A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion

Morten O.A. Sommer[1†], George M. Church[1], Gautam Dantas[2†]

[1] Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA.,

[2] Department of Pathology and Immunology, Center for Genome Sciences, Washington University School of Medicine, 4444 Forest Park Blvd., St Louis, MO 63108 USA.

[†] To whom correspondence should be addressed. E-mail: sommer@genetics.med.harvard.edu (M.O.A.S.), dantas@wustl.edu (G.D.)

Table of Contents: **Page #**

Supplementary methods

Environmental metagenomic DNA Extraction

Soil samples (200-500g) were collected from urban parks (MA), farm land (MA), and bogs (NH). Metagenomic DNA was extracted from 10g of soil using the PowerMax Soil DNA Isolation Kit (catalog# 12988-10, Mobio Laboratories Inc.). The suggested protocol (http://www.mobio.com/files/protocol/12988.pdf) was followed with the following modifications:

1. All vortexing steps were eliminated to prevent shearing of high-molecular weight DNA.
2. Cell lysis was achieved by shaking the soil suspension at 250rpm in a 65°C water bath for 1hr, with mixing by gentle inversion every 15 minutes.
3. Inhibitor precipitation solutions C2 and C3 volumes were doubled (10mL used instead of 5mL).
4. High salt DNA-silica binding solution C4 volume was doubled (60mL used instead of 30mL).
5. Metagenomic DNA was eluted from the purification column using 7.5mL of 10mM TRIS pH 8.0.
6. Eluted metagenomic DNA (7.5mL in 50mL falcon tube) was subsequently ethanol precipitated:
    a. Add 750uL (0.1x volume) 3M sodium acetate pH 5.2, 2uL pellet paint (catalog# 69049, Novagen) and 15mL (2x volume) ice-cold 100% ethanol
    b. Invert 3-5 times to mix
    c. Incubate at room temp for 2min
    d. Centrifuge at 5000rcf (in swinging bucket table-top centrifuge) at 4°C for 45min.
    e. Discard supernatant
    f. Air-dry at 65°C for 15min
    g. Add 500uL 10mM Tris pH 8.0 and swirl to mix
    h. Dissolve by incubation at 65°C for 1hr
7. Estimate DNA concentration on Nanodrop 1000 (Thermo Scientific)

Gel purification/size selection

High molecular weight (40-50KB) metagenomic DNA was size selected and purified using a pulse-field gel apparatus (CHEF MAPPER, Biorad). Metagenomic DNA was loaded at approximately 1ug per cm width of gel (total of ~5-10ug of metagenomic DNA per library), with High Molecular Weight DNA Markers (catalog# 15618-010, Invitrogen) and 1KB DNA Extension Ladder (catalog# 10511-012, Invitrogen) as molecular weight standards. Gel conditions were:

- % Agarose: 1% Low Melting Temperature Agarose
- Buffer: 1X TBE
- Temperature: 14°C
- Voltage: 4.5 V/cm
- Pulse: initial 1.0 - final 7.0 sec
- Run Time: 13 hrs
- Angle: 120°

After electrophoresis, gel was stained with 1x SYBR-Gold Gel Stain (Molecular Probes) for 30min, imaged on dark transilluminator, and bands corresponding to 40-50KB were excised. Metagenomic DNA was extracted from the gel slices using the GELase Agarose Gel-Digesting Preparation (catalog# G09200, Epicentre Biotechnologies), using the following modified GELase protocol:

1. Transfer 0.5-2g of gel slice to a 15mL falcon tube
2. Add 2mL of 1X GELase digestion buffer per 0.5g of gel
3. Shake gently on platform shaker for ~30min
4. Discard GELase digestion buffer
5. Repeat steps 2-4
6. Divide gel slices into 1.8g of gel slice per 2mL microcentrifuge tube
7. Melt gel slices at 70°C (~10 min)
8. Cool to 45°C (~20 min)
9. Add GELase enzyme at 1U for every 200uL (approximately 200 mg) of gel
10. Incubate at 45°C overnight
11. Spin down tubes in microcentrifuge for 20min at max speed

12. Transfer top 90% of supernatant to new microcentrifuge tubes, 500uL per tube
13. Ethanol precipitate DNA:
    a. To each tube (with 500uL), add 1uL pellet paint, 50uL 3M sodium acetate pH 5.2, 1000uL 100% ice-cold ethanol
    b. Centrifuge at max speed for 5min
    c. Discard supernatant
    d. Wash with 70% ethanol, centrifuge, discard supernatant
    e. Air-dry 10 min with open cap
    f. Add 5uL 10mM Tris pH 8.0 to each DNA pellet
    g. Incubate at 55°C for 15min
    h. Resuspend pellet and pool any fractions corresponding to the same metagenomic DNA source
14. Estimate DNA concentration on Nanodrop 1000 (Thermo Scientific)

End-Repair of metagenomic DNA

Size-selected gel-purified metagenomic DNA was blunt-end repaired using the End-It DNA End-Repair Kit (catalog# ER0720, Epicentre Biotechnologies). Approximately 0.5-5ug of metagenomic DNA was end repaired in a standard 50uL reaction:

