US 20250376706A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2025/0376706 A1**

Donohue et al. (43) **Pub. Date: Dec. 11, 2025**

(54) **RECOMBINANT MICROORGANISMS THAT CATABOLIZE LIGNIN AROMATICS AND METHODS OF USING SAME**

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

(72) Inventors: **Timothy Donohue**, Middleton, WI (US); **Fletcher Metz**, Madison, WI (US); **Kevin Myers**, Madison, WI (US); **Fachuang Lu**, Madison, WI (US); **Daniel Noguera**, Madison, WI (US)

(73) Assignee: **Wisconsin Alumni Research Foundation**, Madison, WI (US)

(21) Appl. No.: **18/737,647**

(22) Filed: **Jun. 7, 2024**

Publication Classification

(51) **Int. Cl.**

C12P 17/04	(2006.01)
C12N 9/02	(2006.01)
C12N 9/88	(2006.01)

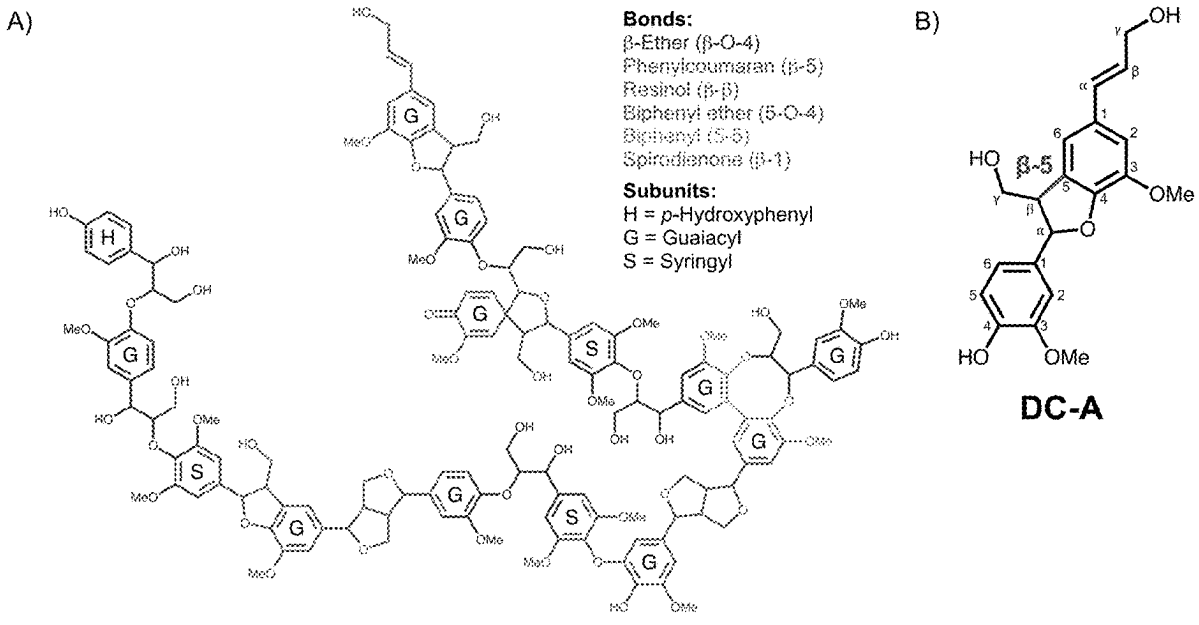
(52) **U.S. Cl.**

CPC ***C12P 17/04*** (2013.01); ***C12N 9/0004*** (2013.01); ***C12N 9/0069*** (2013.01); ***C12N 9/0093*** (2013.01); ***C12N 9/88*** (2013.01); ***C12Y 102/01071*** (2013.01); ***C12Y 113/11043*** (2013.01); ***C12Y 117/01*** (2013.01); ***C12Y 401/01028*** (2013.01); ***C12Y 402/01*** (2013.01)

(57) **ABSTRACT**

Recombinant microorganisms that catabolize lignin aromatics, such as β-5 linked lignin aromatics, and methods of using same to catabolize the lignin aromatics.

Specification includes a Sequence Listing.



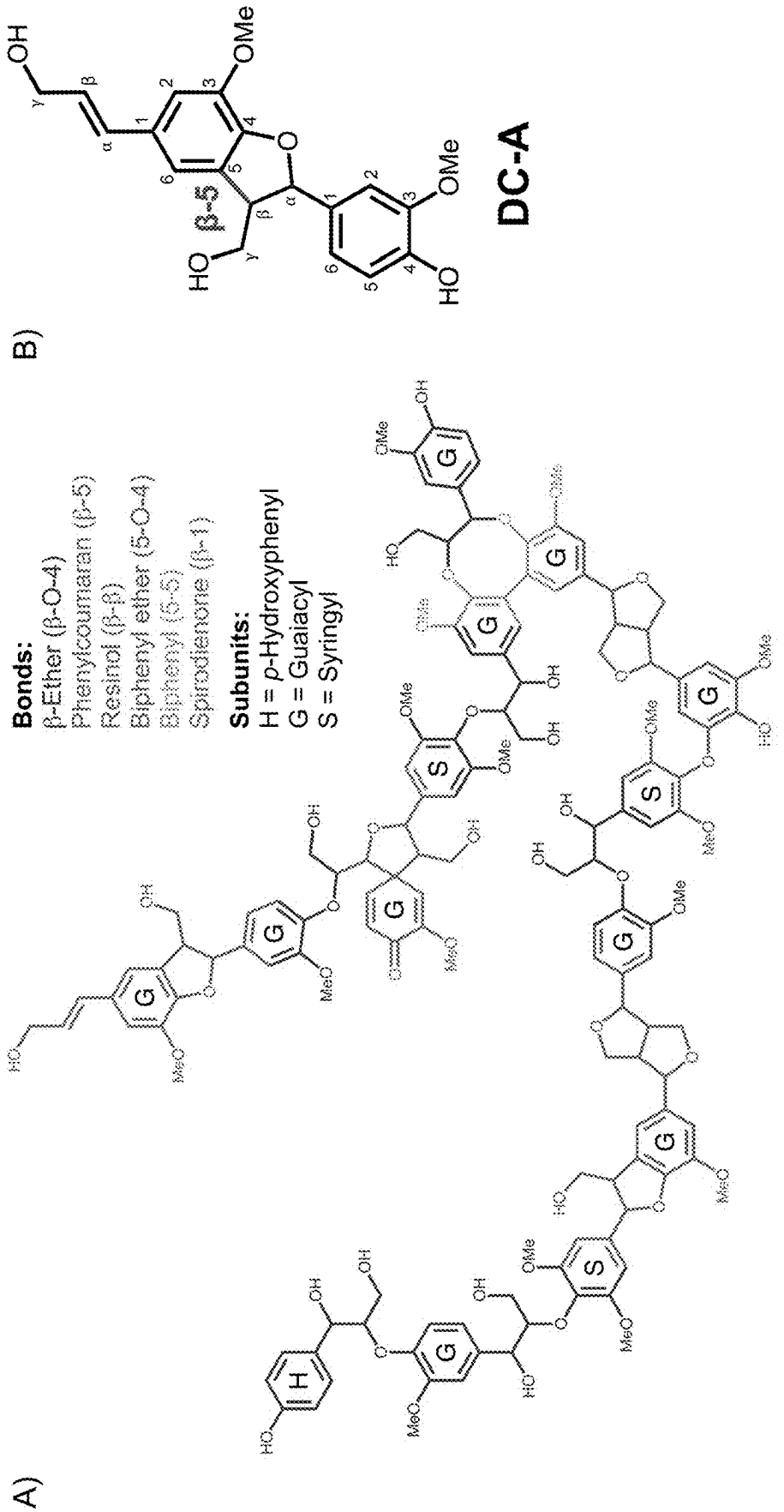

A)
Bonds:
β-Ether (β-O-4)
Phenylcoumaran (β-5)
Resinol (β-β)
Biphenyl ether (5-O-4)
Biphenyl (5-5)
Spirodienone (β-1)
Subunits:
H = p-Hydroxyphenyl
G = Guaiacyl
S = Syringyl
B)
β-5
DC-A

FIG. 1

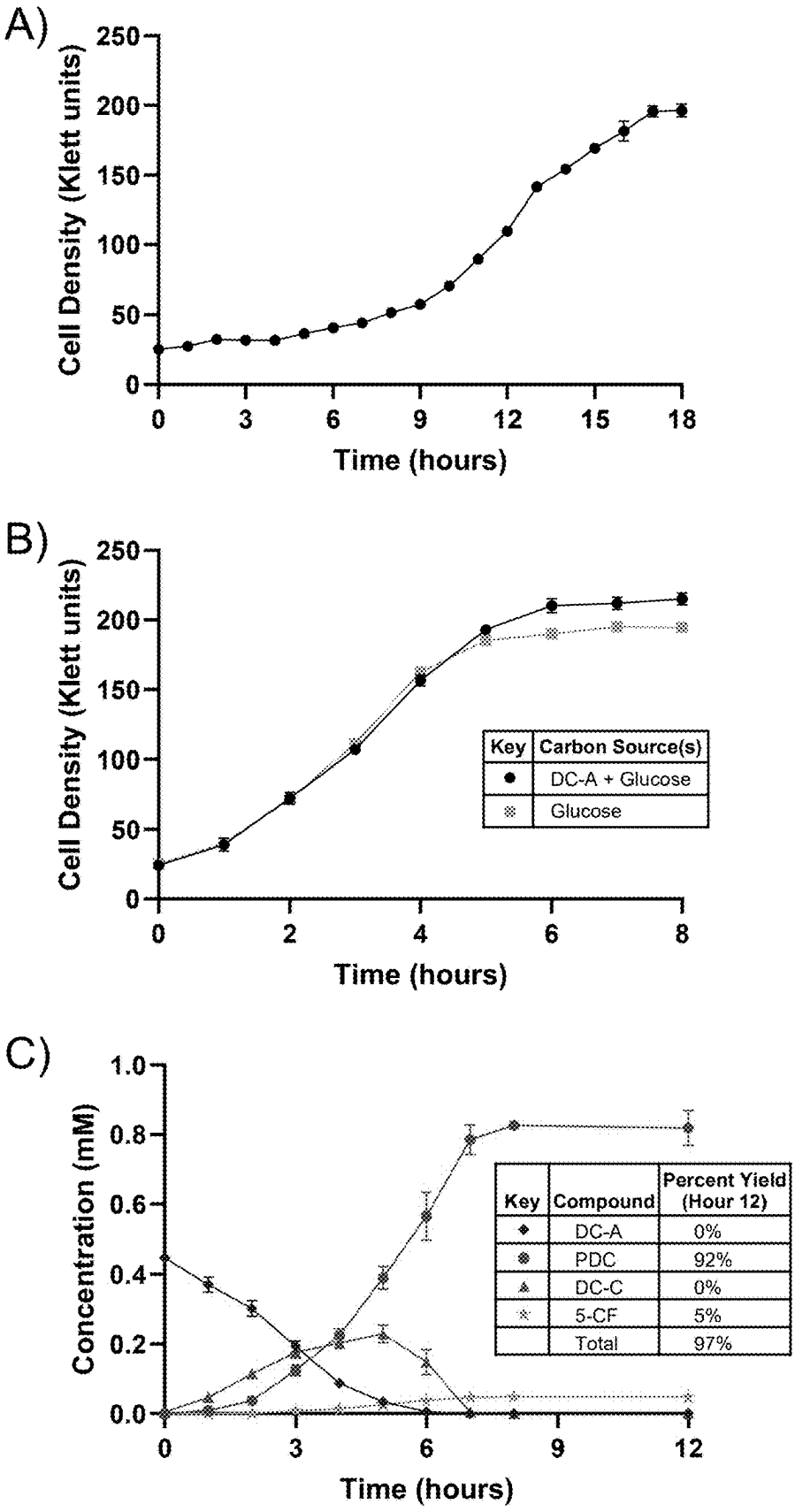

A)
Cell Density (Klett units)
250
200
150
100
50
0
0
3
6
9
12
15
18
Time (hours)
B)
Cell Density (Klett units)
250
200
150
100
50
0
0
2
4
6
8
Time (hours)
Key
Carbon Source(s)
DC-A + Glucose
Glucose
C)
Concentration (mM)
1.0
0.8
0.6
0.4
0.2
0.0
0
3
6
9
12
Time (hours)
Key
Compound
Percent Yield (Hour 12)
DC-A
0%
PDC
92%
DC-C
0%
5-CF
5%
Total
97%

FIG. 2

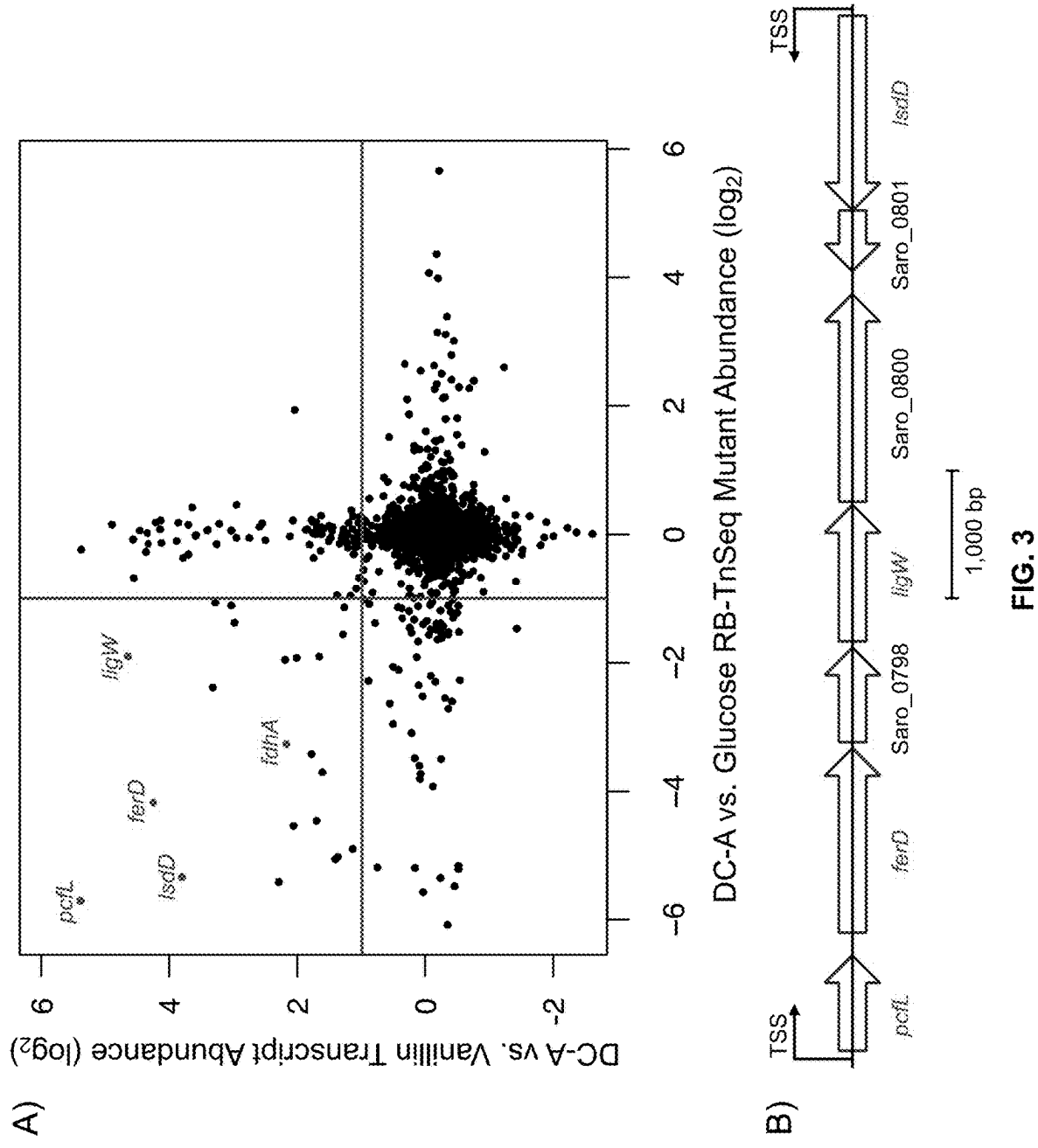

A)
DC-A vs. Vanillin Transcript Abundance (log_2)
DC-A vs. Glucose RB-TnSeq Mutant Abundance (log_2)
pcfL
lsdD
ferD
fdhA
ligW
B)
TSS
pcfL
ferD
Saro_0798
ligW
Saro_0800
Saro_0801
lsdD
1,000 bp

FIG. 3

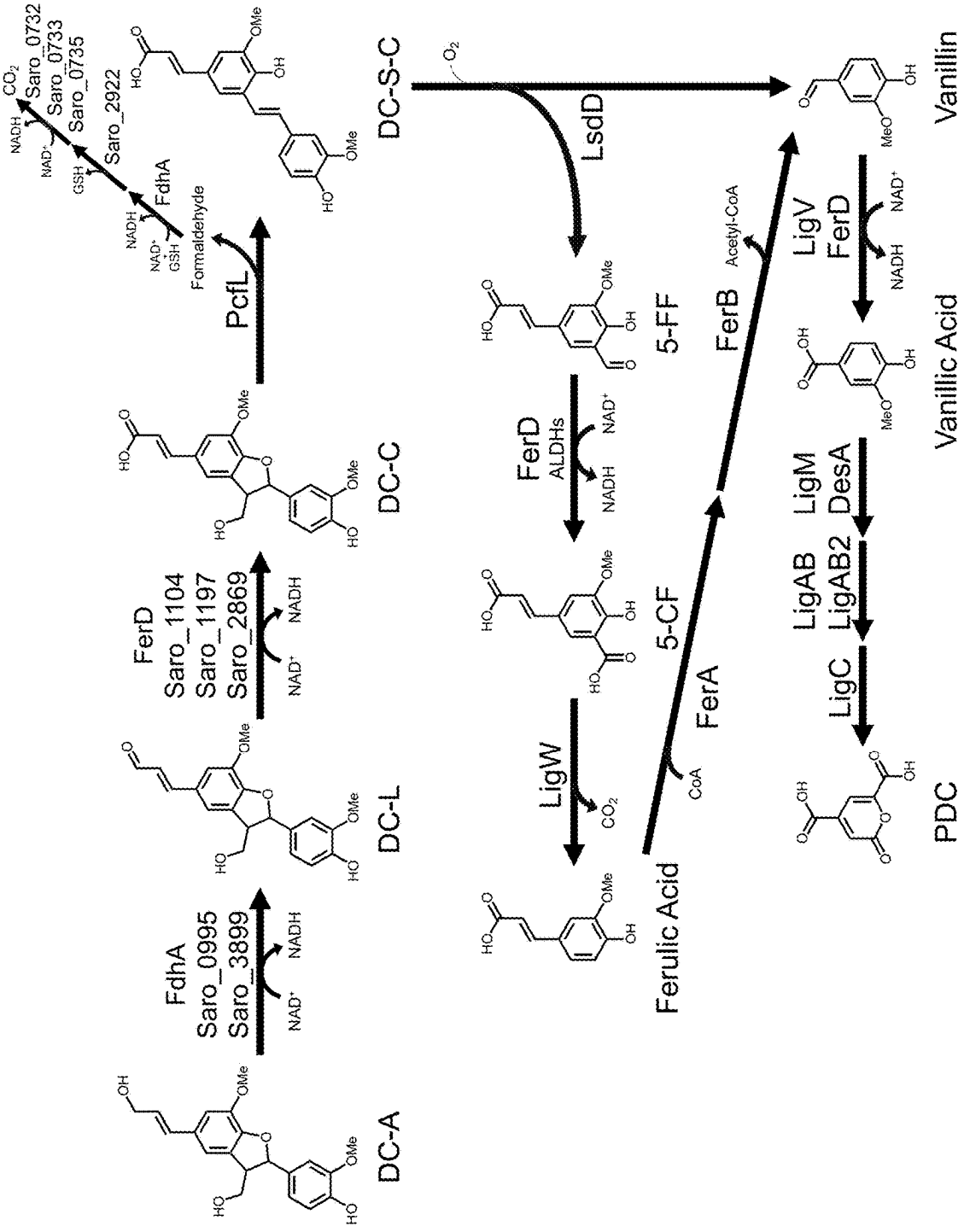

DC-A
FdhA
Saro_0995
Saro_3899
NAD+
NADH
DC-L
FerD
Saro_1104
Saro_1197
Saro_2869
NAD+
NADH
DC-C
PcfL
Formaldehyde
FdhA
NADH
NAD+
GSH
Saro_2922
GSH
NAD+
NADH
Saro_0735
Saro_0733
Saro_0732
CO2
DC-S-C
O2
LsdD
Ferulic Acid
LigW
CO2
5-CF
FerA
CoA
FerD
ALDHs
NAD+
NADH
5-FF
FerB
Acetyl-CoA
LigV
FerD
NADH
NAD+
Vanillin
LigAB
LigAB2
LigC
LigM
DesA
Vanillic Acid
PDC

FIG. 4

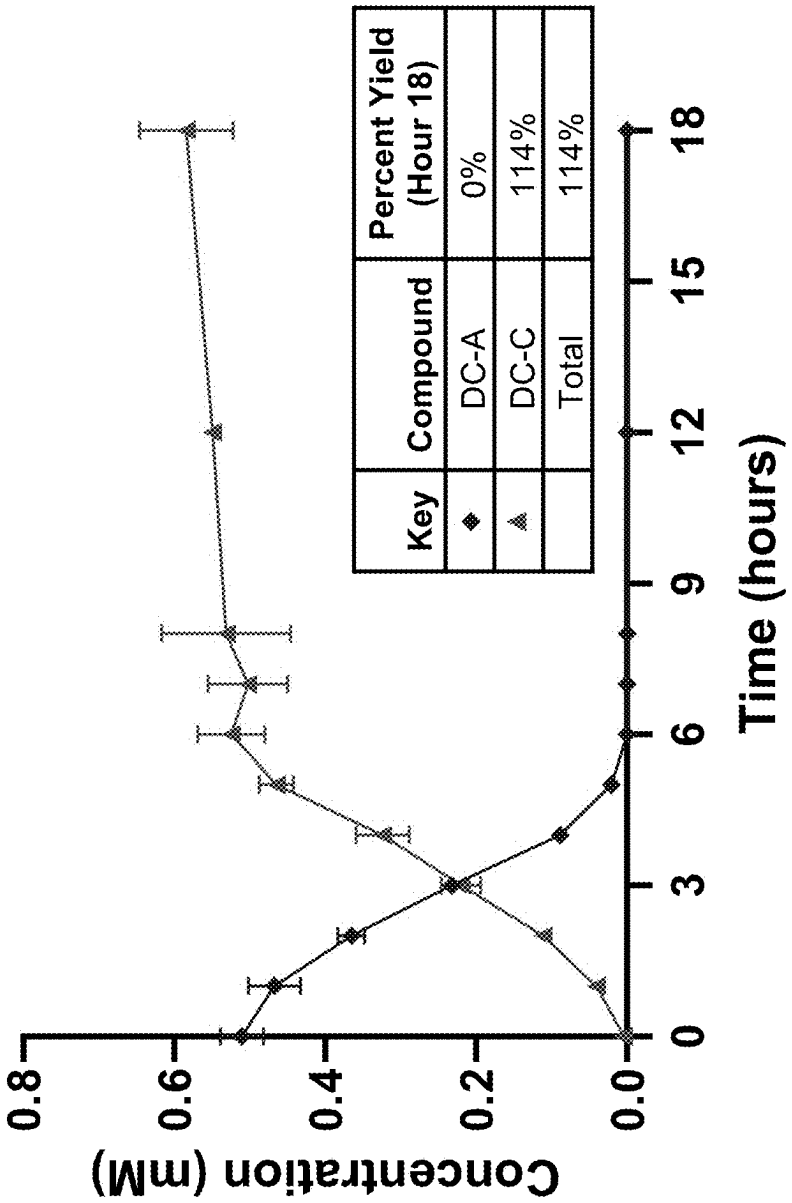

Concentration (mM)
0.0
0.2
0.4
0.6
0.8
0
3
6
9
12
15
18
Time (hours)
Key
Compound
Percent Yield (Hour 18)
DC-A
0%
DC-C
114%
Total
114%

FIG. 5A

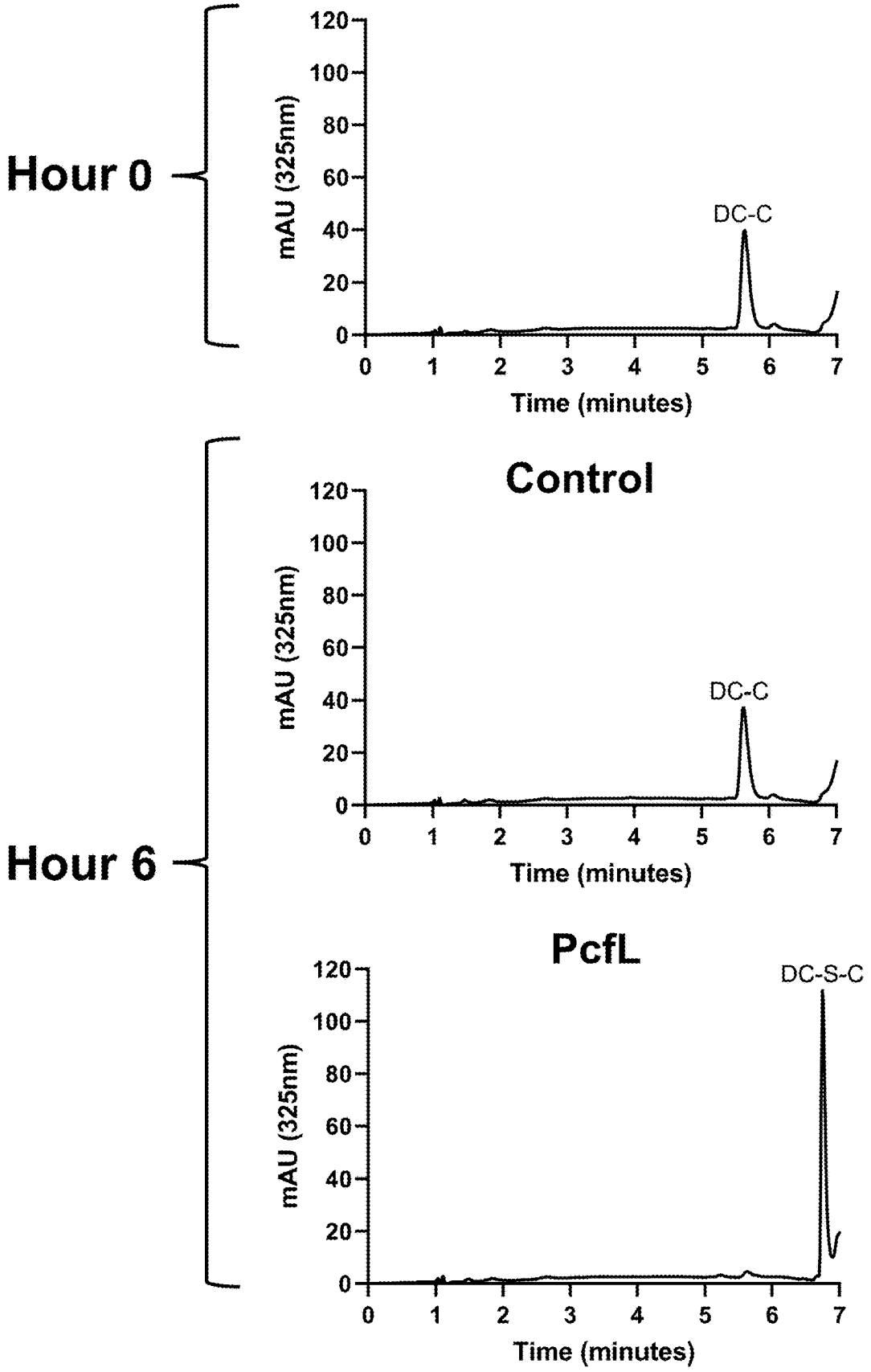

Hour 0
mAU (325nm)
120
100
80
60
40
20
0
DC-C
0 1 2 3 4 5 6 7
Time (minutes)
Hour 6
Control
mAU (325nm)
DC-C
Time (minutes)
PcfL
mAU (325nm)
DC-S-C
Time (minutes)

FIG. 5B

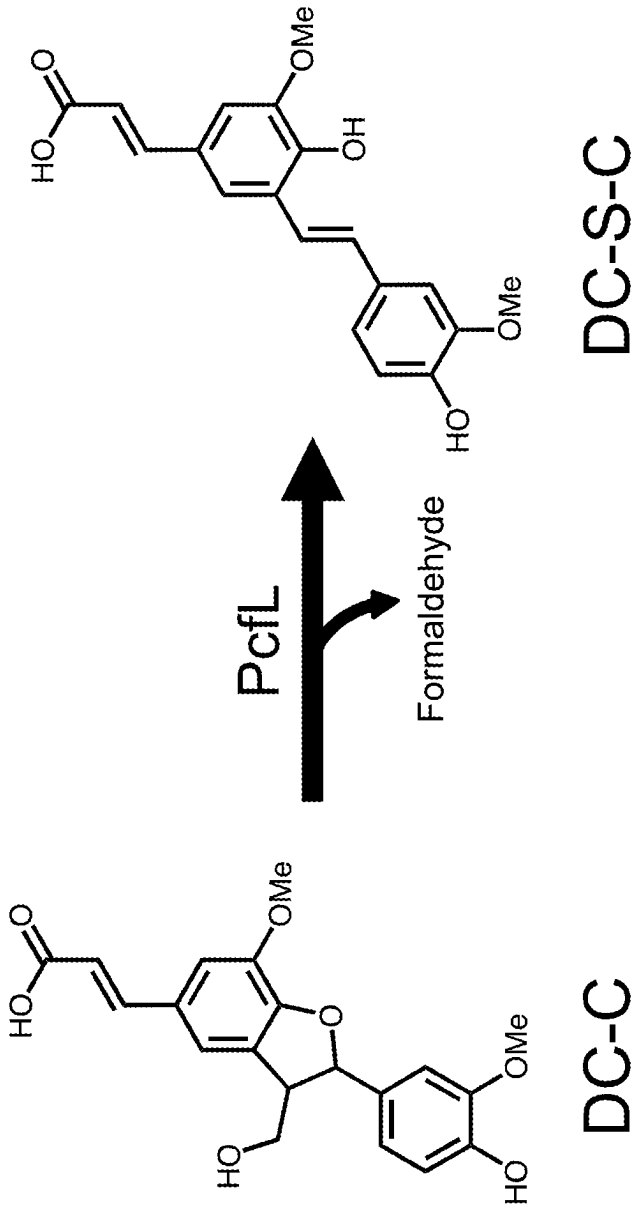


FIG. 5C

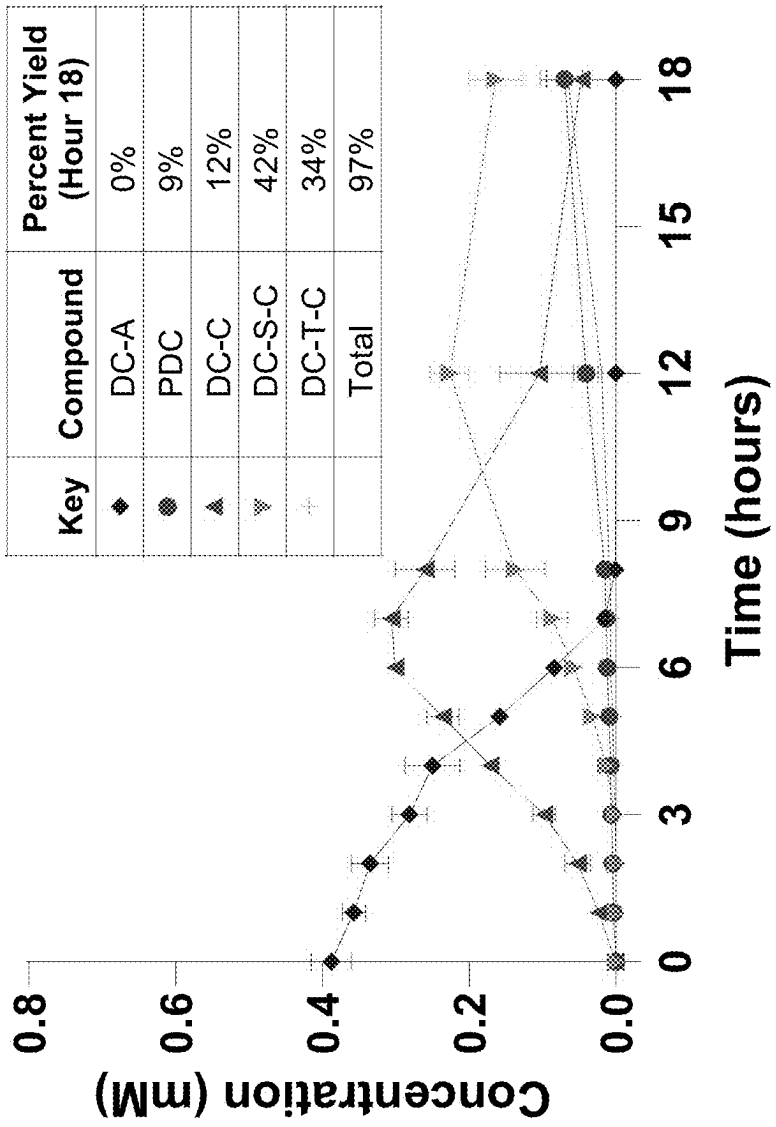

Key
Compound
Percent Yield (Hour 18)
DC-A
0%
PDC
9%
DC-C
12%
DC-S-C
42%
DC-T-C
34%
Total
97%
Concentration (mM)
0.0
0.2
0.4
0.6
0.8
0
3
6
9
12
15
18
Time (hours)

FIG. 6A

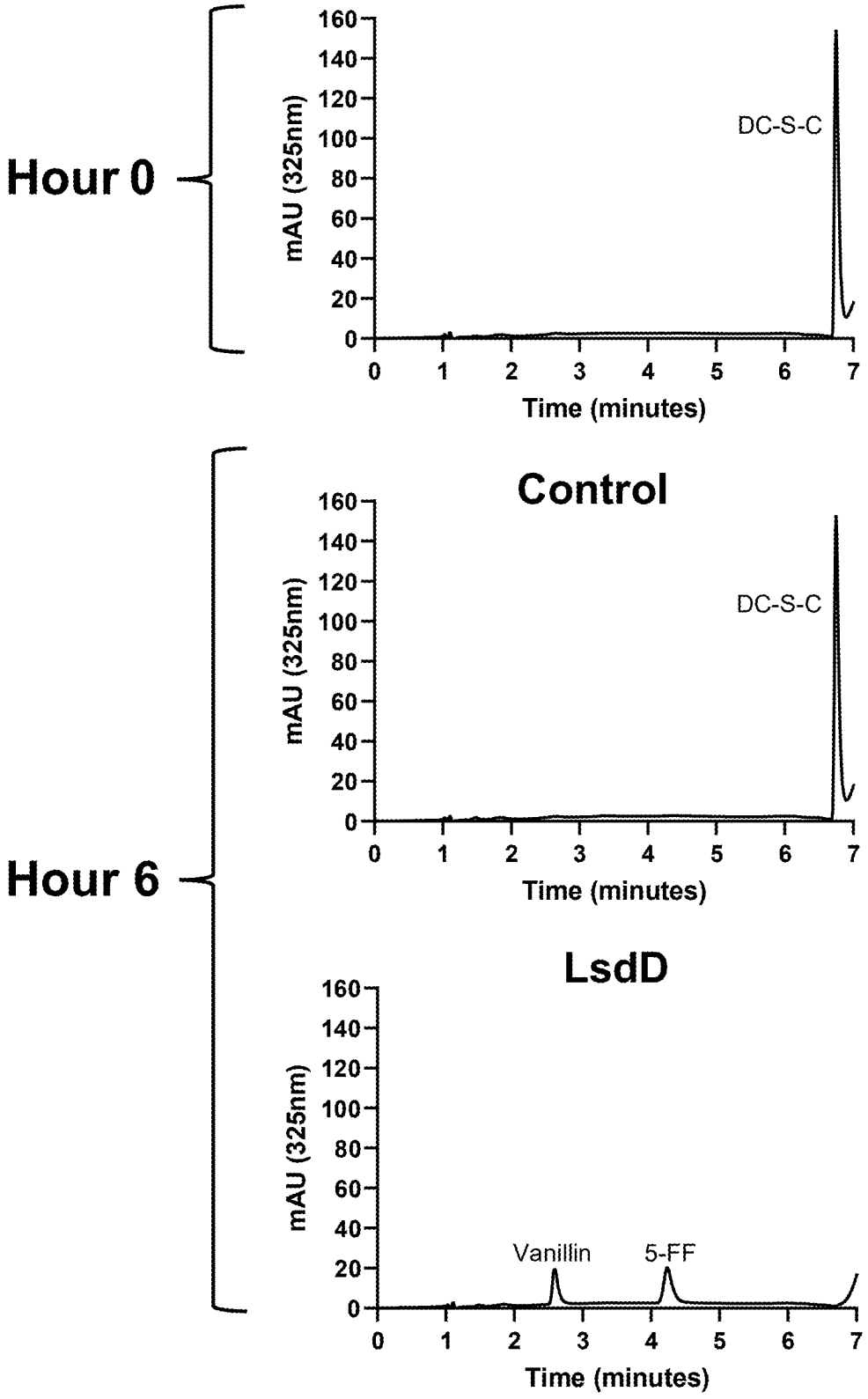

Hour 0
mAU (325nm)
DC-S-C
Time (minutes)
Hour 6
Control
mAU (325nm)
DC-S-C
Time (minutes)
LsdD
mAU (325nm)
Vanillin
5-FF
Time (minutes)

FIG. 6B

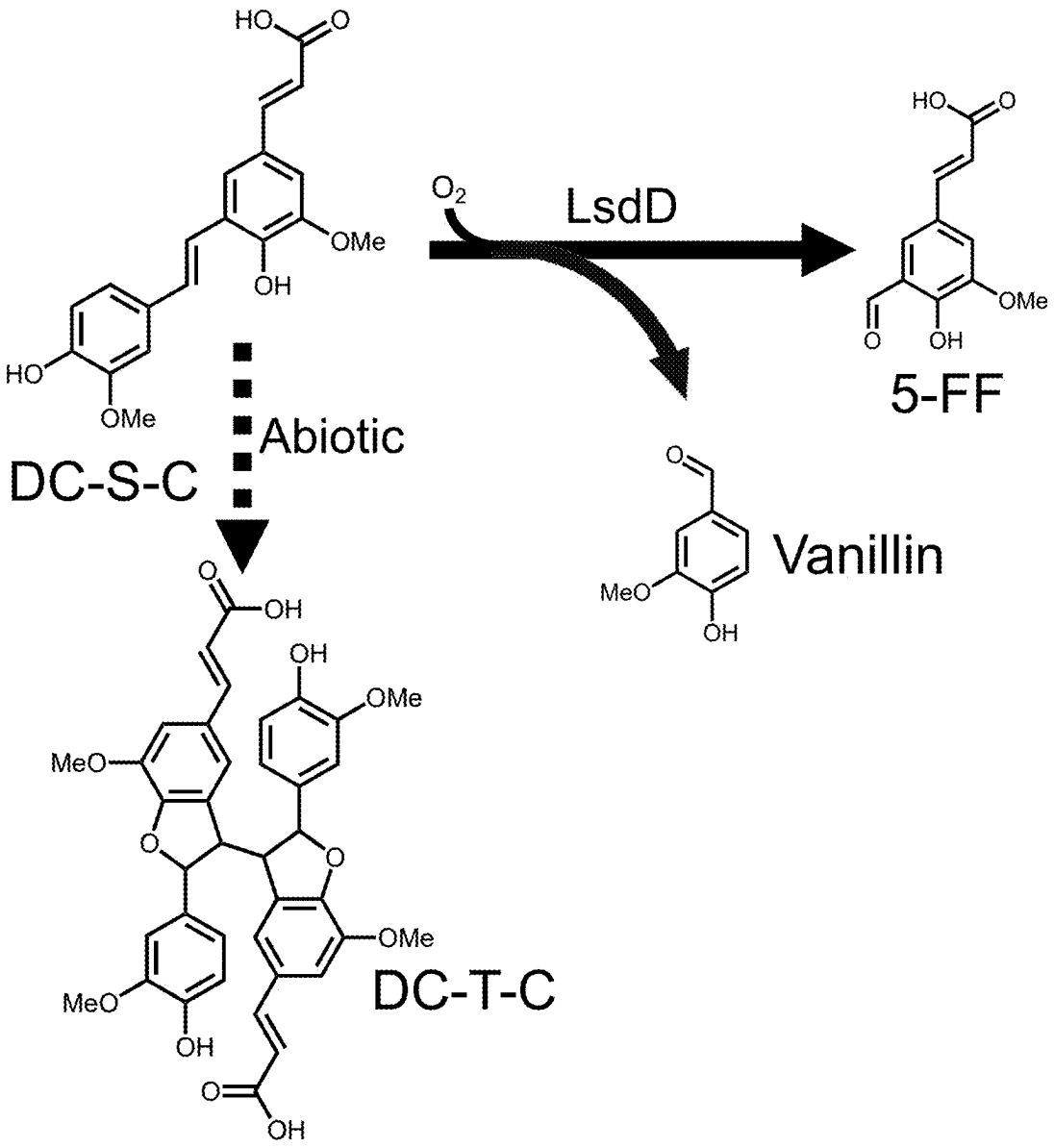

HO
O
OMe
OH
HO
OMe
DC-S-C
Abiotic
O2
LsdD
HO
O
OMe
O
OH
5-FF
O
MeO
OH
Vanillin
OH
OH
OMe
MeO
O
O
OMe
MeO
OH
O
OH
DC-T-C

FIG. 6C

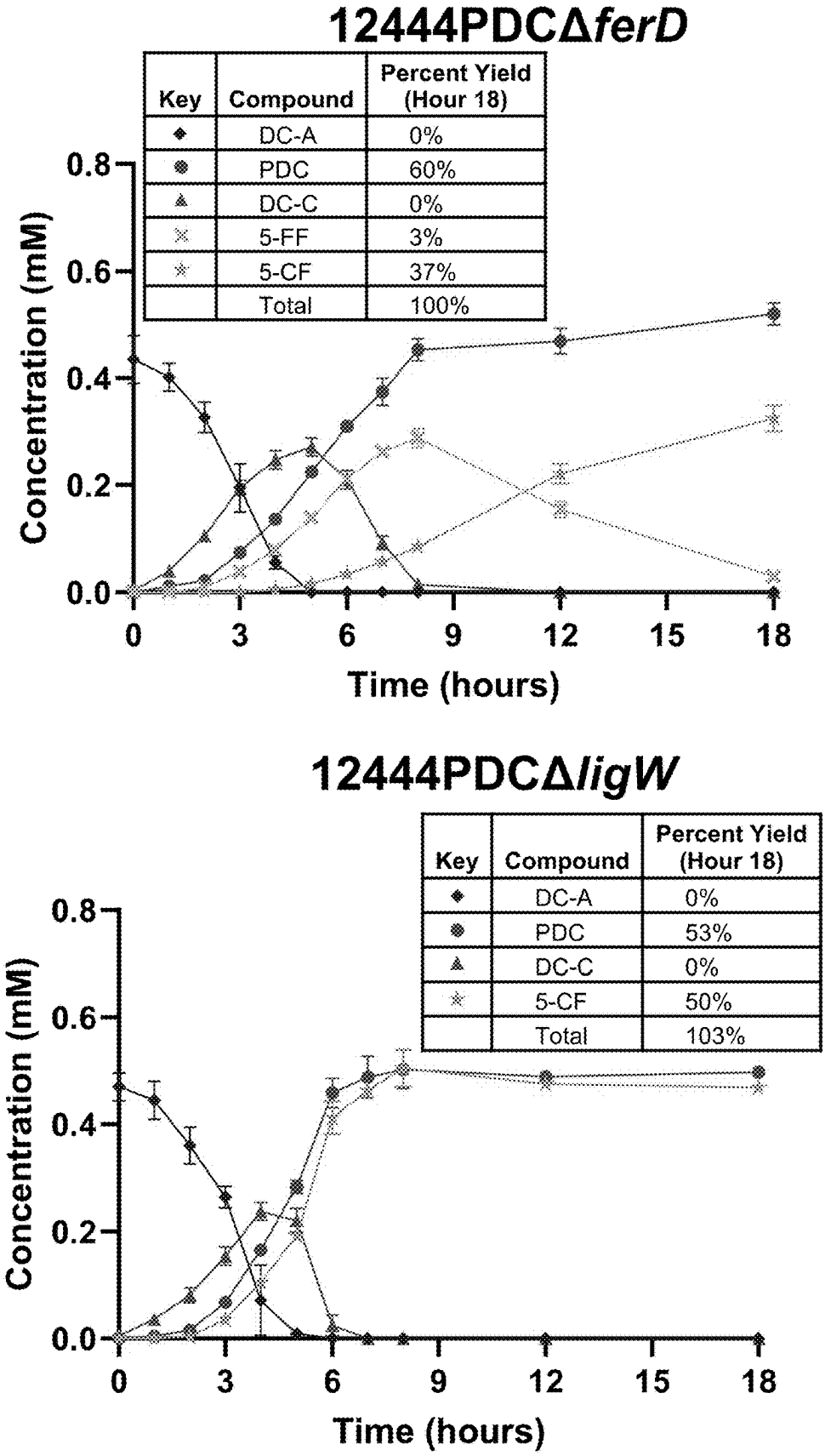

12444PDCΔferD
Key
Compound
Percent Yield (Hour 18)
DC-A
0%
PDC
60%
DC-C
0%
5-FF
3%
5-CF
37%
Total
100%
Concentration (mM)
0.0
0.2
0.4
0.6
0.8
0
3
6
9
12
15
18
Time (hours)
12444PDCΔligW
Key
Compound
Percent Yield (Hour 18)
DC-A
0%
PDC
53%
DC-C
0%
5-CF
50%
Total
103%
Concentration (mM)
0.0
0.2
0.4
0.6
0.8
0
3
6
9
12
15
18
Time (hours)

FIG. 7A

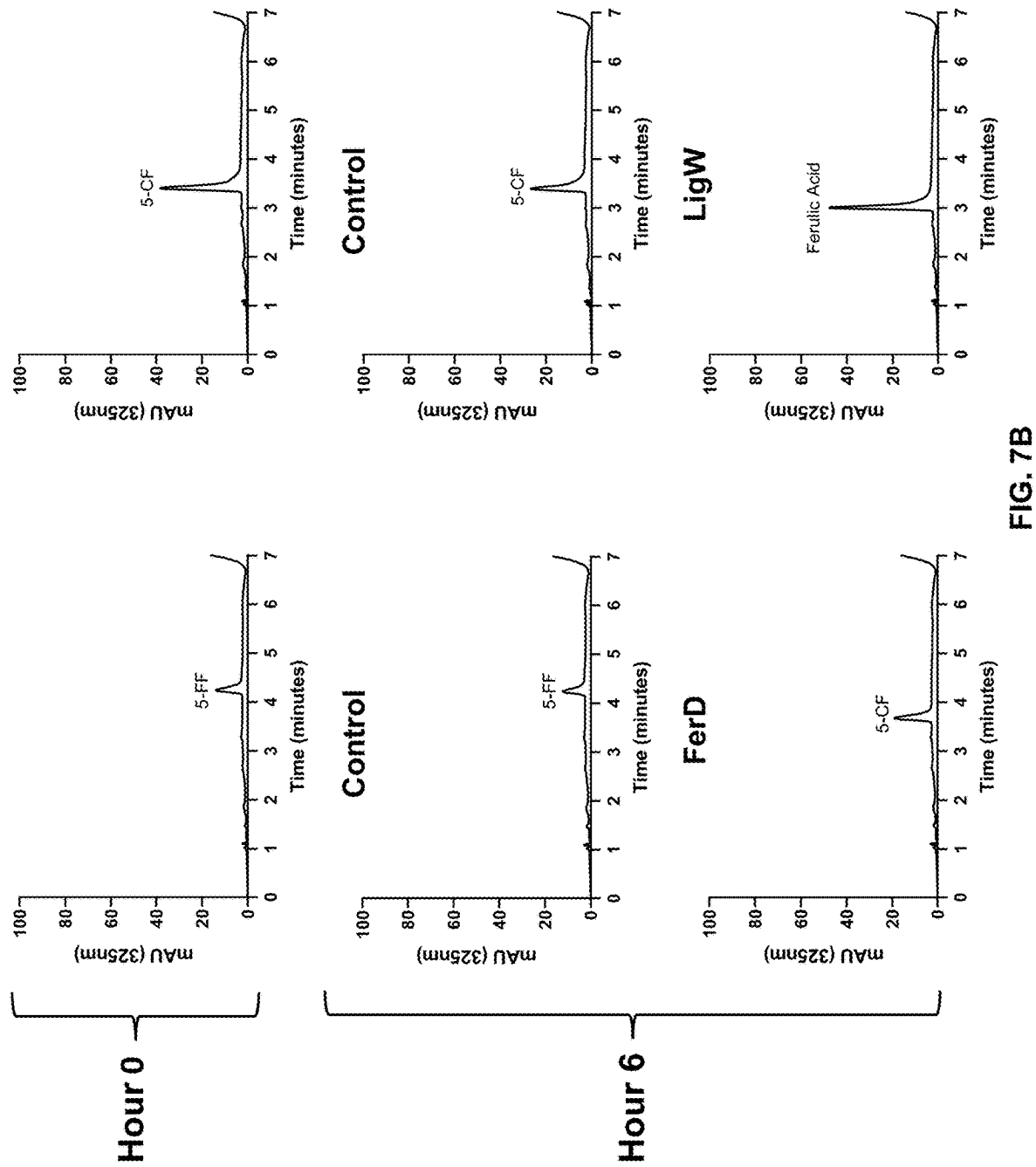

Hour 0
Hour 6
Control
FerD
Control
LigW
5-FF
5-CF
Ferulic Acid
mAU (325nm)
Time (minutes)

FIG. 7B

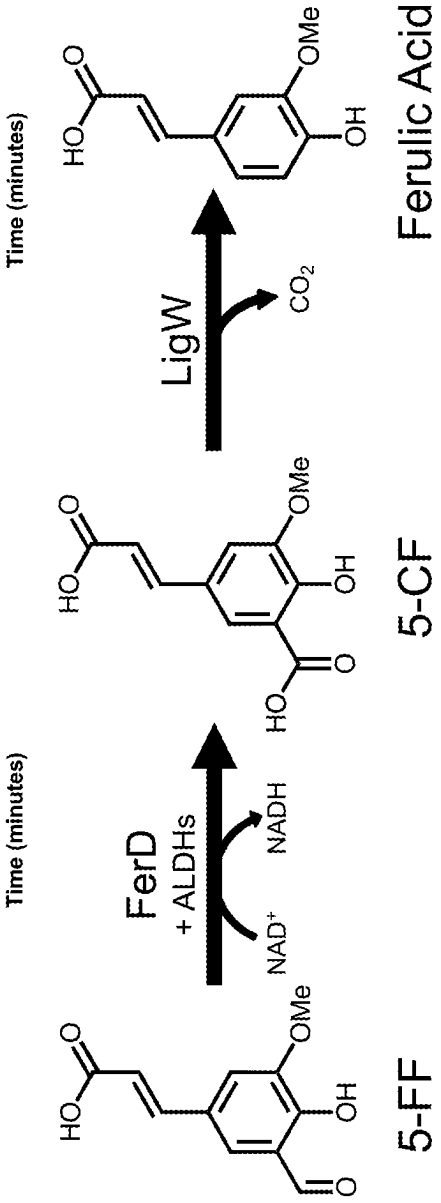

Time (minutes)
FerD
+ ALDHs
NAD+
NADH
5-FF
Time (minutes)
LigW
CO2
5-CF
Ferulic Acid
HO
O
OMe
OH

FIG. 7C

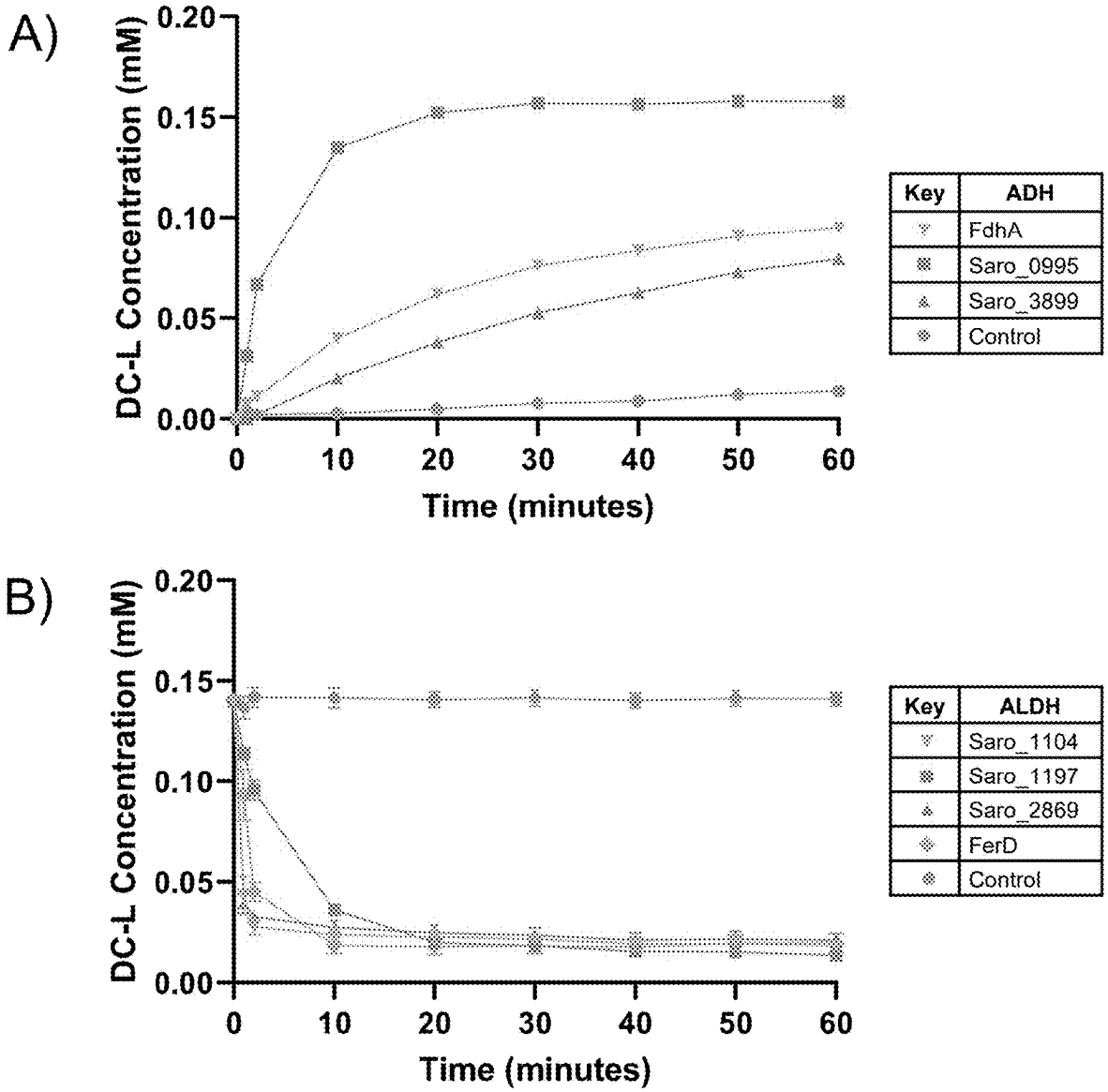

A)
DC-L Concentration (mM)
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
Time (minutes)
Key
ADH
FdhA
Saro_0995
Saro_3899
Control
B)
DC-L Concentration (mM)
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
Time (minutes)
Key
ALDH
Saro_1104
Saro_1197
Saro_2869
FerD
Control

FIG. 8

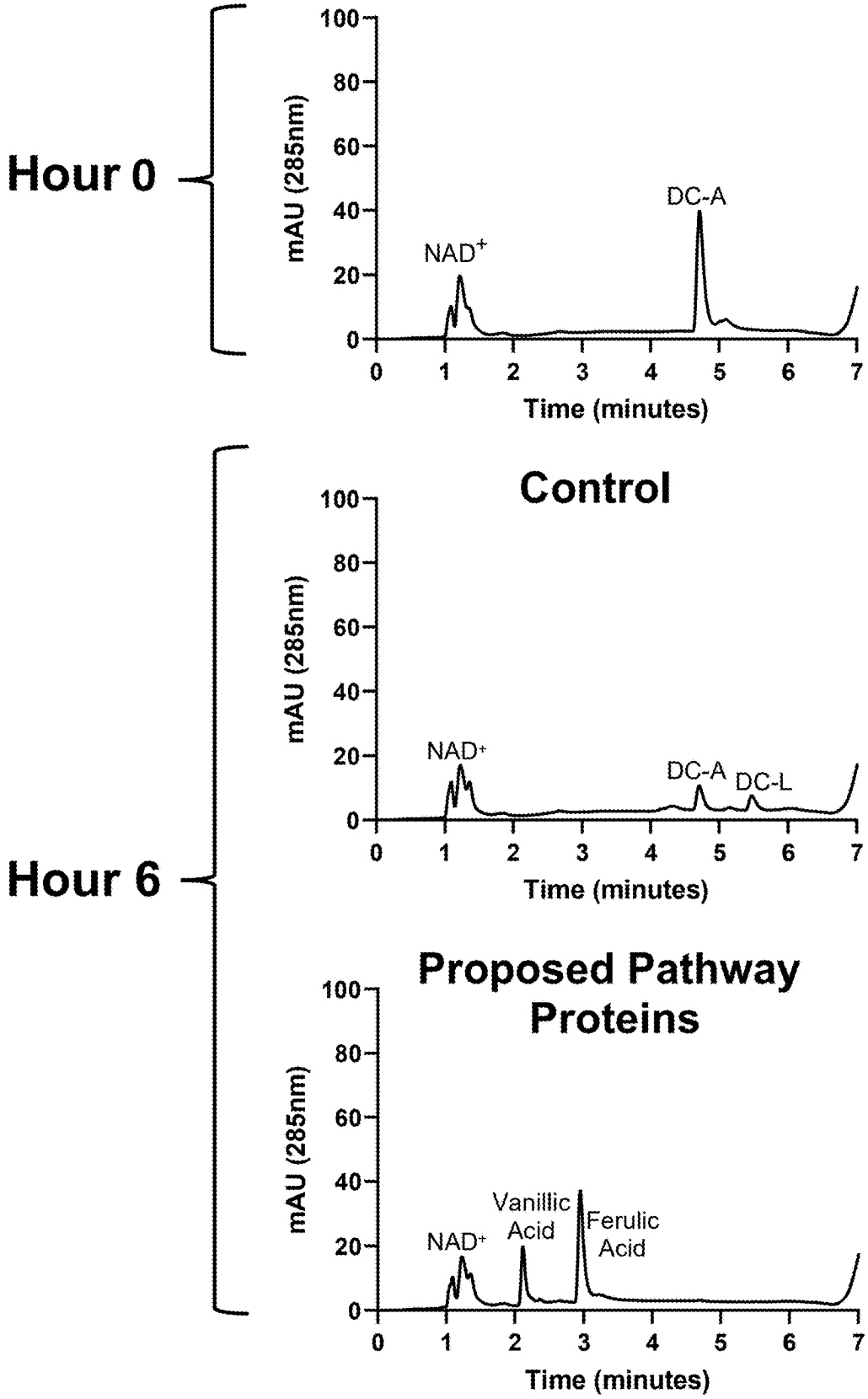

Hour 0
Hour 6
Control
Proposed Pathway Proteins
mAU (285nm)
Time (minutes)
NAD+
DC-A
DC-L
Vanillic Acid
Ferulic Acid
0
20
40
60
80
100
0
1
2
3
4
5
6
7

FIG. 9

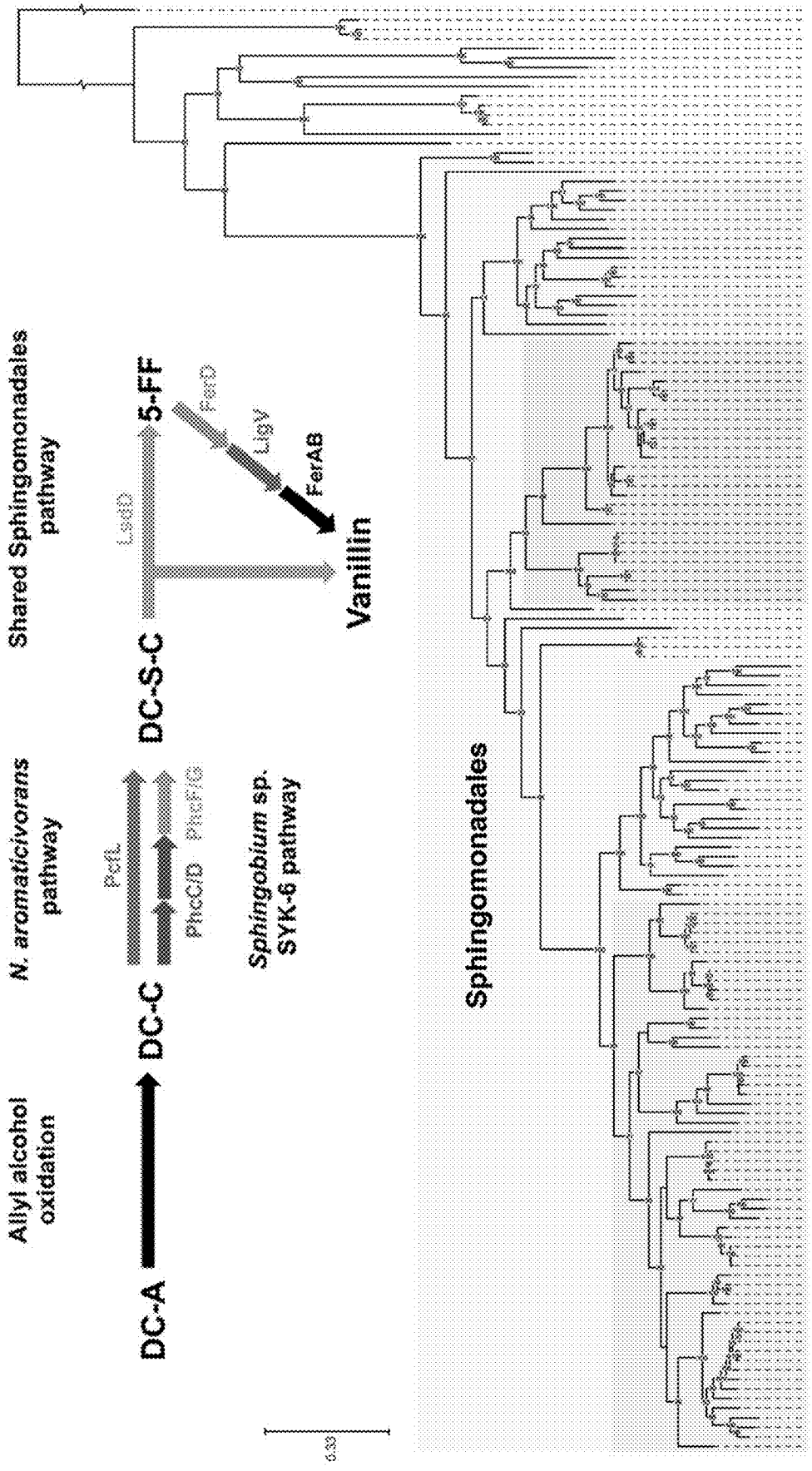

Allyl alcohol oxidation
DC-A
DC-C
N. aromaticivorans pathway
PcfL
PhcC/D
PhcF/G
Sphingobium sp. SYK-6 pathway
DC-S-C
LsdD
5-FF
Shared Sphingomonadales pathway
FerD
LigV
FerAB
Vanillin
Sphingomonadales
0.33

FIG. 10A

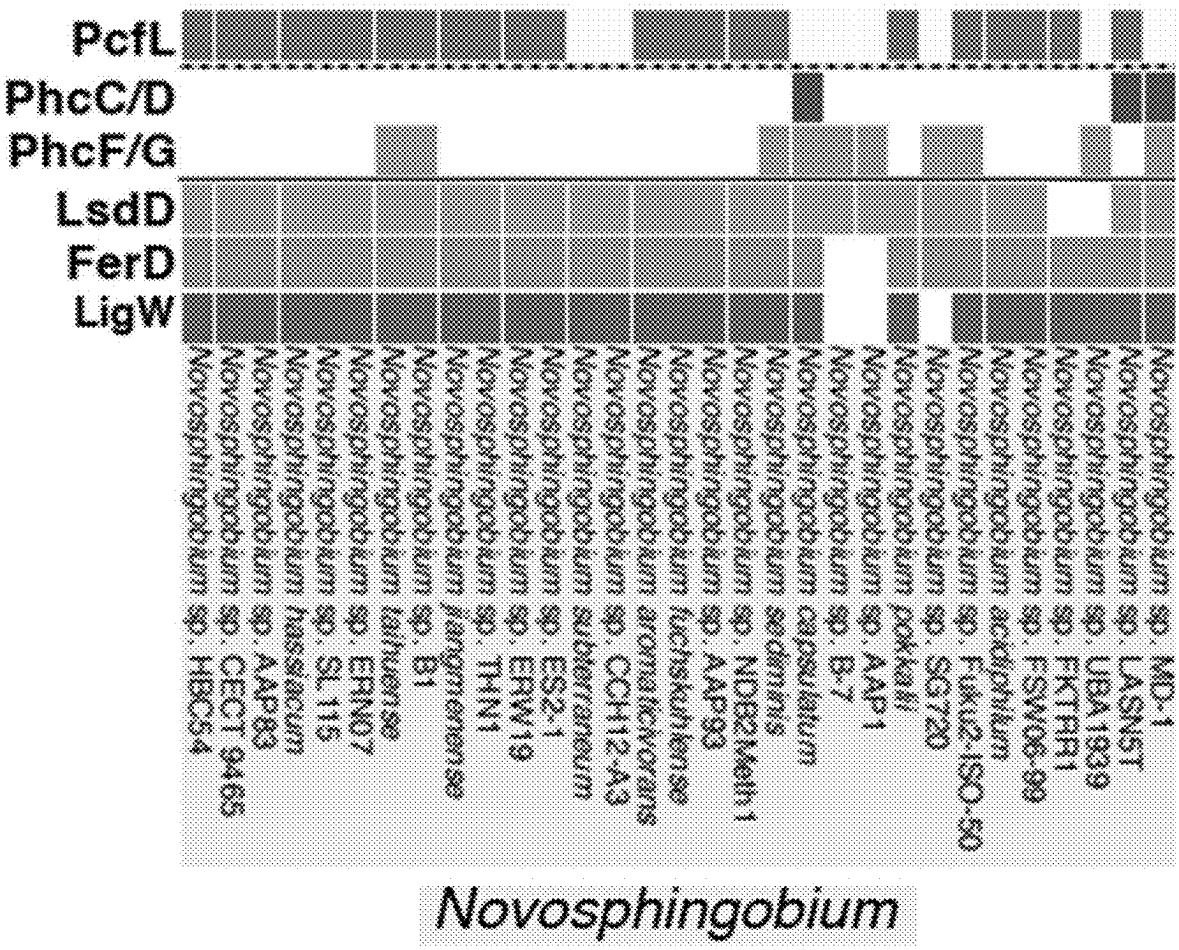

PcfL
PhcC/D
PhcF/G
LsdD
FerD
LigW
Novosphingobium sp. HBC54
Novosphingobium sp. CECT 9465
Novosphingobium sp. AAP83
Novosphingobium hassiacum
Novosphingobium sp. SL 115
Novosphingobium sp. ERN07
Novosphingobium lahuense
Novosphingobium sp. B1
Novosphingobium pentomonense
Novosphingobium sp. THN1
Novosphingobium sp. ES2-1
Novosphingobium sp. ERW19
Novosphingobium subterraneum
Novosphingobium sp. CCH12-A3
Novosphingobium aromaticivorans
Novosphingobium fuchskuhlense
Novosphingobium sp. AAP93
Novosphingobium sp. NDB2Meth1
Novosphingobium sediminis
Novosphingobium capsulatum
Novosphingobium sp. B-7
Novosphingobium sp. AAP1
Novosphingobium pokkalii
Novosphingobium sp. SG720
Novosphingobium sp. Fuku2-ISO-50
Novosphingobium acidiphilum
Novosphingobium sp. FSW06-99
Novosphingobium sp. FKTRR1
Novosphingobium sp. UBA1939
Novosphingobium sp. LASNST
Novosphingobium sp. MD-1
Novosphingobium

