synthase from *Petunia×hybrida* was used to target the TyrA proteins into the plastid[73]. GUS, AtTyrA2 and BvTyrAα expression vectors were assembled by Dr. Ray Collier.

[0112] For the simultaneous expression of *Brachypodium* TyrA and DHS genes in *Nicotiana benthamiana*, the genes were first cloned into the level 0 backbone pAGM1287 (Addgene plasmid #47996, courtesy of Sylvestre Marillonnet[70] by In-Fusion cloning (Clontech). The level 0 modules were assembled into the level 1 binary vector pICH47831 for TyrA enzymes, or into pICH47822 for DHS enzymes (Addgene plasmids #48009 and #48010, courtesy of Sylvestre Marillonnet[70] as illustrated in FIG. **13**, using the MoClo Plant Parts Kit (Addgene Kit #1000000047, courtesy of Nicola Patron[51]). The level 1 modules were then transferred into the level 2 binary backbone pAGM4673 (FIG. **13**). All constructs were checked by restriction digestion and Sanger sequencing prior to being transformed into *Agrobacterium tumefaciens GV*3101. All primers used are listed in Table 3.

Protein Expression and Purification

[0113] Recombinant proteins were produced using the *E. coli* strains Rosetta-2 (DE3) (Millipore Sigma) for TyrAnc enzymes, ArcticExpress (Agilent) for TyrA1 and TyrA2 enzymes, and KRX (Promega) for DHS enzymes. In all cases, starter cultures were grown overnight at 37° C., 200 rpm in 10 mL terrific broth (TB) medium containing the corresponding pET28a antibiotic (50 μg/mL kanamycin) and 0.1% glucose. The next day, flasks containing 200 or 400 mL of TB medium with 50 μg/ml kanamycin and without glucose were inoculated with a 1:100 dilution of the starter cultures and kept at 37° C., 200 rpm, until the $OD_{600\ nm}$ reached ~0.5-0.6. For TyrAnc production in Rosetta-2, the cultures were cooled down to room temperature for ~15 minutes, induced with 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG), and kept at 22° C., 200 rpm, for 8 to 10 hours. For production of DHS proteins in KRX, the cultures were cooled down to room temperature for ~15 minutes, induced with 0.5 mM IPTG and 0.1% rhamnose, and kept at 22° C., 200 rpm, for 16-20 hours. For production of TyrA1 and TyrA2, ArcticExpress cultures were cooled down in a mixture of water and ice for ~10 minutes, induced with 0.5 mM IPTG and kept at 15° C., 200 rpm, for 16 to 20 hours. All cultures were pelleted at 5000 g for 10 minutes and stored at −80° C. until purification.

[0114] Frozen bacterial pellets were thawed on ice and resuspended into 2 to 4 mL of LEW buffer (Lysis-Equilibration-Washing buffer; 50 mM sodium phosphate buffer pH 8.3, 300 mM NaCl 300 mM and 10% v/v glycerol) supplemented with 1 mM PMSF and 1 mg/mL lysozyme and sonicated on ice for 5 minutes in 30 second cycles. Cell lysate was centrifuged at 15,000 g, 4° C., for 15 minutes. The supernatant was recovered, mixed with 100 μL of PureProteome Nickel Magnetic Beads (Millipore) previously washed with LEW buffer, and kept in the cold under gentle shaking for 30 minutes for binding. After that, the magnetic beads were washed twice with 1 mL of LEW buffer. Proteins were eluted with LEW buffer with 250 mM imidazole into four fractions of 100 μL each. The fraction(s) with the highest protein concentration (usually two) were combined and exchanged into the corresponding storage buffer using Sephadex G-50 resin (GE Healthcare): for TyrA proteins, 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer pH 7.5, 50 mM KCl, 10% glycerol, and 1 mM dithiothreitol (DTT); for DHS proteins, 50 mM HEPES buffer pH 7.5, 300 mM NaCl, and 0.2% Triton X-100. In the case of DHS proteins, keeping NaCl concentration at >150 mM in the storage buffer was found to be critical to prevent protein precipitation. Buffer-exchanged proteins were frozen immediately in liquid nitrogen and stored at −80° C. The concentration of total protein was determined using Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-Rad). Purity level of the recombinant enzymes was determined in ImageJ (v1.52a) upon staining of SDS-PAGE gel with Coomassie Brilliant Blue R250. Enzymatic assays were carried out within no longer than 2 weeks of protein storage at −80° C., although we found many TyrA and DHS enzymes to be stable for longer periods (i.e., a few months) under these conditions.

Plastid Targeting Assay in *Arabidopsis* Protoplasts

[0115] Localization studies for BdTyrA1, BdTyrA2, and BdTyrAnc were performed in *Arabidopsis* protoplasts using c-terminal fusion to EGFP. Plasmid DNA was isolated from *E. coli* cell cultures with the PureYield™ Plasmid Maxiprep System (Promega). Protoplasts were isolated from two-weeks-old *Arabidopsis thaliana* leaves, transfected with plasmid DNA, and incubated for 16 hours to allow for protein expression and maturation. Samples were analyzed by laser scanning confocal microscopy using a Zeiss LSM 780 ELYRA PS1. The light path included a 488 nm laser, a 561 nm laser, and a 488/561 dichroic mirror. Fluorescence was detected in two tracks in the range of 578 nm-696 nm and 493 nm-574 nm to record chlorophyll autofluorescence and EGFP signal, respectively. All images were captured with an LDC-Apochromat 40×/1.1 W Korr M27 objective. Images were processed using Zen software (Zeiss).

Extraction of Plant Proteins and Western Blot

[0116] Total protein from *Nicotiana benthamiana* samples was extracted from ~10 mg of pulverized frozen tissue into 75 μL of 1× denaturing protein sample buffer (60 mM Tris [tris(hydroxymethyl)aminomethane] buffer pH 6.8, 2% sodium dodecyl sulfate, 10% glycerol, 3% β-mercaptoethanol, and 0.01% bromophenol blue) by vigorous vortexing for 30 seconds and was boiled immediately at 95° C. for 7 minutes. Tubes were centrifuged at 15,000 g for 5 minutes and 5 μL of the supernatant were applied per lane to the SDS-PAGE gel. Proteins were transferred to a PVDF membrane and blocked for 1 hour in 5% skimmed milk in Tris Saline Buffer with 0.05% Tween-20 before incubation with the corresponding antibodies. HA-tagged fusion proteins were detected using an anti-HA tag monoclonal antibody conjugated to HRP at a 1:1,000 dilution (HA-Probe HRP conjugated mouse monoclonal antibody clone F-7, Sta. Cruz Biotechnology, cat. no. SC-7392). Antibody dilutions were

prepared in Tris Buffer Saline with 0.05% Tween-20 and 0.5% BSA. Immunoblot signal was quantified in non-saturating conditions using ImageJ (version 1.52a) and pure recombinant BdDHS1b-3×HA as a standard, which was mixed with total protein extracts of not-infiltrated *Nicotiana* leaves to ensure homogenous transfer for all lanes. Independent western blot membranes were exposed in parallel to ensure quantitative results. For details about the generation of the recombinant protein standards, see the section of the Materials and Methods titled "Protein expression and purification".

REFERENCES

[0117] 1. Maeda H A, and Dudareva N (2012) The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. Annu Rev Plant Biol, 63:73-105.

[0118] 2. Lynch J H, and Dudareva N (2020) Aromatic Amino Acids: A Complex Network Ripe for Future Exploration. Trends Plant Sci, 25:670-681.

[0119] 3. Boerjan W, Ralph J, and Baucher M (2003) Lignin Biosynthesis. Annu Rev Plant Biol, 54:519-546.

[0120] 4. Cochrane F C, Davin L B, and Lewis N G (2004) The *Arabidopsis* Phenylalanine Ammonia Lyase Gene Family: Kinetic Characterization of the Four PAL Isoforms. Phytochemistry, 65:1557-1564.

[0121] 5. Barros J, and Dixon R A (2020) Plant Phenylalanine/Tyrosine Ammonia-Lyases. Trends Plant Sci, 25:66-79.

[0122] 6. Neish A C (1961) Formation of m- and p-Coumaric Acids by Enzymatic Deamination of the Corresponding Isomers of Tyrosine. Phytochemistry, 1:1-24.

[0123] 7. Young M R, Towers G H N, and Neish A C (1966) Taxonomic Distribution of Ammonia-lyases for L-Phenylalanine and L-Tyrosine in Relation to Lignification. Can J Bot, 44:341:349

[0124] 8. Reid P D, Havir E A, and Marsh H V (1972) L-Phenylalanine Ammonia-Lyase (Maize). Plant Physiol, 50:480-484.

[0125] 9. Rösler J, Krekel F, Amrhein N, and Schmid J (1997) Maize Phenylalanine Ammonia-lyase has Tyrosine Ammonia-Lyase Activity. Plant Physiol, 113:175-179.

[0126] 10. Barros J, Serrani-Yarce J C, Chen F, Baxter D, Venables B J, and Dixon R A (2016) Role of Bifunctional Ammonia-lyase in Grass Cell Wall Biosynthesis. Nat Plants, 2:16050.

[0127] 11. Jun S Y, Sattler S A, Cortez G S, Vermerris W, Sattler S E, Kang C (2018) Biochemical and Structural Analysis of Substrate Specificity of a Phenylalanine Ammonia-Lyase. Plant Physiol, 176:1452-1468.

[0128] 12. Higuchi T, Ito Y, and Kawamura I (1967). p-Hydroxyphenylpropane Component of Grass Lignin and Role of Tyrosine Ammonia-Lyase in its Formation. Phytochemistry, 6:875-881.

[0129] 13. Simpson J P, Olson J, Dilkes B, and Chapple C (2021) Identification of the Tyrosine- and Phenylalanine-Derived Soluble Metabolomes of Sorghum. Front Plant Sci, 12:714164.

[0130] 14. Tzin V, and Galili G (2010) New Insights into the Shikimate and Aromatic Amino Acids Biosynthesis Pathways in Plants. Mol Plant, 3:956-972.

[0131] 15. Yokoyama R, de Oliveira M V V, Kleven B, and Maeda H A (2021) The Entry Reaction of the Plant Shikimate Pathway is Subjected to Highly Complex Metabolite-Mediated Regulation. Plant Cell, 33:671-696.

[0132] 16. Yokoyama R, de Oliveira M V V, Takeda-Kimura Y, Ishihara H, Alseekh S, Arrivault S, Kukshal V, Jez J M, Stitt M, Fernie A R, and Maeda H A (2022) Point Mutations that Boost Aromatic Amino acid Production and CO_2 Assimilation in Plants. Sci Adv, 8: eabo3416.

[0133] 17. Byng G S, Whitaker R J, Flick C, and Jensen R A (1981) Enzymology of L-Tyrosine Biosynthesis in Corn (*Zea mays*). Phytochem, 208:1289-1292.

[0134] 18. Gaines C G, Byng G S, Whitaker R J, and Jensen R A (1982) L-Tyrosine Regulation and Biosynthesis via Arogenate Dehydrogenase in Suspension-Cultured cells of *Nicotiana silvestris*. Planta, 1568:233-240.

[0135] 19. Connelly J A, and Conn E E (1986) Tyrosine Biosynthesis in *Sorghum bicolor*: Isolation and Regulatory Properties of Arogenate Dehydrogenase. Z Naturforsch C, 41:69-78.

[0136] 20. Rippert P, and Matringe M (2002) Purification and Kinetic Analysis of the Two Recombinant Arogenate Dehydrogenase Isoforms of *Arabidopsis thaliana*. Eur J Biochem, 269:4753-4761.

[0137] 21. Yoo H, Shrivastava S, Lynch J H, Huang X Q, Widhalm J R, Guo L, Carter B C, Qian Y, Maeda H A, Ogas J P, Morgan J A, Marshall-Colón A, Dudareva N (2021) Overexpression of Arogenate Dehydratase Reveals an Upstream Point of Metabolic Control in Phenylalanine Biosynthesis. Plant J, 108:737-751.

[0138] 22. El-Azaz J, Cánovas F M, Barcelona B, Ávila C, and de la Torre F (2022) Deregulation of Phenylalanine Biosynthesis Evolved with the Emergence of Vascular Plants. Plant Physiol, 188:134-150.

[0139] 23. Herrmann, K M (1995) The Shikimate Pathway: Early Steps in the Biosynthesis of Aromatic Compounds. Plant Cell, 7:907-919.

[0140] 24. Yokoyama R, Kleven B, Gupta A, Wang Y, and Maeda H A (2022) 3-Deoxy-D-Arabino-Heptulosonate 7-Phosphate Synthase as the Gatekeeper of Plant Aromatic Natural Product Biosynthesis. Curr Opin Plant Biol, 67:102219.

[0141] 25. Verdonk J C, Haring M A, van Tunen A J, and Schuurink R C (2005) ODORANT1 Regulates Fragrance Biosynthesis in Petunia Flowers. Plant Cell, 17:1612-1624.

[0142] 26. Maeda H A, Yoo H, and Dudareva N (2011) Prephenate Aminotransferase Directs Plant Phenylalanine Biosynthesis via Arogenate. Nat Chem Biol, 7:19-21.

[0143] 27. Dal Cin V, Tieman D M, Tohge T, McQuinn R, de Vos R C, Osorio S, Schmelz E A, Taylor M G, Smits-Kroon M T, Schuurink R C, Haring M A, Giovannoni J, Fernie A R, and Klee H J (2011) Identification of Genes in the Phenylalanine Metabolic Pathway by Ectopic Expression of a MYB Transcription Factor in Tomato Fruit. Plant Cell, 23:2738-2753.

[0144] 28. Corea O R, Ki C, Cardenas C L, Kim S J, Brewer S E, Patten A M, Davin L B, and Lewis N G (2012) Arogenate Dehydratase Isoenzymes Profoundly and Differentially Modulate Carbon Flux into Lignins. J Biol Chem, 287:11446-11459.

[0145] 29. Obayashi T, Aoki Y, Tadaka S, Kagaya Y, and Kinoshita K (2018) ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. Plant Cell Physiol, 59: e3.

[0146] 30. Voll L, Häusler R E, Hecker R, Weber A, Weissenböck G, Fiene G, Waffenschmidt S, and Flügge

U-I (2003) The Phenotype of the *Arabidopsis* cue1 Mutant is not Simply Caused by a General Restriction of the Shikimate Pathway. Plant J, 36:301-317.

[0147] 31. Voll L M, Allaire E E, Fiene G, and Weber A P (2004) The *Arabidopsis* Phenylalanine Insensitive Growth Mutant Exhibits a Deregulated Amino Acid Metabolism. Plant Physiol, 136:3058-3069.

[0148] 32. Mancla N, Oliva M, Ovadia R, Sikron-Persi N, Ayenew B, Fait A, Galili G, Perl A, Weiss D, and Oren-Shamir M (2015) Phenylalanine and Tyrosine Levels are Rate-Limiting Factors in Production of Health Promoting Metabolites in *Vitis vinifera* cv. Gamay Red Cell Suspension. Front Plant Sci, 6:538.

[0149] 33. Qian Y, Lynch J H, Guo L, Rhodes D, Morgan J A, and Dudareva N (2019) Completion of the Cytosolic Post-Chorismate Phenylalanine Biosynthetic Pathway in Plants. Nat Commun, 10:15.

[0150] 35. Higuchi T, and Shimada M (1969) Metabolism of Phenylalanine and Tyrosine During Lignification of Bamboos. Phytochemistry, 8:1185-1192.

[0151] 36. Robinson P W, and Hodges C F (1977) Effect of Nitrogen Fertilization on Free Amino Acid and Soluble Sugar Content of *Poa pratensis* and on Infection and Disease Severity by *Drechslera sorokiniana*. Phytopathology, 67:1239-1244.

[0152] 37. Seebauer J R, Moose S P, Fabbri B J, Crossland L D, and Below F E (2004) Amino Acid Metabolism in Maize Earshoots. Implications for Assimilate Preconditioning and Nitrogen Signaling. Plant Physiol, 136:4326-4334.

[0153] 38. de Souza L P, Szecówka M, Fernie A R, and Tohge T (2017) $^{13}CO_2$ Labeling and Mass Spectral Analysis of Photorespiration. In: Fernie A, Bauwe H, and Weber A (eds) Photorespiration. Methods in Molecular Biology, vol 1653. Humana Press, New York, N Y.

[0154] 39. Schenck C A, and Maeda H A (2018) Tyrosine Biosynthesis, Metabolism, and Catabolismin Plants. Phytochemistry, 149:82-102.

[0155] 40. Schenck C A, Chen S, Siehl D L, and Maeda H A (2015) Non-plastidic, Tyrosine-Insensitive Prephenate Dehydrogenases from Legumes. Nat Chem Biol, 11:52-57.

[0156] 41. Ruprecht C, Mendrinna A, Tohge T, Sampathkumar A, Klie S, Fernie A R, Nikoloski Z, Persson S. and Mutwil M. (2016) FamNet: a Framework to Identify Multiplied Modules Driving Pathway Expansion in Plants. Plant Physiol, 170:1878-1894.

[0157] 42. Makita Y, Shimada S, Kawashima M, Kondou-Kuriyama T, Toyoda T, and Matsui M (2015) MOROKOSHI: Transcriptome Database in *Sorghum bicolor*. Plant Cell Physiol, 56: e6.