1. 10uL metagenomic DNA (0.5-5ug of pulse-field gel-purified DNA)
2. 5uL 10x End-It Buffer
3. 5uL 2.5mM dNTP (End-It)
4. 5uL 10mM ATP (End-It)
5. 24uL sterile H_2O (up to 50uL)
6. 1uL End-It Enzyme Mix
7. Incubate at room temp for 45min
8. Incubate at 70°C for 10min

Phenol/Chloroform Extraction and Ethanol Precipitation of End-Repaired metagenomic DNA

End-repaired metagenomic DNA was phenol/chloroform extracted in two steps and concentrated by ethanol precipitation:

1. Phenol-Chloroform-Isoamyl alcohol (PCI) Extraction
    a. 50uL: end-repaired-metagenomic DNA
    b. 350uL: H_2O
    c. 400uL: PCI
    d. Invert to mix, ~3-5mins
    e. Centrifuge at max speed for 10min
    f. Save supernatant (PCI extracted end-repaired metagenomic DNA)
2. Chloroform-Isoamyl alcohol (CI) Extraction
    a. 400uL: PCI extracted end-repaired metagenomic DNA
    b. 400uL: CI
    c. Invert to mix, ~3-5mins
    d. Centrifuge at max speed for 10min
    e. Save supernatant (CI extracted end-repaired –metagenomic DNA)
3. Ethanol precipitation
    a. 400uL: CI extracted end-repaired metagenomic DNA
    b. 40uL: 3M ammonium acetate pH 5.2
    c. 800uL: ice-cold EtOH
    d. Invert to mix
    e. RT for ~2min
    f. Centrifuge at max speed for 5min
    g. Remove supernatant, save pellet
    h. 1000uL 70% EtOH
    i. Invert to wash, ~3-5min
    j. Centrifuge at max speed for 5min
    k. Remove supernatant, save pellet
    l. Air-dry ~15min
    m. Dissolve pellet in 10uL 10mM Tris pH 8
4. Estimate metagenomic DNA concentration on Nanodrop 1000 (Thermo Scientific)

Determination of inhibitory concentrations of 7 biomass chemicals to *E. coli*

The inhibitory concentrations of lignocellulosic compounds in LB-agar (Supplementary Table I) were determined for two versions of the *E. coli* strain used to create the metagenomic DNA libraries – a strain with a control pCC1FOS fosmid insert containing *E. coli* genomic DNA, and an untransformed strain. In all cases, the lignocellulosic compound inhibitory effects, growth

rates and biomass yields were found to be identical between these two strains. Accordingly, to control for the effect of the fosmid vector backbone, the *E.coli* strain with the control fosmid was used for all subsequent control comparisons against the metagenomic DNA fosmid library clones. Four concentrations for each lignocellulosic compound were tested (Supplementary Table I), based on inhibitory concentrations for *E. coli* previously reported (Zaldivar and Ingram, 1999a; Zaldivar *et al*, 1999b, 2000). Approximately 1 x 10^6 *E. coli* cells were spread on each LB-agar plate containing each inhibitor at each specific concentration (and 12.5ug/mL chloramphenicol for the strain containing the control fosmid insert), and growth of colonies was assayed after 48hrs of growth at 37°C. The lowest concentration of each compound tested which prevented colony formation at this time was denoted the selective inhibitory concentration (Supplementary Table I).

Additional details on subcloning metagenomic genes implicated in tolerance phenotypes

All primers for subcloning metagenomic genes were designed to ligate into the multiple cloning site of pZE21-MCS-1:

pZE21-MCS-1

```
            EcoRI        RBS     KpnI  ApaI      SciI  SalI
ACGCACTGACCGAATTCATTAAAGAGGAGAAAGGTACCGGGCCCCCCCTCGAGGTCGACG
TGCGTGACTGGCTTAAGTAATTTCTCCTCTTTCCATGGCCCGGGGGGGGTGCTCCAGCTGC

        HindIII                    SmaI
   ClaI         EcoRV        PstI  XmaI  BamHI        MluI
GTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCCATGGTACGCGTGCTA
CATAGCCATTCGAACTATAGCTTAAGGACGTCGGGCCCCCTAGGGTACCATGCGCACGAT
```

The monocistronic constructs were created by amplifying from the mgSyrAld and mgFurAc fosmids using the following primers:

All Forward primers have a KpnI site to allow for in frame expression in the pZE21_MCS1 vector: *ATCGTC***GGTACC** (*Filler***KpnI site**)

All reverse primers have a HindIII site for cloning into the pZE21_MCS1 vector: *ATCGTC***AAGCTT** (*Filler***HindIII site**)

mgUdpE:

Amplicon size: 1044

>UDP_ATG_For_60C

*ATCGTC***GGTACC**ATGgcgatctcctctggatcg

>UDP_Rev_59C_II

*ATCGTC***AAGCTT**ttagccggcggcg

mgRecA:

Amplicon size: 1026

>RecA_For_60C

*ATCGTC***GGTACC**atgggattatcgacggataagc

>RecA_Rev_60C_II

*ATCGTC***AAGCTT**ctacgcctccttctcaacg

mgOrfX:

Amplicon size: 358

>ORFX_ATG_For_59C

*ATCGTC***GGTACC**ATGagaccgaaattttatcgaaacga

>ORFX_Rev_59C_II

*ATCGTC***AAGCTT**ttattcggacttcttctccgt

The bicistronic mgRecA_mgOrfX construct was created by amplifying the RBS (underlined in red above) and mgOrfX using the following primers, which also add XbaI at the 3’ of mgOrfX:

>EcoRV_pzeRBS_ORFX_FOR_65C

5’-*ATCTTC***GATATC**attaaagaggagaaaggtacc-3’

>XbaIMluI_ORFX_REV_65C

5’-*GATAAT***ACGCGT***AAATAT***TCTAGA**ttattcggacttct -3’

The tricistronic mgRecA_mgOrfX_mgUdpE construct was created by amplifying the RBS and mgUdpE using the following primers:

>XbaI_pZERBS_UDPE_FOR_64C

*ATCTTC***TCTAGA**attaaagaggagaaaggtacc

>MluI_UDPE_REV_74C

*GATAAT***ACGCGT**ttagccggcggcggc

Supplementary discussion

Structural modeling of gene products required for improved phenotypes

We attempted to model the tertiary structure of the 3 gene products from mgFurAc and mgSyrAld identified to be responsible for the improved tolerance. The 3D-Jury structure prediction meta-server (Ginalski *et al*, 2003) returned high-quality consensus predictions for the mgSyrAld gene annotated to be a UDP-glucose 4-epimerase and the mgFurAc gene annotated to be a RecA protein. The best scoring consensus structure prediction for the mgSyrAld UDP-glucose 4-epimerase was obtained from the FFAS03 structure prediction server (Jaroszewski *et al*, 2005), which computed a model with significant homology to chain A of the 2.37Å x-ray crystal structure of the *Thermus thermophilus* HB8 UDP-glucose 4-epimerase (PDB ID: 2P5U) (Berman *et al*, 2000), with a 3D-Jury consensus similarity J-score of 244.86 for the 348 amino-acid query. The best scoring consensus structure prediction for the mgFurAc RecA was obtained from the SAM-T02 HMM-based structure prediction server (Karplus *et al*, 2003),which computed a model with significant homology to chain A of the 3.10Å x-ray crystal structure *Mycobacterium smegmatis* RecA protein (PDB ID: 2OEP) (Berman *et al*, 2000), with a 3D-Jury consensus similarity J-score of 269.88 for the 342 amino-acid query.

None of the protein structure prediction servers queried by the 3D-Jury structure prediction server were able to return significant models for the mgFurAc hypothetical protein mgOrfX. We were able to obtain high confidence topology predictions from the Phobius server (Kall *et al*, 2004), indicating that the protein contains two transmembrane helices (Supplementary Figure 3).

Putative roles of selected metagenomic genes for inhibitor tolerance

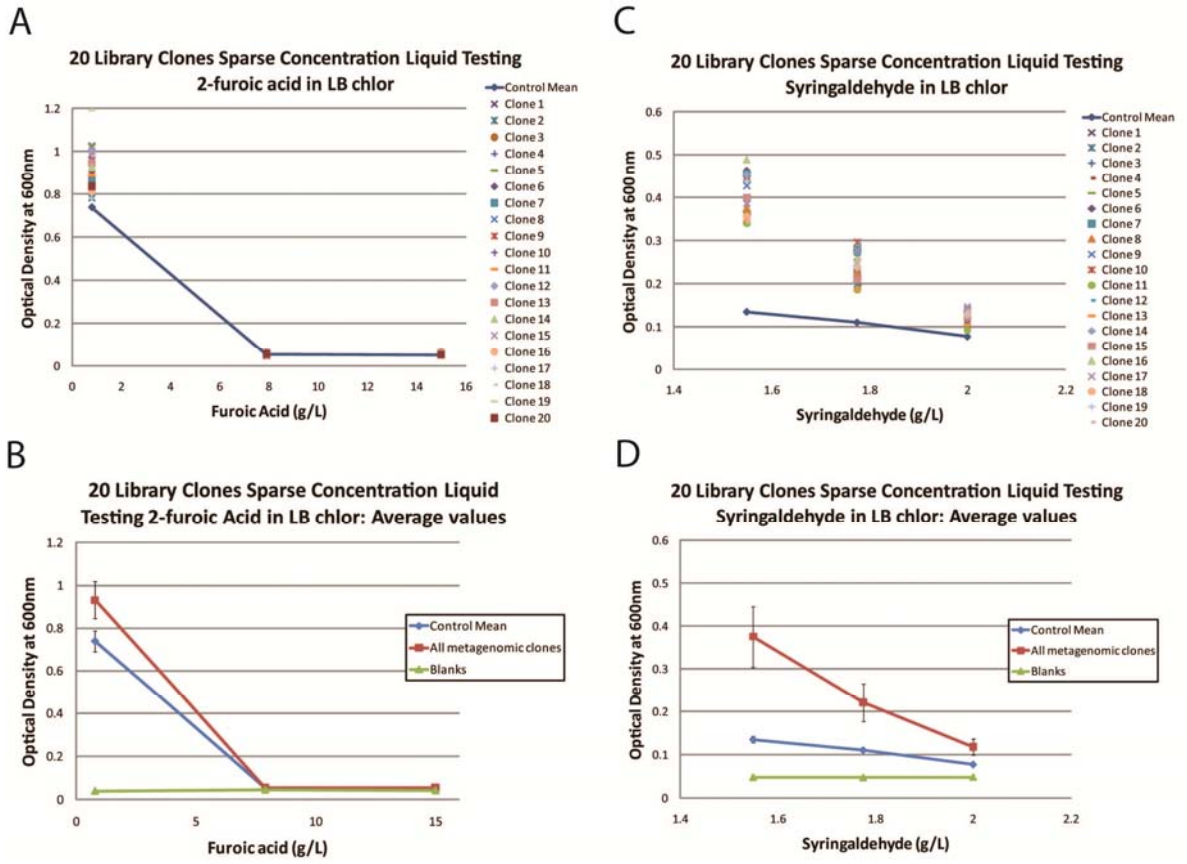
The *E. coli* UDP-glucose 4-epimerase, *galE*, is a key metabolic enzyme required for the interconversion of UDP-glucose and UDP-galactose. While the exact mode of toxicity of syringaldehyde is unknown, a number of substituted phenolic compounds have been found to inhibit UDP-glucose 4-epimerases (Thoden *et al*, 1996; Urbaniak *et al*, 2006). Deficiency of this enzyme leads to compromised cell wall biosynthesis in the absence of galactose (Nikaido, 1961), or to cell death in the presence of galactose (Yarmolinsky *et al*, 1959). Galactose is a major constituent of the biomass polymer hemicellulose (Zaldivar *et al*, 2001). Since syringaldehyde is a substituted phenolic compound, its mode of toxicity in *E.coli* may involve inhibition of *galE*, which would compromise cellular integrity or convert a biomass substrate into a toxin. The improved phenotype conferred by our selected metagenomic insert may therefore function through rescue of a compromised *E.coli* UDP-glucose 4-epimerase.

