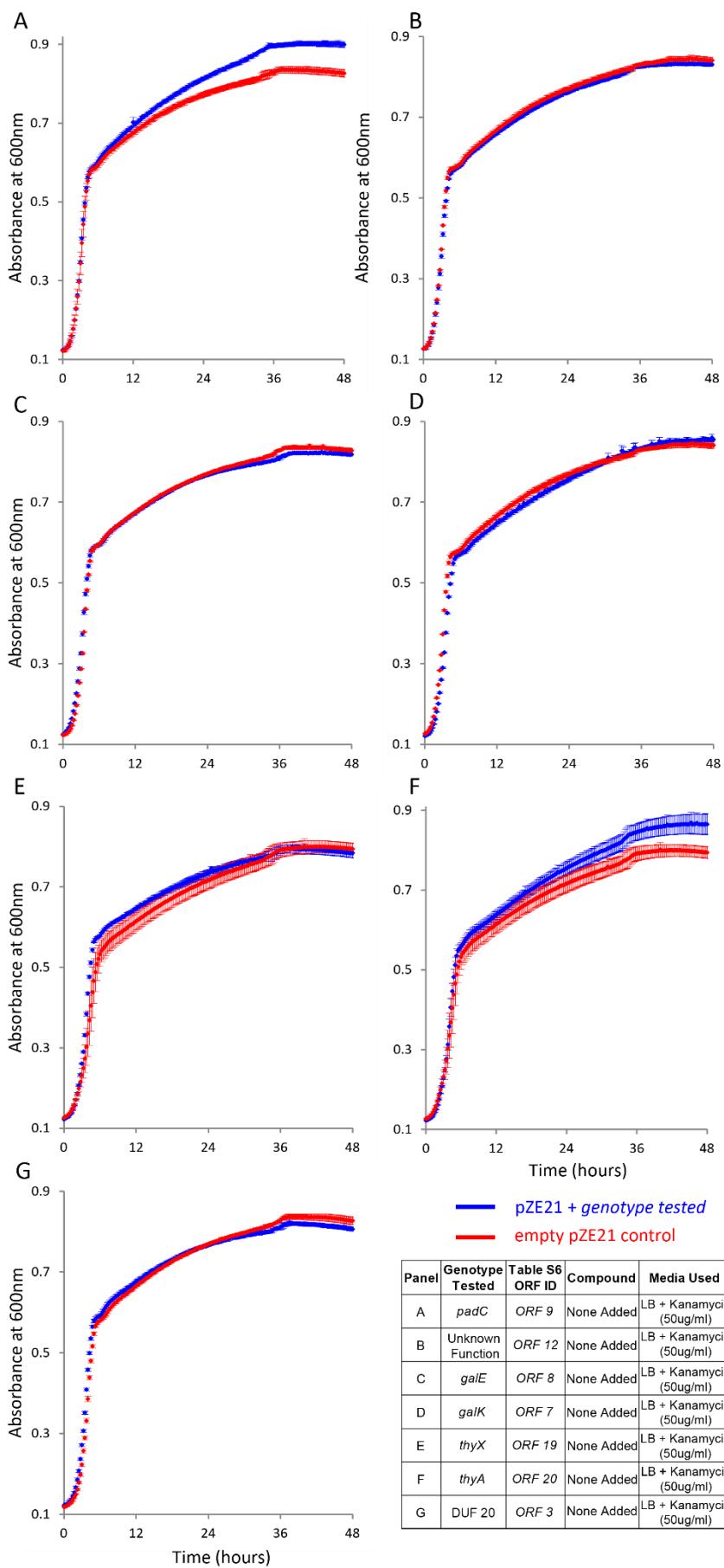


Figure S1: Selections may permit the growth of cheater transformants that lack a true tolerance phenotype. (A) A representative selection depicting local regions of extensive growth centered on large transformant colonies. This may indicate compound detoxification by 'altruistic' clones, facilitating the growth of other, non-tolerant transformants. (B) The matched control experiment performed with an empty-vector control rather than a metagenomic library.



| Panel | Genotype Tested | Table S6 ORF ID | Compound | Media Used |
|-------|------------------|-----------------|------------|--------------------------|
| A | <i>padC</i> | ORF 9 | None Added | LB + Kanamycin (50ug/ml) |
| B | Unknown Function | ORF 12 | None Added | LB + Kanamycin (50ug/ml) |
| C | <i>galE</i> | ORF 8 | None Added | LB + Kanamycin (50ug/ml) |
| D | <i>galK</i> | ORF 7 | None Added | LB + Kanamycin (50ug/ml) |
| E | <i>thyX</i> | ORF 19 | None Added | LB + Kanamycin (50ug/ml) |
| F | <i>thyA</i> | ORF 20 | None Added | LB + Kanamycin (50ug/ml) |
| G | DUF 20 | ORF 3 | None Added | LB + Kanamycin (50ug/ml) |

Figure S2: Open reading frames (ORFs) are not deleterious to *E. coli* in the absence of lignocellulosic inhibitor. The Y-axes of each panel give the absorbance at 600nm, values are the average of three trials and error bars depict standard error. Growth of *E. coli* expressing each genotype is in blue and an empty vector control (pZE21) is in red.