[0158] 43. Kellogg E A (2001) Evolutionary History of the Grasses. Plant Physiol, 125:1198-1205.

[0159] 44. Soreng R J, Peterson P M, Romaschenko K, Davidse G, Zuloaga F O, Judziewicz E J, Filgueiras T S, Davis J I, and Morrone O (2015) A Worldwide Phylogenetic Classification of the Poaceae (Gramineae) J Syst Evol, 53:117-137.

[0160] 45. Lopez-Nieves S, Yang Y, Timoneda A, Wang M, Feng T, Smith S A, Brockington S F, and Maeda H A (2018) Relaxation of Tyrosine Pathway Regulation Underlies the Evolution of Betalain Pigmentation in Caryophyllales. New Phytol, 217:896-908.

[0161] 46. Schenck C A, Holland C K, Schneider M R, Men Y, Lee S G, Jez J M, and Maeda H A (2017) Molecular Basis of the Evolution of Alternative Tyrosine Biosynthetic Routes in Plants. Nat Chem Biol, 13:1029-1035.

[0162] 47. Hudson G S, Wong V, and Davidson B E (1984) Chorismate Mutase/Prephenate Dehydrogenase from *Escherichia coli* K12: Purification, Characterization, and Identification of a Reactive Cysteine. Biochemistry, 23:6240-6249.

[0163] 48. Lütke-Eversloh T, and Stephanopoulos G (1985) Feedback Inhibition of Chorismate Mutase/Prephenate Dehydrogenase (TyrA) of *Escherichia coli*: Generation and Characterization of Tyrosine-Insensitive Mutants. Appl Environ Microbiol, 71:7224-7228.

[0164] 49. Schoner R, and Herrmann K M (1976) 3-Deoxy-D-Arabino-Heptulosonate 7-Phosphate Synthase. Purification, Properties, and Kinetics of the Tyrosine-Sensitive Isoenzyme from *Escherichia coli*. J Biol Chem, 251:5440-5447

[0165] 50. Zamir L O, Jense R A, Arison B H, Douglas A W, Albers-Schoenberg G, and Bowen J R (1980) Structure of Arogenate (pretyrosine), an Amino Acid Intermediate of Aromatic Biosynthesis. J Am Chem Soc, 102:4499-4504

[0166] 51. Webby C J, Jiao W, Hutton R D, Blackmore N J, Baker H M, Baker E N, Jameson G B, and Parker E J (2010) Synergistic Allostery, a Sophisticated Regulatory Network for the Control of Aromatic Amino Acid Biosynthesis in *Mycobacterium tuberculosis*. J Biol Chem, 285:30567-30576.

[0167] 52. Engler C, Youles M, Gruetzner R, Ehnert T-M, Werner S, Jones J D G, Patron N J, and Marillonnet S (2014) A Golden Gate Modular Cloning Toolbox for Plants. ACS Synthetic Biology, 3:839-843.

[0168] 53. Vogt T (2010) Phenylpropanoid Biosynthesis. Mol Plant, 3:2-20.

[0169] 54. Maeda H A, and Fernie, A R (2021) Evolutionary History of Plant Metabolism. Annu Rev Plant Biol, 72:185-216.

[0170] 55. Maeda, H (2016) Lignin biosynthesis: Tyrosine Shortcut in Grasses. Nat Plants 2, 16080.

[0171] 56. Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, Cumming M, Kelley L A, Sternberg M J, Krishnakumar V, Ferlanti E, Miller J, Town C, Stuerzlinger W, and Provart N J (2017) ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. Plant Cell, 29:1806-1821.

[0172] 57. de Oliveira M V V, Jin X, Chen X, Griffith D, Batchu S, and Maeda H A (2019) Imbalance of Tyrosine by Modulating TyrA Arogenate Dehydrogenases Impacts Growth and Development of *Arabidopsis thaliana*. Plant J, 97:901-922.

[0173] 58. Lopez-Nieves S, El-Azaz J, Men Y, Holland C K, Feng T, Brockington S F, Jez J M, and Maeda H A (2022) Two Independently Evolved Natural Mutations Additively Deregulate TyrA Enzymes and Boost Tyrosine Production in planta. Plant J, 109:844-855.

[0174] 59. Tzin V, Malitsky S, Zvi M M B, Bedair M, Sumner L, Aharoni A, and Galili G (2012) Expression of a Bacterial Feedback-Insensitive 3-Deoxy-D-Arabino-Heptulosonate 7-Phosphate Synthase of the Shikimate Pathway in *Arabidopsis* Elucidates Potential Metabolic

Bottlenecks Between Primary and Secondary Metabolism. New Phytol, 194:430-439.

[0175] 60. Oliva M, Ovadia R, Perl A, Bar E, Lewinsohn E, Galili G, and Oren-Shamir M (2015) Enhanced Formation of Aromatic Amino Acids Increases Fragrance without Affecting Flower Longevity or Pigmentation in *Petunia×hybrida*. Plant Biotechnol J, 13:125-136.

[0176] 61. Oliva M, Guy A, Galili G, Dor E, Schweitzer R, Amir R, and Hacham Y (2021) Enhanced Production of Aromatic Amino Acids in Tobacco Plants Leads to Increased Phenylpropanoid Metabolites and Tolerance to Stresses. Front Plant Sci, 11:604349.

[0177] 62. Tohge T, Watanabe M, Hoefgen R, and Fernie A R (2013) Shikimate and Phenylalanine Biosynthesis in the Green Lineage. Front Plant Sci, 4:62.

[0178] 63. Graziana A, and Boudet A M (1980) 3-Deoxy-D-Arabino-Heptulosonate 7-Phosphate Synthase from *Zea mays*: General Properties and Regulation by Tryptophan. Plant Cell Physiol, 21:793-802.

[0179] 64. Jung E, Zamir L O and Jensen R A (1986) Chloroplasts of Higher Plants Synthesize L-Phenylalanine via L-Arogenate. PNAS, 83:7231-7235.

[0180] 65. Ruijter J M, Ramakers C, Hoogaars W H M, Karlen Y, Bakker O, van den Hoff M J B, and Moorman A F M (2009) Amplification Efficiency: Linking Baseline and Bias in the Analysis of Quantitative PCR data. Nucleic Acids Res, 37: e45.

[0181] 66. Hong S Y, Seo P J, Yang M S, Xiang F, and Park C M (2008) Exploring Valid Reference Genes for Gene Expression Studies in *Brachypodium distachyon* by Real-Time PCR. BMC Plant Biol, 8:112.

[0182] 67. Maeda H A, Shasany A K, Schnepp J, Orlova I, Taguchi G, Cooper B R, Rhodes D, Pichersky E, and Dudareva N (2010) RNAi Suppression of Arogenate Dehydratase 1 Reveals that Phenylalanine is Synthesized Predominantly via the Arogenate Pathway in Petunia Petals. Plant Cell, 22:832-849.

[0183] 68. Goodstein D M, Shu S, Howson R, Neupane R, Hayes R D, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, and Rokhsar D S (2012) Phytozome: a Comparative Platform for Green Plant Genomics. Nucleic Acids Res, 40: D1178-D1186.

[0184] 69. Tamura K, Stecher G, and Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. Mol Biol Evol, 38:3022-3027

[0185] 70. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLOS One, 6: e16765.

[0186] 71. Liao Z, Chen M, Guo L, Gong Y, Tang F, Sun X, and Tang K (2004) Rapid Isolation of High-Quality total RNA from Taxus and Ginkgo. Pre Biochem Biotechnol, 34:209-214.

[0187] 72. Canales J, Rueda-López M, Craven-Bartle B, Avila C, and Cánovas F M (2012) Novel Insights into Regulation of Asparagine Synthetase in Conifers. Front Plant Sci, 3:100.

[0188] 73. Della-Cioppa G, Bauer S C, Klein B K, Shah D M, Fraley R T, and Kishore G M (1986) Translocation of the Precursor of 5-Enolpyruvylshikimate-3-Phosphate Synthase into Chloroplasts of Higher Plants in vitro. Proc Natl Acad Sci USA, 83:6873-6877.

Example 2

[0189] In the following example, the inventors demonstrate that transiently expressing DHS1b enzymes from three additional grass species (i.e., *Oryza sativa, Sorghum bicolor*, and *Setaria viridis*) in *Nicotiana benthamiana* leaves increases the production of aromatic amino acids and their precursors to varying degrees (FIG. **21**, FIG. **22**).

[0190] To determine whether the grass DHS1b gene encodes a deregulated DHS in other grass species besides *Brachypodium distachyon*, BdDHS1b orthologs from *Oryza sativa* (OsDHS1b) and *Sorghum bicolor* (SbDHS1b) were cloned into the Golden Gate plant expression vector pICH47822 under control of the RuBisCO small subunit 3B promoter from *Arabidopsis thaliana*.[52] A 3× human influenza hemagglutinin tag was fused in frame to the C-terminus of the DHS genes to confirm the production of heterologous DHS protein.

[0191] *Agrobacterium tumefaciens* (strain GV3101) clones transformed with the OsDHS1b and SbDHS1b plant expression constructs were infiltrated into the leaves of *Nicotiana benthamiana* side-by-side with *Agrobacterium* clones expressing BdDHS1a, BdDHS1b, and the fluorescent protein YPet (negative control). The levels of phenylalanine, tyrosine, tryptophan, and three of their biosynthetic intermediates (shikimate, prephenate, and arogenate) were determined at 72 hours post-infiltration by liquid chromatography coupled to mass spectrometry (LCMS). The results of this analysis showed that, as for BdDHS1b, the engineered plants expressing OsDHS1b had increased levels of all measured metabolites except for tryptophan (FIG. **21**). On average, expression of BdDHS1b or OsDHS1b increased phenylalanine, tyrosine, and shikimate levels by 5 to 20-times compared to the YPet negative control (FIG. **21**), whereas prephenate and arogenate rose to ~100-times the control (FIG. **21**). In contrast, expression of SbDHS1b had little or no significant effect on the levels of aromatic amino acid levels and their precursor molecules (FIG. **21**). Based on these findings, BdDHS1b and OsDHS1b, but not SbDHS1b, encode DHS enzymes that can boost the accumulation of aromatic amino acids in planta upon heterologous expression.

[0192] To further explore the distribution of deregulated DHS activity in grasses, this experimental approach was extended to the four members of the grass DHS gene family (DHS1a, DHS1b, DHS2, and DHSnc) (FIG. **4**A) from four different grass species: *Brachypodium distachyon* (Bd), *Oryza sativa* (Os), *Sorghum bicolor* (Sb), and *Setaria viridis* (Sv). These 16 genes, in addition of the aforementioned YPet-expressing negative control, were transiently expressed side-by-side in the leaves of *Nicotiana benthamiana* by the *Agrobacterium tumefaciens* infiltration method. LCMS was used to determine the levels of the three aromatic amino acids and their biosynthetic intermediates at 72 hours post-infiltration. This experiment showed that BdDHS1b caused the highest increase in the levels of these metabolites, followed by (from highest to lowest effect) OsDHS1b, SvDHS1b, and SbDHS1b (FIG. **22**). Overall, little or no significant effect was observed upon expression of any of the other three grass DHS isoforms (i.e., DHS1a, DHS2, and DHSnc) (FIG. **22**). Hence, DHS1b enzymes from different grass species have a varied impact on aromatic amino acid production in planta.

SEQUENCE LISTING

Sequence total quantity: 98
SEQ ID NO: 1 moltype = AA length = 537
FEATURE Location/Qualifiers
source 1..537
mol_type = protein
organism = Brachypodium distachyon
SEQUENCE: 1
MALAINSAAA AAAAALSNGG AAQPPRPRAG LLPMKRRGSI SAVHAADPAR GTGSVPAAAK 60
TSSPTVAPEA AASPAASWKV DSWKAKKALQ LPEYPSKEEL DTVLQTIETF PPIVFAGEAR 120
HLEERLAEAA MGRAFVLQGG DCAESFKEFN ANNIRDTFRV LLQMGAVLMF GGQVPVVKVG 180
RMAGQFAKPR SDNLEERDGV KLPSYRGDNV NGDAFDVKSR TPDPERMIRA YAQSVATLNL 240
LRAFATGGYA AMQRVIQWNL DFMDHNEQGD RYRELAHRVD EALGFMTAAG LGIDHPIMTT 300
TDFWTSHECL LLPYEQALTR EDSTSGLFYD CSAHMLWVGE RTRQLDGAHV EFLRGIANPL 360
GIKVSDKMNP AELVKLIEIL NPSNKPGRIT IITRMGAENM RVKLPHLIRA VRNSGQIVTW 420
ITDPMHGNTI KAPCGLKTRP FDSIMNEVRA FFDVHDQEGS HPGGIHLEMT GQNVTECIGG 480
SRTVTFDDLG DRYHTHCDPR LNASQSLELA FIIAERLRKR RMQSGLTNNL PLPPLAF 537

SEQ ID NO: 2 moltype = DNA length = 1614
FEATURE Location/Qualifiers
source 1..1614
mol_type = other DNA
organism = Brachypodium distachyon
SEQUENCE: 2
atggcgctcg ccatcaactc cgccgccgcc gcagcggccg ccgccctctc caacggcggg 60
gcagcccagc cgccgcgccc cagagcgggg ttactcccga tgaagaggcg gggcagcatc 120
tctgccgtgc acgccgccga ccccgccagg ggcaccgggt ccgtcccggc ggccgcgaag 180
acctcctccc cgacggtggc gcccgaggcg gcggcttctc cggcggcgag ctggaaggtc 240
gacagctgga aggcgaagaa ggccctgcag ctgcccgagt acccgagcaa ggaggagctc 300
gacaccgtgc tgcagaccat cgagacgttc ccgcccatcg tgttcgccgg ggaggcgcgc 360
cacctcgagg agcgcctcgc cgaggccgcc atgggccggg ctttcgtcct ccagggcggc 420
gactgcgccg agagcttcaa ggagttcaac gccaacaaca tccgtgacac cttccgtgtc 480
ctcctccaga tgggcgccgt cctcatgttc ggcggccagg tgcccgtcgt caaggtaggt 540
aggatggctg gccagttcgc caagcctagg tccgacaacc tcgaggagag ggacggggtt 600
aagctgccca gttacagggg agacaacgtc aacggcgacg ccttcgacgt caagagccgc 660
accccggacc ctgagaggat gatcagggcc tatgcccagt cggtggcgac gctcaacttg 720
ctccgcgcgt tcgccacagg aggctacgct gctatgcagc gggttatcca gtggaacctt 780
gatttcatgg atcacaacga gcagggagac aggtaccgtg agttggcaca tagggtggat 840
gaggcccttg gcttcatgac tgcagcaggg ctggggatcg atcatccgat aatgacaact 900
actgacttct ggacatccca tgagtgtctc ctcttgccat acgagcaggc tcttacccgt 960
gaggactcca caagtggcct tttctatgac tgctcggcac acatgctgtg ggttggtgag 1020
cgcacccggc aacttgatgg ggctcatgtt gagttcctcc gtggtattgc gaaccctctt 1080
ggcattaagg tgagtgataa aatgaaccct gctgaattgg tgaagctgat tgagatttta 1140
aacccttcaa acaagccggg aagaataacc ataattacaa ggatgggtgc tgagaacatg 1200
agggtgaagt tgcctcatct tatccgtgct gtccgcaatt ctggacagat tgttacatgg 1260
attactgatc ccatgcatgg gaacaccatc aaggccccat gtggtctgaa gactcgtcca 1320
tttgattcca ttatgaatga ggtgcgtgcg ttcttcgacg tgcatgacca agaaggaagc 1380
cacccaggag gtatccacct cgagatgacc ggacagaatg taaccgagtg catcggtggg 1440
tcacggaccg tgaccttcga tgacctgggt gaccgctacc acactcactg cgacccaagg 1500
ctgaatgctt cccaatctct ggagctcgcc ttcatcatcg ccgagagact caggaagagg 1560
aggatgcagt cagggctcac aaacaacctg ccactgccgc cattggcctt ctaa 1614

SEQ ID NO: 3 moltype = AA length = 542
FEATURE Location/Qualifiers
source 1..542
mol_type = protein
organism = Brachypodium distachyon
SEQUENCE: 3
MALATNHAAA AISSGAAAPQ PRRAPSFLPL KRRTICAVHA AEPSKSAAAA PAAAKTSSPS 60
VAPEKSAIPD PKPEAPAVPA KWTVDSWRAK KALQLPEYPN AAELESALKT IEAFPPIVFA 120
GEARHLEERL ADAAMGRAFL LQGGDCAESF KEFNGNNIRD TFRVLLQMSA VLTFGGQMPV 180
IKVGRMAGQF AKPRSDSFEV RDGVKLPSYR GDNINGDAFN EKSRIPDPQR MIRAYTQSAA 240
TLNLLRAFAM GGYAAMQRVT QWNLDFTENS EQGDRYRELA HRVDEALGFM SAAGLTLDHP 300
VMSSTEFWTS HECLLLPYEQ ALTRQDSTSG LFYDCSAHML WVGERTRQLD GAHVEFLRGV 360
ANPLGIKVSD KMNPADLVKL IDILNPTNKP GRITIITRMG AENMRVKLPH LIRAVRHAGQ 420
IVTWITDPMH GNTIKAPCGL KTRPFDSILA EVRAFFDVHE QEGSHAGGVH LEMTGQNVTE 480
CIGGSRTVTF DDLGDRYHTH CDPRLNASQS LELSFIIAEK LRKRRIRSSK LNSVLPLPTY 540
GF 542

