

Supplementary table. Representations of pathways in best (most overlapping) clusters.

* Content of a pathway:

- Records: the number of independent records in the Palsson collection.
- Genes: the number of listed genes (operons encoding multisubunit complexes are listed as single genes).

• Reactions: the number of different reactions.

Example. The following set of records contains:

- 6 records: 165, 166, 174, 628, 660, 661
- 6 genes: *speC*, *speF*, *artPMQJI*, *argT*, *hisMQP*, *potABCD* (occurs twice)
- 5 reactions (ORN => PTRC + CO2 occurs twice)

165	<i>speC</i>	ORN => PTRC + CO2	4.1.1.17	Arginine, Putrescine, and Spermidine Biosynthesis
166	<i>speF</i>	ORN => PTRC + CO2	4.1.1.17	Arginine, Putrescine, and Spermidine Biosynthesis
174	-	KMB+ GLN => GLU + MET		Arginine, Putrescine, and Spermidine Biosynthesis
628	<i>artPMQJI argT hisMQP</i>	ARGxt+ ATP => ARG + ADP + PI		Membrane Transport: Amino Acids
660	<i>potABCD</i>	SPMDxt+ ATP => SPMD + ADP + PI		Membrane Transport: Amino Acids
661	<i>potABCD</i>	PTRCxt+ ATP => PTRC + ADP + PI		Membrane Transport: Amino Acids

	pathway	Content* records/ genes/ reactions	Clusters	Pathway & cluster	Pathway – cluster (lost)	Cluster – pathway (false)	comments
1	EMP		Three clusters	1: <i>pgi</i> , 17: <i>tpiA</i> , <i>gapA</i> , <i>pgk</i> , 170: <i>glgC</i> , <i>glgA</i> , <i>glgP</i> , <i>malP</i> .	<i>glk</i> , <i>galM</i> , <i>agp</i> , <i>pfkA</i> , <i>pfkB</i> , <i>fbp</i> , <i>fbaA</i> , <i>gapC</i> , <i>gpmA</i> , <i>gpmB</i> , <i>eno</i> , <i>ppsA</i> , <i>pykA</i> , <i>pykF</i> , <i>lpdA</i> , <i>aceEF</i>		
2	Pentose phosphate (and Embden- Doudoroff)		18	<i>tktA</i> , <i>tktB</i> , <i>talB</i>	<i>zwf</i> , <i>pgl</i> , <i>gnd</i> , <i>rpiA</i> , <i>rpiB</i> , <i>rpe</i> ,		

					<i>edd, eda</i>		
3	The Tricarboxylic Acid Cycle		40	<i>sucAB, lpdA, sucCD, sdhABCD, frdABCD</i>	<i>gltA, acnA, acnB, icdA, fumA, fumB, fumC, mdh</i>	<i>glpD, glpABC, glpK, glpF</i>	Contains also respiration and alternative carbon source <i>glp</i> genes.
4	Pyruvate metabolism		171	<i>pflAB, pflCD, focA</i>	<i>dld, ldhA, adhE, fdhF, hycBEFG, pta, ackA, purT, acs</i>		
5	Anapleurotic reactions		No cluster				
6	Respiration		21	<i>fdnGHI, fdolHG, fdhF, hyaABC, hybAC</i>	<i>ndh, nuo, poxB, dld, glpD, glpABC, cyoABCD, cycBC, cydAB, cydABCD, appBC, cycBC, sdhABCD, trxB, pntAB, hycFGBE, atp, glpD, glpABC</i>		<i>glpD, glpABC</i> are in the TCA cluster
7-27	Alternative carbon sources						Mainly small pathways, difficult resolution of orthology and specificity of enzymes and transporters
8	gslsetose	9/9/10	No cluster				
17	fucose	13/11/9	Two clusters	5: <i>fucO, aldA, aldB, adhE, aldH</i> 174: <i>fucI, fucK, fucA, fucP.</i>	<i>adhC, lldP</i>	174: <i>rhaB, rhaD</i>	<i>rha</i> are rhamnose catabolism genes
19	rhamnose	4/4/4	In cluster 174 (path. 17, fucose)				
20	arabinose	6/5/6	No cluster				
21	xylose	4/5/4	No cluster				
25	galacitol	7/7/4	No cluster				
26	glucerosol	3/3/3	In cluster 40 (path. 3, TCA)				