FIG. 10B

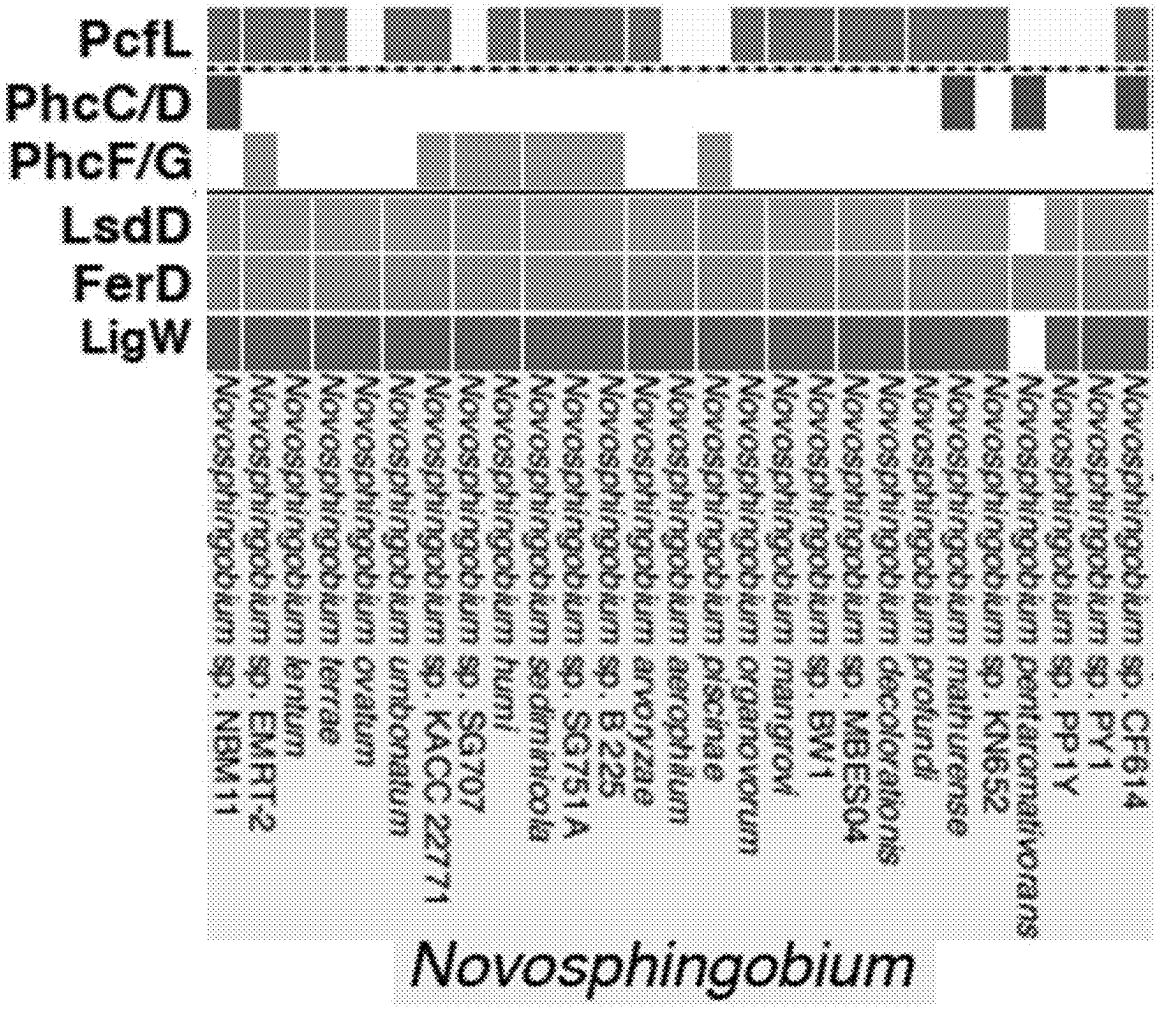

PcfL
PhcC/D
PhcF/G
LsdD
FerD
LigW
Novosphingobium sp. NBM11
Novosphingobium sp. EMRT-2
Novosphingobium kunmingense
Novosphingobium terrae
Novosphingobium ovatum
Novosphingobium umbonatum
Novosphingobium sp. SG707
Novosphingobium humi
Novosphingobium sediminicola
Novosphingobium sp. SG751A
Novosphingobium sp. B225
Novosphingobium aromaticivorans
Novosphingobium arvum
Novosphingobium pisci
Novosphingobium organovorum
Novosphingobium mangrovi
Novosphingobium sp. BW1
Novosphingobium sp. MBES04
Novosphingobium decolorationis
Novosphingobium profundi
Novosphingobium mathurense
Novosphingobium sp. KN65.2
Novosphingobium pentaromativorans
Novosphingobium sp. PP1Y
Novosphingobium sp. PY1
Novosphingobium sp. CF614
Novosphingobium

FIG. 10C

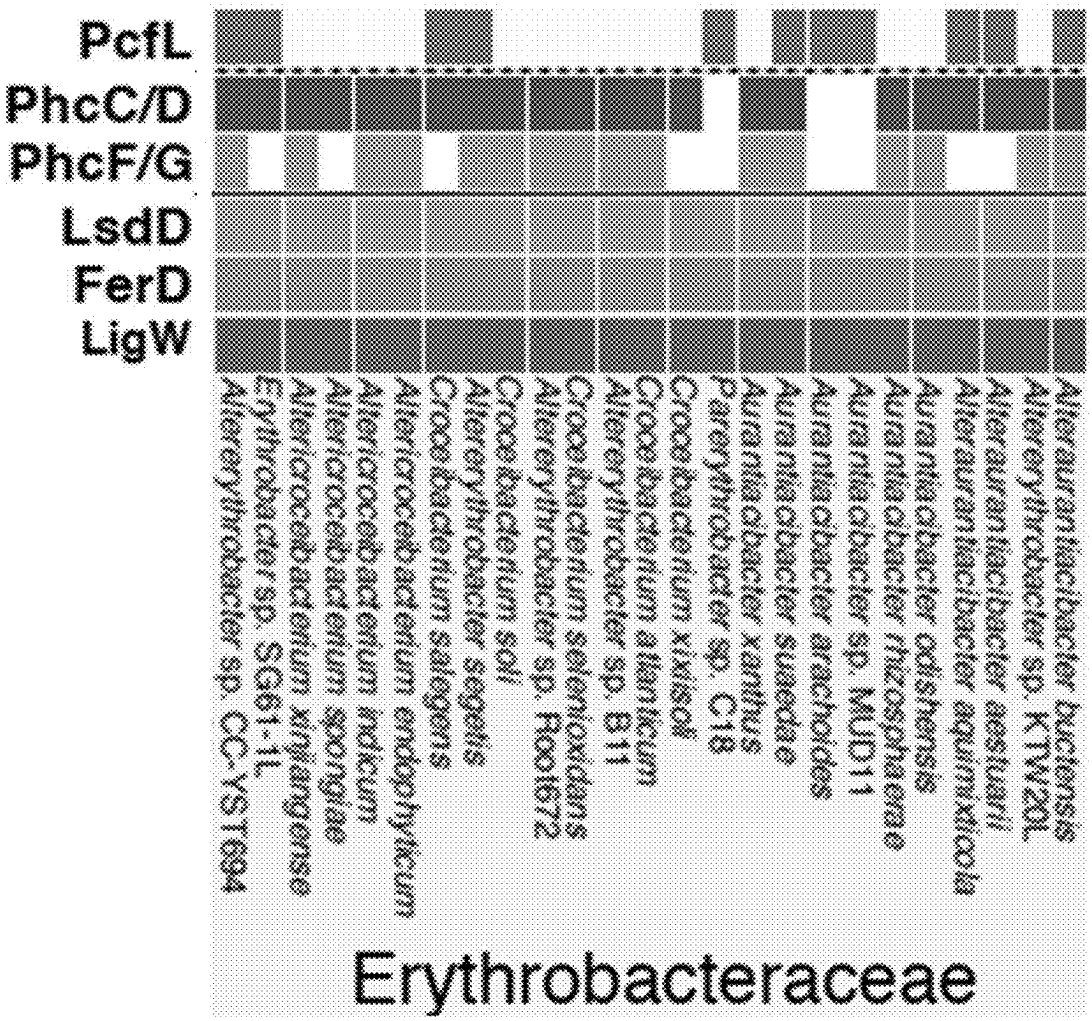

PcfL
PhcC/D
PhcF/G
LsdD
FerD
LigW
Alterythrobacter sp. CC-YST694
Erythrobacter sp. SG61-1L
Altericroceibacterium xinjiangense
Altericroceibacterium spongiae
Altericroceibacterium indicum
Altericroceibacterium endophyticum
Croceibacterium salegens
Croceibacterium soli
Alterythrobacter segetis
Croceibacterium selenioxidans
Alterythrobacter sp. Root672
Croceibacterium atlanticum
Alterythrobacter sp. B11
Croceibacterium xixiisoli
Parerythrobacter sp. C18
Aurantiacibacter xanthus
Aurantiacibacter suaedae
Aurantiacibacter arachoides
Aurantiacibacter sp. MUD11
Aurantiacibacter rhizosphaerae
Aurantiacibacter odishensis
Aurantiacibacter aquimaritimus
Alteraurantiacibacter aestuarii
Alterythrobacter sp. KTW20L
Alteraurantiacibacter luctensis
Erythrobacteraceae

FIG. 10D

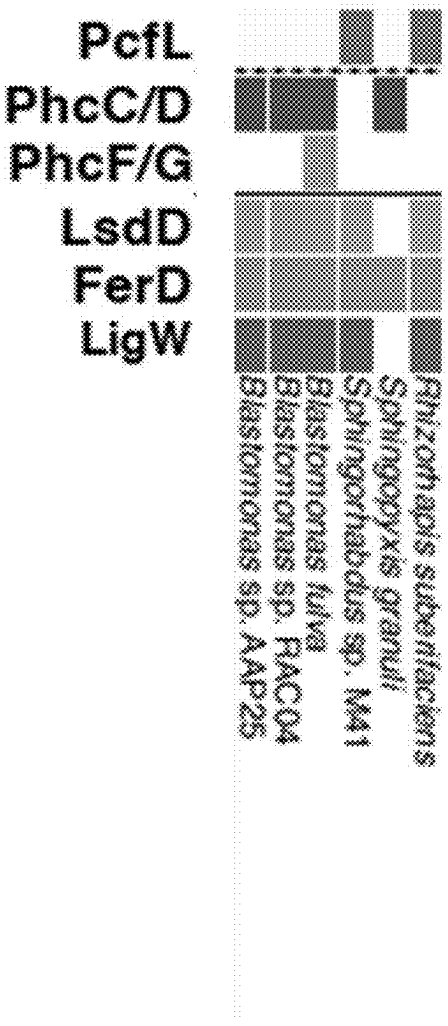

PcfL
PhcC/D
PhcF/G
LsdD
FerD
LigW
Rhizorhapis suberifaciens
Sphingopyxis granuli
Sphingorhabdus sp. M41
Blastomonas fulva
Blastomonas sp. RAC04
Blastomonas sp. AAP25

FIG. 10E

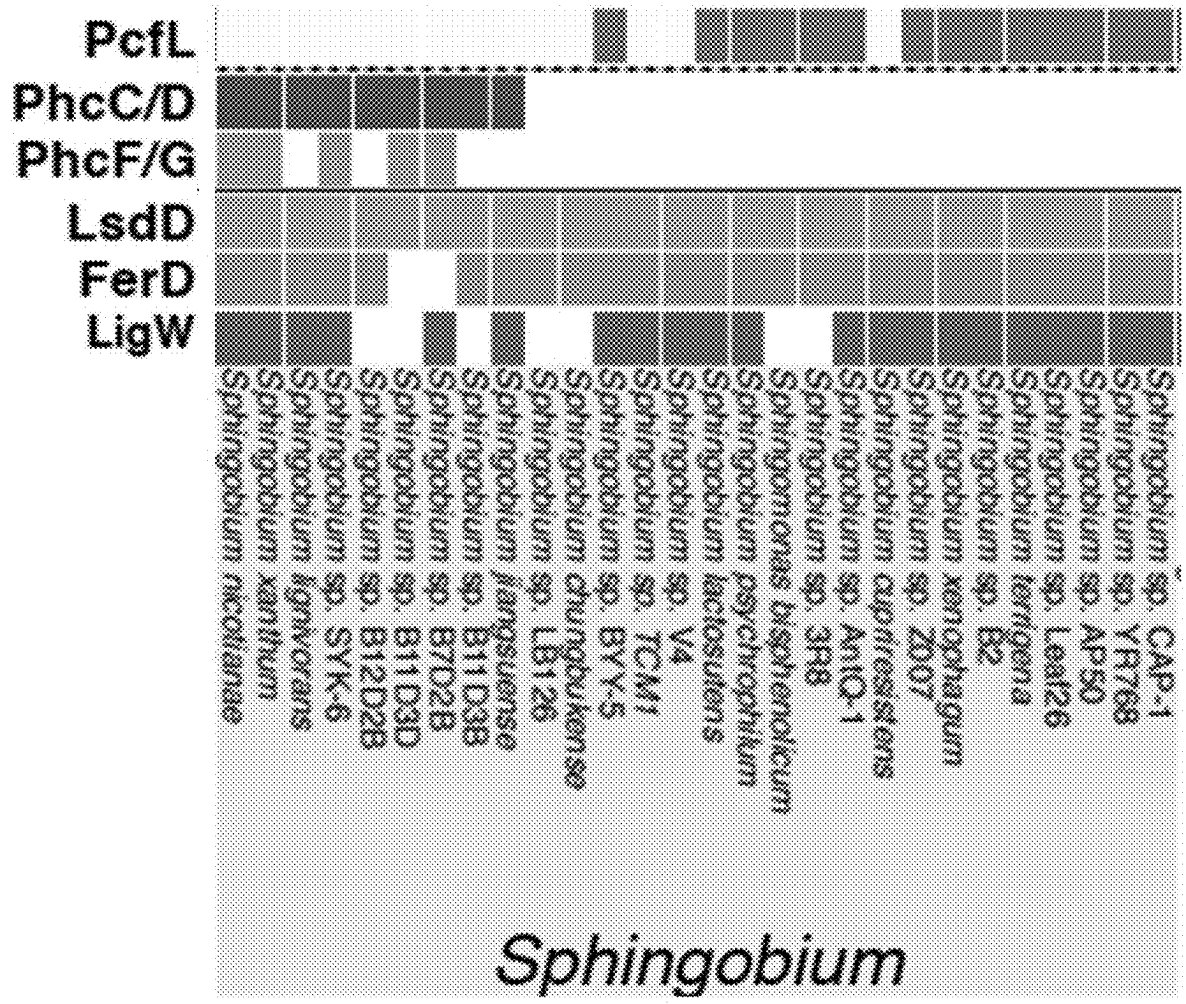

PcfL
PhcC/D
PhcF/G
LsdD
FerD
LigW
Sphingobium sp. CAP-1
Sphingobium sp. YR768
Sphingobium sp. AP50
Sphingobium sp. Leaf26
Sphingobium terrae
Sphingobium sp. B2
Sphingobium xenophagum
Sphingobium sp. 207
Sphingobium cupriresistens
Sphingobium sp. AntiO-1
Sphingobium sp. 3P8
Sphingomonas bisphenolicum
Sphingobium psychrophilum
Sphingobium sp. LA
Sphingobium sp. V4
Sphingobium sp. TCM1
Sphingobium sp. BYY-5
Sphingobium chungbukense
Sphingobium sp. LB126
Sphingobium fuliginis ATCC 27551
Sphingobium sp. B11D3B
Sphingobium sp. B11D3D
Sphingobium sp. B7D2B
Sphingobium sp. B12D2B
Sphingobium sp. SYK-6
Sphingobium lignivorans
Sphingobium xanthum
Sphingobium yanoikuyae
Sphingobium

FIG. 10F

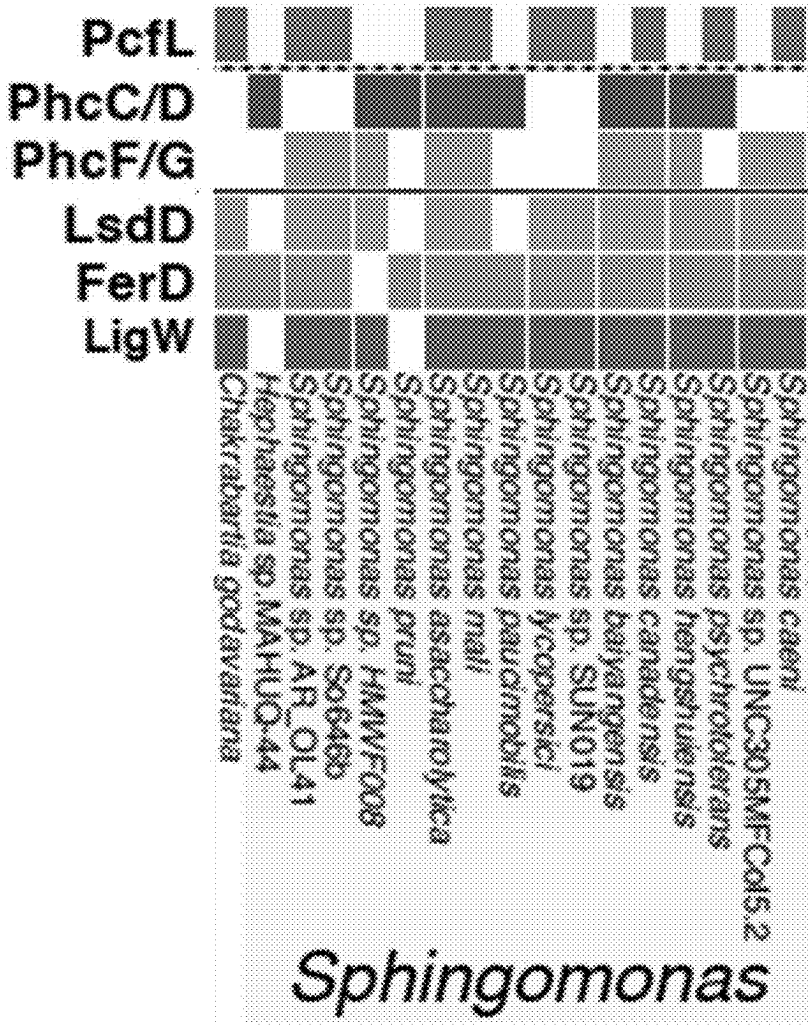

PcfL
PhcC/D
PhcF/G
LsdD
FerD
LigW
Chakrabartia godavariana
Hoeflea sp. MAH-UC-44
Sphingomonas sp. AR_OL41
Sphingomonas sp. Sc6420
Sphingomonas sp. HMWF008
Sphingomonas prunI
Sphingomonas asaccharolytica
Sphingomonas mali
Sphingomonas paucimobilis
Sphingomonas lycopersici
Sphingomonas sp. SUN019
Sphingomonas baiyangensis
Sphingomonas canadensis
Sphingomonas hengshuiensis
Sphingomonas psychrotolerans
Sphingomonas sp. UNC305MFCol5.2
Sphingomonas caeni
Sphingomonas

FIG. 10G

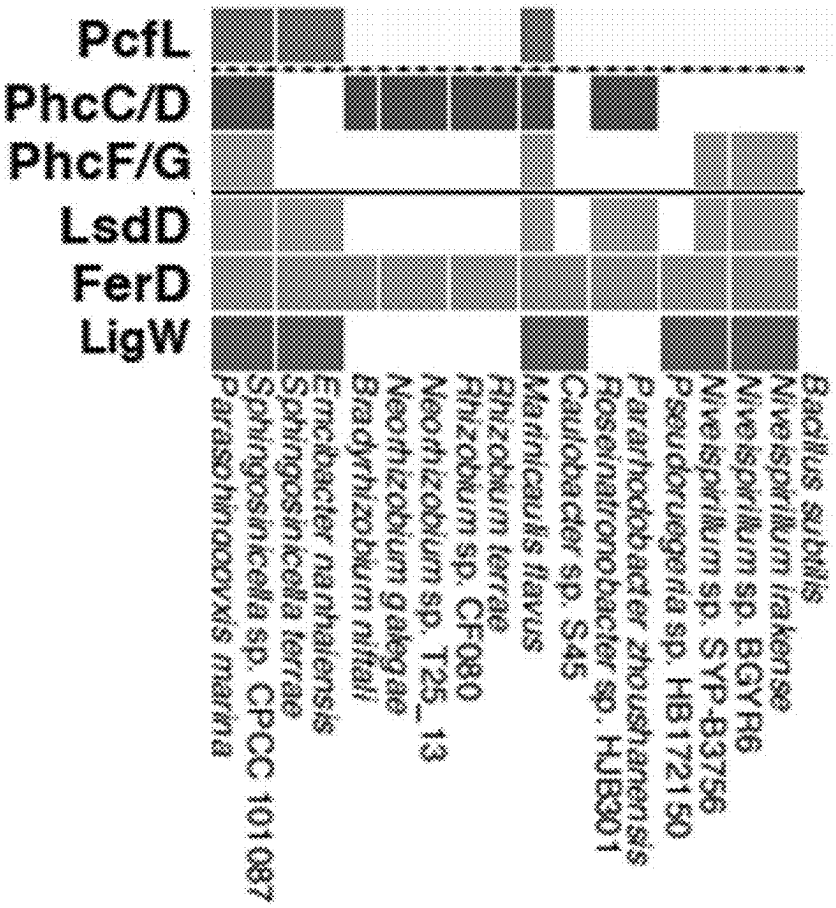

PcfL
PhcC/D
PhcF/G
LsdD
FerD
LigW
Bacillus subtilis
Niveisporium rakense
Niveisporium sp. BGYR6
Niveisporium sp. SYP-B3756
Pseudoruegeria sp. HB172150
Parahaliclobacter zhoushanensis
Roseinatronobacter sp. HUB301
Caulobacter sp. S45
Marinicaulis flavus
Rhizobium terrae
Rhizobium sp. CF080
Neorhizobium sp. T25_13
Bradyrhizobium algae
Bradyrhizobium niftali
Emcibacter nanhaiensis
Sphingosinicella terrae
Sphingosinicella sp. CPCC 101087
Parasphingopyxis marina

FIG. 10G

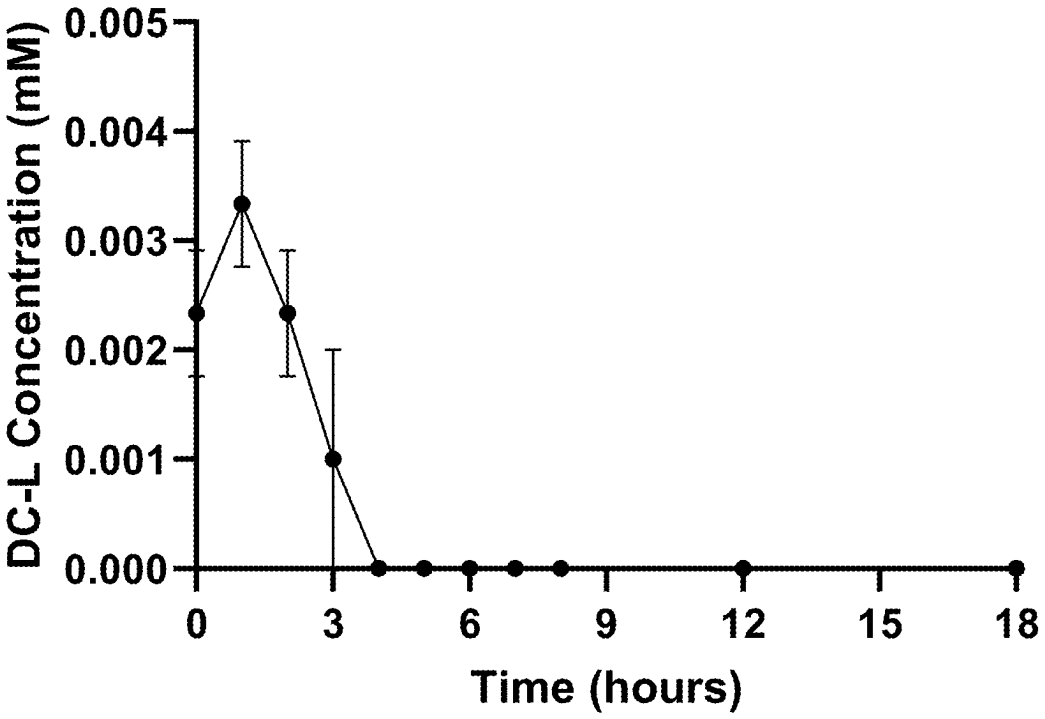

DC-L Concentration (mM)
0.005
0.004
0.003
0.002
0.001
0.000
0
3
6
9
12
15
18
Time (hours)

FIG. 11

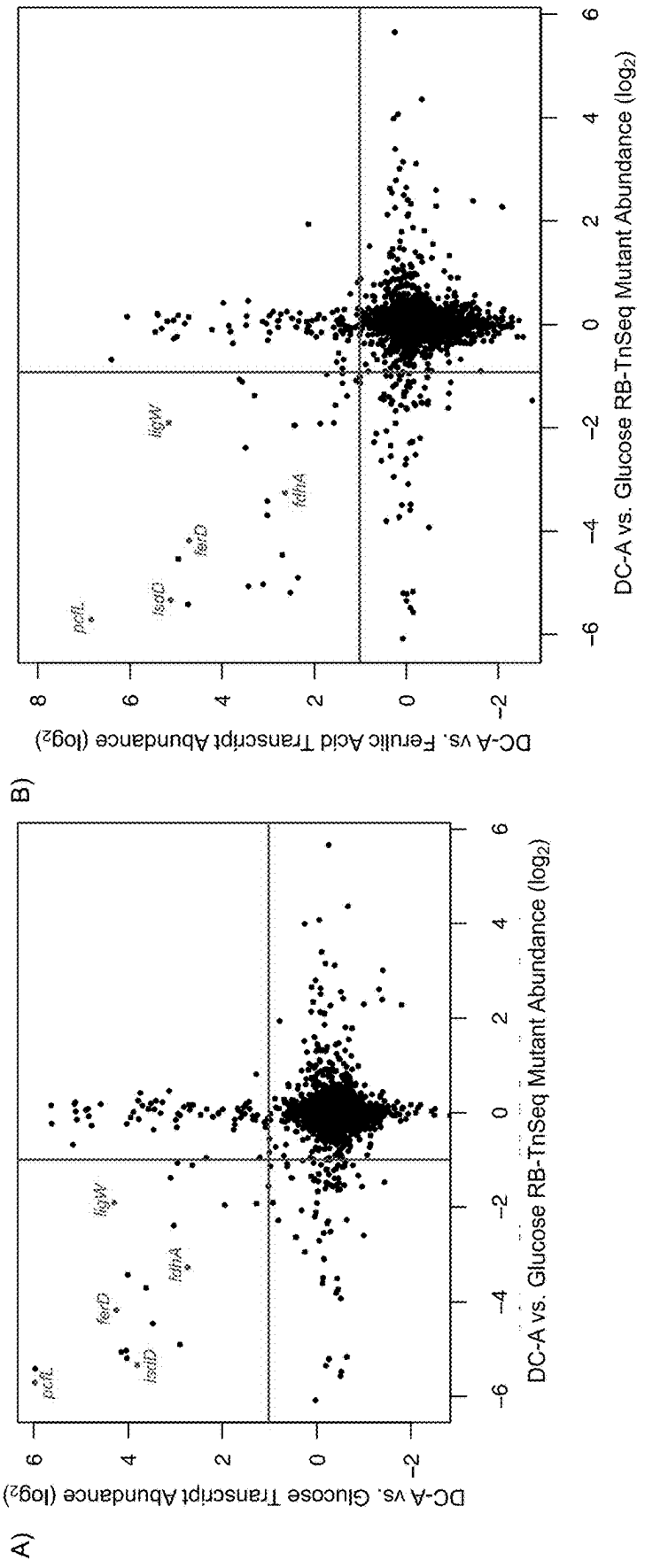

A)
DC-A vs. Glucose Transcript Abundance (log2)
DC-A vs. Glucose RB-TnSeq Mutant Abundance (log2)
pcfL
lstD
ferD
fdhA
ligW
B)
DC-A vs. Ferulic Acid Transcript Abundance (log2)
DC-A vs. Glucose RB-TnSeq Mutant Abundance (log2)
pcfL
lstD
ferD
fdhA
ligW

FIG. 12

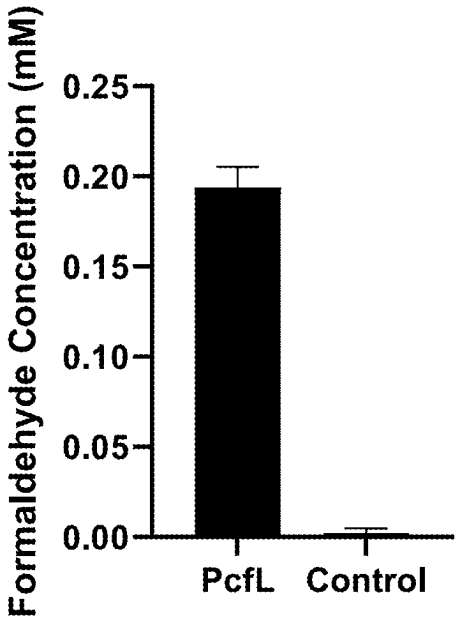

Formaldehyde Concentration (mM)
0.25
0.20
0.15
0.10
0.05
0.00
PcfL
Control

FIG. 13

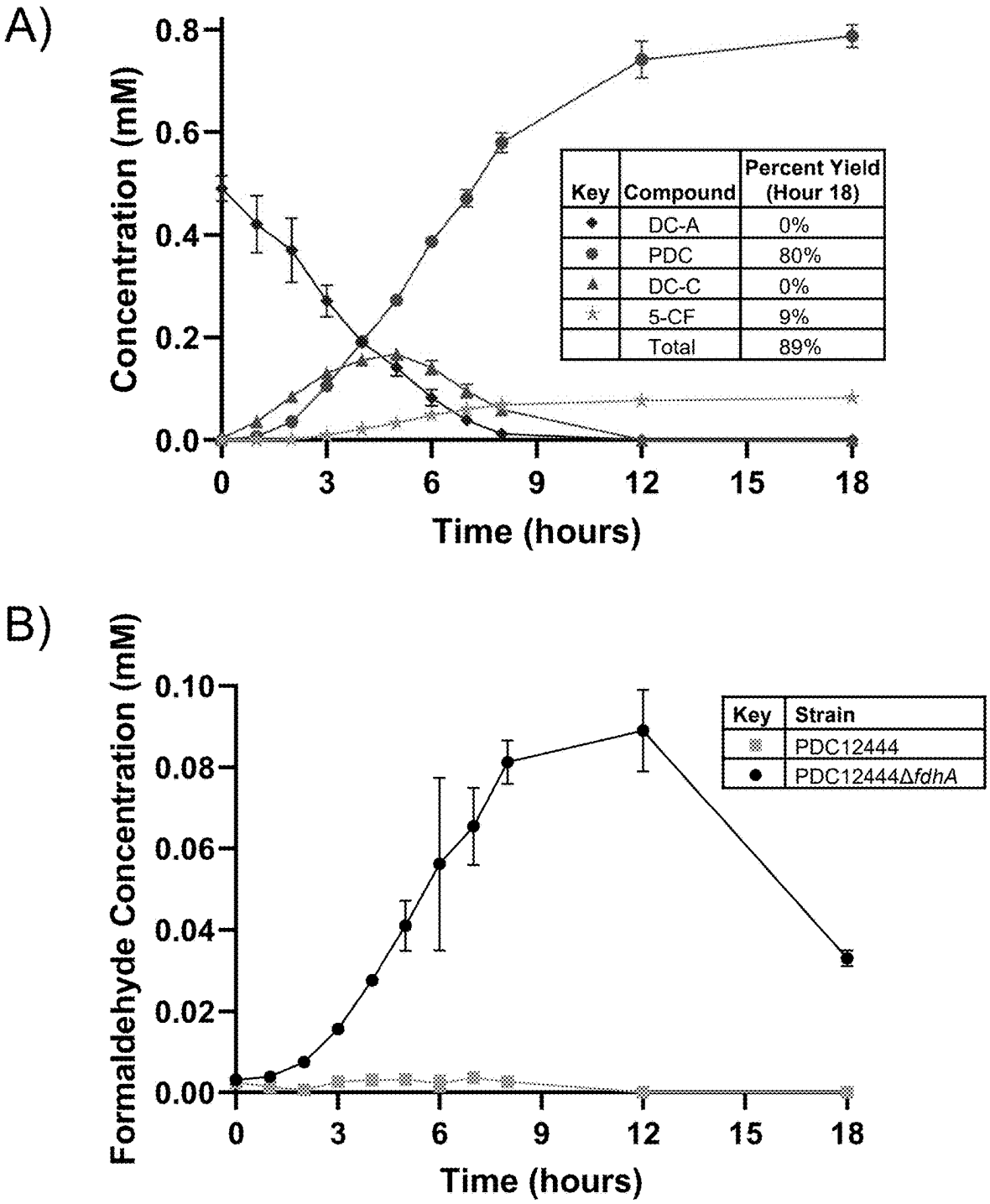

A)
Concentration (mM)
0.0
0.2
0.4
0.6
0.8
0
3
6
9
12
15
18
Time (hours)
Key
Compound
Percent Yield (Hour 18)
DC-A
0%
PDC
80%
DC-C
0%
5-CF
9%
Total
89%
B)
Formaldehyde Concentration (mM)
0.00
0.02
0.04
0.06
0.08
0.10
0
3
6
9
12
15
18
Time (hours)
Key
Strain
PDC12444
PDC12444ΔfdhA

FIG. 14

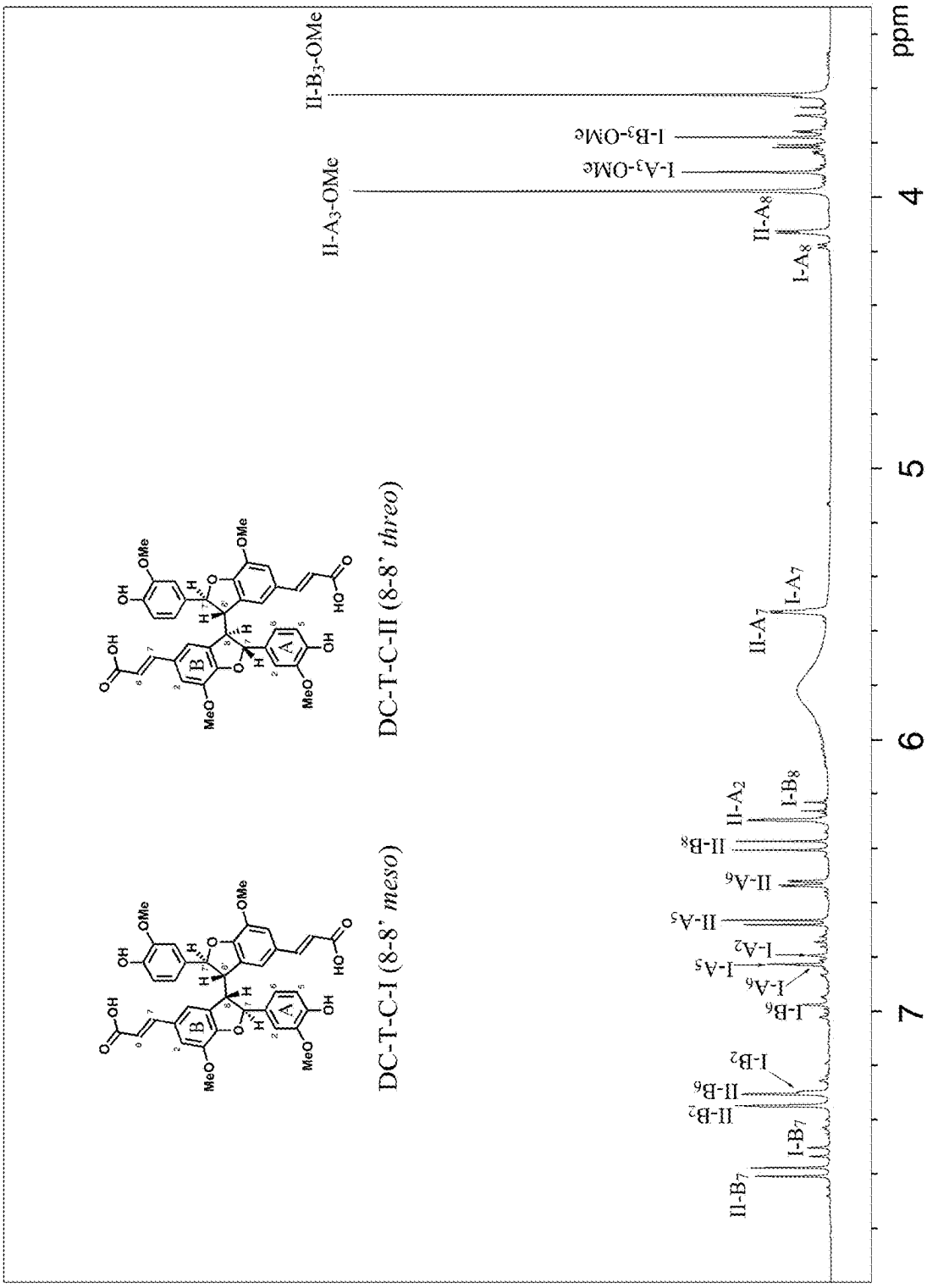

DC-T-C-I (8-8' meso)
DC-T-C-II (8-8' threo)
II-B3-OMe
I-B3-OMe
I-A3-OMe
II-A3-OMe
II-A8
I-A8
II-A7
I-A7
II-A2
I-B8
II-B8
II-A6
II-A5
I-A2
I-A5
I-A6
I-B6
I-B2
II-B6
II-B2
I-B7
II-B7
ppm
4
5
6
7

FIG. 15A

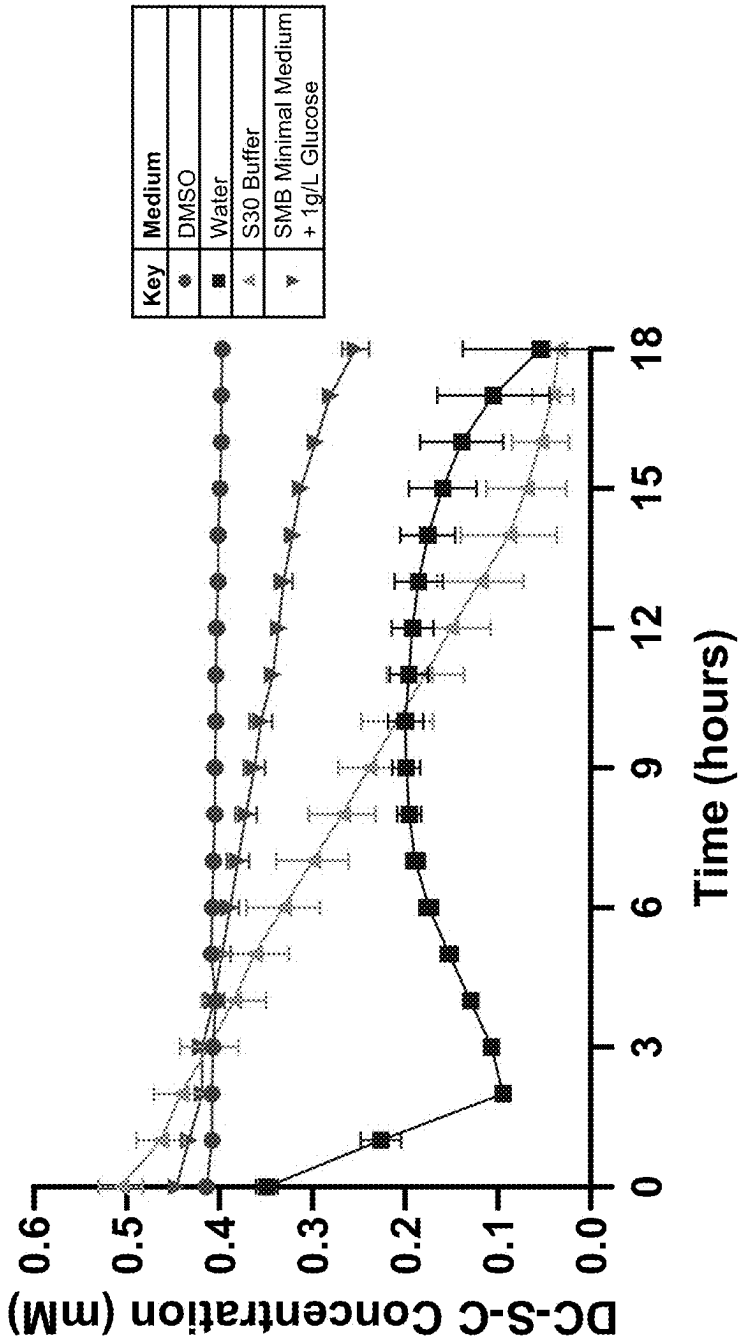

Key
Medium
DMSO
Water
S30 Buffer
SMB Minimal Medium + 1g/L Glucose
DC-S-C Concentration (mM)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
3
6
9
12
15
18
Time (hours)

FIG. 15B

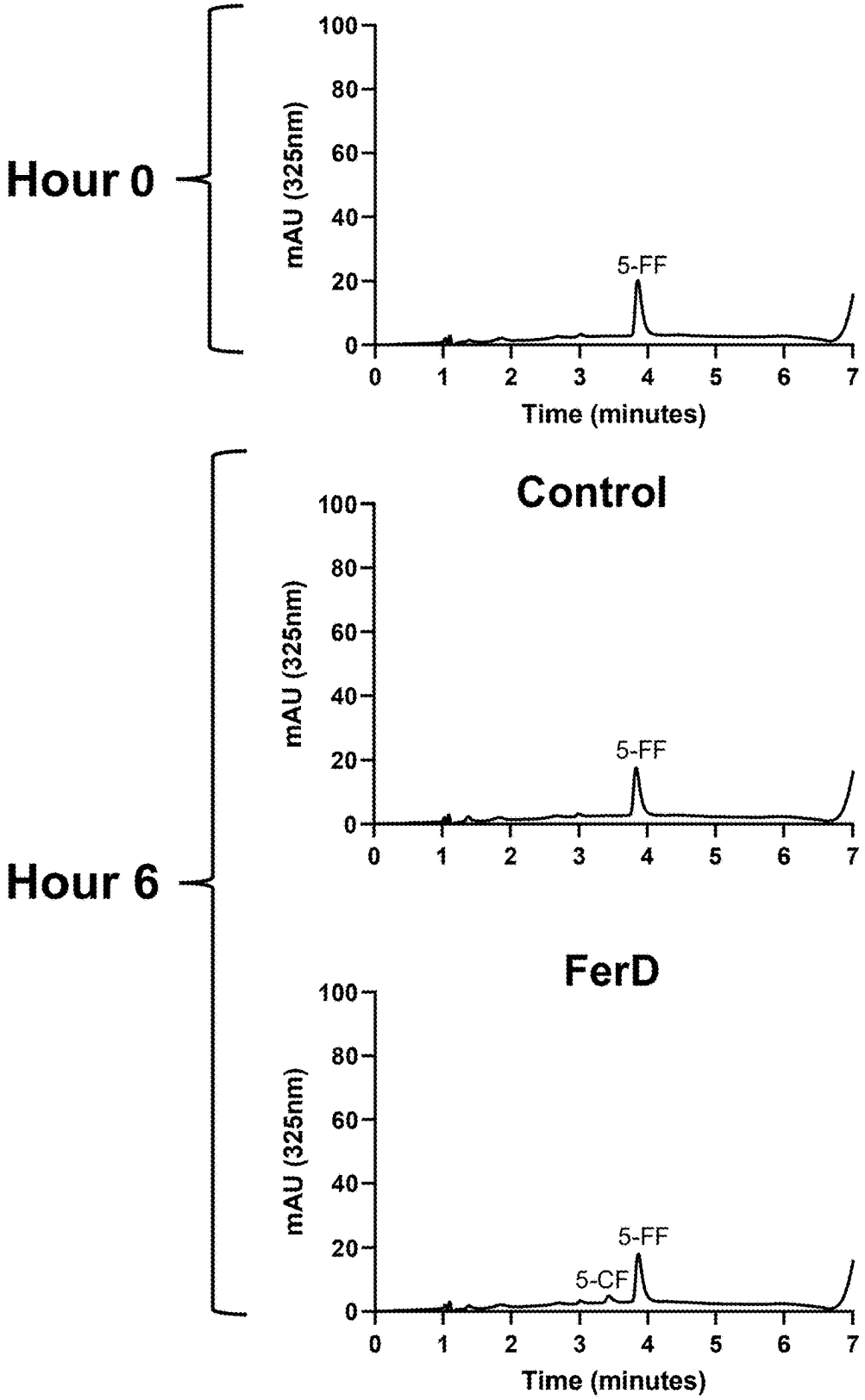

Hour 0
Hour 6
Control
FerD
mAU (325nm)
100
80
60
40
20
0
5-FF
5-CF
0 1 2 3 4 5 6 7
Time (minutes)

FIG. 16A

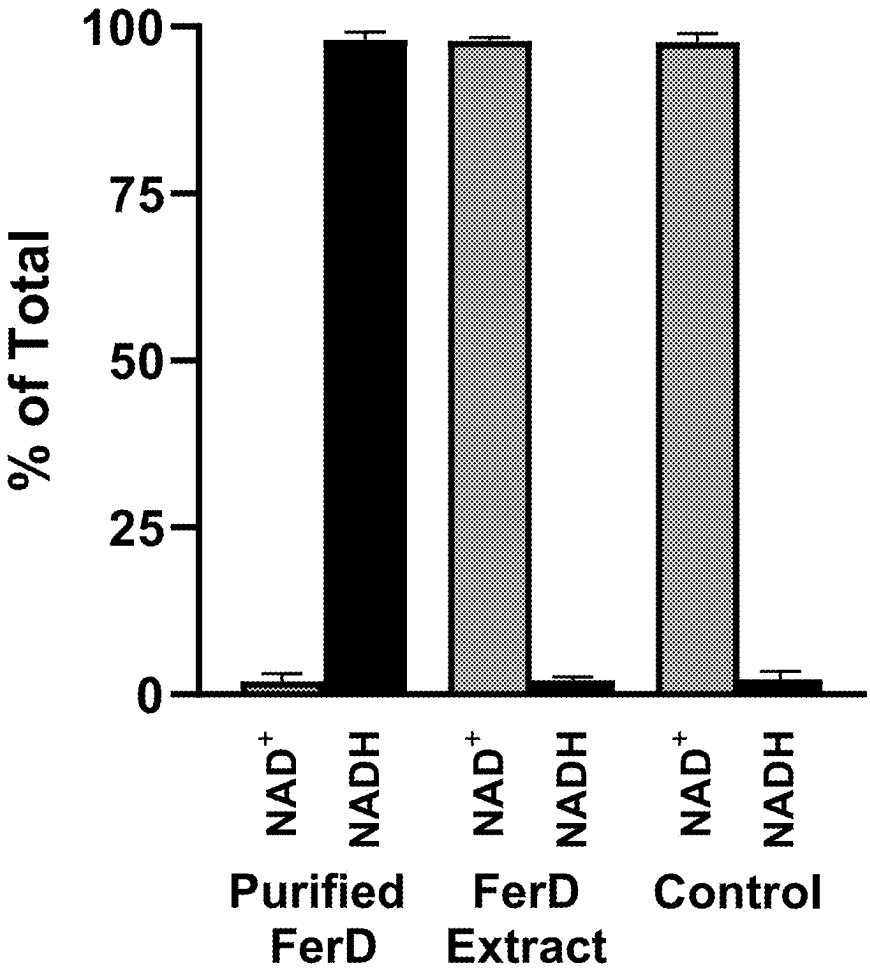

100
75
50
25
0
% of Total
NAD+
NADH
NAD+
NADH
NAD+
NADH
Purified FerD
FerD Extract
Control

FIG. 16B

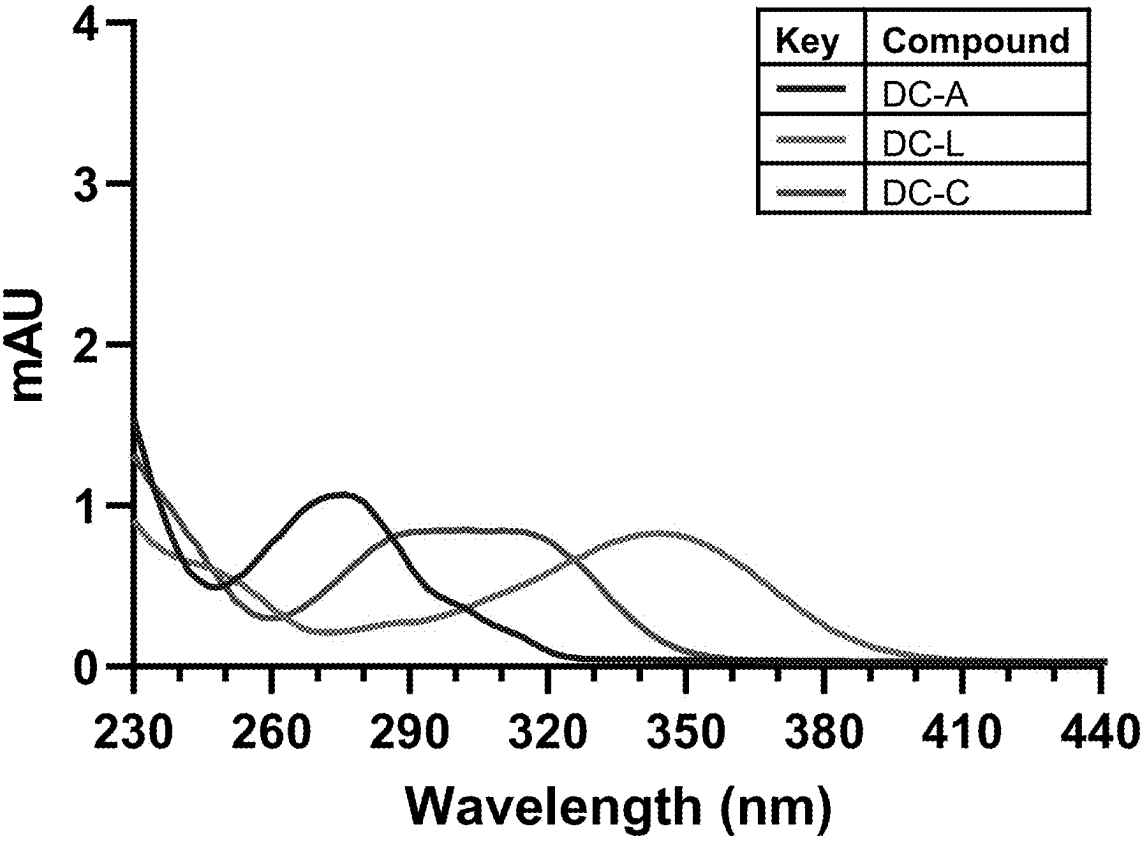

Key
Compound
DC-A
DC-L
DC-C
mAU
0
1
2
3
4
230
260
290
320
350
380
410
440
Wavelength (nm)

FIG. 17

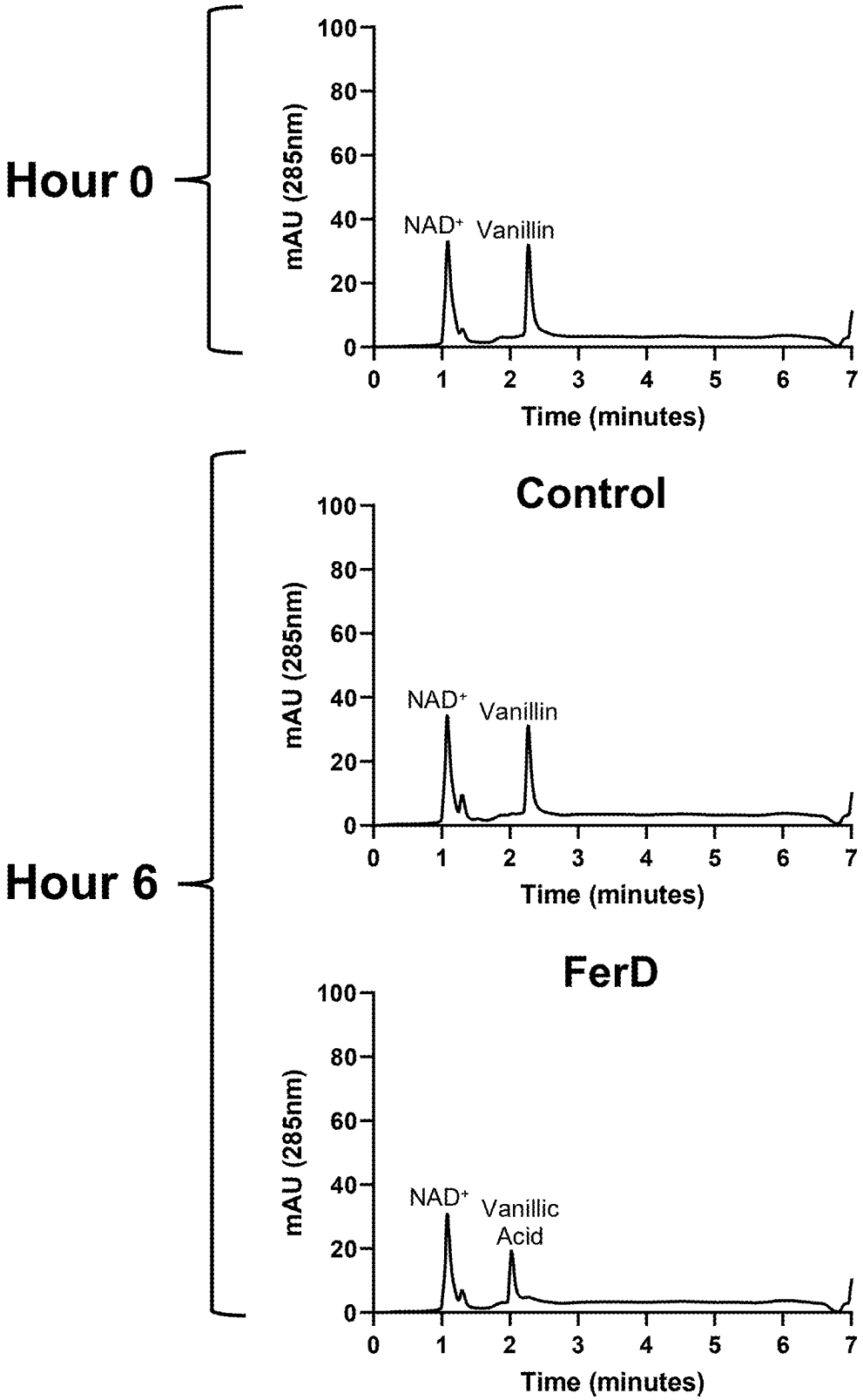

Hour 0
mAU (285nm)
NAD+
Vanillin
Time (minutes)
Hour 6
Control
mAU (285nm)
NAD+
Vanillin
Time (minutes)
FerD
mAU (285nm)
NAD+
Vanillic Acid
Time (minutes)

FIG. 18

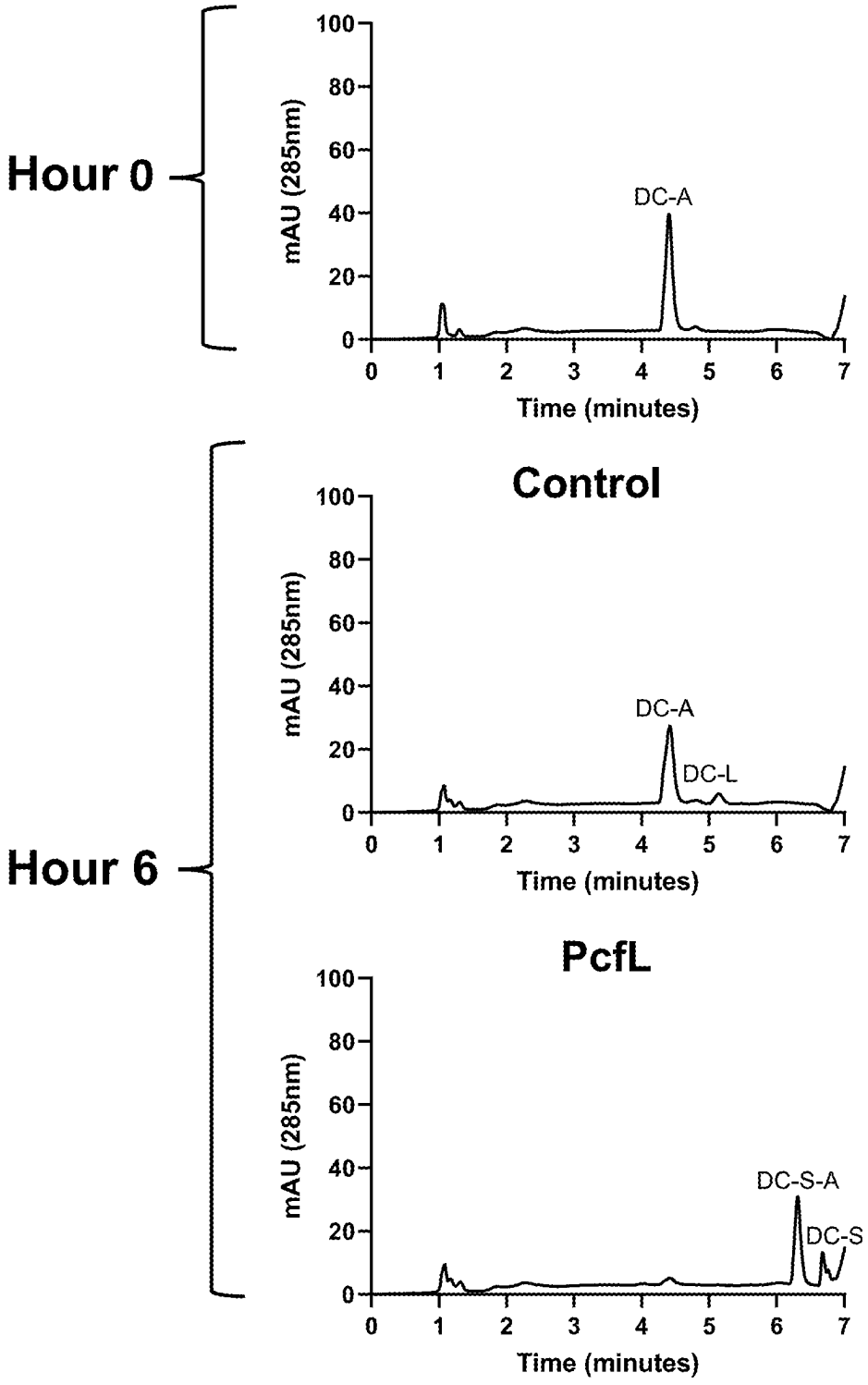

Hour 0
mAU (285nm)
100
80
60
40
20
0
DC-A
0 1 2 3 4 5 6 7
Time (minutes)
Hour 6
Control
mAU (285nm)
DC-A
DC-L
Time (minutes)
PcfL
mAU (285nm)
DC-S-A
DC-S
Time (minutes)

FIG. 19A

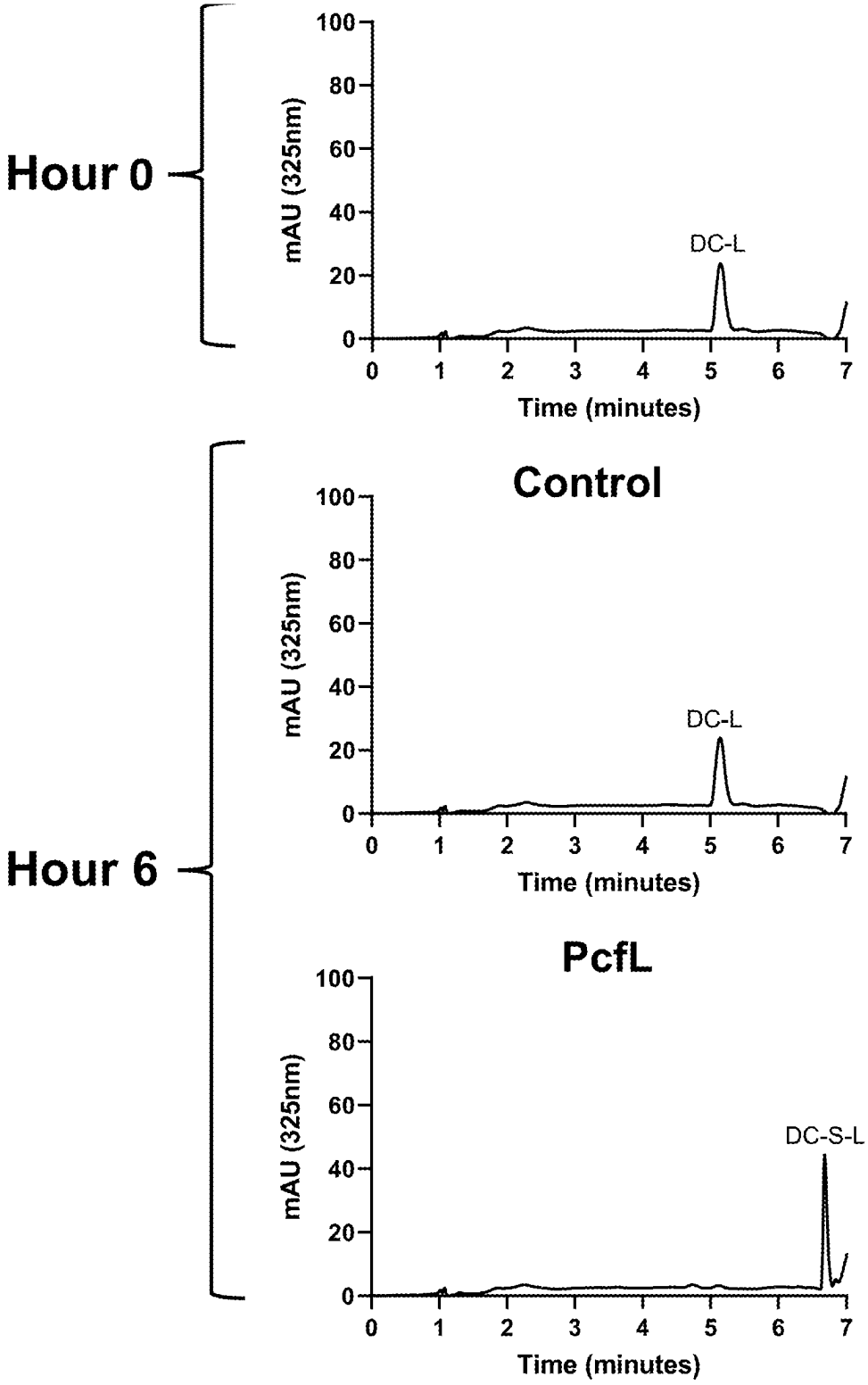
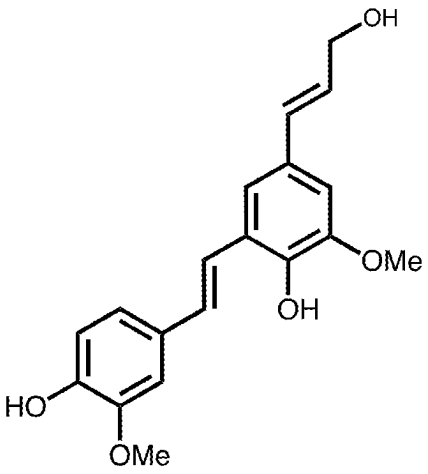
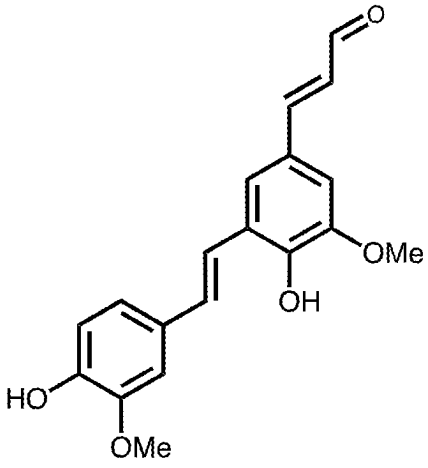

Hour 0
Hour 6
Control
PcfL
mAU (325nm)
Time (minutes)
DC-L
DC-S-L
0
20
40
60
80
100
0
1
2
3
4
5
6
7