SEQ ID NO: 4 moltype = DNA length = 1629
FEATURE Location/Qualifiers
source 1..1629
mol_type = other DNA
organism = Brachypodium distachyon
SEQUENCE: 4
atggcactcg ccaccaacca cgccgccgcc gccatatcgt ccggagccgc cgcccctcag 60

-continued

```
ccccgccgcg ccccttcgtt cctcccgctg aagcgccgca caatctgcgc cgtccacgcc  120
gcggagccct cgaagagcgc cgccgccgcc ccggccgcgg cgaagacctc ttcgccgtcg  180
gtggcgccgg agaagtccgc gataccggac ccgaaaccgg aggcgcccgc ggtgccagca  240
aagtggacgg tggacagctg gagggcgaag aaggctcttc agctgccgga gtaccccaac  300
gccgcggagc tggagtcggc cctcaagacg atcgaggcgt tcccgccgat cgtcttcgcc  360
ggggaggcgc ggcacctgga ggagcgcctc gccgacgccg ccatgggccg ggccttcctt  420
ctccaaggag gcgactgcgc cgagagcttc aaggagttca acggcaacaa tatccgcgat  480
accttccgcg tcctgcttca gatgtctgcc gtcctcacct ttggcggtca aatgcccgtc  540
atcaaggttg ggagaatggc cggccaattc gcaaagccga ggtcggattc gttcgaggta  600
agggatgggg tgaagctgcc gagctaccga ggggacaaca tcaacggtga tgctttcaac  660
gaaaagagcc gcatccccga tccccagagg atgatcaggg cctacaccca gtcggccgcg  720
acactcaacc tcctccgcgc gttcgccatg ggagggtacg ctgccatgca acgggtcacc  780
cagtggaacc tcgatttcac tgaaaacagc gagcagggcg acaggtaccg agaattggca  840
cacagggttg atgaagccct tggcttcatg tctgcagctg ggctaacatt ggatcaccct  900
gttatgtcga gtactgaatt ctggacctct catgagtgcc tactcctacc gtatgagcaa  960
gctttgaccc gccaggactc gacctctggt cttttctatg actgttctgc ccacatgctt 1020
tgggtcggtg aacgcactcg ccagctcgat ggtgctcatg ttgagttcct caggggtgtt 1080
gctaaccctc ttggtatcaa ggtgagtgac aaaatgaacc ctgctgactt ggtaaaactg 1140
attgacatat tgaacccaac aaacaagcct gggagaatta ccatcattac aagaatgggg 1200
gcggagaaca tgagggtcaa gttacctcat cttatccgtg ctgtccgcca tgctggtcaa 1260
attgtcacct ggatcactga tcccatgcat gggaacacta tcaaggctcc ttgcggtctg 1320
aagacacgcc cctttgactc cattctggct gaggtcaggg cgttcttcga tgttcatgag 1380
caagaaggta gccacgcagg aggcgtccac cttgagatga ctgggcagaa tgtgacagag 1440
tgcattggtg gatcccgaac cgtgaccttc gatgacctgg gcgaccgcta ccacacccac 1500
tgtgacccga ggctgaatgc atcccagtct ctggagctct ccttcatcat tgctgagaag 1560
ctgaggaaga ggaggatccg ctcgtcgaag cttaacagcg tcttgccatt gccaacctat 1620
ggtttctga                                                         1629

SEQ ID NO: 5           moltype = AA  length = 503
FEATURE                Location/Qualifiers
source                 1..503
                       mol_type = protein
                       organism = Brachypodium distachyon
SEQUENCE: 5
MPLAPPAAAV RNPLLPSPAL APARRGGLIR AHAVRAAPSQ WAPGSWRARP ALQQPEYPDK  60
AGLDEVLRTV ESFPPIVFAG EARKLEERLA DAALGRAFLL QGGDCAESFK EFNANNIRDT  120
FRVLLQMSVV LMFGGQMPII KVGRMAGQFA KPRSDGFEER DGVKLPSYRG DNINGDVFDE  180
KSRVPDPQRM IRAYSQSAAT LNLLRAFATG GYAAMQRVTQ WNLDFTEHCE QGDRYMELAH  240
RVDEALGFMS AAGLTVDHPI MTTTEFWTSH ECLLLPYEQA LTREDSTSGL YYDCSAHFLW  300
AGERTRQLDG AHVEFLRGIA NPLGIKVSDK MDPKELVKLI DILNPENRPG RITIITRMGP  360
ENMRVKLPHL IRAVRGAGQI VTWVTDPMHG NTMKAPCGLK TRSFDKILAE VRAFFDVHEQ  420
EGSHPGGVHL EMTGQNVTEC IGGSRTVTFD DLSSRYHTHC DPRLNASQSL ELVFIIAERL  480
RRRRAASWAL DNQPGTIPSS MGL                                          503

SEQ ID NO: 6           moltype = DNA  length = 1512
FEATURE                Location/Qualifiers
source                 1..1512
                       mol_type = other DNA
                       organism = Brachypodium distachyon
SEQUENCE: 6
atgcccctcg cgccacccgc cgccgccgtg cgcaacccac tcctcccttc cccggcgctc  60
gccccggcgc ggcgcggcgg gctcatccgc gcgcacgcgg tgcgcgcggc gccgagccag  120
tgggcgccgg ggagctggcg ggcgcggcct gcgctgcagc agccggagta cccggacaag  180
gccgggctgg acgaggtgct gcggacggtg gagtcgttcc cgccgatcgt cttcgcaggg  240
gaggcgcgca aactggagga gcggctcgcg gatgccgctc tcggtcgcgc gttccttctc  300
cagggcggcg actgcgccga gagcttcaag gagttcaacg ccaacaacat ccgggacacc  360
ttccgtgtcc tcctgcagat gtccgtcgtg ctcatgttcg gtggacagat gcctataatc  420
aaggtaggaa gaatggcagg tcaatttgca aagccaaggt cagatggttt tgaagagagg  480
gatggagtga agttgccaag ctacagagga gacaatatca atggggatgt atttgatgag  540
aagtcaagag tgccagatcc acagcgcatg atcagggcat actcacagtc tgcagcaaca  600
ctgaatttgc tgcgggcatt tgccacagga ggttatgctg caatgcagag ggtaacacag  660
tggaaccttg acttcacaga gcattgcgaa cagggtgata ggtacatgga gttggctcat  720
cgagttgacg aggctttggg gttcatgtca gctgctgggc tcactgtaga tcacccaata  780
atgacaacaa cagaattctg gacatcacat gaatgccttc ttcttcccta cgagcaggca  840
cttactcgtg aggattccac atctggcctc tactatgact gttcggccca cttcctatgg  900
gctggagagc gaacccgtca gttggatggt gcccatgtgg agttcctccg aggcattgcc  960
aacccctttgg gtatcaaggt tagcgacaaa atggacccaa aagaacttgt gaaattgatt 1020
gatatcttga atcccgaaaa caggccagga agaataacta tcattacgag aatgggacct 1080
gaaaacatga gagtgaaact cccccacctc atacgtgctg ttcgtggtgc tggccagata 1140
gtaacatggg taactgaccc aatgcacggg aacacgatga aggccccttg tggcctcaag 1200
actcgctcct tcgacaaaat tttggctgag gtgcgcgcat tctttgatgt ccacgagcaa 1260
gaagggagcc acccaggagg ggtgcatctg gagatgaccg ggcaaaacgt gacggagtgc 1320
atcggcgggt cacgcacggt gacgttcgat gatctgagct ctcgctacca cacgcactgt 1380
gacccgaggc tcaatgcctc acagtcgctg gagctggtgt tcatcatcgc cgagcgcctc 1440
agaagaagaa gggccgcctc gtgggcattg gacaaccagc cgggcaccat tccatcatcc 1500
atgggactct ga                                                     1512
```

-continued

SEQ ID NO: 7 moltype = AA length = 511
FEATURE Location/Qualifiers
source 1..511
mol_type = protein
organism = Brachypodium distachyon
SEQUENCE: 7
MAAPALPVAP PVPAHAPLVL ATRRRSPPSA SDPPRRPRAG PLVRATSVAA AESGWAPGSW 60
RSRPVRQIPE YPDAAALEEA ERVLASFPPL VFAGEARMLE ERLGDAAVGR AFLLQGGDCA 120
ESFKEFGANN IRDTFRLMLQ MAVVLTFGGQ MPTIKVGRMA GQFAKPRSNP VETIDGVTLP 180
SYQGDIINND AFDEKSRAPD PQRLIRAYSQ SASTLNLLRG FAHGGYADLQ RVTQWNLDFL 240
RHSLQGERYV ELAQRVQDAI GFMFAAGLPR QHPMMTTAEF WTSHECLHLP YEQALTREDS 300
ISGLYYDCSA HMLWVGERTR QLDGAHVEFL RGISNPLGVK VSDKLEPSEL VELCEILNPH 360
NKPGRLTLIT RMGAENMRVK LPHMIRAVRQ AGIIVTWVSD PMHGNTISAP CGLKTRSFDA 420
IRAELRAFFD VHEQEGSYPG GVHLEMTGQN VTECIGGSNT VTFDDLSSRY RTHCDPRLNA 480
SQSLELAFAI AERLRNKRDR TWNSLISRVE A 511

SEQ ID NO: 8 moltype = DNA length = 1536
FEATURE Location/Qualifiers
source 1..1536
mol_type = other DNA
organism = Brachypodium distachyon
SEQUENCE: 8
atggcggctc ccgctctccc cgtcgcgccg cccgttcccg cgcacgcgcc gctcgtcctc 60
gccacccgcc ggaggtcgcc gccgtcggcc tccgatccgc cgcggcgccc tagggcaggc 120
ccgctggtac gcgcgacgtc ggtcgcggcg gcggagagtg ggtgggcgcc gggcagctgg 180
aggtcgcggc cggtacggca gatcccggag tacccggacg cggcggcgct ggaggaggcg 240
gagcgcgtgc tggcgtcgtt cccgccgctg gtgttcgcgg gggaggcacg gatgctggag 300
gagcggctcg gggatgccgc cgtgggtcgc gccttcctcc tgcagggcgg cgactgcgcc 360
gagagcttca aggagttcgg cgccaacaac atccgcgaca ccttccgcct catgctccag 420
atggccgtcg tcctcacctt cggtggccag atgcccacca tcaaggttgg gaggatggct 480
ggccaatttg caaagccaag atcaaaccca gttgagacta tagatggagt gacacttcct 540
tcctatcaag gggatatcat caataacgat gcttttgacg agaaatcgcg cgcaccagat 600
cctcaaaggt tgatcagagc ctacagccag tctgcgagca ccctgaatct tttgagagga 660
tttgcccatg gaggatatgc cgatcttcag agagtcaccc agtggaacct tgacttcttg 720
aggcacagct tgcagggaga aaggtatgtg gagcttgccc agagggttca agacgccatt 780
gggttcatgt ttgctgctgg tctgcctcgt cagcacccca tgatgaccac agctgaattc 840
tggacatctc atgagtgtct tcacttgcca tacgagcagg cactgaccag ggaggactcc 900
atttctggcc tctactatga ctgctctgcc cacatgctct gggttggaga gaggactagg 960
cagctggatg gtgctcatgt tgaattcctc cgtggcattt ccaatcctct tggtgtaaag 1020
gtgagtgata agcttgagcc atcagagctt gtggaactgt gtgaaatttt gaatcctcac 1080
aacaagcctg gtaggctgac acttatcaca agaatggggg ctgagaacat gcgtgtcaag 1140
ctcccccata tgatcagagc agtgcgccaa gctgggataa ttgtcacctg ggtcagcgat 1200
cccatgcacg ggaacaccat cagtgcaccg tgcgggctca agacaagatc atttgacgca 1260
atcagggctg agcttagggc tttctttgat gttcatgagc aagaggggag ctaccccgga 1320
ggagttcacc tggagatgac agggcagaac gttacagagt gcattggtgg atcaaatacg 1380
gtgaccttcg atgatctcag ctcccgatat cgcacgcact gcgaccctag gctgaatgcg 1440
tcacagtcgc tcgagctggc cttcgccatt gctgagaggc taaggaataa gagagacagg 1500
acatggaata gcttgatatc tagagtagaa gcttaa 1536

SEQ ID NO: 9 moltype = AA length = 540
FEATURE Location/Qualifiers
source 1..540
mol_type = protein
organism = Sorghum bicolor
SEQUENCE: 9
MALATNSAAA AAAAAAVSGG ASSQPRRAAV FLPLKRRTIS AIHAADPSKN NGPAVPAAAA 60
AKSSASAVAT PEKKPAAPGK WAVDSWKSKK ALQLPEYPNQ EELDTVLKTI ETFPPVVFAG 120
EARHLEERMA EAAMGRAFVL QGGDCAESFK EFHANNIRDT FRILLQMGAV LMFGGQVPVV 180
KVGRMAGQFA KPRSEPPEER DGVKLPSYRG DNVNGDDFTE KSRVPDPQRM IRAYAQSVAT 240
LNLLRAFATG GYAAMQRVTQ WNLDFMDHSE QGDRYRELAH RVDEALGFMT AAGLTVDHPI 300
MTTTDFWTSH ECLLLPYEQA LTREDSTSGL FYDCSAHMLW VGERTRQLDG AHVEFLRGVA 360
NPLGIKVSDK MNPSDLVKLI EILNPSNKPG RITIITRMGA ENMRVKLPHL IRAVRNAGLI 420
VTWITDPMHG NTIKAPCGLK TRPFDSILAE VRAFFDVHDQ EGSHPGGIHL EMTGQNVTEC 480
IGGSRTVTFD DLSDRYHTHC DPRLNASQSL ELAFIIAERL RKRRMRSGLN NSLPLPPLAF 540

SEQ ID NO: 10 moltype = DNA length = 1623
FEATURE Location/Qualifiers
source 1..1623
mol_type = other DNA
organism = Sorghum bicolor
SEQUENCE: 10
atggcgctgg ccaccaactc cgccgccgcc gccgcagcag cagcggccgt atccggtggc 60
gcgtcatccc agccgcgccg cgcggccgtg ttcctcccgc tgaagaggcg caccatctcc 120
gccatccacg ccgccgaccc gtcaaagaac aacggacccg ccgtccccgc ggcggccgct 180
gccaagtctt ccgcctcggc ggtggccacg ccggagaaga agccggcggc tccggggaag 240