The RecA family of proteins function in recombinational DNA repair in bacteria and are required for the initiation and regulation of the SOS DNA damage response (Cox, 2007). RecA has recently been shown to remediate hydroxyl radical damage resulting from the action of bactericidal antibiotics targeting different and unrelated cellular pathways (Kohanski *et al*, 2007). 2-furoic acid and its derivatives have previously been shown to have mutagenic and antimicrobial activities (Grunberg and Titsworth, 1973; Wang *et al*, 1975) as well as forming reactive oxygen species resulting in damage of mitochondrial membranes, chromatin and actin in yeast (Almeida *et al*, 2007). Hence the metagenomically selected RecA protein may function to remediate DNA damage resulting from 2-furoic acid.

The other gene that is responsible for the improved 2-furoic acid tolerance is a 111 amino acid membrane protein of unknown function. Interestingly, a transposon insertion at residue 82 in a region predicted to be cytoplasmic did not affect the phenotype. It has previously been hypothesized that 2-furoic acid also affects membrane integrity (Zaldivar *et al*, 1999a), and these data suggest that the membrane traversing regions of the metagenomically selected hypothetical protein may contribute to the improved tolerance to 2-furoic acid.

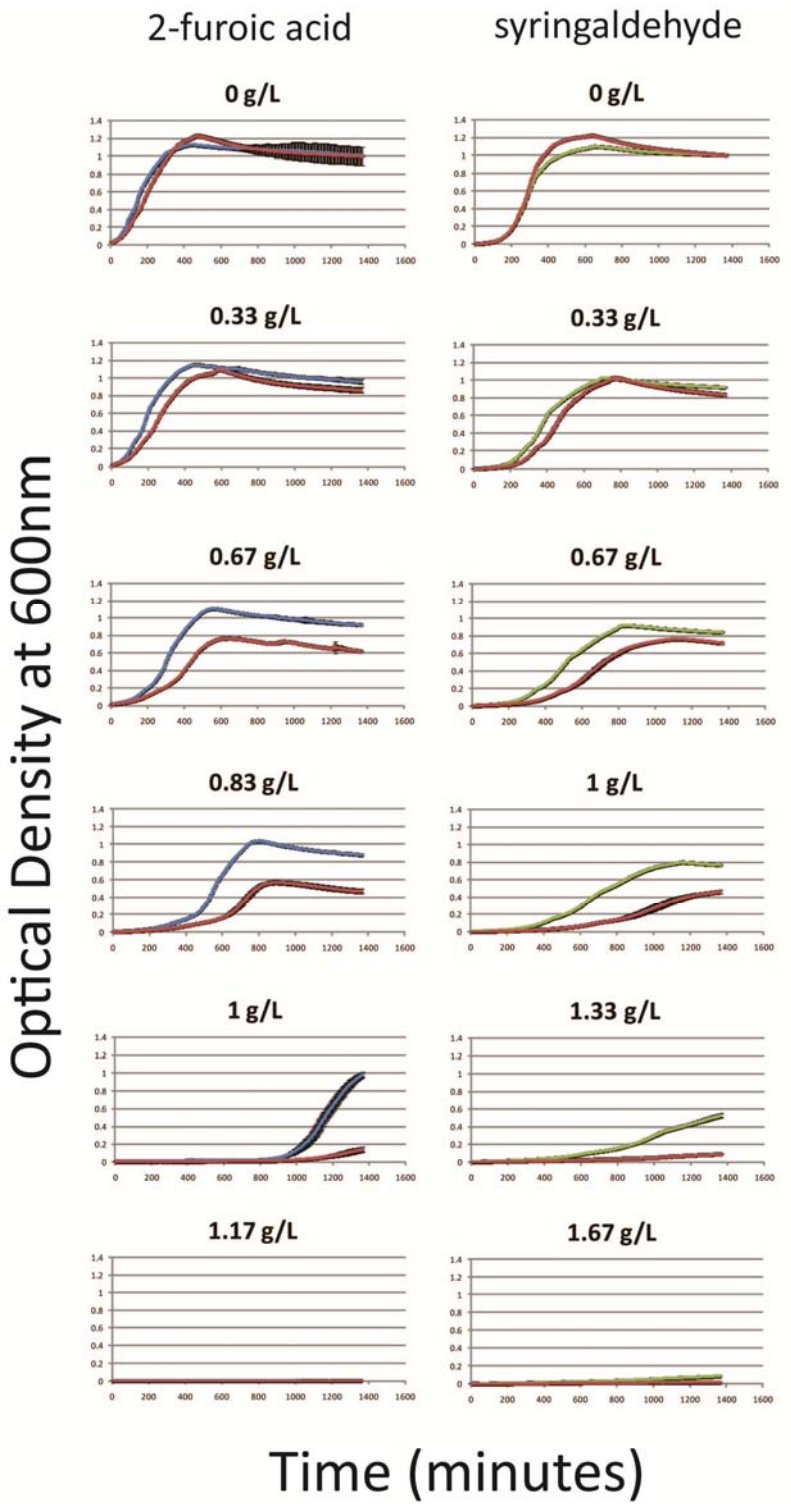
Supplementary figures:

Supplementary figure 1:



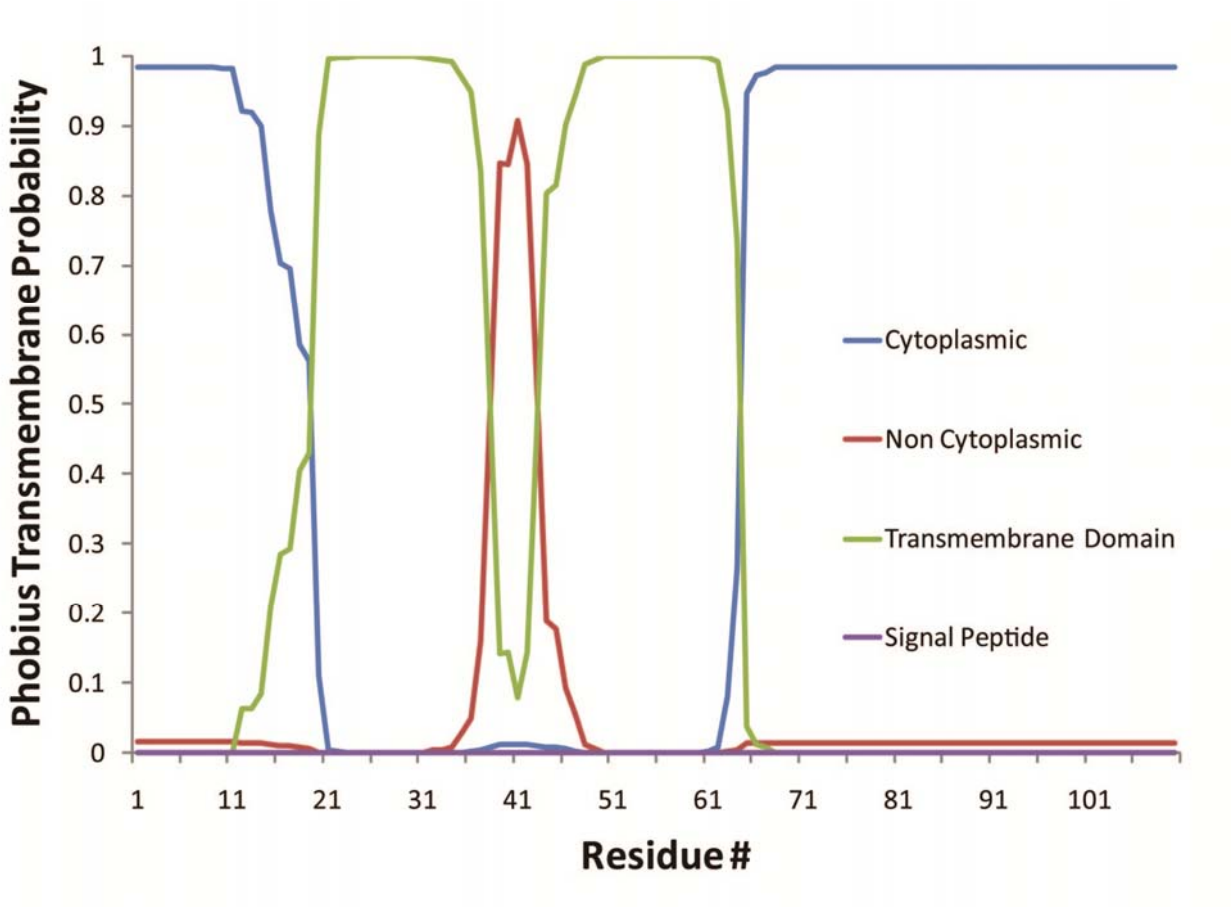
Supplementary Figure 1: End-point growth measurements of *E. coli* containing selected metagenomic functional genes encoding resistance to two lignocellulosic inhibitors. End-point growth measurements of 20 metagenomic clones each conferring *E. coli* containing mgFurAc (left panels, blue) and mgSyrAld (right panels, green) versus the *E. coli* control (all panels, red) are shown at 6 different concentrations of 2-furoic acid and syringaldehyde. Each plot shows the mean of triplicate readings with standard-deviation shown as error-bars for each metagenomic DNA clone and *E. coli* control, measured as 600nm readings every 5min over 24hrs at 37°C with shaking in a Versamax microplate reader.