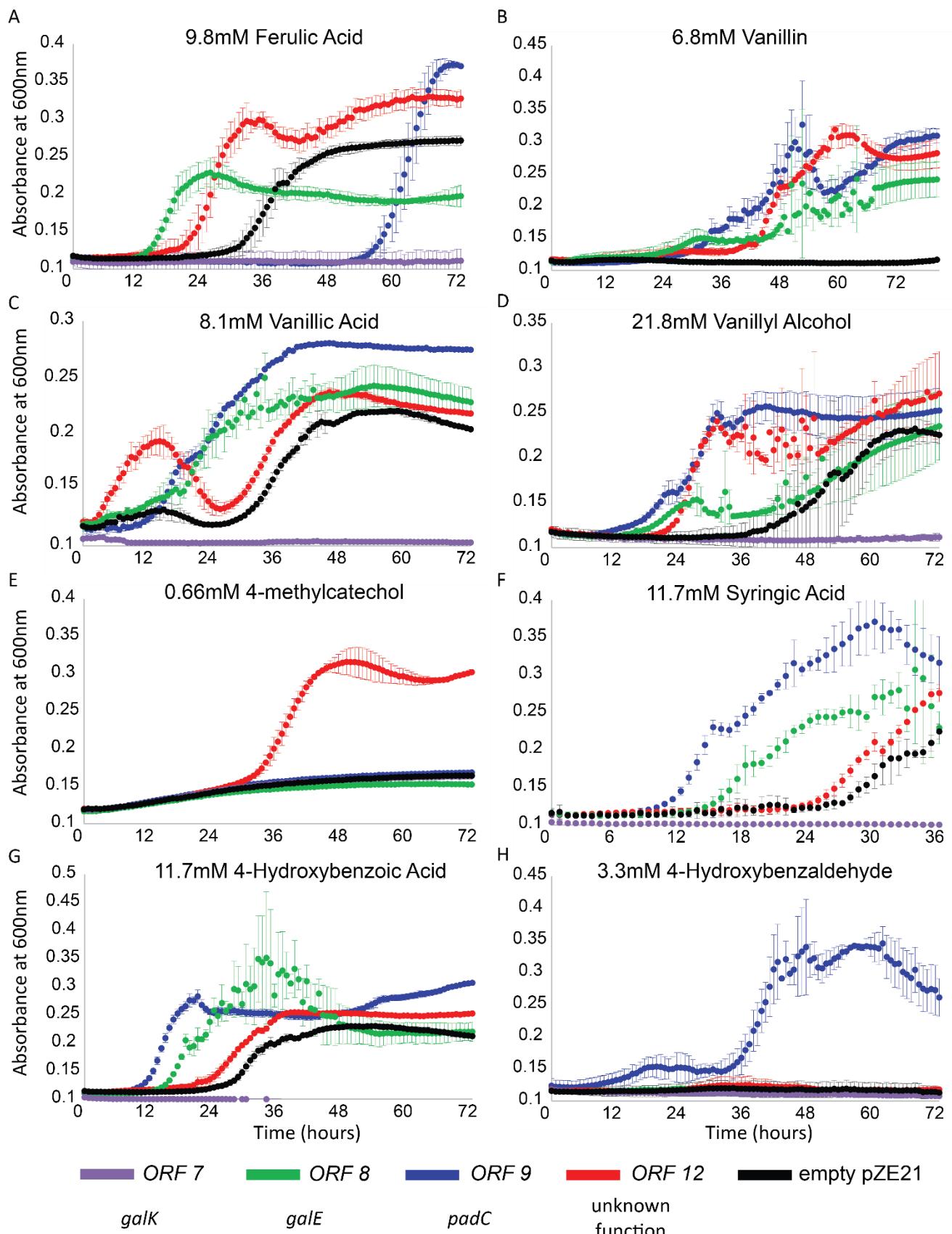


Figure S3. Ferulic acid tolerance ORFs confer tolerance against diverse phenolic inhibitors in defined (M9) medium. (A-H) The Y-axes of each panel give the absorbance at 600nm, values are the average of three trials and error bars depict standard deviation. Inhibitor concentrations are given above each panel and genotypes correspond to the colors depicted in the legend at the bottom of the figure.

Table S1: Sample information on the 16 soils used in this study. The soil identifier refers to the soil names used in Fierer *et al*, 2012 (supplemental reference 1).

| Site | Sample Name | Soil Identifier | Estimated Library Size (Gb) | Soil Type | Added N (kg ha ⁻¹ y ⁻¹) | pH |
|----------------------------|-------------|-----------------|-----------------------------|--------------|--|-----|
| Cedar Creek | S04 | CC A22 | 16.9 | Grassland | 0 | 7.3 |
| Cedar Creek | S05 | CC A45 | 7.9 | Grassland | 0 | 7.2 |
| Cedar Creek | S06 | CC A54 | 6.9 | Grassland | 0 | 7.2 |
| Cedar Creek | S07 | CC D08 | 12.2 | Grassland | 34 | 6.8 |
| Cedar Creek | S08 | CC D23 | 12.7 | Grassland | 34 | 6.9 |
| Cedar Creek | S09 | CC D38 | 13.5 | Grassland | 34 | 7 |
| Cedar Creek | S10 | CC H17 | 13.3 | Grassland | 272 | 6.1 |
| Cedar Creek | S11 | CC H40 | 7.8 | Grassland | 272 | 6.1 |
| Cedar Creek | S12 | CC H52 | 11.7 | Grassland | 272 | 6 |
| Kellogg Biological Station | S13 | KBS 201 | 6.7 | Agricultural | 0 | 6.1 |
| Kellogg Biological Station | S15 | KBS 209 | 1.4 | Agricultural | 291 | 5.5 |
| Kellogg Biological Station | S16 | KBS 301 | 3.2 | Agricultural | 0 | 6.6 |
| Kellogg Biological Station | S18 | KBS 309 | 1.5 | Agricultural | 291 | 5 |
| Kellogg Biological Station | S19 | KBS 401 | 8.0 | Agricultural | 0 | 6.2 |
| Kellogg Biological Station | S20 | KBS 404 | 6.4 | Agricultural | 101 | 6 |
| Kellogg Biological Station | S21 | KBS 409 | 3.4 | Agricultural | 291 | 5.1 |

Table S2: Compounds used in this study. MIC; minimum inhibitory concentration in LB. Solid-phase values used in initial selections with agar plates, liquid-phase values used to validate individual colonies picked from initial selections. N/A; not applicable, no tolerant clones were observed for selections with these compounds.