-continued

```
tgggcggtcg acagctggaa atcgaagaag gcgctgcagc tccccgaata cccgaaccag  300
gaggagctgg acacggtgct caagaccatc gaaacgttcc cgccggtcgt gttcgccgga  360
gaggcgcgcc acctcgagga gcgcatggca gaggctgcca tgggccgcgc cttcgtcctc  420
cagggcggcg actgcgccga gagcttcaag gagttccacg ccaacaatat ccgtgacacc  480
ttccgtattc tgctccagat gggcgccgtg ctcatgttcg gtggtcaggt gccggtcgtc  540
aaggtgggga ggatggctgg ccagttcgcc aagccaaggt ccgaaccgtt cgaggagagg  600
gatggtgtta agctgccgag ctacaggggt gacaacgtca acggcgatga cttcaccgag  660
aagagccgcg tgccagaccc gcagaggatg atccgcgcct acgcgcagtc ggtggcgacg  720
ctcaacctgc tccgcgcgtt cgccacagga gggtatgctg ccatgcagcg tgtcacacaa  780
tggaacctcg acttcatgga tcacagcgag caaggtgata ggtaccgtga attggcccat  840
agggtggatg aggctcttgg gttcatgact gcagcaggac ttaccgttga ccacccgata  900
atgacgacta ctgacttctg gacctcgcac gagtgccttc tcttacccta tgagcaggct  960
cttacccgtg aggactccac cagtggcctt ttctatgatt gttcagccca catgttgtgg  1020
gttggtgagc gcactcgtca acttgatgga gctcatgttg aattcctccg tggtgtggcc  1080
aaccctcttg gcataaaggt gagtgacaaa atgaatccca gtgacttggt gaagctgatt  1140
gagattctga acccttcaaa caaacccgga aggatcacca taattacaag gatgggggca  1200
gagaacatga gagtaaagtt gcctcatctc atccgtgctg tccgcaatgc tggattgatt  1260
gtcacatgga ttactgatcc tatgcatgga aacaccataa aggccccttg tggcttgaag  1320
actcgtcctt tcgattccat tctggctgaa gtgcgcgcat tcttcgacgt gcatgaccaa  1380
gaaggaagtc acccaggagg tatccacctt gaaatgactg ggcagaacgt gaccgagtgc  1440
attggtggat cacggactgt gacctttgac gacctgagcg accgctacca cacccactgt  1500
gacccaagac tgaacgcctc ccagtccctg gagcttgcct tcatcattgc agagaggctc  1560
aggaagagga ggatgcggtc agggctcaac aacagcctgc cgctgccacc actggctttc  1620
tga                                                                1623

SEQ ID NO: 11          moltype = AA  length = 551
FEATURE                Location/Qualifiers
source                 1..551
                       mol_type = protein
                       organism = Sorghum bicolor
SEQUENCE: 11
MALSTNAAAA AAAISGGSAS ASQPARRTPS SLLPLTRRRR AVVRAVHAAE PSKNPGVVVP  60
AAAKASSPTT VAPENDAAPA PAPARAPAKW AVDSWRTKKA LQLPEYPNPA ELEAVLKTIE  120
AFPPIVFAGE ARHLEERLAD AAMGRAFLLQ GGDCAESFKE FNSNNIRDTF RVLLQMSAVL  180
MFGAQMPVVK VGRMAGQFAK PRSDPFEVRD GVKLPSYRGD NINGEAFDEK SRVPDPQRMI  240
RAYAQSAATL NLLRAFATGG YAAMQRVTQW NLDFTEHSEQ GDRYRELAHR VDEALGFMSA  300
AGLTPDHPLT TTTEFWTSHE CLLLPYEQAL TRQDSTSGLF YDCSAHMLWV GERTRQLDGA  360
HVEFLRGIAN PLGIKVSDKM NPSDLVKLID ILNPTNKPGR ITVITRMGAE NMRVKLPHLI  420
RAVRQAGQIV TWITDPMHGN TIKAPCGLKT RPFDNILAEV RAFFDVHEQE GSHPGGVHLE  480
MTGQNVTECI GGSRTVTFDD LSDRYHTHCD PRLNASQSLE LAFIIAERLR KRRIRSSSGL  540
NNILPLPPFG F                                                       551

SEQ ID NO: 12          moltype = DNA  length = 1656
FEATURE                Location/Qualifiers
source                 1..1656
                       mol_type = other DNA
                       organism = Sorghum bicolor
SEQUENCE: 12
atggcgctct ccaccaacgc cgccgccgcc gccgccgcca tctccggcgg gtccgcgtcc  60
gcgtcgcagc cggcccgtcg cacgccctcg tctctcctcc cgctgacgcg ccgccgccgc  120
gccgtcgtcc gcgccgtgca cgccgcggag ccgtcgaaga accccggcgt cgtcgtcccg  180
gccgccgcca aggcctcgtc gccgacgacg gtcgcgcctg agaacgacgc ggcccctgct  240
ccggccccgg cccgtgcgcc ggcgaagtgg gcggtggaca gctggaggac gaagaaagcg  300
ctacagctgc cggagtatcc gaacccggca gagctagagg cggtgctcaa gaccatcgag  360
gcgttcccac ccatcgtctt cgctggggag gcgcgccacc tggaagagcg cctcgccgac  420
gccgccatgg gacgcgcctt cctcctccag ggcggcgact gcgctgagag cttcaaggag  480
ttcaacagca acaacatccg cgacaccttc cgcgtcctcc tgcagatgtc cgccgttctc  540
atgttcggcg cccagatgcc cgtcgtcaag gttggtagga tggccggcca attcgcgaag  600
ccaaggtcgg atccgttcga ggttagagat ggcgtcaagc tgcccagcta ccggggcgat  660
aacattaatg gcgaggcatt cgacgagaag agccgtgtcc ccgacccaca gaggatgatc  720
cgcgcgtatg cgcagtctgc cgcaacgctg aacttgctcc gcgccttcgc caccggaggc  780
tacgctgcca tgcagcgcgt cacgcagtgg aacctcgatt tcaccgaaca cagcgagcag  840
ggcgacaggt accgtgaatt ggcacatcgg gttgatgaag cccttggctt catgtctgca  900
gctggcctaa caccggacca ccctttgacg acgaccactg agttctggac ctcacatgag  960
tgcctcctcc taccttacga gcaagcttta acccgtcaag actccacctc tggtcttttc  1020
tacgattgct ctgcccacat gctctgggtt ggtgagcgca ctcgccagct tgatggcgcc  1080
catgttgagt tcctcagggg tattgccaac cctcttggca tcaaggtaag tgacaaaatg  1140
aaccccagcg acttggtgaa gctgattgat atactgaacc caacaaacaa gcccgggagg  1200
atcaccgtta ttacaaggat gggggcagag aacatgaggg tgaagttacc tcaccttatc  1260
cgtgcggtcc gccaggctgg acaaattgtc acctggatca ctgacccgat gcacggcaac  1320
accatcaagg ctccttgtgg tctaaagact cgtccatttg acaatattct ggctgaggta  1380
cgggccttct ttgatgtgca cgagcaagag gggagccacc ctggaggtgt ccaccttgag  1440
atgactgggc aaaacgtaac tgaatgcatt ggcgggtcac gcaccgtgac cttcgatgac  1500
ctgagcgacc gctaccacac ccactgtgac cccaggctga acgcgtccca gtctctggag  1560
ctcgcgttca ttatcgccga gaggctgagg aagaggagga tccggtcatc gtccgggctc  1620
aacaacatct tgcccttgcc gccttttggt ttctga                             1656
```

-continued

```
SEQ ID NO: 13          moltype = AA  length = 508
FEATURE                Location/Qualifiers
source                 1..508
                       mol_type = protein
                       organism = Sorghum bicolor
SEQUENCE: 13
MPLAPSTPAL PPNPALPSPC RGQGRGRPGR GALLRARAVR AAPRPPSQWS VGSWRGRPAL  60
QQPEYPDKAD LNEVLRTVEA FPPIVFAGEA RTLEERLAEA AVGRAFLLQG GDCAESFKEF  120
NANNIRDTFR VLLQMSVVLM FGGQMPVVKV GRMAGQFAKP RSDGFEERDG VKLPSYRGDN  180
INGDAFDEKS RLPDPHRMIS AYSQSAATLN LLRAFATGGY AAMQRVTQWN LDFTEHSEQG  240
DRYMELAHRV DEALGFMSAA GLTLDHPIMT TTEFWTSHEC LLLPYEQALT REDSTSGLYY  300
DCSAHFLWVG ERTRQLDGAH VEFLRGIANP LGIKVSDKMD PAELVRLIDI LNPENRAGRI  360
TIIARMGPEN MRVKLPHLIR AVRGAGQIVT WVTDPMHGNT MKAPCGLKTR SFDRILAEVR  420
AFFDVHEQEG SHPGGVHLEM TGQNVTECIG GSRTVTFDDL GSRYHTHCDP RLNASQSLEM  480
AFIIAERLRK RRIASSPLYT NQLGSIRL                                      508

SEQ ID NO: 14          moltype = DNA  length = 1527
FEATURE                Location/Qualifiers
source                 1..1527
                       mol_type = other DNA
                       organism = Sorghum bicolor
SEQUENCE: 14
atgccactcg cgccatcaac ccccgcgctg ccgcctaacc cggccctccc ctcgccgtgc  60
cgggggcagg ggcggggccg gcccggccgc ggggcgctcc tccgcgcccg cgccgtccgc  120
gcggcgccgc ggcccccgag ccagtggtcc gtcggcagct ggcggggccg cccggcgctg  180
cagcagcccg aatacccgga caaggcggac ctgaacgagg tgctgcggac ggtggaggcg  240
ttcccgccca tcgtgtttgc cggcgaggcg cgcaccctcg aggagcgcct cgcggaggcc  300
gccgtcggca gagctttcct cctccagggc ggtgactgcg ctgagagctt caaggagttc  360
aacgccaaca atattaggga caccttcgc gtcctcctgc aaatgtctgt tgtgctcatg  420
ttcggaggcc agatgcctgt cgtcaaggtg ggaagaatgg caggtcagtt tgcaaagcca  480
agatcagatg gttttgagga gcgggatgga gtgaagttgc caagctatag aggggacaat  540
atcaatgggg atgcatttga tgagaagtca agattgccag atccacaccg catgataagt  600
gcctactcac agtctgcagc aacgctgaat ctgctgcggg cgttcgcaac tggaggttat  660
gctgcaatgc agagggtaac acaatggaac cttgatttca cagagcatag tgaacaaggg  720
gataggtaca tggaattggc tcaccgagtt gacgaagctt tggggttcat gtcagctgct  780
ggactcactt tagatcaccc tataatgacg acaacagaat tctggacgtc acatgagtgc  840
cttcttctgc cttatgagca agcgctcacc cgtgaagact ccaccagtgg cctctattat  900
gattgctctg ctcacttcct atgggttgga gagcgcactc gccagcttga tggtgctcat  960
gtggagttcc ttcgaggcat cgccaaccct cttggtatca aggttagtga caagatggac  1020
ccagcagaac ttgtgcggtt gattgatata ttgaatcccg aaaacagggc tgggagaata  1080
accatcatcg caagaatggg acctgaaaac atgagggtga aacttccaca cctgatacgc  1140
gccgtccgtg gggccggcca gatagtaaca tgggtcactg acccaatgca tgggaacacc  1200
atgaaggccc cttgcggact caaaacccgc tcatttgaca gaattttggc cgaggtgcgt  1260
gcgttctttg atgtccacga acaagaaggg agccaccctg gaggagtgca tctagagatg  1320
actgggcaaa atgttacaga gtgcattgga gggtcacgta ccgtgacatt tgatgatctg  1380
ggctcacgct accacacgca ctgcgaccca aggctcaacg cctcacagtc tctggagatg  1440
gcattcatta tcgcggagcg ccttaggaaa aggaggattg cctcgtcgcc tttgtacacg  1500
aaccagctgg gttccattcg tctgtga                                      1527

SEQ ID NO: 15          moltype = AA  length = 499
FEATURE                Location/Qualifiers
source                 1..499
                       mol_type = protein
                       organism = Sorghum bicolor
SEQUENCE: 15
MAPLATAPPP TPHAPRSLVP RVRPRRGLTA VRAASQGRTT DGWTPGSWRA RPARQIPEYP  60
DTAALEATER TLAEFPPLVF TGEVRKLEER LGEAAMGRAF LLQGGDCAES FREFNAQKIR  120
DTFRLILQMA VVLTFGGQMP TIKVGRMGGQ FAKPRSNPTE TVDGVTLPSY RGDIINDQAF  180
DEKSRVPDPE RLIRAYTQSA STLNLLRAFA HGGFADLQRV TQWNLDFLRH STQGDRYLEL  240
SQRVHEAIGF MVAAGLTPQH PIMTTAELWT SHESLHLPYE QALTREDSIS GRYYDCSAHM  300
LWVGERTRQL DGAHVEFLRG ISNPLGIKVS DKLDPSELVK LCETLNPHNK PGRLTIITRM  360
GAENMRVKLP QMIRAVRQAG MIVTWVSDPM HGNTISAPCG LKTRSFDSIM AELRAFFDVH  420
GLEGSYPGGV HLEMTGQNVT ECIGGSKAVT FDDLSARYHT HCDPRLNASQ SLELAFAIAD  480
RLRNKRDKTW HNLTSRVVA                                               499

SEQ ID NO: 16          moltype = DNA  length = 1500
FEATURE                Location/Qualifiers
source                 1..1500
                       mol_type = other DNA
                       organism = Sorghum bicolor
SEQUENCE: 16
atggcgcctc tcgccaccgc gccgcctcct actccccatg ccccgcgctc tctcgttcca  60
cgggtaaggc cgcggcgagg gctgacggca gtgcgcgcgg cctcacaggg aagaacaacg  120
gatgggtgga cgccgggaag ctggcgtgcg cgccctgcgc ggcagatccc ggagtacccc  180
gacacggcgg cgctggaggc tacagagcga actctggcgg agtttccacc gctagtgttc  240
```

-continued

```
acgggtgagg tgcggaagct ggaggagcgg ctcggggagg cggccatggg tcgggcattc  300
ctcctgcagg gcggcgactg cgccgagagc tttagggagt tcaatgcaca gaaaatccgg  360
gacaccttcc gactcatact gcagatggcc gtcgttctca ccttcggcgg acagatgccc  420
accatcaagg tgggaaggat gggtggtcaa tttgcgaagc caagatcgaa cccaactgag  480
actgttgatg gggtaaccct tccctcctat agaggggata tcatcaacga tcaggctttt  540
gatgagaaat ctcgtgtacc agatcctgaa aggttgatta gagcctatac ccagtctgcg  600
agcaccctga atcttctcag agcatttgcc catggaggtt ttgcggatct tcagagagtc  660
acccagtgga atctcgactt tttgaggcac agcacacaag gagacaggta tctggagctt  720
tctcagaggg ttcatgaagc cattgggttt atggttgccg ctgggctgac ccctcaacac  780
cccatcatga ccacagctga gttatggaca tcccatgagt cccttcactt accatacgag  840
caagcactga ctagggagga ttccattagt ggccggtact atgactgttc cgcacacatg  900
ctttgggtcg gtgagaggac tcggcagctg gatggtgctc atgttgaatt ccttcgtggt  960
atttcaaatc ctctgggtat aaaggtaagt gataagcttg acccatcaga gcttgtgaag 1020
ttgtgtgaga ctctgaatcc tcacaacaag cctgggcgat tgacgattat cacaagaatg 1080
ggggctgaga acatgcgtgt taagcttcca caaatgatca gagcagtgcg ccaagctggg 1140
atgattgtta cttgggttag tgatcctatg cacggaaaca ctatcagtgc accatgcgga 1200
ctgaagacaa gatcatttga ttcaatcatg gctgaactca gggctttctt cgatgttcat 1260
gggctagagg gaagctaccc tggaggagtt cacctggaga tgacggggca gaatgttaca 1320
gagtgtatcg ggggatctaa ggcggtgaca tttgatgatc tcagtgcccg ctaccacaca 1380
cactgtgacc ccagactgaa cgcgtcgcag tccctcgagc tggctttcgc cattgctgac 1440
aggttaagga acaagagaga caagacatgg cataatttga catccagagt tgttgcttaa 1500

SEQ ID NO: 17          moltype = AA  length = 537
FEATURE                Location/Qualifiers
source                 1..537
                       mol_type = protein
                       organism = Oryza sativa
SEQUENCE: 17
MALATNSAAV SGGAAAAASS APQPRLAATF LPMRRRTVSA VHAADPAKSN GPVQAAAKAS  60
SPSTVAAPEK KPVGLGKWTV DSWKAKKALQ LPEYPSQEEL DSVLKTIETF PPVVFAGEAR  120
HLEERLADAA MGRAFVLQGG DCAESFKEFN ANNIRDTFRI LLQMGAVLMF GGQMPVVKVG  180
RMAGQFAKPR SDSFEERDGV KLPSYRGDNI NGDTFDEKSR VPDPQRMIRA YAQSVATLNL  240
LRAFATGGYA AMQRVTQWNL DFMDHSEQGD RYRELAHRVD EALGFMTAAG LTVDHPIMTT  300
TDFWTSHECL LLPYEQSLTR EDSTSGLFYD CSAHMLWVGE RTRQLDGAHV EFLRGVANPL  360
GIKVSDKMNP RDLVKLIEIL NPSNKPGRIT IITRMGAENM RVKLPHLIRA VRNSGQIVTW  420
ITDPMHGNTI KAPCGLKTRP FDSILAEVRA FFDVHDQEGS HPGGIHLEMT GQNVTECIGG  480
SRTVTFDDLS DRYHTHCDPR LNASQSLELA FIIAERLRRR RMRSGVNSNL PLPPLAF     537

SEQ ID NO: 18          moltype = DNA  length = 1614
FEATURE                Location/Qualifiers
source                 1..1614
                       mol_type = other DNA
                       organism = Oryza sativa
SEQUENCE: 18
atggcgctcg ccaccaactc cgccgctgtt tccggtggcg ctgcggccgc ggcgtcgtcg  60
gcgccccagc cgcggctcgc cgccacgttc ctcccgatga ggaggcgaac cgtctcggcg  120
gtccacgcgg ccgacccggc gaagagcaac gggcccgtgc aggccgcggc gaaggcctcg  180
tccccgtcga cggtggcggc gccggagaag aagccggtgg ggttggggaa gtggacggtg  240
gatagctgga aggcgaagaa ggcgctgcag ctgcccgagt acccgagcca ggaggagctc  300
gactccgtgc tcaagacgat cgagacgttc ccgcccgtgg tgttcgccgg ggaggcgcgc  360
cacctcgagg agcgcctcgc cgacgccgcc atgggccgcg ccttcgtcct ccagggcggc  420
gactgcgccg agagcttcaa ggagttcaac gccaacaaca tccgtgacac cttccgcatc  480
ctgctccaga tgggcgccgt cctcatgttc ggcggccaga tgcccgtcgt caaggtcggg  540
aggatggctg gccagttcgc caagccgagg tctgattcgt tcgaggagag ggacggggtt  600
aagctgccga gttacagggg agacaacatc aatggcgaca ccttcgacga gaagagccgc  660
gtgccggacc cgcagcggat gatccgcgcg tacgcgcagt cggtggcaac gctcaacctg  720
ctccgcgcct tcgccacggg agggtacgcc gccatgcagc gcgtcacgca gtggaacctc  780
gatttcatgg atcacagcga acaaggagac aggtaccgtg aattggcaca ccgggtggat  840
gaggcacttg gcttcatgac tgcagcaggg ctaacagtcg accatccgat aatgacaaca  900
actgacttct ggacatccca tgagtgcctc ctcttaccat atgagcagtc tcttacacgt  960
gaggattcca ccagtggcct cttctatgac tgctcagccc acatgctatg ggttggtgag 1020
cgcactcgcc agctcgatgg agctcatgtt gaatttctcc gtggtgttgc caaccctctt 1080
ggcataaagg tgagtgacaa aatgaacccc cgtgacttgg tgaagctcat tgagattttg 1140
aacccatcaa acaagcctgg aaggataacc ataattacaa gaatgggggc agagaacatg 1200
agggtgaaat tgcctcatct tatccgtgct gtccgcaact ctggacagat tgtcacatgg 1260
attactgatc ccatgcatgg aaacaccatc aaggctccat gtggtctgaa gactcgtcca 1320
ttcgattcca ttctggctga agtacgtgcg ttcttcgatg tccatgatca agaaggtagc 1380
cacccaggag gtatccacct ggaaatgact gggcagaacg tgactgaatg catcggtgga 1440
tcaaggaccg tcaccttcga tgacctgagc gaccgctacc acacccactg tgacccaagg 1500
ctgaacgcgt cgcaatccct ggagctcgcc ttcatcatcg ccgagagact gaggcggagg 1560
aggatgcggt ccggggtcaa cagcaacctg ccattgcccc cattggcttt ctaa        1614

SEQ ID NO: 19          moltype = AA  length = 555
FEATURE                Location/Qualifiers
source                 1..555
                       mol_type = protein
```