FIG. 19B



DC-S-A



DC-S-L

FIG. 19C

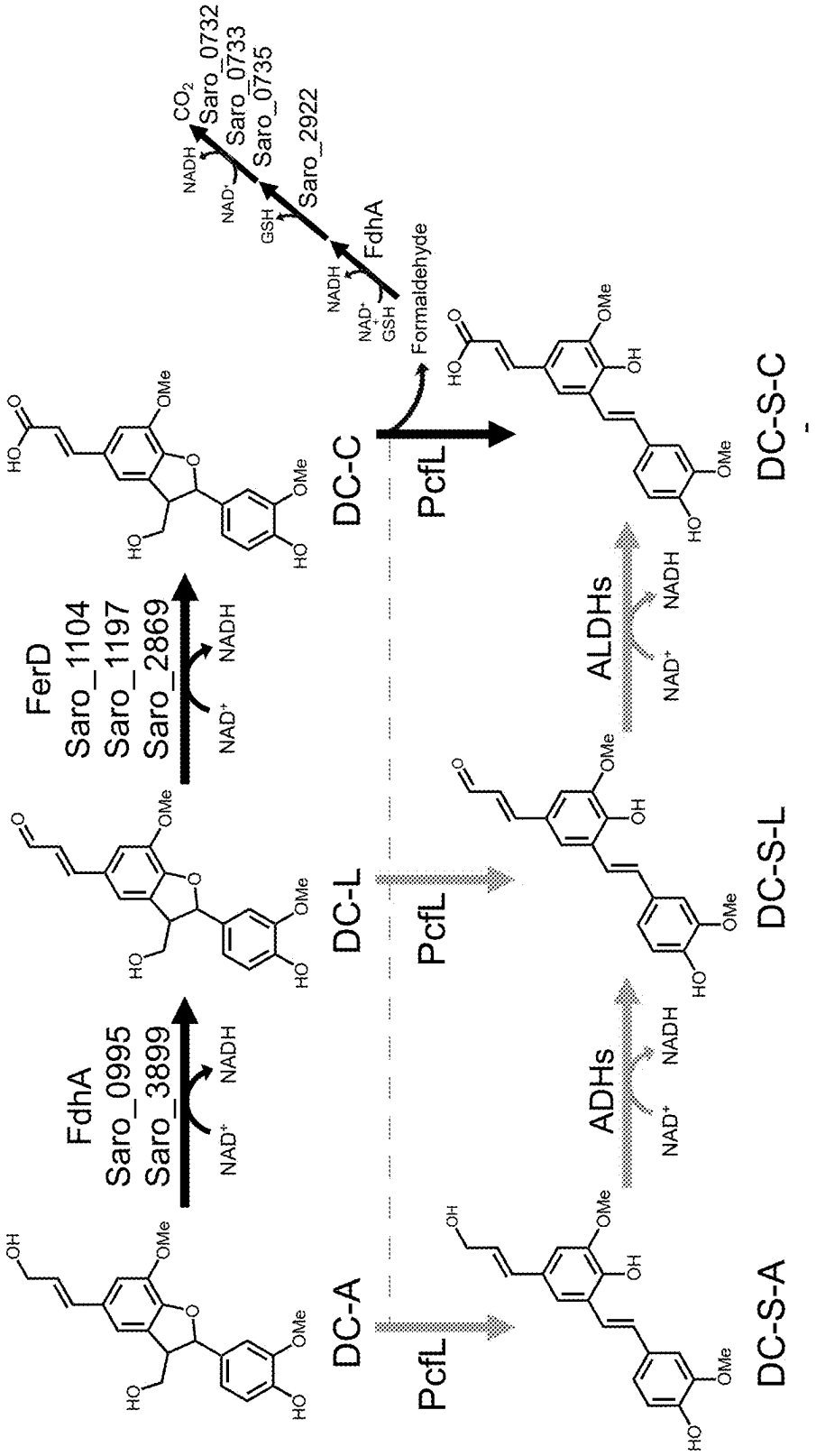

DC-A
PcfL
DC-S-A
FdhA
Saro_0995
Saro_3899
NAD+
NADH
DC-L
PcfL
DC-S-L
ADHs
NAD+
NADH
FerD
Saro_1104
Saro_1197
Saro_2869
NAD+
NADH
DC-C
PcfL
DC-S-C
ALDHs
NAD+
NADH
FdhA
NADH
NAD+
GSH
Formaldehyde
Saro_2922
GSH
NAD+
Saro_0735
Saro_0733
Saro_0732
NADH
CO2

FIG. 20

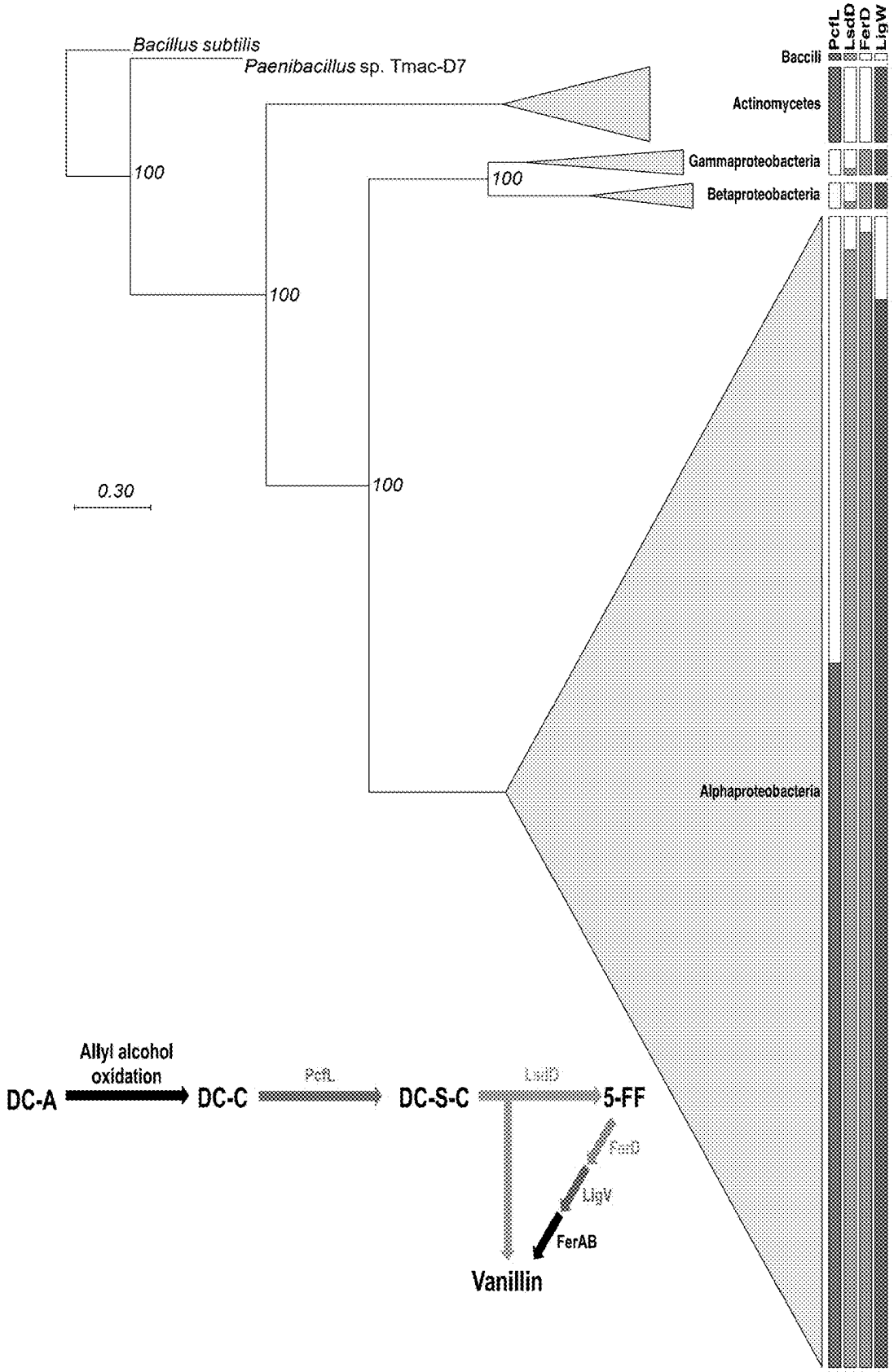

PcfL
LsdD
FerD
LigW
Bacillus subtilis
Paenibacillus sp. Tmac-D7
Bacilli
Actinomycetes
Gammaproteobacteria
Betaproteobacteria
Alphaproteobacteria
100
100
100
100
0.30
Allyl alcohol oxidation
DC-A
DC-C
PcfL
DC-S-C
LsdD
5-FF
FerD
LigV
FerAB
Vanillin

FIG. 21A

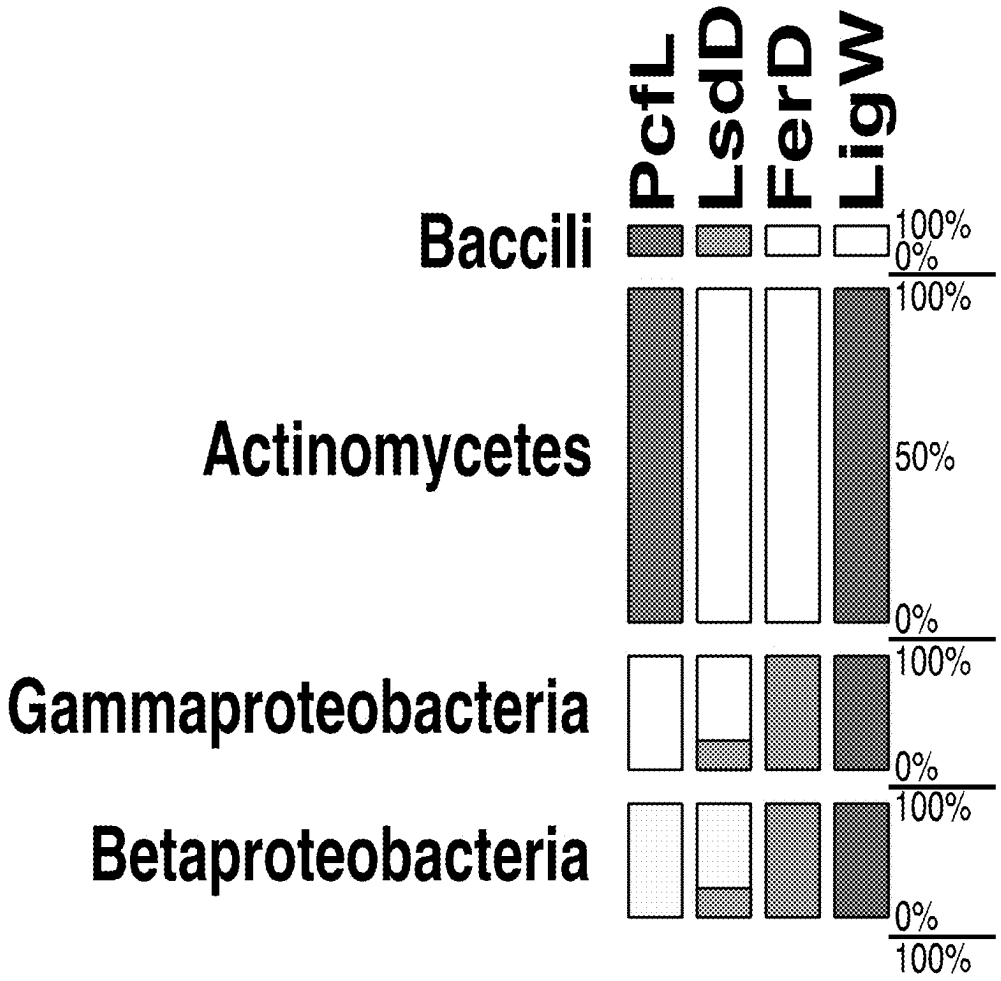

PcfL
LsdD
FerD
LigW
Baccili
Actinomycetes
Gammaproteobacteria
Betaproteobacteria
100%
0%
100%
50%
0%
100%
0%
100%
0%
100%

FIG. 21B

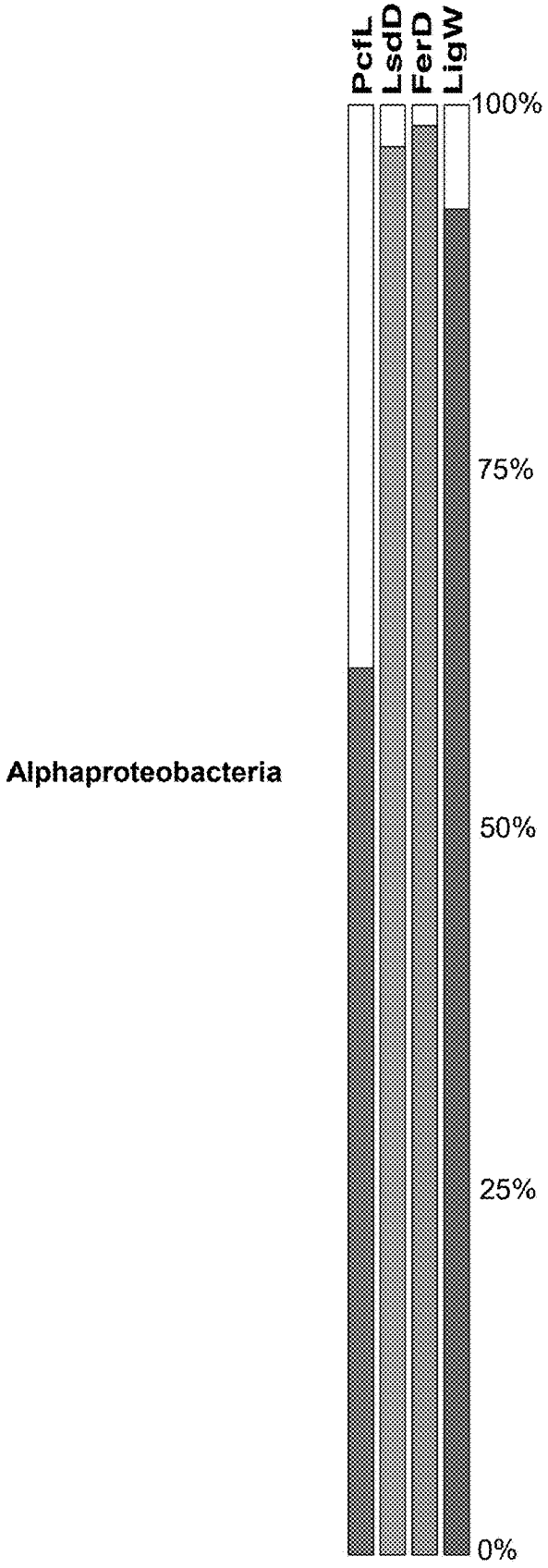


FIG. 21C

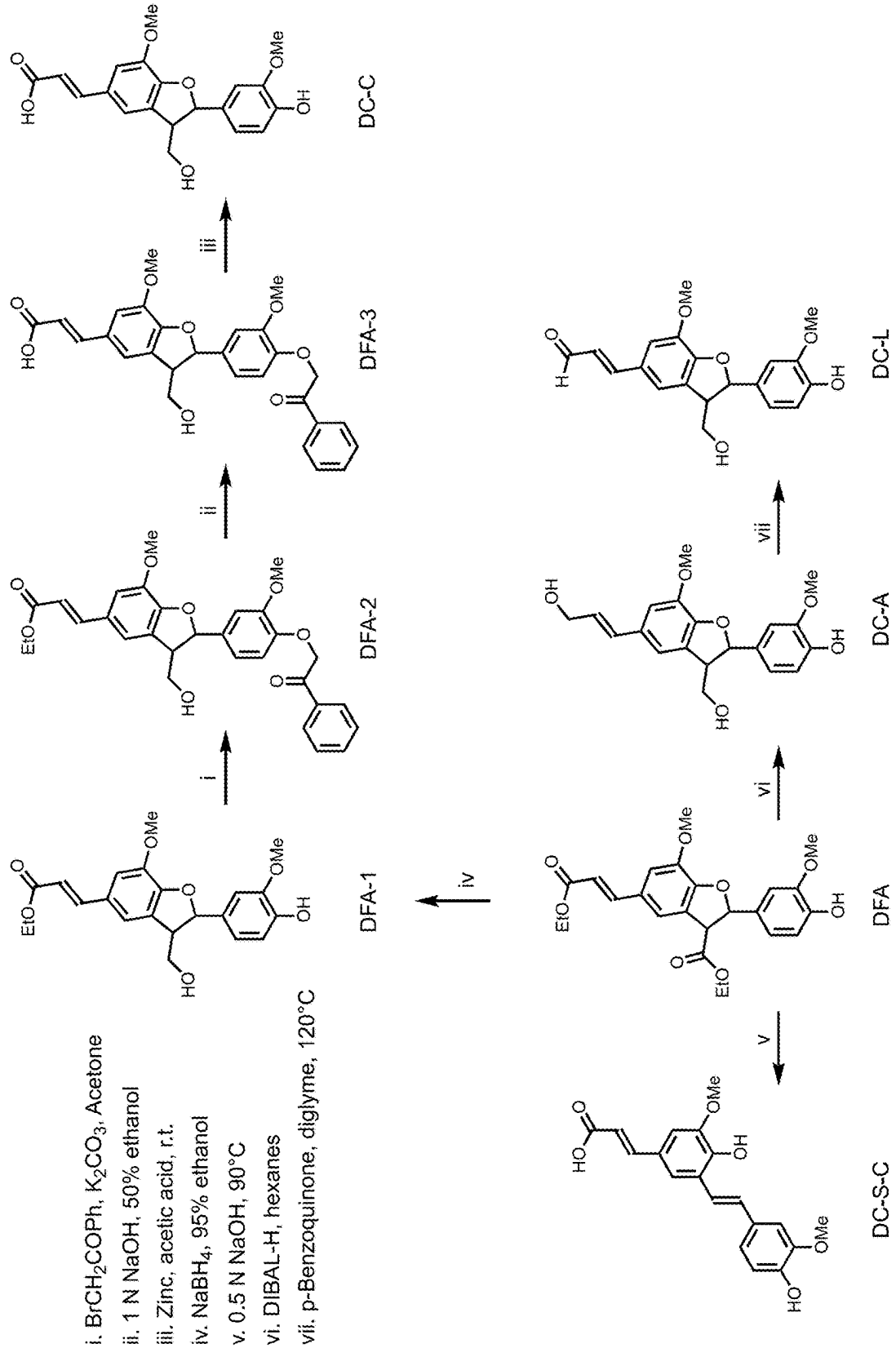

i. $BrCH_2COPh$, K_2CO_3, Acetone
ii. 1 N NaOH, 50% ethanol
iii. Zinc, acetic acid, r.t.
iv. $NaBH_4$, 95% ethanol
v. 0.5 N NaOH, 90°C
vi. DIBAL-H, hexanes
vii. p-Benzoquinone, diglyme, 120°C
DFA-1
DFA-2
DFA-3
DC-C
DFA
DC-A
DC-L
DC-S-C

FIG. 22A

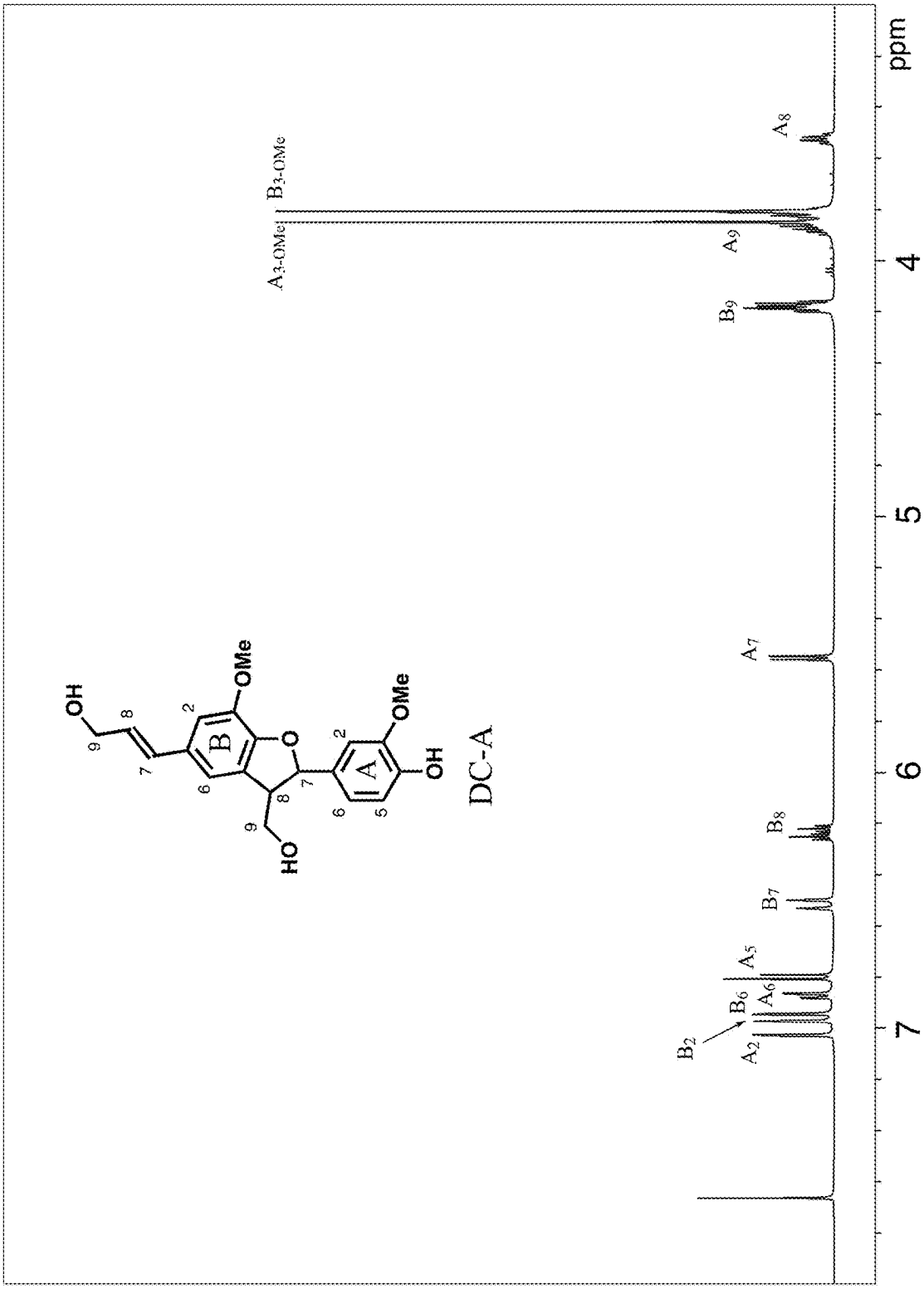

OH
OMe
OMe
OH
HO
B
A
DC-A
A3-OMe
B3-OMe
A8
A9
B9
A7
B8
B7
A5
B6
A6
A2
B2
ppm
7
6
5
4

FIG. 22B

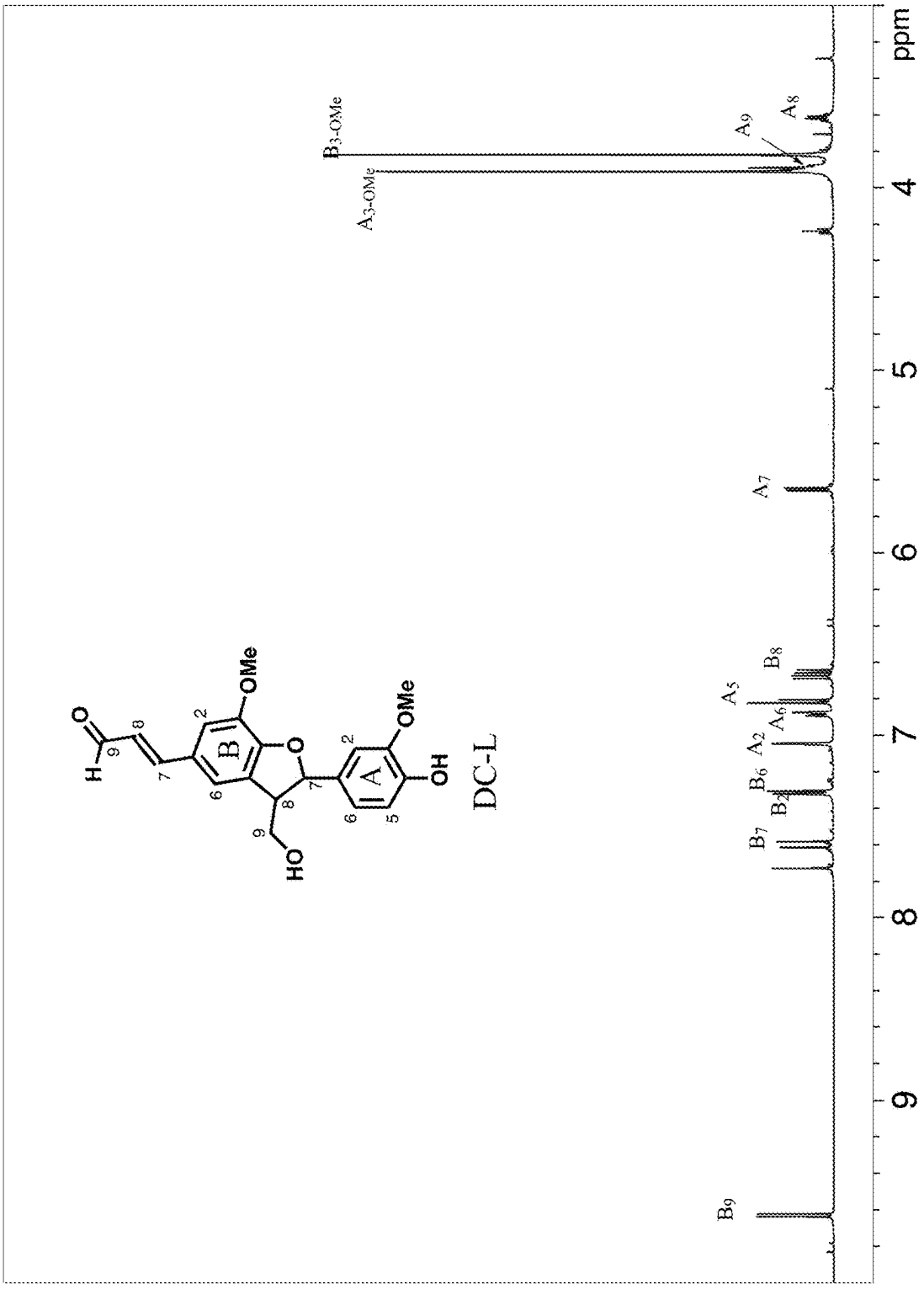


FIG. 22C

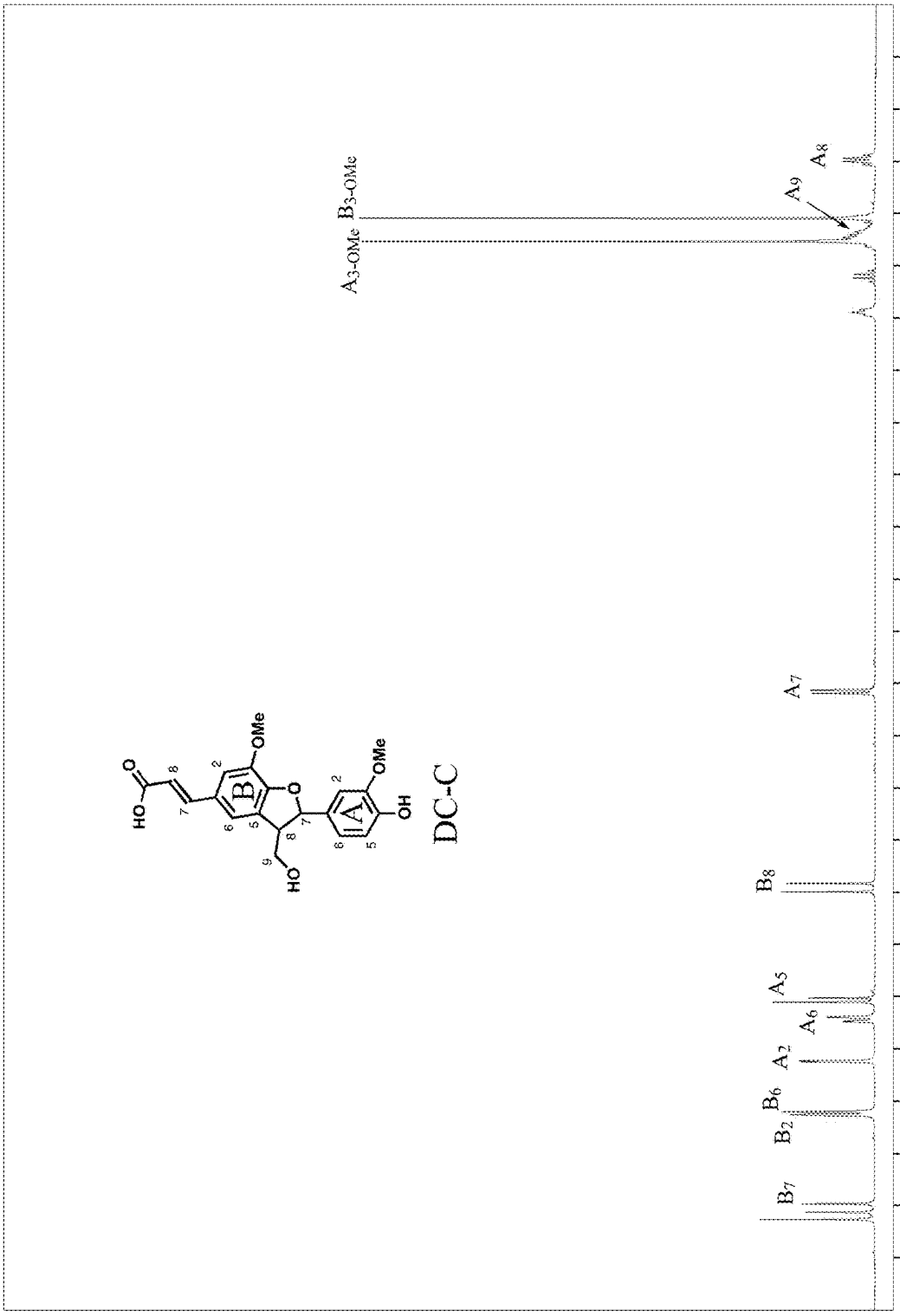


FIG. 22D

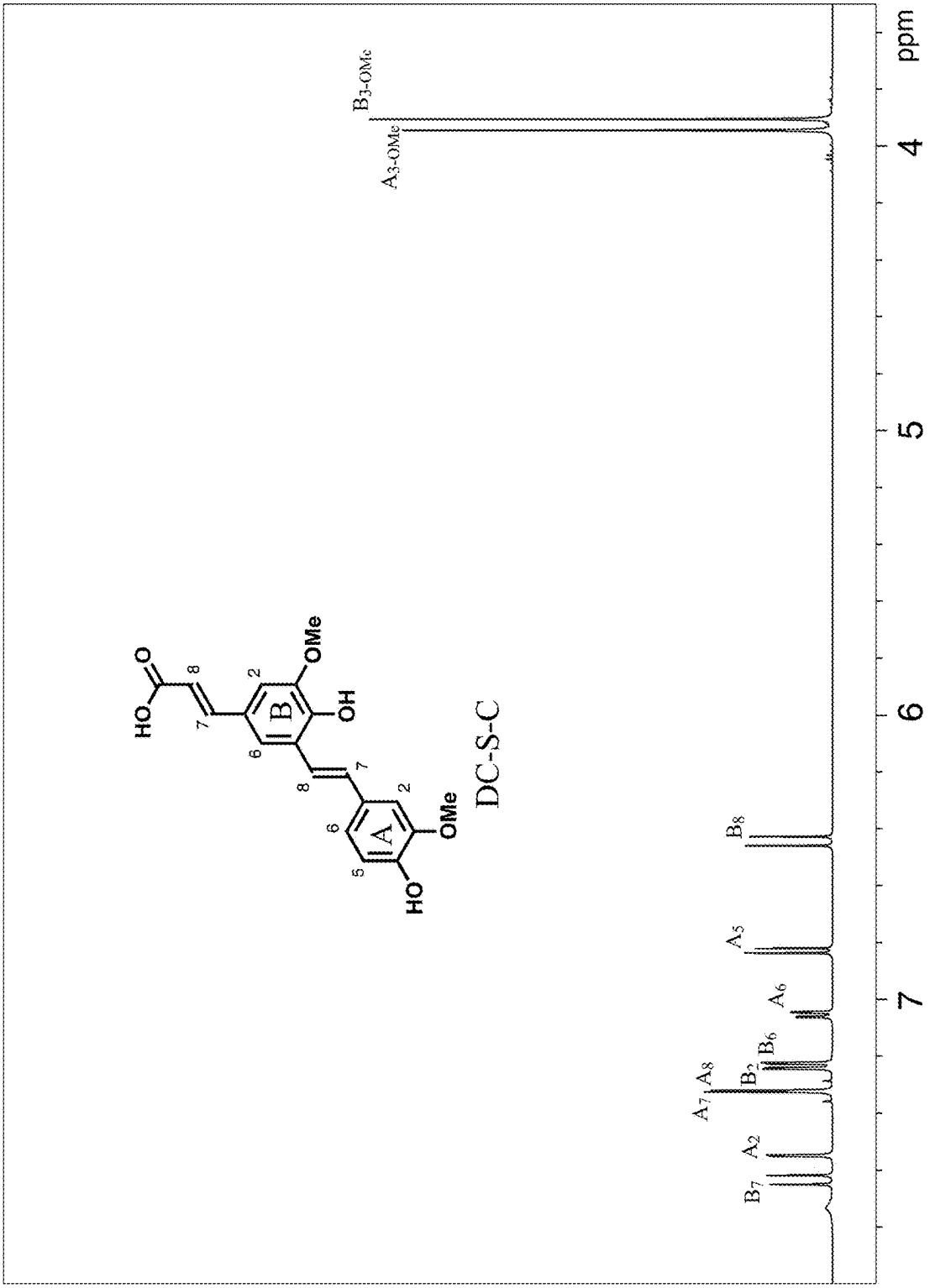

DC-S-C
HO
O
OMe
OH
OMe
HO
A
B
$B_{3\text{-}OMe}$
$A_{3\text{-}OMe}$
B_8
A_5
A_6
B_6
B_2
A_8
A_7
A_2
B_7
7
6
5
4
ppm

FIG. 22E

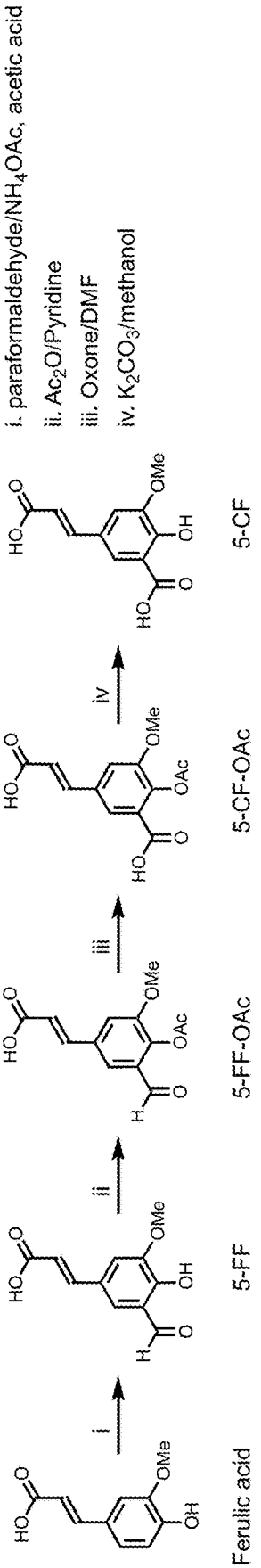

i. paraformaldehyde/NH_4OAc, acetic acid
ii. Ac_2O/Pyridine
iii. Oxone/DMF
iv. K_2CO_3/methanol
Ferulic acid
5-FF
5-FF-OAc
5-CF-OAc
5-CF

FIG. 23A

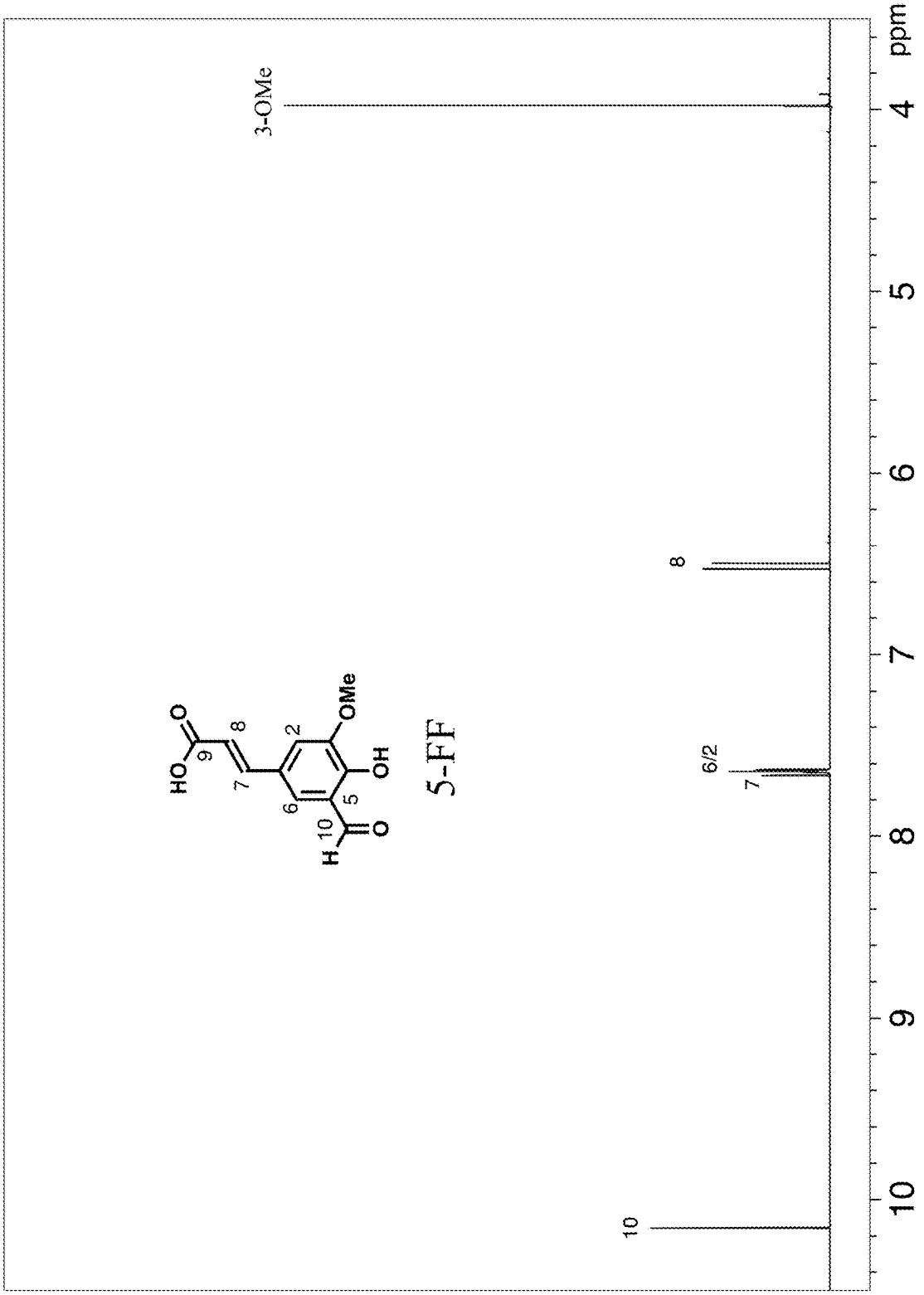


FIG. 23B

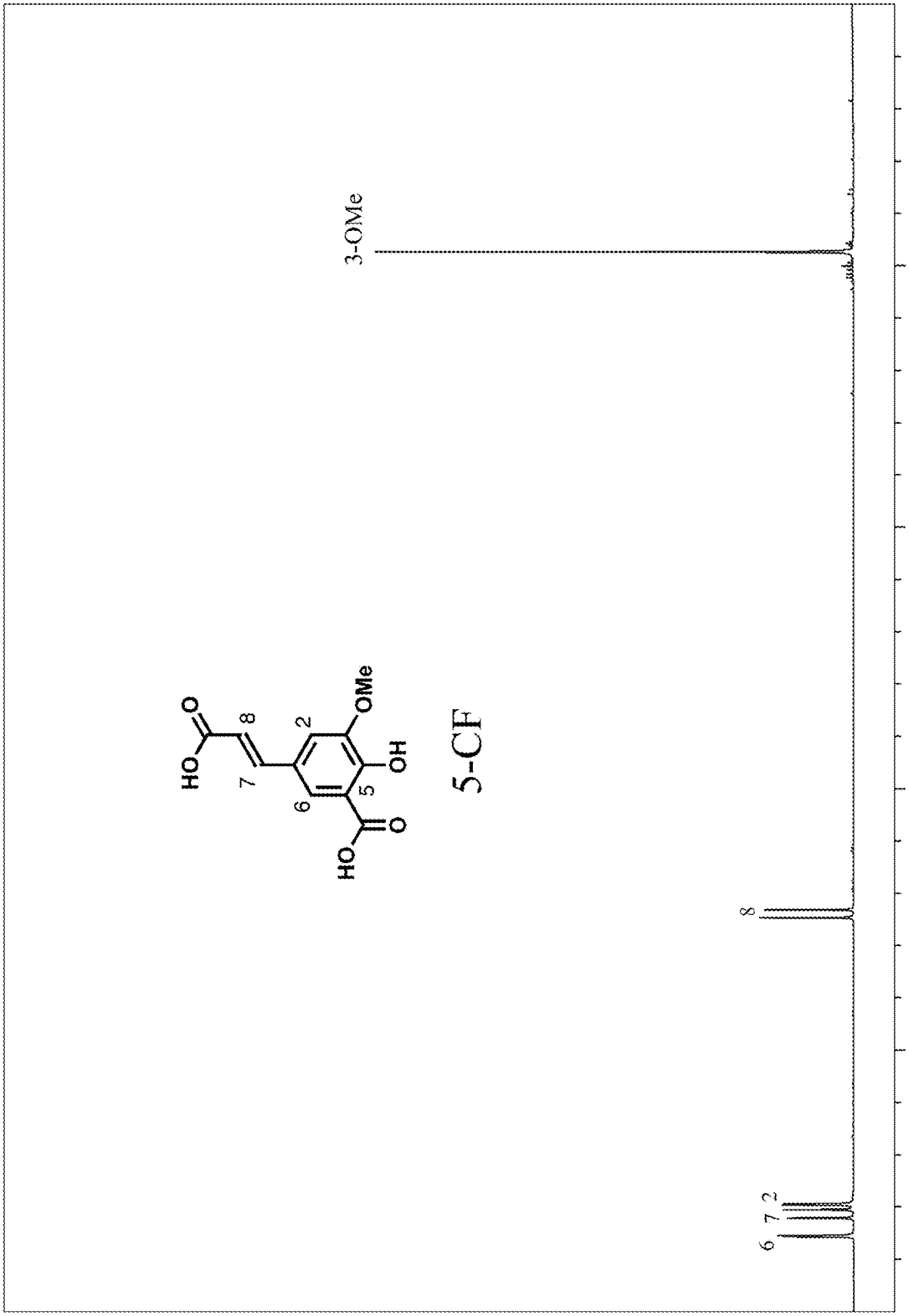


FIG. 23C

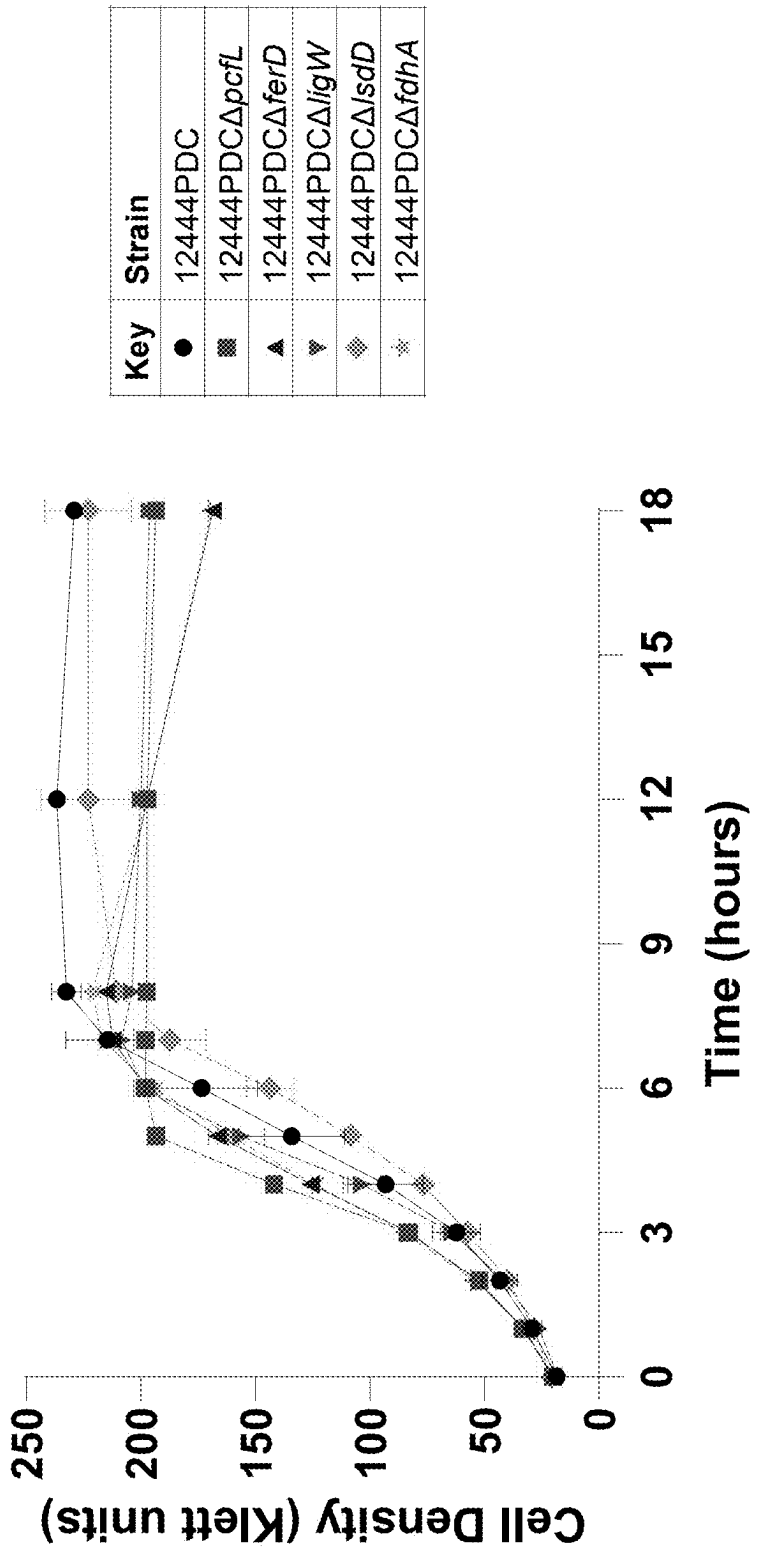

Key
Strain
12444PDC
12444PDCΔpcfL
12444PDCΔferD
12444PDCΔligW
12444PDCΔlsdD
12444PDCΔfdhA
Cell Density (Klett units)
250
200
150
100
50
0
0
3
6
9
12
15
18
Time (hours)

FIG. 24

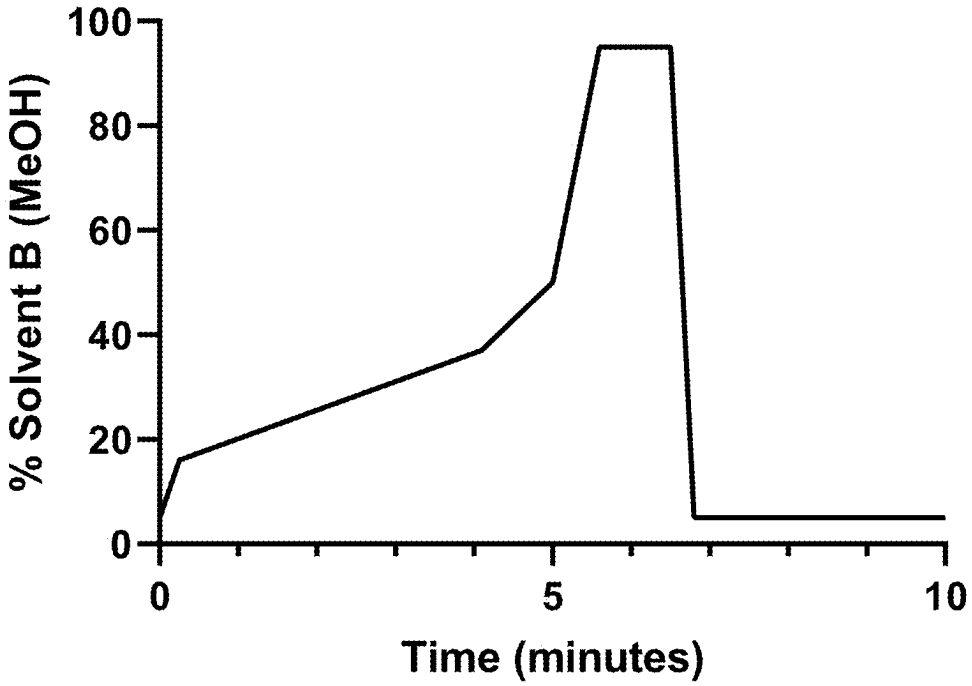

100
80
60
40
20
0
% Solvent B (MeOH)
0
5
10
Time (minutes)

FIG. 25

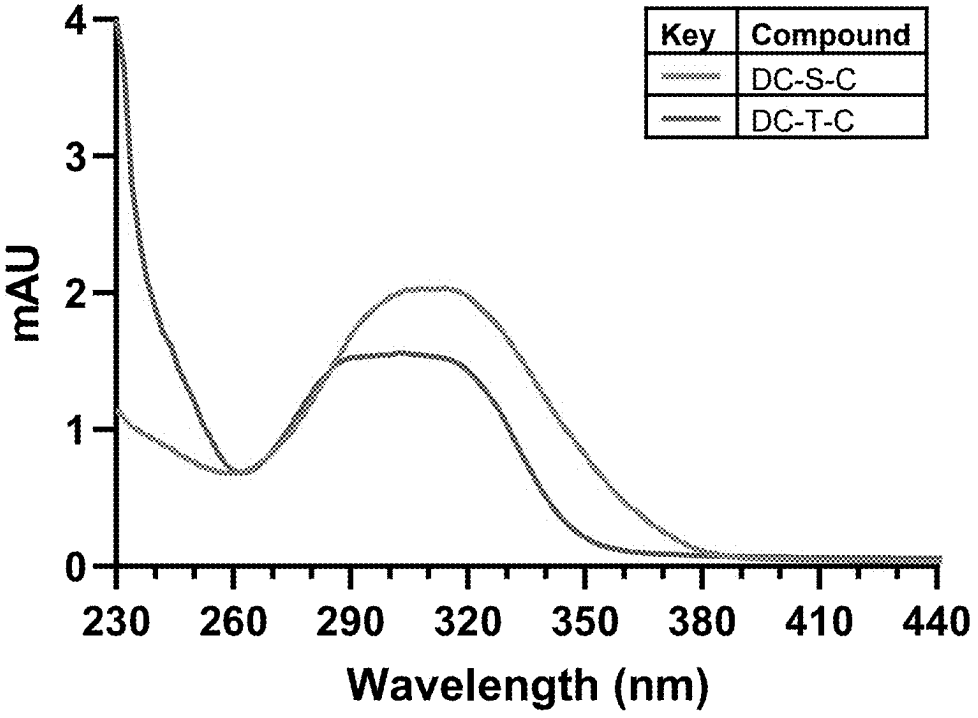

Key
Compound
DC-S-C
DC-T-C
4
3
2
1
0
mAU
230
260
290
320
350
380
410
440
Wavelength (nm)

FIG. 26

RECOMBINANT MICROORGANISMS THAT CATABOLIZE LIGNIN AROMATICS AND METHODS OF USING SAME

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0001] This invention was made with government support under DE-SC0018409 awarded by the US Department of Energy. The government has certain rights in the invention.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in XML format and is hereby incorporated by reference in its entirety. The XML copy, created on May 31, 2024, is named USPTO-24607-09824544-P240270US01-SEQ_LIST.xml and is 140,384 bytes in size.

FIELD OF THE INVENTION

[0003] The invention is directed to recombinant microorganisms that catabolize lignin aromatics, such as 3-5 linked lignin aromatics, and methods of using same to catabolize the lignin aromatics.

BACKGROUND

[0004] Over the past century, aromatic compounds have proven integral to industries that generate critical chemicals and materials for society. For example, aromatic compounds are precursors for the production of plastics, adhesives, medicinal compounds, and flavorings. Most of today's industrial aromatics are derived from fossil fuels. However, there is increasing interest in identifying renewable raw materials that can serve as alternative sources of these valuable chemicals.

[0005] The plant polymer lignin can comprise up to 40% of the dry weight of plant biomass, making it the second most abundant biopolymer on the planet (1) and an attractive source of renewable aromatics for producing chemicals. Lignin is a heteropolymer composed of syringyl (S), guaiacyl (G), and p-hydroxyphenyl (H) aromatic subunits which differ in the number of methoxy groups attached to the aromatic ring (two, one, or zero, respectively) (2, 3). Since lignin polymers are synthesized via radical chemistry in plants, the aromatic subunits are joined by a variety of interunit bonds (FIG. **1** (A)) (4-6). The chemical heterogeneity of its inter-aromatic linkages makes lignin recalcitrant to break down, so it has traditionally been burned for fuel (1, 7, 8). However, strategies are emerging to convert the aromatic subunits of lignin to commodity chemicals and materials that are needed by society (2, 8).

[0006] One promising strategy is to use the aromatic compounds resulting from depolymerization of lignin as carbon sources that microbes can funnel into valuable products (9-12). Microbes suitable for this purpose are needed.

SUMMARY OF THE INVENTION

[0007] One aspect of the invention is directed recombinant microorganisms. The recombinant microorganisms can comprise any one or more, any two or more, any three or more, any four or more, or each of: one or more recombinant alcohol dehydrogenase genes; one or more recombinant aldehyde dehydrogenase genes; a recombinant T-formaldehyde lyase gene; a recombinant lignostilbene dioxygenase gene; and a recombinant aromatic acid decarboxylase gene.

[0008] In some versions, the recombinant microorganism comprises any two or more, any three or more, any four or more, or each of: the one or more recombinant alcohol dehydrogenase genes; the one or more recombinant aldehyde dehydrogenase genes; the recombinant T-formaldehyde lyase gene; the recombinant lignostilbene dioxygenase gene; and the recombinant aromatic acid decarboxylase gene. In some versions, the recombinant microorganism comprises any three or more, any four or more, or each of: the one or more recombinant alcohol dehydrogenase genes; the one or more recombinant aldehyde dehydrogenase genes; the recombinant T-formaldehyde lyase gene; the recombinant lignostilbene dioxygenase gene; and the recombinant aromatic acid decarboxylase gene. In some versions, the recombinant microorganism comprises any four or more or each of: the one or more recombinant alcohol dehydrogenase genes; the one or more recombinant aldehyde dehydrogenase genes; the recombinant T-formaldehyde lyase gene; the recombinant lignostilbene dioxygenase gene; and the recombinant aromatic acid decarboxylase gene. In some versions, the recombinant microorganism comprises each of: the one or more recombinant alcohol dehydrogenase genes; the one or more recombinant aldehyde dehydrogenase genes; the recombinant T-formaldehyde lyase gene; the recombinant lignostilbene dioxygenase gene; and the recombinant aromatic acid decarboxylase gene.

[0009] In some versions, the one or more recombinant alcohol dehydrogenase genes encode FdhA of *Novosphingobium aromaticivorans* (SEQ ID NO:2) or a homolog thereof. In some versions, the one or more recombinant alcohol dehydrogenase genes encode FdhA of *Novosphingobium aromaticivorans* (SEQ ID NO:2), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:2, an ortholog of FdhA of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of FdhA of *Novosphingobium aromaticivorans*.

[0010] In some versions, the one or more recombinant alcohol dehydrogenase genes encode Saro_0995 of *Novosphingobium aromaticivorans* (SEQ ID NO:4) or a homolog thereof. In some versions, the one or more recombinant alcohol dehydrogenase genes encode Saro_0995 of *Novosphingobium aromaticivorans* (SEQ ID NO:4), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:4, an ortholog of Saro_0995 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_0995 of *Novosphingobium aromaticivorans*.

[0011] In some versions, the one or more recombinant alcohol dehydrogenase genes encode Saro_3899 of *Novosphingobium aromaticivorans* (SEQ ID NO:6) or a homolog thereof. In some versions, the one or more recombinant alcohol dehydrogenase genes encode Saro_3899 of *Novosphingobium aromaticivorans* (SEQ ID NO:6), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:6, an ortholog of Saro_3899 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_3899 of *Novosphingobium aromaticivorans*.

[0012] In some versions, the one or more recombinant aldehyde dehydrogenase genes encode FerD of *Novosphin-*

gobium aromaticivorans (SEQ ID NO:8) or a homolog thereof. In some versions, the one or more recombinant aldehyde dehydrogenase genes encode FerD of *Novosphingobium aromaticivorans* (SEQ ID NO:8), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:8, an ortholog of FerD of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of FerD of *Novosphingobium aromaticivorans*.

[0013] In some versions, the one or more recombinant aldehyde dehydrogenase genes encode Saro_1104 of *Novosphingobium aromaticivorans* (SEQ ID NO:10) or a homolog thereof. In some versions, the one or more recombinant aldehyde dehydrogenase genes encode Saro_1104 of *Novosphingobium aromaticivorans* (SEQ ID NO:10), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:10, an ortholog of Saro_1104 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_1104 of *Novosphingobium aromaticivorans*.

[0014] In some versions, the one or more recombinant aldehyde dehydrogenase genes encode Saro_1197 of *Novosphingobium aromaticivorans* (SEQ ID NO:12) or a homolog thereof. In some versions, the one or more recombinant aldehyde dehydrogenase genes encode Saro_1197 of *Novosphingobium aromaticivorans* (SEQ ID NO:12), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:12, an ortholog of Saro_1197 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_1197 of *Novosphingobium aromaticivorans*.

[0015] In some versions, the one or more recombinant aldehyde dehydrogenase genes encode Saro_2869 of *Novosphingobium aromaticivorans* (SEQ ID NO:14) or a homolog thereof. In some versions, the one or more recombinant aldehyde dehydrogenase genes encode Saro_2869 of *Novosphingobium aromaticivorans* (SEQ ID NO:14), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:14, an ortholog of Saro_2869 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_2869 of *Novosphingobium aromaticivorans*.

[0016] In some versions, the recombinant y-formaldehyde lyase gene encodes PcfL of *Novosphingobium aromaticivorans* (SEQ ID NO:16) or a homolog thereof. In some versions, the recombinant y-formaldehyde lyase gene encodes PcfL of *Novosphingobium aromaticivorans* (SEQ ID NO:16), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:16, an ortholog of PcfL of *Novosphingobium aromaticivorans*, a recombinant variant of the ortholog of PcfL of *Novosphingobium aromaticivorans*.

[0017] In some versions, the recombinant lignostilbene dioxygenase gene encodes LsdD of *Novosphingobium aromaticivorans* (SEQ ID NO:18) or a homolog thereof. In some versions, the recombinant lignostilbene dioxygenase gene encodes LsdD of *Novosphingobium aromaticivorans* (SEQ ID NO:18), a protein comprising a sequence at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, or at least 99% identical to SEQ ID NO:18, an ortholog of LsdD of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of LsdD of *Novosphingobium aromaticivorans*.

[0018] In some versions, the recombinant aromatic acid decarboxylase gene encodes LigW of *Novosphingobium aromaticivorans* (SEQ ID NO:20) or a homolog thereof. In some versions, the recombinant aromatic acid decarboxylase gene encodes LigW of *Novosphingobium aromaticivorans* (SEQ ID NO:20), a protein comprising a sequence at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, or at least 99% identical to SEQ ID NO:20, an ortholog of LigW of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of LigW of *Novosphingobium aromaticivorans*.

[0019] In some versions, the orthologs of FdhA, Saro_0995, Saro_3899, FerD, Saro_1104, Saro_1197, Saro_2869, PcfL, LsdD, and/or LigW are from a bacterium. In some versions, the orthologs of FdhA, Saro_0995, Saro_3899, FerD, Saro_1104, Saro_1197, Saro_2869, PcfL, LsdD, and/or LigW are from an Alphaproteobacterium. In some versions, the orthologs of FdhA, Saro_0995, Saro_3899, FerD, Saro_1104, Saro_1197, Saro_2869, PcfL, LsdD, and/or LigW are from an order selected from the group consisting of Sphingomonadales, *Actinomyces*, Gammaproteobacteria, Betaproteobacteria, and Bacilli. In some versions, the orthologs of FdhA, Saro_0995, Saro_3899, FerD, Saro_1104, Saro_1197, Saro_2869, PcfL, LsdD, and/or LigW are from the group consisting of *Novosphingobium*, Erythrobacteraceae, *Sphingobium*, and *Sphingomonas*.

[0020] In some versions, the recombinant microorganism is a bacterium. In some versions, the recombinant microorganism is an Alphaproteobacterium. In some versions, the recombinant microorganism is from an order selected from the group consisting of Sphingomonadales, *Actinomyces*, Gammaproteobacteria, Betaproteobacteria, and Bacilli. In some versions, the recombinant microorganism is from the group consisting of *Novosphingobium*, Erythrobacteraceae, *Sphingobium*, and *Sphingomonas*.

[0021] Another aspect of the invention is directed to methods of catabolizing a lignin aromatic. The methods can comprise culturing the recombinant microorganism of the invention in a medium comprising the lignin aromatic to thereby catabolize the lignin aromatic. In some versions, the lignin aromatic comprises a β-5 linked lignin aromatic. In some versions, the lignin aromatic comprises one or more of dehydrodiconiferyl alcohol (DC-A), dehydrodiconiferyl aldehyde (DC-L), dehydrodiconiferyl carboxylic acid (DC-C), dehydrodiconiferyl stilbene carboxylic acid (DC-S-C), 5-formyl ferulate (5-FF), 5-carboxyferulate (5-CF), and 4-hydroxyphenyl and syringyl analogs thereof.

[0022] The objects and advantages of the invention will appear more fully from the following detailed description of the preferred embodiment of the invention made in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0024] FIG. **1**. DC-A models β-5 linked lignin aromatics. A) Model lignin polymer that illustrates major interunit linkages and aromatic subunits. B) Structure of dehydrodiconiferyl alcohol (DC-A), a β-5 linked aromatic dimer composed of two G-family aromatic subunits. The β-5 bond is highlighted in red.

[0025] FIG. **2**. *N. aromaticivorans* funnels DC-A into central aromatic metabolism. A) Growth of WT *N. aromaticivorans* in SMB minimal medium with DC-A as the sole carbon source. B) Growth of 12444PDC in SMB minimal medium containing either DC-A plus glucose or glucose alone as carbon sources. C) Metabolite concentrations in extracellular medium of 12444PDC grown in SMB minimal medium with DC-A plus glucose as carbon sources. Error bars represent standard deviation across biological triplicates.

[0026] FIG. **3**. Genome-wide screens identify candidate genes for DC-A catabolism. A) Dot plot (log_2 scale) of RNA-Seq (y-axis) and RB-TnSeq (x-axis) data sets, with each dot representing a single gene. The horizontal and vertical red lines mark a 2-fold increase in transcript abundance when *N. aromaticivorans* PDC12444 is grown on DC-A compared to vanillin and a 2-fold abundance reduction of a disrupted gene when a *N. aromaticivorans* DSM12444 RB-TnSeq library is grown on DC-A compared to glucose, respectively. The five candidate genes investigated in this study are labeled in red. B) Genomic region containing four of the five candidate genes. Candidate genes are labeled in red. Experimentally determined transcription start sites (TSS) are labeled (34).

[0027] FIG. **4**. Proposed catabolic pathway for DC-A in *N. aromaticivorans*. The allylic alcohol side chain of DC-A is oxidized to DC-L and then to DC-C by dehydrogenases. The five-member ring of DC-C is opened by PcfL to form DC-S-C, which is then cleaved by LsdD into vanillin and 5-FF. 5-FF is oxidized to 5-CF by FerD and other dehydrogenases before it is decarboxylated by LigW to form ferulic acid. Metabolism of ferulic acid and vanillin to PDC by *N. aromaticivorans* has been previously described (10, 21). The gene products predicted to be involved in metabolism of formaldehyde following oxidation by FdhA are based on homology of *N. aromaticivorans* gene products with known S-glutathione hydrolases (Saro_2822) (35) and the subunits of a formate dehydrogenase complex (Saro_0732, Saro_0733, and Saro_0735) (36).

[0028] FIGS. **5**A-**5**C. PcfL converts DC-C to DC-S-C. FIG. **5**A) Metabolite concentrations in extracellular medium of 12444PDCΔpcfL grown in SMB minimal medium with DC-A plus glucose as carbon sources. Error bars represent standard deviation across biological triplicates. FIG. **5** B) Representative HPLC chromatograms of in vitro reactions containing DC-C and either control *E. coli* B834 cell extract or cell extract from *E. coli* B834 expressing recombinant PcfL. FIG. **5**C) Conversion of DC-C to DC-S-C by PcfL.

[0029] FIGS. **6**A-**6**C. LsdD cleaves DC-S-C to form 5-FF and vanillin. FIG. **6**A) Metabolite concentrations in extracellular medium of 12444PDCΔlsdD grown in SMB minimal medium with DC-A plus glucose as carbon sources. Error bars represent standard deviation across biological triplicates. FIG. **6**B) Representative HPLC chromatograms of in vitro reactions containing DC-S-C and either control *E. coli* cell extract or cell extract from *E. coli* expressing recombinant LsdD. FIG. **6**C) Cleavage of DC-S-C to 5-FF and vanillin by LsdD and abiotic dimerization of DC-S-C to DC-T-C.

[0030] FIGS. **7**A-**7**C. FerD and LigW convert 5-FF to 5-CF and then ferulic acid. FIG. **7**A) Metabolite concentrations in extracellular medium of 12444PDCΔferD and 12444PDCΔligW grown in SMB minimal medium with DC-A plus glucose as carbon sources. Error bars represent standard deviation across biological triplicates. FIG. **7**B) Representative HPLC chromatograms of in vitro reactions (left) containing 5-FF plus NAD^+ and either control *E. coli* B834 cell extract or cell extract of *E. coli* B834 expressing recombinant FerD or reactions (right) containing 5-CF and either control *E. coli* B834 cell extract or cell extract of *E. coli* B834 expressing recombinant LigW. FIG. **7**C) Oxidation of 5-FF to 5-CF by FerD and decarboxylation of 5-CF to ferulic acid by LigW.

[0031] FIG. **8**. Multiple partially redundant ADHs and ALDHs can oxidize the allylic side chain of DC-A. Concentration of DC-L over the course of 1 hour long in vitro assays containing A) DC-A, NAD^+, and a control *E. coli* B834 cell extract or cell extracts of *E. coli* B834 expressing recombinant candidate ADHs or B) DC-L, NAD^+, and control *E. coli* B834 cell extract or cell extracts of *E. coli* B834 expressing recombinant candidate ALDHs. For clarity of presentation, only dehydrogenases exhibiting activity on the tested substrates are shown. Error bars represent standard deviation across triplicates.

[0032] FIG. **9**. The proposed catabolic pathway enzymes can convert DC-A to ferulic acid and vanillic acid in vitro. Representative HPLC chromatograms of in vitro reactions containing DC-A plus NAD^+ and either control *E. coli* B834 cell extract or cell extracts from *E. coli* B834 expressing recombinant Saro_0995, PcfL, LsdD, FerD, and LigW.

[0033] FIGS. **10**A-**10**G. Order Sphingomonadales contains two pathways for conversion of DC-C to DC-S-C and a conserved pathway for DC-S-C catabolism. Phylogeny constructed based on the bacterial reference genes of Alphaproteobacteria containing homologs (>50% amino acid identity, >70% query coverage) of at least two enzymes found in the β-5 linked aromatic catabolic pathways characterized in *N. aromaticivorans* or *Sphingobium* sp. SYK-6. Homologs found in each species are marked by colored boxes. Clades are labeled and color-coded. The scale bar indicates the number of nucleotide substitutions per sequence site. The gap in the outgroup corresponds to 1.5 on the scale bar. A simplified diagram of the DC-A catabolic pathways in N. *aromaticivorans* and *Sphingobium* sp. SYK-6 is shown. Phylogeny presented in FIG. **10**A represents the bacteria from left to right as they appear in the order in which they appear in FIGS. **10**B-**10**G.

[0034] FIG. **11** Trace amounts of DC-L transiently accumulate during DC-A catabolism. DC-L concentration in extracellular medium of 12444PDC grown in SMB minimal medium with DC-A plus glucose as carbon sources. Error bars represent standard deviation across biological triplicates.