| Compound | Compound Classification | Solid-Phase MIC [mM] | Liquid-Phase MIC [mM] |
|-----------------------|----------------------------|----------------------|-----------------------|
| Furfuryl alcohol | Alcohol (Inhibitor) | 96.8 | 91.7 |
| 4-Methylcatechol | Alcohol (Inhibitor) | 3.6 | N/A |
| Vanillyl alcohol | Alcohol (Inhibitor) | 73.0 | N/A |
| Furfural | Aldehyde (Inhibitor) | 46.8 | 72.9 |
| 4-hydroxybenzaldehyde | Aldehyde (Inhibitor) | 9.8 | N/A |
| Vanillin | Aldehyde (Inhibitor) | 12.5 | N/A |
| Acetic Acid | Organic Acid (Inhibitor) | 14.2 | N/A |
| Caproic Acid | Organic Acid (Inhibitor) | 9.5 | N/A |
| Ferulic Acid | Organic Acid (Inhibitor) | 9.8 | 11.6 |
| Formic acid | Organic Acid (Inhibitor) | 11.9 | N/A |
| 2-Furoic acid | Organic Acid (Inhibitor) | 11.6 | N/A |
| Gallic Acid | Organic Acid (Inhibitor) | 21.7 | N/A |
| 4-hydroxybenzoic acid | Organic Acid (Inhibitor) | 9.4 | N/A |
| Syringic acid | Organic Acid (Inhibitor) | 17.7 | N/A |
| Vanillic acid | Organic Acid (Inhibitor) | 10.7 | 9.5 |
| 1-butanol | Short-Chain Alcohol (Fuel) | 229.4 | N/A |
| Ethanol | Short-Chain Alcohol (Fuel) | 1302.4 | 651.2 |
| 2-methyl-1-butanol | Short-Chain Alcohol (Fuel) | 170.2 | N/A |
| 3-methyl-1-butanol | Short-Chain Alcohol (Fuel) | 113.4 | 124.8 |
| Isobutanol | Short-Chain Alcohol (Fuel) | 269.8 | 283.3 |

Table S3: Custom primer mix used to amplify metagenomic fragments from selected transformants. Primers target the HinclI site in the pZE21 MCS1 vector and were staggered to ensure diverse nucleotide composition during early Illumina sequencing cycles.

| Primer | Sequence | Final Concentration (nM) |
|--------|---------------------------------|--------------------------|
| F1 | 5'-CCGAATTCAATTAAAGAGGGAGAAAG | 20 |
| F2 | 5'-CGAATTCAATTAAAGAGGGAGAAAGG | 20 |
| F3 | 5'-GAATTCAATTAAAGAGGGAGAAAGGTAC | 20 |
| R1 | 5'-GATATCAAGCTTATCGATAACCGTC | 8.4 |
| R2 | 5'-CGATATCAAGCTTATCGATAACCG | 17.2 |
| R3 | 5'-TCGATATCAAGCTTATCGATAACC | 34.4 |

Table S4: Primers used to amplify putative tolerance-conferring ORFs for follow-up. ORF information is given in table S6.

| Cloned ORF # | Forward Primer (5' → 3') | Reverse Primer (5' → 3') |
|--------------|---------------------------|--------------------------|
| 1 | ATGTTGATTCTTATCGGACTTTTC | TTAATCGAGTTCGAGCGCAGG |
| 2 | ATGACCTTGCGCCTTGT | TTACTGTTGACTGTCACCTGTTG |
| 3 | ATGATCACCCCCACGC | TTAACGATTCAACGCTTCAAC |
| 4 | ATGTTCAAAAAAGTACTCATTGC | TTAAGGCCGGTTCTACC |
| 5 | ATGGGCATACTCGACAAACTC | TTAAGGCCGGTTCTACC |
| 6 | ATGCTGACAACACTGATTGAG | TTACCTCGCGAAATCATG |
| 7 | ATGAAATTCCAACGAATATTGGTAC | TTAGCTGTAATTCCGGGAACC |
| 8 | TTGCACACCACCACTACGTT | TTAGGATCGGTAGCCGTGT |
| 9 | ATGAGCACTTCGATAAACATG | TTAATTTCAGATTCCGG |
| 10 | ATGTACAAGACAACACTCGAACAG | CTAACACGCATCTCCAT |
| 11 | ATGGATTATCGACGAAAGTTTT | TTACAGCGTCGTTTCATTTC |
| 12 | ATGACTGCCGTTCAACCC | TTAGCTCGGGTGGGTTGTTT |
| 13 | ATGACAAAACAACAAGCGG | TTAGTAGCGGCGGTGTTG |
| 14 | ATGTCGAAGCGCACGTT | CTAACATTCCGATGGGGATCTTC |
| 15 | GTGGCCGACAACAAGAACG | TTACGTCATCGGTGGCG |
| 16 | ATCGTTATCTCCACACCAT | CTACCAGTGGCCGGTGT |
| 17 | ATGCGCGCGGTTGC | CTACGCCACCTGCGACC |
| 18 | ATGAAATGTCCAATTGC | CGTCATAGTGCAGATGCC |
| 19 | ATGAGCGATTCAATGAGTTCTT | TCATGGACTGACCCGACC |
| 20 | ATGAACCAGTACGACGATTG | TCACACCGCGACCTC |
| 21 | ATGAACCAGTACGACGATTGCTC | TCAAATAGAGGTGCGCGTCG |
| 22 | ATGAAACGTGCTGGTGTGG | TTACAAGGAAGAGTTGCGCGTTA |
| 23 | ATGAGGAATGTCCCCATT | TCACACCAGCGCCTAAC |

Table S5: The ten most prevalent annotations by functional metagenomic selection, Pfam and TIGRFam protein family identifiers are given.