-continued

```
                     organism = Oryza sativa
SEQUENCE: 19
MSLATSSSMA GGAAVVPRSA TATTASAFVT MKRRATAVRA VHAAEPSKNP PVGVPSAAKT  60
SSPSVAAPEK APVAAAPAPV APAPAATKQV APARWAVDSW RTKKALQLPE YPNAAELEAV  120
LKTIEAFPPI VFAGEARHLE ERLADAAMGR AFLLQGGDCA ESFKEFNGNN IRDTFRVLLQ  180
MSAVLTFGGQ MPVIKVGRMA GQFAKPRSEA FEERDGVKLP SYRGDNINGD AFNEKSRIPD  240
PQRMVRAYAQ SAATLNLLRA FATGGYAAMQ RVTQWNLDFT QHSEQGDRYR ELAHRVDEAL  300
GFMSAAGLTV DHPLMTSTDF WTSHECLLLP YEQSLTRQDS TTGHFYDCSA HMLWVGERTR  360
QLDGAHVEFL RGVANPLGIK VSDKMNPTEL VKLIEILNPS NKPGRITIIT RMGAENMRVK  420
LPHLIRAVRH AGQIVTWITD PMHGNTIKAP CGLKTRPFDS ILAEVRAFFD VHDQEGSHPG  480
GVHLEMTGQN VTECIGGSRT VTFDDLGDRY HTHCDPRLNA SQSLELSFII AERLRRKRIR  540
SSKLNNMLPL PPFGV                                                   555

SEQ ID NO: 20          moltype = DNA  length = 1668
FEATURE                Location/Qualifiers
source                 1..1668
                       mol_type = other DNA
                       organism = Oryza sativa
SEQUENCE: 20
atgtcgctcg ccaccagctc ctcgatggcc ggtggggcgg cggtggttcc acgcagcgcg  60
acggcgacga cggcctcggc gttcgtcacg atgaagcggc gcgccacggc cgtgcgcgcc  120
gtacacgcgg cggagccgtc gaagaacccg cccgtcggcg tcccgtccgc ggcgaagacg  180
tcgtcgccgt cggtggcggc gccggagaag gcccccgtgg cggccgcgcc cgcgcctgtg  240
gctccggctc cggcggccac gaagcaggtg gcgccggcga ggtgggcggt ggacagctgg  300
aggacgaaga aggcgctgca gctgccggag taccccaacg cggcggagct ggaggcggtg  360
ctcaagacga tcgaggcgtt cccgccgatc gtgttcgccg gggaggcgag gcacctggag  420
gagcgcctcg ccgacgccgc catgggccgc gccttcctcc tccagggcgg cgactgcgcc  480
gagagcttca aggagttcaa cggcaacaac atccgcgaca ccttccgcgt cctcctccag  540
atgtccgccg tcctcacctt cggcggccaa atgcccgtca tcaaggttgg gagaatggcg  600
gggcaattcg cgaagccgag gtcggaggcg ttcgaggaga gggacggggt gaagctgccg  660
agctacaggg gcgacaacat caatggcgac gcgttcaacg agaagagccg catccctgac  720
ccgcagagga tggtccgggc ctatgcgcag tccgccgcca cgctcaacct cctccgcgct  780
ttcgccaccg gaggatacgc cgccatgcag cgcgtgacgc agtggaacct cgatttcacc  840
cagcacagcg agcagggcga caggtaccgt gaattggcac acagggttga tgaagccctt  900
ggctttatgt ctgctgctgg gctaacagtg gaccacccgt tgatgacaag tactgatttc  960
tggacctcac atgagtgcct tctcctaccc tatgaacaat ctctgacccg tcaggactcc  1020
acaactggtc atttctacga ctgctctgcc cacatgctat gggtgggcga acgcactcgc  1080
cagctcgatg gtgctcatgt tgagttcctc aggggtgtgg ccaaccctct tggcatcaag  1140
gtgagtgaca aaatgaaccc aaccgagttg gtgaagctga ttgagatttt gaacccatca  1200
aacaagcctg ggagaattac catcatcaca aggatggggg cagagaacat gagggtcaag  1260
ttacctcatc ttatccgtgc ggtccgccac gctggacaaa ttgtcacatg gattactgat  1320
ccaatgcacg ggaacaccat caaggctcct tgtggtctga agacacgccc ctttgactcc  1380
attctggccg aggtacgagc attcttcgat gtgcatgatc aagaaggaag ccacccagga  1440
ggtgtccacc tcgagatgac tgggcagaac gtaaccgagt gcatcggcgg atcccgaacc  1500
gtcaccttcg atgacctggg tgaccgctac cacacccact gcgacccgag gctgaacgcg  1560
tcccagtcgc tggagctctc cttcatcatc gccgagaggc tgaggaggaa gaggatccgg  1620
tcgtcgaagc tgaacaacat gttgccgttg ccacccttcg gtgtctga               1668

SEQ ID NO: 21          moltype = AA  length = 505
FEATURE                Location/Qualifiers
source                 1..505
                       mol_type = protein
                       organism = Oryza sativa
SEQUENCE: 21
MPLAPCPSPP LPSSPWPARA PRRGGLLRAR AVRAAPRPPS KWSLGSWRSL TALQQPEYPD  60
KAELDEVLRT VEAFPPIVFA GEARKLEERL AEAAVGRAFL LQGGDCAESF KEFNANNIRD  120
TFRVLLQMSV VLMFGGQMPI IKVGRMAGQF AKPRSDGFEE RDGVKLPSYR GDNINGDSFD  180
EKSRLPDPHR MIRAYSQSAA TLNLLRAFAT GGYAAMQRVT QWNLDFTEHS EQGDRYMELA  240
HRVDEALGFM AAAGLTMDHP IMTTTEFWTS HECLLLPYEQ ALTREDSTSG LYYDCSAHFL  300
WVGERTRQLD CAHVEFLRGI ANPLGIKVSD KMDPKELVKL IDILNPQNKP GRITIITRMG  360
PENMRVKLPH LIRAVRGAGQ IVTWVTDPMH GNTMKAPCGL KTRSFDRILA EVRAFFDVHE  420
QEGSHPGGVH LEMTGQNVTE CIGGSRTVTF DDLGSRYHTH CDPRLNASQS LELAFIIAER  480
LRKRRIASWQ LNKNSHLGNI PSLGL                                        505

SEQ ID NO: 22          moltype = DNA  length = 1518
FEATURE                Location/Qualifiers
source                 1..1518
                       mol_type = other DNA
                       organism = Oryza sativa
SEQUENCE: 22
atgcccctcg cgccatgccc ctcgccgccc ctcccctcct ccccgtggcc ggcgcgcgcc  60
ccgcgccggg gcggcctcct ccgcgcccgc gcggtgcggg cggcgccccg gccgccgagc  120
aagtggtcgc tgggtagctg gcgcagcctg acggcgctgc agcagccgga gtaccccgac  180
aaggcggagc tggatgaggt gctccggacg gtggaggcgt tcccgccgat tgtcttcgcc  240
ggcgaggcgc gcaagctgga ggagcggctc gcggaggccg ccgtcgggcg cgcgttcctc  300
ctccagggcg gcgactgcgc cgagagcttc aaggagttta acgcgaacaa catccgggac  360
```

-continued

```
accttccgtg tgctcctcca gatgtccgtg gtgctcatgt ttggaggcca gatgcctatc  420
atcaaggtag gaagaatggc aggtcagttt gcaaagccaa ggtcagatgg ctttgaggag  480
agggatggag tgaagttgcc aagctacaga ggggataaca ttaatgggga ttcattcgat  540
gagaaatcaa gattgccaga tccacaccgc atgatcaggg catactcaca gtctgcagca  600
acactgaatt tgctgcgggc ttttgctact ggaggttatg ctgccatgca gagggtaaca  660
caatggaacc ttgacttcac agagcatagt gaacagggtg acaggtacat ggagctggct  720
caccgagttg atgaggcttt ggggttcatg gcagctgctg gtctcactat ggaccatcct  780
attatgacaa caacagaatt ctggacatca catgagtgcc ttcttcttcc ctatgagcaa  840
gcacttactc gcgaggattc cacatctggc ctctattacg actgttctgc tcacttcctt  900
tgggttggag agcgtacacg tcagcttgat tgtgcccatg tggagtttct ccgaggaatt  960
gcgaaccctc tgggtatcaa ggtcagtgac aagatggacc caaaagaact tgtgaagttg 1020
attgatatct tgaatcccca gaacaaacca gggagaatta ctatcattac aagaatggga 1080
cctgaaaaca tgagagtgaa actccctcac ctaatacgtg ctgtccgtgg tgctggccag 1140
atagtaacat gggttactga tccgatgcat ggtaacacaa tgaaggctcc ttgtggcctc 1200
aagactcgct cctttgatag aatcttggct gaggtgcgcg cattctttga tgtgcacgaa 1260
caagaaggca gccacccagg aggggtgcat ctggagatga ctggacaaaa tgtgacagaa 1320
tgcatcggcg ggtcacgcac ggtgacattc gacgatctgg gctcacgata ccacacacac 1380
tgcgacccga ggctcaacgc atcgcagtct ctggagttgg cgttcatcat cgccgagcgg 1440
ctcagaaaga ggaggatcgc ctcatggcag ttgaacaaga acagtcatct gggcaacatc 1500
ccatctttgg ggctctga                                              1518

SEQ ID NO: 23          moltype = AA  length = 508
FEATURE                Location/Qualifiers
source                 1..508
                       mol_type = protein
                       organism = Oryza sativa
SEQUENCE: 23
MTCHSAMAAL TVGHAAIVHA TTRLEDARST GRRRRRRGMI TVRAAAAATS GWEPGSWRAR  60
PARQIPEYPD AAALEGAERE LASFPPLVFA GEARKLEERL GDAAMGRAFL LQGGDCAESF  120
KEFAANNIRD TFRLMLQMAV VLTFGGQMPT IKVGRMAGQF AKPRSNPTET IDGVTLPSYR  180
GDIINSDGFD EKSRAPDPER LIRAYSQSAS TLNLLRGFAH GGYADLQRVT QWNLDFLRDS  240
TQGDRYMELS ERVHDAIGFM VAAGLTPQHP IMTTAEFWTS HECLHLPYEQ ALTRVDSISG  300
LYYDCSAHML WVGERTRQLD GAHVEFLRGI SNPLGVKVSD KLEPSELVKL CEILNPHNKP  360
GRLTIITRMG AENTRVKLPH MIRAVRQAGL IVTWVSDPMH GNTISAPCGL KTRSFDAIRC  420
ELRAFFDVHE QEGSYPGGIH LEMTGQNVTE CIGGSKTVTL DDLSSRYRTH CDPRLNASQS  480
LELAFAIADR LRKKRDRAWN RLVYRAVA                                     508

SEQ ID NO: 24          moltype = DNA  length = 1527
FEATURE                Location/Qualifiers
source                 1..1527
                       mol_type = other DNA
                       organism = Oryza sativa
SEQUENCE: 24
atgacgtgtc actctgccat ggcggctctc accgttgggc acgcggcgat cgtccatgcc  60
accacgcgtc tcgaggacgc ccggtccacc ggccgtcggc ggcggcggcg ggggatgatc  120
acggtgcgcg cggcggcggc ggcgacgagc gggtgggagc cgggcagctg gagggcgcgg  180
ccggcgaggc aaatcccgga gtacccggac gcggcggcgc tggagggcgc ggagcgcgag  240
ctggcgtcgt tcccgccgct ggtgttcgcc ggggaggcgc ggaagctgga ggagcggctc  300
ggggacgccg ccatggggcg ggccttcctc ctgcagggcg gcgactgcgc cgagagcttc  360
aaagagttcg ccgccaacaa catccgcgac acctttcgcc tcatgctcca gatggccgtc  420
gtcctcacct ttggcggcca gatgcccacc atcaaggttg gaaggatggc tggccaattt  480
gcaaagccaa gatcaaaccc aactgagact atagatggag tgacacttcc ttcctatcga  540
ggtgatatta ttaacagcga tggttttgat gagaagtcgc gtgcaccaga tcctgaaagg  600
ttaattagag cctacagcca gtccgcgagc accctgaatc ttctgagagg atttgctcat  660
ggagggtatg cagatctgca gagagtcacc cagtggaatc ttgacttctt gagggacagc  720
acgcaagggg acaggtatat ggagctgtcc gagagggttc acgatgccat cggatttatg  780
gttgctgctg gtctgactcc tcagcatccc atcatgacga cggctgaatt ctggacatct  840
catgagtgcc ttcatttgcc atatgagcaa gcattgacta gggtggactc catttctggg  900
ctttactacg attgctctgc tcatatgctt tgggttgggg agaggactcg acaactggat  960
ggtgctcatg ttgaattcct tcgtggtatt tccaaccctc taggtgtaaa ggtgagtgac 1020
aagctcgaac cttcagagct tgttaaattg tgtgagattt tgaatcctca caacaagccc 1080
ggaagactga caatcatcac aagaatgggt gctgagaaca cgcgcgttaa gctcccacat 1140
atgatcagag cagtccgcca agctgggttg attgtcactt gggtcagtga tcccatgcat 1200
gggaacacca tcagcgcacc atgtggtctc aagacaagat cattcgacgc gatcaggtgc 1260
gagctgaggg ctttcttcga tgtccatgag caagagggaa gctaccctgg agggattcac 1320
ctggagatga cagggcagaa cgttacagag tgcattggtg gatccaagac ggtgaccctt 1380
gatgatctca gctctcgcta ccgcacgcac tgcgacccca ggctgaatgc atcgcagtcg 1440
cttgaactgg ctttcgccat tgctgacagg ctaaggaaga aacgagacag ggcttggaat 1500
aggttggtgt acagggcggt agcttaa                                     1527

SEQ ID NO: 25          moltype = AA  length = 539
FEATURE                Location/Qualifiers
source                 1..539
                       mol_type = protein
                       organism = Setaria viridis
SEQUENCE: 25
```