[0035] FIG. **12**. Genome-wide screens identify candidate genes for DC-A catabolism. Dot plot (log_2 scale) of RNA-Seq (y-axis) and RB-TnSeq (x-axis) data sets, with each dot representing a single gene. The horizontal and vertical red lines mark a 2-fold increase in transcript abundance when *N. aromaticivorans* PDC12444 is grown on DC-A compared to A) glucose or B) ferulic acid and a 2-fold abundance reduction of a disrupted gene when a *N. aromaticivorans* DSM12444 RB-TnSeq library is grown on DC-A compared to glucose, respectively. The five candidate genes investigated in this study are labeled in red.

[0036] FIG. **13**. Formaldehyde is released when PcfL converts DC-C to DC-S-C. Concentration of formaldehyde after 6 hours of incubating in vitro reactions containing DC-C and either cell extract of *E. coli* B834 expressing

recombinant PcfL or control *E. coli* B834 cell extract. Error bars represent standard deviation across triplicates.

[0037] FIG. **14**. FdhA acts on formaldehyde released during DC-A catabolism. A) Metabolite concentrations in extracellular medium of 12444PDCΔfdhA grown in SMB minimal medium with DC-A plus glucose as carbon sources. B) Formaldehyde concentration in extracellular medium of 12444PDC or 12444PDCΔfdhA grown in SMB minimal medium with DC-A plus glucose as carbon sources. Error bars represent standard deviation across biological triplicates.

[0038] FIGS. **15**A and **15**B. DC-S-C abiotically homodimerizes in aqueous solutions to form DC-T-C. FIG. **15**A) ^{13}C NMR spectrum of the product obtained when DC-S-C is incubated in SMB minimal medium supplemented with 1 g/L glucose. The structure of the resulting compound, DC-T-C, is shown. FIG. **15**B) Loss of DC-S-C over time in various solutions. Note that some DC-S-C visually precipitated in the water condition. Error bars represent standard deviation across triplicates.

[0039] FIGS. **16**A and **16**B. FerD is an NAD^+-dependent aldehyde dehydrogenase. FIG. **16**A) Representative HPLC chromatograms of in vitro reactions containing 5-FF and either control *E. coli* B834 cell extract or cell extract of *E. coli* B834 expressing recombinant FerD without added NAD^+. FIG. **16**B) Ratio of NAD^+ to NADH after 6 hours incubating in vitro reactions containing 5-FF and NAD^+ along with purified FerD, cell extract of *E. coli* B834 expressing recombinant FerD, or control *E. coli* B834 cell extract. Error bars represent standard deviation across triplicates.

[0040] FIG. **17**. Differences in DC-A, DC-L, and DC-C absorbance can be leveraged in colorimetric assays. UV-Vis traces of 0.2 mM solutions of DC-A, DC-L, and DC-C in S30 buffer.

[0041] FIG. **18**. FerD converts vanillin to vanillic acid. Representative HPLC chromatograms of in vitro reactions containing vanillin and either control *E. coli* B834 cell extract or cell extract of *E. coli* B834 expressing recombinant FerD.

[0042] FIGS. **19**A-**19**C. PcfL exhibits activity on DC-A and DC-L in vitro. Representative HPLC chromatograms of in vitro reactions containing DC-A (FIG. **19**A) or DC-L (FIG. **19**B) and either control *E. coli* B834 cell extract or cell extract of *E. coli* B834 expressing recombinant PcfL. FIG. **19**C) Structures of proposed stilbene compounds based on m/z of the in vitro reaction products.

[0043] FIG. **20**. Proposed *N. aromaticivorans* catabolic pathway for DC-A, accounting for the ability of PcfL to act on DC-A, DC-L, and DC-C. The allylic alcohol is oxidized to an aldehyde and then to a carboxylic acid by dehydrogenases. The five-member ring of DC-C is opened by PcfL to form DC-S-C, which is then cleaved by LsdD into vanillin and 5-FF. 5-FF is oxidized to 5-CF by FerD and other dehydrogenases before it is decarboxylated by LigW to form ferulic acid. Metabolism of ferulic acid and vanillin to PDC by *N. aromaticivorans* has been previously described (10, 21). The gene products involved in metabolism of formaldehyde following oxidation by FdhA represent a hypothetical pathway based on homology with known S-glutathione hydrolases (Saro_2822) (35) and the subunits of a formate dehydrogenase complex (Saro_0732, Saro_0733, and Saro_0735) (36). Steps that differ from those proposed in FIG. **4** are marked with blue arrows.

[0044] FIGS. **21**A-**21**C. The full *N. aromaticivorans* DC-A catabolic pathway is exclusive to Alphaproteobacteria. Phylogeny constructed based on the bacterial reference genes of bacteria containing homologs (>50% amino acid identity, >70% query coverage) of at least two enzymes found in the *N. aromaticivorans* β-5 linked aromatic pathway. The bacterial species are sorted by class. The colored bars to the right of the tree indicate the proportion of each class containing a homolog of each enzyme. The scale bar indicates the number of nucleotide substitutions per sequence site. A simplified diagram of the DC-A catabolic pathway in *N. aromaticivorans* is shown in FIG. **21**A. FIGS. **21**B and **21**C show a closeups of FIG. **21**A with relevant percentages.

[0045] FIGS. **22**A-**22**E. DC-A, DC-L, DC-C, and DC-S-C synthesis. FIG. **22**A) Synthetic routes to DC-A, DC-L, DC-C, and DC-S-C. FIGS. **22**B-**22**E) ^{13}C NMR (acetone-d_6) spectra and structures of synthetic DC-A (FIG. **22**B), DC-L (FIG. **22**C), DC-C (FIG. **22**D), and DC-S-C (FIG. **22**E).

[0046] FIGS. **23**A-**23**C. DC-S-C and DC-T-C synthesis. FIG. A) Synthetic routes to 5-FF and 5-CF. B-C) ^{13}C NMR (acetone-d_6) spectra and structures of synthetic FIG. B) 5-FF and FIG. C) 5-CF.

[0047] FIG. **24**. Growth of 12444PDC and 12444PDC mutant strains. Growth curves of 12444PDC and 12444PDC mutant strains in SMB minimal medium containing 0.5 mM DC-A and 1 g/L glucose as carbon sources. Error bars represent standard deviation across biological triplicates.

[0048] FIG. **25**. Solvent B (MeOH) percent protocol for HPLC method. Trace of percent solvent B over time. Solvent A was 0.2% formic acid in water.

[0049] FIG. **26**. Differences in DC-S-C and DC-T-C can be leveraged in colorimetric assays. UV-Vis traces of 0.2 mM solutions of DC-S-C and DC-T-C in S30 buffer.

DETAILED DESCRIPTION OF THE INVENTION

[0050] The recombinant microorganisms of the invention can comprise one or more recombinant genes. The recombinant genes can comprise one or more recombinant alcohol dehydrogenase genes, one or more recombinant aldehyde dehydrogenase genes, a recombinant 7-formaldehyde lyase gene, a recombinant lignostilbene dioxygenase gene, and/or a recombinant aromatic acid decarboxylase gene.

[0051] The recombinant alcohol dehydrogenase genes of the invention are preferably capable of catalyzing the conversion of dehydrodiconiferyl alcohol (DC-A) to dehydrodiconiferyl aldehyde (DC-L). See, e.g., FIG. **4**. The recombinant alcohol dehydrogenase genes of the invention may also be capable of catalyzing the conversion of phenolic analogs (such as 4-hydroxyphenyl or syringyl analogs) of dehydrodiconiferyl alcohol (DC-A) (a guaiacyl aromatic) to phenolic analogs (such as 4-hydroxyphenyl or syringyl analogs) of dehydrodiconiferyl aldehyde (DC-L) (a guaiacyl aromatic). Exemplary recombinant alcohol dehydrogenase genes include those encoding FdhA of *Novosphingobium aromaticivorans* (Saro_0874) (SEQ ID NO:2 (exemplary coding sequence is SEQ ID NO:1)) or a homolog thereof, Saro_0995 of *Novosphingobium aromaticivorans* (SEQ ID NO:4 (exemplary coding sequence is SEQ ID NO:3)) or a homolog thereof, and Saro_3899 of *Novosphingobium aromaticivorans* (SEQ ID NO:6 (exemplary coding sequence is SEQ ID NO:5)) or a homolog thereof. The homolog of FdhA can comprise a protein comprising a sequence at least 80%,

at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:2, an ortholog of FdhA, or a recombinant variant of the ortholog of FdhA. The homolog of Saro_0995 can comprise a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:4, an ortholog of Saro_0995, or a recombinant variant of the ortholog of Saro_0995. The homolog of Saro_3899 can comprise a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:6, an ortholog of Saro_3899, or a recombinant variant of the ortholog of Saro_3899.

[0052] The recombinant aldehyde dehydrogenase genes of the invention are preferably capable of catalyzing the conversion of dehydrodiconiferyl aldehyde (DC-L) (a guaiacyl aromatic) or a 4-hydroxyphenyl or syringyl analog thereof to dehydrodiconiferyl carboxylic acid (DC-C) (a guaiacyl aromatic) or a 4-hydroxyphenyl or syringyl analog thereof. See, e.g., FIG. **4**. The recombinant aldehyde dehydrogenase genes of the invention may also be capable of catalyzing the conversion of phenolic analogs (such as 4-hydroxyphenyl or syringyl analogs) of dehydrodiconiferyl aldehyde (DC-L) (a guaiacyl aromatic) to phenolic analogs (such as 4-hydroxyphenyl or syringyl analogs) of dehydrodiconiferyl carboxylic acid (DC-C) (a guaiacyl aromatic). Exemplary recombinant aldehyde dehydrogenase genes include those encoding FerD of *Novosphingobium aromaticivorans* (Saro_0797) (SEQ ID NO:8 (exemplary coding sequence is SEQ ID NO:7)) or a homolog thereof, Saro_1104 of *Novosphingobium aromaticivorans* (SEQ ID NO:10 (exemplary coding sequence is SEQ ID NO:9)) or a homolog thereof, Saro_1197 of *Novosphingobium aromaticivorans* (SEQ ID NO:12 (exemplary coding sequence is SEQ ID NO:11)) or a homolog thereof, and Saro_2869 of *Novosphingobium aromaticivorans* (SEQ ID NO:14 (exemplary coding sequence is SEQ ID NO:13)) or a homolog thereof. The homolog of FerD can comprise a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:8, an ortholog of FerD, or a recombinant variant of the ortholog of FerD. The homolog of Saro_1104 can comprise a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:10, an ortholog of Saro_1104, or a recombinant variant of the ortholog of Saro_1104. The homolog of Saro_1197 can comprise a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:12, an ortholog of Saro_1197, or a recombinant variant of the ortholog of Saro_1197. The homolog of Saro_2869 can comprise a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:14, an ortholog of Saro_2869, or a recombinant variant of the ortholog of Saro_2869. The FerD of *Novosphingobium aromaticivorans* (Saro_0797) can also convert 5-formyl ferulate (5-FF) to 5-carboxyferulate (5-CF) and vanillin to vanillic acid.

[0053] The recombinant γ-formaldehyde lyase genes of the invention are preferably capable of catalyzing the conversion of dehydrodiconiferyl carboxylic acid (DC-C) to dehydrodiconiferyl stilbene carboxylic acid (DC-S-C). See, e.g., FIG. **4**. The recombinant γ-formaldehyde lyase genes of the invention may also be capable of catalyzing the conversion of phenolic analogs (such as 4-hydroxyphenyl or syringyl analogs) of dehydrodiconiferyl carboxylic acid (DC-C) (a guaiacyl aromatic) to phenolic analogs (such as 4-hydroxyphenyl or syringyl analogs) of dehydrodiconiferyl stilbene carboxylic acid (DC-S-C) (a guaiacyl aromatic). Exemplary recombinant aldehyde dehydrogenase genes include those encoding PcfL of *Novosphingobium aromaticivorans* (Saro_0796) (SEQ ID NO:16 (exemplary coding sequence is SEQ ID NO:15)) or a homolog thereof. The homolog of PcfL can comprise a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:16, an ortholog of PcfL, a recombinant variant of the ortholog of PcfL.

[0054] The recombinant lignostilbene dioxygenase genes of the invention are preferably capable of catalyzing the conversion of dehydrodiconiferyl stilbene carboxylic acid (DC-S-C) to 5-formyl ferulate (5-FF) and/or vanillin. See, e.g., FIG. **4**. The recombinant lignostilbene dioxygenase genes of the invention may also be capable of catalyzing the conversion of phenolic analogs (such as a 4-hydroxyphenyl analog) of dehydrodiconiferyl stilbene carboxylic acid (DC-S-C) to phenolic analogs (such as a 4-hydroxyphenyl analog) of dehydrodiconiferyl stilbene carboxylic acid (DC-S-C) (a guaiacyl aromatic). Exemplary recombinant lignostilbene dioxygenase genes include those encoding LsdD of *Novosphingobium aromaticivorans* (Saro_0802) (SEQ ID NO:18 (exemplary coding sequence is SEQ ID NO:17)) or a homolog thereof. The homolog of LsdD can comprise a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:18, an ortholog of LsdD, a recombinant variant of the ortholog of LsdD.

[0055] The recombinant aromatic acid decarboxylase genes of the invention are preferably capable of catalyzing the conversion of 5-carboxyferulate (5-CF) to ferulic acid. See, e.g., FIG. **4**. The recombinant aromatic acid decarboxylase genes of the invention may also be capable of catalyzing the conversion of phenolic analogs (such as a 4-hydroxyphenyl analog) of 5-carboxyferulate (5-CF) to phenolic analogs (such as a 4-hydroxyphenyl analog) of ferulic acid. Exemplary recombinant aromatic acid decarboxylase genes include those encoding LigW of *Novosphingobium aromaticivorans* (Saro_0799) (SEQ ID NO:20 (exemplary coding sequence is SEQ ID NO:19)) or a homolog thereof. The homolog of LigW can comprise a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:20, an ortholog of LigW, a recombinant variant of the ortholog of LigW.

[0056] The recombinant genes of the invention can be configured to be expressed or overexpressed in the microorganism. If a microorganism endogenously comprises a particular gene, the gene may be modified to exchange or optimize promoters, exchange or optimize enhancers, or exchange or optimize any other genetic element to result in increased expression of the gene. Alternatively, one or more additional copies of the gene or coding sequence thereof may be introduced to the cell for enhanced expression of the gene product. If a microorganism does not endogenously comprise a particular gene, the gene or coding sequence thereof may be introduced to the microorganism for heterologous expression of the gene product. The gene or coding sequence may be incorporated into the genome of the microorganism or may be contained on an extra-chromosomal plasmid. The gene or coding sequence may be introduced to the microorganism individually or may be included on an operon. Techniques for genetic manipulation are described in further detail below.

[0057] The recombinant microorganisms of the invention may be genetically altered to express or overexpress any of the specific genes or gene products explicitly described herein or homologs thereof. Proteins and/or protein sequences are "homologous" when they are derived, naturally or artificially, from a common ancestral protein or protein sequence. Similarly, nucleic acids and/or nucleic acid sequences are homologous when they are derived, naturally or artificially, from a common ancestral nucleic acid or nucleic acid sequence. Nucleic acid or gene product (amino acid) sequences of any known gene, including the genes or gene products described herein, can be determined by searching any sequence databases known in the art using the gene name or accession number as a search term. Common sequence databases include GenBank (www.ncbi.nlm.nih.gov), ExPASy (expasy.org), KEGG (www.genome.jp), among others. Homology is generally inferred from sequence similarity between two or more nucleic acids or proteins (or sequences thereof). The precise percentage of similarity between sequences that is useful in establishing homology varies with the nucleic acid and protein at issue, but as little as 25% sequence similarity (e.g., identity) over 50, 100, 150 or more residues (nucleotides or amino acids) is routinely used to establish homology (e.g., over the full length of the two sequences to be compared). Higher levels of sequence similarity (e.g., identity), e.g., 30%, 35% 40%, 45% 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% or more, can also be used to establish homology. Accordingly, homologs of the genes or gene products described herein include genes or gene products having at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to the genes or gene products described herein. Methods for determining sequence similarity percentages (e.g., BLASTP and BLASTN using default parameters) are described herein and are generally available. The homologous proteins should demonstrate comparable activities and, if an enzyme, participate in the same or analogous pathways. Homologs include orthologs and paralogs. "Orthologs" are genes and products thereof in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same or similar function in the course of evolution. Paralogs are genes and products thereof related by duplication within a genome. As used herein, "orthologs" and "paralogs" are included in the term "homologs."

[0058] For sequence comparison and homology determination, one sequence typically acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence based on the designated program parameters. A typical reference sequence of the invention is a nucleic acid or amino acid sequence corresponding to the genes or gene products described herein.

[0059] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see Current Protocols in Molecular Biology, F. M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2008)).

[0060] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity for purposes of defining homologs is the BLAST algorithm, which is described in Altschul et al., J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=−4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

[0061] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. The above-described techniques are useful in identifying homologous sequences for use in the methods described herein.

[0062] The terms "identical" or "percent identity", in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when com-

pared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described above (or other algorithms available to persons of skill) or by visual inspection.

[0063] The phrase "substantially identical" in the context of two nucleic acids or polypeptides refers to two or more sequences or subsequences that have at least about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90, about 95%, about 98%, or about 99% or more nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm or by visual inspection. Such "substantially identical" sequences are typically considered to be "homologous," without reference to actual ancestry. Preferably, the "substantial identity" exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably, the sequences are substantially identical over at least about 150 residues, at least about 250 residues, or over the full length of the two sequences to be compared.

[0064] Derived: When used with reference to a nucleic acid or protein, "derived" means that the nucleic acid or polypeptide is isolated from a described source or is at least 70%, 80%, 90%, 95%, 99%, or more identical to a nucleic acid or polypeptide included in the described source.

[0065] Endogenous: As used herein with reference to a nucleic acid molecule, genetic element (e.g., gene, promoter, etc.), or polypeptide in a particular cell, "endogenous" refers to a nucleic acid molecule, genetic element, or polypeptide that is in the cell and was not introduced into the cell or transferred within the genome of the cell using recombinant engineering techniques. For example, an endogenous genetic element is a genetic element that was present in a cell in its particular locus in the genome when the cell was originally isolated from nature.

[0066] Exogenous: As used herein with reference to a nucleic acid molecule, genetic element (e.g., gene, promoter, etc.), or polypeptide in a particular cell, "exogenous" refers to any nucleic acid molecule, genetic element, or polypeptide that was introduced into the cell or transferred within the genome of the cell using recombinant engineering techniques. For example, an exogenous genetic element is a genetic element that was not present in its particular locus in the genome when the cell was originally isolated from nature.

[0067] Expression: The process by which a gene's coded information is converted into the structures and functions of a cell, such as a protein, transfer RNA, or ribosomal RNA. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (for example, transfer and ribosomal RNAs).

[0068] Introduce: When used with reference to genetic material, such as a nucleic acid, and a cell, "introduce" refers to the delivery of the genetic material to the cell in a manner such that the genetic material is capable of being expressed within the cell. Introduction of genetic material includes both transformation and transfection. Transformation encompasses techniques by which a nucleic acid molecule can be introduced into cells such as prokaryotic cells or non-animal eukaryotic cells. Transfection encompasses techniques by which a nucleic acid molecule can be introduced into cells such as animal cells. These techniques include but are not limited to introduction of a nucleic acid via conjugation, electroporation, lipofection, infection, and particle gun acceleration.

[0069] Isolated: An "isolated" biological component (such as a nucleic acid molecule, polypeptide, or cell) has been substantially separated or purified away from other biological components in which the component naturally occurs, such as other chromosomal and extrachromosomal DNA and RNA and proteins. Nucleic acid molecules and polypeptides that have been "isolated" include nucleic acid molecules and polypeptides purified by standard purification methods. The term also includes nucleic acid molecules and polypeptides prepared by recombinant expression in a cell as well as chemically synthesized nucleic acid molecules and polypeptides. In one example, "isolated" refers to a naturally occurring nucleic acid molecule that is not immediately contiguous with both of the sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally-occurring genome of the organism from which it is derived.

[0070] Gene: Genes minmally include a promoter operationally linked to a coding sequence, and can include other elements that facilitate or regulate the transcription and/or translation of the coding sequence.

[0071] Heterologous: The term "heterologous" refers to an element in an arrangement with another element that does not occur in nature. For example, a gene or protein that is heterologous to a given cell is a gene or protein that does not occur in the cell in nature. A promoter that is heterologous to a given coding sequence is a promoter that is not operably linked to the coding sequence in nature.

[0072] Nucleic acid: Encompasses both RNA and DNA molecules including, without limitation, cDNA, genomic DNA, and mRNA. Nucleic acids also include synthetic nucleic acid molecules, such as those that are chemically synthesized or recombinantly produced. The nucleic acid can be double-stranded or single-stranded. Where single-stranded, the nucleic acid molecule can be the sense strand, the antisense strand, or both. In addition, the nucleic acid can be circular or linear.

[0073] Operably linked: A first element is operably linked with a second element when the first element is placed in a functional relationship with the second element. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. A secretion signal sequence is operably linked to a protein (such as an enzyme) when the secretion signal sequence affects secretion of the protein from a cell.

[0074] Overexpress: When a gene is caused to be transcribed at an elevated rate compared to the endogenous or basal transcription rate for that gene. In some examples, overexpression additionally includes an elevated rate of translation of the gene compared to the endogenous translation rate for that gene. Methods of testing for overexpression are well known in the art, for example transcribed RNA levels can be assessed using RT-PCR and protein levels can be assessed using SDS-PAGE gel analysis.

[0075] Recombinant: A recombinant nucleic acid or polypeptide is one comprising a sequence that is not naturally occurring. A recombinant gene is a gene that comprises a recombinant nucleic acid sequence, is present within a cell in which it does not naturally occur, and/or is present in a different locus (e.g., genetic locus or on an extrachromosomal plasmid) within a particular cell than in a correspond-

ing native cell. A recombinant cell (such as a recombinant microorganism) is one that comprises a recombinant nucleic acid, a recombinant gene, or a recombinant polypeptide. An example of a recombinant gene is a gene that has a coding sequence operably linked to a heterologous promoter.

[0076] Recombinant variant: Used with reference to an ortholog, "recombinant variant" refers to a variant of the ortholog that comprises one or more modifications to amino acid sequence of the ortholog. Exemplary modifications include substitutions, deletions, and insertions. The recombinant variant preferably comprises an amino acid sequence at least 95% identical to the amino acid sequence of the ortholog.

[0077] Another aspect of the invention is directed to methods of catabolizing a lignin aromatic. The methods can comprise culturing the recombinant microorganism of the invention in a medium comprising the lignin aromatic to thereby catabolize the lignin aromatic.

[0078] "Lignin aromatic" as used herein refers to an aromatic present in or derived from lignin. The lignin aromatics can be a monomer, a dimer, an oligomer, or a polymer. The lignin aromatics can comprise syringyl aromatics, guaiacyl aromatics, p-hydroxyphenyl aromatics, or any combinations thereof. Syringyl, guaiacyl, and p-hydroxyphenyl aromatics differ in their degree of methoxilation of the aromatic ring. Syringyl aromatics comprise methoxy groups at the 3 and 5 positions of the aromatic ring. Guaiacyl aromatics comprise a methoxy group on only one of the 3 and 5 positions on the aromatic ring. p-Hydroxyphenyl aromatics are devoid of methoxy groups on either of the 3 and 5 positions of the aromatic ring.

[0079] In some versions, the lignin aromatic comprises a β-5 linked lignin aromatic. β-5 linked lignin aromatics include lignin aromatics that comprise at least one β-5 linkage.

[0080] In some versions, the lignin aromatic comprises one or more of dehydrodiconiferyl alcohol (DC-A), dehydrodiconiferyl aldehyde (DC-L), dehydrodiconiferyl carboxylic acid (DC-C), dehydrodiconiferyl stilbene carboxylic acid (DC-S-C), 5-formyl ferulate (5-FF), 5-carboxyferulate (5-CF) or a 4-hydroxyphenyl or syringyl analog thereof. The 4-hydroxyphenyl or syringyl analogs of these compounds lack methoxy groups at both of the 3 and 5 positions of the aromatic ring or comprise methoxy groups at both of the 3 and 5 positions of the aromatic ring, respectively.

[0081] In some versions, the lignin aromatic can be derived from (and optionally isolated from) and/or provided in the form of depolymerized lignin, such as chemically depolymerized lignin. Methods of depolymerizing lignin are well known in the art. See Pandey et al. 2010 (Pandey M P, Kim C S. Lignin Depolymerization and Conversion: A Review of Thermochemical Methods. *Chemical & Engineering Technology,* 2010, Vol. 34, Issue 1, pp. 3-145) and Wang et al. 2013 (Wang H, Tucker M, Ji Y. Recent Development in Chemical Depolymerization of Lignin: A Review. *Journal of Applied Chemistry,* 2013, Volume 2013, Article ID 838645).

[0082] The depolymerized lignin can be derived from pretreated lignocellulosic biomass. Methods of pretreating lignocellulosic biomass are well known in the art. See Kumar et al. 2017 (Kumar A K and Sharma S. Recent Updates on Different Methods of Pretreatment of Lignocellulosic Feedstocks: A Review. Bioresour. Bioprocess. (2017) 4:7); Kumar et al. 2009 (Kumar, P.; Barrett, D. M.; Delwiche, M. J.; Stroeve, P., Methods for Pretreatment of lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. Industrial & *Engineering Chemistry Research* 2009, 48, (8), 3713-3729); Wang et al. 2013 (Wang H, Tucker M, Ji Y. Recent Development in Chemical Depolymerization of Lignin: A Review. (2013) *Journal of Applied Chemistry.* 2013:1-9), and Karlen et al. 2020 (Karlen S D, Fasahati P, Mazaheri M, Serate J, Smith R A, Sirobhushanam S, Chen M, Tymkhin V I, Cass C L, Liu S, Padmakshan D, Xie D, Zhang Y, McGee M A, Russell J D, Coon J J, Kaeppler H F, de Leon N, Maravelias C T, Runge T M, Kaeppler S M, Sedbrook J C, Ralph J. Assessing the viability of recovering hydroxycinnamic acids from lignocellulosic biorefinery alkaline pretreatment waste streams. *ChemSusChem.* 2020 Jan. 26). Examples include chipping, grinding, milling, steam pretreatment, ammonia fiber expansion (AFEX, also referred to as ammonia fiber explosion), ammonia recycle percolation (ARP), CO_2 explosion, steam explosion, ozonolysis, wet oxidation, acid hydrolysis, dilute-acid hydrolysis, alkaline hydrolysis, organosolv, ionic liquids, gamma-valerolactone, enzymatic pretreatment, biological pretreatment, and pulsed electrical field treatment, among others.

[0083] The lignocellulosic biomass can be derived from any source, such as corn cobs, corn stover, cotton seed hairs, grasses, hardwood stems, leaves, newspaper, nut shells, paper, softwood stems, sorghum, switchgrass, waste papers from chemical pulps, wheat straw, wood, woody residues, mixed biomass species such as those produced by native prairie, and other sources. Sources that maintain β-5 bonds in lignin are preferred.

[0084] It is noted that the aromatic analogs of the compounds described herein will have modifications to aromatic groups only at positions on the aromatic groups where they are chemically possible. For example, only one of the two aromatic groups in DC-A, DC-L, DC-C, and DC-S-C permit the presence of syringyl analogs due to the β-5 bonds or other bonding at the relevant position on the aromatic ring. Similarly, 5-FF and 5-CF do not permit the presence of syringyl analogs due to the presence of the aldehyde and carboxy groups, respectively, at the relevant position on the aromatic ring. Mixed type β-5 aromatics (e.g., those containing one syringyl type aromatic and one 4-hydroxyphenyl type aromatic) are contemplated as examples of aromatic analogs of the compounds herein.

[0085] Unless explained otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below.

[0086] The elements and method steps described herein can be used in any combination whether explicitly described or not.

[0087] All combinations of method steps as used herein can be performed in any order, unless otherwise specified or clearly implied to the contrary by the context in which the referenced combination is made.

[0088] As used herein, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise.

[0089] Numerical ranges as used herein are intended to include every number and subset of numbers contained

within that range, whether specifically disclosed or not. Further, these numerical ranges should be construed as providing support for a claim directed to any number or subset of numbers in that range. For example, a disclosure of from 1 to 10 should be construed as supporting a range of from 2 to 8, from 3 to 7, from 5 to 6, from 1 to 9, from 3.6 to 4.6, from 3.5 to 9.9, and so forth.

[0090] All patents, patent publications, and peer-reviewed publications (i.e., "references") cited herein are expressly incorporated by reference to the same extent as if each individual reference were specifically and individually indicated as being incorporated by reference. In case of conflict between the present disclosure and the incorporated references, the present disclosure controls.

[0091] It is understood that the invention is not confined to the particular construction and arrangement of parts herein illustrated and described, but embraces such modified forms thereof as come within the scope of the claims.

Examples

Catabolism of β-5 Linked Aromatics by *Novosphingobium aromaticivorans*

Summary

[0092] Aromatic compounds are an important source of commodity chemicals traditionally produced from fossil fuels. Aromatics derived from plant lignin can potentially be converted into commodity chemicals through depolymerization followed by microbial funneling of monomers and low molecular weight oligomers. This study investigates the catabolism of the β-5 linked aromatic dimer dehydrodiconiferyl alcohol (DC-A) by the bacterium *Novosphingobium aromaticivorans*. We used genome-wide screens to identify candidate genes involved in DC-A catabolism. Subsequent in vivo and in vitro analyses of these candidate genes elucidated a catabolic pathway composed of four required gene products and several partially redundant dehydrogenases that convert DC-A to aromatic monomers that can be funneled into the central aromatic metabolic pathway of *N. aromaticivorans*. Specifically, a newly identified γ-formaldehyde lyase, PcfL, opens the phenylcoumaran ring to form a stilbene and formaldehyde. A lignostilbene dioxygenase, LsdD, then cleaves the stilbene to generate the aromatic monomers vanillin and 5-formylferulate (5-FF). We also show that the aldehyde dehydrogenase FerD oxidizes 5-FF before it is decarboxylated by LigW, yielding ferulic acid. We found that some enzymes involved in the β-5 catabolism pathway can act on multiple substrates and that some steps in the pathway can be mediated by multiple enzymes, providing new insights into the robust flexibility of aromatic catabolism in N. *aromaticivorans*. A comparative genomic analysis predicted that the newly discovered β-5 aromatic catabolic pathway is common within the order Sphingomonadales.

[0093] In the transition to a circular bioeconomy, the plant polymer lignin holds promise as a renewable source of industrially important aromatic chemicals. However, since lignin contains aromatic subunits joined by various chemical linkages, producing single chemical products from this polymer can be challenging. One strategy to overcome this challenge is using microbes to funnel a mixture of lignin-derived aromatics into target chemical products. This approach requires strategies to cleave the major inter-unit linkages of lignin to release monomers for funneling into valuable products. In this study, we report newly discovered aspects of a pathway by which the *Novosphingobium aromaticivorans* DSM12444 catabolizes aromatics joined by the second most common inter-unit linkage in lignin, the β-5 linkage. This work advances our knowledge of aromatic catabolic pathways, laying the groundwork for future metabolic engineering of this and other microbes for optimized conversion of lignin into products.

Introduction

[0094] *Novosphingobium aromaticivorans* DSM12444 is an Alphaproteobacterium with properties that make it a potential microbial chassis for lignin valorization. *N. aromaticivorans* can metabolize a variety of natural and chemically modified aromatic monomers and oligomers and it can co-metabolize aromatic compounds with other carbon sources (13, 14). Additionally, native metabolic pathways enable engineered strains of this bacterium to funnel the products of depolymerized lignin into commodity chemicals such as 2-pyrone-4,6-dicarboxylic acid (PDC) (10, 15), cis-cis-muconic acid (16), and carotenoids (17). This study uses a previously engineered strain of *N. aromaticivorans* (12444PDC), in which ligI, desC, and desD have been deleted so that it converts S-, G- and H-aromatics into PDC (10), which is a potential platform chemical for industrial valorization (18, 19).

[0095] While metabolic pathways by which *N. aromaticivorans* funnels aromatic monomers into central aromatic metabolism have been characterized (10, 20, 21), less is known about how it catabolizes aromatics joined by the various interunit bonds present in lignin. To date, only the pathways for catabolism of the most abundant interunit bond, the 3-O-4 linkage (22, 23), as well as the R-1 linkage (24) have been elucidated in *N. aromaticivorans*. Catabolic pathways for aromatic oligomers containing other abundant interunit linkages have been reported in some organisms, but knowledge gaps remain in the pathways used by this bacterium.

[0096] This work sought to investigate the ability of *N. aromaticivorans* to catabolize β-5 (phenylcoumaran) linked aromatics. β-5 linked aromatics represent the second most abundant interunit linkage in lignin, accounting for up to 12% of the total interunit bonds depending on the biomass source (25, 26). The only pathway for the catabolism of β-5 linked aromatics has been proposed in *Sphingomonas paucimobilis* TMY10009 (27) and characterized in *Sphingobium* sp. SYK-6 (28-32), while one enzyme with activity on β-5 linked aromatics has been identified in *Agrobacterium* sp. (33). However, there are reports of significant differences in either the ability to catabolize aromatic compounds or the enzymes involved in the catabolic pathways of members of the order Sphingomonadales (11, 12, 20). Thus, it is important to identify similarities and differences in aromatic catabolism among different bacteria when developing strategies to valorize lignin.

[0097] The goal of this study was to determine if and how *N. aromaticivorans* catabolizes aromatics joined by a β-5 linkage. To do this, we synthesized dehydrodiconiferyl alcohol (DC-A), a dimer composed of two G-aromatic monomers connected by a β-5 interunit linkage (FIG. **1** (B)). We found that *N. aromaticivorans* can grow on DC-A and funnel it through its central aromatic metabolism. We combined data from two genome-wide screens to identify candidate

genes involved in DC-A catabolism, followed by in vivo analysis of defined mutants and in vitro enzyme activity assays to test the roles of candidate genes and proteins in catabolism of this β-5 linked aromatic dimer. This approach defined a pathway for *N. aromaticivorans* DC-A catabolism that contains enzymes not previously known to be involved in aromatic dimer catabolism. Furthermore, comparative genomic analysis allows us to predict that gene products involved in this catabolic pathway are widespread among the order Sphingomonadales.

Results

N. aromaticivorans Catabolizes DC-A

[0098] To test whether *N. aromaticivorans* can catabolize the β-5 linked dimer DC-A, we used a sacB− strain (23) as the wild-type (WT) and grew it in standard mineral base (SMB) minimal medium with DC-A as the sole carbon source. We found that WT *N. aromaticivorans* grows on DC-A under these conditions (FIG. **2** (A)). This led us to predict that the *N. aromaticivorans* genome encodes enzymes that cleave the β-5 linkage and metabolize the resulting G-family aromatic monomers.

[0099] We then asked whether *N. aromaticivorans* funnels these monomers through the known central aromatic metabolic pathway. To answer this question, we took advantage of the properties of *N. aromaticivorans* strain 12444PDC, which contains mutations in the central aromatic catabolic pathway that allow it to produce PDC when grown in the presence of many G-family aromatics (10). However, since G-aromatics are funneled into PDC in this strain, glucose or another alternative carbon source is required for growth. 12444PDC grown in the presence of 1 g/L glucose and 0.4 mM DC-A grows at a similar rate but to a slightly higher density than when it uses glucose as a sole carbon source (FIG. **2** (B)), suggesting that both the glucose and some of the DC-A are used to produce biomass.

[0100] We used high pressure liquid chromatography-mass spectrometry (HPLC-MS) to analyze the culture medium of 12444PDC grown in the presence of DC-A and glucose for consumption of DC-A and accumulation of PDC or other aromatic intermediates (see FIG. **4** for chemical structures). We found that DC-A disappears from the culture medium and PDC accumulates at 92% of the expected yield, assuming that one mole of DC-A would generate two moles of PDC (FIG. **2** (C)). We used HPLC-MS to identify unknown aromatics (Table 1), including 5-carboxyferulate (5-CF), which represents 5% of the aromatics present in the medium at the end of the incubation period (FIG. **2** (C)). Finally, we observed the transient extracellular accumulation of trace amounts of a compound that was subsequently identified as dehydrodiconiferyl aldehyde (DC-L) (FIG. **11**) and the accumulation of a compound identified as dehydrodiconiferyl carboxylic acid (DC-C), suggesting the side chain of DC-A is oxidized from an alcohol to an aldehyde and then to a carboxylic acid. These results led us to conclude that *N. aromaticivorans* possesses the ability to funnel both G-family monomers of the β-5 linked DC-A dimer through its central aromatic metabolic pathway.

TABLE 1

HPLC-MS multiple reaction monitoring conditions and elution times for the compounds analyzed in this study.

Compound	MW (g/mol)	Parent Ion (—) m/z	Transition 1 m/z	Transition 2 m/z	Transition 3 m/z	Elution Time (min)[1]
PDC	184.10	183.30	111.00	139.05	95.00	1.11
Vanillic Acid	168.14	167.25	152.05	108.05	123.05	2.13
Vanillin	152.15	151.15	136.00	92.00	108.00	2.41
Ferulic Acid	194.18	193.25	134.15	178.00	149.10	2.99
5-carboxyferulate	238.19	237.10	134.10	178.10	149.15	3.36
5-formylferulate	222.19	221.10	206.10	134.10	162.10	3.87
DC-A	358.38	357.15	203.10	339.15	221.20	5.25
DC-C	372.37	371.15	352.30	341.20	191.05	5.62
DC-L	356.37	355.15	337.15	219.05	190.05	5.97
DC-S-C	342.34	341.15	267.15	326.15	282.10	6.72
DC-T-C	682.68	681.25	339.20	637.25	324.15	6.84

[1]Elution times can differ when measurements are taken on different days. The elution times listed are those that are found in the HPLC chromatograms shown in this study.

Genome-Wide Screens Identify Candidate Genes Involved in DC-A Catabolism

[0101] Based on the above results, we sought to identify potential gene products involved in the catabolic pathway for β-5 linked aromatics in *N. aromaticivorans*. To do this, we integrated data from a pair of genome-wide screens. In one approach, we used RNA-Seq to compare mid-log phase transcript abundances of *N. aromaticivorans* 12444PDC grown on glucose plus either DC-A or the G-family aromatic monomer vanillin, which was used as a control because we predicted this aromatic monomer to be a product of DC-A catabolism that is further metabolized by known pathways (20, 21). We focused on the 126 transcripts that exhibited a greater than 2-fold, statistically significant increase in abundance when grown in the presence of DC-A compared to cells grown in the presence of vanillin (FIG. **3** (A)). Additionally, we performed RNA-Seq experiments using glucose alone (FIG. **12** (A)) and glucose plus the G-family monomer ferulic acid (FIG. **12** (B)) as controls, which yielded similar results.

[0102] In a second genome-wide screen, we used an existing *N. aromaticivorans* randomly barcoded transposon insertion sequencing (RB-TnSeq) library (21) to identify insertions that led to fitness defects when cells were grown on DC-A as a sole carbon source compared to those grown on glucose alone. In this screen, we found 91 genes for which transposon insertions led to a greater than 2-fold

reduced abundance (>50% fitness decrease) after ~6.5 doublings when using DC-A compared to glucose as sole carbon sources (FIG. **3** (A)).

[0103] Of the 91 transposon insertions that met the 2-fold abundance reduction threshold in the RB-TnSeq screen, 22 were also among the candidates from the DC-A vs. vanillin RNA-Seq screen. Subsequent analysis centered on five candidate genes annotated as encoding proteins with predicted enzymatic activity (Table 2). Four of these five genes are found in two adjacent predicted transcription units (FIG. **3** (B)), leading us to hypothesize that the gene products encoded by this region of the genome play a key role in DC-A catabolism.

[0104] Below, we present data from in vivo and in vitro experiments used to test this hypothesis. Combined, the data from these experiments identify dehydrogenases that can oxidize the allylic side chain of DC-A in a stepwise manner as well as gene products that open the phenylcoumaran ring in the β-5 interunit linkage of DC-C, cleave the resulting dehydrodiconiferyl stilbene carboxylic acid (DC-S-C), and funnel the monomeric G-family cleavage product 5-formyl ferulate (5-FF) into the *N. aromaticivorans* central aromatic metabolic pathway (FIG. **4**).

TABLE 2

DC-A catabolismcandidate genes identified from RNA-Seq and RB-TnSeq data.

Name	Locus Tag	Transcript Increase[1]	Abundance Reduction[2]	Annotation	Function in DC-A Catabolism
pcfL	Saro_0796	5.39	−5.71	Nuclear transport factor 2 family protein	Phenylcoumaran ring opening
fdhA	Saro_0874	2.17	−3.27	S-(hydroxymethyl) glutathione dehydrogenase	Formaldehyde metabolism; Allylic alcohol oxidation
lsdD	Saro_0802	3.80	−5.34	Carotenoid oxygenase family protein	Stilbene cleavage
ferD	Saro_0797	4.25	−4.18	NAD^+-dependent succinate-semialdehyde dehydrogenase	Allylic aldehyde 5-FF oxidation; oxidation
ligW	Saro_0799	4.65	−1.90	Amidohydrolase	5-CF decarboxylation

[1]$\log_2$ comparing transcript abundance when *N. aromaticivorans* PDC12444 is grown on DC-A plus glucose compared and vanillin plus glucose.
[2]$\log_2$ comparing abundance of *N. aromaticivorans* DSM12444 transposon mutants grown on DC-A to those grown on glucose.

PcfL Opens the DC-A Phenylcoumaran Ring

[0105] We examined the role of PcfL (Saro_0796) in DC-A catabolism by comparing metabolism of this β-5 linked aromatic dimer in the 12444PDC strain with a ΔpcfL in-frame deletion strain (12444PDCΔpcfL). We found that DC-A disappears from the growth medium of this mutant (FIG. **5**A), but unlike the parent strain (FIG. **2** (C)), it does not accumulate PDC. Instead, when grown in the presence of DC-A and glucose, 12444PDCΔpcfL accumulates a compound which we were able to identify as DC-C using a synthetic DC-C standard. In addition, when we quantified DC-C in the 12444PDCΔpcfL medium, we found that one mole of DC-C accumulates per mole of DC-A. Since DC-A catabolism does not progress past DC-C in cells that lack pcfL, we proposed that DC-C is a substrate for this enzyme.

[0106] To evaluate this hypothesis, we incubated *E. coli* cell extracts containing a recombinant PcfL enzyme with pure DC-C. We found that PcfL-containing cell extract converts DC-C to another compound that matches synthetic DC-S-C, while a control extract exhibits no detectable conversion of DC-C under the same conditions (FIG. **5**B). Based on these data and the 44% amino acid identity between PcfL and the γ-formaldehyde lyase LdpA that contributes to 3-1 linked aromatic catabolism in *N. aromaticivorans* (24, 37), we proposed that PcfL removes formaldehyde from DC-C to form the stilbene DC-S-C. We further predicted that the formaldehyde released during this reaction is oxidized by the putative glutathione-dependent dehydrogenase Saro_0874, which we named FdhA (formaldehyde dehydrogenase A), based on homology with an enzyme found in *Rhodobacter sphaeroides* (38, 39). Upon testing these hypotheses, we found that PcfL produces formaldehyde from DC-C in vitro (FIG. **13**) and that a 12444PDCΔfdhA mutant accumulates more extracellular formaldehyde than the parent strain when grown in the presence of DC-A and glucose (FIG. **14**). In sum, our data indicate that PcfL is a newly identified γ-formaldehyde lyase that deformylates DC-C, yielding DC-S-C and formaldehyde (FIG. **5**C). Based on these results, we named this gene product PcfL to denote its activity as a phenylcoumaran γ-formaldehyde lyase.

LsdD Cleaves DC-S-C into Two Aromatic Monomers

[0107] Our results suggest that *N. aromaticivorans* contains one or more gene products that use the stilbene DC-S-C as a substrate. LsdD (Saro_0802) is a candidate for cleavage of DC-S-C since this gene product shares 80% amino acid identity with the *Sphingobium* sp. SYK-6 enzyme LsdD, which has been reported to convert DC-S-C into vanillin and 5-FF (30). Furthermore, *N. aromaticivorans* LsdD (named NOV1 in other work) has been shown to be an iron-dependent dioxygenase that cleaves stilbenes such as resveratrol in vitro (40, 41).

[0108] As predicted by this hypothesis, we found that 12444PDCΔlsdD grown in the presence of DC-A and glucose accumulates DC-S-C in the medium (FIG. **6**A). This strain also accumulates more DC-C than the parent strain (FIG. **2** (B)) before it is metabolized to DC-S-C, with a detectable amount of DC-C still present in the medium after the 18-hour incubation. In addition, HPLC-MS analysis of extracellular compounds in the 12444PDCΔlsdD strain indicated the presence of another unknown aromatic compound in the medium. In control experiments, we found that DC-S-C is subject to abiotic homodimerization to form the dehydroconiferyl tetramer carboxylic acid DC-T-C when incubated in SMB minimal medium (FIG. **15** (A,B)). At the end of the incubation, 76% of the extracellular aromatics produced from DC-A by 12444PDCΔlsdD are found in the

sum of DC-S-C and DC-T-C, while only 9% are converted into PDC. We propose that the low amount of PDC excreted by this strain is derived from the activity of one or more enzymes besides LsdD in cleaving DC-S-C (see Discussion).

[0109] We tested the predicted activity of LsdD by incubating *E. coli* cell extracts containing a recombinant LsdD enzyme with synthetic DC-S-C. When incubated with DC-S-C in the absence of any cofactors, LsdD converts this substrate to 5-FF and vanillin (FIG. **6**B). Therefore, we concluded that LsdD cleaves the β-5 linked stilbene DC-S-C into two G-family monomers (FIG. **6**C) that can then be funneled into the central pathway for aromatic metabolism.

FerD and LigW Convert 5-FF to Ferulic Acid

[0110] Our data indicate that the two monomeric products of DC-A catabolism are the G-aromatic monomers vanillin and 5-FF. In *N. aromaticivorans*, vanillin is known to be oxidized to vanillic acid by LigV before entering central G-aromatic metabolism (21). However, the enzymes that metabolize 5-FF have not been identified in this organism. Based on the data from our genome-wide screens, we hypothesized that the putative pyridine nucleotide-dependent ALDH FerD (Saro_0797) oxidizes 5-FF to 5-CF, which is then decarboxylated by LigW (Saro_0799) to form ferulic acid. Ferulic acid is known to be converted into vanillin via a previously described pathway in *N. aromaticivorans* (21).

[0111] Since the conversion of 5-FF to 5-CF occurs after DC-S-C cleavage, we predicted that growing 12444PDCΔferD in the presence of DC-A and glucose would result in the accumulation of one mole of both 5-FF and PDC per mole of DC-A. We found that 12444PDCΔferD cells transiently accumulate 5-FF in the medium. However, at later time points, as the concentration of 5-FF decreases, the concentration of 5-CF increases. 5-CF can then be funneled into PDC production, leading to the accumulation of 1.17 moles of PDC per mole of DC-A by the end of the incubation (FIG. **7**A). To explain these results, we hypothesize that one or more other N. *aromaticivorans* dehydrogenases can oxidize 5-FF to 5-CF, albeit at a slower rate than FerD. Additionally, *E. coli* cell extract containing recombinant FerD converts 5-FF into 5-CF (FIG. **7**B). As expected, FerD-containing cell extract requires NAD^+ to convert 5-FF to 5-CF (FIG. **16**A) and a purified recombinant FerD protein reduces NAD^+ to NADH during this reaction (FIG. **16**B). From these data, we propose that the NAD^+-dependent dehydrogenase FerD is the major gene product responsible for 5-FF to 5-CF conversion (FIG. **7**C) when cells are grown on DC-A, but that other yet uncharacterized enzymes can also catalyze this reaction.

[0112] We investigated the predicted role of LigW in decarboxylation of 5-CF to ferulic acid by growing a 12444PDCΔligW strain in medium containing DC-A and glucose. Under these conditions, we found that cells lacking ligW accumulate ~1 mole of both PDC and 5-CF per mole of DC-A (FIG. **7**A), suggesting that this gene product is responsible for decarboxylation of 5-CF. As predicted, we found that *E. coli* cell extracts expressing recombinant LigW are able to convert 5-CF into ferulic acid in vitro (FIG. **7**B). We therefore concluded that LigW decarboxylates 5-CF in *N. aromaticivorans* (FIG. **7**C).

Multiple Dehydrogenases can Oxidize the DC-A Allylic Alcohol Side Chain

[0113] Given the predicted intermediates of DC-A catabolism (FIG. **4**), we hypothesized that N. *aromaticivorans* contains enzymes that oxidize the allylic alcohol to an aldehyde and then to a carboxylic acid. The only proteins annotated as either alcohol dehydrogenases (ADH) or aldehyde dehydrogenases (ALDH) that were identified as candidates in our genome-wide screens were FdhA and FerD, respectively. However, in the 12444PDCΔferD and 12444PDCΔfdhA strains, the DC-A allylic side chain was still oxidized to a carboxylic acid (FIG. **7**A, FIG. **14** (A)). Based on these findings, we hypothesized that *N. aromaticivorans* contains multiple partially redundant ADHs and ALDHs that convert DC-A to DC-L and DC-L to DC-C.

[0114] We tested this hypothesis by analyzing the activity of 8 putative ADHs and 9 putative ALDHs for which transcripts represented >2% of the total RNA coding for ADHs or ALDHs when *N. aromaticivorans* is grown in the presence of DC-A (Table 3). We performed enzyme assays to determine the activity of these gene products by expressing recombinant versions of the proteins in *E. coli* and incubating cell extracts normalized to the same protein concentration with either DC-A or DC-L with and without NAD^+ (or PQQ for Saro_2870). We used differences in absorption spectra (FIG. **17**) to monitor conversion from DC-A to DC-L and DC-L to DC-C. Control experiments show that none of the cell extracts containing recombinant ADHs or ALDHs were active on these substrates in the absence of NAD^+.

TABLE 3

Candidate ADHs and ALDHs identified from RNA-Seq data.

Name/ Locus Tag	Enzyme Class	Percent of Total ADH or ALDH Transcripts[1]	Activity on DC-A or DC-L
FdhA	ADH	46.65%	Yes
Saro_0995	ADH	2.16%	Yes
Saro_1431	ADH	2.95%	No
Saro_1476	ADH	2.38%	No
Saro_2795	ADH	2.17%	No
Saro_2870	ADH	30.89%	No
Saro_3899	ADH	3.41%	Yes
Saro_3463	ADH	3.84%	No
Saro_0060	ALDH	2.36%	No
FerD	ALDH	7.43%	Yes
Saro_1104	ALDH	16.02%	Yes
Saro_1197	ALDH	12.16%	Yes
Saro_1410	ALDH	10.16%	No
LigV	ALDH	2.04%	No
Saro_1967	ALDH	22.20%	No
Saro_2869	ALDH	14.74%	Yes
Saro_3848	ALDH	4.76%	No

[1]Percent of total putative ADH or ALDH transcripts when *N. aromaticivorans* 12444PDC is grown in the presence of DC-A.

[0115] We found that the putative ADHs FdhA, Saro_0995, and Saro_3899 convert DC-A to DC-L in vitro, with Saro_0995 exhibiting the highest activity under our assay conditions (FIG. **8** (A)). There was some conversion of DC-A to DC-L when a control *E. coli* extract was incubated with DC-A, suggesting that one or more native *E. coli* enzymes have limited activity on DC-A. However, the conversion of DC-A to DC-L was much faster when using extracts prepared from cells expressing the ADHs listed above.

[0116] Using the same approach, we found that the cell extracts containing recombinant versions of the putative ALDHs FerD, Saro_1104, Saro_1197, and Saro_2869 are able to convert DC-L to DC-C in vitro (FIG. **8** (B)). The similar activity of extracts containing these ALDHs on DC-L suggests that they could each make a significant contribution to the metabolism of DC-L in vivo. Combined, the results of these experiments predict that multiple *N. aromaticivorans* enzymes can oxidize the DC-A allylic alcohol side chain to an aldehyde and then to a carboxylic acid.

Reconstructing the DC-A Catabolic Pathway In Vitro

[0117] As an independent test of whether the enzymes described above are sufficient for the catabolism of DC-A to G-family aromatic monomers, we sought to reconstruct the entire N. *aromaticivorans* DC-A catabolic pathway in vitro. Based on the above results, we predicted that a mixture of cell extracts containing NAD^+, the γ-formaldehyde lyase PcfL, the stilbene cleaving dioxygenase LsdD, the ALDH FerD, the decarboxylase LigW, and the ADH Saro_0995 would be able to convert DC-A to G-family aromatics. After incubating DC-A with these five cell extracts and NAD^+, we observed complete conversion of DC-A to ferulic and vanillic acid (FIG. **9**). When incubated with a control *E. coli* cell extract containing none of these *N. aromaticivorans* enzymes, ferulic acid and vanillic acid do not accumulate. However, DC-A is slowly converted to DC-L by the control extract, resulting in a mixture of DC-A and DC-L, in agreement with observations that some native *E. coli* enzymes have limited activity on DC-A (FIG. **8**A). Overall, this experiment confirms that the *N. aromaticivorans* enzymes we identified are sufficient for the catabolism of DC-A to aromatic monomers that are funneled through known pathways into *N. aromaticivorans* central aromatic metabolism.

Discussion

[0118] Aromatic compounds are an important source of industrial products and there is increasing interest in renewable sources of these compounds. The abundant plant polymer lignin is a potential source of aromatics that could be used in the production of commodity chemicals. To valorize lignin, the various interunit linkages between aromatic subunits of this polymer must be cleaved and the resulting mixture of monomers funneled into products (9, 10, 12). Recently, progress has been made in the biological funneling of aromatics into valuable chemicals using the Alphaproteobacterium *N. aromaticivorans* (15). In this study, we found that *N. aromaticivorans* contains enzymes capable of catabolizing aromatic dimers with β-5 linkages, which is the second most abundant interunit linkage in lignin (25, 26).

[0119] Specifically, we showed that *N. aromaticivorans* can grow on the model β-5 linked G-family aromatic dimer DC-A and that the engineered 12444PDC strain funnels both of its aromatic monomers into PDC production. By combining genomic, genetic, and biochemical assays, we identified gene products that are necessary and sufficient for catabolism of DC-A. Based on these studies, we proposed a catabolic pathway for conversion of DC-A to intermediates in the known *N. aromaticivorans* central aromatic metabolic pathway.

Oxidation of the DC-A Allylic Side Chain

[0120] We identified enzymes that oxidize the allylic alcohol side chain of DC-A to an aldehyde and the aldehyde to a carboxylic acid. Our data show that three *N. aromaticivorans* pyridine nucleotide-dependent ADHs (FdhA, Saro_0995, and Saro_3899) can oxidize the allylic alcohol side chain of DC-A, producing the aldehyde DC-L. We also identified four pyridine nucleotide-dependent ALDHs (FerD, Saro_1104, Saro_1197, and Saro_2869) that can oxidize the aldehyde side chain of DC-L to generate the carboxylic acid DC-C. These findings are consistent with RNA-Seq and RB-TnSeq data that indicate increased transcript abundance for multiple ADHs and ALDHs but small or no fitness defects when these dehydrogenases are mutated, suggesting that oxidization of the allylic alcohol side chain of DC-A could be performed by multiple ADHs and ALDHs in vivo (FIG. **3**A). Additional biochemical and genetic analyses would be needed to quantify the activity of each ADH and ALDH enzyme on DC-A or DC-L and their relative contribution to catabolism of these and other β-5 linked aromatics in vivo.

Cleavage of the β-5 Linkage

[0121] We found that the phenylcoumaran DC-C is converted to the stilbene DC-S-C and formaldehyde by the newly identified γ-formaldehyde lyase PcfL. This strategy for catabolism of a phenylcoumaran by *N. aromaticivorans* diverges from the one reported in another aromatic metabolizing member of the order Sphingomonadales, *Sphingobium* sp. SYK-6 (28, 29). In this bacterium, a pair of enantiospecific oxidoreductases, PhcC and PhcD, as well as other partially redundant dehydrogenases, were shown to sequentially oxidize the phenylcoumaran alcohol to an aldehyde and then a carboxylic acid (28). Next, a pair of enantiospecific decarboxylases, PhcF and PhcG, decarboxylate and open the phenylcoumaran ring on DC-C to produce DC-S-C and CO_2 (29). By comparison, the *N. aromaticivorans* pathway for generating a stilbene from DC-C requires only a single enzyme as PcfL opens the phenylcoumaran ring and releases formaldehyde in a single step. In addition, our finding that recombinant PcfL can completely convert DC-C into DC-S-C indicates that this enzyme is agnostic to the enantiomeric state of its substrate. Additionally, an *Agrobacterium* sp. enzyme catalyzes a similar reaction in which it converts a phenylcoumaran to a stilbene, but this enzyme is a glutathione-dependent LigE family enzyme rather than a γ-formaldehyde lyase like PclF.

[0122] To our knowledge, the only homolog of PcfL that has been characterized is LdpA, which is another *N. aromaticivorans* gene product that converts a dimeric aromatic substrate into a stilbene and releases formaldehyde (24, 37). While we found that PcfL has activity with a phenylcoumaran substrate, LdpA acts on a diarylpropane dimer which is a reported intermediate in the *N. aromaticivorans* β-1 linked aromatic catabolic pathway (24). Since PcfL shares eight of the eleven active site residues of LdpA, future work should test if and how these amino acid differences contribute to the substrate preferences of these two enzymes.

[0123] Once DC-S-C forms, our data show this aromatic dimer is cleaved to form 5-FF and vanillin by the lignostilbene dioxygenase LsdD, a homolog of an enzyme previously reported in *Sphingobium* sp. SYK-6 (30). Cleavage of this β-5 linked stilbene by *N. aromaticivorans* mirrors the

process in 3-1 aromatic dimer metabolism, in which the stilbene produced by LdpA is then cleaved by the dioxygenase NOV2. This combination of a γ-formaldehyde lyase followed by a lignostilbene dioxygenase is a newly described strategy for breaking both β-5 and 3-1 interunit linkages in lignin.

Funneling of Monomers into Central Aromatic Metabolism

[0124] Once the β-5 linked dimer DC-A is cleaved into monomeric products, vanillin and 5-FF are funneled into the *N. aromaticivorans* central G-aromatic metabolic pathway and can be converted into PDC. While vanillin is metabolized through a known pathway (21), our experiments identified enzymes involved in the conversion of 5-FF to 5-CF and then to ferulic acid. We found that 5-FF is oxidized to 5-CF by FerD with minor contributions from one or more uncharacterized ALDHs. We also found that LigW decarboxylates 5-CF to ferulic acid, which is metabolized to vanillin through a known pathway (21). A recently published analysis of 5-FF metabolism in *Sphingobium* sp. SYK-6 reports the same functions for FerD and LigW (31). *N. aromaticivorans* LigW has previously been shown to decarboxylate 5-carboxyvanillate (5-CV) (42), which contains a simple carboxylic acid in place of the allylic acid side chain of 5-CF. Thus, it appears that N. *aromaticivorans* LigW is a relatively broad specificity manganese-dependent aromatic decarboxylase that can function in the metabolism of both the β-5 linked aromatic catabolic pathway intermediate 5-CF and the predicted 5-5 linked aromatic catabolic pathway intermediate 5-CV (43).

Redundant Enzymes in Catabolism of β-5 Linked Aromatics

[0125] *N. aromaticivorans* is known to contain several enzymes with multiple functions in aromatic metabolism (20, 44), so it is not surprising for us to find that LigW is not the only enzyme in this pathway with activity on multiple aromatics. We also showed that the dehydrogenases FerD and FdhA display activity on multiple intermediates in the DC-A catabolic pathway. While FdhA is active in conversion of DC-A to DC-L and in the catabolism of formaldehyde, FerD is a promiscuous ALDH that plays a crucial role in the oxidation of 5-FF to 5-CF but is also able to oxidize both DC-L to DC-C and vanillin to vanillic acid (FIG. **18**).

[0126] In addition, PcfL deformylates not only DC-C, but also DC-A and DC-L in vitro (FIGS. **19**A and **19**B), forming products that match the m/z of predicted allylic alcohol and allylic aldehyde stilbenes (FIG. **19**C). While we propose that side chain oxidation precedes conversion of the phenylcoumaran to a stilbene based on the transient accumulation of DC-C in the medium when 12444PDC is grown on DC-A (FIG. **2**B), it is possible that PcfL converts some DC-A or DC-L to a stilbene prior to side chain oxidation (FIG. **20**).

[0127] In addition to *N. aromaticivorans* enzymes acting on multiple aromatic substrates, it is known that multiple enzymes often mediate the same reaction in aromatic metabolism. Consistent with this, we found that allylic side chain oxidation of DC-A and oxidation of 5-FF are performed by multiple dehydrogenases. While our data indicate that LsdD plays a major role in cleavage of DC-S-C into monomers, it is possible that one or both of two other *N. aromaticivorans* homologs of this dioxygenase (NOV2 (Saro_2809) and Saro_3580) can also perform this reaction. Overall, our findings showcase the robust and flexible strategies *N. aromaticivorans* uses for funneling a range of aromatics into a central metabolic pathway.

Conservation of β-5 Linked Aromatic Catabolic Pathways in the Order Sphingomonadales

[0128] After uncovering the pathway for β-5 linked aromatic catabolism in *N. aromaticivorans*, we asked whether other organisms contain enzymes predicted to function in this pathway. To do so, we searched for homologs (>50% amino acid identity, >70% query coverage) of PcfL, LsdD, FerD, and LigW across all bacteria. We found that 82 organisms, all Alphaproteobacteria, are predicted to contain all four of these enzymes. Of those 82, all but *Maricaulis flavus* are members of the order Sphingomonadales. We also identified organisms with at least two homologs of β-5 linked aromatic catabolism enzymes, which are distributed across both gram-negative and gram-positive bacteria, including members of the orders *Actinomyces*, Gammaproteobacteria, Betaproteobacteria, and Bacilli (FIGS. **21**A-**21**C). Thus, we concluded that the complete N. *aromaticivorans* pathway for β-5 linked aromatics is almost exclusively found in Sphingomonadales, but that other bacteria are predicted to contain some of the enzymes described in this study.

[0129] We also used comparative genomics to analyze the distribution of the β-5 linked aromatic catabolic pathways found in *N. aromaticivorans* and *Sphingobium* sp. SYK-6 (FIG. **10**). For this analysis, we included the two pairs of enantiospecific enzymes (PhcC/PhcD and PhcF/PhcG) from the *Sphingobium* sp. SYK-6 pathway that are not shared by *N. aromaticivorans*. We found that most species predicted to have the enzymes needed for β-5 linked aromatic catabolism contain homologs of LsdD, FerD, and LigW, but they differ in whether they are predicted to convert DC-C to DC-S-C using a PcfL homolog (*N. aromaticivorans* pathway) or through oxidation and decarboxylation of DC-C (*Sphingobium* sp. SYK-6 pathway). Most of the organisms identified by our search contain homologs of either PcfL or PhcC/PhcD and/or PhcF/PhcG, but ten species contain homologs of all of these enzymes, suggesting they can convert a phenylcoumaran to a stilbene via both of these pathways.

[0130] The largest clades of Alphaproteobacteria with predicted β-5 catabolism capabilities are members of the genera *Novosphingobium, Sphingobium*, and *Sphingomonas*, and other members of the family Erythrobacteraceae aside from *Novosphingobium*. Our analysis predicts that the PcfL-dependent formaldehyde releasing pathway found in *N. aromaticivorans* is common in the genus *Novosphingobium*, while the phenylcoumaran oxidation and decarboxylation pathway discovered in *Sphingobium* sp. SYK-6 is common in other Erythrobacteraceae. The *Sphingobium* clade can be split into two groups, one of which is predicted to use either pathway. By contrast, the *Sphingomonas* clade is comprised of organisms predicted to contain either or both pathways for β-5 linked aromatic catabolism. In total, while the PcfL-dependent pathway is found in 82 Alphaproteobacteria, homologs of both PhcC/PhcD and PhcF/PhcG are found in 32 organisms. Overall, this analysis has revealed a conserved core pathway among the Sphingomonadales for metabolism of a β-5 linked stilbene and a pair of diverging pathways for the conversion of a phenylcoumaran to a stilbene.

[0131] In sum, we identified a catabolic pathway for β-5 linked aromatics in *N. aromaticivorans* that uses four conserved enzymes in addition to several partially redundant enzymes to funnel each monomeric unit into the *N. aromaticivorans* central aromatic pathway. Notably, this work

showed that *N. aromaticivorans* uses a heretofore undescribed γ-formaldehyde lyase, PcfL, for converting phenylcoumarans to stilbenes. Future studies should focus on biochemically and mechanistically characterizing PcfL, as well as comparing it to its homolog, LdpA (24, 37), which is reported to generate a stilbene from a R-1 linked aromatic dimer.

[0132] The results of this analysis have expanded our knowledge of the aromatic metabolism of *N. aromaticivorans* and the order Sphingomonadales, laying the groundwork for future metabolic engineering to optimize the production of commodity chemicals from additional major components of deconstructed lignin. This *N. aromaticivorans* pathway holds promise for industrial applications since its catabolism of β-5 linked aromatics to vanillic acid and ferulic acid requires a minimal set of five gene products, as we demonstrated in vitro. These five genes could confer β-5 linked aromatic catabolism on other industrially relevant species. To increase the impact of our findings, future work is needed to assess whether β-5 linked aromatics that have been subjected to different pretreatment conditions are catabolized by *N. aromaticivorans* through a similar pathway to the one elucidated in this study.

Methods

Chemicals

[0133] Other than those noted below, all chemicals used were analytical grade and were purchased commercially.

[0134] (E)-4-(3-(hydroxymethyl)-5-(3-hydroxyprop-1-en-1-yl)-7-methoxy-2,3-dihydrobenzofuran-2-yl)-2-methoxyphenol (DC-A) was synthesized in 65% yield by DIBAL-H reduction of 8-5-coupled diferulate (DFA) (45), which was synthesized from ethyl ferulate through peroxidase-H_2O_2 oxidative coupling reaction (46). (E)-3-(2-(4-hydroxy-3-methoxyphenyl)-3-(hydroxymethyl)-7-methoxy-2,3-dihydrobenzofuran-5-yl)acrylaldehyde (DC-L) was synthesized in 80% yield from DC-A by p-benzoquinone oxidation as previously described (47). (E)-3-(4-hydroxy-3-((E)-4-hydroxy-3-methoxystyryl)-5-methoxyphenyl)acrylic acid (DC-S-C) was synthesized in 23% yield from DFA by alkali hydrolysis at 90° C. as previously described (48). To synthesize (E)-3-(2-(4-hydroxy-3-methoxyphenyl)-3-(hydroxymethyl)-7-methoxy-2, 3-dihydrobenzofuran-5-yl) acrylic acid (DC-C), DFA was selectively reduced in 95% ethanol by $NaBH_4$ to produce the alcohol DFA-1 (32% yield). Protection of phenolic hydroxyl in DFA-1 by phenacyl ether was accomplished in 90% yield. Alkali hydrolysis of the ester group in DFA-2 was performed in 1N NaOH/ethanol (1/1, v/v) solution, producing the acid DFA-3 in 85% yield. Finally, deprotection of the phenacyl ether in DFA-3 by Zinc dust in acetic acid resulted in DC-C in 70% yield. The synthesis of DC-A, DC-L, DC-C, and DC-S-C is depicted in FIG. **12** (A). Each product was confirmed by NMR (FIGS. **12**B-**12**E, Table 4).