Supplementary figure 2:



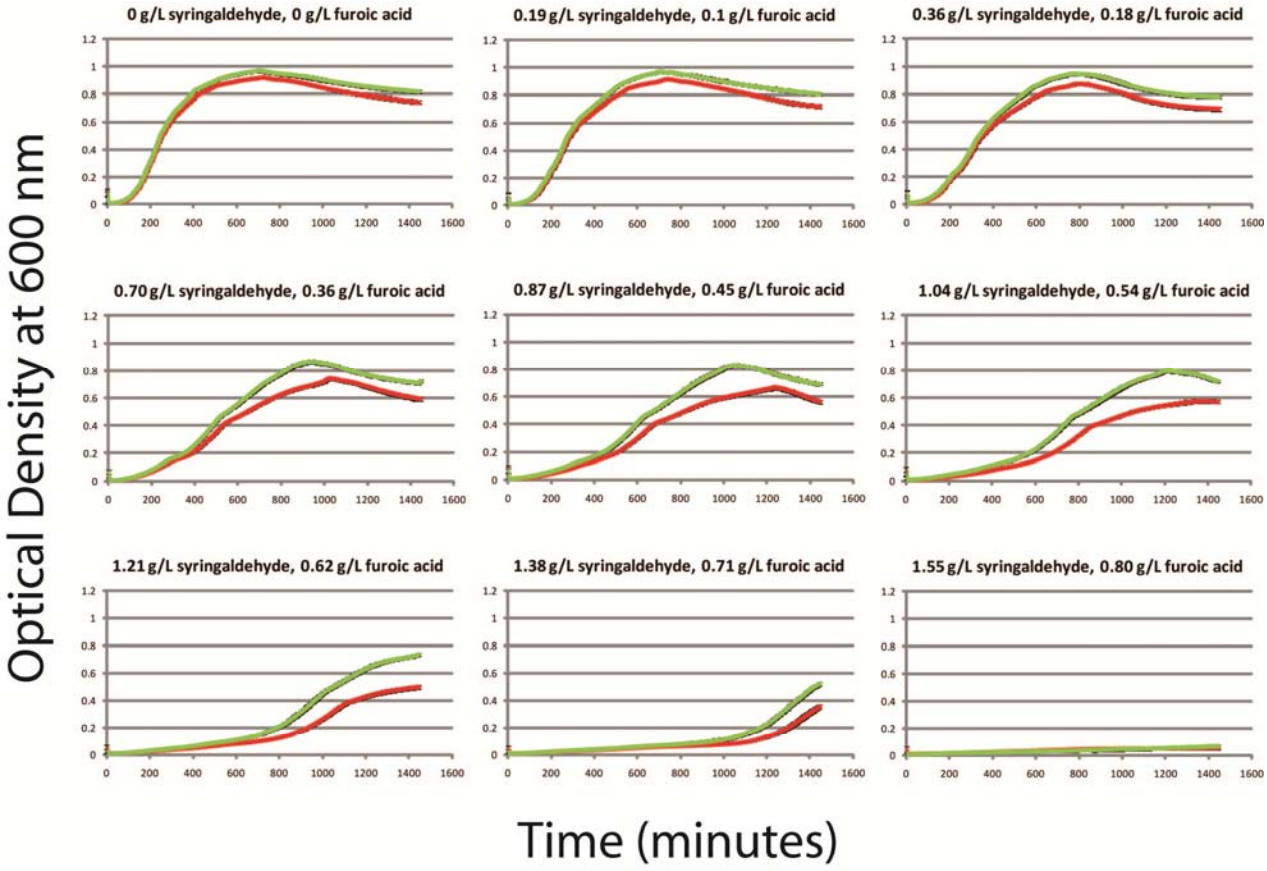
Supplementary Figure 2: Growth kinetics of *E. coli* containing selected metagenomic functional genetic elements encoding resistance to two lignocellulosic inhibitors. Growth kinetics of *E. coli* containing mgFurAc (left panels, blue) and mgSyrAld (right panels, green) versus the *E. coli* control (all panels, red) are shown at 6 different concentrations of 2-furoic acid and syringaldehyde. Each plot shows the mean of triplicate readings with standard-deviation shown as error-bars for each metagenomic DNA clone and *E. coli* control, measured as 600nm readings every 5min over 24hrs at 37°C with shaking in a Versamax microplate reader.