| Compound | Rank | Protein Family | Annotation | Count |
|------------------|------|----------------|---|-------|
| Furfuryl Alcohol | 1 | PF13023 | HD domain; metal-dependent phosphohydrolase | 45 |
| | 2 | PF09335 | SNARE associated Golgi protein | 25 |
| | 3 | PF01594 | Domain of unknown function DUF20 | 16 |
| | T-4 | PF09656 | Putative Actinobacterial transmembrane protein (PGPGW) | 13 |
| | T-4 | PF10173 | Mitochondrial K+-H+ exchange-related | 13 |
| | T-6 | PF00375 | Sodium:dicarboxylate symporter family | 11 |
| | T-6 | PF14125 | Domain of unknown function (DUF4292) | 11 |
| | T-6 | TIGR02611 | Putative Actinobacterial transmembrane protein (PGPGW) | 11 |
| | 9 | TIGR00387 | glcD: glycolate oxidase, subunit | 10 |
| | T-10 | TIGR00957 | multi drug resistance-associated protein | 9 |
| | T-10 | TIGR01205 | D-alanine--D-alanine ligase | 9 |
| | T-10 | PF01066 | CDP-alcohol phosphatidyltransferase | 9 |
| Ferulic Acid | 1 | TIGR01179 | galE: UDP-glucose 4-epimerase | 15 |
| | T-2 | TIGR02967 | guaD: guanine deaminase | 7 |
| | T-2 | TIGR00131 | galk: galactokinase | 7 |
| | T-2 | TIGR00904 | mreB: cell shape determining protein, MreB/Mrl | 7 |
| | 5 | PF00665 | Integrase core domain | 6 |
| | 6 | PF04389 | Peptidase family M28 | 5 |
| | T-7 | PF02687 | FtsX-like permease family | 4 |
| | T-7 | TIGR00956 | 3a01205: pleiotropic drug resistance family | 4 |
| | T-7 | PF00126 | Bacterial regulatory helix-turn-helix protein, lysR family | 4 |
| | T-7 | PF00459 | Inositol monophosphatase family | 4 |
| Furfural | 1 | PF00903 | Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily | 10 |
| | T-2 | TIGR00957 | multi drug resistance-associated protein | 8 |
| | T-2 | PF13620 | Carboxypeptidase regulatory-like domain | 8 |
| | T-4 | PF12697 | Alpha/beta hydrolase family | 6 |
| | T-4 | PF13517 | VCBS repeat domain | 6 |
| | T-4 | TIGR01734 | D-alanine--poly(phosphoribitol) ligase, subunit | 6 |
| | T-7 | PF00753 | Metallo-beta-lactamase superfamily | 5 |
| | T-7 | TIGR01830 | fabG: 3-oxoacyl-[acyl-carrier-protein] | 5 |
| | T-7 | PF13518 | Helix-turn-helix domain | 5 |
| | T-7 | PF00069 | protein kinase domain | 5 |
| | T-7 | PF05593 | RHS Repeat | 5 |
| | T-7 | PF01261 | Xylose isomerase-like TIM barrel | 5 |
| | T-7 | TIGR00915 | RND transporter, hydrophobe/amphiphile efflux-1 (HAE1) | 5 |
| | T-7 | PF03401 | Tripartite tricarboxylate transporter family receptor | 5 |

Table S6: Cloned open reading frames (ORFs) tested for tolerance against the indicated compounds. An asterisk (*) indicates a gene selected on 3-methyl-1-butanol, but tested for furfural tolerance (this amplicon could not be confirmed for alcohol or furfural tolerance).

| Cloned ORF | Compound | Soil | Contig ID | Gene Start_Stop | Annotation | NCBI Accession No. | Pfam / TIGRFam ID | KEGG ID | Confirmation of tolerance? | Evidence for Tolerance |
|------------|------------------|------|-----------|-----------------|--|--------------------|-------------------|---------|----------------------------|------------------------|
| 1 | Furfuryl Alcohol | S07 | Contig_1 | 382_807 | Potassium-transporting ATPase A subunit | KT941849 | PF03814 | -- | No | -- |
| 2 | Furfuryl Alcohol | S07 | Contig_1 | 807_1937 | predicted gene, unknown function | KT941849 | -- | -- | No | -- |
| 3 | Furfuryl Alcohol | S09 | Contig_5 | 290_1294 | Domain of unknown function DUF20 | KT942498 | PF01594 | -- | Yes | figure 3G |
| 4 | Furfuryl Alcohol | S20 | Contig_1 | 348_1844 | putative carboxylase (also similarity to: Dala_Dala Ligase) | KT943337 | TIGR01205 | K01961 | No | -- |
| 5 | Furfuryl Alcohol | S20 | Contig_1 | 348_2129 | putative carboxylase + Protein of unknown function (DUF3592) | KT943337 | PF12158 | -- | No | -- |
| 6 | Furfuryl Alcohol | S20 | Contig_3 | 565_1152 | HD domain; metal-dependent phosphohydrolase | KT943338 | PF13023 | K07023 | No | -- |
| 7 | Ferulic Acid | S08 | Contig_2 | 367_1440 | galK; galactokinase | KT941975 | TIGR00131 | K00849 | Yes | figure 3D |
| 8 | Ferulic Acid | S11 | Contig_1 | 1002_2036 | galE; UDP-glucose 4-epimerase | KT942560 | TIGR01179 | K01784 | Yes | figure 3C |
| 9 | Ferulic Acid | S16 | Contig_6 | 1038_1547 | Phenolic acid decarboxylase (PAD) | KT942848 | PF05870 | K13727 | Yes | figure 3A |
| 10 | Ferulic Acid | S16 | Contig_6 | 1653_2552 | Bacterial regulatory helix-turn-helix protein, lysR family | KT942848 | PF00126 | -- | No | -- |
| 11 | Ferulic Acid | S16 | Contig_6 | 94_951 | Metallo-beta-lactamase superfamily | KT942848 | PF00753 | -- | No | -- |
| 12 | Ferulic Acid | S19 | Contig_1 | 1628_2317 | predicted gene, unknown function | KT943335 | -- | -- | Yes | figure 3B |
| 13 | Ferulic Acid | S19 | Contig_1 | 349_1371 | galE; UDP-glucose 4-epimerase | KT943335 | TIGR01179 | K01784 | No | -- |
| 14 | Ferulic Acid | S19 | Contig_4 | 357_1364 | predicted gene, unknown function | KT943336 | -- | -- | No | -- |
| 15 | Furfural | S04 | Contig_2 | 1205_1852 | Glutathione S-transferase | KT941425 | PF13417 | K00799 | No | -- |
| 16 | Furfural | S04 | Contig_2 | 338_790 | GLO1; lactoylglutathione lyase | KT941425 | PF00903 | K01759 | No | -- |
| 17 | Furfural | S04 | Contig_2 | 613_1149 | unknown function (alternate ORF) | KT941425 | -- | -- | No | -- |
| 18* | Furfural | S06 | Contig_1 | 641_3217 | TOMM system kinase/cyclase fusion | KT941799 | TIGR03903 | -- | No | -- |
| 19 | Furfural | S12 | Contig_1 | 1171_1881 | thyX: thymidylate synthase | KT942684 | TIGR02170 | K03465 | Yes | figure 3E |
| 20 | Furfural | S12 | Contig_2 | 499_1295 | thyA; thymidylate synthase | KT942685 | PF00303 | K00560 | Yes | figure 3F |
| 21 | Furfural | S12 | Contig_2 | 663_1295 | thyA; thymidylate synthase (alternate ORF) | KT942685 | PF00303 | K00560 | No | -- |
| 22 | Furfural | S13 | Contig_1 | 1411_1971 | Domain of unknown function (DUF1851) | KT942791 | PF08906 | -- | No | -- |
| 23 | Furfural | S13 | Contig_1 | 413_1447 | Domain of unknown function DUF20 | KT942791 | PF01594 | -- | No | -- |