[0135] (E)-3-(3-formyl-4-hydroxy-5-methoxyphenyl) acrylic acid (5-FF) was synthesized in 38% yield from ferulic acid by ortho formylation with paraformaldehyde and ammonium acetate in acetic acid as previously described (49). To synthesize (E)-5-(2-carboxyvinyl)-2-hydroxy-3-methoxybenzoic acid (5-CF), the phenolic hydroxyl of 5-FF was protected by acetylation in acetic anhydride/pyridine (1/1, v/v) to produce acetylated 5-FF. The aldehyde group was then converted to carboxylic acid in 85% yield by Oxone oxidation in DMF as previously described (50). Finally, the acetylated 5-CF was transferred in 95% yield to 5-CF by hydrolysis of the acetate with K_2CO_3 in 60% aqueous ethanol. The synthesis of 5-FF and 5-CF is depicted in FIG. **23**A. Each product was confirmed by NMR (FIGS. **23**B and **23**C), Table 4).

[0136] To generate DC-T-C, DC-S-C was incubated under abiotic conditions in SMB minimal medium supplemented with 1 g/L glucose at 30° C. for 2 weeks. DMSO was then added to a 30% final concentration (v/v). The resulting product was recovered by ethyl acetate extraction of the SMB buffer solution. After removing the solvent, the crude residue was directly examined by NMR. It was found that the DC-S-C was completely converted and the majority of products were two stereoisomers of 8-8-coupled dimer DC-T-C, which was identified by comparison of their NMR data with those published (FIG. **15**A, Table 4) (51). This material was used as a 1 mM DC-T-C standard. All other standards were created by dissolving the appropriate compound in DMSO at a final concentration of 100 mM.

TABLE 4

^1H and ^{13}C NMR (acetone-d_6) analysis of indicated compounds.

Compound	^1H NMR Data	^{13}C NMR Data
DC-A	3.52, 3.78-3.88, 3.81,3.85, 4.19, 5.56, 6.23, 6.52, 6.80, 6.87, 6.94, 6.97, 7.03	54.70, 56.13, 56.21, 63.33, 64.49, 88.45, 110.30, 111.41, 115.58, 115.96, 119.51, 128.28, 130.29, 130.42, 131.82, 134.28, 145.09, 147.19, 148.28, 148.82
DC-L	3.61, 3.82, 3.91, 3.87-3.91, 5.65, 6.65, 6.81, 6.88, 7.04, 7.29, 7.32, 7.59, 9.63	54.25, 56.29, 56.46, 64.32, 89.39, 110.59, 113.56, 115.76, 119.64, 119.73, 127.14, 129.00, 131.24, 133.75, 145.65, 147.55, 148.46, 152.41, 154.10, 193.77
DC-C	3.59 (m, 1H), 3.82 (s, 3H, —OMe), 3.83-3.92 (m, 2H), 3.90 (s, 3H, —OMe), 4.18, 5.63, 6.38 (d, J = 15.92 Hz), 6.81 (d, J = 8.15 Hz), 6.88 (dd, J = 8.15, 1.93 Hz), 7.05 (d, J = 1.93 Hz), 7.23 (br-s), 7.25 (br-s), 7.61(d, J = 15.92 Hz)	54.36, 56.20, 56.33, 64.28, 89.14, 110.45, 113.12, 115.67, 116.00, 118.73, 119.67, 129.01, 130.88, 133.86, 145.46, 145.98, 147.41, 148.38, 151.54, 168.04.
DC-S-C	3.91 (s, OMe), 3.95 (s, OMe), 6.44 (d, J = 15.9 Hz),6.83(d, J = 8.1 Hz), 7.05 (dd, J = 8.1, 2.0,), 7.22 (d, J = 2.0 Hz), 7.23 (d, J = 1.9 Hz), 7.31 and 7.33 (ABqt, AVAB = 7.39 Hz, JAB = 16.5 Hz), 7.54 (d, J = 1.9 Hz), 7.63 (1 H, d, J = 15.9 Hz)	56.10, 56.44, 108.96, 109.89, 115.90, 116.18, 120.41, 120.82, 121.10, 125.33, 126.83, 130.57, 130.77, 146.21, 146.88, 147.46, 148.49, 148.71, 168.35

TABLE 4-continued

^{1}H and ^{13}C NMR (acetone-d_6) analysis of indicated compounds.

Compound	^{1}H NMR Data	^{13}C NMR Data
5-FF	3.98 (s, 3H, OMe), 6.52 (d, J = 16.0 Hz), 7.64 (d, J = 16.0 Hz), 7.64 and 7.64 (ABqt, AVAB = 3.56 Hz, JAB = 2.15 Hz), 10.15 (s, —CHO)	56.68, 116.36, 118.06, 122.11, 125.31, 127.39, 144.34, 149.74, 154.02, 167.70, 196.04 (—CHO)
5-CF	3.95 (s, OMe), 6.48 (d, J = 15.95 Hz), 7.59 (d, J = 2.0 Hz), 7.62 (d, J = 15.95 Hz), 7.71 (d, J = 2.0 Hz)	56.50 (OMe), 113.17, 115.43, 117.60, 123.87, 126.30, 144.75, 150.12, 155.52, 167.78, 172.64
DC-T-C (threo isomer)	3.62(s), 3.98 (s), 4.13 (d, J = 3.64 Hz), 5.53 (d, J = 3.64 Hz), 6.30 (d, J = 1.90 Hz), 6.39 (d, J = 15.90 Hz), 6.53 (dd, J = 8.15, 1.90 Hz), 6.67 (d, J = 8.15 Hz), 7.30 (d, J = 1.50 Hz), 7.35 (d, J = 1.50 Hz), 7.59 (d, J = 15.90 Hz)	55.76, 55.98, 56.48, 87.12, 109.10, 113.15, 115.59, 117.72, 118.56, 118.77, 129.60, 130.13, 133.63, 144.20, 145.65, 146.96, 148.30, 151.41, 169.60
DC-T-C (meso isomer)	3.78 (s, OMe), 3.91 (s, OMe), 4.18 (d, J = 6.15 Hz), 5.52 (d, J = 6.15 Hz), 6.25 (d, J = 15.90 Hz), 6.80 (d, J = 1.2 Hz), 6.82 (d, J = 8.10 Hz), 6.84 (dd, J = 8.10, 1.36 Hz), 6.98 (d, J = 1.56 Hz), 7.30 (d, J = 1.56 Hz), 7.52 (d, J = 15.90 Hz)	53.50 (C-8), 56.22, 56.38, 88.67, 110.83, 113.57, 115.85, 116.43, 118.48, 120.12, 129.35, 130.11, 132.91, 145.65, 145.70, 147.81, 148.50, 151.92, 167.93

Bacterial Strains and Growth Media

[0137] *N. aromaticivorans* strain 12444A1879 is referred to as the wild-type elsewhere in this paper. In 12444A1879, a putative sacB homolog (Saro_1879) has been deleted (23) to allow for genomic modifications to be made using the pK18mobsacB plasmid system (52). The 12444PDC strain harbors several gene deletions that allow it to funnel aromatics into production of the aromatic metabolic pathway intermediate PDC (10). 12444PDC was used as a parent strain for the construction of the deletion mutants used to study DC-A catabolism. All *N. aromaticivorans* strains (Table 5) were grown at 30° C. and shaking at 200 rpm in SMB minimal medium supplemented with 1 g/L glucose, except where noted. SMB minimal medium was prepared as previously described (23).

[0138] *E. coli* NEB5a (New England Biolabs, Ipswich, MA) was used as a plasmid host. *E. coli* WM6026 (53) was used as a conjugal donor for mobilizing plasmids into *N. aromaticivorans* while *E. coli* B834 (54) was used to express recombinant proteins. All *E. coli* strains (Table 5) were grown in lysogeny broth (LB) at 37° C. and shaking at 200 rpm, except where noted below.

RNA-Seq Analysis

[0139] Four isolated *N. aromaticivorans* PDC12444 colonies were cultured and grown overnight. The next day, the overnight cultures were diluted 1:1 with SMB minimal medium supplemented with 1 g/L glucose and grown for one hour. The cultures were then diluted 1:100 into separate cultures of SMB minimal medium supplemented with 1 g/L glucose, 1 g/L glucose plus 0.5 mM DC-A, 1 g/L glucose plus 0.5 mM vanillin, or 1 g/L glucose plus 0.5 mM ferulic acid. These cultures were grown until they reached mid-exponential growth phase, at which point growth was stopped by the 1:8 addition of ice cold 5% acid phenol: chloroform (5:1) in ethanol. The cells were pelleted by centrifugation (4,300×g for 10 minutes) at 4° C. and stored at −80° C. RNA was extracted using hot acid phenol: chloroform (5:1), as previously described (55). RNA was purified using the RNeasy Kit (Qiagen, Germantown, MD), checked for purity by NanoDrop spectrophotometry (OD 260:280 ratio >2.0, OD 260:230 ratio >2.0), visualized after electrophoresis on a 1% agarose gel, and quantified with a Qubit fluorometer.

TABLE 5

Bacterial strains used in this study.

Strain	Relevant Characteristics	Source
12444Δ1879	WT *N. aromaticivorans* Δ1879 (sacB-)	(23)
12444PDC	1244441879 Δ2819 (ligI) Δ2864 (desC) Δ2865 (desD)	(10)
12444PDCΔpcfL	12444PDC Δ0796 (pcfL)	This study
12444PDCΔferD	12444PDC Δ0797 (ferD)	This study
12444PDCΔligW	12444PDC Δ0799 (lig W)	This study
12444PDCΔlsdD	12444PDC Δ0802 (lsdD)	This study
12444PDCΔfdhA	12444PDC Δ0874 (fdhA)	This study
E. coli NEB5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endAl thi-1 hsdR17	New England Biolabs
E. coli WM6026	$lacI^{q}$, rrnB3, ΔlacZ4787, hsdR514, ΔaraBAD567, ΔrhaBAD568, rph-1, attλ::pAE12(ΔoriR6K-cat::Frt5), ΔendA::Frt, uidA(ΔMluI)::pir, attHK::pJK1006D(oriR6K-cat::Frt5; trfA::Frt) dap	(53)
E. coli B834	F^{-} hsdS metE gal ompT	(54)

[0140] RNA-Seq library preparation and sequencing was performed by the Joint Genome Institute (JGI) using default parameters. rRNA in the samples was depleted using the QIAseq FastSelect kit (Qiagen, Germantown, MD). Libraries were constructed using the TruSeq stranded mRNA kit (Illumina, San Diego, CA) following standard JGI protocols. The libraries were sequenced on an Illumina NovaSeq to produce 2×150 reads. All paired-end FASTQ files were processed through the same pipeline. Reads were trimmed using Trimmomatic version 0.3 with the default settings except for a HEADCROP of 5, LEADING of 3, TRAILING of 3, SLIDINGWINDOW of 3:30, and MINLEN of 36 (56). After trimming, the reads were aligned to the *N. aromaticivorans* DSM12444 genome sequence (GenBank accession GCF_000013325.1) using bwa-mem (version 0.7.17-h5bf99c6_8) with default settings (57). Alignment files were further processed with Picard-tools (version 2.26.10) (https://broadinstitute.github.io/picard/) (CleanSAM and AddOrReplaceReadGroups commands) and samtools (version 1.2) (sort and index commands) (58). Paired aligned reads were mapped to gene locations using HTSeq version 0.6.0 (59). The R package edgeR (version 3.30.3) (60) with default settings was used to identify significantly differentially expressed genes from pairwise analyses, using Benjamini and Hochberg false discovery rate (FDR) less than 0.05 as a significance threshold (61). Raw sequencing reads were normalized using the fragments per kilobase per million mapped reads method (FPKM). Fold change, FPKM, and FDR for all genes are described elsewhere herein.

Screening a Genome-Scale RB-TnSeq Library

[0141] A previously generated RB-TnSeq library in wild-type *N. aromaticivorans* was used to screen for fitness (21). An aliquot of the library was thawed and cultured in LB supplemented with 50 mg/L kanamycin and grown overnight. The culture was diluted 1:100 into three flasks containing 2 g/L glucose in SMB minimal medium and grown to saturation (~6.5 doublings). Each culture was then diluted to a starting cell density of 40 Klett units in SMB minimal medium with 1 g/L glucose or 1 g/L DC-A as the sole carbon source. The cultures were grown to saturation (~6.5 doublings), split into 0.6 mL aliquots, frozen, and stored at −80° C. The cells were harvested by centrifugation (2,300×g for 5 minutes) at 4° C., resuspended in lysis buffer (0.16 mM EDTA and 2% SDS), and incubated at 65° C. for 5 minutes. Genomic DNA was extracted using 25:24:1 phenol:chloroform:isoamyl alcohol. Barcode DNA sequences were amplified from the genome using custom indexing primers BarSeq_P1 and BarSeq_P2_ITO01 to BarSeq_P2_IT009 (62). Barcode amplicons were quantified using a Qubit fluorometer and pooled before being sequenced at Azenta/GENEWIZ on an Illumina MiSeq with paired-end 150 bp reads (Illumina, San Diego, CA). Barcode frequencies and fitness values were calculated as previously described (62).

Heterologous Protein Expression

[0142] To express recombinant proteins, a single isolated colony of each *E. coli* B834 expression strain was cultured in LB medium containing kanamycin (50 mg/L). The next day, the overnight cultures were diluted 1:1 in LB medium and grown for one hour at 37° C. Next, flasks containing either 48 ml, 2×YPTG medium (16 g/L, tryptone, 10 g/L yeast extract, 5 g/L NaCl, 7 g/L, KH_2PO_4, 3 g/L K_2HPO_4, 18 g/L glucose) or 49.5 mL ZMS-80155 auto-inducing medium (63) were inoculated with 2_mL or 0.5 mL of *E. coli* B834 culture, respectively. The 2×YPTG cultures were allowed to grow until their OD600 reached 0.6-0.8, at which point expression of the recombinant protein was induced via addition of 1 mM isopropyl β-D-1-thiogalactopyranosid (IPTG). Since significant recombinant FdhA was present in inclusion bodies, we added 0.5 M sorbitol and 0.2 M arginine to its culture at the same time we added IPTG (64). 2×YPTG and ZMS-801555 cultures were both grown overnight at room temperature (~24 hours). The cultures were washed twice with cold S30 buffer supplemented with 2 mM dithiothreitol (DTT) (65) and the cells were harvested by centrifugation (3000×g for 10 minutes) at 4° C. The cell pellets were flash frozen in a dry ice-ethanol bath and stored at −80° C. Heterologous expression of His-tagged proteins for purification was performed as described above except the cultures contained 990 mL ZMS-80155 auto-inducing medium and were inoculated with 10 mL *E. coli* B834 culture.

Harvesting Cell Extracts

[0143] Harvested *E. coli* B834 cells containing the recombinant proteins were resuspended in 12 mL ice-cold S30 buffer supplemented with 2 mM DTT for untagged constructs or in 2.5 mL/g pellet lysis buffer (50 mM $NatPO_4*H_2O$, 0.5 mM tris(2-carboxyethyl)phosphine, 5 mM imidazole, 100 mM NaCl, 10% glycerol, and 1% Triton-X-100, pH 8.0) for His-tagged constructs. Cells were sonicated on ice using a QSonic sonicator set to amplitude 40 with 20 seconds on and 40 seconds off cycles for 15 minutes. The sonicated solutions were then centrifuged (7,600×g for 20 minutes) at 4° C. and the supernatant was collected as a crude cell extract, flash frozen in a dry ice-ethanol bath, and stored at −80° C.

Growth Experiments

[0144] All *N. aromaticivorans* strains were cultured in triplicate from three isolated colonies and grown overnight. The next day, the cultures were diluted 1:1 in SMB minimal medium supplemented with 1 g/L glucose and incubated for one hour before being diluted with additional 1 g/L glucose in SMB minimal medium to the same cell density. A portion of these cultures were centrifuged (2,300×g for 5 minutes), the supernatant was discarded, and the cell pellets were diluted in the appropriate growth medium (SMB minimal medium with 1 g/L glucose and with or without 0.5 mM DC-A). One mL aliquots of the resuspended cells were used to inoculate triplicate flasks containing 19 mL of the appropriate medium, giving a starting cell density of 20-25 Klett units. The cultures were grown for 18 hours and growth was monitored using a Klett-Summerson colorimeter (FIG. **24**). At indicated time points, 0.8 mL of the cultures were removed, the cells were pelleted by centrifugation (2,300×g for 5 minutes) at 4° C., and the supernatants were passed through a 0.22 m PVDF syringe filter to collect extracellular samples that were stored at −80° C. for subsequent analysis.

[0145] Since DC-A has low solubility in SMB minimal medium, a 100 mM DC-A stock in DMSO was added to SMB minimal medium that was heated to 65° C. to achieve final concentrations of ~0.45 mM DC-A and 0.5% DMSO after filtering the medium.

Analysis of Extracellular Aromatic Metabolites

[0146] The aromatics in extracellular samples were analyzed on a Shimadzu triple quadrupole liquid chromatography mass spectrometer (Nexera XR HPLC-8045 MS/MS). The mobile phase was a binary gradient with solvent A (0.2% formic acid in water) and solvent B (methanol) using the protocol in FIG. **25** and flowing at a rate of 0.4 mL/min. The stationary phase was a Phemonenex Kinetex F5 column (2.6 μm pore size, 2.1 mm ID, 150 mm length, P/N: H18-105937). The m/z of peaks was determined using a negative ion mode scan. Aromatic compound standards were generated as described above and used to confirm the identity of unknown chemicals through elution and multiple-reaction monitoring (MRM).

[0147] A series of 2-fold dilutions were performed to create a standard curve of eight concentrations of each compound. The standard curves were then used to quantify extracellular concentrations of aromatics via MRM (Table 2). The percent yields of individual compounds were calculated using equation (1).

$$\text{percent yield} = \frac{([\text{aromatic}]_{final} \times n)}{([DC\text{-}A]_{initial} \times 2)} \times 100 \qquad \text{Equation (1)}$$

Where n = number of aromatic rings in the compound

In Vitro Enzyme Activity Assays

[0148] Crude cell extracts containing individual recombinant proteins were prepared as described above. The cell extracts expressing candidate DC-A catabolism proteins and control *E. coli* B834 cell extract or control extract alone were added to 3 separate reaction mixtures containing S30 buffer (pH 8.2) supplemented with aromatic substrate and NAD^+, where appropriate. In candidate test conditions, candidate protein and control extracts each comprised 15% of the final volume and the aromatic and NAD^+ (where appropriate) concentrations were 0.25 mM and 1 mM, respectively. For the in vitro reconstruction of the DC-A catabolic pathway experiment, each of the five protein expression cell extracts made up 5% of the final reaction volume instead. For control reactions, the crude extract from *E. coli* B834 comprised 30% of the final mixture. These reactions were incubated at 30° C. for 6 hours and then diluted 1:1 with 40% acetonitrile, 40% methanol, and 100 mM formic acid in water to terminate enzyme activity. The samples were centrifuged (21,000×g for 5 minutes) at 4° C. and the supernatants were passed through a 0.22 m PVDF syringe filter and stored at −80° C. for further analysis. Experiments testing in vitro activity of purified PcfL and FerD were performed in the same fashion, except HEPES buffer (pH 7.66) was used in placed of S30 buffer and control experiments were conducted by adding additional HEPES buffer instead of crude *E. coli* B834 cell extract.

[0149] Analysis of the in vitro reaction products was performed on a Shimadzu triple quadrupole liquid chromatography mass spectrometer as described above. LC traces were collected and reaction products were identified using MRM methods developed from synthetic standards (Table 2).

[0150] To assay the relative rate of conversion of substrates to products by candidate ADHs and ALDHs, absorbance at 370 nm was used for measuring DC-L concentration since DC-L absorbs at this wavelength while DC-A and DC-C do not (FIG. **17**). *E. coli* B834 cell extracts expressing candidate ADHs or ALDHs as well as control extracts were collected as described above and diluted with S30 buffer plus 2 mM DTT to a total protein concentration of 2 mg/mL. The dehydrogenase and control *E. coli* B834 cell extracts were each added to triplicate wells of a 96-well plate containing S30 buffer (pH 8.2) supplemented with 0.15 mM DC-A or 0.15 mM DC-L, as well as 1 mM electron acceptor (NAD^+ or PQQ, where appropriate). The diluted extracts comprised 5% of the final reaction volume. Each enzyme was tested for activity in assays with and without added electron acceptor. After addition of cell extract to the wells, the 96-well plate was immediately placed in a Tecan Infinite M1000 reader set to maintain a temperature of 30° C. At indicated timepoints over the course of one hour, absorbance of DC-L was measured at 370 nm. Control experiments show that NADH does not accumulate significantly in this cell extract system, potentially due to the activity of native *E. coli* dehydrogenases (FIG. **16**B). A series of standards created by 2-fold dilutions of DC-L in S30 buffer plus 2 mM DTT were used to generate an 8-point standard curve and quantify the concentration of DC-L in the reactions based on absorbance at 370 nm.

[0151] Due to absorbance of PQQ at 370 nm, the activity assay for the putative PQQ-dependent ALDH Saro_2870 was performed as described above except 15 L samples were collected from the reaction at each indicated time point and diluted 1:1 with 40% acetonitrile, 40% methanol, and 100 mM formic acid in water to terminate enzyme activity. These samples were then diluted 5:1 with S30 buffer and analyzed by LC-MS as described above.

[0152] Formaldehyde was measured as a product of PcfL activity by using small aliquots of the cell extract reaction 2mixtures and the Invitrogen Formaldehyde Fluorescent Detection Kit (Invitrogen, Carlsbad, CA). To test for conversion of NAD^+ to NADH by FerD, assays were performed as described above for both the purified FerD and FerD-containing cell extract, except the S30 or HEPES buffer was supplemented with 0.4 mM NAD^+ and 0.4 mM 5-FF. NAD^+ and NADH were quantified using small aliquots of the reactions and the Sigma Aldrich NAD/NADH Quantitation Kit (Sigma Aldrich, St. Louis, MO).

Phylogenetic Analysis

[0153] Predicted homologs of DC-A catabolism genes were identified using NCBI protein-protein BLAST to search all genomes in the NCBI database as of July 2023, excluding uncultured/environmental sample sequences and using cut-offs of 50% amino acid identity and 70% query coverage. All bacteria containing homologs of at least two *N. aromaticivorans* DC-A catabolism enzymes (PcfL, FerD, LigW, and LsdD) were used to create a phylogenetic tree. Alphaproteobacteria containing homologs of at least two *N. aromaticivorans* DC-A catabolism enzymes (PcfL, FerD, LigW, and LsdD) and/or *Sphingobium* sp. SYK-6 DC-A catabolism enzymes that differ from *N. aromaticivorans* (PhcC/PhcD and PhcF/PhcG) were used to create an additional phylogenetic tree.

[0154] Phylogenetic analysis was performed on genomes identified in these BLAST searches (Table 6) using GDTB-Tk (version 2.1.1, release 207_v2) to identify and align the bacterial reference genes using default parameters (66). The multiple sequence alignment file was used to construct maximum likelihood trees using RAxML-ng (version 0.9.0) using model LG+G8+F and default parameters (67). *Bacillus subtilis* subsp. *subtilis* str. 168 was used as an outgroup. Trees were visualized in TreeViewer (version 2.2.0) (68).

TABLE 6

Organisms included in the phylogenetic analyses in FIGS. 10A-10G and FIGS. 21A-21C.

Scientific Name	Assembly Accession Number	Class
Alteraurantiacibacter aestuarii	GCF_009827405.1	Alphaproteobacteria
Alteraurantiacibacter aquimixticola	GCF_004965515.1	Alphaproteobacteria
Alteraurantiacibacter buctensis	GCF_009827655.1	Alphaproteobacteria
Altererythrobacter segetis	GCF_011320115.1	Alphaproteobacteria
Altererythrobacter sp. B11	GCF_003569745.1	Alphaproteobacteria
Altererythrobacter sp. CC-YST694	GCF_020539485.1	Alphaproteobacteria
Altererythrobacter sp. KTW20L	GCF_023501975.1	Alphaproteobacteria
Altererythrobacter sp. Root672	GCF_001427865.1	Alphaproteobacteria
Altericroceibacterium endophyticum	GCF_009827595.1	Alphaproteobacteria
Altericroceibacterium indicum	GCF_009828105.1	Alphaproteobacteria
Altericroceibacterium spongiae	GCF_003610805.1	Alphaproteobacteria
Altericroceibacterium xinjiangense	GCF_003958635.1	Alphaproteobacteria
Aurantiacibacter arachoides	GCF_009827335.1	Alphaproteobacteria
Aurantiacibacter odishensis	GCF_003605195.1	Alphaproteobacteria
Aurantiacibacter rhizosphaerae	GCF_009807005.1	Alphaproteobacteria
Aurantiacibacter sp. MUD11	GCF_026967575.1	Alphaproteobacteria
Aurantiacibacter suaedae	GCF_005434915.1	Alphaproteobacteria
Aurantiacibacter xanthus	GCF_003584015.1	Alphaproteobacteria
Blastomonas fulva	GCF_003431825.1	Alphaproteobacteria
Blastomonas sp. AAP25	GCF_001295965.1	Alphaproteobacteria
Blastomonas sp. RAC04	GCF_001713435.1	Alphaproteobacteria
Bradyrhizobium niftali	GCF_004571025.1	Alphaproteobacteria
Caulobacter sp. S45	GCF_009765965.1	Alphaproteobacteria
Chakrabartia godavariana	GCA 023260075.1	Alphaproteobacteria
Croceibacterium atlanticum	GCF_001008165.2	Alphaproteobacteria
Croceibacterium salegens	GCF_009827435.1	Alphaproteobacteria
Croceibacterium selenioxidans	GCF_018599195.1	Alphaproteobacteria
Croceibacterium soli	GCF_009828065.1	Alphaproteobacteria
Croceibacterium xixiisoli	GCF_009827305.1	Alphaproteobacteria
Emcibacter nanhaiensis	GCF_006385175.1	Alphaproteobacteria
Erythrobacter sp. SG61-1L	GCF_001305965.1	Alphaproteobacteria
Hephaestia sp. MAHUQ-44	GCF_023806085.1	Alphaproteobacteria
Marinicaulis flavus	GCF_002943565.1	Alphaproteobacteria
Neorhizobium galegae	GCF_008806425.1	Alphaproteobacteria
Neorhizobium sp. T25_13	GCF_002968675.1	Alphaproteobacteria
Niveispirillum irakense	GCF_000429645.1	Alphaproteobacteria
Niveispirillum sp. BGYR6	GCF_027568365.1	Alphaproteobacteria
Niveispirillum sp. SYP-B3756	GCF_009495745.1	Alphaproteobacteria
Novosphingobium acidiphilum	GCF_000429005.1	Alphaproteobacteria
Novosphingobium aerophilum	GCF_014230345.1	Alphaproteobacteria
Novosphingobium aromaticivorans	GCF_900102455.1	Alphaproteobacteria
Novosphingobium arvoryzae	GCF_014652615.1	Alphaproteobacteria
Novosphingobium capsulatum	GCF_031454595.1	Alphaproteobacteria
Novosphingobium decolorationis	GCF_018417475.1	Alphaproteobacteria
Novosphingobium fuchskuhlense	GCF_001519075.1	Alphaproteobacteria
Novosphingobium hassiacum	GCF_014196055.1	Alphaproteobacteria
Novosphingobium humi	GCF_028607105.1	Alphaproteobacteria
Novosphingobium jiangmenense	GCF_015694345.1	Alphaproteobacteria
Novosphingobium lentum	GCF_001590965.1	Alphaproteobacteria
Novosphingobium mangrovi	GCF_022818885.1	Alphaproteobacteria
Novosphingobium mathurense	GCF_900168325.1	Alphaproteobacteria
Novosphingobium organovorum	GCF_022832435.1	Alphaproteobacteria
Novosphingobium ovatum	GCF_009909235.1	Alphaproteobacteria
Novosphingobium pentaromativorans	GCA 003241455.1	Alphaproteobacteria
Novosphingobium piscinae	GCF_014230355.1	Alphaproteobacteria
Novosphingobium pokkalii	GCF_014652855.1	Alphaproteobacteria
Novosphingobium profundi	GCF_018491765.1	Alphaproteobacteria
Novosphingobium sediminicola	GCF_014196525.1	Alphaproteobacteria
Novosphingobium sediminis	GCF_007991615.1	Alphaproteobacteria
Novosphingobium sp. AAP1	GCF_001295765.1	Alphaproteobacteria
Novosphingobium sp. AAP83	GCF_001295795.1	Alphaproteobacteria
Novosphingobium sp. AAP93	GCF_001296055.1	Alphaproteobacteria
Novosphingobium sp. B 225	GCF_002198665.1	Alphaproteobacteria
Novosphingobium sp. B-7	GCF_000410615.1	Alphaproteobacteria
Novosphingobium sp. B1	GCF_900176395.1	Alphaproteobacteria
Novosphingobium sp. BW1	GCF_008107685.1	Alphaproteobacteria

TABLE 6-continued

Organisms included in the phylogenetic analyses in FIGS. 10A-10G and FIGS. 21A-21C.

Scientific Name	Assembly Accession Number	Class
Novosphingobium sp. CCH12-A3	GCF_001556015.1	Alphaproteobacteria
Novosphingobium sp. CECT 9465	GCF_920987055.1	Alphaproteobacteria
Novosphingobium sp. CF614	GCF_900113255.1	Alphaproteobacteria
Novosphingobium sp. EMRT-2	GCF_005145025.1	Alphaproteobacteria
Novosphingobium sp. ERN07	GCF_012641335.1	Alphaproteobacteria
Novosphingobium sp. ERW19	GCF_012641315.1	Alphaproteobacteria
Novosphingobium sp. ES2-1	GCF_015169775.1	Alphaproteobacteria
Novosphingobium sp. FKTRR1	GCF_020404405.1	Alphaproteobacteria
Novosphingobium sp. FSW06-99	GCF_001519065.1	Alphaproteobacteria
Novosphingobium sp. Fuku2-ISO-50	GCF_001519055.1	Alphaproteobacteria
Novosphingobium sp. HBC54	GCF_029436685.1	Alphaproteobacteria
Novosphingobium sp. KACC 22771	GCF_028736195.1	Alphaproteobacteria
Novosphingobium sp. KN65.2	GCF_001368935.1	Alphaproteobacteria
Novosphingobium sp. LASN5T	GCF_003856955.1	Alphaproteobacteria
Novosphingobium sp. MBES04	GCF_000813185.1	Alphaproteobacteria
Novosphingobium sp. MD-1	GCF_001014975.1	Alphaproteobacteria
Novosphingobium sp. NBM11	GCF_015390225.1	Alphaproteobacteria
Novosphingobium sp. NDB2Meth1	GCF_900117425.1	Alphaproteobacteria
Novosphingobium sp. PP1Y	GCF_000253255.1	Alphaproteobacteria
Novosphingobium sp. PY1	GCF_017312445.1	Alphaproteobacteria
Novosphingobium sp. SG707	GCF_012275515.1	Alphaproteobacteria
Novosphingobium sp. SG720	GCF_012275365.1	Alphaproteobacteria
Novosphingobium sp. SG751A	GCF_013149295.1	Alphaproteobacteria
Novosphingobium sp. SL115	GCF_026672515.1	Alphaproteobacteria
Novosphingobium sp. THN1	GCF_003454795.1	Alphaproteobacteria
Novosphingobium sp. UBA1939	GCF_002336885.1	Alphaproteobacteria
Novosphingobium subterraneum	GCF_000807925.1	Alphaproteobacteria
Novosphingobium taihuense	GCF_007830315.1	Alphaproteobacteria
Novosphingobium terrae	GCF_017163935.1	Alphaproteobacteria
Novosphingobium umbonatum	GCF_004005905.1	Alphaproteobacteria
Pararhodobacter zhoushanensis	GCF_003990445.1	Alphaproteobacteria
Parasphingopyxis marina	GCF_014237875.1	Alphaproteobacteria
Parerythrobacter sp. C18	GCF_030140925.1	Alphaproteobacteria
Pseudoruegeria sp. HB172150	GCF_013184805.1	Alphaproteobacteria
Rhizobium sp. CF080	GCF_000282095.2	Alphaproteobacteria
Rhizobium terrae	GCF_003425685.1	Alphaproteobacteria
Rhizorhapis suberifaciens	GCF_014200045.1	Alphaproteobacteria
Roseinatronobacter sp. HJB301	GCF_028745735.1	Alphaproteobacteria
Sphingobium chungbukense	GCF_001005725.1	Alphaproteobacteria
Sphingobium cupriresistens	GCF_004152865.1	Alphaproteobacteria
Sphingobium jiangsuense	GCF_014196495.1	Alphaproteobacteria
Sphingobium lactosutens	GCF_013393185.1	Alphaproteobacteria
Sphingobium lignivorans	GCF_014203955.1	Alphaproteobacteria
Sphingobium nicotianae	GCF_018603885.1	Alphaproteobacteria
Sphingobium psychrophilum	GCF_012927105.1	Alphaproteobacteria
Sphingobium sp. 3R8	GCF_020166615.1	Alphaproteobacteria
Sphingobium sp. AntQ-1	GCF_028538045.1	Alphaproteobacteria
Sphingobium sp. AP50	GCF_900109095.1	Alphaproteobacteria
Sphingobium sp. B11D3B	GCF_025961735.1	Alphaproteobacteria
Sphingobium sp. B11D3D	GCF_025961755.1	Alphaproteobacteria
Sphingobium sp. B12D2B	GCF_025961775.1	Alphaproteobacteria
Sphingobium sp. B2	GCF_007693735.1	Alphaproteobacteria
Sphingobium sp. B7D2B	GCF_025961895.1	Alphaproteobacteria
Sphingobium sp. BYY-5	GCF_022758885.1	Alphaproteobacteria
Sphingobium sp. CAP-1	GCF_009720145.1	Alphaproteobacteria
Sphingobium sp. LB126	GCF_002795205.1	Alphaproteobacteria
Sphingobium sp. Leaf26	GCF_001421665.1	Alphaproteobacteria
Sphingobium sp. SYK-6	GCF_000283515.1	Alphaproteobacteria
Sphingobium sp. TCM1	GCF_001650725.1	Alphaproteobacteria
Sphingobium sp. V4	GCF_029590555.1	Alphaproteobacteria
Sphingobium sp. YR768	GCF_900111125.1	Alphaproteobacteria
Sphingobium sp. Z007	GCF_900013445.1	Alphaproteobacteria
Sphingobium terrigena	GCF_003591655.1	Alphaproteobacteria
Sphingobium xanthum	GCF_019737615.1	Alphaproteobacteria
Sphingobium xenophagum	GCF_002288285.1	Alphaproteobacteria
Sphingomonas asaccharolytica	GCF_001598355.1	Alphaproteobacteria
Sphingomonas baiyangensis	GCF_005144715.1	Alphaproteobacteria
Sphingomonas bisphenolicum	GCF_024349785.1	Alphaproteobacteria
Sphingomonas caeni	GCF_026013415.1	Alphaproteobacteria
Sphingomonas canadensis	GCF_026013525.1	Alphaproteobacteria
Sphingomonas hengshuiensis	GCF_000935025.1	Alphaproteobacteria

TABLE 6-continued

Organisms included in the phylogenetic analyses in FIGS. 10A-10G and FIGS. 21A-21C.

Scientific Name	Assembly Accession Number	Class
Sphingomonas lycopersici	GCF_026130585.1	Alphaproteobacteria
Sphingomonas mali	GCF_001598415.1	Alphaproteobacteria
Sphingomonas paucimobilis	GCF_001029575.1	Alphaproteobacteria
Sphingomonas pruni	GCF_001598455.1	Alphaproteobacteria
Sphingomonas psychrotolerans	GCF_002796605.1	Alphaproteobacteria
Sphingomonas sp. AR_OL41	GCF_029911635.1	Alphaproteobacteria
Sphingomonas sp. HMWF008	GCA 003061185.1	Alphaproteobacteria
Sphingomonas sp. So64.6b	GCF_014171475.1	Alphaproteobacteria
Sphingomonas sp. SUN019	GCF_024758705.1	Alphaproteobacteria
Sphingomonas sp. UNC305MFCol5.2	GCF_000712135.1	Alphaproteobacteria
Sphingopyxis granuli	GCF_001956775.1	Alphaproteobacteria
Sphingorhabdus sp. M41	GCF_001586275.1	Alphaproteobacteria
Sphingosinicella sp. CPCC 101087	GCF_004151485.1	Alphaproteobacteria
Sphingosinicella terrae	GCF_003347635.1	Alphaproteobacteria
Caldimonas tepidiphila	GCF_003569765.1	Betaproteobacteria
Glaciimonas soli	GCF_009497155.1	Betaproteobacteria
Massilia cavernae	GCF_003590855.1	Betaproteobacteria
Noviherbaspirillum humi	GCF_900188095.1	Betaproteobacteria
Luteimonas sp. BDR2-5	GCF_021191695.1	Gammaproteobacteria
Pseudomonas capeferrum	GCF_000731675.1	Gammaproteobacteria
Pseudomonas sp. LS1212	GCF_024741815.1	Gammaproteobacteria
Pseudomonas sp. R5(2019)	GCF_009905435.1	Gammaproteobacteria
Geodermatophilus sabuli	GCF_900215145.1	Actinomycetes
Lipingzhangella halophila	GCF_014203805.1	Actinomycetes
Pseudonocardia sp. CNS-004	GCF_001942185.1	Actinomycetes
Pseudonocardia sp. DSM 110487	GCF_019468565.1	Actinomycetes
Pseudonocardia hierapolitana	GCF_007994075.1	Actinomycetes
Rhodococcus jostii	GCF_900105375.1	Actinomycetes
Rhodococcus opacus	GCF_019856255.1	Actinomycetes
Streptomyces sp. NRRL S-813	GCF_000718945.1	Actinomycetes
Streptomyces spiralis	GCF_014654675.1	Actinomycetes
Thermopolyspora flexuosa	GCF_006716785.1	Actinomycetes
Bacillus subtilis subsp. *subtilis* str. 168	GCF_000155325.1	Bacilli
Paenibacillus sp. tmac-D7	GCF_006519665.1	Bacilli

Construction of in-Frame Deletion Mutants

[0155] Gene deletion mutants were constructed using 12444PDC as a parent strain and the pK18mobsacB suicide plasmid. This plasmid was linearized via polymerase chain reaction (PCR) as previously described (23). Regions of *N. aromaticivorans* genomic DNA ~1,000 bp upstream and downstream of each gene of interest (Table 7) were amplified via PCR using the primers listed in Table 8 that contain overhanging regions complementary to the ends of linearized pK18mobsacB. NEBuilder HiFi Assembly system (New England Biolabs, Ipswich, MA) was used to insert the amplified fragments into the linearized plasmid, creating a construct in which the genomic regions upstream and downstream of the gene to be deleted are adjacent to each other with no coding region between them. All plasmids used are listed in Table 9.

TABLE 7

N. aromaticivorans genes analyzed in this study and their associated locus tags. Unnamed alcohol dehydrogenase gene products (ADHs) and aldehyde dehydrogenase gene products (ALDHs) investigated are labeled by enzyme class.

N. aromaticivorans gene	Saro_Locus Tag	SARO_RS Locus Tag
PcfL	Saro_0796	SARO_RS03975
FerD	Saro_0797	SARO_RS03980
LigW	Saro_0799	SARO_RS03990

TABLE 7-continued

N. aromaticivorans genes analyzed in this study and their associated locus tags. Unnamed alcohol dehydrogenase gene products (ADHs) and aldehyde dehydrogenase gene products (ALDHs) investigated are labeled by enzyme class.

N. aromaticivorans gene	Saro_Locus Tag	SARO_RS Locus Tag
LsdD	Saro_0802	SARO_RS04005
FdhA	Saro_0874	SARO_RS04375
LigV	Saro_1668	SARO_RS08360
Putative ADH	Saro_0995	SARO_RS04970
Putative ADH	Saro_1431	SARO_RS07175
Putative ADH	Saro_1476	SARO_RS07405
Putative ADH	Saro_2795	SARO_RS14810
Putative ADH	Saro_2870	SARO_RS14555
Putative ADH	Saro_3463	SARO_RS18190
Putative ADH	Saro_3899	SARO_RS17300
Putative ALDH	Saro_0060	SARO_RS02990
Putative ALDH	Saro_1104	SARO_RS05510
Putative ALDH	Saro_1197	SARO_RS05980
Putative ALDH	Saro_1410	SARO RS07070
Putative ALDH	Saro_1967	SARO_RS09870
Putative ALDH	Saro_2869	SARO_RS14550
Putative ALDH	Saro_3848	SARO_RS17045

TABLE 8

Primers used to create gcne deletion mutants. Capitalized regions are complementary to the end of linearized pK18mobsacB. Underlined bases do not match template.

PCR Reaction	Primers
Linearize pK18mobsacB	pK18msB AseI ampl F: ctgtcgtgccagctgcattaatg (SEQ ID NO: 21) pK18msB -MCS XbaI R: gaacatctagaaagccagtccgcagaaac (SEQ ID NO: 22)
Amplify region upstream of pcfL	PcfL pk18 F: CGATTCATTAATGCAGCTGGCACGACAGcttttcgcttctccagctcgg (SEQ ID NO: 23) PcfL Del R.2: cccacccgcaatctcttatttccggtccaactcccatcaatttagtttgtc (SEQ ID NO: 24)
Amplify region downstream of pcfL	PcfL pk18 R.2: GTTTCTGCGGACTGGCTTTCTAGATGTTCcttccacgatgaagcgggttgg (SEQ ID NO: 25) PcfL Del F.2: gacaaactaaattgatgggagttggaccggaaataagagattgcgggtggg (SEQ ID NO: 26)
Amplify region upstream of ferD	FerD pk18 F: CGATTCATTAATGCAGCTGGCACGACAGcggctcgcgcaatttgttagtaag (SEQ ID NO: 27) FerD Del R.3: ctgccgaccgacaccgcaattatatttaatctccggaagccttttgcctg (SEQ ID NO: 28)
Amplify region downstream of ferD	FerD pk18 R.2: GTTTCTGCGGACTGGCTTTCTAGATGTTCcggatcatgcgcaggtagacgtc (SEQ ID NO: 29) FerD Del F.3: caggcaaaaggcttccggagattaaatataattgcggtgtcggtcggcag (SEQ ID NO: 30)
Amplify region upstream of ligW	LigW pk18 F: CGATTCATTAATGCAGCTGGCACGACAGgaaggcgcaatccggagttctcc (SEQ ID NO: 31) LigW Del R: ccctcccggcgctggtcaaaggcaggcttccttcccgggaag (SEQ ID NO: 32)
Amplify region downstream of ligW	LigW pk18 R: GTTTCTGCGGACTGGCTTTCTAGATGTTCtccagtggaagccgggagtgacc (SEQ ID NO: 33) LigW Del F: cttcccgggaaggaagcctgcctttgaccagcgccgggaggg (SEQ ID NO: 34)
Amplify region upstream of lsdD	LsdD pk18 F.4: CGATTCATTAATGCAGCTGGCACGACAGgggggctaaccgccagtctctatcttc (SEQ ID NO: 35) LsdD Del R.4: gcaatacatacaatattgcaaggaggatgccgccgcatgatccagcccggag (SEQ ID NO: 36)
Amplify region downstream of lsdD	LsdD pk18 R.3: GTTTCTGCGGACTGGCTTTCTAGATGTTCccaacaggcagccgaggatag (SEQ ID NO: 37) LsdD Del F.4: ctccgggctggatcatgcggcggcatcctccttgcaatattgtatgtattgc (SEQ ID NO: 38)
Amplify region upstream of fdhA	FdhA pk18 F: CGATTCATTAATGCAGCTGGCACGACAGctgacacggatotctcctcaacc (SEQ ID NO: 39) FdhA Del R: gtaaaccgtgtaaacccgttcaggtattgctacagccctgttaaattgcg (SEQ ID NO: 40)
Amplify region downstream of fdhA	FdhA pk18 R: cgcaatttaacagggctgtagcaatacctgaacgggtttacacggtttac (SEQ ID NO: 41) FdhA Del F: cgcaatttaacagggctgtagcaatacctgaacgggtttacacggtttac (SEQ ID NO: 42)

TABLE 9

Plasmids used in this study.

Plasmid	Relevant Characteristics	Source
pK18mobsacB	pMB1ori sacB kan^R mobT oriT(RP4) lacZa	(52)
PVP302K	lac promoter lacI, Tev site rtxA (*V. cholera*) kan^R; coding sequence for 8 × His-tag	(8)
pK18mobsacBΔpcfL	pK18mobsacB containing genomic regions flanking pcfL	This study
pK18mobsacBΔlsdD	pK18mobsacB containing genomic regions flanking lsdD	This study
pK18mobsacBΔferD	pK18mobsacB containing genomic regions flanking ferD	This study
pK18mobsacBΔligW	pK18mobsacB containing genomic regions flanking ligW	This study
pK18mobsacBΔfdhA	pK18mobsacB containing genomic regions flanking fdhA	This study
PVP302K-PcfL	pVP302K containing codon optimized PcfL	This study
PVP302K-PcfL-NTag	pVP302K containing codon optimized PcfL downstream of His-tag coding sequence and Tev protease site	This study
PVP302K-LsdD	pVP302K containing codon optimized LsdD	This study
PVP302K-FerD	pVP302K containing codon optimized FerD	This study
PVP302K-FerD-NTag	pVP302K containing codon optimized FerD downstream of His-tag coding sequence and Tev protease site	This study
PVP302K-LigW	PVP302K containing codon optimized LigW	This study
PVP302K-FdhA	pVP302K containing codon optimized FdhA	This study
pVP302K-LigV	pVP302K containing codon optimized LigV	This study
PVP302K-0995	pVP302K containing codon optimized Saro_0995	This study
PVP302K-1431	pVP302K containing codon optimized Saro_1431	This study
PVP302K-1476	pVP302K containing codon optimized Saro_1476	This study
PVP302K-2795	pVP302K containing codon optimized Saro_2795	This study
pVP302K-2870	pVP302K containing codon optimized Saro_2870	This study
pVP302K-3463	pVP302K containing codon optimized Saro_3463	This study
PVP302K-3899	pVP302K containing codon optimized Saro_3899	This study
pVP302K-0060	pVP302K containing codon optimized Saro_0060	This study
PVP302K-1104	pVP302K containing codon optimized Saro_1104	This study
PVP302K-1197	pVP302K containing codon optimized Saro_1197	This study
PVP302K-1410	pVP302K containing codon optimized Saro_1410	This study
PVP302K-1967	pVP302K containing codon optimized Saro_1967	This study
PVP302K-2869	pVP302K containing codon optimized Saro_2869	This study
PVP302K-3848	pVP302K containing codon optimized Saro_3848	This study

[0156] These plasmids were transformed into *E. coli* NEB5α by heat shock. Plasmids were isolated from NEB5αcultures using the QIAprep Miniprep Kit (Q)iagen, Germantown, NID) and the insert regions of the plasmids were amplified and submitted for Sanger sequencing at Functional Biosciences (Madison, WI) or the, University of Wisconsin-Madison DNA Sequencing core facility. Once the sequences of these plasmids were verified, they were transformed via heat shock into *E. coli* WM46026, which served as a conjugal donor to mobilize the plasmids into *N. aromaticivorans* as previously described (16), except that the SMB minimal medium contained 1 g/L glucose.

Construction of Protein Expression Strains

[0157] Plasmids for recombinant protein expression were constructed using pVP302K, which was linearized via PCR using the primers listed in Table 10. Codon optimized (Benchling Biological Software) gBlocks (Table 11) of genes of interest (Table 7) for heterologous recombinant protein expression were obtained from Integrated DNA Technologies (San Diego, California) and amplified by PCR using the primers in Table 9 that contain overhanging regions complementary to the ends of linearized pVP302K. NEBuilder HiFi Assembly system was used to insert the amplified gBlocks into the linearized plasmid, yielding untagged expression plasmids for all genes as well as N-terminal His-tagged constructs with a TEV-protease cleavage site between the tag and the protein for PcfL and FerD. All plasmids used are listed in Table 9.

[0158] These pVP302K derivatives were transformed into *E. coli* NEB5α and their sequences were verified as described above. They were then transformed into *E. coli* B834 by heat shock.

TABLE 10

Primers used to create recombinant protein expression plasmids. Capitalized DNA sequences are complementary to the end of linearized pVP302K.

PCR Reaction	Primers
Linearize PVP302K with no His-tag	PVP302K No His Lin F: taacagaaagccgaaaataacaaagttagc (SEQ ID NO: 43)
	PVP302K No His Lin R: catggttaatttctcctctttaatgaattctgtg (SEQ ID NO: 44)

TABLE 10-continued

Primers used to create recombinant protein expression plasmids. Capitalized DNA sequences are complementary to the end of linearized pVP302K.

PCR Reaction	Primers
Linearize PVP302K with an N-terminal His-tag	PVP302K N-Term Lin F: cagaaagccgaaaataacaaagttagcctgag (SEQ ID NO: 45) PVP302K N-Term Lin R: tgcgatcgcgctctgaaaatacag (SEQ ID NO: 46)
Amplify PcfL gBlock (no His-tag construct)	pVP302K No His PcfL HiFi F: TAAAGAGGAGAAATTAACCATGtccgatagcaatcagattgcc (SEQ ID NO: 47) pVP302K No His PcfL HiFi R: TGTTATTTTCGGCTTTCTGTTAtttccgcgcattttcgc (SEQ ID NO: 48)
Amplify FerD gBlock (no His-tag construct)	PVP302K No His FerD HiFi F: TAAAGAGGAGAAATTAACCATGactgcgtacccttctctcc (SEQ ID NO: 49) pVP302K No His FerD HiFi R: TGTTATTTTCGGCTTTCTGTTAcccttcatgtaccgctttgg (SEQ ID NO: 50)
Amplify LigW gBlock	PVP302K No His LigW HiFi F: TAAAGAGGAGAAATTAACCATGacacaagacctgaagaccgg (SEQ ID NO: 51) pVP302K No His LigW HiFi R: TGTTATTTTCGGCTTTCTGTTAaagtttaaaccatttttcagcgttgg (SEQ ID NO: 52)
Amplify LsdD gBlock	PVP302K No His LsdD HiFi F: TAAAGAGGAGAAATTAACCATGgctcaatttccgaataccccaag (SEQ ID NO: 53) PVP302K No His LsdD HiFi R: TGTTATTTTCGGCTTTCTGTTAtgcggccaggacctttc (SEQ ID NO: 54)
Amplify FdhA gBlock	PVP302K No His LsdD HiFi F: TAAAGAGGAGAAATTAACCATGctaagcgacaggcacgtcaaag (SEQ ID NO: 55) PVP302K No His LsdD HiFi R: TGTTATTTTCGGCTTTCTGTTAgaacaccactactgaacgaatcgatttac (SEQ ID NO: 56)
Amplify PcfL gBlock (N-terminal His-tag construct)	pVP302K-N PcfL HiFi F: AAATCTGTATTTTCAGAGCGCGATCGCAtccgatagcaatcagattgccg (SEQ ID NO: 57) PVP302K-N PcfL HiFi R: GGCTAACTTTGTTATTTTCGGCTTTCTGttatttccgcgcattttcgcg (SEQ ID NO: 58)
Amplify FerD gBlock (N-terminal His-tag construct)	PVP302K-N FerD HiFi F: AAATCTGTATTTTCAGAGCGCGATCGCAactgcgtacccttctctccacatg (SEQ ID NO: 59) PVP302K-N FerD HiFi R: GGCTAACTTTGTTATTTTCGGCTTTCTGttacccttcatgtaccgctttggtgac (SEQ ID NO: 60)
Amplify LigV gBlock	LigV Exp LigV F: CATTAAAGAGGAGAAATTAACCatgcagtttgaacgtatcaatccgatg (SEQ ID NO: 61) Exp LigV R: GTTTAAACTATTAATGATGATGttaaattggatagtgacctggttggg (SEQ ID NO: 62)
Amplify Saro_0995 gBlock	0995 Exp F: CATTAAAGAGGAGAAATTAACCatgaaagccgccgtactc (SEQ ID NO: 63) 0995 Exp R: GTTTAAACTATTAATGATGATGttattgatcaaacacaataacagaacg (SEQ ID NO: 64)
Amplify Saro_1431 gBlock	1431 Exp F: CATTAAAGAGGAGAAATTAACCatgacaatcaatacaattcgcgtacg (SEQ ID NO: 65) 1431 Exp R: CGTTTAAACTATTAATGATGATttaacaaaaatgacggcagctctg (SEQ ID NO: 66)

TABLE 10-continued

Primers used to create recombinant protein expression plasmids. Capitalized DNA sequences are complementary to the end of linearized pVP302K.

PCR Reaction	Primers
Amplify Saro_1476 gBlock	1476 Exp F: CATTAAAGAGGAGAAATTAACCatgttgggacgtgcatcgg (SEQ ID NO: 67) 1476 Exp R: GTTTAAACTATTAATGATGATGttacgtgatcgtoggatcgatc (SEQ ID NO: 68)
Amplify Saro_2795 gBlock	Exp 2795 F: CATTAAAGAGGAGAAATTAACCatggcggcaattaatcttccccg (SEQ ID NO: 69) Exp 2795 R: GTTTAAACTATTAATGATGATGttagccaaagacttcggcatagaggc (SEQ ID NO: 70)
Amplify Saro_2870 gBlock	Exp 2870x F: CATTAAAGAGGAGAAATTAACCatgcgattgaaagtactgggacttatgg (SEQ ID NO: 71) Exp 2870 R: GTTTAAACTATTAATGATGATGttagccacctttggcttctaaag (SEQ ID NO: 72)
Amplify Saro_3463 gBlock	Exp 3463 F: CATTAAAGAGGAGAAATTAACCatgattccgcatggtgaacattcaatgctg (SEQ ID NO: 73) Exp 3463 R: GTTTAAACTATTAATGATGATGttatggcaccaaaaccagagcgccac (SEQ ID NO: 74)
Amplify Saro_3899 gBlock	Exp 3899 F: CATTAAAGAGGAGAAATTAACCatggacgcatacgctgcaattatc (SEQ ID NO: 75) Exp 3899 R: GTTTAAACTATTAATGATGATGttacattttgagaatggcttttatcgcttttc (SEQ ID NO: 76)
Amplify Saro_0060 gBlock	Exp 0060 F: CATTAAAGAGGAGAAATTAACCatgtctacacagcctgcaaccatagctg (SEQ ID NO: 77) Exp 0060 R: GTTTAAACTATTAATGATGATGttatggacgagtttgcccgcttcc (SEQ ID NO: 78)
Amplify Saro_1104 gBlock	Exp 1104 F: CATTAAAGAGGAGAAATTAACCatgcgcgaacggctacagcaatacattg (SEQ ID NO: 79) Exp 1104 R: GTTTAAACTATTAATGATGATGttaggcaggcaggccgctgatcg (SEQ ID NO: 80)
Amplify Saro_1197 gBlock	Exp 1197 F: CATTAAAGAGGAGAAATTAACCatgactgcccctaccgcc (SEQ ID NO: 81) Exp 1197 R: GTTTAAACTATTAATGATGATGttactgctgatgacgatatacagcc (SEQ ID NO: 82)
Amplify Saro_1410 gBlock	Exp 1410 F: CATTAAAGAGGAGAAATTAACCatgggttaccgggttgtagtggtg (SEQ ID NO: 83) Exp 1410 R: CATTAAAGAGGAGAAATTAACCatgcagtttgaacgtatcaatccgatg (SEQ ID NO: 84)
Amplify Saro_1967 gBlock	Exp 1967 F: CATTAAAGAGGAGAAATTAACCatggcgatcaaagttgcgataaac (SEQ ID NO: 85) Exp 1967 R: GTTTAAACTATTAATGATGATGttaaaggaatttcgccattgctcc (SEQ ID NO: 86)

TABLE 10-continued

Primers used to create recombinant protein expression plasmids. Capitalized DNA sequences are complementary to the end of linearized pVP302K.

PCR Reaction	Primers
Amplify Saro_2869 gBlock	Exp 2869 F: CATTAAAGAGGAGAAATTAACCatgaatgacatgactaccatctc (SEQ ID NO: 87) Exp 2869 R: GTTTAAACTATTAATGATGATGttacatttgaataattactgttttagtctc (SEQ ID NO: 88)
Amplify Saro_3848 gBlock	Exp 3848 F: CATTAAAGAGGAGAAATTAACCatggctacgcagttgagaagtgcag (SEQ ID NO: 89) Exp 3848 R: GTTTAAACTATTAATGATGATGttactgatcgaacattccggtacgacc (SEQ ID NO: 90)

TABLE 11

gBlocks of *N. aromaticivorans* genes codon optimized for *E. coli* and used to create heterologous protein expression constructs.

gBlock	Sequence
PcfL gBlock	ccgatagcaatcagattgccgcgcttgaaagtcgcctgaatgacctcgaa aggcgactgacggttagagaggacgagctggacgtacgcaaactccagca tttatacggttatctgattgataaatgcatgtataacgagacagttgacc tgttcacagaagatggggaagtgcggttctttggtggcgtatggaaaggc aaggagggcatccgccgctttgtacgttgaacgttttcagaaacgtttcac ctatggcaataacggcccgattgatgggttcctgttagatcatccacaac ttcaagatattattcacgtgcaggatgatggggtcacggctttgggccgc gcgcgttccatgatgcaagccggtcgccacaaggattatgagggagatgc acctcatctgaaagcgcgtcagtggtgggaaggtggtatatacgaaaaca cttataaaaaagtggatggcgtgtggcgtatgcatatcctaaactacatg ccgatctggcacgcagattttgaaagcggctgggccaataccccgcacga atacgttccttttcccaaagtcacctatccagaagacccgactggaccgg atgaactgattgctgaccattggttatggccgacccataagctgaacccc tttcacatgaaacatccggtgacgggtgaggaaatggtcgcacagcgctg gcagggtgacatcgatcgcgaaaatgcgcggaaataa (SEQ ID NO: 91)
FerD gBlock	actgcgtacccttctctccacatgattattgacggtgcccgtgtcagcgg cggaggacgtcgcacccacgcggtcgtcaatccggctaccggagagacca tcggtgaactgccgctggcagaagttgcagatctggatcgagcgttagaa gtagcggcgaagggcttccgtatttggcgtgacagcacaccgcagcagcg cgcagccgtgttacagggcgcggcccggctgatgctggaacggcaagagg atctcgctcgcatagccacgatggaagaaggtaaaaccctgcccgaggcg cgcatcgaagttctgatgaacgtgggcctgttcaatttttacgctggaga agtatttcgtttatatggccgaaccctagtgcgccctgcgggtcagagaa gcacgatcacgcatgaaccggtagggccggtggccgcctttgctccgtgg aactttccgcttgggaatccaggtcgcaaactgggcgcgccaattgccgc cggttgctcggtgattctaaaagcggcggaagaaacgccggcttcagcgt taggggtgctgcaatgtctgctggatgctggcctgcctaaagaagtggcc caggctgtgttcggtgtgcctgacgaggtgagtcgccacctgttgggcag ttccgttatccgcaagctctcgtttacaggttctaccgtcatcggcaagc atctgatgcgacttgcagccgacaacatgttgcgtacaactatggagctt ggcggccatggtcctgtcttagttttcggtgatgcagatattgacaaagc gctcgataccatggcagcttccaaatatcgtaacgcgggccaagtttgtg tttcaccaaccagatttatagtggaagaaagcgtgttcgaacgttttcgt gatggttttgcagagcgtgtcggtcggatcaaagttggaaatggtttgga tcaggatgcgcagatgggaccgatggcaaatgcccgccgcccggaggcga tggatcgtctgatcggggacgccgtgactcgcggcgcaaggttgcatact gggggcgaacgtgtcggcaacgccggctatttttatgcccccacggttct gagtgaagtaccgctggacgcggctattatgaacgaagaaccgtttggcc cggtagctctgattaatccattcggcggtgaggaagcgatgatcgccgaa gcaaaccgtctgccgtatggcttggcagcctacgcatggacagatagcgc ggcgcgggcaaaacgcttagcacgcgagattgagacggggatgctggggc ttaattctaccatgattggcggcgcggattcgccattcggtggggtgaaa tggtccggacacggttcagaggacggtcccgaaggtgttatggcctgcct tgtcaccaaagcggtacatgaagggtaa (SEQ ID NO: 92)

TABLE 11-continued

gBlocks of *N. aromaticivorans* genes codon optimized for *E. coli* and used to create heterologous protein expression constructs.

gBlock	Sequence
LigW gBlock	acacaagacctgaagaccggcggggagcagggttacctgcgtatcgccac cgaagaagctttcgccacgcgagaaatcattgatgtctacctgcgcatga tacgcgatggaactgctgataaaggtatggtatcattgtggggcttttat gcccagtccccttcagagcgcgccacccagatcttagaacgtctgttaga tcttggcgagcggcgtattgcagatatggatgcgacaggcattgacaagg ctattctagcgctgacctcgccgggcgtacagccgctgcatgacttagat gaagcacggacgctcgcaacccgtgcaaatgatactcttgccgatgcgtg ccaaaagtatccagaccgatttattggaatgggcaccgtggccccgcagg atccggaatggagtgcgcgcgaaattcatcgtggtgcaagggaactgggt tttaagggcatccagatcaacagccacacgcaagggcgctacttggatga ggaattctttgatccgatattccgtgccctcgttgaagtcgaccagccgc tgtatattcatcctgccacttcgccagattccatgatcgatccgatgttg gaagcgggcctggacggtgcaatcttcggcttcggtgtggagacgggcat gcatctgctgcgcctgatcacgattgggattttcgacaaatatcccagct tgcaaattatggttgggcacatgggcgaggcgctgccctactggctctat agactggattatatgcaccaggctggtgtgcgctctcagcgctatgaacg tatgaaaccactgaaaaaaaccatcgaaggttatcttaaaagcaacgtgt tagtgacaaattctggagtcgcgtgggaacctgcgattaaattttgtcag caagtaatgggtgaggatcgggttatgtacgcgatggactacccgtatca gtacgttgcagacgaagtgcgtgcgatggatgccatggacatgagtgcgc aaacgaaaaaaaaattttttcagaccaacgctgaaaaatggtttaaactt taa (SEQ ID NO: 93)
LsdD gBlock	atggctcaatttccgaataccccaagcttcacgggattcaacacgccgtc tcggattgaggcggatattgcagatctggcccacgaaggtacgattccgc aagggttaaacggcgcatttatcgtgtccagcccgatccgcagtttcct ccacgcctcgatgatgacattgcctttaacggagacgggatgattacccg attccatatacatgatggccaggtcgacttccgtcaacgttgggcgaaaa ccgataaatggaaactggaaaacgcggccggaaaagccctgtttggtgcc taccgcaacccactgaccgatgacgaggcggttaaaggcgagatccgttc gaccgccaacactaacgccttcgttttcggtggcaaactgtgggcgatga aagaggacagtccagcactcgtaatggatccggcgacgatggaaaccttc gggttcgaaaagttcggcggtaaaatgacaggccagacctttactgccca tccgaaggtagatccgaaaaccggcaatatggtagcgatcggttatgctg caagcgggttgtgcacagatgatgtgacctacatggaagttagtccggag ggtgaattagtacgcgaagtgtggttcaaagtgccgtattattgcatgat gcacgacttcggcattacagaggattacctcgtgctgcacattgttcctt ccatcggaagctgggaaagattagaacagggcaaaccgcactttggcttt gatactactatgccggttcacctaggtatcattccgaggcgtgacggtgt gcgccaggaagatatccgttggttcacgcgggataattgttttgccagtc atgtactgaatgcttggcaagaagggaccaaaattcactttgtgacttgc gaagcgaaaaacaacatgtttccttttctttccagatgtccatggcgcgcc ctttaacggtatggaggcaatgtcacatcctacggactgggtggtcgaca tggcaagcaacggcgaggactttgctgggatcgtgaagctttccgataca gctgcagaatttcctcgcatcgacgaccggtttaccggccagaaaacccg ccatggttggttcttagaaatggatatgaaacgaccagtggaattgcgcg gtggttcagcgggcggcctgctgatgaattgtctgtttcacaaggacttc gaaacgggtcgtgaacagcattggtggtgcggcccggtttcgtctcttca ggagccgtgttttgttccgcgcgcgaaagatgcccccgaaggtgatggat ggattgtgcaagtttgtaatcgtctggaagaacagcgttccgatttgctg atatttgatgcgctggatattgagaaaggcccggtggctacggtcaatat ccccatccgcctgcgctttggcttgcatggtaattgggcgaatgcagacg aaattgggcttgcggaaaaggtcctggccgcagcgatcgcaggaagcgaa aatctgtattttcagagcgcattggcacatcaccatcatcaccatcacca ttaa (SEQ ID NO: 94)
FdhA gBlock	ctaagcgacaggcacgtcaaagggagaccgcatgaaatgaaaacacgcgc cgcagttgcgtttgcgccaaagcaaccgttggaaattgtagaactggatc tggaaggtcccaaagctggggaagttctggttgagattatggcgactgga gtgtgtcacaccgatgcatatacgttagacgggttcgacagcgaaggcat tttccctagcgtgctgggtcatgaaggtgccggtatcgtgcgcgaagtgg gccctggggtaacttccgtgaaacctggcgatcatgtgatcccgctctat acgccggaatgtcgccagtgcaaatcgtgcttgtcgggtaagaccaacct gtgcaccgctattcgcgccacgcaagggcagggcctgatgcccgatggca ccagtcgtttttcttacaaaggccagaccgtgttccactacatgggttgc agtacattctctaattttacagttctgccagagatcgcggttgcaaagat tcgcgaggatgcgccgtttaaaacctcatgttatattggctgtggcgtga cgacgggtgttggcgcggtgattaacactgctaaagtacaggtcggtgac aacgtcgtggtctttggattaggcggcataggtctcaatgttattcaggg agcgcggcttgccggtgcagggaaaatcattggcgtcgatatcaatccag atcgggaggaatggggccgtaaatttggcatgactgactttctgaatagt aagggcatgagccgcgaggacgtagttgctaaagtcgtcgccatgaccga

TABLE 11-continued

gBlocks of *N. aromaticivorans* genes codon optimized for *E. coli* and used to create heterologous protein expression constructs.