-continued

```
MALATNSAAT AAAAAAVSGG AASQPRRAAA FVPLKRRTIS AIHAADPSKN NGSAVPAASS  60
SKASSFAVAT PEKKPAAQGK WTVDSWKSKK ALQLPEYPSQ EELDSVLKTI ETFPPVVFAG  120
EARHLEERLA EAAMGRAFVL QGGDCAESFK EFHANNIRDT FRILLQMGAV LMFGGQVPVV  180
KVGRMAGQFA KPRSEAPEER DGVKLPSYRG DNVNGDDFTE KSRVPDPQRM IRAYAQSVAT  240
LNLLRAFATG GYAAMQRVTQ WNLDFMDHSE QGDRYRELAH RVDEALGFMT AAGLTVDHPI  300
MTTTDFWTSH ECLLLPYEQA LTREDSTSGL FYDCSAHMLW VGERTRQLDG AHVEFLRGVA  360
NPLGIKVSDK MNPSDLVKLI EILNPSNKPG RITIITRMGA ENMRVKLPHL IRAVRNAGLI  420
VTWITDPMHG NTIKAPCGLK TRPFDSILAE VRAFFDVHDQ EGSHPGGIHL EMTGQNVTEC  480
IGGSRTVTFD DLSDRYHTHC DPRLNASQSL ELAFIIAERL RKRRMRSGLN NSLPLPLAF   539

SEQ ID NO: 26          moltype = DNA  length = 1620
FEATURE                Location/Qualifiers
source                 1..1620
                       mol_type = other DNA
                       organism = Setaria viridis
SEQUENCE: 26
atggcgctcg ccaccaactc cgcggcaacc gccgccgccg cggcggccgt atccggcggc  60
gcggcatccc agccgcgccg cgcggccgcg ttcgtcccgc tgaagaggcg caccatctcc  120
gccatccacg ccgccgaccc gtcgaagaac aacgggtccg ccgtccccgc cgcctcctcc  180
tccaaggcct cgtcctttgc ggtggcgacg ccggagaaga agccggcggc gcaagggaag  240
tggacggtag acagctggaa gtcgaagaag gcgctgcagc tccccgagta cccgagccag  300
gaggagctgg actccgtgct caagacgatc gagacgttcc cgccggtggt gttcgccggg  360
gaggcgcgcc acctcgagga gcgcctcgcc gaggccgcca tgggccgcgc cttcgtcctc  420
cagggtggag attgcgctga gagcttcaag gagttccacg ccaacaacat ccgtgacacc  480
ttccgcatcc tgctccagat gggcgctgtc ctcatgttcg gcggtcaggt cccggtcgtc  540
aaggtgggga ggatggctgg ccagttcgcc aagccaaggt ccgaggcatt cgaggagagg  600
gacggcgtta agctgccgag ctacaggggc gacaacgtga acggagacga tttcaccgag  660
aagagccgcg tgccggaccc gcagaggatg atccgcgcct acgcgcagtc ggtggccacg  720
ctcaaccttc tccgcgcgtt cgcgactgga ggctacgctg ctatgcagcg cgtcacacag  780
tggaatctcg atttcatgga tcacagcgag caaggtgata ggtaccgtga attggcccat  840
agggtggatg aggctcttgg attcatgact gcagcggggc ttacagttga ccacccgata  900
atgacgacta ctgacttctg gacctcgcac gagtgccttc tcttaccata cgagcaggct  960
cttacccgtg aggactccac cagtggcctt ttctatgatt gttcggccca catgctgtgg  1020
gttggggagc gcactcgcca acttgatggg gctcatgttg aattcctcag aggtgttgcc  1080
aaccctcttg gcataaaggt gagcgacaaa atgaacccca gtgacttggt gaagctgatt  1140
gagattttga acccttcaaa caaacctgga agaatcacca taattacaag gatgggggca  1200
gagaacatga gagtgaagtt gcctcatctt atccgtgctg tccgcaatgc tggactgatt  1260
gtcacatgga ttactgatcc tatgcatgga aacacaatca aggccccttg tggtttgaag  1320
actcgtccat tcgactccat tcttgctgaa gtacgtgcat tcttcgatgt gcatgaccaa  1380
gaagggagcc acccaggagg tatccacctg gagatgactg ggcagaacgt gactgagtgc  1440
attggtggat cacggaccgt gaccttcgat gacctgagcg accgctacca cacccactgt  1500
gacccaaggc tgaatgcctc ccagtccctt gagcttgcct tcatcattgc cgagaggctc  1560
aggaagagga ggatgcggtc agggctcaac aacagcctgc cgctgccact ggctttctaa  1620

SEQ ID NO: 27          moltype = AA  length = 548
FEATURE                Location/Qualifiers
source                 1..548
                       mol_type = protein
                       organism = Setaria viridis
SEQUENCE: 27
MALATNAAAA AAAAAAISGA AASQPSRAPS FLPMRRRCAV RAVHAAEPSK SHGVPAAAKT  60
SAPTVAPEKE AAPVAAPAPA PKAPAKWAVD SWRSKKALQL PEYPNAAELE AVLKTIEAFP  120
PIVFAGEARH LEERLADAAM GRAFLLQGGD CAESFKEFNS NNIRDTFRVL LQMSAVLMFG  180
AQMPVVKVGR MAGQFAKPRS DPFEVRDGVK LPSYRGDNIN GEAFDEKSRV PDPQRMIRAY  240
AQSAATLNLI RAFATGGYAA MQRVTQWNLD FTEHSEQGDR YRELAHRVDE ALGFMSAAGL  300
TADHTLMKTT EFWTSHECLL LPYEQALTRQ DSTSGLFYDC SAHMLWVGER TRQLDGAHVE  360
FLRGIANPLG IKVSDKMNPS DLVKLIEILN PSNKPGRITI ITRMGAENMR VKLPHLIRAV  420
RQAGLIVTWI TDPMHGNTIK APCGLKTRPF DSILAEVRAF FDVHEQEGSH PGGVHLEMTG  480
QNVTECIGGS RTVTFDDLSD RYHTHCDPRL NASQSLELSF IIAERLRKRR IRSSSGLNNI  540
LPLPPFGF                                                           548

SEQ ID NO: 28          moltype = DNA  length = 1647
FEATURE                Location/Qualifiers
source                 1..1647
                       mol_type = other DNA
                       organism = Setaria viridis
SEQUENCE: 28
atggcgctcg ccaccaacgc cgccgccgcc gccgccgcgg cagcagccat ctccggcgcg  60
gccgcgtcgc agcccagccg cgctccgtcg ttcctcccga tgaggcgccg ctgcgccgtc  120
cgcgccgtgc acgccgcgga gccgtcgaag agccacggcg tcccggcggc ggcgaagacc  180
tcggcgccga cggtcgcgcc cgagaaggag gcggcccctg tcgcggcccc ggccccggcc  240
ccgaaggcgc ctgcgaagtg ggcggtggac agctggaggt cgaagaaggc cctgcagctg  300
ccggagtacc cgaacgcggc ggagctggag gccgtgctca agacgatcga ggccttcccg  360
ccgatcgtgt tcgccgggga ggcgcgccac ctcgaggagc gcctcgccga cgccgccatg  420
ggccgcgcct tcctcctcca gggcggcgac tgcgccgaga gcttcaagga gttcaacagc  480
aacaacatcc gcgacacctt ccgcgtcctc ctccagatgt ccgccgtcct catgttcggc  540
```

-continued

```
gcccaaatgc ccgtcgtcaa ggttggtagg atggccgggc aattcgccaa gccgaggtcg  600
gacccgtttg aggtcaggga cggggtgaag ctgcccagct accggggcga caacatcaac  660
ggcgaggcgt tcgacgagaa gagccgcgtc cccgacccgc agaggatgat cagggcgtac  720
gcccaatccg ccgccacgct caacctgatc cgtgcgttcg ccaccggagg gtacgccgcc  780
atgcagcgcg tcacgcagtg gaacctcgat ttcaccgaac acagcgagca gggcgacagg  840
taccgtgaat tggcacatcg ggttgatgaa gcccttggct tcatgtctgc agctgggcta  900
acagcggacc acacgttgat gaagactact gaattctgga cctcacatga atgcctcctc  960
ctaccctacg aacaagctct aacccgtcag gattccacct ctggtctttt ctatgattgc 1020
tctgcccaca tgctctgggt tggcgagcgc actcgccagc ttgatggtgc tcatgttgag 1080
ttcctaaggg gcattgccaa ccccttggc atcaaggtga gtgacaaaat gaaccccagc 1140
gatttggtga agctgattga gatattgaat ccatcaaaca agcctgggag aattaccatc 1200
attacaagga tgggggcgga gaacatgagg gtcaagttac ctcatcttat tcgtgctgtc 1260
cgccaagcag gactaattgt cacatggatc actgatccca tgcacggtaa caccatcaag 1320
gctccttgcg gtctaaagac tcgccccttc gactccattt tggctgaggt acgggccttt 1380
tttgatgtgc acgagcaaga gggaagccac ccaggaggcg tccaccttga gatgactggg 1440
caaaacgtaa ctgaatgcat cggtggatca cgcaccgtga cttttgatga cctgagcgac 1500
cgctaccaca ctcactgcga cccaaggttg aacgcgtccc agtcgctgga gctctcattt 1560
atcattgccg agaggctgag gaagaggagg atccggtcat cgtctggtct caacaacatc 1620
ttgcccttgc cgcctttgg tttctga                                      1647

SEQ ID NO: 29          moltype = AA  length = 506
FEATURE                Location/Qualifiers
source                 1..506
                       mol_type = protein
                       organism = Setaria viridis
SEQUENCE: 29
MPLAPSTPAL LPNPALPSPG RPRSRGALLR ARAVRAAPRP PSRWSVGSWR ERPAQQQPEY  60
PDKAELDEVL RTVEAFPPIV FAGEARTLEE RLAEAAVGRA FLLQGGDCAE SFKEFNANNI  120
RDTFRVLLQM SVVLMFGGQM PIVKVGRMAG QFAKPRSDGI EERDGVKLPS YRGDNINGDT  180
FDEKSRLPDP HRLIRAYSQS AATLNLLRAF ATGGYAAMQR VTQWNLDFTQ NCEQGDRYME  240
LAHRVDEALG FMSAAGLTLD HPIMTTTEFW TSHECLLLPY EQALTREDST TGLYYDCSAH  300
FLWVGERTRQ LDGAHVEFLR GIANPLGIKV SDKMDPAELV RLIDILNPEN RAGRITIITR  360
MGPENMRVKL PHLIRAVRGA GQIVTWVTDP MHGNTMKAPC GLKTRSFDRI LAEVRAFFDV  420
HEQEGSYPGG VHLEMTGQNV TECIGGSRTV TFDDLTSRYH THCDPRLNAS QSLEMAFIIA  480
ERLRKRRIAS WNLNGNQLGS IPSLGL                                       506

SEQ ID NO: 30          moltype = DNA  length = 1521
FEATURE                Location/Qualifiers
source                 1..1521
                       mol_type = other DNA
                       organism = Setaria viridis
SEQUENCE: 30
atgccactcg cgccatcaac ccccgcgctg ctgcccaacc cggccctccc ctcgccgggg  60
cggccgcgct cccgcggggc gctcctccgc gcccgcgccg tgcgggcggc gcctcggccc  120
ccgagccggt ggtcggtggg gagctggcgg gaacgcccgg cgcaacagca gccggagtac  180
ccggacaagg cggagctgga tgaggtgctg cggacggtgg aggcgttccc gcccatcgtc  240
ttcgccgggg aggcgcgcac cctcgaggag cggctcgcgg aggcggccgt gggccgggcg  300
ttcctcctcc agggcggcga ctgcgccgag agcttcaagg agttcaacgc caacaacatc  360
agggatacct tccgggtcct tctgcaaatg tccgtcgtgc tcatgttcgg aggacagatg  420
cctatcgtca aggtgggaag aatggcaggc cagtttgcaa agccaagatc agatggtatt  480
gaggaacggg atggagtgaa gctgccaagc tatagagggg acaatataaa tggggacaca  540
ttcgatgaga agtcaagatt gccagatcca caccgcctga taagggcgta ctcacaatct  600
gcagctacac tgaatttgtt gcgggcattc gctactggag gttatgctgc aatgcagagg  660
gtaacacagt ggaaccttga cttcacccag aattgtgaac agggcgatag gtacatggag  720
ttggcccacc gagttgacga ggctttggga ttcatgtcag ctgctggact cactttagat  780
caccctataa tgacaacaac agaattctgg acatcgcatg agtgccttct tcttccttac  840
gagcaggcgc ttactcgtga agactccacc acaggcctct attatgactg ctctgcccac  900
ttcctatggg ttggagagcg tactcgtcaa cttgatggtg cccatgtgga gtttcttcgg  960
ggcattgcca accctcttgg tatcaaggtc agtgacaaga tggacccagc agaacttgtg 1020
cggttgattg atatcttgaa tcctgaaaac agggcaggga gaataaccat catcacaaga 1080
atgggacctg aaaacatgag ggtgaaactt cctcacctga tacgtgctgt ccgtggggct 1140
ggccagatag taacatgggt cactgaccca atgcatggga acaccatgaa ggccccttgt 1200
ggactcaaaa cccgctcctt cgacagaatt ttggctgagg tgcgtgcatt ctttgatgtt 1260
cacgaacaag aagggagcta tccaggaggc gtgcatctgg agatgaccgg acagaatgtg 1320
acagaatgca ttggtggatc acggacagtg acatttgatg atctgacctc acgctaccac 1380
acacactgcg acccaaggct caatgcctcg caatcgctgg agatggcatt catcatcgct 1440
gagcgcttga ggaaaaggag gattgcctca tggaatttga acgggaacca gctgggttcc 1500
attccatcat tggggctgta a                                           1521

SEQ ID NO: 31          moltype = AA  length = 489
FEATURE                Location/Qualifiers
source                 1..489
                       mol_type = protein
                       organism = Setaria viridis
SEQUENCE: 31
MYCMRSVSRA APASFPLAVR SPRFDLPSLP TAPPAPPSAP RLLVPQTRPR RGPTMYPDPA  60
```

-continued

```
ALEATERALE TFPPLVFAGE ARKLEERLGE AFLLQGGDCA ESFKEFSANN IRDTFRLMLQ  120
MAVILTFGGQ MPTIKVGRLG GQFAKPRSNP TETRDGVTLP LYTYRGDIIN GDTFDEKSRV  180
PDPERLIKAY NQSASTLNLL RGFTHGGFAD LQRVTQWNLD FLRHSTQGDR YLELSQRVQD  240
AVGFMAAAGL TTRHPIMTTA ELWTSHECLH LQYEQALTRE DSISGMYYDC SGHMLWVGER  300
TRQLDGAHVE FLRGISNPLG VKVSDKLDPS ELVKLCEILN PHNKPGRLTI ITRMGAENMR  360
VKLPHMIGAV RQAGMIVTWV SNPMHGNTIR APCGLKTRSF DATRAFFDVH EQEGSHPGGV  420
HLEMTGQDVT ECIGGSKAVT FDDLGDRYHT HCDPRLNASQ SLELAFAIAD RLRKKRNMTW  480
NKLMSRAIA                                                          489

SEQ ID NO: 32          moltype = DNA  length = 1470
FEATURE                Location/Qualifiers
source                 1..1470
                       mol_type = other DNA
                       organism = Setaria viridis
SEQUENCE: 32
atgtattgca tgcgaagcgt ttcgcgcgcc gcgcccgcct cctttcctct cgcggttcgc  60
tccccccgat tcgatctccc ctccctgccc accgcgccgc cggctccccc cagtgcccca  120
cggctcctcg tgccgcagac gaggccgcgg cggggcccca cgatgtaccc tgacccggcc  180
gcgctggagg ccacggagcg cgcgctggag acgttcccgc cgctggtgtt cgcgggcgag  240
gcgcggaagc tggaggagcg gctcggggag gcgttcctcc tgcagggcgg cgactgcgcc  300
gagagcttca aggagttcag cgccaacaac atccgcgaca cgttccggct catgctgcag  360
atggccgtca tcctcacctt cggcggccag atgcccacca tcaaggttgg aaggttgggt  420
ggtcaatttg ctaaaccaag atcgaaccca actgagacta gagatggggt aacgcttccc  480
ctctatacct atagagggga tatcatcaat ggggacactt ttgatgagaa atcacgtgta  540
ccagatcctg aaaggctgat taaggcctac aaccagtcag caagcacact gaatcttctc  600
agaggattta cccatggagg gtttgcagat cttcagagag tcacccagtg gaatctcgat  660
ttcttgaggc acagcacaca aggagacagg tatctggagc tttctcagag ggttcaggat  720
gccgttgggt ttatggctgc tgctggtctg actactcgac accccatcat gaccacagct  780
gagttatgga catcccatga gtgccttcac ttacaatatg agcaagcact gactagggag  840
gactccattt ctgggatgta ctatgactgt tctggacaca tgctttgggt cggtgagagg  900
actaggcagc tggatggtgc tcatgttgaa ttccttcgtg gtatttcaaa tcctctgggt  960
gtaaaggtaa gtgataagct tgatccatca gagctggtga agttgtgtga gattctgaac  1020
cctcacaaca agcccggtag attgaccatt atcacaagaa tgggggctga gaacatgcgt  1080
gtcaagcttc cacatatgat cggagcagtt cgccaagctg ggatgattgt tacttgggtt  1140
agcaatccaa tgcacgggaa caccatcagg gcaccgtgtg ggctgaagac aagatcattc  1200
gatgcaacca gggctttctt tgatgttcat gagcaagagg gaagccaccc cggaggtgtt  1260
cacttggaga tgacggggca ggacgttaca gaatgtattg gtggatccaa ggcagtgaca  1320
tttgatgatc tcggtgatcg ctaccacaca cactgtgacc ccaggctgaa cgcatcgcag  1380
tccctcgagc tggcttttgc cattgctgac aggctaagga agaagagaaa catgacatgg  1440
aacaaattga tgtccagggc tattgcttaa                                   1470

SEQ ID NO: 33          moltype = AA  length = 387
FEATURE                Location/Qualifiers
source                 1..387
                       mol_type = protein
                       organism = Brachypodium distachyon
SEQUENCE: 33
MLSSSTSHLR LHQPSRAARR LPPAPAPATY QFQSPASRRS LPAPAVGARI RPNRGGTIRA  60
IDAAQPFDYE SRAAGLLEER QRLKIAIVGF GNFGQFLART FARQGHTLLA HSRSDHSSLA  120
ASLGAAYFQD PHDLCECHPD VVLLATSILS AEAVLRSLPV HRLRRNTLFV DVLSVKEFPK  180
NLLLTTLPEG FDIICTHPMF GPESARDGWD GLPFVFDKVR VGDCPARRAR ADTFLNIFER  240
EGCRMVEMSC AEHDAHAAET QFLTHTVGRM LATLELQSTP INTKGYETLL RLVDNTCSDS  300
FDLYNGLFMY NKNSTDLLNR LESAMDSVKK RLFDGLHDVL RKQLFEGKAS PPATSSNTKS  360
DVHRGQLLLE GKASPAPLNA NNTTVRN                                       387

SEQ ID NO: 34          moltype = DNA  length = 1164
FEATURE                Location/Qualifiers
source                 1..1164
                       mol_type = other DNA
                       organism = Brachypodium distachyon
SEQUENCE: 34
atgctgtcgt cttccacctc ccacctccgc ctccaccaac catccagagc ggcgcgccgc  60
ctcccgccgg ctcctgctcc ggcgacctac cagttccagt cgcccgcatc ccgtcgctcc  120
ctccccgcgc cggccgtcgg cgcccgcatc cgccccaacc gcggcggcac catccgcgcc  180
atcgacgccg cccagccgtt cgactacgag tcccgcgcgg cggggctgct ggaggagcgg  240
cagcgcctga agatcgccat cgtcgggttc ggcaacttcg ggcagttcct ggcccggacc  300
ttcgcgcggc agggccacac cctgctcgcc cactcccgct ccgaccactc ctccctcgcc  360
gcctccctgg gcgccgccta cttccaggac ccgcacgacc tctgcgagtg ccacccggac  420
gtggtcctcc tggccacctc catcctctcc gccgaggccg tcctccgctc gctccccgtc  480
caccgcctcc gccgcaacac cctcttcgtc gacgtcctct ccgtcaagga attccccaag  540
aacctcctcc tcaccacgct cccggagggc ttcgacatca tctgcaccca ccccatgttc  600
ggcccggagt cggcccgcga cggctgggac ggcctccctt tcgtcttcga caaggtccgc  660
gtcggcgact gccccgcccg ccgcgcccgc gccgacacct tcctcaacat cttcgagcgc  720
gagggctgcc gcatggtgga gatgtcctgc gccgagcacg acgcgcacgc cgccgagacg  780
cagttcctca cgcacaccgt cggccgcatg ctcgccacgc ttgagctcca gtccacccgc  840
atcaacacca agggttacga gacgctgctc cgcctcgtcg acaacacctg cagcgacagc  900
```