Table S7: Full length *galK* open reading frames from ferulic acid selections, all GalK enzymes have mutated residues that are predicted to ablate galactokinase activity. An asterisk denotes the *galK* variant expressed in figure 3D. *S. cerevisiae*; *Saccharomyces cerevisiae*.

| Gene ID | Selected GalK gene name | Mutations in conserved motifs (a.a. coordinates per human GalK) | Motif with Mutation | Consequences in GalK homologs | References |
|---------|-------------------------------|--|-----------------------------|--|----------------------------|
| A | S8_FerAcid_Contig_6_1612_3072 | (38-41) VNLI → LDVM | Motif I, sugar-binding | Unknown | |
| A | S8_FerAcid_Contig_6_1612_3072 | (44) H → I | Motif I, sugar-binding | Human H → Y results 33-fold lower Kcat/Km | Timson, 2003 |
| A | S8_FerAcid_Contig_6_1612_3072 | (139) L → V | Motif II, phosphate-binding | Unknown | |
| A | S8_FerAcid_Contig_6_1612_3072 | (350) ΔC | Motif III, sugar-binding | Unknown | |
| B | S8_FerAcid_Contig_9_71_1234 | (139) L → V | Motif II, phosphate-binding | Unknown | |
| B | S8_FerAcid_Contig_9_71_1234 | (143) A → S | Motif II, phosphate-binding | Part of 2 a.a. deletion that eliminates galactokinase activity in GAL3 of <i>S. cerevisiae</i> ; restoration of 2 a.a. returns galK activity | Platt, 2000; Hartley, 2004 |
| C* | S8_FerAcid_Contig_2_367_1440 | (346) G → R | Motif III, sugar-binding | Human G → S results in 22-fold lower Kcat/Km | Timson, 2003 |
| C* | S8_FerAcid_Contig_2_367_1440 | (347) G → R | Motif III, sugar-binding | Human G → S results in 105-fold lower Kcat/Km | Timson, 2003 |
| C* | S8_FerAcid_Contig_2_367_1440 | (349) G → W | Motif III, sugar-binding | Human G → S results in 8-fold lower Kcat/Km | Timson, 2003 |
| D | S8_FerAcid_Contig_1_192_1391 | (143) A → H | Motif II, phosphate-binding | Part of 2 a.a. deletion that eliminates galactokinase activity in GAL3 of <i>S. cerevisiae</i> ; restoration of 2 a.a. returns galK activity | Platt, 2000; Hartley, 2004 |
| D | S8_FerAcid_Contig_1_192_1391 | (350) C → A | Motif III, sugar-binding | Unknown | |
| E | S7_FerAcid_Contig_19_4_1155 | Strong alternative translation start site putatively deletes the first 39 a.a. of the enzyme | Motif I, sugar-binding | Motif I would be partially deleted due to expression of truncated protein. | Hyatt, 2010 |

File S1: Multiple sequence alignment of 13 galactokinases, including five assembled from ferulic acid selections. NCBI GenInfo numbers follow each database protein.

| | |
|-------------------------------|--|
| Scerevisiae_GAL3_1346086 | -----MNTNVPIFSSPVRLPLRSFEQKHLAVVDAFFQTY-HVKPD-FIARSPGRVNL |
| Scerevisiae_GAL1_585166 | --MTKSHSEEVIVPEFNSSAKELPRPLAEKCPSIIKKFISAY-DAKPD-FVARSPGRVNL |
| S8_FerAcid_Contig_6_1612_3072 | -----MYRITDGSTQNLSDVEAFIGTLSSIEARAFVDSNAEIFVARAPGRLDV |
| S8_FerAcid_Contig_1_192_1391 | -----MLSDLIHQWHSHFGPDTPO-VIVRSPGRVNI |
| Pfuriosus_24211722 | -----MSK-ITVKSPGRVNL |
| S8_FerA_Contig_2_367_1440 | -----MKFQRIF-GTLPK-TTASAPGRVNL |
| Spneumoniae_38604858 | -----MTQHLTAETLRKDFLAVF-GQEVD-QTFFSPGRINL |
| Llactis_1PIE_A | MGSSHHHHHHSSGLVPRGSHMSIVVENSTVLSALTEKFAEVF-GDTKEVEYFFSPGRINL |
| S8_FerAcid_Contig_9_71_1234 | -----MSASPDPLTAPLASPTEPGTADLAARFEQAF-GRLPD-GVWQAPGRVNL |
| Scoelicolor_8894769 | -----MGEAVAGTVGERFRELY-GAEPE-GVWAAPGRVNL |
| S7_FerAcid_Contig_19_4_1155 | -----MCELARHFEERF-GNAPR--IFRAPGRVNL |
| Ecoli_62288105 | -----MSLKEKTQSLFANAF-GYPAT-HTIQAPGRVNL |
| Hsapiens_4503895 | -----MAALRQPQVAELLAEARRAFREEF-GAEPE-LAVSAPGRVNL |