Supplementary figure 3:



Supplementary Figure 3: Membrane topology prediction of the 111 amino-acid mgFurAc hypothetical protein responsible for the 2-furoic acid tolerance phenotype, as predicted by the Phobius Server (Kall *et al*, 2004).

Supplementary figure 4:



Supplementary Figure 4: Growth kinetics of *E. coli* containing the three tolerance genes mgRecA, mgOrfX and mgUdpE in the presence of the inhibitor combination syringaldehyde and 2-furoic acid. Growth kinetics of *E. coli* containing mgRecA_mgOrfX_mgUdpE (green) versus the *E. coli* control (red) are shown at 9 different concentrations of 2-furoic acid and syringaldehyde. Each plot shows the mean of triplicate readings with standard-deviation shown as error-bars for each metagenomic DNA clone and *E. coli* control, measured as 600nm readings every 5min over 24hrs at 37°C with shaking in a Versamax microplate reader.

Supplementary Tables:

Supplementary Table I: Range of concentrations of lignocellulosic compounds scanned to determine inhibitory concentration (highlighted in red and bold border for each compound) for *E. coli*.

Lignocellulosic compound	Conc-1 [g/L]	Conc-2 [g/L]	Conc-3 [g/L]	Conc-4 [g/L]
Hydroquinone	0.5	0.7	**1.85**	3
Methylcatechol	**0.2**	0.4	0.95	1.5
4-hydroxybenzaldehyde	0.15	0.3	0.775	**1.25**
Syringaldehyde	0.3	0.6	**1.55**	2.5
2-Furoic acid	0.4	**0.8**	9.15	17.5
Furfural	2	2.4	**2.95**	3.5
Ethanol	20	30	42.5	**55**

Supplementary Table II: Annotated features in metagenomic insert mgFurAc

Feature #	Start	Stop	Length (bp)	Function
1	35	760	726	alternate gene name: yzbB
2	854	1555	702	Ribonuclease HI (EC 3.1.26.4)
3	2031	1699	333	hypothetical protein
4	2396	4915	2520	Glycerol-3-phosphate acyltransferase (EC 2.3.1.15)
5	4809	5060	252	hypothetical protein
6	5073	6320	1248	Mannose-1-phosphate guanylyltransferase (EC 2.7.7.13)
7	6106	6816	711	COG3178: Predicted phosphotransferase related to Ser/Thr protein kinases
8	6913	7743	831	Dihydrodipicolinate reductase (EC 1.3.1.26)
9	7872	10268	2397	hypothetical protein
10	12957	10330	2628	Alanyl-tRNA synthetase (EC 6.1.1.7)
11	13986	12961	1026	RecA protein
12	14888	14313	576	2'-5' RNA ligase
13	16072	14903	1170	Ferredoxin oxidoreductase
14	17512	16079	1434	Membrane proteins related to metalloendopeptidases
15	17668	17567	102	hypothetical protein
16	19506	18238	1269	hypothetical protein
17	19500	20222	723	Multidrug resistance ABC transporter ATP-binding and permease protein
18	20176	21663	1488	RNA polymerase sigma-54 factor rpoN
19	21674	22201	528	Ribosomal subunit interface protein
20	22216	22713	498	PTS system, fructose-specific IIA component (EC 2.7.1.69) / PTS system, fructose-specific IIB component (EC 2.7.1.69) / PTS system, fructose-specific IIC component (EC 2.7.1.69)
21	22739	23347	609	UPF0042 protein SYNAS_12170
22	23779	23910	132	hypothetical protein
23	24038	24457	420	PTS system, mannose-specific IIA component (EC 2.7.1.69)
24	24454	24927	474	Ribosomal-protein-S18p-alanine acetyltransferase (EC 2.3.1.-)
25	24924	25928	1005	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)

26	25956	26711	756	Triosephosphate isomerase (EC 5.3.1.1)
27	26731	27096	366	Preprotein translocase subunit SecG (TC 3.A.5.1.1)
28	27692	28336	645	Acyl carrier protein phosphodiesterase (EC 3.1.4.14)
29	28553	28350	204	hypothetical protein
30	34587	36143	1557	Serine phosphatase RsbU, regulator of sigma subunit
31	36250	37329	1080	hypothetical protein
32	39479	38472	1008	hypothetical protein
33	40098	39517	582	pXO1-120 homology; transposase for IS660
34	41616	40507	1110	Rhs family protein
35	41809	43314	1506	Transposase, IS4
36	27121	27206	86	tRNA-Leu-GAG