-continued

```
ttcgacctct acaacggact cttcatgtac aacaagaact ccaccgacct gctcaaccgc  960
ctcgagtccg ccatggattc cgtcaagaag aggctcttcg acggcctcca cgacgtgctc  1020
cggaagcagc tcttcgaggg caaggcgtcg ccgcccgcca ccagcagcaa caccaaatct  1080
gacgtgcacc gggggcagct cctcttggag ggcaaggcct cgccggcgcc gctcaacgcc  1140
aacaacacca ccgtccggaa ttag                                         1164

SEQ ID NO: 35          moltype = AA  length = 365
FEATURE                Location/Qualifiers
source                 1..365
                       mol_type = protein
                       organism = Brachypodium distachyon
SEQUENCE: 35
MSSSGRFHQP PSCRRPAALA QSLRASTTTV TCRWRQQYHH PLAAAPLRLR AVPVRATDAA  60
RQPFDHLSGA VKSEEGTHPR LKIAIVGFGN YGQFLARTMV QQGHTVLAHS RSDHSAAAAT  120
IGASFYADAH DLCECQPDVV LLSTSILSAE AVLRSLPVHR FRRSTLFADV LSVKEFPKNL  180
LLAYLPGDFD VICTHPMFGP ESARDGWAGL PFVFDEVRVG DGPARRARAD AFLDVFAREG  240
CRMVEMSCAE HDAHAAETQF LTHTVGRMLA TLDLKSTPIN TKGYETLLRL VDNTCSDSFD  300
LYNGLFMYNN NATELLHRLE SAMDSVKRRL FDGLHEVLRR QLFEGSPPLN RDSSFPAESS  360
LDDGR                                                              365

SEQ ID NO: 36          moltype = DNA  length = 1098
FEATURE                Location/Qualifiers
source                 1..1098
                       mol_type = other DNA
                       organism = Brachypodium distachyon
SEQUENCE: 36
atgtcttcct ccggtcgctt ccaccaaccc ccttcctgcc gacggccggc ggctctagcc  60
cagtctctac gcgcgtccac caccacggtg acatgccggt ggcggcagca gtaccatcat  120
cccctagctg ctgctccgct ccgcctccgc gctgtgcctg tgcgtgccac ggacgccgca  180
cgccaaccgt tcgaccattt gtccggcgcc gttaaaagcg aggagggcac gcatccgcgc  240
ctcaagatcg ccatcgtcgg gttcggcaac tacggccagt ttctggcgcg gacgatggtg  300
cagcaggggc acaccgtgct ggcccactcc cgctccgacc actccgccgc cgcggccacc  360
atcggcgcgt ccttctacgc cgacgcgcac gacctctgcg agtgccagcc cgacgtggtc  420
ctcctatcca cctcgatcct ctccgcggag gccgtcctcc gctcgctccc cgtccaccgc  480
ttccgccgca gcaccctctt cgccgacgtc ctctccgtca aggagttccc caagaacctc  540
ctcctcgcct acctcccggg ggacttcgac gtgatctgca cccaccccat gttcggcccg  600
gagtcggccc gcgacggctg ggccgggctc cccttcgtct tcgacgaggt ccgcgtcggc  660
gacggcccgg cccgccgcgc ccgcgccgac gccttcctcg acgtcttcgc gcgcgagggc  720
tgccgcatgg tggagatgtc ctgcgcggag cacgacgcgc acgccgccga gacgcagttc  780
ctgacgcaca ccgtcggccg gatgctcgcc acgctggacc tcaagtccac gccgatcaac  840
accaagggat acgagacgct gctccgcctc gtcgacaaca cctgcagcga cagcttcgac  900
ctctacaacg gcctcttcat gtacaacaac aacgccacgg agctgctcca ccgcctggaa  960
tccgccatgg actccgtcaa gaggaggctc ttcgacggcc tccacgaggt gctcaggagg  1020
cagctcttcg aaggctcgcc gccgttgaac agggattcct ctttcccagc cgagtcatcg  1080
ttggatgacg gacggtga                                                1098

SEQ ID NO: 37          moltype = AA  length = 361
FEATURE                Location/Qualifiers
source                 1..361
                       mol_type = protein
                       organism = Brachypodium distachyon
SEQUENCE: 37
MASSLVHLSS PAGGRTAAAP PSLLSRFSPN FCAFATLRPG PTRPTAATPK HARARTCAEQ  60
EQKQEVAAPC RDAYEKPAVW NTTAADSAEA APPLRVGIIG FGNFGQFIAR GIQRQGHAVL  120
ATSRSDYSAY CSAQGIRYFR SLEALCEEQP NVLLVCSSIL STEAVVRAIP FHKLRSDTIV  180
ADVLSVKQFP RNLLLEILPP EFGIVCTHPM FGPESGKHGW STLPFVYDKV RLADKGDQKA  240
NCGQFLSIFE GEGCRMVEMS CAEHDRHAAA SQFITHTIGR VLAQLNLKST PINTKGFEAL  300
LKLTENTVSD SFDLYYGLFM YNVNATEQIE KLERAFEKVK QMLYGRLHDI LRKQIVERVP  360
I                                                                  361

SEQ ID NO: 38          moltype = DNA  length = 1086
FEATURE                Location/Qualifiers
source                 1..1086
                       mol_type = other DNA
                       organism = Brachypodium distachyon
SEQUENCE: 38
atggcttcct cccttgtcca cctaagctcc cccgccggtg gccgcaccgc cgccgcgccg  60
cccagcttgc taagccgctt cagccccaac ttctgcgcct tcgctacact ccgcccaggc  120
ccgaccagac cgaccgccgc caccccccaag cacgcgagag ccaggacctg cgccgagcag  180
gagcaaaagc aggaggtcgc tgccccgtgc cgtgacgcct acgagaagcc ggccgtatgg  240
aatacgaccg cggcagactc cgcggaggcg gcgccgccgc tacgcgtggg gatcatcggc  300
ttcggcaact tcgggcagtt catcgccagg ggcatccagc gccagggcca tgccgtgctg  360
gccacttcca gatctgacta ctccgcctac tgctccgccc aagggattcg ctacttcagg  420
agcttggagg cgctgtgcga ggagcagccc aacgtgctgc tggtgtgcag ctcaatcctc  480
tcgacggagg ccgtcgtccg ggcaatcccc ttccacaagc tccgctctga caccatcgta  540
gctgacgtgc tctccgtcaa gcagtttccc cgtaacctcc tcctcgagat cctgccgccg  600
```

-continued

```
gagtttggga tcgtctgcac acaccccatg tttgggccgg agagtggtaa gcatggctgg  660
agcacactgc ccttcgtcta tgacaaggtt cgccttgcgg acaaaggaga tcagaaagcc  720
aattgcggcc agttcttgag catctttgag ggagagggat gtcggatggt ggagatgtca  780
tgcgcagaac atgatcgcca cgctgcggca agtcaattca tcacacacac tattgggagg  840
gttctggcgc aactaaatct caagtccacg ccaatcaaca ccaaaggttt tgaggccctc  900
ctgaaactta cagaaaacac cgtgagcgat agtttcgatc tatactacgg gctcttcatg  960
tacaatgtga atgccacaga gcagattgag aaactggaaa gggcatttga gaaggtgaag 1020
cagatgctgt atggtagact ccatgacata ctaagaaagc agatcgtaga gagggtcccc 1080
atttaa                                                            1086

SEQ ID NO: 39          moltype = AA  length = 362
FEATURE                Location/Qualifiers
source                 1..362
                       mol_type = protein
                       organism = Sorghum bicolor
SEQUENCE: 39
MLSSSTSTLR LHQPTRPHRH HPPAPAAGGA THLASPRRWR GHAPGASSPP ALRARAQRIR  60
ALDAAQPFDF ESRAAGLLEE RQRLKIAIVG FGNFGQFLAR TFARQGHTLL AHSRTDHSAL  120
ASTLGASFFT DPHDLCECHP DVVLLATSIL SAEAVLRSLP VHRLRRNTLF VDVLSVKEFP  180
RNLLLSSLPP DFDVICTHPM FGPESARDGW DGLPFVFDKV RVGDCPARRA RAEAFLNIFE  240
REGCRMVEMS CAEHDAHAAE TQFLTHTVGR MLAMLELRST PINTKGYETL LRLVDNTCSD  300
SFDLYNGLFM YNKNSTELLN RVEWAMDSVK KKLFDGLHDV LRKQLFEGSP HAPNVSSSNV  360
RK                                                                362

SEQ ID NO: 40          moltype = DNA  length = 1089
FEATURE                Location/Qualifiers
source                 1..1089
                       mol_type = other DNA
                       organism = Sorghum bicolor
SEQUENCE: 40
atgctgtcgt cttccacctc caccctccgc ctccaccaac caacccgtcc ccaccgccac  60
cacccgccgg ccccggccgc cgggggcgcc acccacctag cctccccgcg gcggtggcgc  120
ggccacgccc cgggggcgtc gtccccgccg gcgctccgcg cacgggccca gcgcatccgc  180
gcgctggacg ccgcccagcc gttcgacttc gagtcccgcg cggcggggct gctggaggag  240
cgtcagcgcc tgaagatcgc catcgtcggg ttcggcaact tcgggcagtt cctggcgcgc  300
accttcgcgc ggcaggggca cacgctgctc gcccactccc gcaccgacca ctccgcgctg  360
gcgtccacgc tgggggcctc cttcttcacc gacccgcacg acctgtgcga gtgccacccg  420
gacgtggtgc tcctcgccac ctccatcctc tccgcggagg ccgtgctccg ctcgctgccc  480
gtccaccgtc tccgccgcaa cacgctcttc gtcgacgtgc tctccgtcaa ggagttcccc  540
aggaacctgc tgctcagctc gttaccgccc gacttcgacg tcatctgcac ccaccccatg  600
ttcgggccgg agtcggcgcg cgacggctgg gacggccttc ccttcgtgtt cgacaaggtg  660
cgcgtgggcg actgcccggc ccgccgcgcc cgcgccgagg cgttcctcaa catcttcgag  720
cgggaagggt gccggatggt ggagatgtcg tgcgccgagc acgacgcgca cgccgccgag  780
acccagttcc tgacgcacac cgtcgggagg atgctcgcca tgctggagct ccgctccacg  840
cccatcaaca ccaaggggta cgagacgctg ctccgcctgg tcgacaacac ctgcagcgac  900
agcttcgacc tctacaacgg gctcttcatg tacaacaaga actccaccga gctgctcaac  960
cgcgtcgagt gggccatgga ctccgtcaag aagaagctct tcgacggcct ccacgacgtg 1020
ctccggaagc agctcttcga ggggtcgccg cacgccccca atgtctcctc ctccaacgtc 1080
cgcaagtaa                                                         1089

SEQ ID NO: 41          moltype = AA  length = 389
FEATURE                Location/Qualifiers
source                 1..389
                       mol_type = protein
                       organism = Sorghum bicolor
SEQUENCE: 41
MASSSCVRLH HLPSTSRRAT APTNLQFRAA SCRWRRLSIP VPGVSPPPLR LRATGASQPL  60
GTDSNEEVEK VEEQPQPPPR LKIAVVGFGT FGQFLARTLV AQGHTVLAHS RSDHSAAAAS  120
MGALFFSDPH DLCECHPDVV LLATSILSAE SVVRSLPLHR LRRDTLFADV LSVKEFPKRL  180
LLGLLPEEMD ILCTHPMFGP ESARAGWAGL PFMFDKVRVR DTVPARRARA EAFLDVFAQE  240
GCRMVEMSCA EHDAHAAETQ FLTHTVGRML AALELRATPI DTRGYETLLR LVENTCSDSF  300
DLYNGLFMYN NNSTELLNRL DWAMDAVKRR LFDGLHDVLR RQLFHVGEVE GEAERMEELA  360
VVPGGGPDTD DDGCATATAT SIKSGEENN                                   389

SEQ ID NO: 42          moltype = DNA  length = 1170
FEATURE                Location/Qualifiers
source                 1..1170
                       mol_type = other DNA
                       organism = Sorghum bicolor
SEQUENCE: 42
atggcgtcat cttcctgcgt ccgcctccat cacctgccgt ccacctctcg ccgggcgacc  60
gctccaacca accttcagtt ccgcgcggcg agctgccggt ggcggcggct ttccattccc  120
gttcccggcg tgtcccctcc tcccctccgc ctccgcgcca cgggtgcctc gcagccgctt  180
ggtactgaca gcaacgagga ggtcgagaag gtggaggagc agcctcagcc tccgccgcgg  240
ctcaagatcg ccgtggtcgg gttcggcacc ttcggacagt tcctggcgcg cacgctggtg  300
gcgcaggggc acacggtgct ggcgcactcc cggtccgacc actccgccgc ggcggcgtcc  360
```