:****:::

| | |
|-------------------------------|--|
| Scerevisiae_GAL3_1346086 | IGEHIDYCDFSVLPLAIDVDMCAVKI-----LDEKNPSITLTNADP-KFAQRKFDLPL |
| Scerevisiae_GAL1_585166 | IGEHIDYCDFSVLPLAIDFDMLCAVKV-----LNEKNPSITLINADP-KFAQRKFDLPL |
| S8_FerAcid_Contig_6_1612_3072 | MGGIADYSGSLALELPIREATLVALQR-----NESRELKIVSLSAEE-NNQARSFEMPP |
| S8_FerAcid_Contig_1_192_1391 | IGEHTDYNEGWVMPGAMNLCLYIVVSRSLAAQNNASLHHWVAADLN--ETIVFELDKEL |
| Pfuriosus_24211722 | IGEHTDYTYGYVMPMAIDLTYTITAEK-----YD--KVQLYSEHFN--EEKFTLDNLT |
| S8_FerA_Contig_2_367_1440 | LGEHTDYNDFVLPTAIPQRTTVQIGF-----SSDDRHHFYSAEYD--ELINYSDDDTI |
| Spneumoniae_38604858 | IGEHTDYNNGHVFPVAISLGTYGAARK-----RDDQVLRFYSANFEDKGIIIEVPLADLK |
| Llactis_1PIE_A | IGEHTDYNGGYVFPASITIGTTGLARL-----REDKKVKLYSENFPKLGVIEFDLDEVE |
| S8_FerAcid_Contig_9_71_1234 | IGEHTDYNEGFLPFAIDKTARVAVGV-----RADSTVRLLSTYGDH-GMVSADLGSLA |
| Scoelicolor_8894769 | IGEHTDYNDFVMPFALPHQAVAASR-----RDDGILRLHSADIDA-DPVELRVADLA |
| S7_FerAcid_Contig_19_4_1155 | IGEHTDYNEGFMPPAAVEFSSLVIAIP-----RKDRKLLIQSNQFP--GDLEFDVDRLP |
| Ecoli_62288105 | IGEHTDYNDFVLPCAIDYQTIVSCAP-----RDDRKVRVMAADYEN-QLDEFSLDAPI |
| Hsapiens_4503895 | IGEHTDYNQGLVLPMALELMTVLVGSP-----RKDGLVSLTTS-EG-ADEPQRLQFPL |

:* ** . : . :

| | |
|-------------------------------|---|
| Scerevisiae_GAL3_1346086 | DGSYMAIDPSVSE-----WSNYFKCG-LHVAHSYLKKIAPERFNNTPLVGAQ |
| Scerevisiae_GAL1_585166 | DGSYVTIDPSVSD-----WSNYFKCG-LHVAHSFLKKLAPERFASAPLAGLQ |
| S8_FerAcid_Contig_6_1612_3072 | ADFESDGKPISYEKARALFRREASQSWAAYVAGAFLVLMRERGVCFQ-----GAR |
| S8_FerAcid_Contig_1_192_1391 | PDTRTI-----WPKYIHGA-IHIFEGALG-----PLN |
| Pfuriosus_24211722 | KEGS-----WIDYVKGV-LWVLIQEGLYKI-G-----GLK |
| S8_FerA_Contig_2_367_1440 | PQK-----FVSYICGC-LRLLEKEGYKI-P-----PIN |
| Spneumoniae_38604858 | FEKEHN-----WTNYPKGV-LHFLQEAGHVIDK-----GFD |
| Llactis_1PIE_A | KKDGE-----WSNYVKGM-IVMLKGAGYEIDK-----GFE |
| S8_FerAcid_Contig_9_71_1234 | AASAKG-----WTKYPLGV-MWALQQQGTTV-P-----GVD |
| Scoelicolor_8894769 | PGSDKS-----WTAYPSGV-LWALREAGHEL-T-----GAD |
| S7_FerAcid_Contig_19_4_1155 | AQRTGT-----WRDYVLGV-AVLVREHSDL-Q-----GAN |
| Ecoli_62288105 | VAHENYQ-----WANYVRGV-VKHLQLRNNSF-G-----GVD |
| Hsapiens_4503895 | PTAQRSLEPGTPT-----WANYVKGV-IQYYPAAPL---P-----GFS |

: *

| | |
|-------------------------------|---|
| Scerevisiae_GAL3_1346086 | IFCQSIDIPTGGGLSS--AFTCAAALATIRANMGKNFDISKDLTRITAVAE-HYVGVNNG |
| Scerevisiae_GAL1_585166 | VFCEGDVPTGSGLSSSAAFICAVALAVVAKANMGPGYHMSKQNLMRITVVAE-HYVGVNNG |
| S8_FerAcid_Contig_6_1612_3072 | ILIDSERVPEGKGVSSSAIEVAVMNAVAAAF---DVSIAPRQMALLCQIVENHVGAPCG |
| S8_FerAcid_Contig_1_192_1391 | ILIGGNLPIGAGVSSSSLVGGLLLAFQQLT---GDTRTKQELANLSSRVEKEVIGLQGG |
| Pfuriosus_24211722 | GKITGDLPLGAGLSSSASFEVGILEVLNQLY---NLNIDPLKKALLAKKAENEFGVPCG |
| S8_FerA_Contig_2_367_1440 | LYITSNVPIGAGLSSSAALEIATLRGLRALL---NLPLDDVRLAQIGQVEIRYAGLNCG |
| Spneumoniae_38604858 | FYVYGNIPNGSGLSSSASLEILTGVVAEHLF---DLKLERLDLVKIGKQTENNFIGVNSG |
| Llactis_1PIE_A | LLIKGEIPTASGLSSSASLELLVGVVLDDLF---NLNVPRLLEVQLGQKTENDYIGVNSG |
| S8_FerAcid_Contig_9_71_1234 | LLLDSNVPLGAGLSSSHAIECAVISALNDLT---GAGLEAEEMVLTQRAENDFVGAPTG |
| Scoelicolor_8894769 | VHLASTVPSGAGLSSSAALEVVVLALAMNDLY---ALGLRGWQLARLCQRAENVYVGAPVG |
| S7_FerAcid_Contig_19_4_1155 | LLVHGNNPIGSGLSSSASIEVASALAFISLD---SKRFLSLPEIAKICRQAENEYVGARVG |
| Ecoli_62288105 | MVISGNVPQGAGLSSSASLEVAVGTVLQQLY---HLPLDGAQIALNGQEAENQFVGNCG |
| Hsapiens_4503895 | AVVVSSVPLGGGLSSSASLEVATYTFLQQLC---PDSGTIAARAQVCQQAEHSFAGMPCG |