gBlock	Sequence
	tggcggtgcggactatacctttgatgccaccggtaataccgaagtgatgc
	gtacggcgcttgaagcatgccatcgtggttggggaacctccataatcatt
	ggtgtggcagaggcgggtaaagaaattagcacgcgtccgttccaattagt
	tactggccgtaactggcgaggcacggccttcggaggcgccaaggggcgca
	cagatgttccgaaaattgtagatatgtacatgaccggaaaaatcgaaatc
	gatccgatgatcacccatgtcatggggctggaagagatcaacacagcatt
	tgatctgatgcacgctggtaaatcgattcgttcagtagtggtgttctaa
	(SEQ ID NO: 95)
LigV	cagtttgaacgtatcaatccgatgacaggggcagtagcctcgcaggcaga
gBlock	ggccatgaaagcgtcggacattccttccattgctgcccgcgcaggacagg
	cctttccggcgtgggcagcgatgggccccaacgcacgtcgcggcgtactg
	atgaaggggctgcggcgttggaagcgcgggctgatgctttcgtcgaagcc
	atgatgggcgaaatcggcgcgactagagggtgggcgctgtttaaccttgg
	ccttgcagcaagcatggtgcgcgaagccgccgcgctgaccactcaaatct
	ctggagaggttattccatctgacaaaccggggtgtatttcgatggctctg
	cgcgaaccggttggtgtgattttgggcatcgcgccgtggaatgcgccgat
	tatccttggggtgcgcgcaattgccgtgccgcttgcctgcggtaacgcgg
	tgatattaaaagcaagcgaaacatgtccgcgaacccacgcgctcatcatc
	gaggcctttgctgaagcaggtttcccagaaggcgtggttaatgtagtgac
	gaacgcgcctgcagatgcagcggaagtggtcggggcgctgattgatgcgc
	cggaagtgcgtcgtataaacttaccggtagtactaatgtaggcaggatt
	atcgcaaaacgggggccgagcatttgaaaccctgtttactcgaactgggc
	ggtaaagcaccgttaatagttctggatgatgcggatctagacgaagcggt
	caaagctgcggcttttggcgccttcatgaaccaagggcagatttgcatgt
	caacggagcggatcatcgttgtagatgccgttgccgatgcattcgcagat
	aaattcaaggccaaggtcgcctccatggctgtaggcgacccgcgtgaggg
	tacgaccccgttgggtgcagttgtcgacgctaaaactgtcgctcattgcc
	gtagcttaattgacgatgccctggcaaaaggtgcccgtctgctgaccggc
	ggtgaaaccacgcacaatgtgctcatgcccgcccatgtcgtagatggcgt
	gacgcaggatatgaagctgttccgcgatgagagctttggcccagtggtgg
	gcgtgattcgcgcgcgcgacgaagctcatgccattgaactggcgaacgac
	agtgaatatggactgtcagcggctgttttcacacgtgacacagcgcgcgg
	cctgcgagttgcccgccagatccgtagcggtatttgccatgttaatggac
	ctaccgtccacgatgaggcgcagatgccttttggtggagtgggtgcgtcc
	ggctacggtcgttttgggggtaaagccggcatcgatagttttaccgagct
	gagatggattacgatggaaacccaaccaggtcactatccaatttaa
	(SEQ ID NO: 96)
Saro_0995	aaagccgccgtactcgtcgaaccgggtaaaccgctggatattcagcattt
gBlock	aagcgtgagtaaacccggccctcatgaagtccttatacgcacagcagcct
	gcgggctgtgccatagtgacttgcacttcatcgaaggtgcctatccacat
	ccgctgccggctgtgccagggcacgaggctgctgggattgtggaagcggt
	aggttcagaagtgcgcacagtaaaagtgggtgacgctgttgttacctgcc
	tgtccgcgttctgtggtcattgcgagttttgcgtgaccggccggatgtcg
	ctgtgtcttggtggcgatactcggcgcggtgcgggtgaggcacctcgctt
	gacacgcaccgacgatggaagcgcagtgaaccagatgctcaacctatcgg
	cctttgcagaacaaatgctggttcacgaacatgcctgtgttgcgatcaat
	cccgagatgccgctcgatagagctgcggttatcggctgtgcggtaaccac
	tggcgcgggtgcggtgtttaatgctgcgaaactgaccccaggagagacgg
	tatgcgttgtcggctgtggcggcgtaggcttagcaacggtcaatgccgcg
	aaaattgccggggcaggccgtattatcgctgtggatccgatgccggaaaa
	acgcgaactggccatgaaactgggtgcgaccgatgtgatggacgcgggac
	ccgatgctgcggcacagatcgttgaaatgacgaaaggcggcgttcaccat
	gcgatcgaggccgtggggcgtcctgcatctggcgaccttgcggtcgcgac
	gctgcgtcgtgggggcaccgccacgattttaggtatgatgccgctggcac
	acaaggtcggattatcagcgatggatctgctgagcgataagaagctgcag
	ggtgcaattatgggccgcaaccacttcccagtggatctgccgcgactggt
	cgacttctacatgcgtggcttgttggatctagacactatcattgccgaaa
	ggattccgcttgaagggataaacgatggttttgaaaaaatgaaacaggga
	cattccgcccgttctgttattgtgtttgatcaataa
	(SEQ ID NO: 97)
Saro_1431	acaatcaatacaattcgcgtacgttcgccggccactctcgacaccttaaa
gBlock	tttcgatacgctgacggattgtggacaaccgggaacgagcgaaatccgca
	ttcgtctgcgcgcaacttctctgaacttccactactacgcgatgattacc
	agaatgctgccggctgcaacaagtcgaattcctatgtctaacggcgcctg
	acaggttttcggggtgtgcgatggcgtgaccaaattccaggcgcgtaacg
	cagttatctcgacctttttcaccgacaggaacgccggtccgccacagtca
	gccgcgtttacgaccgtcacggctgatgggattaatcgctacgcgcgga
	agaagtggtggccccggctcattggtttacccgcgcgccgttatgctata
	gtcacgcaaaagccgccacgctgacctgcgcgggccttactgcatggcgt

TABLE 11-continued

gBlocks of *N. aromaticivorans* genes codon optimized for *E. coli* and used to create heterologous protein expression constructs.

gBlock	Sequence
	gctttgttcatagataacgctatcaagccgggcgacacggtcttggtgca
	gggcactggcagcgtttcggttttcgcgctgcagttaacaaaggcggcat
	gcgcgcgtgtcatcgcaacgagttcctcccaccagtaactgaaacgcctg
	cgcagccttagagcgaataaaaccataaactataaaacgcaaacctcacg
	ggggatgcagacactagatttcactgccggtatttgtgtacactgtattg
	tcgagattagccggcccggtacgtttcatcaagcgatgatgtccacccgc
	gtgcgtgctcatatcgcgctgatcggtgttctcgcgcgttttgcgggtcc
	agtttaaaccactttgctgatggcacagaatctgcgcgtataaggccttа
	ccgtggcctcacgtaccaatcatctgcgaatgattcccggtatcgaggca
	aaccgtatccaacctgtcattcaccgccatttttccatttccgtattttgc
	cgctgcctttcgccatcaacagagctgccgtcatttttgttaaatcgtga
	ttgacatttga (SEQ ID NO: 98)
Saro_1476	ttgggacgtgcatcggtgctggtaaaaccgaaccaactggagacgtggga
gBlock	tgttaaagtagccgatccggaaccgggcggtgccttagtttcgattgtgc
	tgggtggggtatgcgggagcgacgtccatatattgaccggcgaggctggc
	gtgatgccgtttccgatcattctgggacatgagggcgtgggaaggatcga
	aaaactggggcacggcgtcagcactgattacgctggtgaggaacttaaac
	ccggcgatctggtatattggtcgccgattgctctgtgtcatcgatgttat
	tcctgcaatgttctcgatgaaacaccttgcgaaaatacccagtttttcga
	agatgcttccaagccgaactggggttcatacgcagattatgcatggctgc
	ccaacggtatgccgttctataaactgccagcccaagcgcagcctgaagcg
	gttgctgcgcttggctgtgcacttccaaccgccctgcgcggctttgatcg
	ctgcggcagtgttagagtgggtgaaactgtggttgtccaaggtgcaggcc
	ctgtcggcctgtctgcagtgctcgtggcggcgcaggccggggcgcgtgac
	gtgattgttattgacggttcaccacttcgtcgcgaagcggctaccgcatt
	gggtgcctctctgacgattggcttagatgtcgcgcctgaggaacggcgcc
	ggatgatttacgatcgcgttggtcgcaatggtcccaatgtagtcatcgag
	gcagccggagttctgccagcgtttccggaaggggtggacctgaccggtaa
	ccacggccgttacattgtgctaggattgtggggcgcaatagggacccagc
	cgatcagcccgcgcgacttaacaatcaaaaacctgactatcgctggtgcg
	accttccctaaaccaaaacattattatcaggccttgcatttagcgacggc
	cctgcaggaccgtgtaccgttagccggtctggtgagccaccgttttggcg
	tcagccaggcgggcgaagcgctgagtctcaccaagagtgggacagcgatt
	aaggccgtgatcgatccgacgatcacgtaa (SEQ ID NO: 99)
Saro_2795	gcggcaattaatcttccccgcgtgattcgtgctggtgggggtgcattagc
gBlock	cgaactgcccgatgcaatggcgcagtgcggcctttcacgcccgttcgtgg
	tgaccgatgcattcttagtgcaaagcgggatggtcgctcggatgttagag
	gttctggacggcgctgggattgcggccacggtcttcgatgctacggtacc
	tgatccgactgttgctgtggtagaacaggcgcttggcgcattgcgagagg
	cggaatgtgattgtgtgatcgggtttggaggtggtagcccgatcgacacc
	agtaaagccattgccgccctggcgctggaaccgcgtgcagttcaatccat
	gaaggcaccagcgacgaccgacgtcccgggtctgccgatcattgccgtcc
	cgacgaccgccggcaccggctcggaggcgactaaatttacaatcgtgacc
	gatgaggcgacgagtgaaaaaatgctctgcgcaggtctggccttcctgcc
	tactatagccattgtagatttcgagctgaccatgggcaaaccggctcggc
	taactgccgacacaggtattgattcgctgacacatgcgattgaggcctat
	gtttctaagaaagccaatccgtttagtgatgctatggcgatctcggcgat
	gaaactgatcgcgccgaacattcgcaccgcctgcgccgaacccggaaacc
	gtgctgcacgcgaagcgatgatgattggcgcgcaccatgccggtattgcg
	ttttccaacgctagcgttgcactggtgcacggtatgagccgcccaatcgg
	cgcattctttcatgtgccgcacggattgtccaacgcaatgttgctgcctg
	cgattaccgcgttttccgctccgtcagcgttaccacgttacgccgattgt
	gcccgtgcgatgggtgtagctttggaaagcgaaggcgaccagtctgccgt
	tgcaaggctgctcgacgaactggcggcgctgaacgcagaccttagtgtcc
	cgacgccgcagtcgcatgggatcagcgctgatcgttggtttgaagtagtg
	cctgaaatggcgagacaggcaatagcatcaggctctccaggcaataatcc
	acgcgttcctgatgcggcggaaatcgagcgcctctatgccgaagtctttg
	gctaa (SEQ ID NO: 100)
Saro_2870	cgattgaaagttctgggacttatggcagcactgctgccgctggcggcttg
gBlock	taacatcaaaagcgagggtggaggggatgcagtcgccaacgctggagtca
	cagatgccctgattgcccaagcgcccgaaggcgaatggctgagctatggc
	cgcgattatggggaacaacgcttttcaccgttgacccaaattaatgatgg
	taacgtcgggcagttgggtcttgcctggtttcatgacctggagactgcgc
	gcgggcaagaagcgacgccgctgatgcatgatggtacgttatatatctcg
	actgcgtggtcaatggtgaaagcgttcgatgcaaaaaccggcgcgctgaa
	atggagttacgatcccgaagtaccgcgtgaaacgctggtgcgcgcatgct
	gcgacgcggtcaatcgtggcgtcgcgctgtatggagataaagtttttgta
	ggtacgctcgatggtcgtctagtagcgttagatcagaagaccggaaaagt
	agtttggtccaaggtagtagtgcccaatcaggaggactacaccataactg

TABLE 11-continued

gBlocks of *N. aromaticivorans* genes codon optimized for *E. coli* and used to create heterologous protein expression constructs.

gBlock	Sequence
	gtgccccgcgcgtggtgaaaggcaaagttctgattggtagcggtggctcg
	gagtacaaagctcgaggctatattgccgcctatgacgttaacacaggcaa
	cgaagtgtggaaattccacaccgtccctggcaatccagcggatgggtttg
	agaacaaagcgatggaaaatgccgctcgcacttgggctggtgaatggtgg
	aaactcggtgggggtggcacggtgtgggattccatcacctatgatccagc
	caccaacctagttctgttcggcacaggcaatgcagaaccatggaacccgg
	cagcagccgggggagggagacagcttgtacacgtcctctattgtagcgg
	tgaatgccgatactggcgactatgtatggcattttcaagaaaccccggaa
	gaccgttgggacttcgattccgcgcagcagattacgctggccgacctgac
	aattgatgggcagcggcgccacgtgatccttcatgcgcctaagaacggtc
	atgtttatgtgttggacgcaagaaccgggcagtttctgtcggcaacgccc
	tttgtgatggtgaactgggcgaccggtattgatcctaaaacgggcaaggc
	cactgtcaatccagaagcccgttatgaaaaaaccggcaaaacctttcgtta
	gcctgccaggtgcggtaggcgcacattcatggcagccgcagagtttcagc
	ccgaaaaccggcctgctgtaccttccggtgaacaatgcggcatttcctta
	tgcagccgccaaagactggaaagcaaccgatattggtttccagaccggtc
	tcgacggctatgttaccagtatgccagccgacgcaaaggtccagggcgca
	gcgatgaaagcgaccactggtacgttagtggcgtgggacccggttgcgaa
	gaaagccgcttggaaagtcgaactgccgagcccgagtaacggtggcattt
	tatcgacagctggcaatttagtgtttcaaggtaccgcgggcggtgatttt
	gttgcatacaacgccgataagggcaaacaattatggtcttttccggcgca
	gagtggcatccttgccgcgccgatgacctatgctatcgatggggaacagt
	acgttgcggtcatggtgggctggggaggtgtgtgggacgtcgccacaggt
	gtgctcgctcataaggccaaaaaacagaggaacataagccgcctggtagt
	gttcaaactgggcgggaaagccacgctgccggctgctcctccgatggcaa
	aaatggttttggatccgccgccgtttacaggtacgcccgaacaagctaag
	gccggtggcgaattatacggacgttactgcaacgtttgtcatggtgatgc
	tgcggttgcgggcggcgtgaatccagatctgcgtcactcagctgcgctta
	atgcaccagaggcgatccggtctgtggtgattgaggggggcgctgcagcac
	aacgggatggtctcgttcaaatctgcgctgaagcctgaggatgcggataa
	tatccgccactacttgatcaaacgtgcaaatgaagacaaagctctcgaag
	ccaaaggaggctaa (SEQ ID NO: 101)
Saro_3463 gBlock	attccgcatggtgaacattcaatgctggcaatgcagttggatggtccagg
	caaacggctgcacccagtcgtgcgccctctgccgttaccggggcgaggtg
	aagtgcgggtaaaagtgcatgcctgtggtgtttgccgtacggacctgcac
	gttgcagatggcgatattcacggtctgctacctattgtgccggggcacga
	agtgataggcgttgtcgatgcactggggccgggggtgacggatgttgaac
	ctggtgcgcgtgtaggtgtcccgtggctcggccatgcctgtggcacctgc
	ccatattgcgacagcgggagggaaaaccttttgtgatgcgccgctgttcac
	cggttttactcgcgatggcggatacgctacccatgtgattgcagatgcgc
	gcttttgctttcctattccagagggttttgacgatctgcacgcggcgccg
	ctcctgtgcgcgggcttgatcggctatcgcgctcttcggcttgccggcga
	tgcacctgtactcggattctatggttttggagcggcggcgcatattttag
	ctcaggtggccctgtggcagggtagaacggtttacgcgtttactcgcgat
	ggcgacgctaaggcccaggcctttgctcgtgacatcggttgccaatgggc
	cggaccctctggcgctgcgccgccgcaagctctggacgcagcgatcatct
	tcgcctccgcgggagaattggtgccgacagccctgcgtgcagtgcgcaaa
	ggcgggcgtgttgtctgtgccggtattcatatgagcgatatcccggcatt
	cccctacgccgatttatgggaggaacgtcagatcctgtcggtagcgaatt
	taacccgacgcgatggcgtagaattcctgccccttgcagcgcgtgcaggc
	gttcgcacacatgtcgaggccatgccgttaatgaaagcgaacgaggccct
	ggaccgcctgcgtcgtggcgacgtcagtggcgctctggttttggtgccat
	aa (SEQ ID NO: 102)
Saro_3899 gBlock	gacgcatacgctgcaattatcgagcgtcagggtggagaattcgttctgga
	taacgtatctatcgaggatccgcgcgatggcgaagtgctggttaaggttg
	ccgcagctggcatgtgtcataccgatctgacggttcgcgatcaatattac
	ccgacgccgcttccggcggtgctgggccacgaaggtagcggcgttgttga
	aaaagtgggacgtggcgtcaccactgtcaaaccaggtgacaaagtagtgt
	tatccttcagctattgcggtacttgtccttcgtgcctcaaagggcatcag
	gcatactgtccgagcctgttcccgttaaatttcatgggccgtcgcctgga
	tggttcaacgcccattacacgcaacggtcaagaggtcaacgcctgctttt
	tcgggcaatcctcttttgcgacctatagtattgcgtcagaaaacaattgc
	gtcaaggttgccgacgatgcacagattgaacttttgggcccactgggctg
	cggcattcagaccggtgcgggaagtattttaaatgctctttgtcccgaac
	ctggttcctctatagcgatctttgggggggagtgtaggcttaagcgccg
	tgatggctgctaaagcatcgggctgcttgaagatcatcgcggttgacaga
	aatgcaggtcgcttggaactggcgcgtgaactgggcgccaccgatgtgat
	tgacgccaacacggtcaatgctcaggaagcgatcgtcgcgatgactggtg
	gcggcgccgactatgcaatggataccacagccattccagcggtgctgcgg
	agtgcggtggatagcacgcacaatatgggtgaaacagcagtggtgggcgg

TABLE 11-continued

gBlocks of *N. aromaticivorans* genes codon optimized for *E. coli* and used to create heterologous protein expression constructs.

gBlock	Sequence
	ggcgaaactgggtaccgagttttcactagacatgaataacatgctgtttg
	gtcgaaaattgcgtggcgtagtcgaaggatcgagcacgcctcaggtgttc
	atcccgcaactgattgcgatgcagaaagccgggctgtttccgtttgagaa
	actctgtacctttttatgatctggatcagatcaaccaggccgtagaggata
	ccgaaaagactggaaaagcgataaaagccattctcaaaatgtaa
	(SEQ ID NO: 103)
Saro_0060	tctacacagcctgcaaccatagctgattccgcgaccgatctggttgaggg
gBlock	tcttgcacgtgcagcccgttctgcgcagcgccagttggcgcggatggatt
	caccggtaaaagaacgcgcgctgacgttagccgctgcagcgctgcgtgcc
	gctgaggccgaaatttagccgctaacgcgcaggatatggcgaatggcgc
	agcaaacggcctgtcctcggccatgctcgaccggctgaagttaacgccag
	agcgtctggccggcattgccgatgctgtggcgcaagtcgccgggctggcc
	gatccggtcggcgaggtgatcagtgaagctgcgcgtccgaatggcatggt
	gctgcagagagtgcgtattccggtcggagttatcggcatcatttacgaaa
	gccgccccaacgttaccgccgatgcagcagcgctctgcgtgcgttcaggt
	aatgcggcgattctgcgcggtggctcggaagcggttcatagtaaccgtgc
	gatccataaagcgctggttgctgggcttgccgaaggcggagtgccggcag
	aagcggtgcagcttgtacctacgcaggaccgtgctgccgtaggggcaatg
	etaggtgccgcgggactgatcgacatgatcgttccgcgcggcggaaaaag
	ccttgtcgctcgcgtccaggcagatgcccgcgtgccggtgttagcacact
	tggacggtatcaaccacacgtttgttcatgccagtgcagatccggcgatg
	gcccaagcgatagtgttgaatgccaaaatgcgtcgcaccggcgtttgtgg
	tgcgatggaaaccctgctgattgacgcgacttatccagatccccacggcc
	tggtcgaaccgctgctagacgccggttgcgagctgcgcggcgatgctcga
	gcgagagcaattgatccgaggattgcgccagctgccgacaacgactggga
	tacagaatatttggaagcgattctttcggttgcagtggtcgacggtttgg
	atgaagcgctcgcccacatcgcgcgccatgcctctggtcataccgatgca
	atcgtcgcggcggaccaagatgtggcagaccgattcttagctgaagtaga
	tagcgcaattgtaatgcataatgcatccagccagtttgctgatggcggtg
	agttcggcctgggtgctgagattggtattgccacgggggctgcacgcgc
	gcggccctgtagcgctcgaagggctgactacctacaaatggctggtgcgc
	ggaagcgggcaaactcgtccataa (SEQ ID NO: 104)
Saro_1104	cgcgaacggctacagcaatacattgatggaaagtgggtagacagtgaagg
gBlock	tggcaaacgtcacgaagtcattaatccgactacagaggaaccctgttgtg
	tgattacgctgggcacgcaagcagatgtcgacaaagcagtggccgcggca
	cagcgcgcctttaaaaccttcagcaaaacgacgcgtgaggaacgactggc
	gctgcttgaacgcatcgtagaagaatacaagaagcgtgtccctgatttag
	ccgccgcgatggccgaggaaatgggagctccggtaagctttgccagcacc
	gcgcaagttggcgccggaatcggagcatttctgggcaccatggccgcgct
	ccgtaatttctccttttgttgaggacaacggtgcgtttaaagtggcctacg
	aaccgataggtgttgtgggtatgattacgccatggaactggccactgaat
	cagatagctctgaaagtagcaccggcgctggccgcggggaataccatgat
	cctgaaaccgtccgaggaatgcccaaccaacgcagcgatctttaccgaaa
	ttttggatgccgcaggggttccgccaggggtttttaacctgattcagggc
	gatggtcctggtgtaggcactgcgatcagtagtcatccgggcattgatat
	ggttagtttcaccggttcgacccgtgcgggcatcctcgtggcgaaagctg
	cggccgataccgtcaagcgggtgcatcaggaacttggcggtaaatctccc
	aatgtggtgctgcccgatgcagacttcgcaaaatatctgccgtctaccgc
	gtcaggcccgttggtgaacagcggccagagctgcatttcgccaacccgta
	ttttagtaccaagagaacgcgaagcagaagccgcggcttttgtttctgcg
	atgtactccgcaacaccggtcggggatccgatgcaagaaggtgcgcacat
	tgggccggtggttaacaaagctcagtttgacaagatccgcggtctgattc
	aatcggcaatagacgaaggcgcgaaactcgagacaggggggcccgacttac
	cggccaatgtgaaccgcggctattatatcaaaccaacggtcttttcaggc
	gttactcctgatatgcgcattgctcaggaagaaatcttcggcccggtggc
	gacgattatggcgtacgattcattagaggaggccattgagatcgcaaatg
	atacagcctatggactgtcggcctgcattactggtgatccggcgaaagcg
	gctgaagtcgctcctgagcttcgtgcaggtatggtggctatcaataactg
	gggccctactccgggtgctccgttcggtggctataaacagtccggtaacg
	gtaggggggagggttgtatgggttgaaagacttcatggaaatgaaagcga
	tcagcggcctgcctgcctaa (SEQ ID NO: 105)
Saro_1197	actgcccctaccgccgcagacctttccgccgatattgcacgggtttttgc
gBlock	actgcaacaagcgcacatgtgggaggccaaggcgtccaccgcggcggagc
	gcaaagaaaaattggcgcgtctgaaggccgcggttgaagcacacgcggat
	gacattgtggccgcggttctggaagatacgcgcaaacctgttggtgaaat
	aagggtgaccgaagttctgaatgtaaccgccaatatccagcgaaacatcg
	ataatctcgatgaatggatgaaaccggtcgaggtcgctacctcactgaat
	ccagcggaccgcgcgcagataattcatgaagcgcgcggcgtatgcctgat
	tcttggcccatggaatttccccttaggtctggcgctgggtccggtcgccg

TABLE 11-continued

gBlocks of *N. aromaticivorans* genes codon optimized for *E. coli* and used to create heterologous protein expression constructs.

gBlock	Sequence
	ctgctatcgccgcaggcaatacttgtatcgtgaaattaacggacttgtgt
	ccagcgaccgcaagagtggcatcggtgatcgtgcgtgaagcgttcgatga
	aaaagatgtggctctgtttgagggagacgttagtgtagctaccgcgcttt
	tggatctgccgtttaatcatgtatttttacaggctctccacgtgtaggc
	aaaattgtgatggctgctgcggcaaagcatctgaccagcgtcacgttaga
	gcttggtgggaagtctcccgttattgtcgatgatagcgcagatatcgatc
	aagttgctgcccagttagccgcggccaaacaattcaacggcgggcaggcc
	tgcatttccccggactatgtgtttgtgaaagaagacaaaaaagctgcgct
	ggtagaaggtttccgtgccaatgtgcagaaaaacttgtatgatgatgcag
	gcaacctgaaaaaagacagtattgcacaggtggtcaacaaagcgaacttt
	gatcgtgtgaaagccatgttcgacgatgcagtcgcaaaaggcgcgaccgt
	cgccgctggtggaacgtttgaagcggatgacttgactattcatccgacaa
	tgctgacaggcgtaaccccgcagatgactattctccaggatgagatcttt
	gcccctgtcattccggtgatgacctacgacacgctggatcaagcgatcgg
	gtatatcgaagcacgcgacaaaccgctagcactctatgtttacagtaaag
	atgaagcgaacgttgaaaaggtcttagcccgcacgtcatcgggtggtgtt
	acggtgaatggtgtgttctcgcactacctggaaaacaacctgccgttcgg
	gggggttaacacaagcggtatgggcagctaccatggcgtgttcggattta
	agtgctttagccacgagcgggctgtatatcgtcatcagcagtaa
	(SEQ ID NO: 106)
Saro_1410	ggttaccgggttgtagtggtgggtgcgactgggaatgtggggcgtgaaat
gBlock	gctgaacattctggcagaacgcgagtttccttgtgacgagatcgcagcgg
	ttgctagctctcgttcgcagggcaccgaaatagaatttggcgaaactggc
	cggaagctgaaagtacagaatgttgaaaattttgattttaccggatggga
	cattgcactgtttgcggcgggatcaggcccgacgcagatccatgctccac
	gtgccgcttctcagggctgcgtggtgatcgataacagtagcttataccgc
	atggacccggacgtgcctctgatcgtgcccgaggtgaatccggatgcgat
	tgatggctataccaaaaaaaacattattgccaatccaaactgttccaccg
	cgcaaatggtcgtggcgctgaaaccgttacatgatgccgccaaaattaaa
	agagttgtcgtctccacgtatcaaagcgtttccggcgcgggtaaagaagg
	gatggatgaactgttcgaacaaagccgcgcgatatttgtcggggacccgg
	tggaaccgaaaaaattcaccaaacagatcgcattcaacgtgatccctcat
	atcgatgtattcctagacgatggttcgactaaagaagagtggaaaatggt
	cgccgaaaccaaaaaaattttggaccccaaggttaaggtaacggcaacct
	gcgtgcgtgtgccggtgttcatcggccactcggaagcgttaaacattgag
	ttcgagaatgaaattagtgccgaggaagcgcagaatatcctgcgcgaagc
	accaggtgtgatgctcgtcgataagcgcgagaacggcggatatgttacgc
	cggtcgaatgcgttggtgattttgccacatttgttagccgcgtacgtgag
	gattcaacagttgataacggccttaatatttggtgtgtcagtgataacct
	gaggaaaggtgctgccttgaacgctgtacagattgcagaactgctcggtc
	gtcgacaccttaaaaagggttaa (SEQ ID NO: 107)
Saro_1967	gcgatcaaagttgcgataaacggttttggacgtatcgggaggaatgtggc
gBlock	ccgcgccattttagaacgtcccgattgtgggttagaactggttagcatta
	acgacctggctgatgccaaggctaacgccctgctgtttaaacgcgacagc
	gttcatggcgcgttcagtggcgaagtatcagtggatggcaatgatctgat
	tgtgaatggcaagcgcattcaggtgactgcagagcgcgatcctgctaacc
	tgccacacggagccaatggtattgacattgcgctggaatgcacgggcttt
	ttcaccaatcgtgatggtggccagaaacacttggacgcgggcgccaaacg
	cgttctgatttccgctccggcaaaaaacgtagacctgacggtcgtctatg
	gtgtgaaccacgacaaactgaccggcgatcataagatcgtgtccaacgcg
	agttgcacgaccaactgtttggcgccgatggcaaaagtcctgcatgaatc
	tatcgggattgagcgtggtctaatgacaacgattcattcgtataccaatg
	atcaaaaaatactcgaccagatccatagcgatcctagacgggctcgggca
	gcggcgatgaatatgatccccacaagcaccggggccgcagttgcagtggg
	tgaagttctgccagacttaaaagggaaacttgatggttcgtcgattcgag
	tcccgaccccgaacgtatctgtcgtggatcttactttcacgccgaagcgt
	gataccagcgtagaggaagtaaatggtctcttgaaagcggctgccgaagg
	cgcattgaaaggcgtgttaggttacaccgacgaaccgctggtttcaatcg
	attttaaccacgatccgcatagttcaacaatcgacagccttgagactgcc
	gtgctcgaaggtaaactggtgcgcgtcctgtcttggtacgataatgagtg
	gggcttttccaaccgtatgctggatacggcgggagcaatggcgaaattcc
	tttaa (SEQ ID NO: 108)
Saro_2869	aatgacatgactaccatctcacgcacgcagcgtgaatactccgaggccgc
gBlock	aaaagctttcctcgcgagaaagccgcaattgtttattaataacgagtggg
	tcgatagcagtcacgatgcagtgatcgaagtggaagacccctcgaatggg
	aggattgtaggtcatgtcgttgatgcctcggacaaagacgttgaccgggc
	ggttgccgctgcgcgggccgctttcgatgatggtcgttggtccaacctgc
	cgccaatggtacgcgatcgtaccatgaatcgcctggccgacctgcttgaa
	gcaaacgcagatctctttgcagagctggaagcgattgataatggtaaacc

TABLE 11-continued

gBlocks of *N. aromaticivorans* genes codon optimized for *E. coli* and used to create heterologous protein expression constructs.

gBlock	Sequence
	gaagggtatggccggcgccgttgatattccaggtgcgataagccaactac
	gcttcatggcaggatgggccagcaaggtagctggcgaaacgacgcagcct
	tacacgatgccgaatggcaccgtgtttagttacaccgtcaaagaacccgt
	cggtgtctgcgcgcagattgtgccgtggaacttcccgctgctgatggcat
	cattgaagatcgccccggcgctggcggctggatgtacactggtgctgaaa
	cctgccgaacagacatcgcttaccgcgttaaaactggcagatttggtggt
	tgaggctggctttcctgcgggagtgatcaacattatcacagggaacggcc
	acaccgcaggtgatcgcatggtcaaacatcccgacgtagacaaagtcgcc
	tttactggctccaccgaaatcgggaaactgataaatcgaaacgcaaccac
	cacgcttaaacgggttacgctcgaactgggggggaaaagtcccgtagtggt
	tatgccagacgtagatgtggcgcagaccgcgcctggcgttgccggtgcga
	ttttttttcaacgctggccaggtttgtgttgccggtagtcgtttatatgcg
	caccgttcggtgttcgattccgtgttagaaggtatgacccagactgcgcc
	gttttgggcgccgcgcccgagcctggatccagaagcacacatgggaccgt
	tggtcagcaaagagcaacatgaccgtgtgatgggatatatcgaggcgggc
	aagcgtgatggcgccagcgtagtgatgggcggtgattgcccaagcgctga
	tggagggtactatgttaatccgacgattctggcagacgtgaatccgcaga
	tgtctgtcgtgcgcgaggaaatttttggtccggttgtcgtcgcccaacgc
	ttcgacgatttagatgaagtggcgaaaatggcaaacgacacctgttttgg
	cttaggtgcgggcgtgtggacgcgcgatgttgcggtgatgcataaacttg
	ctttcaaagatcaaatctggcactgtgtggggcaactgccatgccctgatc
	gatacagcgctgccttttggcggctataaagaatctgggctgggtcgaga
	acaggggcgtgccggtattgatgcttatttggagactaaaacagtaatta
	ttcaaatgtaa (SEQ ID NO: 109)
Saro_3848	gctacgcagttgagaagtgcagaaaatgaatatgggatcaaatccgagta
gBlock	tggtcattatataggaggtgagtggattgcaggggatagcggcaagacca
	tagatttactaaatccctctaccggtaaagtgctgaccaaaattcaagcc
	ggcaacgcaaaagatattgaacgcgcgattgccgctgcaaaagcggcgtt
	tccgaagtggagccagagcctgccaggggagcgccaagaaatcctgatag
	aggttgcgcgtcgtctgaaagcacgccattcgcactatgcaaccttagaa
	acgctcaataacggtaaaccgatgcgcgaatcaatgtatttcgatatgcc
	tcaaacgatcgggcaatttgagctgttcgccggtgccgcctatggcctgc
	atggccagacgctggattatccagacgcgattggcatcgtccaccgtgaa
	ccgttaggcgtatgcgcgcagattattccatggaacgtgccgatgttgat
	gatggcgtgcaaaatcgcgcccgcgctggcctctggcaacactgtcgttc
	tgaaaccggccgaaacggtgtgcctttctgtgattgaatttttcgtggaa
	atggctgatctgttgcctccgggtgtgatcaacgttgttaccgggtatgg
	tgctgacgttggcgaggcgcttgtaacaagccctgatgtagctaaagtgg
	cctttaccggttcgattgctacggcgcgccggattattcagtatgcctcg
	gccaatatcattccacagacgctcgagttgggcggtaaatcagcgcatat
	cgtgtgtggcgatgccgatattgacgcggcggtggaaagtgcgactatgt
	ccaccgttttaaataaaggtgaagtctgtctggctggttcacgcctgttt
	ctgcatcagtccatccaggatgaattcctggccaaatttaaaacagcgct
	tgaaggcattcgccaaggcgacccgctagatatggcgactcaacttggag
	cgcaggcatcgaagatgcagtttgacaaggtgcaaagctacttaaggctg
	gctacagaggaaggggcagaggtactgaccggcggtagtcgttcagatgc
	cgcagatctggcagatggcaatttttatcaaaccgacggtttttactaacg
	tcaataactccatgcggatcgcgcaggaagagattttcggaccggttacc
	agcgtaattacatggagcgacgaagacgacatgatgaaacaggccaacaa
	tacaacttacggcttggctggcggtgtctggaccaaggacatcgcacgag
	cacaccgtattgcgcgtaaactcgaaactggcacggtctggatcaatcgc
	tactacaacctgaaagccaacatgccgctgggaggttacaagcaaagtgg
	ctttgggcgtgaattcagccatgaagtgctgaatcactacacccagacca
	aatctgtggttgtcaacctccaggaaggtcgtaccggaatgttcgatcag
	taa (SEQ ID NO: 110)

Protein Purification

[0159] PcfL and FerD were purified from the crude cell extract by fast protein liquid chromatography. The crude cell extracts were applied directly to a Ni-NTA column and washed with buffer A (50 mM $NaH_2PO_4*H_2O$, 0.5 mM tris(2-carboxyethyl) phosphine, 25 mM imidazole, and 200 mM NaCl, pH 7.5). The His-tagged proteins bound to the resin were eluted with Buffer B (50 mM $NaH_2PO_4*H_2O$, 0.5 mM tris(2-carboxyethyl) phosphine, 500 mM imidazole, and 300 mM NaCl, pH 7.5). The eluted proteins were collected and concentrated in Buffer C (50 mM $NaH_2PO_4*H_2O$, 0.5 mM tris(2-carboxyethyl) phosphine, 10 mM imidazole, and 100 mM NaCl, pH 7.5) using a 10 kDA MWCO centrifugal filter and hanging basket centrifugation (3,000×g) at 4° C. Protein concentration was quantified by Bradford protein assay measuring absorbance at 595 nm and the purified proteins were diluted to ~2 mg/mL protein by addition of buffer C. They were then treated overnight at 4° C. with 1 mg TEV-protease per ~30 mg of protein. The protease-treated samples were applied to a Ni-NTA column and the proteins were eluted with buffer C and the high imidazole buffer B was used afterwards to elute any remaining protein. A 10 kDA MWCO centrifugal filter and hanging

basket centrifugation (3,000×g) at 4° C. was used to concentrate the proteins, wash them twice with HEPES buffer (50 mM HEPES, 20 mM NaCl, pH 7.5), and concentrate them again, Fractions were saved throughout the purification process and protein content in each fraction was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis. Glycerol was added to the purified, concentrated proteins to a final concentration of 20% before they were flash frozen in a dry ice-ethanol bath and stored at −80° C. A Bradford protein assay measuring absorbance at 595 nm was used to determine the final protein concentration.

Analysis of Extracellular Formaldehyde

[0160] Extracellular medium samples were collected as described in the Materials and Methods and analyzed for extracellular formaldehyde by the Great Lakes Bioenergy Research Center Metabolomics Lab. Formaldehyde concentrations were measured by headspace analysis using an Agilent 7890 Gas Chromatogram equipped with a LECO Pegasus BT time-of-flight mass spectrometry and controlled using LECO's ChromTOF software v4.72.0.0. The samples were prepared in 20 mL headspace vials (Restek, Cat #23082) by diluting 100 μL of filtered medium into 5 mL of water containing p-TSA as the internal standard. The diluted samples were loaded onto a L-PAL 3 auto-sampler equipped with a 2.5 mL headspace syringe (PAL system, Cat #PAL3-Sys-008655). Prior to injection, each sample was transferred to an agitator preheated to 70° C. and incubated for 40 minutes at 350 rpm prior to loading 500 μL of the headspace gas into the syringe. The sample was injected into a 120° C. inlet with a 50:1 split ratio onto a Stabilwax-DA column (Restek, 30 m×0.25 mm×0.5 μm, Cat #11038) with helium as the mobile phase flowing at a constant 1 mL/min. The temperature program was set at 40° C. for 4.20 minutes, followed by a 40° C./minute ramp up to 200° C. The transfer line to the MS was set to 210° C. The MS source was set to 200° C. and had an acquisition delay of 135 seconds. The chromatogram data was collected from 135-55 seconds at 10 spectra/see covering the mass range of 10-350 m/z. Quantification was performed using p-TSA as the internal standard with a 10-point calibration curve.

DC-S-C Abiotic Dimerization Assay

[0161] The time-dependent abiotic conversion of DC-S-C to DC-T-C was measured in water, DMSO, S30 buffer, and SMB minimal medium supplemented with 1 g/L glucose in a 96-well plate. DC-S-C was added in triplicate to each medium to a concentration of 0.2 mM and the 96-well plate was immediately placed in a Tecan Infinite M1000 reader set to maintain a temperature of 30° C. Every hour for 18 hours, absorbance of DC-S-C was measured at 370 nm since DC-S-C absorbs at 370 nm while DC-T-C does not (FIG. **26**). A series of 2-fold dilutions were performed to create a standard curve of eight concentrations of DC-S-C and of DC-T-C in each medium. The standard curves were then used to quantify extracellular concentrations of these aromatics based on absorbance at 370 nm.

Absorbance Spectra of Standards

[0162] To identify the wavelengths at which to measure absorbance in the ADH and ALDH in vitro assays and DC-S-C abiotic dimerization assay, the absorbance of standards was determined with the goal of identifying wavelengths at which either solely a substrate or solely a product absorbs. Triplicate 0.2 mM mixtures of DC-A, DC-L, and DC-C in S30 buffer and 0.2 mM standards of DC-S-C and DC-T-C in SMB minimal medium supplemented with 1 g/L glucose were created and their absorbance was measured from 230 nm to 500 nm in a Tecan Infinite M1000 reader.

REFERENCES

[0163] 1. Ragauskas A J, Beckham G T, Biddy M J, Chandra R, Chen F, Davis M F, Davison B H, Dixon R A, Gilna P, Keller M, Langan P, Naskar A K, Saddler J N, Tschaplinski T J, Tuskan G A, Wyman C E. 2014. Lignin valorization: improving lignin processing in the biorefinery. Science 344:1246843.

[0164] 2. Sun Z, Fridrich B, de Santi A, Elangovan S, Barta K. 2018. Bright Side of Lignin Depolymerization: Toward New Platform Chemicals. Chem Rev 118:614-678.

[0165] 3. Abu-Omar M M, Barta K, Beckham G T, Luterbacher J S, Ralph J, Rinaldi R, Romin-Leshkov Y, Samec J S M, Sels B F, Wang F. 2021. Guidelines for performing lignin-first biorefining. Energy & Environmental Science 14:262-292.

[0166] 4. Ralph J, Lapierre C, Boerjan W. 2019. Lignin structure and its engineering. Curr Opin Biotechnol 56:240-249.

[0167] 5. Vanholme R, De Meester B, Ralph J, Boerjan W. 2019. Lignin biosynthesis and its integration into metabolism. Curr Opin Biotechnol 56:230-239.

[0168] 6. Sangha A K, Parks J M, Standaert R F, Ziebell A, Davis M, Smith J C. 2012. Radical coupling reactions in lignin synthesis: a density functional theory study. J Phys Chem B 116:4760-8.

[0169] 7. Zakzeski J, Jongerius A L, Bruijnincx P C, Weckhuysen B M. 2012. Catalytic lignin valorization process for the production of aromatic chemicals and hydrogen. ChemSusChem 5:1602-9.

[0170] 8. Gall D L, Ralph J, Donohue T J, Noguera D R. 2017. Biochemical transformation of lignin for deriving valued commodities from lignocellulose. Current Opinion in Biotechnology 45:120-126.

[0171] 9. Linger J G, Vardon D R, Guarnieri M T, Karp E M, Hunsinger G B, Franden M A, Johnson C W, Chupka G, Strathmann T J, Pienkos P T, Beckham G T. 2014. Lignin valorization through integrated biological funneling and chemical catalysis. Proc Natl Acad Sci USA 111:12013-8.

[0172] 10. Perez J M, Kontur W S, Alherech M, Coplien J, Karlen S D, Stahl S S, Donohue T J, Noguera D R. 2019. Funneling aromatic products of chemically depolymerized lignin into 2-pyrone-4-6-dicarboxylic acid with. Green Chemistry 21:1340-1350.

[0173] 11. Kamimura N, Takahashi K, Mori K, Araki T, Fujita M, Higuchi Y, Masai E. 2017. Bacterial catabolism of lignin-derived aromatics: New findings in a recent decade: Update on bacterial lignin catabolism. Environ Microbiol Rep 9:679-705.

[0174] 12. Becker J, Wittmann C. 2019. A field of dreams: Lignin valorization into chemicals, materials, fuels, and health-care products. Biotechnol Adv 37:107360.

[0175] 13. Fredrickson J K, Brockman F J, Workman D J, Li S W, Stevens T O. 1991. Isolation and characterization

of a subsurface bacterium capable of growth on toluene, naphthalene, and other aromatic compounds. Appl Environ Microbiol 57:796-803.

[0176] 14. Fredrickson J K, Balkwill D L, Drake G R, Romine M F, Ringelberg D B, White D C. 1995. Aromatic-degrading *Sphingomonas* isolates from the deep subsurface. Appl Environ Microbiol 61:1917-22.

[0177] 15. Perez J M, Sener C, Misra S, Umana G E, Coplien J, Haak D, Li Y D, Maravelias C T, Karlen S D, Ralph J, Donohue T J, Noguera D R. 2022. Integrating lignin depolymerization with microbial funneling processes using agronomically relevant feedstocks. Green Chemistry 24:2795-2811.

[0178] 16. Vilbert A C, Kontur W S, Gille D, Noguera D R, Donohue T J. 2024. Engineering *Novosphingobium aromaticivorans* to produce cis,cis-muconic acid from biomass aromatics. Appl Environ Microbiol 90:e0166023.

[0179] 17. Hall B W, Kontur W S, Neri J C, Gille D M, Noguera D R, Donohue T J. 2023. Production of carotenoids from aromatics and pretreated lignocellulosic biomass by *Novosphingobium aromaticivorans*. Appl Environ Microbiol 89:e0126823.

[0180] 18. Otsuka Y, Nakamura M, Shigehara K, Sugimura K, Masai E, Ohara S, Katayama Y. 2006. Efficient production of 2-pyrone 4,6-dicarboxylic acid as a novel polymer-based material from protocatechuate by microbial function. Appl Microbiol Biotechnol 71:608-14.

[0181] 19. Shikinaka K, Otsuka Y, Nakamura M, Masai E, Katayama Y. 2018. Utilization of Lignocellulosic Biomass via Novel Sustainable Process. J Oleo Sci 67:1059-1070.

[0182] 20. Perez J M, Kontur W S, Gehl C, Gille D M, Ma Y, Niles A V, Umana G, Donohue T J, Noguera D R. 2021. Redundancy in aromatic O-demethylation and ring opening reactions in *Novosphingobium aromaticivorans* and their impact in the metabolism of plant derived phenolics. Appl Environ Microbiol 87.

[0183] 21. Cecil J H, Garcia D C, Giannone R J, Michener J K. 2018. Rapid, Parallel Identification of Catabolism Pathways of Lignin-Derived Aromatic Compounds in *Novosphingobium aromaticivorans*. Appl Environ Microbiol 84.

[0184] 22. Gall D L, Ralph J, Donohue T J, Noguera D R. 2014. A group of sequence-related sphingomonad enzymes catalyzes cleavage of beta-aryl ether linkages in lignin beta-guaiacyl and beta-syringyl ether dimers. Environ Sci Technol 48:12454-63.

[0185] 23. Kontur W S, Bingman C A, Olmsted C N, Wassarman D R, Ulbrich A, Gall D L, Smith R W, Yusko L M, Fox B G, Noguera D R, Coon J J, Donohue T J. 2018. *Novosphingobium aromaticivorans* uses a Nu-class glutathione S-transferase as a glutathione lyase in breaking the beta-aryl ether bond of lignin. J Biol Chem 293:4955-4968.

[0186] 24. Presley G N, Werner A Z, Katahira R, Garcia D C, Haugen S J, Ramirez K J, Giannone R J, Beckham G T, Michener J K. 2021. Pathway discovery and engineering for cleavage of a β-1 lignin-derived biaryl compound. Metabolic Engineering 65:1-10.

[0187] 25. Chen Z, Wan C X. 2017. Biological valorization strategies for converting lignin into fuels and chemicals. Renewable & Sustainable Energy Reviews 73:610-621.

[0188] 26. Guadix-Montero S, Sankar M. 2018. Review on Catalytic Cleavage of C-C Inter-unit Linkages in Lignin Model Compounds: Towards Lignin Depolymerisation. Topics in Catalysis 61:183-198.

[0189] 27. Habu N, Samejima M, Yoshimoto T. 1988. Metabolic Pathway of Dehydrodiconiferyl Alcohol by *Pseudomonas* Sp Tmy1009. Mokuzai Gakkaishi 34:1026-1034.

[0190] 28. Takahashi K, Hirose Y, Kamimura N, Hishiyama S, Hara H, Araki T, Kasai D, Kajita S, Katayama Y, Fukuda M, Masai E. 2015. Membrane-Associated Glucose-Methanol-Choline Oxidoreductase Family Enzymes PhcC and PhcD Are Essential for Enantioselective Catabolism of Dehydrodiconiferyl Alcohol. Applied and Environmental Microbiology 81:8022-8036.

[0191] 29. Takahashi K, Miyake K, Hishiyama S, Kamimura N, Masai E. 2018. Two novel decarboxylase genes play a key role in the stereospecific catabolism of dehydrodiconiferyl alcohol in sp strain SYK-6. Environmental Microbiology 20:1739-1750.

[0192] 30. Kamimura N, Hirose Y, Masuba R, Kato R, Takahashi K, Higuchi Y, Hishiyama S, Masai E. 2021. LsdD has a critical role in the dehydrodiconiferyl alcohol catabolism among eight lignostilbene α,β-dioxygenase isozymes in sp. strain SYK-6. International Biodeterioration & Biodegradation 159.

[0193] 31. Kawazoe M, Takahashi K, Tokue Y, Hishiyama S, Seki H, Higuchi Y, Kamimura N, Masai E. 2023. Catabolic System of 5-Formylferulic Acid, a Downstream Metabolite of a-5-Type Lignin-Derived Dimer, in SYK-6. Journal of Agricultural and Food Chemistry 71:19663-19671.

[0194] 32. Takahashi K, Kamimura N, Hishiyama S, Hara H, Kasai D, Katayama Y, Fukuda M, Kajita S, Masai E. 2014. Characterization of the catabolic pathway for a phenylcoumaran-type lignin-derived biaryl in *Sphingobium* sp. strain SYK-6. Biodegradation 25:735-45.

[0195] 33. Rashid G M M, Riviere G, Cottyn-Boitte B, Majira A, Cezard L, Sodre V, Lam R, Fairbairn J A, Baumberger S, Bugg T D. 2024. Ether Bond Cleavage of a Phenylcoumaran beta-5 Lignin Model Compound and Polymeric Lignin Catalysed by a LigE-type Etherase from *Agrobacterium* sp. Chembiochem doi:10.1002/cbic.202400132:e202400132.

[0196] 34. Myers K S, Vera J M, Lemmer K C, Linz A M, Landick R, Noguera D R, Donohue T J. 2020. Genome-Wide Identification of Transcription Start Sites in Two Alphaproteobacteria, *Rhodobacter sphaeroides* 2.4.1 and *Novosphingobium aromaticivorans* DSM 12444. Microbiol Resour Announc 9.

[0197] 35. Gonzalez C F, Proudfoot M, Brown G, Korniyenko Y, Mori H, Savchenko A V, Yakunin A F. 2006. Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from *Escherichia coli*, FrmB and YeiG. J Biol Chem 281:14514-22.

[0198] 36. Leonhartsberger S, Korsa I, Bock A. 2002. The molecular biology of formate metabolism in enterobacteria. J Mol Microbiol Biotechnol 4:269-76.

[0199] 37. Kuatsjah E, Zahn M, Chen X, Kato R, Hinchen D J, Konev M O, Katahira R, Orr C, Wagner A, Zou Y, Haugen S J, Ramirez K J, Michener J K, Pickford A R, Kamimura N, Masai E, Houk K N, McGeehan J E, Beckham G T. 2023. Biochemical and structural charac-

terization of a sphingomonad diarylpropane lyase for cofactorless deformylation. Proc Natl Acad Sci USA 120:e2212246120.

[0200] 38. Barber R D, Rott M A, Donohue T J. 1996. Characterization of a glutathione-dependent formaldehyde dehydrogenase from *Rhodobacter sphaeroides*. J Bacteriol 178:1386-93.

[0201] 39. Barber R D, Donohue T J. 1998. Function of a glutathione-dependent formaldehyde dehydrogenase in *Rhodobacter sphaeroides* formaldehyde oxidation and assimilation. Biochemistry 37:530-7.

[0202] 40. Marasco E K, Schmidt-Dannert C. 2008. Identification of bacterial carotenoid cleavage dioxygenase homologues that cleave the interphenyl alpha,beta double bond of stilbene derivatives via a monooxygenase reaction. Chembiochem 9:1450-61.

[0203] 41. McAndrew R P, Sathitsuksanoh N, Mbughuni M M, Heins R A, Pereira J H, George A, Sale K L, Fox B G, Simmons B A, Adams P D. 2016. Structure and mechanism of NOV1, a resveratrol-cleaving dioxygenase. Proc Natl Acad Sci USA 113:14324-14329.

[0204] 42. Vladimirova A, Patskovsky Y, Fedorov A A, Bonanno J B, Fedorov E V, Toro R, Hillerich B, Seidel R D, Richards N G, Almo S C, Raushel F M. 2016. Substrate Distortion and the Catalytic Reaction Mechanism of 5-Carboxyvanillate Decarboxylase. J Am Chem Soc 138: 826-36.

[0205] 43. Peng X, Masai E, Kitayama H, Harada K, Katayama Y, Fukuda M. 2002. Characterization of the 5-carboxyvanillate decarboxylase gene and its role in lignin-related biphenyl catabolism in *Sphingomonas paucimobilis* SYK-6. *Appl Environ Microbiol* 68:4407-15.

[0206] 44. Linz A M, Ma Y, Perez J M, Myers K S, Kontur W S, Noguera D R, Donohue T J. 2021.

[0207] Aromatic Dimer Dehydrogenases from *Novosphingobium aromaticivorans* Reduce Monoaromatic Diketones. Appl Environ Microbiol 87:e0174221.

[0208] 45. Quideau S, Ralph J. 1992. Facile Large-Scale Synthesis of Coniferyl, Sinapyl, and Para-Coumaryl Alcohol. Journal of Agricultural and Food Chemistry 40:1108-1110.

[0209] 46. Ralph J, Conesa MTG, Williamson G. 1998. Simple preparation of 8-5-coupled diferulate. Journal of Agricultural and Food Chemistry 46:2531-2532.

[0210] 47. Kulkarni M G, Mathew S. 1990. 1,4-Benzoquinone—a New Selective Reagent for Oxidation of Alcohols. Tetrahedron Letters 31:4497-4500.

[0211] 48. Yue F X, Gao R L, Piotrowski J S, Kabbage M, Lu F C, Ralph J. 2017. Scaled-up production of poacic acid, a plant-derived antifungal agent. Industrial Crops and Products 103:240-243.

[0212] 49. Li Q, Li Y, Liu W, Wang T Y, Zhu Y J, Du Z Y. 2021. Formylation of Phenols and Paraformaldehyde Catalyzed by Ammonium Acetate. Chinese Journal of Organic Chemistry 41:2038-2044.

[0213] 50. Travis B R, Sivakumar M, Hollist G O, Borhan B. 2003. Facile oxidation of aldehydes to acids and esters with Oxone. Org Lett 5:1031-4.

[0214] 51. Huber R, Marcourt L, Koval A, Schnee S, Righi D, Michellod E, Katanaev V L, Wolfender J L, Gindro K, Queiroz E F. 2021. Chemoenzymatic Synthesis of Complex Phenylpropanoid Derivatives by the *Botrytis cinerea* Secretome and Evaluation of Their Wnt Inhibition Activity. Front Plant Sci 12:805610.

[0215] 52. Schafer A, Tauch A, Jager W, Kalinowski J, Thierbach G, Puhler A. 1994. Small mobilizable multipurpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. Gene 145:69-73.

[0216] 53. Blodgett J A, Thomas P M, Li G, Velasquez J E, van der Donk W A, Kelleher N L, Metcalf W W. 2007. Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. Nat Chem Biol 3:480-5.

[0217] 54. Doherty A J, Ashford S R, Brannigan J A, Wigley D B. 1995. A superior host strain for the overexpression of cloned genes using the T7 promoter based vectors. Nucleic Acids Res 23:2074-5.

[0218] 55. Lakey B D, Myers K S, Alberge F, Mettert E L, Kiley P J, Noguera D R, Donohue T J. 2022. The essential *Rhodobacter sphaeroides* CenKR two-component system regulates cell division and envelope biosynthesis. PLoS Genet 18:e1010270.

[0219] 56. Bolger A M, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.

[0220] 57. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.

[0221] 58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-9.

[0222] 59. Anders S, Pyl P T, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166-9.

[0223] 60. Robinson M D, McCarthy D J, Smyth G K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-40.

[0224] 61. Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate—a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Statistical Methodology 57:289-300.

[0225] 62. Wetmore K M, Price M N, Waters R J, Lamson J S, He J, Hoover C A, Blow M J, Bristow J, Butland G, Arkin A P, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. mBio 6:e00306-15.

[0226] 63. Studier F W. 2005. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207-34.

[0227] 64. Prasad S, Khadatare P B, Roy I. 2011. Effect of chemical chaperones in improving the solubility of recombinant proteins in *Escherichia coli*. Appl Environ Microbiol 77:4603-9.

[0228] 65. Kigawa T, Yabuki T, Matsuda N, Matsuda T, Nakajima R, Tanaka A, Yokoyama S. 2004. Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. J Struct Funct Genomics 5:63-8.

[0229] 66. Chaumeil P A, Mussig A J, Hugenholtz P, Parks D H. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925-7.

[0230] 67. Kozlov A M, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-N G: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35:4453-4455.

[0231] 68. Bianchini G, Sanchez-Baracaldo P. 2024. TreeViewer: Flexible, modular software to visualise and manipulate phylogenetic trees. Ecol Evol 14:e10873.

Enzyme Sequences

FdhA (Saro_0874) Coding Sequence (SEQ ID NO: 1)

Atgctatcggaccgccacgtcaaagggagaccgcacgaaatgaag

acccgcgccgcagttgcgttcgcgcccaagcagccgctcgagatc

gtcgaactggacctcgaaggccccaaggctggcgaagtgctggtc

gagatcatggcgaccggcgtgtgccacaccgatgcctacacgctc

gacgggttcgacagcgaaggcatcttccccagcgtgctgggccac

gaaggcgccggtatcgtgcgcgaggtgggccctggggtcacttcg

gtgaagcccggcgatcacgtgatcccgctctacacgccggaatgc

cgccagtgcaaatcgtgcctctcgggcaagaccaacctgtgcacc

gcgatccgcgccacgcaagggcagggcctgatgcccgacggcacc

agccgcttttcgtacaagggccagaccgtgttccactacatgggc

tgctcgaccttctctaacttcaccgtcctgcccgagatcgcggtt

gccaagatccgcgaggacgcgccgttcaagacctcgtgctatatc

ggctgcggcgtgacgacgggcgtcggcgcggtgatcaacaccgcc

aaggtccaggtcggtgacaacgtcgtggtcttcggcctcggcggc

atcggcctcaacgtgatccagggcgcgcggcttgccggtgccggc

aagatcatcggcgtcgacatcaaccccgaccgcgaggaatggggc

cgcaagttcggcatgaccgacttcctcaacagcaagggcatgagc

cgcgaggacgtcgtcgccaaggtcgtcgccatgaccgacggcggc

gcggactacaccttcgacgccaccggcaacaccgaagtgatgcgc

acggcgcttgaagcctgccatcgcggctggggcacctccatcatc

atcggcgtggccgaggcgggcaaggaaatcagcacgcgtccgttc

cagctcgtcaccggccgcaactggcgcggcacggccttcggcggc

gccaagggccgcaccgacgtgcccaagatcgtcgacatgtacatg

accggcaagatcgagatcgacccgatgatcacccatgtcatgggc

ctggaagagatcaacaccgccttcgacctgatgcacgccggcaag

tcgatccgttcagtcgtggtgttctga

FdhA (Saro_0874) Protein Sequence (SEQ ID NO: 2)

MLSDRHVKGRPHEMKTRAAVAFAPKQPLEIVELDLEGPKAGEVLV

EIMATGVCHTDAYTLDGFDSEGIFPSVLGHEGAGIVREVGPGVTS

VKPGDHVIPLYTPECRQCKSCLSGKTNLCTAIRATQGQGLMPDGT

-continued

SRFSYKGQTVFHYMGCSTFSNFTVLPEIAVAKIREDAPFKTSCYI

GCGVTTGVGAVINTAKVQVGDNVVVFGLGGIGLNVIQGARLAGAG

KIIGVDINPDREEWGRKFGMTDFLNSKGMSREDVVAKVVAMTDGG

ADYTFDATGNTEVMRTALEACHRGWGTSIIIGVAEAGKEISTRPF

QLVTGRNWRGTAFGGAKGRTDVPKIVDMYMTGKIEIDPMITHVMG

LEEINTAFDLMHAGKSIRSVVVF*

Saro_0995 Coding Sequence (SEQ ID NO: 3)

Atgaaagccgccgtactcgtcgaaccgggcaagccgctggatatt

cagcatctcagcgtgtccaagcccggcccgcatgaagtccttatc

cgcaccgcagcctgcgggctgtgccattcggacttgcacttcatc

gaaggtgcctatccccatccgctgcccgcggtgccggggcacgag

gcggcggggatcgtcgaggcggtcggctcggaagtgcgcacggtc

aaggtgggtgacgcggtcgtcacctgcctgtccgcgttctgcggt

cattgcgagttctgcgtgaccggccggatgtcgctgtgccttggc

ggcgacacccggcgcggcgcgggcgaggcacctcgccttacccgc

accgacgacggcagcgccgtgaaccagatgctcaacctctcggcc

tttgccgaacagatgctggtgcacgaacatgcctgcgtggcgatc

aatcccgagatgccgctcgaccgcgcggcggtgatcggctgcgcg

gtcaccactggcgcgggtgcggtgttcaacgcggcgaagctgacc

ccgggcgagacggtctgcgtggtcggctgtggcggcgtcggcctt

gccacggtcaacgccgcgaagatcgccggcgcaggccggatcatc

gcggtggacccgatgccggaaaagcgcgaactggccatgaagctg

ggcgcgaccgatgtgatggacgcgggacccgatgcggcggcacag

atcgtcgagatgacgaaaggcggcgtccaccatgcgatcgaggcc

gtggggcgtccggcatcgggcgaccttgcggtcgcgacgctgcgc

cgcggcggcaccgccacgatccttggcatgatgccgctggcacac

aaggtcggactttccgcgatggacctgctgtcggacaagaagctg

cagggcgccatcatgggccgcaaccacttcccggtggacctgccg

cgcctggtcgacttctacatgcgcggcttgctcgatctcgacacg

atcattgccgaacgcatcccgctcgaagggatcaacgatggcttc

gagaagatgaagcagggccattccgcccgctctgtcatcgtgttc

gaccaatga

Saro_0995 Protein Sequence (SEQ ID NO: 4)

MKAAVLVEPGKPLDIQHLSVSKPGPHEVLIRTAACGLCHSDLHFI

EGAYPHPLPAVPGHEAAGIVEAVGSEVRTVKVGDAVVTCLSAFCG

HCEFCVTGRMSLCLGGDTRRGAGEAPRLTRTDDGSAVNQMLNLSA

FAEQMLVHEHACVAINPEMPLDRAAVIGCAVTTGAGAVENAAKLT

PGETVCVVGCGGVGLATVNAAKIAGAGRIIAVDPMPEKRELAMKL

GATDVMDAGPDAAAQIVEMTKGGVHHAIEAVGRPASGDLAVATLR

-continued

RGGTATILGMMPLAHKVGLSAMDLLSDKKLQGAIMGRNHFPVDLP

RLVDFYMRGLLDLDTIIAERIPLEGINDGFEKMKQGHSARSVIVF

DQ*

Saro_3899 Coding Sequence

(SEQ ID NO: 5)

Atggacgcatacgcggcaattatcgagcgtcaaggcggcgaattc

gttctggataacgtctctatcgaggatccgcgcgacggcgaagtg

ctggtcaaggttgccgcagctggcatgtgtcataccgacctgacg

gttcgcgatcaatattacccgacgccgctgccggcggtgctgggc

catgaaggttcgggcgttgtcgaaaaggtcggacgtggcgtcacc

actgtcaagccaggcgacaaggtcgtgctctccttcagctattgc

ggcacctgtccatcgtgcctcaaggggcatcaggcctattgtccg

agcctgttcccgctcaatttcatgggccgccgcctggatggttcg

acgccgattacccgcaacggccaagaggtcaacgcctgcttcttc

gggcaatcctcgttcgcgacctattcgatcgcgtcggaaaacaac

tgcgtcaaggttgccgacgacgcacagatcgaacttttgggccca

ctgggctgcggcatccagaccggggcgggcagcatcctcaatgcg

ctttgtcccgaacctggctcctcgatcgcgatcttcggggtcggg

tcggtcggcctcagcgccgtgatggccgccaaggcctcgggctgc

ctcaagatcatcgcggttgaccgcaacgcaggccgcttggaactg

gcgcgtgaactgggcgccaccgatgtgatcgacgccaacacggtc

aacgctcaggaagcgatcgtcgcgatgaccggtggcggcgccgac

tatgccatggataccaccgccattccagcggtgctgcgctcggcg

gtggacagcacgcacaacatgggtgaaaccgcagtggtcggcggg

gcgaagctgggcaccgagttttcgctagacatgaacaacatgctg

tttggccgcaagttgcgcggcgtagtcgaaggatcgagcaccccg

caggtcttcatcccgcaactgattgcgatgcagaaggccgggctg

ttcccgttcgagaagctctgcaccttctatgatctcgaccagatc

aaccaggccgtcgaggataccgaaaagaccggcaaggcgatcaag

gccattctcaaaatgtag

Saro_3899 Protein Sequence

(SEQ ID NO: 6)

MDAYAAIIERQGGEFVLDNVSIEDPRDGEVLVKVAAAGMCHTDLT

VRDQYYPTPLPAVLGHEGSGVVEKVGRGVTTVKPGDKVVLSFSYC

GTCPSCLKGHQAYCPSLFPLNFMGRRLDGSTPITRNGQEVNACFF

GQSSFATYSIASENNCVKVADDAQIELLGPLGCGIQTGAGSILNA

LCPEPGSSIAIFGVGSVGLSAVMAAKASGCLKIIAVDRNAGRLEL

ARELGATDVIDANTVNAQEAIVAMTGGGADYAMDTTAIPAVLRSA

VDSTHNMGETAVVGGAKLGTEFSLDMNNMLFGRKLRGVVEGSSTP

QVFIPQLIAMQKAGLFPFEKLCTFYDLDQINQAVEDTEKTGKAIK

AILKM*

-continued

FerD (Saro_0797) Coding Sequence

(SEQ ID NO: 7)

gtgactgcgtacccttcgctccacatgatcatcgacggcgcccgc

gtcagcggcggcggacgtcgcacccacgcggtcgtcaatcccgct

accggagagaccatcggcgaactgccgctggccgaagtcgccgat

ctcgaccgcgcgctcgaagtcgcggcgaagggcttccgcatctgg

cgcgacagcacgccgcagcagcgcgcagccgtgctccagggcgcg

gcccgcctgatgctggaacggcaggaggacctcgcccgcatcgcc

acgatggaagaaggcaagaccctgcccgaggcgcgcatcgaagtc

ctgatgaacgtgggcctgttcaacttctacgccggcgaggtattc

cggctctatggccgcaccctcgtgcgccctgcgggtcagcgcagc

acgatcacgcatgaaccggtcgggcccgtggccgcctttgcgccg

tggaactttccgctcggcaaccccggccgcaagctcggcgcgccc

attgccgccggttgctcggtgatcctcaaggcggcggaagaaacg

ccggcctccgcgctcggggtgctgcaatgcctgctcgatgcgggc

ctgcccaaggaagtggcccaggccgtgttcggtgtgcctgacgag

gtgagtcgccacctgctcggctcgtccgtcatccgcaagctctcg

ttcaccggctcgaccgtcatcggcaagcacctcatgcgccttgcc

gccgacaacatgttgcgcacaacgatggagcttggcggccacggc

cctgtcctcgtcttcggcgatgccgatatcgacaaggcgctcgat

accatggccgcgtccaagtatcgcaacgcgggccaggtctgcgtc

tcgccaacccgcttcatcgtggaagagagcgtgttcgaacgcttc

cgcgacggttttgccgagcgcgtcggccggatcaaggtcggcaac

ggcctcgatcaggatgcgcagatgggccccatggccaacgcccgc

cgccccgaggcgatggatcgcctgatcggggacgccgtgacccgc

ggcgcaaggctccacaccgggggcgagcgcgtcggcaacgccggc

tatttctacgcccccacggtcctgtccgaagtcccgctcgacgcg

gcgatcatgaacgaggagccgttcggcccggtcgcgctgatcaat

cccttcggcggcgaggaagcgatgatcgccgaggccaaccgcctg

ccctacggcctcgccgcctacgcctggaccgacagcgcggcgcgg

gccaagcgcctcgcccgcgagatcgagacggggatgctcgggctt

aactcgaccatgatcggcggcgcggattcgcccttcggcggggtc

aagtggtccggccacggctccgaggacggtcccgaaggcgtcatg

gcctgccttgtcaccaaggcggtccacgaagggtaa

FerD (Saro_0797) Protein Sequence

(SEQ ID NO: 8)