-continued

```
atgggcgcgc tcttcttctc ggacccgcac gacctgtgcg agtgccaccc ggacgtggtg  420
ctcctggcca cctccatcct gtccgcggag tccgtggtcc ggtcgctccc gctgcaccgc  480
ctccgccgcg acaccctctt cgccgacgtg ctctccgtga aggagttccc caagaggctc  540
ctcctgggct tgctcccgga ggagatggac atcctctgca cgcacccgat gttcgggccg  600
gagtcggcgc gcgccggctg ggcgggcctc cccttcatgt tcgacaaggt gcgcgtccgg  660
gacaccgtcc cggcgcgccg cgcccgcgcc gaggcgttcc tggacgtgtt cgcgcaggaa  720
gggtgccgga tggtggagat gtcgtgcgcc gagcacgacg cgcacgccgc cgagacgcag  780
ttcctgacgc acaccgtcgg taggatgctg gcggcgctgg agctccgggc gacgccgatc  840
gacacgagag ggtacgagac gctgctccgg ctggtggaga acacgtgcag cgacagcttc  900
gacctctaca acggcctgtt catgtacaac aacaactcca ccgagctgct caaccgcctc  960
gactgggcca tggacgccgt caagaggagg ctcttcgacg gcctccacga cgtgctccgg 1020
aggcagttgt tccacgtcgg ggaggtggag ggggaggcgg agcgcatgga ggagctggcg 1080
gtggtgccgg gtggtggtcc ggacaccgat gacgacgggt gcgcgacggc gacggcgacg 1140
agcatcaagt cgggagagga gaataattga                                  1170

SEQ ID NO: 43          moltype = AA  length = 382
FEATURE                Location/Qualifiers
source                 1..382
                       mol_type = protein
                       organism = Sorghum bicolor
SEQUENCE: 43
MASSLRHFAG LGCFPAVAST SGSTCCLRRY TPNFCAFVAL RPISPPAATA TATAKALTSP  60
SPVEQHLQAV VPCHGISDPP AASSAAAVPA APLRVGIVGF GNFGQFIAGG LQRQGHVVLA  120
ASRSDYSVYC ASHGIRFFRS VDALCEEQPD VLLICSSILS TEGVVRAIPF RKLRHDTIVA  180
DVLSVKEFPR NLLLEVLPPG FGIICTHPMF GPESGKHGWG KLPFVFDKVR VAEDGDQAAK  240
CDQFLSIFEQ EGCRMVEMSC AEHDRYAAGS QFITHTIGRV LSQLNLKSTP INTKGYETLL  300
QLTKNTVSDS FDLYYGLFMY NVNATEQLDK LEMAFEKVRQ MLSGRLHDFI RKQIVERAAH  360
VPADPSGKLA NGLSSSPAAR LL                                           382

SEQ ID NO: 44          moltype = DNA  length = 1149
FEATURE                Location/Qualifiers
source                 1..1149
                       mol_type = other DNA
                       organism = Sorghum bicolor
SEQUENCE: 44
atggcctcct cgctccgcca tttcgccggc ctcggctgtt tccccgccgt ggcctctacc  60
tctggctcca cctgctgcct gcgccgctac accccaaact tctgcgcctt cgtggcgctc  120
cgtccgatca gcccgcccgc cgccaccgcc acggcgactg ccaaggccct cacctctccg  180
tctccggtgg agcagcacct acaggccgtc gtcccgtgcc acggcatcag cgacccgccg  240
gcggcgtcat cggcagcagc ggtcccggcg gcgccgctgc gggtggggat cgtcgggttc  300
ggcaacttcg ggcagttcat cgcgggcggg ctgcagcggc agggccacgt cgtgctggcg  360
gcttccagat ccgactactc cgtctactgc gccagccatg gcattcgctt cttcaggagc  420
gtcgacgcgc tgtgcgagga gcagccggac gtgctgctca tctgcagctc catcctgtcc  480
acggagggcg tcgtccgagc catccccttc cgcaagctcc gccacgacac catcgtcgcc  540
gacgtgctct ccgtcaagga gttccctcgc aacctcctcc tcgaggttct cccaccggga  600
tttgggatca tctgtacgca ccccatgttt gggccggaga gtggtaaaca cggctggggc  660
aagcttccct tcgtctttga caaggtccgt gtcgcggaag acggggatca ggcagcaaag  720
tgcgaccagt tcttgagcat atttgaacag gagggatgta ggatggtgga gatgtcatgc  780
gcggagcatg atcgctacgc tgcgggaagt caattcatca cgcacacaat tgggagggtt  840
ttatcacaac tgaacctcaa gtcaacgcca atcaacacca agggttatga aaccttgctg  900
caacttacca agaacaccgt aagcgacagt ttcgatctgt actatgggct cttcatgtac  960
aatgtcaatg ccacagagca gctcgacaaa ctggaaatgg catttgagaa ggtgagacag 1020
atgctgtctg gcaggctcca cgacttcata cgaaagcaga ttgtggagag ggcagcccat 1080
gtgccagcag atccttcagg aaaattggca aatggtttgt ccagttctcc tgcagctcgc 1140
ctcttatag                                                         1149

SEQ ID NO: 45          moltype = DNA  length = 35
FEATURE                Location/Qualifiers
source                 1..35
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 45
gatctagact cgagggtacc atgctgtcgt cttcc                             35

SEQ ID NO: 46          moltype = DNA  length = 31
FEATURE                Location/Qualifiers
source                 1..31
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 46
ctagtgcatg cggccgcatt ccggacggtg g                                 31

SEQ ID NO: 47          moltype = DNA  length = 31
FEATURE                Location/Qualifiers
source                 1..31
                       mol_type = other DNA
```

-continued

```
                     organism = synthetic construct
SEQUENCE: 47
cgcgcggcag ccatatgcgc gccatcgacg c                                  31

SEQ ID NO: 48        moltype = DNA  length = 33
FEATURE              Location/Qualifiers
source               1..33
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 48
gctcgaattc ggatccctaa ttccggacgg tgg                                33

SEQ ID NO: 49        moltype = DNA  length = 36
FEATURE              Location/Qualifiers
source               1..36
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 49
gatctagact cgagggtacc atgtcttcct ccggtc                             36

SEQ ID NO: 50        moltype = DNA  length = 32
FEATURE              Location/Qualifiers
source               1..32
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 50
ctagtgcatg cggccgcccg tccgtcatcc aa                                 32

SEQ ID NO: 51        moltype = DNA  length = 31
FEATURE              Location/Qualifiers
source               1..31
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 51
cgcgcggcag ccatatgcgt gccacggacg c                                  31

SEQ ID NO: 52        moltype = DNA  length = 32
FEATURE              Location/Qualifiers
source               1..32
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 52
gctcgaattc ggatcctcac cgtccgtcat cc                                 32

SEQ ID NO: 53        moltype = DNA  length = 36
FEATURE              Location/Qualifiers
source               1..36
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 53
gatctagact cgagggtacc atggcttcct cccttg                             36

SEQ ID NO: 54        moltype = DNA  length = 33
FEATURE              Location/Qualifiers
source               1..33
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 54
ctagtgcatg cggccgcaat ggggaccctc tct                                33

SEQ ID NO: 55        moltype = DNA  length = 32
FEATURE              Location/Qualifiers
source               1..32
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 55
cgcgcggcag ccatatggcc gagcaggagc aa                                 32

SEQ ID NO: 56        moltype = DNA  length = 34
FEATURE              Location/Qualifiers
source               1..34
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 56
gctcgaattc ggatccttaa atggggaccc tctc                               34
```

-continued

```
SEQ ID NO: 57          moltype = DNA  length = 39
FEATURE                Location/Qualifiers
source                 1..39
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 57
tcactctgtg gtctcaaatg cgcgcgctgg acgccgccc                          39

SEQ ID NO: 58          moltype = DNA  length = 41
FEATURE                Location/Qualifiers
source                 1..41
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 58
ccacttcgtg gtctcacgaa ctcttgcgga cgttggagga g                       41

SEQ ID NO: 59          moltype = DNA  length = 39
FEATURE                Location/Qualifiers
source                 1..39
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 59
tcactctgtg gtctcaaatg cgcgccacgg gtgcctcgc                          39

SEQ ID NO: 60          moltype = DNA  length = 44
FEATURE                Location/Qualifiers
source                 1..44
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 60
ccacttcgtg gtctcacgaa ctattattct cctctcccga cttg                    44

SEQ ID NO: 61          moltype = DNA  length = 40
FEATURE                Location/Qualifiers
source                 1..40
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 61
tcactctgtg gtctcaaatg agcccgcccg ccgccaccgc                         40

SEQ ID NO: 62          moltype = DNA  length = 42
FEATURE                Location/Qualifiers
source                 1..42
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 62
ccacttcgtg gtctcacgaa cttaagaggc gagctgcagg ag                      42

SEQ ID NO: 63          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 63
atggcctcct cgctccgcc                                                19

SEQ ID NO: 64          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 64
gtatccggtt gaagtgtagg                                               20

SEQ ID NO: 65          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 65
caccaccgtc cggaattagc                                               20

SEQ ID NO: 66          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
source                 1..20
                       mol_type = other DNA
```

-continued

```
                     organism = synthetic construct
SEQUENCE: 66
gcaccagttt ctccccaaag                                              20

SEQ ID NO: 67        moltype = DNA  length = 20
FEATURE              Location/Qualifiers
source               1..20
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 67
gatgacggac ggtgatctcg                                              20

SEQ ID NO: 68        moltype = DNA  length = 20
FEATURE              Location/Qualifiers
source               1..20
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 68
ttcgtaccgc ttgttggtcg                                              20

SEQ ID NO: 69        moltype = DNA  length = 19
FEATURE              Location/Qualifiers
source               1..19
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 69
tgctgtgtcc cctctcctc                                               19

SEQ ID NO: 70        moltype = DNA  length = 20
FEATURE              Location/Qualifiers
source               1..20
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 70
agggctgaaa gacactgggc                                              20

SEQ ID NO: 71        moltype = DNA  length = 21
FEATURE              Location/Qualifiers
source               1..21
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 71
cagacaatgc ggagatgatc g                                            21

SEQ ID NO: 72        moltype = DNA  length = 22
FEATURE              Location/Qualifiers
source               1..22
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 72
tttgcttcag aaaccatgtc ac                                           22

SEQ ID NO: 73        moltype = DNA  length = 21
FEATURE              Location/Qualifiers
source               1..21
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 73
ggtagtaatt cagtgcgtcg g                                            21

SEQ ID NO: 74        moltype = DNA  length = 21
FEATURE              Location/Qualifiers
source               1..21
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 74
ggtgttcttc ttccagagag g                                            21

SEQ ID NO: 75        moltype = DNA  length = 20
FEATURE              Location/Qualifiers
source               1..20
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 75
gatcgcttcc atcccaaggc                                              20
```

-continued

```
SEQ ID NO: 76          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 76
caggcggtct gaaaggaagg                                              20

SEQ ID NO: 77          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 77
agttgtcgcg tgtctgagtc                                              20

SEQ ID NO: 78          moltype = DNA  length = 21
FEATURE                Location/Qualifiers
source                 1..21
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 78
acacgggctc acttattcat c                                            21

SEQ ID NO: 79          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 79
gggctcattg tgctgctgtc                                              20

SEQ ID NO: 80          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 80
ccggaggaca taggacttgc                                              20

SEQ ID NO: 81          moltype = DNA  length = 42
FEATURE                Location/Qualifiers
source                 1..42
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 81
ggtgccgcgc ggcagccata tggccgtcca cgccgcggag cc                     42

SEQ ID NO: 82          moltype = DNA  length = 43
FEATURE                Location/Qualifiers
source                 1..43
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 82
cggagctcga attcggatcc tcagaaacca taggttggca atg                    43

SEQ ID NO: 83          moltype = DNA  length = 42
FEATURE                Location/Qualifiers
source                 1..42
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 83
ggtgccgcgc ggcagccata tggccgtgca cgccgccgac cc                     42

SEQ ID NO: 84          moltype = DNA  length = 43
FEATURE                Location/Qualifiers
source                 1..43
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 84
cggagctcga attcggatcc ttagaaggcc aatggcggca gtg                    43

SEQ ID NO: 85          moltype = DNA  length = 42
FEATURE                Location/Qualifiers
source                 1..42
                       mol_type = other DNA
```

-continued

```
                     organism = synthetic construct
SEQUENCE: 85
ggtgccgcgc ggcagccata tgatccgcgc gcacgcggtg cg                    42

SEQ ID NO: 86        moltype = DNA  length = 41
FEATURE              Location/Qualifiers
source               1..41
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 86
cggagctcga attcggatcc tcagagtccc atggatgatg g                     41

SEQ ID NO: 87        moltype = DNA  length = 42
FEATURE              Location/Qualifiers
source               1..42
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 87
ggtgccgcgc ggcagccata tgcgcgcgac gtcggtcgcg gc                    42

SEQ ID NO: 88        moltype = DNA  length = 46
FEATURE              Location/Qualifiers
source               1..46
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 88
cggagctcga attcggatcc ttaagcttct actctagata tcaagc                46

SEQ ID NO: 89        moltype = DNA  length = 42
FEATURE              Location/Qualifiers
source               1..42
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 89
ggtgccgcgc ggcagccata tggccatcca cgccgccgac cc                    42

SEQ ID NO: 90        moltype = DNA  length = 42
FEATURE              Location/Qualifiers
source               1..42
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 90
cggagctcga attcggatcc tcagaaagcc agtggtggca gc                    42

SEQ ID NO: 91        moltype = DNA  length = 41
FEATURE              Location/Qualifiers
source               1..41
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 91
ggtgccgcgc ggcagccata tgctccgcgc ccgcgccgtc c                     41

SEQ ID NO: 92        moltype = DNA  length = 42
FEATURE              Location/Qualifiers
source               1..42
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 92
cggagctcga attcggatcc tcacagacga atggaaccca gc                    42

SEQ ID NO: 93        moltype = DNA  length = 36
FEATURE              Location/Qualifiers
source               1..36
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 93
tcactctgtg gtctcaaatg gccgtgcacg ccgccg                           36

SEQ ID NO: 94        moltype = DNA  length = 38
FEATURE              Location/Qualifiers
source               1..38
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 94
ccacttcgtg gtctcacgaa ctgaaggcca atggcggc                         38
```

-continued

```
SEQ ID NO: 95          moltype = DNA  length = 40
FEATURE                Location/Qualifiers
source                 1..40
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 95
tcactctgtg gtctcaaatg gccgtccacg ccgcggagcc                        40

SEQ ID NO: 96          moltype = DNA  length = 42
FEATURE                Location/Qualifiers
source                 1..42
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 96
ccacttcgtg gtctcacgaa ctgaaaccat aggttggcaa tg                     42

SEQ ID NO: 97          moltype = DNA  length = 36
FEATURE                Location/Qualifiers
source                 1..36
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 97
tcactctgtg gtctcaaatg atccgcgcgc acgcgg                            36

SEQ ID NO: 98          moltype = DNA  length = 39
FEATURE                Location/Qualifiers
source                 1..39
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 98
ccacttcgtg gtctcacgaa ctgagtccca tggatgatg                         39
```

What is claimed:

1. A cell engineered to express or overexpress an enzyme selected from:

a) a 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1b (DHS1b) enzyme comprising SEQ ID NO: 3 (BdDHS1b), SEQ ID NO: 19 (OsDHS1b), SEQ ID NO: 27 (SvDHS1b) or a DHS1b enzyme having at least 95% identity to one of SEQ ID NOs: 3, 19 or 27; or

b) a noncanonical TyrA enzyme comprising SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 43 (SbTyrAnc) or a TyrA enzyme having at least 95% identity to SEQ ID NO: 37 or 43;

wherein (i) the cell is engineered to express an enzyme that is not native to the cell, or (ii) the cell is engineered to overexpress an enzyme that is native to the cell as compared to a control cell.

2. The cell of claim **1**, wherein the cell is engineered to express or overexpress both:

a) the DHS1b enzyme; and

b) a TyrA enzyme selected from the group consisting of SEQ ID NO: 33 (BdTyrA1), SEQ ID NO: 35 (BdTyrA2), SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 39 (SbTyrA1), SEQ ID NO: 41 (SbTyrA2), SEQ ID NO: 43 (SbTyrAnc) and an enzyme having at least 95% sequence identity to one of SEQ ID NOs: 33, 35, 37, 39, 41 and 43.

3. The cell of claim **1**, wherein the cell comprises an exogenous nucleic acid encoding the enzyme.

4. The cell of claim **3**, wherein the exogenous nucleic acid is inserted in the genome of the cell.

5. The cell of claim **1**, wherein the cell is a plant cell.

6. A plant comprising the cell of claim **5**.

7. The plant of claim **6**, wherein the plant is a tomato plant, tobacco plant, soybean plant, cotton plant, poplar plant, sorghum plant, corn plant, beet plant, mung bean plant, opium poppy plant, alfalfa plant, wheat plant, barley plant, millet plant, oat plant, rye plant, rapeseed plant, miscanthus plant, or grass plant.

8. The plant of claim **7**, wherein the grass plant is a *Sorghum bicolor* plant, *Oryza sativa* plant, or *Brachypodium distachyon* plant.

9. The plant of claim **6**, wherein the quantity of aromatic amino acids produced by the plant is greater than the quantity produced by a control plant.

10. A method for increasing production of one or more aromatic amino acid or derivative thereof in a cell, the method comprising engineering the cell to express or overexpress an enzyme selected from:

c) a DHS1b enzyme comprising SEQ ID NO: 3 (BdDHS1b), SEQ ID NO: 19 (OsDHS1b), SEQ ID NO: 27 (SvDHS1b) or a DHS1b enzyme having at least 95% identity to one of SEQ ID NOs: 3, 19 or 27; or

d) a noncanonical TyrA enzyme comprising SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 43 (SbTyrAnc), or a TyrA enzyme having at least 95% identity to SEQ ID NO: 37 or 43;

wherein (i) the cell is engineered to express an enzyme that is not native to the cell, or (ii) the cell is engineered to overexpress an enzyme that is native to the cell as compared to a control cell.

11. The method of claim **10**, wherein the cell is engineered to express or overexpress both:

a) the DHS1b enzyme; and

b) a TyrA enzyme selected from the group consisting of SEQ ID NO: 33 (BdTyrA1), SEQ ID NO: 35 (BdTyrA2), SEQ ID NO: 37 (BdTyrAnc), SEQ ID NO: 39 (SbTyrA1), SEQ ID NO: 41 (SbTyrA2), SEQ ID NO: 43 (SbTyrAnc) and an enzyme having at least 95% sequence identity to one of SEQ ID NOs: 33, 35, 37, 39, 41 and 43.

12. The method of claim **10**, wherein the cell is engineered by introducing an exogenous nucleic acid encoding the enzyme into the cell.

13. The method of claim **12**, wherein the exogenous nucleic acid is introduced via insertion into the genome of the cell.

14. The method of claim **10**, wherein the cell is a plant cell.

15. The method of claim **14**, wherein the plant cell is part of a plant.

16. The method of claim **15**, wherein the plant is a tomato plant, tobacco plant, soybean plant, cotton plant, poplar plant, sorghum plant, corn plant, beet plant, mung bean plant, opium poppy plant, alfalfa plant, wheat plant, barley plant, millet plant, oat plant, rye plant, rapeseed plant, miscanthus plant, or grass plant.

17. The method of claim **16**, wherein the grass plant is a *Sorghum bicolor* plant, *Oryza sativa* plant, or *Brachypodium distachyon* plant.

18. The method of claim **10** further comprising purifying aromatic amino acids or derivatives thereof from the cell.

19. A method for producing aromatic amino acids or derivatives thereof, the method comprising:

a) growing the plant of claim **6**; and

b) purifying aromatic amino acids or derivatives thereof produced by the plant.

20. A method for sequestering carbon dioxide, the method comprising growing the plant of claim **6**.

* * * * *