. :* . *:** ::

. * * *

Scerevisiae_GAL3_1346086 GMDQATSVYGEEDHALYVEFRPKLKATPFKFPQLKNHEISFVIANTLVKSNKFETAPTN-
Scerevisiae_GAL1_585166 GMDQAASVCGEEDHALYVEFKPQLKATPFKFPQLKNHEISFVIANLTVVSNKFETAPTN-
S8_FerAcid_Contig_6_1612_3072 VMDQMAASCGEANQLALLCOP-AELEAMRR--IPE-EIAVGIDSGI---RHSVGSGD-
S8_FerAcid_Contig_1_192_1391 IMDQYAIMLSQDKVMMMLDRRD-QTYTHIPA---AIPGTKWLLINTKV---KHQLIDS-
Pfuriosus_24211722 ILDQFAVVFHKDNVIFLDTQT-LQYEYIPF---PK-DVSVLVFTGTV---KRELASSE-
S8_FerA_Contig_2_367_1440 IMDQMASSLADTNMLFIDTRS-LECRPIPF---PT-STEILIIDSGE---SHQLAAGSG
Spneumoniae_38604858 IMDQFAIGMGADQRAYLDTNT-LEYDLVPL---DLKDNNVVIMNTNK---RRELADSK-
Llactis_1PIE_A ILDQFAIGFGEVKKAIELDCNT-LKYEMVPV---ELRDYDIVIMNTNK---PRALTESK-
S8_FerAcid_Contig_9_71_1234 IMDQSASLRGSKGHAVFLDCRD-QSQLVPFEAEPA-GLVMLVIDTKV---SHSHSDGG-
Scoelicolor_8894769 IMDQTASACCEAGHALFLDTRD-LSQRQIPF-DLAAEGMRLLVVDTRV---KHSHSEGE-
S7_FerAcid_Contig_19_4_1155 IMDQFVSCMGRAGHALLDCRS-LDFQFAPI---PA-GIELVVCNTMV---KHDLATGA-
Ecoli_62288105 IMDQLISALGKKDHALLIDCRS-LGTKAVSM---PK-GVAVVIINSNF---KRTLVGSE-
Hsapiens_4503895 IMDQFISLMGQKGHALLIDCRS-LETSLVPL---SDPKLAVLITNSNV---RHSLASSE-
: ** : : . :
:

Scerevisiae_GAL3_1346086 YNLRVIEVTVAANALATRYSVALPSHKDNSNSERGNILRD-----FMDAYYARYENQAQP
Scerevisiae_GAL1_585166 YNLRVVEVTAAVLAATYGVVLLSGKEGSSTNKGNLRD-----FMNVYYARYHNISTP
S8_FerAcid_Contig_6_1612_3072 YGSVRTGAFMGTIIAELAGLPIDETDVKGVV---SIRDPNWRGYLANIAPSQFEQHYGP
S8_FerAcid_Contig_1_192_1391 YNQRANQCHQAVIMLQQMFKTVK-----SLRD-----
Pfuriosus_24211722 YAERKRIAEEESLRILGKE-----SSKE-----
S8_FerA_Contig_2_367_1440 YNQRRAECEESARLLGVK-----ALRD-----
Spneumoniae_38604858 YNERRAECEKAVEELQVALDIQ-----TLGE-----
Llactis_1PIE_A YNERFAETREALKRMQTRLDIQ-----SLGE-----
S8_FerAcid_Contig_9_71_1234 YASRRASCELGAEVMGVK-----ALRD-----
Scoelicolor_8894769 YGKRRAGCEKGAAALLGVD-----ALRD-----
S7_FerAcid_Contig_19_4_1155 YNQRRAECEEGVRYFAKADPAIR-----ALRD-----
Ecoli_62288105 YNTRREQCETGARFFQQP-----ALRD-----
Hsapiens_4503895 YPVRRRQCEEVARALGKE-----SLRE-----
* : :
:

| | |
|-------------------------------|--|
| Scerevisiae_GAL3_1346086 | WNGDIGTGIERLLKMLQLV--EESFSRKSGFTVHEASTALNCREEFTRDYLTTFPVRF |
| Scerevisiae_GAL1_585166 | WNGDIESGIERLTKMLVLV--EESLANKKQGFSVDDVAQSLNCSREEFTRDYLTTSPVRF |
| S8_FerAcid_Contig_6_1612_3072 | HLPERITGAEFLSRYKGIT--DPV-----TRVQKE |
| S8_FerAcid_Contig_1_192_1391 | -----VS MEM LEK CEM---PDL----- |
| Pfuriosus_24211722 | -----VTEKDLGKL----PPL----- |
| S8_FerA_Contig_2_367_1440 | -----IADPQAVEVL----PEP----- |
| Spneumoniae_38604858 | -----LDEWA VDQYSYLIK-DEN----- |
| Llactis_1PIE_A | -----LSNEEF DANTDLIG-DET----- |
| S8_FerAcid_Contig_9_71_1234 | -----VRLEDLEEAAGLL--DEV----- |
| Scoelicolor_8894769 | -----VPYADLDAALERLGDEEE----- |
| S7_FerAcid_Contig_19_4_1155 | -----VSSELLERIGGEL--PPT----- |
| Ecoli_62288105 | -----VTIEEFNAVAHEL--DPI----- |
| Hsapiens_4503895 | -----VQLEELEAARDLV--SKE----- |