VTAYPSLHMIIDGARVSGGGRRTHAVVNPATGETIGELPLAEVAD

LDRALEVAAKGFRIWRDSTPQQRAAVLQGAARLMLERQEDLARIA

TMEEGKTLPEARIEVLMNVGLFNFYAGEVFRLYGRTLVRPAGQRS

TITHEPVGPVAAPAPWNFPLGNPGRKLGAPIAAGCSVILKAAEET

PASALGVLQCLLDAGLPKEVAQAVFGVPDEVSRHLLGSSVIRKLS

-continued

FTGSTVIGKHLMRLAADNMLRTTMELGGHGPVLVFGDADIDKALD

TMAASKYRNAGQVCVSPTRFIVEESVFERFRDGFAERVGRIKVGN

GLDQDAQMGPMANARRPEAMDRLIGDAVTRGARLHTGGERVGNAG

YFYAPTVLSEVPLDAAIMNEEPFGPVALINPFGGEEAMIAEANRL

PYGLAAYAWTDSAARAKRLAREIETGMLGLNSTMIGGADSPFGGV

KWSGHGSEDGPEGVMACLVTKAVHEG*

Saro_1104 Coding Sequence

(SEQ ID NO: 9)

atgcgcgaacggctacagcaatacattgatggcaagtgggtagac

agcgagggtggcaagcgccacgaggtcatcaatccgacgaccgag

gaaccctgctgcgtcatcacgctgggcacgcaggccgatgtcgac

aaggcagtggccgcggcccagcgcgccttcaagaccttcagcaag

acgacgcgcgaggagcgactcgcgctgcttgaacgcatcgtcgag

gaatacaagaagcgcgtccccgatctcgccgccgcgatggccgag

gaaatgggcgctccggtaagcttcgccagcaccgcgcaggtcggc

gccggcatcggcgccttcctcggcaccatggccgcgctccgcaac

ttctccttcgtcgaggacaacggtgcgttcaaggtcgcctacgaa

ccgatcggcgtcgtcggcatgatcacgccatggaactggccctc

aaccagatcgcgctcaaggtcgcaccggcgctggccgcgggcaac

accatgatcctcaagccgtccgaggaatgccccaccaacgccgcg

atctttaccgagatcctcgatgccgccggcgtcccgccaggcgtc

ttcaacctcatccagggcgatggtcccggcgtcggcactgcgatc

agctcgcacccgggcatcgacatggtcagcttcaccggctcgacc

cgcgcgggcatcctcgtggcgaaggctgcggccgataccgtcaag

cgcgtccatcaggagcttggcggcaagtcgcccaacgtcgtcctg

cccgatgcagacttcgccaagtacctgccgtcgaccgcgtccggc

ccgttggtcaacagcggccagagctgcatttcgcccacccgcatt

ctcgtaccccgcgaacgcgaagccgaagccgcggcgttcgtttcg

gcgatgtactcggcaaccccggtcggcgatccgatgcaggaaggt

gcgcacatcggcccggtggtcaacaaggcgcagttcgacaagatc

cgcggcctgatccagtcggcgatcgacgaaggcgcgaagctcgag

accggcggccccgacctcccggccaacgtcaaccgcggctactac

atcaagcccacggtcttctccggcgtcacgcccgacatgcgcatt

gcgcaggaggaaatcttcggcccggtcgcgacgatcatggcgtac

gacagcctcgaggaggccatcgagatcgccaacgacaccgcctat

ggcctgtcggcctgcatcaccggcgatccggcgaaggcggctgaa

gtcgcgcccgagcttcgcgccggcatggtcgcgatcaacaactgg

ggccccaccccgggcgcgccgttcggcggctacaagcagtccggc

aacggccgcgagggggctctatggcctcaaggacttcatggaaa

tgaaggcgatcagcggcctgcctgcctga

-continued

Saro_1104 Protein Sequence

(SEQ ID NO: 10)

MRERLQQYIDGKWVDSEGGKRHEVINPTTEEPCCVITLGTQADVD

KAVAAAQRAFKTFSKTTREERLALLERIVEEYKKRVPDLAAAMAE

EMGAPVSFASTAQVGAGIGAFLGTMAALRNFSFVEDNGAFKVAYE

PIGVVGMITPWNWPLNQIALKVAPALAAGNTMILKPSEECPTNAA

IFTEILDAAGVPPGVFNLIQGDGPGVGTAISSHPGIDMVSFTGST

RAGILVAKAAADTVKRVHQELGGKSPNVVLPDADFAKYLPSTASG

PLVNSGQSCISPTRILVPREREAEAAAFVSAMYSATPVGDPMQEG

AHIGPVVNKAQFDKIRGLIQSAIDEGAKLETGGPDLPANVNRGYY

IKPTVFSGVTPDMRIAQEEIFGPVATIMAYDSLEEAIEIANDTAY

GLSACITGDPAKAAEVAPELRAGMVAINNWGPTPGAPFGGYKQSG

NGREGGLYGLKDFMEMKAISGLPA*

Saro_1197 Coding Sequence

(SEQ ID NO: 11)

atgactgccccgaccgccgccgacctttccgccgacatcgcacgc

gtcttcgcactccagcaggcgcacatgtgggaggccaaggcctcc

accgcggccgagcgcaaggaaaagctcgcgcgcctcaaggccgcc

gtcgaagcccacgccgacgacatcgtcgccgccgtcctcgaagac

acgcgcaagccggttggcgaaatccgcgtgaccgaagtcctcaac

gtcaccgccaacatccagcgcaacatcgacaatctcgatgaatgg

atgaagccggtcgaggtcgccacctcgctcaatcccgccgaccgc

gcgcagatcatccacgaagcgcgcggcgtctgcctgatccttggc

ccctggaacttccccctcggcctcgcgctcggtccggtcgccgct

gccatcgccgcaggcaacacctgcatcgtgaagctcaccgacctc

tgccccgccaccgcaagggtggcctcggtgatcgtcagggaagcg

ttcgacgaaaaggatgtggctctgttcgaaggcgacgtctcggtc

gccaccgcgctcctcgatctgccgttcaaccacgtcttcttcacc

ggctcgccccgcgtcggcaagatcgtgatggccgctgccgcaaag

cacctcaccagcgtcacgctcgaacttgggggaagtcgcccgtca

tcgtcgacgatagcgccgacatcgatcaggtcgccgcccagctcg

ccgcggccaagcagttcaacgggggcaggcctgcatcagcccgga

ctacgtcttcgtgaaggaagacaagaaggccgcgctggtcgaagg

cttccgggccaacgtgcagaagaacctctatgacgatgccggcaa

cctgaagaaggacagcatcgcccaggtggtcaacaaggcgaactt

cgaccgcgtgaaggccatgttcgacgatgccgtcgccaagggcgc

gaccgtcgccgccggcggaacgttcgaagccgatgacctcaccat

ccatccgaccatgctgaccggcgtcaccccgcagatgaccatcct

ccaggacgaaatcttcgcccccgtcatcccggtgatgacctacga

cacgctcgaccaggcgatcggctacatcgaagcccgcgacaagcc

gctcgcactctatgtctacagcaaggacgaagcgaacgtcgaaaa

-continued

ggtcctcgcccgcacctcgtcgggcggtgtcacggtgaatggcgt

gttctcgcactacctggaaaacaacctgccgttcggcggcgtcaa

caccagcggcatgggcagctaccacggcgtgttcggcttcaagtg

cttcagccacgaacgggctgtctaccgccaccagcagtaa

Saro_1197 Protein Sequence

(SEQ ID NO: 12)

MTAPTAADLSADIARVFALQQAHMWEAKASTAAERKEKLARLKAA

VEAHADDIVAAVLEDTRKPVGEIRVTEVLNVTANIQRNIDNLDEW

MKPVEVATSLNPADRAQIIHEARGVCLILGPWNFPLGLALGPVAA

AIAAGNTCIVKLTDLCPATARVASVIVREAFDEKDVALFEGDVSV

ATALLDLPFNHVFFTGSPRVGKIVMAAAAKHLTSVTLELGGKSPV

IVDDSADIDQVAAQLAAAKQFNGGQACISPDYVFVKEDKKAALVE

GFRANVQKNLYDDAGNLKKDSIAQVVNKANFDRVKAMEDDAVAKG

ATVAAGGTFEADDLTIHPTMLTGVTPQMTILQDEIFAPVIPVMTY

DTLDQAIGYIEARDKPLALYVYSKDEANVEKVLARTSSGGVTVNG

VFSHYLENNLPFGGVNTSGMGSYHGVFGFKCFSHERAVYRHQQ*

Saro_2869 Coding Sequence

(SEQ ID NO: 13)

atgaacgacatgaccaccatctcgcgcacgcagcgcgaatactcg

gaggccgccaaggccttcctcgcgcgcaagccgcagttgttcatc

aacaacgagtgggtcgacagcagccacgacgccgtgatcgaggtg

gaagacccctcgaacggcaggatcgtcggtcatgtcgtcgatgcc

tcggacaaggacgtcgaccgggcggttgccgcggcgcgcgccgcg

ttcgacgatggccgctggtccaacctgccgccaatggtccgcgat

cgcaccatgaatcgcctggccgacctgcttgaagccaacgccgat

ctctttgccgagctcgaagcgatcgacaacggcaagcccaagggc

atggccggcgccgtcgacatccccggcgcgatcagccagctccgc

ttcatggccggctgggccagcaaggtcgcgggcgagacgacgcag

ccctacacgatgcccaacggcaccgtgttcagctacaccgtcaag

gaacccgtcggcgtctgcgcgcagatcgtgccgtggaacttcccg

ctgctgatggcctcgctcaagatcgccccggcgctggcggctggc

tgcaccctggtgctgaagcccgccgaacagacctcgcttaccgcg

ctcaagcttgccgatctcgtggtcgaggccggcttccctgcgggc

gtgatcaacatcatcaccggcaacggccacaccgccggtgaccgc

atggtcaagcatcccgacgtcgacaaggtcgccttcaccggctcg

accgagatcggcaagctgatcaatcgcaacgccaccaccacgctc

aagcgggtcacgctcgaactgggggggaagagccccgtcgtggtca

tgcccgacgtcgacgtggcgcagaccgcgcctggcgttgccggcg

cgatcttcttcaacgcgggccaggtctgcgttgccggttcgcgtc

tctatgcgcaccgttcggtgttcgattccgtgctcgaaggcatga

cccagaccgcgccgttctgggcgccgcgcccctcgctggatcccg

-continued

aagcccacatgggcccgttggtcagcaaggagcagcacgaccgcg

tgatgggctacatcgaggcgggcaagcgcgatggcgccagcgtcg

tcatgggcggcgattgccccagcgccgatggcgggtactacgtca

acccgacgatccttgcagacgtgaacccgcagatgtcggtcgtgc

gcgaggaaatcttcggccccgtcgtcgtcgcccagcgcttcgacg

atctcgatgaagtggcgaagatggccaacgacacctgcttcggcc

tcggcgcgggcgtgtggacgcgcgatgtcgcggtgatgcacaagc

ttgcctcgaagatcaaatcgggcaccgtgtggggcaactgccacg

ccctgatcgataccgcgctgcccttggcggctacaaggaatcgg

gcctgggccgcgaacaggggcgcgccggcatcgacgcctacctcg

agaccaagaccgtcatcatccagatgtaa

Saro_2869 Protein Sequence

(SEQ ID NO: 14)

MNDMTTISRTQREYSEAAKAFLARKPQLFINNEWVDSSHDAVIEV

EDPSNGRIVGHVVDASDKDVDRAVAAARAAFDDGRWSNLPPMVRD

RTMNRLADLLEANADLFAELEAIDNGKPKGMAGAVDIPGAISQLR

FMAGWASKVAGETTQPYTMPNGTVFSYTVKEPVGVCAQIVPWNFP

LLMASLKIAPALAAGCTLVLKPAEQTSLTALKLADLVVEAGFPAG

VINIITGNGHTAGDRMVKHPDVDKVAFTGSTEIGKLINRNATTTL

KRVTLELGGKSPVVVMPDVDVAQTAPGVAGAIFFNAGQVCVAGSR

LYAHRSVFDSVLEGMTQTAPFWAPRPSLDPEAHMGPLVSKEQHDR

VMGYIEAGKRDGASVVMGGDCPSADGGYYVNPTILADVNPQMSVV

REEIFGPVVVAQRFDDLDEVAKMANDTCFGLGAGVWTRDVAVMHK

LASKIKSGTVWGNCHALIDTALPEGGYKESGLGREQGRAGIDAYL

ETKTVIIQM*

PcfL (Saro_0796) Coding Sequence

(SEQ ID NO: 15)

Gtgtccgatagcaatcagattgccgcgctcgaaagccgcctgaac

gacctcgaaaggcgcctgacggtgcgcgaggacgagctggacgta

cgcaagctccagcatctctacggctacctgatcgacaagtgcatg

tataacgagaccgtggacctgttcaccgaagatggcgaagtgcgc

ttcttcggcggcgtctggaagggcaaggagggcatccgccgtctc

tacgtcgaacgtttccagaagcgcttcacctacggcaacaacggc

ccgatcgacggcttcctgctcgatcacccccagcttcaggacatc

atccacgtgcaggatgacggggtcaccgctctcggccgcgcgcgg

tcgatgatgcaggccggtcgccacaaggattacgagggcgatgcc

ccgcacctcaaggcgcgccagtggtgggaaggcggcatctacgag

aacacctacaagaaggggacggcgtgtggcggatgcacatcctca

actacatgccgatctggcacgccgatttcgaaagcggctgggcca

acaccccgcacgaatacgtgccgttccccaaggtcacctatcccg

aagacccgaccggaccggacgaactgatcgccgaccactggctct

-continued

ggccgacccacaagctgaacccttccacatgaagcacccggtga

cgggcgaggaaatggtcgcgcagcgctggcagggcgacatcgacc

gcgagaacgcgcggaaataa

PcfL (Saro_0796) Protein Sequence

(SEQ ID NO: 16)

VSDSNQIAALESRLNDLERRLTVREDELDVRKLQHLYGYLIDKCM

YNETVDLFTEDGEVRFFGGVWKGKEGIRRLYVERFQKRFTYGNNG

PIDGFLLDHPQLQDIIHVQDDGVTALGRARSMMQAGRHKDYEGDA

PHLKARQWWEGGIYENTYKKVDGVWRMHILNYMPIWHADFESGWA

NTPHEYVPFPKVTYPEDPTGPDELIADHWLWPTHKLNPFHMKHPV

TGEEMVAQRWQGDIDRENARK*

LsdD (Saro_0802) Coding Sequence

(SEQ ID NO: 17)

atggcccaatttccgaacacccccagcttcacgggattcaacacg

ccgtcgcggatcgaggcggatatcgccgatctggcccacgaaggc

acgattccgcaagggttaaacggcgcattctaccgcgtccagccc

gacccgcagtttcctccccgcctcgacgacgacatcgccttcaac

ggcgacggcatgatcacccgcttccacatccacgacggccaggtc

gacttccgccagcgctgggcgaagaccgacaagtggaagctggag

aacgccgccggaaaggccctgttcggcgcctaccgcaacccgctg

accgacgacgaggcggtcaagggcgagatccgttcgaccgccaac

accaacgccttcgtgttcggcggcaagctgtgggcgatgaaggag

gacagtcccgccctcgtcatggacccggcgacgatggaaaccttc

gggttcgagaagttcggcggcaagatgaccggccagacctttacc

gcccaccccaaggtcgatccgaagaccggcaacatggtcgccatc

ggctatgccgcaagcgggctgtgcaccgacgatgtgacctacatg

gaagtgagcccggagggcgagcttgtccgcgaagtgtggttcaag

gtgccgtactactgcatgatgcacgacttcggcatcaccgaggat

tacctcgtgctgcacatcgtgccttccatcggaagctgggaaagg

ctggaacagggcaagccgcacttcggcttcgacacgaccatgccg

gtgcacctcggcatcatcccgcgccgcgacggcgtgcgccaggaa

gacatccgctggttcacgcgggacaactgctttgccagccatgtc

ctgaacgcctggcaagaggggaccaagatccacttcgtgacctgc

gaggcgaagaacaacatgttcccgttcttccccgacgtccacggc

gcgcccttcaacggcatggaggccatgagccatccgaccgactgg

gtggtcgacatggccagcaacggcgaggactttgccgggatcgtg

aagctttccgacacagccgccgagttcccgcgcatcgacgaccgc

tttaccggccagaagacccgccatggctggttcctcgaaatggac

atgaagcgcccggtggaattgcgcggcggcagcgccggcggcctg

ctgatgaactgcctgttccacaaggacttcgaaacgggtcgcgag

cagcactggtggtgcggcccggtgtcgagccttcaggagccgtgc

-continued

ttcgtgccgcgcgccaaggatgcccccgaaggcgacggctggatc

gtgcaggtttgcaaccggctggaagagcagcgcagcgacttgctg

atcttcgacgcgctcgacatcgagaaaggcccggtggccacggtc

aacatccccatccgcctgcgcttcggccttcacggcaactgggcg

aatgccgacgaaatcggccttgccgagaaggtcctggccgcatga

LsdD (Saro_0802) Protein Sequence

(SEQ ID NO: 18)

MAQFPNTPSFTGFNTPSRIEADIADLAHEGTIPQGLNGAFYRVQP

DPQFPPRLDDDIAFNGDGMITRFHIHDGQVDFRQRWAKTDKWKLE

NAAGKALFGAYRNPLTDDEAVKGEIRSTANTNAFVFGGKLWAMKE

DSPALVMDPATMETFGFEKFGGKMTGQTFTAHPKVDPKTGNMVAI

GYAASGLCTDDVTYMEVSPEGELVREVWFKVPYYCMMHDFGITED

YLVLHIVPSIGSWERLEQGKPHFGFDTTMPVHLGIIPRRDGVRQE

DIRWFTRDNCFASHVLNAWQEGTKIHFVTCEAKNNMFPFFPDVHG

APFNGMEAMSHPTDWVVDMASNGEDFAGIVKLSDTAAEFPRIDDR

FTGQKTRHGWFLEMDMKRPVELRGGSAGGLLMNCLFHKDFETGRE

QHWWCGPVSSLQEPCFVPRAKDAPEGDGWIVQVCNRLEEQRSDLL

IFDALDIEKGPVATVNIPIRLRFGLHGNWANADEIGLAEKVLAA*

LigW (Saro_0799) Coding Sequence

(SEQ ID NO: 19)

atgacacaagaccttaagaccggcggcgagcagggctacctgcgc

atcgccaccgaggaagccttcgccacgcgcgagatcatcgacgtc

tacctgcgcatgatccgcgatggcactgccgacaagggcatggtc

tcgctctggggcttctacgcccagtccccctcagagcgcgccacc

cagatcctcgaacgcctgctcgatcttggcgagcgccgcatcgcc

gacatggacgcgaccggcatcgacaaggctatcctcgcgctgacc

tcgcccggcgtccagccgctgcacgaccttgacgaggccaggacg

ctcgccacccgcgccaacgacacgcttgccgacgcgtgccaaaag

tacccagaccgcttcatcggcatgggcaccgtcgccccgcaggac

ccggaatggtccgcgcgcgagatccatcgtggtgccagggaactg

ggcttcaagggcatccagatcaacagccacacgcaagggcgctac

ctcgacgaggagttcttcgacccgatcttccgcgccctcgttgaa

gtcgaccagccgctctacatccaccctgccacttcgcccgattcc

atgatcgacccgatgctcgaagcgggcctcgacggcgccatcttc

ggcttcggcgtggagacgggcatgcacctgctgcgcctcatcacc

atcggcatcttcgacaagtatcccagccttcagatcatggtcggc

cacatgggcgaggcgctgccctactggctctaccgcctggactac

atgcaccaggccggtgtccgctcgcagcgctacgaacgcatgaag

cccctgaagaagaccatcgagggctacctcaagtccaacgtcctc

gtcaccaattcgggcgtcgcgtgggaacctgcgatcaagttctgc

cagcaggtcatgggcgaggaccgcgttatgtacgcgatggactac

-continued

ccctaccagtacgttgccgacgaggtgcgcgcgatggacgccatg

gacatgagtgcgcaaacgaagaagaagttcttccagaccaacgcg

gagaagtggttcaagctttga

LigW (Saro_0799) Protein Sequence (SEQ ID NO: 20)

MTQDLKTGGEQGYLRIATEEAFATREIIDVYLRMIRDGTADKGMV

SLWGFYAQSPSERATQILERLLDLGERRIADMDATGIDKAILALT

SPGVQPLHDLDEARTLATRANDTLADACQKYPDRFIGMGTVAPQD

PEWSAREIHRGARELGFKGIQINSHTQGRYLDEEFFDPIFRALVE

VDQPLYIHPATSPDSMIDPMLEAGLDGAIFGFGVETGMHLLRLIT

IGIFDKYPSLQIMVGHMGEALPYWLYRLDYMHQAGVRSQRYERMK

PLKKTIEGYLKSNVLVTNSGVAWEPAIKFCQQVMGEDRVMYAMDY

PYQYVADEVRAMDAMDMSAQTKKKFFQTNAEKWFKL*

EXEMPLARY VERSIONS OF THE INVENTION

[0232] 1. A recombinant microorganism comprising any one or more, any two or more, any three or more, any four or more, or each of:

[0233] one or more recombinant alcohol dehydrogenase genes encoding:

[0234] FdhA of *Novosphingobium aromaticivorans* (SEQ ID NO:2) or a homolog thereof;

[0235] Saro_0995 of *Novosphingobium aromaticivorans* (SEQ ID NO:4) or a homolog thereof; and/or

[0236] Saro_3899 of *Novosphingobium aromaticivorans* (SEQ ID NO:6) or a homolog thereof;

[0237] one or more recombinant aldehyde dehydrogenase genes encoding:

[0238] FerD of *Novosphingobium aromaticivorans* (SEQ ID NO:8) or a homolog thereof;

[0239] Saro_1104 of *Novosphingobium aromaticivorans* (SEQ ID NO:10) or a homolog thereof;

[0240] Saro_1197 of *Novosphingobium aromaticivorans* (SEQ ID NO:12) or a homolog thereof; and/or

[0241] Saro_2869 of *Novosphingobium aromaticivorans* (SEQ ID NO:14) or a homolog thereof;

[0242] a recombinant γ-formaldehyde lyase gene encoding PcfL of *Novosphingobium aromaticivorans* (SEQ ID NO:16) or a homolog thereof;

[0243] a recombinant lignostilbene dioxygenase gene encoding LsdD of *Novosphingobium aromaticivorans* (SEQ ID NO:18) or a homolog thereof; and

[0244] a recombinant aromatic acid decarboxylase gene encoding LigW of *Novosphingobium aromaticivorans* (SEQ ID NO:20) or a homolog thereof.

[0245] 2. The recombinant microorganism of version 1, comprising any two or more, any three or more, any four or more, or each of:

[0246] the one or more recombinant alcohol dehydrogenase genes;

[0247] the one or more recombinant aldehyde dehydrogenase genes;

[0248] the recombinant γ-formaldehyde lyase gene;

[0249] the recombinant lignostilbene dioxygenase gene; and

[0250] the recombinant aromatic acid decarboxylase gene.

[0251] 3. The recombinant microorganism of version 1, comprising any three or more, any four or more, or each of:

[0252] the one or more recombinant alcohol dehydrogenase genes;

[0253] the one or more recombinant aldehyde dehydrogenase genes;

[0254] the recombinant γ-formaldehyde lyase gene;

[0255] the recombinant lignostilbene dioxygenase gene; and

[0256] the recombinant aromatic acid decarboxylase gene.

[0257] 4. The recombinant microorganism of version 1, comprising any four or more or each of:

[0258] the one or more recombinant alcohol dehydrogenase genes;

[0259] the one or more recombinant aldehyde dehydrogenase genes;

[0260] the recombinant γ-formaldehyde lyase gene;

[0261] the recombinant lignostilbene dioxygenase gene; and

[0262] the recombinant aromatic acid decarboxylase gene.

[0263] 5. The recombinant microorganism of version 1, comprising each of:

[0264] the one or more recombinant alcohol dehydrogenase genes;

[0265] the one or more recombinant aldehyde dehydrogenase genes;

[0266] the recombinant γ-formaldehyde lyase gene;

[0267] the recombinant lignostilbene dioxygenase gene; and

[0268] the recombinant aromatic acid decarboxylase gene.

[0269] 6. The recombinant microorganism of any prior version, comprising the one or more recombinant alcohol dehydrogenase genes.

[0270] 7. The recombinant microorganism of any prior version, wherein, when present, the one or more recombinant alcohol dehydrogenase genes encode:

[0271] FdhA of *Novosphingobium aromaticivorans* (SEQ ID NO:2), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:2, an ortholog of FdhA of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of FdhA of *Novosphingobium aromaticivorans;*

[0272] Saro_0995 of *Novosphingobium aromaticivorans* (SEQ ID NO:4), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:4, an ortholog of Saro_0995 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_0995 of *Novosphingobium aromaticivorans*; and/or

[0273] Saro_3899 of *Novosphingobium aromaticivorans* (SEQ ID NO:6), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:6, an ortholog of Saro_3899 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_3899 of *Novosphingobium aromaticivorans.*

[0274] 8. The recombinant microorganism of any prior version comprising the one or more recombinant aldehyde dehydrogenase genes.

[0275] 9. The recombinant microorganism of any prior version, wherein, when present, the one or more recombinant aldehyde dehydrogenase genes encode:

[0276] FerD of *Novosphingobium aromaticivorans* (SEQ ID NO:8), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:8, an ortholog of FerD of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of FerD of *Novosphingobium aromaticivorans;*

[0277] Saro_1104 of *Novosphingobium aromaticivorans* (SEQ ID NO:10), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:10, an ortholog of Saro_1104 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_1104 of *Novosphingobium aromaticivorans;*

[0278] Saro_1197 of *Novosphingobium aromaticivorans* (SEQ ID NO:12), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:12, an ortholog of Saro_1197 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_1197 of *Novosphingobium aromaticivorans*; and/or

[0279] Saro_2869 of *Novosphingobium aromaticivorans* (SEQ ID NO:14), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:14, an ortholog of Saro_2869 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_2869 of *Novosphingobium aromaticivorans.*

[0280] 10. The recombinant microorganism of any prior version, comprising the recombinant 7-formaldehyde lyase gene.

[0281] 11. The recombinant microorganism of any prior version, wherein, when present, the recombinant γ-formaldehyde lyase gene encodes PcfL of *Novosphingobium aromaticivorans* (SEQ ID NO:16), a protein comprising a sequence at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identical to SEQ ID NO:16, an ortholog of PcfL of *Novosphingobium aromaticivorans*, a recombinant variant of the ortholog of PcfL of *Novosphingobium aromaticivorans.*

[0282] 12. The recombinant microorganism of any prior version, comprising the recombinant lignostilbene dioxygenase gene.

[0283] 13. The recombinant microorganism of any prior version, wherein, when present, the recombinant lignostilbene dioxygenase gene encodes LsdD of *Novosphingobium aromaticivorans* (SEQ ID NO:18), a protein comprising a sequence at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, or at least 99% identical to SEQ ID NO:18, an ortholog of LsdD of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of LsdD of *Novosphingobium aromaticivorans.*

[0284] 14. The recombinant microorganism of any prior version, comprising the recombinant aromatic acid decarboxylase gene.

[0285] a recombinant aromatic acid decarboxylase gene encoding LigW of *Novosphingobium aromaticivorans* (SEQ ID NO:20) or a homolog thereof.

[0286] 15. The recombinant microorganism of any prior version, wherein, when present, the recombinant aromatic acid decarboxylase gene encodes LigW of *Novosphingobium aromaticivorans* (SEQ ID NO:20), a protein comprising a sequence at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, or at least 99% identical to SEQ ID NO:20, an ortholog of LigW of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of LigW of *Novosphingobium aromaticivorans.*

[0287] 16. The recombinant microorganism of any prior version, wherein the recombinant microorganism is a bacterium.

[0288] 17. The recombinant microorganism of any prior version, wherein the recombinant microorganism is an Alphaproteobacterium.

[0289] 18. The recombinant microorganism of any prior version, wherein the recombinant microorganism is from an order selected from the group consisting of Sphingomonadales, *Actinomyces*, Gammaproteobacteria, Betaproteobacteria, and Bacilli.

[0290] 19. A method of catabolizing a lignin aromatic, the method comprising culturing the recombinant microorganism of any prior version in a medium comprising the lignin aromatic to thereby catabolize the lignin aromatic.

[0291] 20. The method of version 19, wherein the lignin aromatic comprises a β-5 linked lignin aromatic.

[0292] 21. The method of any one of versions 19-20, wherein the lignin aromatic comprises one or more of dehydrodiconiferyl alcohol (DC-A), dehydrodiconiferyl aldehyde (DC-L), dehydrodiconiferyl carboxylic acid (DC-C), dehydrodiconiferyl stilbene carboxylic acid (DC-S-C), 5-formyl ferulate (5-FF), 5-carboxyferulate (5-CF), and 4-hydroxyphenyl and syringyl analogs thereof.

SEQUENCE LISTING

Sequence total quantity: 110
SEQ ID NO: 1 moltype = DNA length = 1152
FEATURE Location/Qualifiers
source 1..1152
mol_type = genomic DNA
organism = Novosphingobium aromaticivorans
SEQUENCE: 1
atgctatcgg accgccacgt caaagggaga ccgcacgaaa tgaagacccg cgccgcagtt 60
gcgttcgcgc ccaagcagcc gctcgagatc gtcgaactgg acctcgaagg ccccaaggct 120
ggcgaagtgc tggtcgagat catggcgacc ggcgtgtgcc acaccgatgc ctacacgctc 180
gacgggttcg acagcgaagg catcttcccc agcgtgctgg gccacgaagg cgccggtatc 240
gtgcgcgagg tgggccctgg ggtcacttcg gtgaagcccg gcgatcacgt gatcccgctc 300
tacacgccgg aatgccgcca gtgcaaatcg tgcctctcgg gcaagaccaa cctgtgcacc 360

-continued

```
gcgatccgcg ccacgcaagg gcagggcctg atgcccgacg gcaccagccg cttttcgtac  420
aagggccaga ccgtgttcca ctacatgggc tgctcgacct tctctaactt caccgtcctg  480
cccgagatcg cggttgccaa gatccgcgag gacgcgccgt tcaagacctc gtgctatatc  540
ggctgcggcg tgacgacggg cgtcggcgcg gtgatcaaca ccgccaaggt ccaggtcggt  600
gacaacgtcg tggtcttcgg cctcggcggc atcggcctca acgtgatcca gggcgcgcgg  660
cttgccggtg ccggcaagat catcggcgtc gacatcaacc ccgaccgcga ggaatggggc  720
cgcaagttcg gcatgaccga cttcctcaac agcaagggca tgagccgcga ggacgtcgtc  780
gccaaggtcg tcgccatgac cgacggcggc gcggactaca ccttcgacgc caccggcaac  840
accgaagtga tgcgcacggc gcttgaagcc tgccatcgcg gctggggcac ctccatcatc  900
atcggcgtgg ccgaggcggg caaggaaatc agcacgcgtc cgttccagct cgtcaccggc  960
cgcaactggc gcggcacggc cttcggcggc gccaagggcc gcaccgacgt gcccaagatc 1020
gtcgacatgt acatgaccgg caagatcgag atcgacccga tgatcaccca tgtcatgggc 1080
ctggaagaga tcaacaccgc cttcgacctg atgcacgccg gcaagtcgat ccgttcagtc 1140
gtggtgttct ga                                                     1152

SEQ ID NO: 2           moltype = AA  length = 383
FEATURE                Location/Qualifiers
source                 1..383
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 2
MLSDRHVKGR PHEMKTRAAV AFAPKQPLEI VELDLEGPKA GEVLVEIMAT GVCHTDAYTL  60
DGFDSEGIFP SVLGHEGAGI VREVGPGVTS VKPGDHVIPL YTPECRQCKS CLSGKTNLCT  120
AIRATQGQGL MPDGTSRFSY KGQTVFHYMG CSTFSNFTVL PEIAVAKIRE DAPFKTSCYI  180
GCGVTTGVGA VINTAKVQVG DNVVVFGLGG IGLNVIQGAR LAGAGKIIGV DINPDREEWG  240
RKFGMTDFLN SKGMSREDVV AKVVAMTDGG ADYTFDATGN TEVMRTALEA CHRGWGTSII  300
IGVAEAGKEI STRPFQLVTG RNWRGTAFGG AKGRTDVPKI VDMYMTGKIE IDPMITHVMG  360
LEEINTAFDL MHAGKSIRSV VVF                                          383

SEQ ID NO: 3           moltype = DNA  length = 1089
FEATURE                Location/Qualifiers
source                 1..1089
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 3
atgaaagccg ccgtactcgt cgaaccgggc aagccgctgg atattcagca tctcagcgtg  60
tccaagcccg gcccgcatga agtccttatc cgcaccgcag cctgcgggct gtgccattcg  120
gacttgcact tcatcgaagg tgcctatccc catccgctgc ccgcggtgcc ggggcacgag  180
gcggcgggga tcgtcgaggc ggtcggctcg gaagtgcgca cggtcaaggt gggtgacgcg  240
gtcgtcacct gcctgtccgc gttctgcggt cattgcgagt tctgcgtgac cggccggatg  300
tcgctgtgcc ttggcggcga cacccggcgc ggcgcgggcg aggcacctcg ccttacccgc  360
accgacgacg gcagcgccgt gaaccagatg ctcaacctct cggcctttgc cgaacagatg  420
ctggtgcacg aacatgcctg cgtggcgatc aatcccgaga tgccgctcga ccgcgcggcg  480
gtgatcggct gcgcggtcac cactggcgcg ggtgcggtgt tcaacgcggc gaagctgacc  540
ccgggcgaga cggtctgcgt ggtcggctgt ggcggcgtcg gccttgccac ggtcaacgcc  600
gcgaagatcg ccggcgcagg ccggatcatc gcggtggacc cgatgccgga aaagcgcgaa  660
ctggccatga agctgggcgc gaccgatgtg atggacgcgg gacccgatgc ggcggcacag  720
atcgtcgaga tgacgaaagg cggcgtccac catgcgatcg aggccgtggg gcgtccggca  780
tcgggcgacc ttgcggtcgc gacgctgcgc cgcggcggca ccgccacgat ccttggcatg  840
atgccgctgg cacacaaggt cggactttcc gcgatggacc tgctgtcgga caagaagctg  900
cagggcgcca tcatgggccg caaccacttc ccggtggacc tgccgcgcct ggtcgacttc  960
tacatgcgcg gcttgctcga tctcgacacg atcattgccg aacgcatccc gctcgaaggg 1020
atcaacgatg gcttcgagaa gatgaagcag ggccattccg cccgctctgt catcgtgttc 1080
gaccaatga                                                         1089

SEQ ID NO: 4           moltype = AA  length = 362
FEATURE                Location/Qualifiers
source                 1..362
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 4
MKAAVLVEPG KPLDIQHLSV SKPGPHEVLI RTAACGLCHS DLHFIEGAYP HPLPAVPGHE  60
AAGIVEAVGS EVRTVKVGDA VVTCLSAFCG HCEFCVTGRM SLCLGGDTRR GAGEAPRLTR  120
TDDGSAVNQM LNLSAFAEQM LVHEHACVAI NPEMPLDRAA VIGCAVTTGA GAVFNAAKLT  180
PGETVCVVGC GGVGLATVNA AKIAGAGRII AVDPMPEKRE LAMKLGATDV MDAGPDAAAQ  240
IVEMTKGGVH HAIEAVGRPA SGDLAVATLR RGGTATILGM MPLAHKVGLS AMDLLSDKKL  300
QGAIMGRNHF PVDLPRLVDF YMRGLLDLDT IIAERIPLEG INDGFEKMKQ GHSARSVIVF  360
DQ                                                                 362

SEQ ID NO: 5           moltype = DNA  length = 1098
FEATURE                Location/Qualifiers
source                 1..1098
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 5
atggacgcat acgcggcaat tatcgagcgt caaggcggcg aattcgttct ggataacgtc  60
```

-continued

```
tctatcgagg atccgcgcga cggcgaagtg ctggtcaagg ttgccgcagc tggcatgtgt  120
cataccgacc tgacggttcg cgatcaatat tacccgacgc cgctgccggc ggtgctgggc  180
catgaaggtt cgggcgttgt cgaaaaggtc ggacgtggcg tcaccactgt caagccaggc  240
gacaaggtcg tgctctcctt cagctattgc ggcacctgtc catcgtgcct caaggggcat  300
caggcctatt gtccgagcct gttcccgctc aatttcatgg gccgccgcct ggatggttcg  360
acgccgatta cccgcaacgg ccaagaggtc aacgcctgct tcttcgggca atcctcgttc  420
gcgacctatt cgatcgcgtc ggaaaacaac tgcgtcaagg ttgccgacga cgcacagatc  480
gaacttttgg gcccactggg ctgcggcatc cagaccgggg cgggcagcat cctcaatgcg  540
ctttgtcccg aacctggctc ctcgatcgcg atcttcgggg tcgggtcggt cggcctcagc  600
gccgtgatgg ccgccaaggc ctcgggctgc ctcaagatca tcgcggttga ccgcaacgca  660
ggccgcttgg aactggcgcg tgaactgggc gccaccgatg tgatcgacgc caacacggtc  720
aacgctcagg aagcgatcgt cgcgatgacc ggtggcggcg ccgactatgc catggatacc  780
accgccattc cagcggtgct gcgctcggcg gtggacagca cgcacaacat gggtgaaacc  840
gcagtggtcg gcggggcgaa gctgggcacc gagttttcgc tagacatgaa caacatgctg  900
tttggccgca agttgcgcgg cgtagtcgaa ggatcgagca ccccgcaggt cttcatcccg  960
caactgattg cgatgcagaa ggccgggctg ttcccgttcg agaagctctg caccttctat 1020
gatctcgacc agatcaacca ggccgtcgag gataccgaaa agaccggcaa ggcgatcaag 1080
gccattctca aaatgtag                                              1098

SEQ ID NO: 6           moltype = AA  length = 365
FEATURE                Location/Qualifiers
source                 1..365
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 6
MDAYAAIIER QGGEFVLDNV SIEDPRDGEV LVKVAAAGMC HTDLTVRDQY YPTPLPAVLG  60
HEGSGVVEKV GRGVTTVKPG DKVVLSFSYC GTCPSCLKGH QAYCPSLFPL NFMGRRLDGS  120
TPITRNGQEV NACFFGQSSF ATYSIASENN CVKVADDAQI ELLGPLGCGI QTGAGSILNA  180
LCPEPGSSIA IFGVGSVGLS AVMAAKASGC LKIIAVDRNA GRLELARELG ATDVIDANTV  240
NAQEAIVAMT GGGADYAMDT TAIPAVLRSA VDSTHNMGET AVVGGAKLGT EFSLDMNNML  300
FGRKLRGVVE GSSTPQVFIP QLIAMQKAGL FPFEKLCTFY DLDQINQAVE DTEKTGKAIK  360
AILKM                                                            365

SEQ ID NO: 7           moltype = DNA  length = 1431
FEATURE                Location/Qualifiers
source                 1..1431
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 7
gtgactgcgt acccttcgct ccacatgatc atcgacggcg cccgcgtcag cggcggcgga  60
cgtcgcaccc acgcggtcgt caatcccgct accggagaga ccatcggcga actgccgctg  120
gccgaagtcg ccgatctcga ccgcgcgctc gaagtcgcgg cgaagggctt ccgcatctgg  180
cgcgacagca cgccgcagca gcgcgcagcc gtgctccagg gcgcggcccg cctgatgctg  240
gaacggcagg aggacctcgc ccgcatcgcc acgatggaag aaggcaagac cctgcccgag  300
gcgcgcatcg aagtcctgat gaacgtgggc ctgttcaact tctacgccgg cgaggtattc  360
cggctctatg gccgcaccct cgtgcgccct gcgggtcagc gcagcacgat cacgcatgaa  420
ccggtcgggc ccgtggccgc ctttgcgccg tggaactttc cgctcggcaa ccccggccgc  480
aagctcggcg cgcccattgc cgccggttgc tcggtgatcc tcaaggcggc ggaagaaacg  540
ccggcctccg cgctcggggt gctgcaatgc ctgctcgatg cgggcctgcc caaggaagtg  600
gcccaggccg tgttcggtgt gcctgacgag gtgagtcgcc acctgctcgg ctcgtccgtc  660
atccgcaagc tctcgttcac cggctcgacc gtcatcggca agcacctcat gcgccttgcc  720
gccgacaaca tgttgcgcac aacgatggag cttggcggcc acggccctgt cctcgtcttc  780
ggcgatgccg atatcgacaa ggcgctcgat accatggccg cgtccaagta tcgcaacgcg  840
ggccaggtct gcgtctcgcc aacccgcttc atcgtggaag agagcgtgtt cgaacgcttc  900
cgcgacggtt ttgccgagcg cgtcggccgg atcaaggtcg gcaacggcct cgatcaggat  960
gcgcagatgg gccccatggc caacgcccgc cgccccgagg cgatggatcg cctgatcggg 1020
gacgccgtga cccgcggcgc aaggctccac accgggggcg agcgcgtcgg caacgccggc 1080
tatttctacg cccccacggt cctgtccgaa gtcccgctcg acgcggcgat catgaacgag 1140
gagccgttcg gcccggtcgc gctgatcaat cccttcggcg gcgaggaagc gatgatcgcc 1200
gaggccaacc gcctgcccta cggcctcgcc gcctacgcct ggaccgacag cgcggcgcgg 1260
gccaagcgcc tcgcccgcga gatcgagacg gggatgctcg ggcttaactc gaccatgatc 1320
ggcggcgcgg attcgccctt cggcggggtc aagtggtccg gccacggctc cgaggacggt 1380
cccgaaggcg tcatggcctg ccttgtcacc aaggcggtcc acgaagggta a          1431

SEQ ID NO: 8           moltype = AA  length = 476
FEATURE                Location/Qualifiers
source                 1..476
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 8
VTAYPSLHMI IDGARVSGGG RRTHAVVNPA TGETIGELPL AEVADLDRAL EVAAKGFRIW  60
RDSTPQQRAA VLQGAARLML ERQEDLARIA TMEEGKTLPE ARIEVLMNVG LFNFYAGEVF  120
RLYGRTLVRP AGQRSTITHE PVGPVAAFAP WNFPLGNPGR KLGAPIAAGC SVILKAAEET  180
PASALGVLQC LLDAGLPKEV AQAVFGVPDE VSRHLLGSSV IRKLSFTGST VIGKHLMRLA  240
ADNMLRTTME LGGHGPVLVF GDADIDKALD TMAASKYRNA GQVCVSPTRF IVEESVFERF  300
RDGFAERVGR IKVGNGLDQD AQMGPMANAR RPEAMDRLIG DAVTRGARLH TGGERVGNAG  360
```

-continued

```
YFYAPTVLSE VPLDAAIMNE EPFGPVALIN PFGGEEAMIA EANRLPYGLA AYAWTDSAAR  420
AKRLAREIET GMLGLNSTMI GGADSPFGGV KWSGHGSEDG PEGVMACLVT KAVHEG      476

SEQ ID NO: 9           moltype = DNA  length = 1425
FEATURE                Location/Qualifiers
source                 1..1425
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 9
atgcgcgaac ggctacagca atacattgat ggcaagtggg tagacagcga gggtggcaag  60
cgccacgagg tcatcaatcc gacgaccgag gaaccctgct gcgtcatcac gctgggcacg  120
caggccgatg tcgacaaggc agtggccgcg gcccagcgcg ccttcaagac cttcagcaag  180
acgacgcgcg aggagcgact cgcgctgctt gaacgcatcg tcgaggaata caagaagcgc  240
gtccccgatc tcgccgccgc gatggccgag gaaatgggcg ctccggtaag cttcgccagc  300
accgcgcagg tcggcgccgg catcggcgcc ttcctcggca ccatggccgc gctccgcaac  360
ttctccttcg tcgaggacaa cggtgcgttc aaggtcgcct acgaaccgat cggcgtcgtc  420
ggcatgatca cgccatggaa ctggcccctc aaccagatcg cgctcaaggt cgcaccggcg  480
ctggccgcgg gcaacaccat gatcctcaag ccgtccgagg aatgccccac caacgccgcg  540
atctttaccg agatcctcga tgccgccggc gtcccgccag gcgtcttcaa cctcatccag  600
ggcgatggtc ccggcgtcgg cactgcgatc agctcgcacc cgggcatcga catggtcagc  660
ttcaccggct cgacccgcgc gggcatcctc gtggcgaagg ctgcggccga taccgtcaag  720
cgcgtccatc aggagcttgg cggcaagtcg cccaacgtcg tcctgcccga tgcagacttc  780
gccaagtacc tgccgtcgac cgcgtccggc ccgttggtca acagcggcca gagctgcatt  840
tcgcccaccc gcattctcgt accccgcgaa cgcgaagccg aagccgcggc gttcgtttcg  900
gcgatgtact cggcaacccc ggtcggcgat ccgatgcagg aaggtgcgca catcggcccg  960
gtggtcaaca aggcgcagtt cgacaagatc cgcggcctga tccagtcggc gatcgacgaa  1020
ggcgcgaagc tcgagaccgg cggccccgac ctcccggcca acgtcaaccg cggctactac  1080
atcaagccca cggtcttctc cggcgtcacg cccgacatgc gcattgcgca ggaggaaatc  1140
ttcggcccgg tcgcgacgat catggcgtac gacagcctcg aggaggccat cgagatcgcc  1200
aacgacaccg cctatggcct gtcggcctgc atcaccggcg atccggcgaa ggcggctgaa  1260
gtcgcgcccg agcttcgcgc cggcatggtc gcgatcaaca actggggccc caccccgggc  1320
gcgccgttcg gcggctacaa gcagtccggc aacggccgcg agggcgggct ctatggcctc  1380
aaggacttca tggaaatgaa ggcgatcagc ggcctgcctg cctga                  1425

SEQ ID NO: 10          moltype = AA  length = 474
FEATURE                Location/Qualifiers
source                 1..474
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 10
MRERLQQYID GKWVDSEGGK RHEVINPTTE EPCCVITLGT QADVDKAVAA AQRAFKTFSK  60
TTREERLALL ERIVEEYKKR VPDLAAAMAE EMGAPVSFAS TAQVGAGIGA FLGTMAALRN  120
FSPVEDNGAF KVAYEPIGVV GMITPWNWPL NQIALKVAPA LAAGNTMILK PSEECPTNAA  180
IFTEILDAAG VPPGVFNLIQ GDGPGVGTAI SSHPGIDMVS FTGSTRAGIL VAKAAADTVK  240
RVHQELGGKS PNVVLPDADF AKYLPSTASG PLVNSGQSCI SPTRILVPRE REAEAAAFVS  300
AMYSATPVGD PMQEGAHIGP VVNKAQFDKI RGLIQSAIDE GAKLETGGPD LPANVNRGYY  360
IKPTVFSGVT PDMRIAQEEI FGPVATIMAY DSLEEAIEIA NDTAYGLSAC ITGDPAKAAE  420
VAPELRAGMV AINNWGPTPG APFGGYKQSG NGREGGLYGL KDFMEMKAIS GLPA        474

SEQ ID NO: 11          moltype = DNA  length = 1347
FEATURE                Location/Qualifiers
source                 1..1347
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 11
atgactgccc cgaccgccgc cgacctttcc gccgacatcg cacgcgtctt cgcactccag  60
caggcgcaca tgtgggaggc caaggcctcc accgcggccg agcgcaagga aaagctcgcg  120
cgcctcaagg ccgccgtcga agcccacgcc gacgacatcg tcgccgccgt cctcgaagac  180
acgcgcaagc cggttggcga aatccgcgtg accgaagtcc tcaacgtcac cgccaacatc  240
cagcgcaaca tcgacaatct cgatgaatgg atgaagccgg tcgaggtcgc cacctcgctc  300
aatcccgccg accgcgcgca gatcatccac gaagcgcgcg gcgtctgcct gatccttggc  360
ccctggaact tcccctcgg cctcgcgctc ggtccggtcg ccgctgccat cgccgcaggc  420
aacacctgca tcgtgaagct caccgacctc tgccccgcca ccgcaagggt ggcctcggtg  480
atcgtcaggg aagcgttcga cgaaaaggat gtggctctgt tcgaaggcga cgtctcggtc  540
gccaccgcgc tcctcgatct gccgttcaac cacgtcttct tcaccggctc gccccgcgtc  600
ggcaagatcg tgatggccgc tgccgcaaag cacctcacca gcgtcacgct cgaacttggc  660
gggaagtcgc ccgtcatcgt cgacgatagc gccgacatcg atcaggtcgc cgcccagctc  720
gccgcggcca agcagttcaa cggcgggcag gcctgcatca gcccggacta cgtcttcgtg  780
aaggaagaca agaaggccgc gctggtcgaa ggcttccggg ccaacgtgca gaagaacctc  840
tatgacgatg ccggcaacct gaagaaggac agcatcgccc aggtggtcaa caaggcgaac  900
ttcgaccgcg tgaaggccat gttcgacgat gccgtcgcca agggcgcgac cgtcgccgcc  960
ggcggaacgt tcgaagccga tgacctcacc atccatccga ccatgctgac cggcgtcacc  1020
ccgcagatga ccatcctcca ggacgaaatc ttcgcccccg tcatcccggt gatgacctac  1080
gacacgctcg accaggcgat cggctacatc gaagcccgcg acaagccgct cgcactctat  1140
gtctacagca aggacgaagc gaacgtcgaa aaggtcctcg cccgcacctc gtcgggcggt  1200
gtcacggtga atggcgtgtt ctcgcactac ctggaaaaca acctgccgtt cggcggcgtc  1260
```

-continued

```
aacaccagcg gcatgggcag ctaccacggc gtgttcggct tcaagtgctt cagccacgaa  1320
cgggctgtct accgccacca gcagtaa                                      1347

SEQ ID NO: 12          moltype = AA  length = 448
FEATURE                Location/Qualifiers
source                 1..448
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 12
MTAPTAADLS ADIARVFALQ QAHMWEAKAS TAAERKEKLA RLKAAVEAHA DDIVAAVLED  60
TRKPVGEIRV TEVLNVTANI QRNIDNLDEW MKPVEVATSL NPADRAQIIH EARGVCLILG  120
PWNFPLGLAL GPVAAAIAAG NTCIVKLTDL CPATARVASV IVREAFDEKD VALFEGDVSV  180
ATALLDLPFN HVFFTGSPRV GKIVMAAAAK HLTSVTLELG GKSPVIVDDS ADIDQVAAQL  240
AAAKQFNGGQ ACISPDYVFV KEDKKAALVE GFRANVQKNL YDDAGNLKKD SIAQVVNKAN  300
FDRVKAMFDD AVAKGATVAA GGTFEADDLT IHPTMLTGVT PQMTILQDEI FAPVIPVMTY  360
DTLDQAIGYI EARDKPLALY VYSKDEANVE KVLARTSSGG VTVNGVFSHY LENNLPFGGV  420
NTSGMGSYHG VFGFKCFSHE RAVYRHQQ                                     448

SEQ ID NO: 13          moltype = DNA  length = 1515
FEATURE                Location/Qualifiers
source                 1..1515
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 13
atgaacgaca tgaccaccat ctcgcgcacg cagcgcgaat actcggaggc cgccaaggcc  60
ttcctcgcgc gcaagccgca gttgttcatc aacaacgagt gggtcgacag cagccacgac  120
gccgtgatcg aggtggaaga cccctcgaac ggcaggatcg tcggtcatgt cgtcgatgcc  180
tcggacaagg acgtcgaccg ggcggttgcc gcggcgcgcg ccgcgttcga cgatggccgc  240
tggtccaacc tgccgccaat ggtccgcgat cgcaccatga atcgcctggc cgacctgctt  300
gaagccaacg ccgatctctt tgccgagctc gaagcgatcg acaacggcaa gcccaagggc  360
atggccggcg ccgtcgacat ccccggcgcg atcagccagc tccgcttcat ggccggctgg  420
gccagcaagg tcgcgggcga gacgacgcag ccctacacga tgcccaacgg caccgtgttc  480
agctacaccg tcaaggaacc cgtcggcgtc tgcgcgcaga tcgtgccgtg gaacttcccg  540
ctgctgatgg cctcgctcaa gatcgccccg gcgctggcgg ctggctgcac cctggtgctg  600
aagcccgccg aacagacctc gcttaccgcg ctcaagcttg ccgatctcgt ggtcgaggcc  660
ggcttccctg cgggcgtgat caacatcatc accggcaacg gccacaccgc cggtgaccgc  720
atggtcaagc atcccgacgt cgacaaggtc gccttcaccg gctcgaccga gatcggcaag  780
ctgatcaatc gcaacgccac caccacgctc aagcgggtca cgctcgaact gggcgggaag  840
agccccgtcg tggtcatgcc cgacgtcgac gtggcgcaga ccgcgcctgg cgttgccggc  900
gcgatcttct tcaacgcggg ccaggtctgc gttgccggtt cgcgtctcta tgcgcaccgt  960
tcggtgttcg attccgtgct cgaaggcatg acccagaccg cgccgttctg ggcgccgcgc  1020
ccctcgctgg atcccgaagc ccacatgggc ccgttggtca gcaaggagca gcacgaccgc  1080
gtgatgggct acatcgaggc gggcaagcgc gatggcgcca gcgtcgtcat gggcggcgat  1140
tgccccagcg ccgatggcgg gtactacgtc aacccgacga tccttgcaga cgtgaacccg  1200
cagatgtcgg tcgtgcgcga ggaaaatcttc ggccccgtcg tcgtcgccca gcgcttcgac  1260
gatctcgatg aagtggcgaa gatggccaac gacacctgct tcggcctcgg cgcgggcgtg  1320
tggacgcgcg atgtcgcggt gatgcacaag cttgcctcga agatcaaatc gggcaccgtg  1380
tggggcaact gccacgccct gatcgatacc gcgctgccct ttggcggcta caaggaatcg  1440
ggcctgggcc gcgaacaggg gcgcgccggc atcgacgcct acctcgagac caagaccgtc  1500
atcatccaga tgtaa                                                   1515

SEQ ID NO: 14          moltype = AA  length = 504
FEATURE                Location/Qualifiers
source                 1..504
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 14
MNDMTTISRT QREYSEAAKA FLARKPQLFI NNEWVDSSHD AVIEVEDPSN GRIVGHVVDA  60
SDKDVDRAVA AARAAFDDGR WSNLPPMVRD RTMNRLADLL EANADLFAEL EAIDNGKPKG  120
MAGAVDIPGA ISQLRFMAGW ASKVAGETTQ PYTMPNGTVF SYTVKEPVGV CAQIVPWNFP  180
LLMASLKIAP ALAAGCTLVL KPAEQTSLTA LKLADLVVEA GFPAGVINII TGNGHTAGDR  240
MVKHPDVDKV AFTGSTEIGK LINRNATTTL KRVTLELGGK SPVVVMPDVD VAQTAPGVAG  300
AIFFNAGQVC VAGSRLYAHR SVFDSVLEGM TQTAPFWAPR PSLDPEAHMG PLVSKEQHDR  360
VMGYIEAGKR DGASVVMGGD CPSADGGYYV NPTILADVNP QMSVVREEIF GPVVVAQRFD  420
DLDEVAKMAN DTCFGLGAGV WTRDVAVMHK LASKIKSGTV WGNCHALIDT ALPFGGYKES  480
GLGREQGRAG IDAYLETKTV IIQM                                         504

SEQ ID NO: 15          moltype = DNA  length = 741
FEATURE                Location/Qualifiers
source                 1..741
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 15
gtgtccgata gcaatcagat tgccgcgctc gaaagccgcc tgaacgacct cgaaaggcgc  60
ctgacggtgc gcgaggacga gctggacgta cgcaagctcc agcatctcta cggctacctg  120
atcgacaagt gcatgtataa cgagaccgtg gacctgttca ccgaagatgg cgaagtgcgc  180
```

-continued

```
ttcttcggcg gcgtctggaa gggcaaggag ggcatccgcc gtctctacgt cgaacgtttc  240
cagaagcgct tcacctacgg caacaacggc ccgatcgacg gcttcctgct cgatcacccc  300
cagcttcagg acatcatcca cgtgcaggat gacggggtca ccgctctcgg ccgcgcgcgg  360
tcgatgatgc aggccggtcg ccacaaggat tacgagggcg atgccccgca cctcaaggcg  420
cgccagtggt gggaaggcgg catctacgag aacacctaca agaaggtgga cggcgtgtgg  480
cggatgcaca tcctcaacta catgccgatc tggcacgccg atttcgaaag cggctgggcc  540
aacaccccgc acgaatacgt gccgttcccc aaggtcacct atcccgaaga cccgaccgga  600
ccggacgaac tgatcgccga ccactggctc tggccgaccc acaagctgaa ccccttccac  660
atgaagcacc cggtgacggg cgaggaaatg gtcgcgcagc gctggcaggg cgacatcgac  720
cgcgagaacg cgcggaaata a                                            741

SEQ ID NO: 16          moltype = AA  length = 246
FEATURE                Location/Qualifiers
source                 1..246
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 16
VSDSNQIAAL ESRLNDLERR LTVREDELDV RKLQHLYGYL IDKCMYNETV DLFTEDGEVR  60
FFGGVWKGKE GIRRLYVERF QKRFTYGNNG PIDGFLLDHP QLQDIIHVQD DGVTALGRAR  120
SMMQAGRHKD YEGDAPHLKA RQWWEGGIYE NTYKKVDGVW RMHILNYMPI WHADFESGWA  180
NTPHEYVPFP KVTYPEDPTG PDELIADHWL WPTHKLNPFH MKHPVTGEEM VAQRWQGDID  240
RENARK                                                            246

SEQ ID NO: 17          moltype = DNA  length = 1485
FEATURE                Location/Qualifiers
source                 1..1485
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 17
atggcccaat ttccgaacac ccccagcttc acgggattca acacgccgtc gcggatcgag  60
gcggatatcg ccgatctggc ccacgaaggc acgattccgc aagggttaaa cggcgcattc  120
taccgcgtcc agcccgaccc gcagtttcct ccccgcctcg acgacgacat cgccttcaac  180
ggcgacggca tgatcacccg cttccacatc cacgacggcc aggtcgactt ccgccagcgc  240
tgggcgaaga ccgacaagtg gaagctggag aacgccgccg gaaaggccct gttcggcgcc  300
taccgcaacc cgctgaccga cgacgaggcg gtcaagggcg agatccgttc gaccgccaac  360
accaacgcct tcgtgttcgg cggcaagctg tgggcgatga aggaggacag tcccgccctc  420
gtcatggacc cggcgacgat ggaaaccttc gggttcgaga agttcggcgg caagatgacc  480
ggccagacct ttaccgccca ccccaaggtc gatccgaaga ccggcaacat ggtcgccatc  540
ggctatgccg caagcgggct gtgcaccgac gatgtgacct acatggaagt gagcccggag  600
ggcgagcttg tccgcgaagt gtggttcaag gtgccgtact actgcatgat gcacgacttc  660
ggcatcaccg aggattacct cgtgctgcac atcgtgcctt ccatcggaag ctgggaaagg  720
ctggaacagg gcaagccgca cttcggcttc gacacgacca tgccggtgca cctcggcatc  780
atcccgcgcc gcgacggcgt gcgccaggaa gacatccgct ggttcacgcg ggacaactgc  840
tttgccagcc atgtcctgaa cgcctggcaa gaggggacca agatccactt cgtgacctgc  900
gaggcgaaga acaacatgtt cccgttcttc cccgacgtcc acggcgcgcc cttcaacggc  960
atggaggcca tgagccatcc gaccgactgg gtggtcgaca tggccagcaa cggcgaggac  1020
tttgccggga tcgtgaagct ttccgacaca gccgccgagt tcccgcgcat cgacgaccgc  1080
tttaccggcc agaagacccg ccatggctgg ttcctcgaaa tggacatgaa gcgcccggtg  1140
gaattgcgcg gcggcagcgc cggcggcctg ctgatgaact gcctgttcca caaggacttc  1200
gaaacgggtc gcgagcagca ctggtggtgc ggcccggtgt cgagccttca ggagccgtgc  1260
ttcgtgccgc gcgccaagga tgcccccgaa ggcgacggct ggatcgtgca ggtttgcaac  1320
cggctggaag agcagcgcag cgacttgctg atcttcgacg cgctcgacat cgagaaaggc  1380
ccggtggcca cggtcaacat ccccatccgc ctgcgcttcg gccttcacgg caactgggcg  1440
aatgccgacg aaatcggcct tgccgagaag gtcctggccg catga                  1485

SEQ ID NO: 18          moltype = AA  length = 494
FEATURE                Location/Qualifiers
source                 1..494
                       mol_type = protein
                       organism = Novosphingobium aromaticivorans
SEQUENCE: 18
MAQFPNTPSF TGFNTPSRIE ADIADLAHEG TIPQGLNGAF YRVQPDPQFP PRLDDDIAFN  60
GDGMITRFHI HDGQVDFRQR WAKTDKWKLE NAAGKALFGA YRNPLTDDEA VKGEIRSTAN  120
TNAFVFGGKL WAMKEDSPAL VMDPATMETF GFEKFGGKMT GQTFTAHPKV DPKTGNMVAI  180
GYAASGLCTD DVTYMEVSPE GELVREVWFK VPYYCMMHDF GITEDYLVLH IVPSIGSWER  240
LEQGKPHFGF DTTMPVHLGI IPRRDGVRQE DIRWFTRDNC FASHVLNAWQ EGTKIHFVTC  300
EAKNNMFPFF PDVHGAPFNG MEAMSHPTDW VVDMASNGED FAGIVKLSDT AAEFPRIDDR  360
FTGQKTRHGW FLEMDMKRPV ELRGGSAGGL LMNCLFHKDF ETGREQHWWC GPVSSLQEPC  420
FVPRAKDAPE GDGWIVQVCN RLEEQRSDLL IFDALDIEKG PVATVNIPIR LRFGLHGNWA  480
NADEIGLAEK VLAA                                                   494

SEQ ID NO: 19          moltype = DNA  length = 1056
FEATURE                Location/Qualifiers
source                 1..1056
                       mol_type = genomic DNA
                       organism = Novosphingobium aromaticivorans
```

-continued

SEQUENCE: 19
atgacacaag accttaagac cggcggcgag cagggctacc tgcgcatcgc caccgaggaa 60
gccttcgcca cgcgcgagat catcgacgtc tacctgcgca tgatccgcga tggcactgcc 120
gacaagggca tggtctcgct ctggggcttc tacgcccagt ccccctcaga gcgcgccacc 180
cagatcctcg aacgcctgct cgatcttggc gagcgccgca tcgccgacat ggacgcgacc 240
ggcatcgaca aggctatcct cgcgctgacc tcgcccggcg tccagccgct gcacgacctt 300
gacgaggcca ggacgctcgc cacccgcgcc aacgacacgc ttgccgacgc gtgccaaaag 360
tacccagacc gcttcatcgg catgggcacc gtcgccccgc aggacccgga atggtccgcg 420
cgcgagatcc atcgtggtgc cagggaactg ggcttcaagg gcatccagat caacagccac 480
acgcaagggc gctacctcga cgaggagttc ttcgacccga tcttccgcgc cctcgttgaa 540
gtcgaccagc cgctctacat ccaccctgcc acttcgcccg attccatgat cgacccgatg 600
ctcgaagcgg gcctcgacgg cgccatcttc ggcttcggcg tggagacggg catgcacctg 660
ctgcgcctca tcaccatcgg catcttcgac aagtatccca gccttcagat catggtcggc 720
cacatgggcg aggcgctgcc ctactggctc taccgcctgg actacatgca ccaggccggt 780
gtccgctcgc agcgctacga acgcatgaag cccctgaaga agaccatcga gggctacctc 840
aagtccaacg tcctcgtcac caattcgggc gtcgcgtggg aacctgcgat caagttctgc 900
cagcaggtca tgggcgagga ccgcgttatg tacgcgatgg actaccccta ccagtacgtt 960
gccgacgagg tgcgcgcgat ggacgccatg gacatgagtg cgcaaacgaa gaagaagttc 1020
ttccagacca acgcggagaa gtggttcaag ctttga 1056

SEQ ID NO: 20 moltype = AA length = 351
FEATURE Location/Qualifiers
source 1..351
mol_type = protein
organism = Novosphingobium aromaticivorans
SEQUENCE: 20
MTQDLKTGGE QGYLRIATEE AFATREIIDV YLRMIRDGTA DKGMVSLWGF YAQSPSERAT 60
QILERLLDLG ERRIADMDAT GIDKAILALT SPGVQPLHDL DEARTLATRA NDTLADACQK 120
YPDRFIGMGT VAPQDPEWSA REIHRGAREL GFKGIQINSH TQGRYLDEEF FDPIFRALVE 180
VDQPLYIHPA TSPDSMIDPM LEAGLDGAIF GFGVETGMHL LRLITIGIFD KYPSLQIMVG 240
HMGEALPYWL YRLDYMHQAG VRSQRYERMK PLKKTIEGYL KSNVLVTNSG VAWEPAIKFC 300
QQVMGEDRVM YAMDYPYQYV ADEVRAMDAM DMSAQTKKKF FQTNAEKWFK L 351

SEQ ID NO: 21 moltype = DNA length = 23
FEATURE Location/Qualifiers
source 1..23
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 21
ctgtcgtgcc agctgcatta atg 23

SEQ ID NO: 22 moltype = DNA length = 29
FEATURE Location/Qualifiers
source 1..29
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 22
gaacatctag aaagccagtc cgcagaaac 29

SEQ ID NO: 23 moltype = DNA length = 49
FEATURE Location/Qualifiers
source 1..49
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 23
cgattcatta atgcagctgg cacgacagct tttcgcttct ccagctcgg 49

SEQ ID NO: 24 moltype = DNA length = 51
FEATURE Location/Qualifiers
source 1..51
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 24
cccacccgca atctcttatt tccggtccaa ctcccatcaa tttagtttgt c 51

SEQ ID NO: 25 moltype = DNA length = 51
FEATURE Location/Qualifiers
source 1..51
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 25
gtttctgcgg actggctttc tagatgttcc ttccacgatg aagcgggttg g 51

SEQ ID NO: 26 moltype = DNA length = 51
FEATURE Location/Qualifiers
source 1..51

-continued

```
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 26
gacaaactaa attgatggga gttggaccgg aaataagaga ttgcgggtgg g          51

SEQ ID NO: 27        moltype = DNA  length = 52
FEATURE              Location/Qualifiers
source               1..52
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 27
cgattcatta atgcagctgg cacgacagcg gctcgcgcaa tttgttagta ag         52

SEQ ID NO: 28        moltype = DNA  length = 52
FEATURE              Location/Qualifiers
source               1..52
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 28
gtttctgcgg actggctttc tagatgttcc ggatcatgcg caggtagacg tc         52

SEQ ID NO: 29        moltype = DNA  length = 52
FEATURE              Location/Qualifiers
source               1..52
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 29
gtttctgcgg actggctttc tagatgttcc ggatcatgcg caggtagacg tc         52

SEQ ID NO: 30        moltype = DNA  length = 50
FEATURE              Location/Qualifiers
source               1..50
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 30
caggcaaaag gcttccggag attaaatata attgcggtgt cggtcggcag            50

SEQ ID NO: 31        moltype = DNA  length = 51
FEATURE              Location/Qualifiers
source               1..51
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 31
cgattcatta atgcagctgg cacgacagga aggcgcaatc cggagttctc c          51

SEQ ID NO: 32        moltype = DNA  length = 42
FEATURE              Location/Qualifiers
source               1..42
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 32
ccctcccggc gctggtcaaa ggcaggcttc cttcccggga ag                    42

SEQ ID NO: 33        moltype = DNA  length = 52
FEATURE              Location/Qualifiers
source               1..52
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 33
gtttctgcgg actggctttc tagatgttct ccagtggaag ccgggagtga cc         52

SEQ ID NO: 34        moltype = DNA  length = 42
FEATURE              Location/Qualifiers
source               1..42
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 34
cttcccggga aggaagcctg cctttgacca gcgccgggag gg                    42

SEQ ID NO: 35        moltype = DNA  length = 55
FEATURE              Location/Qualifiers
source               1..55
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 35
cgattcatta atgcagctgg cacgacaggg gggctaaccg ccagtctcta tcttc       55
```