Supplementary Table III: Annotated features in metagenomic insert mgSyrAld

Feature #	Start	Stop	Length (bp)	Function
1	45	3284	3240	Glycosyl hydrolase, BNR repeat-containing protein precursor
2	4388	3333	1056	Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
3	5231	4443	789	Prolipoprotein diacylglyceryl transferase (EC 2.4.99.-)
4	5861	5262	600	Lipoprotein signal peptidase (EC 3.4.23.36)
5	8779	5807	2973	Isoleucyl-tRNA synthetase (EC 6.1.1.5)
6	10595	9159	1437	Potassium efflux system kefA / Small-conductance mechanosensitive channel
7	11203	10592	612	Thiamin-phosphate pyrophosphorylase (EC 2.5.1.3)
8	11463	11200	264	hypothetical protein
9	13738	11888	1851	Conserved domain protein
10	13914	14423	510	Peptide deformylase (EC 3.5.1.88)
11	14566	15960	1395	Cysteinyl-tRNA synthetase (EC 6.1.1.16)
12	16680	17630	951	3-dehydroquinate dehydratase (EC 4.2.1.10) / Shikimate 5-dehydrogenase (EC 1.1.1.25)
13	17627	19144	1518	Anthranilate synthase, aminase component (EC 4.1.3.27)
14	19141	19716	576	Anthranilate synthase, amidotransferase component (EC 4.1.3.27)
15	19713	20750	1038	Anthranilate phosphoribosyltransferase (EC 2.4.2.18)
16	20711	21532	822	Indole-3-glycerol phosphate synthase (EC 4.1.1.48)
17	21529	22146	618	Phosphoribosylanthranilate isomerase (EC 5.3.1.24)
18	22130	23329	1200	Tryptophan synthase beta chain (EC 4.2.1.20)
19	23326	24189	864	Tryptophan synthase alpha chain (EC 4.2.1.20)
20	24206	24862	657	hypothetical protein
21	24820	25677	858	Transcriptional regulator, XRE family
22	26433	26771	339	hypothetical protein
23	27499	28278	780	Protein serine/threonine phosphatase PrpC, regulation of stationary phase
24	29911	28325	1587	hypothetical protein
25	31149	30673	477	ADP-ribose pyrophosphatase (EC 3.6.1.13)
26	33566	31146	2421	DinG family ATP-dependent helicase CPE1197
27	33593	35104	1512	D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4)

28	35850	37067	1218	hypothetical protein
29	37465	38874	1410	hypothetical protein
30	38910	40256	1347	Serine phosphatase RsbU, regulator of sigma subunit
31	40263	41435	1173	Sarcosine oxidase beta subunit (EC 1.5.3.1)
32	42517	41474	1044	UDP-glucose 4-epimerase (EC 5.1.3.2)

Supplementary references:

Almeida JRM, Modig T, Petersson A, Hahn-Hagerdal B, Liden G, Gorwa-Grauslund MF (2007) Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae. *Journal of Chemical Technology and Biotechnology* **82:** 340-349.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28:** 235-242.

Cox MM (2007) Regulation of bacterial RecA protein function. *Crit Rev Biochem Mol Biol* **42:** 41-63.

Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* **19:** 1015-1018.

Grunberg E, Titsworth EH (1973) Chemotherapeutic properties of heterocyclic compounds: monocyclic compounds with five-membered rings. *Annu Rev Microbiol* **27:** 317-346.

Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A (2005) FFAS03: a server for profile--profile sequence alignments. *Nucleic Acids Res* **33:** W284-288.

Kall L, Krogh A, Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338:** 1027-1036.

Karplus K, Karchin R, Draper J, Casper J, Mandel-Gutfreund Y, Diekhans M, Hughey R (2003) Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. *Proteins* **53 Suppl 6:** 491-496.

Kohanski MA, Dwyer DJ, Hayete B, Lawrence CA, Collins JJ (2007) A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* **130:** 797-810.

Nikaido H (1961) Galactose-sensitive mutants of Salmonella. I. Metabolism of galactose. *Biochim Biophys Acta* **48:** 460-469.

Thoden JB, Frey PA, Holden HM (1996) High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol. *Protein Sci* **5:** 2149-2161.

Urbaniak MD, Tabudravu JN, Msaki A, Matera KM, Brenk R, Jaspars M, Ferguson MA (2006) Identification of novel inhibitors of UDP-Glc 4'-epimerase, a validated drug target for african sleeping sickness. *Bioorg Med Chem Lett* **16:** 5744-5747.

Wang CY, Muraoka K, Bryan GT (1975) Mutagenicity of nitrofurans, nitrothiophenes, nitropyrroles, nitroimidazole, aminothiophenes, and aminothiazoles in Salmonella typhimurium. *Cancer Res* **35:** 3611-3617.

Yarmolinsky MB, Wiesmeyer H, Kalckar HM, Jordan E (1959) Hereditary Defects in Galactose Metabolism in Escherichia Coli Mutants, Ii. Galactose-Induced Sensitivity. *Proc Natl Acad Sci U S A* **45:** 1786-1791.

Zaldivar J, Ingram LO (1999a) Effect of organic acids on the growth and fermentation of ethanologenic Escherichia coli LY01. *Biotechnol Bioeng* **66:** 203-210.

Zaldivar J, Martinez A, Ingram LO (1999b) Effect of selected aldehydes on the growth and fermentation of ethanologenic Escherichia coli. *Biotechnol Bioeng* **65:** 24-33.

Zaldivar J, Martinez A, Ingram LO (2000) Effect of alcohol compounds found in hemicellulose hydrolysate on the growth and fermentation of ethanologenic Escherichia coli. *Biotechnol Bioeng* **68:** 524-530.

Zaldivar J, Nielsen J, Olsson L (2001) Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Appl Microbiol Biotechnol* **56:** 17-34.