| | |
|-------------------------------|---|
| Scerevisiae_GAL3_1346086 | QVLKLYQRAKHVYSES L RVL KALKMMTSAT FHTDEDFFTDFGRLMNE SQASCDKLYECSC |
| Scerevisiae_GAL1_585166 | QVLKLYQRAKHVYSES L RVL KAVKLM TTASFTADEDFFKQFGAL MNES QASCDKLYECSC |
| S8_FerAcid_Contig_6_1612_3072 | R S Y S V L A P T A H P I Y E H E R V C R F A E L L G A A Q T E R E --- F I S L G D L M Y G S H A S Y S A - C G L G S |
| S8_FerAcid_Contig_1_192_1391 | --- NKKRCQYVIEENKRVEGMKQALANKD----- AVTAGRLLT ESHQGLQVLYEVSC |
| Pfuriosus_24211722 | --- HRKFFSYIVRENARVLEVR DALKEGD----- IEKVGKILTTAHWDLAENYRVSC |
| S8_FerA_Contig_2_367_1440 | --- LRKRARHVV T ENNRVLEAVKSLP----- AERFGELMNASH ASQRDDYEVSV |
| Spneumoniae_38604858 | --- RLKRARH A VLENQRTLKAQ AALQAGD----- LETFGRLMNASH VSLEHDYEV TG |
| Llactis_1PIE_A | --- LIKRARH A VYENNRTKIAQKAFVAGN----- LTKFGELLNASH ASLKDDYEV TG |
| S8_FerAcid_Contig_9_71_1234 | --- TYRRVRHIVTENDRVLQTVELLASAG----- PGSIGALL DASH ASMRDD FEI SC |
| Scoelicolor_8894769 | --- VRRLVRHV VTEDERVERVVALLES GD----- TRAIGAVLVEGH ASLRDD FRIS C |
| S7_FerAcid_Contig_19_4_1155 | --- ICKRCTHVIQENQRTLDAARALAEGD----- LARMGKLMRESHE SLRDLYEVSC |
| Ecoli_62288105 | --- VAKVRVRHILTENARTVEAASALEQGD----- LKRMGELMAE SHASMRDD FEITV |
| Hsapiens_4503895 | --- GFRRARHVVGEIRRTAQAAAALRRGD----- YRAFGRLMV ESHRS LRDDYEVSC |

: * *. : * : . .

| | |
|-------------------------------|--|
| Scerevisiae_GAL3_1346086 | IETNQICSIALA---NGSFGS-RLTGAGWGGCTIHLVPSGANGNVEQVRKALIEKFYNV |
| Scerevisiae_GAL1_585166 | PEIDKICSIALS---NGSYGS-RLTGAGWGGCTVHLVPGGPNGNIEKVKEALANEFYKV |
| S8_FerAcid_Contig_6_1612_3072 | AGTDLLVELVRSAGYKQGLYGA-KITGGSGG-TVAVLG---RRGADASVEQVADEYERR |
| S8_FerAcid_Contig_1_192_1391 | EELDFLAHLANAT---EGVYGS-RMMGGGFGGCVICLVR---EDAEQSFVEKAISAYDHQ |
| Pfuriosus_24211722 | EELDFFVKKAME---LGAYGA-RLTGAGFGGSAIALVD---KDKAKTIGDAILREYLAK |
| S8_FerA_Contig_2_367_1440 | PAVDALAAILQSI---PGVLRR-KADGRRFWRCL-----RSAGGKWESGCDRIRGDR |
| Spneumoniae_38604858 | LELDTLVHTAWAQ---EGVLGA-RMTGAGFSGCAIALVQ---KDTVEAFKEAVGKHYPEEV |
| Llactis_1PIE_A | LELDTLAETAQKQ---AGVLGA-RMTGAGFGGCIAILVA---HDNSAFRKAVGQVYEEV |
| S8_FerAcid_Contig_9_71_1234 | PELDLAVSTSRA---NGAIGA-RMTGGGFCAAIALTP---VTSEQQVRAAVVRAFSEA |
| Scoelicolor_8894769 | PELDLVVDTALA---SGALGA-RMTGGGF GGSAIVLVE---AAGVDAVTKAVEDAFAAA |
| S7_FerAcid_Contig_19_4_1155 | RELDAMVDAAQGL---PGFIGG-RMTGGGF GGCTVNIVR---EENATDFAAQIAQRYHQA |
| Ecoli_62288105 | PQIDTLVEIVKAV---IGDKGGVRMTGGGF GGCI VALIP---EELVPAVQQAVAEQYEAK |
| Hsapiens_4503895 | PELDQLVEAALAV---PGVYGS-RMTGGGF GGCTVTLLE---ASAAPHAMRHIQEHY--- |

: * . *

| | |
|-------------------------------|------------------------------------|
| Scerevisiae_GAL3_1346086 | RYPDLTDEELKDAIIIVSKPALGTCLYEQ----- |
| Scerevisiae_GAL1_585166 | KYPKITDAELENAAIIVSKPALGSCLYEL----- |
| S8_FerAcid_Contig_6_1612_3072 | MNY-----RPRIFSGSSPGAAARGYKLQRGKLF |
| S8_FerAcid_Contig_1_192_1391 | FGF-----EPDVIYFSLDDGASVISL----- |
| Pfuriosus_24211722 | FSW-----KAKYFVVKPSDGVG----- |
| S8_FerA_Contig_2_367_1440 | SLP-----AHGVYW--PCFGSRNYS----- |
| Spneumoniae_38604858 | VGY-----APSFYIAEVAGGTRVLD----- |
| Llactis_1PIE_A | VGY-----PASFYVAQIGSGSTKLDVE----- |
| S8_FerAcid_Contig_9_71_1234 | GYT-----APDIFTVTPAAGAMRIS----- |
| Scoelicolor_8894769 | GLK-----APRVFEAVPSAGARRLV----- |
| S7_FerAcid_Contig_19_4_1155 | TGI-----TPQIYPCHAANGAQELSLS----- |
| Ecoli_62288105 | TGI-----KETFYVCKPSQGAGQC----- |
| Hsapiens_4503895 | -GG-----TATFYLSSQAADGAKVLCL----- |

Supplemental References

1. **Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R.** 2012. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* **6**:1007-1017.
2. **Platt A, Ross HC, Hankin S, Reece RJ.** 2000. The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase. *Proc Natl Acad Sci U S A* **97**:3154-3159.
3. **Hartley A, Glynn SE, Barynin V, Baker PJ, Sedelnikova SE, Verhees C, de Geus D, van der Oost J, Timson DJ, Reece RJ, Rice DW.** 2004. Substrate specificity and mechanism from the structure of Pyrococcus furiosus galactokinase. *J Mol Biol* **337**:387-398.
4. **Timson DJ, Reece RJ.** 2003. Functional analysis of disease-causing mutations in human galactokinase. *Eur J Biochem* **270**:1767-1774.
5. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119.