-continued

SEQ ID NO: 36 moltype = DNA length = 52
FEATURE Location/Qualifiers
source 1..52
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 36
gcaatacata caatattgca aggaggatgc cgccgcatga tccagcccgg ag 52

SEQ ID NO: 37 moltype = DNA length = 50
FEATURE Location/Qualifiers
source 1..50
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 37
gtttctgcgg actggctttc tagatgttcc caacaggcag ccgaggatag 50

SEQ ID NO: 38 moltype = DNA length = 52
FEATURE Location/Qualifiers
source 1..52
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 38
ctccgggctg gatcatgcgg cggcatcctc cttgcaatat tgtatgtatt gc 52

SEQ ID NO: 39 moltype = DNA length = 51
FEATURE Location/Qualifiers
source 1..51
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 39
cgattcatta atgcagctgg cacgacagct gacacggatc tctcctcaac c 51

SEQ ID NO: 40 moltype = DNA length = 50
FEATURE Location/Qualifiers
source 1..50
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 40
gtaaaccgtg taaacccgtt caggtattgc tacagccctg ttaaattgcg 50

SEQ ID NO: 41 moltype = DNA length = 50
FEATURE Location/Qualifiers
source 1..50
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 41
cgcaatttaa cagggctgta gcaatacctg aacgggttta cacggtttac 50

SEQ ID NO: 42 moltype = DNA length = 50
FEATURE Location/Qualifiers
source 1..50
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 42
cgcaatttaa cagggctgta gcaatacctg aacgggttta cacggtttac 50

SEQ ID NO: 43 moltype = DNA length = 30
FEATURE Location/Qualifiers
source 1..30
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 43
taacagaaag ccgaaaataa caaagttagc 30

SEQ ID NO: 44 moltype = DNA length = 34
FEATURE Location/Qualifiers
source 1..34
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 44
catggttaat ttctcctctt taatgaattc tgtg 34

SEQ ID NO: 45 moltype = DNA length = 32
FEATURE Location/Qualifiers
source 1..32

-continued

```
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 45
cagaaagccg aaaataacaa agttagcctg ag                                  32

SEQ ID NO: 46        moltype = DNA  length = 24
FEATURE              Location/Qualifiers
source               1..24
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 46
tgcgatcgcg ctctgaaaat acag                                           24

SEQ ID NO: 47        moltype = DNA  length = 43
FEATURE              Location/Qualifiers
source               1..43
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 47
taaagaggag aaattaacca tgtccgatag caatcagatt gcc                      43

SEQ ID NO: 48        moltype = DNA  length = 39
FEATURE              Location/Qualifiers
source               1..39
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 48
tgttattttc ggctttctgt tatttccgcg cattttcgc                           39

SEQ ID NO: 49        moltype = DNA  length = 41
FEATURE              Location/Qualifiers
source               1..41
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 49
taaagaggag aaattaacca tgactgcgta cccttctctc c                        41

SEQ ID NO: 50        moltype = DNA  length = 42
FEATURE              Location/Qualifiers
source               1..42
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 50
tgttattttc ggctttctgt taccttcat gtaccgcttt gg                        42

SEQ ID NO: 51        moltype = DNA  length = 42
FEATURE              Location/Qualifiers
source               1..42
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 51
taaagaggag aaattaacca tgacacaaga cctgaagacc gg                       42

SEQ ID NO: 52        moltype = DNA  length = 48
FEATURE              Location/Qualifiers
source               1..48
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 52
tgttattttc ggctttctgt taaagtttaa accatttttc agcgttgg                 48

SEQ ID NO: 53        moltype = DNA  length = 45
FEATURE              Location/Qualifiers
source               1..45
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 53
taaagaggag aaattaacca tggctcaatt tccgaatacc ccaag                    45

SEQ ID NO: 54        moltype = DNA  length = 40
FEATURE              Location/Qualifiers
source               1..40
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 54
tgttattttc ggctttctgt tatgcggcca ggacctttc                           40
```

-continued

```
SEQ ID NO: 55          moltype = DNA  length = 44
FEATURE                Location/Qualifiers
source                 1..44
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 55
taaagaggag aaattaacca tgctaagcga caggcacgtc aaag                    44

SEQ ID NO: 56          moltype = DNA  length = 51
FEATURE                Location/Qualifiers
source                 1..51
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 56
tgttattttc ggctttctgt tagaacacca ctactgaacg aatcgattta c             51

SEQ ID NO: 57          moltype = DNA  length = 50
FEATURE                Location/Qualifiers
source                 1..50
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 57
aaatctgtat tttcagagcg cgatcgcatc cgatagcaat cagattgccg               50

SEQ ID NO: 58          moltype = DNA  length = 49
FEATURE                Location/Qualifiers
source                 1..49
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 58
ggctaacttt gttattttcg gctttctgtt atttccgcgc attttcgcg                49

SEQ ID NO: 59          moltype = DNA  length = 52
FEATURE                Location/Qualifiers
source                 1..52
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 59
aaatctgtat tttcagagcg cgatcgcaac tgcgtaccct tctctccaca tg             52

SEQ ID NO: 60          moltype = DNA  length = 55
FEATURE                Location/Qualifiers
source                 1..55
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 60
ggctaacttt gttattttcg gctttctgtt acccttcatg taccgctttg gtgac          55

SEQ ID NO: 61          moltype = DNA  length = 49
FEATURE                Location/Qualifiers
source                 1..49
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 61
cattaaagag gagaaattaa ccatgcagtt tgaacgtatc aatccgatg                49

SEQ ID NO: 62          moltype = DNA  length = 48
FEATURE                Location/Qualifiers
source                 1..48
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 62
gtttaaacta ttaatgatga tgttaaattg gatagtgacc tggttggg                 48

SEQ ID NO: 63          moltype = DNA  length = 40
FEATURE                Location/Qualifiers
source                 1..40
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 63
cattaaagag gagaaattaa ccatgaaagc cgccgtactc                          40

SEQ ID NO: 64          moltype = DNA  length = 49
FEATURE                Location/Qualifiers
source                 1..49
```

-continued

```
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 64
gtttaaacta ttaatgatga tgttattgat caaacacaat aacagaacg              49

SEQ ID NO: 65        moltype = DNA  length = 48
FEATURE              Location/Qualifiers
source               1..48
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 65
cattaaagag gagaaattaa ccatgacaat caatacaatt cgcgtacg               48

SEQ ID NO: 66        moltype = DNA  length = 46
FEATURE              Location/Qualifiers
source               1..46
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 66
cgtttaaact attaatgatg atttaacaaa aatgacggca gctctg                 46

SEQ ID NO: 67        moltype = DNA  length = 41
FEATURE              Location/Qualifiers
source               1..41
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 67
cattaaagag gagaaattaa ccatgttggg acgtgcatcg g                      41

SEQ ID NO: 68        moltype = DNA  length = 44
FEATURE              Location/Qualifiers
source               1..44
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 68
gtttaaacta ttaatgatga tgttacgtga tcgtcggatc gatc                   44

SEQ ID NO: 69        moltype = DNA  length = 45
FEATURE              Location/Qualifiers
source               1..45
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 69
cattaaagag gagaaattaa ccatggcggc aattaatctt ccccg                  45

SEQ ID NO: 70        moltype = DNA  length = 48
FEATURE              Location/Qualifiers
source               1..48
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 70
gtttaaacta ttaatgatga tgttagccaa agacttcggc atagaggc               48

SEQ ID NO: 71        moltype = DNA  length = 50
FEATURE              Location/Qualifiers
source               1..50
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 71
cattaaagag gagaaattaa ccatgcgatt gaaagtactg ggacttatgg              50

SEQ ID NO: 72        moltype = DNA  length = 45
FEATURE              Location/Qualifiers
source               1..45
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 72
gtttaaacta ttaatgatga tgttagccac ctttggcttc taaag                  45

SEQ ID NO: 73        moltype = DNA  length = 52
FEATURE              Location/Qualifiers
source               1..52
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 73
cattaaagag gagaaattaa ccatgattcc gcatggtgaa cattcaatgc tg           52
```

-continued

SEQ ID NO: 74 moltype = DNA length = 48
FEATURE Location/Qualifiers
source 1..48
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 74
gtttaaacta ttaatgatga tgttatggca ccaaaaccag agcgccac 48

SEQ ID NO: 75 moltype = DNA length = 46
FEATURE Location/Qualifiers
source 1..46
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 75
cattaaagag gagaaattaa ccatggacgc atacgctgca attatc 46

SEQ ID NO: 76 moltype = DNA length = 54
FEATURE Location/Qualifiers
source 1..54
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 76
gtttaaacta ttaatgatga tgttacattt tgagaatggc ttttatcgct tttc 54

SEQ ID NO: 77 moltype = DNA length = 50
FEATURE Location/Qualifiers
source 1..50
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 77
cattaaagag gagaaattaa ccatgtctac acagcctgca accatagctg 50

SEQ ID NO: 78 moltype = DNA length = 46
FEATURE Location/Qualifiers
source 1..46
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 78
gtttaaacta ttaatgatga tgttatggac gagtttgccc gcttcc 46

SEQ ID NO: 79 moltype = DNA length = 50
FEATURE Location/Qualifiers
source 1..50
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 79
cattaaagag gagaaattaa ccatgcgcga acggctacag caatacattg 50

SEQ ID NO: 80 moltype = DNA length = 45
FEATURE Location/Qualifiers
source 1..45
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 80
gtttaaacta ttaatgatga tgttaggcag gcaggccgct gatcg 45

SEQ ID NO: 81 moltype = DNA length = 40
FEATURE Location/Qualifiers
source 1..40
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 81
cattaaagag gagaaattaa ccatgactgc ccctaccgcc 40

SEQ ID NO: 82 moltype = DNA length = 47
FEATURE Location/Qualifiers
source 1..47
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 82
gtttaaacta ttaatgatga tgttactgct gatgacgata tacagcc 47

SEQ ID NO: 83 moltype = DNA length = 46
FEATURE Location/Qualifiers
source 1..46

-continued

```
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 83
cattaaagag gagaaattaa ccatgggtta ccgggttgta gtggtg                46

SEQ ID NO: 84        moltype = DNA  length = 49
FEATURE              Location/Qualifiers
source               1..49
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 84
cattaaagag gagaaattaa ccatgcagtt tgaacgtatc aatccgatg              49

SEQ ID NO: 85        moltype = DNA  length = 46
FEATURE              Location/Qualifiers
source               1..46
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 85
cattaaagag gagaaattaa ccatggcgat caaagttgcg ataaac                46

SEQ ID NO: 86        moltype = DNA  length = 46
FEATURE              Location/Qualifiers
source               1..46
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 86
gtttaaacta ttaatgatga tgttaaagga atttcgccat tgctcc                46

SEQ ID NO: 87        moltype = DNA  length = 45
FEATURE              Location/Qualifiers
source               1..45
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 87
cattaaagag gagaaattaa ccatgaatga catgactacc atctc                 45

SEQ ID NO: 88        moltype = DNA  length = 52
FEATURE              Location/Qualifiers
source               1..52
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 88
gtttaaacta ttaatgatga tgttacattt gaataattac tgttttagtc tc          52

SEQ ID NO: 89        moltype = DNA  length = 47
FEATURE              Location/Qualifiers
source               1..47
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 89
cattaaagag gagaaattaa ccatggctac gcagttgaga agtgcag                47

SEQ ID NO: 90        moltype = DNA  length = 49
FEATURE              Location/Qualifiers
source               1..49
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 90
gtttaaacta ttaatgatga tgttactgat cgaacattcc ggtacgacc              49

SEQ ID NO: 91        moltype = DNA  length = 737
FEATURE              Location/Qualifiers
source               1..737
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 91
ccgatagcaa tcagattgcc gcgcttgaaa gtcgcctgaa tgacctcgaa aggcgactga  60
cggttagaga ggacgagctg gacgtacgca aactccagca tttatacggt tatctgattg  120
ataaatgcat gtataacgag acagttgacc tgttcacaga agatggggaa gtgcggttct  180
ttggtggcgt atggaaaggc aaggagggca tccgccgttt gtacgttgaa cgttttcaga  240
aacgtttcac ctatggcaat aacggcccga ttgatgggtt cctgttagat catccacaac  300
ttcaagatat tattcacgtg caggatgatg gggtcacggc tttgggccgc gcgcgttcca  360
tgatgcaagc cggtcgccac aaggattatg agggagatgc acctcatctg aaagcgcgtc  420
agtggtggga aggtggtata tacgaaaaca cttataaaaa agtggatggc gtgtggcgta  480
tgcatatcct aaactacatg ccgatctggc acgcagattt tgaaagcggc tgggccaata  540
```

-continued

```
ccccgcacga atacgttcct tttcccaaag tcacctatcc agaagacccg actggaccgg  600
atgaactgat tgctgaccat tggttatggc cgacccataa gctgaacccc tttcacatga  660
aacatccggt gacgggtgag gaaatggtcg cacagcgctg gcagggtgac atcgatcgcg  720
aaaatgcgcg gaaataa                                                 737

SEQ ID NO: 92          moltype = DNA  length = 1428
FEATURE                Location/Qualifiers
source                 1..1428
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 92
actgcgtacc cttctctcca catgattatt gacggtgccc gtgtcagcgg cggaggacgt  60
cgcacccacg cggtcgtcaa tccggctacc ggagagacca tcggtgaact gccgctggca  120
gaagttgcag atctggatcg agcgttagaa gtagcggcga agggcttccg tatttggcgt  180
gacagcacac cgcagcagcg cgcagccgtg ttacagggcg cggcccggct gatgctggaa  240
cggcaagagg atctcgctcg catagccacg atggaagaag gtaaaaccct gcccgaggcg  300
cgcatcgaag ttctgatgaa cgtgggcctg ttcaattttt acgctggaga agtatttcgt  360
ttatatggcc gaaccctagt gcgccctgcg ggtcagagaa gcacgatcac gcatgaaccg  420
gtagggccgg tggccgcctt tgctccgtgg aactttccgc ttgggaatcc aggtcgcaaa  480
ctgggcgcgc caattgccgc cggttgctcg gtgattctaa aagcggcgga agaaacgccg  540
gcttcagcgt taggggtgct gcaatgtctg ctggatgctg gcctgcctaa agaagtggcc  600
caggctgtgt tcggtgtgcc tgacgaggtg agtcgccacc tgttgggcag ttccgttatc  660
cgcaagctct cgtttacagg ttctaccgtc atcggcaagc atctgatgcg acttgcagcc  720
gacaacatgt tgcgtacaac tatggagctt ggcggccatg gtcctgtctt agttttcggt  780
gatgcagata ttgacaaagc gctcgatacc atggcagctt ccaaatatcg taacgcgggc  840
caagtttgtg tttcaccaac cagatttata gtggaagaaa gcgtgttcga acgttttcgt  900
gatggttttg cagagcgtgt cggtcggatc aaagttggaa atggtttgga tcaggatgcg  960
cagatgggac cgatggcaaa tgcccgccgc ccggaggcga tggatcgtct gatcggggac  1020
gccgtgactc gcggcgcaag gttgcatact gggggcgaac gtgtcggcaa cgccggctat  1080
ttttatgccc ccacggttct gagtgaagta ccgctggacg cggctattat gaacgaagaa  1140
ccgtttggcc cggtagctct gattaatcca ttcggcggtg aggaagcgat gatcgccgaa  1200
gcaaaccgtc tgccgtatgg cttggcagcc tacgcatgga cagatagcgc ggcgcgggca  1260
aaacgcttag cacgcgagat tgagacgggg atgctggggc ttaattctac catgattggc  1320
ggcgcggatt cgccattcgg tggggtgaaa tggtccggac acggttcaga ggacggtccc  1380
gaaggtgtta tggcctgcct tgtcaccaaa gcggtacatg aagggtaa                1428

SEQ ID NO: 93          moltype = DNA  length = 1053
FEATURE                Location/Qualifiers
source                 1..1053
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 93
acacaagacc tgaagaccgg cggggagcag ggttacctgc gtatcgccac cgaagaagct  60
ttcgccacgc gagaaatcat tgatgtctac ctgcgcatga tacgcgatgg aactgctgat  120
aaaggtatgg tatcattgtg gggcttttat gcccagtccc cttcagagcg cgccacccag  180
atcttagaac gtctgttaga tcttggcgag cggcgtattg cagatatgga tgcgacaggc  240
attgacaagg ctattctagc gctgacctcg ccgggcgtac agccgctgca tgacttagat  300
gaagcacgga cgctcgcaac ccgtgcaaat gatactcttg ccgatgcgtg ccaaaagtat  360
ccagaccgat ttattggaat gggcaccgtg gccccgcagg atccggaatg gagtgcgcgc  420
gaaattcatc gtggtgcaag ggaactgggt tttaagggca tccagatcaa cagccacacg  480
caagggcgct acttggatga ggaattcttt gatccgatat tccgtgccct cgttgaagtc  540
gaccagccgc tgtatattca tcctgccact tcgccagatt ccatgatcga tccgatgttg  600
gaagcgggcc tggacggtgc aatcttcggc ttcggtgtgg agacgggcat gcatctgctg  660
cgcctgatca cgattgggat tttcgacaaa tatcccagct tgcaaattat ggttgggcac  720
atgggcgagg cgctgcccta ctggctctat agactggatt atatgcacca ggctggtgtg  780
cgctctcagc gctatgaacg tatgaaacca ctgaaaaaaa ccatcgaagg ttatcttaaa  840
agcaacgtgt tagtgacaaa ttctggagtc gcgtgggaac ctgcgattaa attttgtcag  900
caagtaatgg gtgaggatcg ggttatgtac gcgatggact acccgtatca gtacgttgca  960
gacgaagtgc gtgcgatgga tgccatggac atgagtgcgc aaacgaaaaa aaaatttttt  1020
cagaccaacg ctgaaaaatg gtttaaactt taa                               1053

SEQ ID NO: 94          moltype = DNA  length = 1554
FEATURE                Location/Qualifiers
source                 1..1554
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 94
atggctcaat ttccgaatac cccaagcttc acgggattca acacgccgtc tcggattgag  60
gcggatattg cagatctggc ccacgaaggt acgattccgc aagggttaaa cggcgcattt  120
tatcgtgtcc agcccgatcc gcagtttcct ccacgcctcg atgatgacat tgcctttaac  180
ggagacggga tgattacccg attccatata catgatggcc aggtcgactt ccgtcaacgt  240
tgggcgaaaa ccgataaatg gaaactggaa aacgcggccg gaaaagccct gtttggtgcc  300
taccgcaacc cactgaccga tgacgaggcg gttaaaggcg agatccgttc gaccgccaac  360
actaacgcct tcgttttcgg tggcaaactg tgggcgatga aagaggacag tccagcactc  420
gtaatggatc cggcgacgat ggaaaccttc gggttcgaaa agttcggcgg taaaatgaca  480
ggccagacct ttactgccca tccgaaggta gatccgaaaa ccggcaatat ggtagcgatc  540
```

-continued

```
ggttatgctg caagcgggtt gtgcacagat gatgtgacct acatggaagt tagtccggag  600
ggtgaattag tacgcgaagt gtggttcaaa gtgccgtatt attgcatgat gcacgacttc  660
ggcattacag aggattacct cgtgctgcac attgttcctt ccatcggaag ctgggaaaga  720
ttagaacagg gcaaaccgca ctttggcttt gatactacta tgccggttca cctaggtatc  780
attccgaggc gtgacggtgt gcgccaggaa gatatccgtt ggttcacgcg ggataattgt  840
tttgccagtc atgtactgaa tgcttggcaa gaagggacca aaattcactt tgtgacttgc  900
gaagcgaaaa acaacatgtt tcctttcttt ccagatgtcc atggcgcgcc ctttaacggt  960
atggaggcaa tgtcacatcc tacggactgg gtggtcgaca tggcaagcaa cggcgaggac  1020
tttgctggga tcgtgaagct ttccgatata gctgcagaat ttcctcgcat cgacgaccgg  1080
tttaccggcc agaaaacccg ccatggttgg ttcttagaaa tggatatgaa acgaccagtg  1140
gaattgcgcg gtggttcagc gggcggcctg ctgatgaatt gtctgtttca caaggacttc  1200
gaaacgggtc gtgaacagca ttggtggtgc ggcccggttt cgtctcttca ggagccgtgt  1260
tttgttccgc gcgcgaaaga tgcccccgaa ggtgatggat ggattgtgca agtttgtaat  1320
cgtctggaag aacagcgttc cgatttgctg atatttgatg cgctggatat tgagaaaggc  1380
ccggtggcta cggtcaatat ccccatccgc ctgcgctttg gcttgcatgg taattgggcg  1440
aatgcagacg aaattgggct tgcggaaaag gtcctggccg cagcgatcgc aggaagcgaa  1500
aatctgtatt ttcagagcgc attggcacat caccatcatc accatcacca ttaa        1554

SEQ ID NO: 95          moltype = DNA  length = 1149
FEATURE                Location/Qualifiers
source                 1..1149
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 95
ctaagcgaca ggcacgtcaa agggagaccg catgaaatga aaacacgcgc cgcagttgcg  60
tttgcgccaa agcaaccgtt ggaaattgta gaactggatc tggaaggtcc caaagctggg  120
gaagttctgg ttgagattat ggcgactgga gtgtgtcaca ccgatgcata tacgttagac  180
gggttcgaca gcgaaggcat tttccctagc gtgctgggtc atgaaggtgc cggtatcgtg  240
cgcgaagtgg gccctggggt aacttccgtg aaacctggcg atcatgtgat cccgctctat  300
acgccggaat gtcgccagtg caaatcgtgc ttgtcgggta agaccaacct gtgcaccgct  360
attcgcgcca cgcaagggca gggcctgatg cccgatggca ccagtcgttt ttcttacaaa  420
ggccagaccg tgttccacta catgggttgc agtacattct ctaattttac agttctgcca  480
gagatcgcgg ttgcaaagat tcgcgaggat gcgccgttta aaacctcatg ttatattggc  540
tgtggcgtga cgacgggtgt tggcgcggtg attaacactg ctaaagtaca ggtcggtgac  600
aacgtcgtgg tctttggatt aggcggcata ggtctcaatg ttattcaggg agcgcggctt  660
gccggtgcag ggaaaatcat tggcgtcgat atcaatccag atcgggagga atggggccgt  720
aaatttggca tgactgactt tctgaatagt aagggcatga gccgcgagga cgtagttgct  780
aaagtcgtcg ccatgaccga tggcggtgcg gactatacct ttgatgccac cggtaatacc  840
gaagtgatgc gtacggcgct tgaagcatgc catcgtggtt ggggaacctc cataatcatt  900
ggtgtggcag aggcgggtaa agaaattagc acgcgtccgt tccaattagt tactggccgt  960
aactggcgag gcacggcctt cggaggcgcc aaggggcgca cagatgttcc gaaaattgta  1020
gatatgtaca tgaccggaaa aatcgaaatc gatccgatga tcacccatgt catggggctg  1080
gaagagatca acacagcatt tgatctgatg cacgctggta aatcgattcg ttcagtagtg  1140
gtgttctaa                                                          1149

SEQ ID NO: 96          moltype = DNA  length = 1398
FEATURE                Location/Qualifiers
source                 1..1398
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 96
cagtttgaac gtatcaatcc gatgacaggg gcagtagcct cgcaggcaga ggccatgaaa  60
gcgtcggaca ttccttccat tgctgcccgc gcaggacagg cctttccggc gtgggcagcg  120
atgggcccca acgcacgtcg cggcgtactg atgaaggcgg ctgcggcgtt ggaagcgcgg  180
gctgatgctt tcgtcgaagc catgatgggc gaaatcggcg cgactagagg gtgggcgctg  240
tttaaccttg gccttgcagc aagcatggtg cgcgaagccg ccgcgctgac cactcaaatc  300
tctggagagg ttattccatc tgacaaaccg gggtgtattt cgatggctct gcgcgaaccg  360
gttggtgtga ttttgggcat cgcgccgtgg aatgcgccga ttatccttgg ggtgcgcgca  420
attgccgtgc cgcttgcctg cggtaacgcg gtgatattaa aagcaagcga aacatgtccg  480
cgaacccacg cgctcatcat cgaggccttt gctgaagcag gtttcccaga aggcgtggtt  540
aatgtagtga cgaacgcgcc tgcagatgca gcggaagtgg tcggggcgct gattgatgcg  600
ccggaagtgc gtcgtataaa ctttaccggt agtactaatg taggcaggat tatcgcaaaa  660
cgggcggccg agcatttgaa accctgttta ctcgaactgg gcggtaaagc accgttaata  720
gttctggatg atgcggatct agacgaagcg gtcaaagctg cggcttttgg cgccttcatg  780
aaccaagggc agatttgcat gtcaacggag cggatcatcg ttgtagatgc cgttgccgat  840
gcattcgcag ataaattcaa ggccaaggtc gcctccatgg ctgtaggcga cccgcgtgag  900
ggtacgaccc cgttgggtgc agttgtcgac gctaaaactg tcgctcattg ccgtagctta  960
attgacgatg ccctggcaaa aggtgcccgt ctgctgaccg gcggtgaaac cacgcacaat  1020
gtgctcatgc ccgcccatgt cgtagatggc gtgacgcagg atatgaagct gttccgcgat  1080
gagagctttg gcccagtggt gggcgtgatt cgcgcgcgcg acgaagctca tgccattgaa  1140
ctggcgaacg acagtgaata tggactgtca gcggctgttt tcacacgtga cacagcgcgc  1200
ggcctgcgag ttgcccgcca gatccgtagc ggtatttgcc atgttaatgg acctaccgtc  1260
cacgatgagg cgcagatgcc ttttggtgga gtgggtgcgt ccggctacgg tcgttttggg  1320
ggtaaagccg gcatcgatag ttttaccgag ctgagatgga ttacgatgga aacccaacca  1380
ggtcactatc caatttaa                                                1398
```

-continued

SEQ ID NO: 97 moltype = DNA length = 1086
FEATURE Location/Qualifiers
source 1..1086
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 97
aaagccgccg tactcgtcga accgggtaaa ccgctggata ttcagcattt aagcgtgagt 60
aaacccggcc ctcatgaagt ccttatacgc acagcagcct gcgggctgtg ccatagtgac 120
ttgcacttca tcgaaggtgc ctatccacat ccgctgccgg ctgtgccagg gcacgaggct 180
gctgggattg tggaagcggt aggttcagaa gtgcgcacag taaaagtggg tgacgctgtt 240
gttacctgcc tgtccgcgtt ctgtggtcat tgcgagtttt gcgtgaccgg ccggatgtcg 300
ctgtgtcttg gtggcgatac tcggcgcggt gcgggtgagg cacctcgctt gacacgcacc 360
gacgatggaa gcgcagtgaa ccagatgctc aacctatcgg cctttgcaga acaaatgctg 420
gttcacgaac atgcctgtgt tgcgatcaat cccgagatgc cgctcgatag agctgcggtt 480
atcggctgtg cggtaaccac tggcgcgggt gcggtgttta atgctgcgaa actgacccca 540
ggagagacgg tatgcgttgt cggctgtggc ggcgtaggct tagcaacggt caatgccgcg 600
aaaattgccg gggcaggccg tattatcgct gtggatccga tgccggaaaa acgcgaactg 660
gccatgaaac tgggtgcgac cgatgtgatg gacgcgggac ccgatgctgc ggcacagatc 720
gttgaaatga cgaaaggcgg cgttcaccat gcgatcgagg ccgtggggcg tcctgcatct 780
ggcgaccttg cggtcgcgac gctgcgtcgt gggggcaccg ccacgatttt aggtatgatg 840
ccgctggcac acaaggtcgg attatcagcg atggatctgc tgagcgataa gaagctgcag 900
ggtgcaatta tgggccgcaa ccacttccca gtggatctgc cgcgactggt cgacttctac 960
atgcgtggct tgttggatct agacactatc attgccgaaa ggattccgct tgaagggata 1020
aacgatggtt ttgaaaaaat gaaacaggga cattccgccc gttctgttat tgtgtttgat 1080
caataa 1086

SEQ ID NO: 98 moltype = DNA length = 1011
FEATURE Location/Qualifiers
source 1..1011
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 98
acaatcaata caattcgcgt acgttcgccg gccactctcg acaccttaaa tttcgatacg 60
ctgacggatt gtggacaacc gggaacgagc gaaatccgca ttcgtctgcg cgcaacttct 120
ctgaacttcc actactacgc gatgattacc agaatgctgc cggctgcaac aagtcgaatt 180
cctatgtcta acggcgcctg acaggttttc ggggtgtgcg atggcgtgac caaattccag 240
gcgcgtaacg cagttatctc gacctttttc accgacagga acgccggtcc gccacagtca 300
gccgcgttta cgaccgtcac ggctgatggg attaatcgct acgcgcggga agaagtggtg 360
gccccggctc attggtttac ccgcgcgccg ttatgctata gtcacgcaaa agccgccacg 420
ctgacctgcg cgggccttac tgcatggcgt gctttgttca tagataacgc tatcaagccg 480
ggcgacacgg tcttggtgca gggcactggc agcgtttcgg ttttcgcgct gcagttaaca 540
aaggcggcat gcgcgcgtgt catcgcaacg agttcctccc accagtaact gaaacgcctg 600
cgcagcctta gagcgaataa aaccataaac tataaaacgc aaacctcacg ggggatgcag 660
acactagatt tcactgccgg tatttgtgta cactgtattg tcgagattag ccggcccggt 720
acgtttcatc aagcgatgat gtccacccgc gtgcgtgctc atatcgcgct gatcggtgtt 780
ctcgcgcgtt ttgcgggtcc agtttaaacc actttgctga tggcacagaa tctgcgcgta 840
taaggcctta ccgtggcctc acgtaccaat catctgcgaa tgattcccgg tatcgaggca 900
aaccgtatcc aacctgtcat tcaccgccat tttccatttc cgtattttgc cgctgccttt 960
cgccatcaac agagctgccg tcatttttgt taaatcgtga ttgacatttg a 1011

SEQ ID NO: 99 moltype = DNA length = 1080
FEATURE Location/Qualifiers
source 1..1080
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 99
ttgggacgtg catcggtgct ggtaaaaccg aaccaactgg agacgtggga tgttaaagta 60
gccgatccgg aaccgggcgg tgccttagtt tcgattgtgc tgggtggggt atgcgggagc 120
gacgtccata tattgaccgg cgaggctggc gtgatgccgt ttccgatcat tctgggacat 180
gagggcgtgg gaaggatcga aaaactgggg cacggcgtca gcactgatta cgctggtgag 240
gaacttaaac ccggcgatct ggtatattgg tcgccgattg ctctgtgtca tcgatgttat 300
tcctgcaatg ttctcgatga aacaccttgc gaaaataccc agtttttcga agatgcttcc 360
aagccgaact ggggttcata cgcagattat gcatggctgc ccaacggtat gccgttctat 420
aaactgccag cccaagcgca gcctgaagcg gttgctgcgc ttggctgtgc acttccaacc 480
gccctgcgcg gctttgatcg ctgcggcagt gttagagtgg gtgaaactgt ggttgtccaa 540
ggtgcaggcc ctgtcggcct gtctgcagtg ctcgtggcgg cgcaggccgg ggcgcgtgac 600
gtgattgtta ttgacggttc accacttcgt cgcgaagcgg ctaccgcatt gggtgcctct 660
ctgacgattg gcttagatgt cgcgcctgag gaacggcgcc ggatgattta cgatcgcgtt 720
ggtcgcaatg gtcccaatgt agtcatcgag gcagccggag ttctgccagc gtttccggaa 780
ggggtggacc tgaccggtaa ccacggccgt tacattgtgc taggattgtg gggcgcaata 840
gggacccagc cgatcagccc gcgcgactta acaatcaaaa acctgactat cgctggtgcg 900
accttcccta aaccaaaaca ttattatcag gccttgcatt tagcgacggc cctgcaggac 960
cgtgtaccgt tagccggtct ggtgagccac cgttttggcg tcagccaggc gggcgaagcg 1020
ctgagtctca ccaagagtgg gacagcgatt aaggccgtga tcgatccgac gatcacgtaa 1080

SEQ ID NO: 100 moltype = DNA length = 1155

-continued

FEATURE Location/Qualifiers
source 1..1155
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 100

```
gcggcaatta atcttccccg cgtgattcgt gctggtgggg gtgcattagc cgaactgccc  60
gatgcaatgg cgcagtgcgg cctttcacgc ccgttcgtgg tgaccgatgc attcttagtg  120
caaagcggga tggtcgctcg gatgttagag gttctggacg gcgctgggat tgcggccacg  180
gtcttcgatg ctacggtacc tgatccgact gttgctgtgg tagaacaggc gcttggcgca  240
ttgcgagagg cggaatgtga ttgtgtgatc gggtttggag gtggtagccc gatcgacacc  300
agtaaagcca ttgccgccct ggcgctggaa ccgcgtgcag ttcaatccat gaaggcacca  360
gcgacgaccg acgtcccggg tctgccgatc attgccgtcc cgacgaccgc cggcaccggc  420
tcggaggcga ctaaatttac aatcgtgacc gatgaggcga cgagtgaaaa aatgctctgc  480
gcaggtctgg ccttcctgcc tactatagcc attgtagatt tcgagctgac catgggcaaa  540
ccggctcggc taactgccga cacaggtatt gattcgctga cacatgcgat tgaggcctat  600
gtttctaaga aagccaatcc gtttagtgat gctatggcga tctcggcgat gaaactgatc  660
gcgccgaaca ttcgcaccgc ctgcgccgaa cccggaaacc gtgctgcacg cgaagcgatg  720
atgattggcg cgcaccatgc cggtattgcg ttttccaacg ctagcgttgc actggtgcac  780
ggtatgagcc gcccaatcgg cgcattcttt catgtgccgc acggattgtc caacgcaatg  840
ttgctgcctg cgattaccgc gttttcgct ccgtcagcgt taccacgtta cgccgattgt  900
gcccgtgcga tgggtgtagc tttggaaagc gaaggcgacc agtctgccgt tgcaaggctg  960
ctcgacgaac tggcggcgct gaacgcagac cttagtgtcc cgacgccgca gtcgcatggg  1020
atcagcgctg atcgttggtt tgaagtagtg cctgaaatgg cgagacaggc aatagcatca  1080
ggctctccag gcaataatcc acgcgttcct gatgcggcgg aaatcgagcg cctctatgcc  1140
gaagtctttg gctaa                                                   1155
```

SEQ ID NO: 101 moltype = DNA length = 2115
FEATURE Location/Qualifiers
source 1..2115
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 101

```
cgattgaaag ttctgggact tatggcagca ctgctgccgc tggcggcttg taacatcaaa  60
agcgagggtg gaggggatgc agtcgccaac gctggagtca cagatgccct gattgcccaa  120
gcgcccgaag gcgaatggct gagctatggc cgcgattatg gggaacaacg cttttcaccg  180
ttgacccaaa ttaatgatgg taacgtcggg cagttgggtc ttgcctggtt tcatgacctg  240
gagactgcgc gcgggcaaga agcgacgccg ctgatgcatg atggtacgtt atatatctcg  300
actgcgtggt caatggtgaa agcgttcgat gcaaaaaccg gcgcgctgaa atggagttac  360
gatcccgaag taccgcgtga aacgctggtg cgcgcatgct gcgacgcggt caatcgtggc  420
gtcgcgctgt atggagataa agtttttgta ggtacgctcg atggtcgtct agtagcgtta  480
gatcagaaga ccggaaaagt agtttggtcc aaggtagtag tgcccaatca ggaggactac  540
accataactg gtgccccgcg cgtggtgaaa ggcaaagttc tgattggtag cggtggctcg  600
gagtacaaag ctcgaggcta tattgccgcc tatgacgtta acacaggcaa cgaagtgtgg  660
aaattccaca ccgtccctgg caatccagcg gatgggtttg agaacaaagc gatggaaaat  720
gccgctcgca cttgggctgg tgaatggtgg aaactcggtg ggggtggcac ggtgtgggat  780
tccatcacct atgatccagc caccaaccta gttctgttcg gcacaggcaa tgcagaacca  840
tggaacccgg cagcagccgg gcgggaggga gacagcttgt acacgtcctc tattgtagcg  900
gtgaatgccg atactggcga ctatgtatgg cattttcaag aaaccccgga agaccgttgg  960
gacttcgatt ccgcgcagca gattacgctg gccgacctga caattgatgg gcagcggcgc  1020
cacgtgatcc ttcatgcgcc taagaacggt catgtttatg tgttggacgc aagaaccggg  1080
cagtttctgt cggcaacgcc ctttgtgatg gtgaactggg cgaccggtat tgatcctaaa  1140
acgggcaagg ccactgtcaa tccagaagcc cgttatgaaa aaaccggcaa acctttcgtt  1200
agcctgccag gtgcggtagg cgcacattca tggcagccgc agagtttcag cccgaaaacc  1260
ggcctgctgt accttccggt gaacaatgcg gcatttcctt atgcagccgc caaagactgg  1320
aaagcaaccg atattggttt ccagaccggt ctcgacggct atgttaccag tatgccagcc  1380
gacgcaaagg tccagggcgc agcgatgaaa gcgaccactg gtacgttagt ggcgtgggac  1440
ccggttgcga agaaagccgc ttggaaagtc gaactgccga gcccgagtaa cggtggcatt  1500
ttatcgacag ctggcaattt agtgtttcaa ggtaccgcgg gcggtgattt tgttgcatac  1560
aacgccgata agggcaaaca attatggtct tttccggcgc agagtggcat ccttgccgcg  1620
ccgatgacct atgctatcga tggggaacag tacgttgcgg tcatggtggg ctggggaggt  1680
gtgtgggacg tcgccacagg tgtgctcgct cataaggcca aaaaacagag gaacataagc  1740
cgcctggtag tgttcaaact gggcgggaaa gccacgctgc cggctgctcc tccgatggca  1800
aaaatggttt tggatccgcc gccgtttaca ggtacgcccg aacaagctaa ggccggtggc  1860
gaattatacg gacgttactg caacgtttgt catggtgatg ctgcggttgc gggcggcgtg  1920
aatccagatc tgcgtcactc agctgcgctt aatgcaccag aggcgatccg gtctgtggtg  1980
attgaggggg cgctgcagca caacgggatg gtctcgttca aatctgcgct gaagcctgag  2040
gatgcggata atatccgcca ctacttgatc aaacgtgcaa atgaagacaa agctctcgaa  2100
gccaaaggag gctaa                                                   2115
```

SEQ ID NO: 102 moltype = DNA length = 1002
FEATURE Location/Qualifiers
source 1..1002
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 102

```
attccgcatg gtgaacattc aatgctggca atgcagttgg atggtccagg caaacggctg  60
```

-continued

```
cacccagtcg tgcgccctct gccgttaccg gggcgaggtg aagtgcgggt aaaagtgcat  120
gcctgtggtg tttgccgtac ggacctgcac gttgcagatg gcgatattca cggtctgcta  180
cctattgtgc cggggcacga agtgataggc gttgtcgatg cactggggcc gggggtgacg  240
gatgttgaac ctggtgcgcg tgtaggtgtc ccgtggctcg gccatgcctg tggcacctgc  300
ccatattgcg acagcgggag ggaaaacctt tgtgatgcgc cgctgttcac cggttttact  360
cgcgatggcg gatacgctac ccatgtgatt gcagatgcgc gcttttgctt tcctattcca  420
gagggttttg acgatctgca cgcggcgccg ctcctgtgcg cgggcttgat cggctatcgc  480
gctcttcggc ttgccggcga tgcacctgta ctcggattct atggttttgg agcggcggcg  540
catattttag ctcaggtggc cctgtggcag ggtagaacgg tttacgcgtt tactcgcgat  600
ggcgacgcta aggcccaggc ctttgctcgt gacatcggtt gccaatgggc cggaccctct  660
ggcgctgcgc cgccgcaagc tctggacgca gcgatcatct tcgcctccgc gggagaattg  720
gtgccgacag ccctgcgtgc agtgcgcaaa ggcgggcgtg ttgtctgtgc cggtattcat  780
atgagcgata tcccggcatt cccctacgcc gatttatggg aggaacgtca gatcctgtcg  840
gtagcgaatt taacccgacg cgatggcgta gaattcctgc cccttgcagc gcgtgcaggc  900
gttcgcacac atgtcgaggc catgccgtta atgaaagcga acgaggccct ggaccgcctg  960
cgtcgtggcg acgtcagtgg cgctctggtt ttggtgccat aa                     1002

SEQ ID NO: 103          moltype = DNA  length = 1095
FEATURE                 Location/Qualifiers
source                  1..1095
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 103
gacgcatacg ctgcaattat cgagcgtcag ggtggagaat tcgttctgga taacgtatct  60
atcgaggatc cgcgcgatgg cgaagtgctg gttaaggttg ccgcagctgg catgtgtcat  120
accgatctga cggttcgcga tcaatattac ccgacgccgc ttccggcggt gctgggccac  180
gaaggtagcg gcgttgttga aaaagtggga cgtggcgtca ccactgtcaa accaggtgac  240
aaagtagtgt tatccttcag ctattgcggt acttgtcctt cgtgcctcaa agggcatcag  300
gcatactgtc cgagcctgtt cccgttaaat ttcatgggcc gtcgcctgga tggttcaacg  360
cccattacac gcaacggtca agaggtcaac gcctgctttt tcgggcaatc ctcttttgcg  420
acctatagta ttgcgtcaga aaacaattgc gtcaaggttg ccgacgatgc acagattgaa  480
cttttgggcc cactgggctg cggcattcag accggtgcgg gaagtatttt aaatgctctt  540
tgtcccgaac ctggttcctc tatagcgatc tttggggtgg ggagtgtagg cttaagcgcc  600
gtgatggctg ctaaagcatc gggctgcttg aagatcatcg cggttgacag aaatgcaggt  660
cgcttggaac tggcgcgtga actgggcgcc accgatgtga ttgacgccaa cacggtcaat  720
gctcaggaag cgatcgtcgc gatgactggt ggcggcgccg actatgcaat ggataccaca  780
gccattccag cggtgctgcg gagtgcggtg gatagcacgc acaatatggg tgaaacagca  840
gtggtgggcg gggcgaaact gggtaccgag ttttcactag acatgaataa catgctgttt  900
ggtcgaaaat tgcgtggcgt agtcgaagga tcgagcacgc ctcaggtgtt catcccgcaa  960
ctgattgcga tgcagaaagc cgggctgttt ccgtttgaga aactctgtac cttttatgat  1020
ctggatcaga tcaaccaggc cgtagaggat accgaaaaga ctggaaaagc gataaaagcc  1080
attctcaaaa tgtaa                                                   1095

SEQ ID NO: 104          moltype = DNA  length = 1275
FEATURE                 Location/Qualifiers
source                  1..1275
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 104
tctacacagc ctgcaaccat agctgattcc gcgaccgatc tggttgaggg tcttgcacgt  60
gcagcccgtt ctgcgcagcg ccagttggcg cggatggatt caccggtaaa agaacgcgcg  120
ctgacgttag ccgctgcagc gctgcgtgcc gctgaggccg aaatttttagc cgctaacgcg  180
caggatatgg cgaatggcgc agcaaacggc ctgtcctcgg ccatgctcga ccggctgaag  240
ttaacgccag agcgtctggc cggcattgcc gatgctgtgg cgcaagtcgc cgggctggcc  300
gatccggtcg gcgaggtgat cagtgaagct gcgcgtccga atggcatggt gctgcagaga  360
gtgcgtattc cggtcggagt tatcggcatc atttacgaaa gccgccccaa cgttaccgcc  420
gatgcagcag cgctctgcgt gcgttcaggt aatgcggcga ttctgcgcgg tggctcggaa  480
gcggttcata gtaaccgtgc gatccataaa gcgctggttg ctgggcttgc cgaaggcgga  540
gtgccggcag aagcggtgca gcttgtacct acgcaggacc gtgctgccgt aggggcaatg  600
ctaggtgccg cgggactgat cgacatgatc gttccgcgcg gcggaaaaag ccttgtcgct  660
cgcgtccagg cagatgcccg cgtgccggtg ttagcacact tggacggtat caaccacacg  720
tttgttcatg ccagtgcaga tccggcgatg gcccaagcga tagtgttgaa tgccaaaatg  780
cgtcgcaccg gcgtttgtgg tgcgatggaa accctgctga ttgacgcgac ttatccagat  840
ccccacggcc tggtcgaacc gctgctagac gccggttgcg agctgcgcgg cgatgctcga  900
gcgagagcaa ttgatccgag gattgcgcca gctgccgaca acgactggga tacagaatat  960
ttggaagcga ttctttcggt tgcagtggtc gacggtttgg atgaagcgct cgcccacatc  1020
gcgcgccatg cctctggtca taccgatgca atcgtcgcgg cggaccaaga tgtggcagac  1080
cgattcttag ctgaagtaga tagcgcaatt gtaatgcata atgcatccag ccagtttgct  1140
gatggcggtg agttcggcct gggtgctgag attggtattg ccacggggcg gctgcacgcg  1200
cgcggccctg tagcgctcga agggctgact acctacaaat ggctggtgcg cggaagcggg  1260
caaactcgtc cataa                                                   1275

SEQ ID NO: 105          moltype = DNA  length = 1422
FEATURE                 Location/Qualifiers
source                  1..1422
                        mol_type = other DNA
```

-continued

organism = synthetic construct

SEQUENCE: 105

```
cgcgaacggc tacagcaata cattgatgga aagtgggtag acagtgaagg tggcaaacgt  60
cacgaagtca ttaatccgac tacagaggaa ccctgttgtg tgattacgct gggcacgcaa  120
gcagatgtcg acaaagcagt ggccgcggca cagcgcgcct ttaaaacctt cagcaaaacg  180
acgcgtgagg aacgactggc gctgcttgaa cgcatcgtag aagaatacaa gaagcgtgtc  240
cctgatttag ccgccgcgat ggccgaggaa atgggagctc cggtaagctt tgccagcacc  300
gcgcaagttg gcgccggaat cggagcattt ctgggcacca tggccgcgct ccgtaatttc  360
tcctttgttg aggacaacgg tgcgtttaaa gtggcctacg aaccgatagg tgttgtgggt  420
atgattacgc catggaactg gccactgaat cagatagctc tgaaagtagc accggcgctg  480
gccgcgggga ataccatgat cctgaaaccg tccgaggaat gcccaaccaa cgcagcgatc  540
tttaccgaaa ttttggatgc cgcaggggtt ccgccagggg tttttaacct gattcagggc  600
gatggtcctg gtgtaggcac tgcgatcagt agtcatccgg gcattgatat ggttagtttc  660
accggttcga cccgtgcggg catcctcgtg gcgaaagctg cggccgatac cgtcaagcgg  720
gtgcatcagg aacttggcgg taaatctccc aatgtggtgc tgcccgatgc agacttcgca  780
aaatatctgc cgtctaccgc gtcaggcccg ttggtgaaca gcggccagag ctgcatttcg  840
ccaacccgta ttttagtacc aagagaacgc gaagcagaag ccgcggcttt tgtttctgcg  900
atgtactccg caacaccggt cggggatccg atgcaagaag gtgcgcacat tgggccggtg  960
gttaacaaag ctcagtttga caagatccgc ggtctgattc aatcggcaat agacgaaggc  1020
gcgaaactcg agacaggcgg gcccgactta ccggccaatg tgaaccgcgg ctattatatc  1080
aaaccaacgg tcttttcagg cgttactcct gatatgcgca ttgctcagga agaaatcttc  1140
ggcccggtgg cgacgattat ggcgtacgat tcattagagg aggccattga gatcgcaaat  1200
gatacagcct atggactgtc ggcctgcatt actggtgatc cggcgaaagc ggctgaagtc  1260
gctcctgagc ttcgtgcagg tatggtggct atcaataact ggggccctac tccgggtgct  1320
ccgttcggtg gctataaaca gtccggtaac ggtagggagg gagggttgta tgggttgaaa  1380
gacttcatgg aaatgaaagc gatcagcggc ctgcctgcct aa                      1422
```

SEQ ID NO: 106 moltype = DNA length = 1344

FEATURE Location/Qualifiers

source 1..1344

mol_type = other DNA

organism = synthetic construct

SEQUENCE: 106

```
actgccccta ccgccgcaga cctttccgcc gatattgcac gggtttttgc actgcaacaa  60
gcgcacatgt gggaggccaa ggcgtccacc gcggcggagc gcaaagaaaa attggcgcgt  120
ctgaaggccg cggttgaagc acacgcggat gacattgtgg ccgcggttct ggaagatacg  180
cgcaaacctg ttggtgaaat aagggtgacc gaagttctga atgtaaccgc caatatccag  240
cgaaacatcg ataatctcga tgaatggatg aaaccggtcg aggtcgctac ctcactgaat  300
ccagcggacc gcgcgcagat aattcatgaa gcgcgcggcg tatgcctgat tcttggccca  360
tggaatttcc ccttaggtct ggcgctgggt ccggtcgccg ctgctatcgc cgcaggcaat  420
acttgtatcg tgaaattaac ggacttgtgt ccagcgaccg caagagtggc atcggtgatc  480
gtgcgtgaag cgttcgatga aaaagatgtg gctctgtttg agggagacgt tagtgtagct  540
accgcgcttt tggatctgcc gtttaatcat gtatttttta caggctctcc acgtgtaggc  600
aaaattgtga tggctgctgc ggcaaagcat ctgaccagcg tcacgttaga gcttggtggg  660
aagtctcccg ttattgtcga tgatagcgca gatatcgatc aagttgctgc ccagttagcc  720
gcggccaaac aattcaacgg cgggcaggcc tgcatttccc cggactatgt gtttgtgaaa  780
gaagacaaaa aagctgcgct ggtagaaggt ttccgtgcca atgtgcagaa aaacttgtat  840
gatgatgcag gcaacctgaa aaaagacagt attgcacagg tggtcaacaa agcgaacttt  900
gatcgtgtga aagccatgtt cgacgatgca gtcgcaaaag gcgcgaccgt cgccgctggt  960
ggaacgtttg aagcggatga cttgactatt catccgacaa tgctgacagg cgtaaccccg  1020
cagatgacta ttctccagga tgagatcttt gcccctgtca ttccggtgat gacctacgac  1080
acgctggatc aagcgatcgg gtatatcgaa gcacgcgaca aaccgctagc actctatgtt  1140
tacagtaaag atgaagcgaa cgttgaaaag gtcttagccc gcacgtcatc gggtggtgtt  1200
acggtgaatg gtgtgttctc gcactacctg gaaaacaacc tgccgttcgg gggggttaac  1260
acaagcggta tgggcagcta ccatggcgtg ttcggattta agtgctttag ccacgagcgg  1320
gctgtatatc gtcatcagca gtaa                                          1344
```

SEQ ID NO: 107 moltype = DNA length = 1023

FEATURE Location/Qualifiers

source 1..1023

mol_type = other DNA

organism = synthetic construct

SEQUENCE: 107

```
ggttaccggg ttgtagtggt gggtgcgact gggaatgtgg ggcgtgaaat gctgaacatt  60
ctggcagaac gcgagtttcc ttgtgacgag atcgcagcgg ttgctagctc tcgttcgcag  120
ggcaccgaaa tagaatttgg cgaaactggc cggaagctga aagtacagaa tgttgaaaat  180
tttgatttta ccggatggga cattgcactg tttgcggcgg gatcaggccc gacgcagatc  240
catgctccac gtgccgcttc tcagggctgc gtggtgatcg ataacagtag cttataccgc  300
atggacccgg acgtgcctct gatcgtgccc gaggtgaatc cggatgcgat tgatggctat  360
accaaaaaaa acattattgc caatccaaac tgttccaccg cgcaaatggt cgtggcgctg  420
aaaccgttac atgatgccgc caaaattaaa agagttgtcg tctccacgta tcaaagcgtt  480
tccggcgcgg gtaaagaagg gatggatgaa ctgttcgaac aaagccgcgc gatatttgtc  540
ggggacccgg tggaaccgaa aaaattcacc aaacagatcg cattcaacgt gatccctcat  600
atcgatgtat tcctagacga tggttcgact aaagaagagt ggaaaatggt cgccgaaacc  660
aaaaaaattt tggaccccaa ggttaaggta acggcaacct gcgtgcgtgt gccggtgttc  720
atcggccact cggaagcgtt aaacattgag ttcgagaatg aaattagtgc cgaggaagcg  780
```

-continued

```
cagaatatcc tgcgcgaagc accaggtgtg atgctcgtcg ataagcgcga gaacggcgga  840
tatgttacgc cggtcgaatg cgttggtgat tttgccacat ttgttagccg cgtacgtgag  900
gattcaacag ttgataacgg ccttaatatt tggtgtgtca gtgataacct gaggaaaggt  960
gctgccttga acgctgtaca gattgcagaa ctgctcggtc gtcgacacct taaaaagggt  1020
taa                                                               1023

SEQ ID NO: 108          moltype = DNA  length = 1005
FEATURE                 Location/Qualifiers
source                  1..1005
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 108
gcgatcaaag ttgcgataaa cggttttgga cgtatcggga ggaatgtggc ccgcgccatt  60
ttagaacgtc ccgattgtgg gttagaactg gttagcatta acgacctggc tgatgccaag  120
gctaacgccc tgctgtttaa acgcgacagc gttcatggcg cgttcagtgg cgaagtatca  180
gtggatggca atgatctgat tgtgaatggc aagcgcattc aggtgactgc agagcgcgat  240
cctgctaacc tgccacacgg agccaatggt attgacattg cgctggaatg cacgggcttt  300
ttcaccaatc gtgatggtgg ccagaaacac ttggacgcgg gcgccaaacg cgttctgatt  360
tccgctccgg caaaaaacgt agacctgacg gtcgtctatg gtgtgaacca cgacaaactg  420
accggcgatc ataagatcgt gtccaacgcg agttgcacga ccaactgttt ggcgccgatg  480
gcaaaagtcc tgcatgaatc tatcgggatt gagcgtggtc taatgacaac gattcattcg  540
tataccaatg atcaaaaaat actcgaccag atccatagcg atcctagacg ggctcgggca  600
gcggcgatga atatgatccc cacaagcacc ggggccgcag ttgcagtggg tgaagttctg  660
ccagacttaa aagggaaact tgatggttcg tcgattcgag tcccgacccc gaacgtatct  720
gtcgtggatc ttactttcac gccgaagcgt gataccagcg tagaggaagt aaatggtctc  780
ttgaaagcgg ctgccgaagg cgcattgaaa ggcgtgttag gttacaccga cgaaccgctg  840
gtttcaatcg attttaacca cgatccgcat agttcaacaa tcgacagcct tgagactgcc  900
gtgctcgaag gtaaactggt gcgcgtcctg tcttggtacg ataatgagtg gggcttttcc  960
aaccgtatgc tggatacggc gggagcaatg gcgaaattcc tttaa                  1005

SEQ ID NO: 109          moltype = DNA  length = 1512
FEATURE                 Location/Qualifiers
source                  1..1512
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 109
aatgacatga ctaccatctc acgcacgcag cgtgaatact ccgaggccgc aaaagctttc  60
ctcgcgagaa agccgcaatt gtttattaat aacgagtggg tcgatagcag tcacgatgca  120
gtgatcgaag tggaagaccc ctcgaatggg aggattgtag gtcatgtcgt tgatgcctcg  180
gacaaagacg ttgaccgggc ggttgccgct gcgcgggccg ctttcgatga tggtcgttgg  240
tccaacctgc cgccaatggt acgcgatcgt accatgaatc gcctggccga cctgcttgaa  300
gcaaacgcag atctctttgc agagctggaa gcgattgata atggtaaacc gaagggtatg  360
gccggcgccg ttgatattcc aggtgcgata agccaactac gcttcatggc aggatgggcc  420
agcaaggtag ctggcgaaac gacgcagcct tacacgatgc cgaatggcac cgtgtttagt  480
tacaccgtca aagaacccgt cggtgtctgc gcgcagattg tgccgtggaa cttcccgctg  540
ctgatggcat cattgaagat cgccccggcg ctggcggctg gatgtacact ggtgctgaaa  600
cctgccgaac agacatcgct taccgcgtta aaactggcag atttggtggt tgaggctggc  660
tttcctgcgg gagtgatcaa cattatcaca gggaacggcc acaccgcagg tgatcgcatg  720
gtcaaacatc ccgacgtaga caaagtcgcc tttactggct ccaccgaaat cgggaaactg  780
ataaatcgaa acgcaaccac cacgcttaaa cgggttacgc tcgaactggg cgggaaaagt  840
cccgtagtgg ttatgccaga cgtagatgtg gcgcagaccg cgcctggcgt tgccggtgcg  900
attttttca acgctggcca ggtttgtgtt gccggtagtc gtttatatgc gcaccgttcg  960
gtgttcgatt ccgtgttaga aggtatgacc cagactgcgc cgttttgggc gccgcgcccg  1020
agcctggatc cagaagcaca catgggaccg ttggtcagca aagagcaaca tgaccgtgtg  1080
atgggatata tcgaggcggg caagcgtgat ggcgccagcg tagtgatggg cggtgattgc  1140
ccaagcgctg atggagggta ctatgttaat ccgacgattc tggcagacgt gaatccgcag  1200
atgtctgtcg tgcgcgagga aatttttggt ccggttgtcg tcgcccaacg cttcgacgat  1260
ttagatgaag tggcgaaaat ggcaaacgac acctgttttg gcttaggtgc gggcgtgtgg  1320
acgcgcgatg ttgcggtgat gcataaactt gcttcaaaga tcaaatctgg cactgtgtgg  1380
ggcaactgcc atgccctgat cgatacagcg ctgccttttg gcggctataa agaatctggg  1440
ctgggtcgag aacaggggcg tgccggtatt gatgcttatt tggagactaa aacagtaatt  1500
attcaaatgt aa                                                      1512

SEQ ID NO: 110          moltype = DNA  length = 1503
FEATURE                 Location/Qualifiers
source                  1..1503
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 110
gctacgcagt tgagaagtgc agaaaatgaa tatgggatca aatccgagta tggtcattat  60
ataggaggtg agtggattgc aggggatagc ggcaagacca tagatttact aaatccctct  120
accggtaaag tgctgaccaa aattcaagcc ggcaacgcaa aagatattga acgcgcgatt  180
gccgctgcaa aagcggcgtt tccgaagtgg agccagagcc tgccagggga gcgccaagaa  240
atcctgatag aggttgcgcg tcgtctgaaa gcacgccatt cgcactatgc aaccttagaa  300
acgctcaata acggtaaacc gatgcgcgaa tcaatgtatt tcgatatgcc tcaaacgatc  360
gggcaatttg agctgttcgc cggtgccgcc tatggcctgc atggccagac gctggattat  420
```

-continued

```
ccagacgcga ttggcatcgt ccaccgtgaa ccgttaggcg tatgcgcgca gattattcca  480
tggaacgtgc cgatgttgat gatggcgtgc aaaatcgcgc ccgcgctggc ctctggcaac  540
actgtcgttc tgaaaccggc cgaaacggtg tgcctttctg tgattgaatt tttcgtggaa  600
atggctgatc tgttgcctcc gggtgtgatc aacgttgtta ccgggtatgg tgctgacgtt  660
ggcgaggcgc ttgtaacaag ccctgatgta gctaaagtgg cctttaccgg ttcgattgct  720
acggcgcgcc ggattattca gtatgcctcg gccaatatca ttccacagac gctcgagttg  780
ggcggtaaat cagcgcatat cgtgtgtggc gatgccgata ttgacgcggc ggtggaaagt  840
gcgactatgt ccaccgtttt aaataaaggt gaagtctgtc tggctggttc acgcctgttt  900
ctgcatcagt ccatccagga tgaattcctg gccaaattta aaacagcgct tgaaggcatt  960
cgccaaggcg acccgctaga tatggcgact caacttggag cgcaggcatc gaagatgcag 1020
tttgacaagg tgcaaagcta cttaaggctg gctacagagg aaggggcaga ggtactgacc 1080
ggcggtagtc gttcagatgc cgcagatctg gcagatggca attttatcaa accgacggtt 1140
tttactaacg tcaataactc catgcggatc gcgcaggaag agattttcgg accggttacc 1200
agcgtaatta catggagcga cgaagacgac atgatgaaac aggccaacaa tacaacttac 1260
ggcttggctg gcggtgtctg gaccaaggac atcgcacgag cacaccgtat tgcgcgtaaa 1320
ctcgaaactg gcacggtctg gatcaatcgc tactacaacc tgaaagccaa catgccgctg 1380
ggaggttaca agcaaagtgg ctttgggcgt gaattcagcc atgaagtgct gaatcactac 1440
acccagacca aatctgtggt tgtcaacctc caggaaggtc gtaccggaat gttcgatcag 1500
taa                                                               1503
```

What is claimed is:

1. A recombinant microorganism comprising any one or more of:

one or more recombinant alcohol dehydrogenase genes encoding:

FdhA of *Novosphingobium aromaticivorans* (SEQ ID NO:2) or a homolog thereof;

Saro_0995 of *Novosphingobium aromaticivorans* (SEQ ID NO:4) or a homolog thereof; and/or

Saro_3899 of *Novosphingobium aromaticivorans* (SEQ ID NO:6) or a homolog thereof;

one or more recombinant aldehyde dehydrogenase genes encoding:

FerD of *Novosphingobium aromaticivorans* (SEQ ID NO:8) or a homolog thereof;

Saro_1104 of *Novosphingobium aromaticivorans* (SEQ ID NO:10) or a homolog thereof;

Saro_1197 of *Novosphingobium aromaticivorans* (SEQ ID NO:12) or a homolog thereof; and/or

Saro_2869 of *Novosphingobium aromaticivorans* (SEQ ID NO:14) or a homolog thereof;

a recombinant γ-formaldehyde lyase gene encoding PcfL of *Novosphingobium aromaticivorans* (SEQ ID NO:16) or a homolog thereof;

a recombinant lignostilbene dioxygenase gene encoding LsdD of *Novosphingobium aromaticivorans* (SEQ ID NO:18) or a homolog thereof; and

a recombinant aromatic acid decarboxylase gene encoding LigW of *Novosphingobium aromaticivorans* (SEQ ID NO:20) or a homolog thereof.

2. The recombinant microorganism of claim **1**, comprising any two or more of:

the one or more recombinant alcohol dehydrogenase genes;

the one or more recombinant aldehyde dehydrogenase genes;

the recombinant γ-formaldehyde lyase gene;

the recombinant lignostilbene dioxygenase gene; and

the recombinant aromatic acid decarboxylase gene.

3. The recombinant microorganism of claim **1**, comprising any three or more of:

the one or more recombinant alcohol dehydrogenase genes;

the one or more recombinant aldehyde dehydrogenase genes;

the recombinant γ-formaldehyde lyase gene;

the recombinant lignostilbene dioxygenase gene; and

the recombinant aromatic acid decarboxylase gene.

4. The recombinant microorganism of claim **1**, comprising any four or more of:

the one or more recombinant alcohol dehydrogenase genes;

the one or more recombinant aldehyde dehydrogenase genes;

the recombinant γ-formaldehyde lyase gene;

the recombinant lignostilbene dioxygenase gene; and

the recombinant aromatic acid decarboxylase gene.

5. The recombinant microorganism of claim **1**, comprising each of:

the one or more recombinant alcohol dehydrogenase genes;

the one or more recombinant aldehyde dehydrogenase genes;

the recombinant γ-formaldehyde lyase gene;

the recombinant lignostilbene dioxygenase gene; and

the recombinant aromatic acid decarboxylase gene.

6. The recombinant microorganism of claim **1**, comprising the one or more recombinant alcohol dehydrogenase genes.

7. The recombinant microorganism of claim **6**, wherein the one or more recombinant alcohol dehydrogenase genes encode:

FdhA of *Novosphingobium aromaticivorans* (SEQ ID NO:2), a protein comprising a sequence at least 95% identical to SEQ ID NO:2, an ortholog of FdhA of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of FdhA of *Novosphingobium aromaticivorans;*

Saro_0995 of *Novosphingobium aromaticivorans* (SEQ ID NO:4), a protein comprising a sequence at least 95% identical to SEQ ID NO:4, an ortholog of Saro_0995 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_0995 of *Novosphingobium aromaticivorans*; and/or

Saro_3899 of *Novosphingobium aromaticivorans* (SEQ ID NO:6), a protein comprising a sequence at least 95% identical to SEQ ID NO:6, an ortholog of Saro_3899 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_3899 of *Novosphingobium aromaticivorans.*

8. The recombinant microorganism of claim **1** comprising the one or more recombinant aldehyde dehydrogenase genes.

9. The recombinant microorganism of claim **8**, wherein, when present, the one or more recombinant aldehyde dehydrogenase genes encode:

FerD of *Novosphingobium aromaticivorans* (SEQ ID NO:8), a protein comprising a sequence at least 95% identical to SEQ ID NO:8, an ortholog of FerD of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of FerD of *Novosphingobium aromaticivorans;*

Saro_1104 of *Novosphingobium aromaticivorans* (SEQ ID NO:10), a protein comprising a sequence at least 95% identical to SEQ ID NO:10, an ortholog of Saro_1104 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_1104 of *Novosphingobium aromaticivorans;*

Saro_1197 of *Novosphingobium aromaticivorans* (SEQ ID NO:12), a protein comprising a sequence at least 95% identical to SEQ ID NO:12, an ortholog of Saro_1197 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_1197 of *Novosphingobium aromaticivorans*; and/or

Saro_2869 of *Novosphingobium aromaticivorans* (SEQ ID NO:14), a protein comprising a sequence at least 95% identical to SEQ ID NO:14, an ortholog of Saro_2869 of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of Saro_2869 of *Novosphingobium aromaticivorans*.

10. The recombinant microorganism of claim **1**, comprising the recombinant T-formaldehyde lyase gene.

11. The recombinant microorganism of claim **10**, wherein, when present, the recombinant T-formaldehyde lyase gene encodes PcfL of *Novosphingobium aromaticivorans* (SEQ ID NO:16), a protein comprising a sequence at least 95% identical to SEQ ID NO:16, an ortholog of PcfL of *Novosphingobium aromaticivorans*, a recombinant variant of the ortholog of PcfL of *Novosphingobium aromaticivorans.*

12. The recombinant microorganism of claim **1**, comprising the recombinant lignostilbene dioxygenase gene.

13. The recombinant microorganism of claim **12**, wherein, when present, the recombinant lignostilbene dioxygenase gene encodes LsdD of *Novosphingobium aromaticivorans* (SEQ ID NO:18), a protein comprising a sequence at least 95% identical to SEQ ID NO:18, an ortholog of LsdD of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of LsdD of *Novosphingobium aromaticivorans.*

14. The recombinant microorganism of claim **1**, comprising the recombinant aromatic acid decarboxylase gene.

15. The recombinant microorganism of claim **14**, wherein, when present, the recombinant aromatic acid decarboxylase gene encodes LigW of *Novosphingobium aromaticivorans* (SEQ ID NO:20), a protein comprising a sequence at least 95% identical to SEQ ID NO:20, an ortholog of LigW of *Novosphingobium aromaticivorans*, or a recombinant variant of the ortholog of LigW of *Novosphingobium aromaticivorans.*

16. The recombinant microorganism of claim **1**, wherein the recombinant microorganism is a bacterium.

17. The recombinant microorganism of claim **1**, wherein the recombinant microorganism is an Alphaproteobacterium.

18. The recombinant microorganism of claim **1**, wherein the recombinant microorganism is from an order selected from the group consisting of Sphingomonadales, *Actinomyces*, Gammaproteobacteria, Betaproteobacteria, and Bacilli.

19. A method of catabolizing a lignin aromatic, the method comprising culturing the recombinant microorganism of claim **1** in a medium comprising the lignin aromatic to thereby catabolize the lignin aromatic.

20. The method of claim **19**, wherein the lignin aromatic comprises a β-5 linked lignin aromatic.

* * * * *