27	(deoxy)nucleosides	3/3/3	No cluster				
28-36	Amino acid biosynthesis and transport						
28	ASP and ASN	7/5/7	No cluster				No genomic links.
29	GLU and GLN	7/7/6	No cluster				Genomic links weak or absent.
30	ALA	7/5/6	No cluster				Genomic links weak or absent.
31	ARG, putrescine, spermidine	30/25/26	41	<i>argA, argB, argC, argD, argF, argI, argG, argH, yjGH</i>	<i>carAB</i> , all <i>spe</i> genes, all transporters	<i>gltBD</i>	11th most tightly linked pathway; strong link to the histidine pathway. <i>carAB</i> is side input branch, in KEGG it is in the glutamate and pyrimidine pathways. Regulated by ArgR and other factors. All <i>spe</i> genes are lost: evidence against the arginine-polyamine superpathway. <i>gltBD</i> – glutamate synthase, the last reaction in the glutamate pathway; the first reaction in the arginine pathway (<i>argA</i>) uses glutamate as the main substrate.
32	PRO	6/8/6	181	<i>proA, proB, proC</i>	<i>argE</i> , transporters		The pathway is recovered. <i>argE</i> also in the arginine pathway.
33a	branched chain aa	24/16/18	26	<i>ilvBN, ilvIH, ilvC, ilvD, leuA, leuCD, leuB.</i>	<i>ilvG(12)M, ilvA, tdcB, avtA, ilvE, tyrB,</i> transporters		The VAL+ILE cluster. 3rd most tightly linked pathway; strong links to the histidine pathway.
33b	branched chain aa	24/16/18	25	<i>ilvA, tdcB, ilvBN, ilvIH, ilvC, ilvD</i>	<i>ilvG(12)M, avtA, ilvE, tyrB,</i> transporters		The LEU cluster. 3rd most tightly linked pathway; strong links to the histidine pathway.
34	aromatic aa (TRP, TYR, PHE)	25/22/22	42	<i>aroB, aroD, aroE, aroL, aroA, aroC, trpDE</i>	<i>aroF, aroG, aroH, aroK, trpC, trpAB,</i> the TYR-PHE branch (<i>pheA, tyrA, tyrB</i>), transporters	<i>pabAB</i> from the folate pathway	The cluster covers the chorismate pathway and the tryptophan branch (upper part). All DAPH-synthases (<i>aroF, aroG, aroH</i>) are lost; they sometimes form parts of larger MC clusters including other <i>aro</i> genes. The <i>tyr</i> and <i>phe</i> genes are missing: evidence towards a specific subpatway? <i>pabAB</i> is homologous to <i>trpDE</i> , sometimes bifunctional, sometimes within aromatic aa operons [1].
35	HIS	11/9/11	27	<i>hisG, hisIE, hisA, hisFH, hisB, hisC, hisD,</i>	<i>prsA, hisJMPQ</i>		2nd most tightly linked pathway; strong links to the branched chain amino acid and arginine pathways. <i>prsA</i> in KEGG is in the pentose phosphate pathway. <i>hisJMPQ</i> is a transporter.
36	SER and GLY	10/9/9	No cluster				Genomic links weak or absent.
37	CYS	10/9/9	160	<i>cysDN, cysC,</i>	<i>cysE, cysM,</i>		The sulfate assimilation subpathway (exactly). The

				<i>cysH, cysIJ</i>	<i>cysK</i> , transporters		H ₂ S_CYS subpathway is in the methionine cluster.
37 / 39	MET+CYS	N/A	158	<i>metB, metC</i>	<i>metA, metH, metE, metK, metD</i>	<i>cysE, cysM, cysK</i>	The core cysteine-methionine superpathway. <i>metB</i> and <i>metC</i> have weak links to <i>metA</i> and <i>metE</i> , no links to other <i>met</i> genes.
39	MET	8/7/7	29	<i>metH, metE</i>	<i>metA, metB, metC, metK, metD</i>	<i>metF</i>	The downstream methionine subpathway. <i>metF</i> belongs to the one-carbon pathway, but has clear links with the methionine pathway (operons, co-regulation) [2].
38	THR and LYS	23/21/18	28	<i>thrA, metL, lysC, asd, thrA, metL, thrB, thrC, dapA, dapB</i>	<i>dapD, dapC, dapE, dapF, lysA, cadA, ldcC, 5 transporters</i>		Few strong and medium links. <i>dapD</i> has only weak links, and <i>dapC</i> has no links, thus downstream <i>dap</i> and <i>lys</i> genes are lost. Co-regulation of <i>dap</i> and <i>lys</i> genes indeed is rather weak [3].
40-50	Amino acid degradation						
41	ARG	9/9/7	176	<i>aldH, gabD, gabT, goaG,</i>	<i>pat, tynA, prr, sad, gabP</i>	<i>gadA, gadB</i>	Isozymes <i>gadA</i> and <i>gadB</i> form the complete glutamate catabolism pathway.
48	SER	5/5/3	32	<i>sdaA, sdaB</i>	<i>tnaA, dsdA, cycA</i>		<i>sdaC</i> is serine transporter, should be in the pathway
49	THR	5/5/3	33	<i>sdaA, sdaB, tdcC</i>	<i>tdh, kbl, tdcB</i>		<i>tdcC</i> is threonine transporter, should be in the pathway
51-52	Biosynthesis of purines and pyrimidines						
51	Purine biosynthesis	17/15/17	Upper: 43, Middle: 16.	Upper: <i>purF, purD, purN, purL, purM, purH</i> . Middle: <i>purK, purE, purC</i> .		<i>purT</i> . Adenine: <i>purA, purB</i> . Guanine: <i>guaA, guaB, guaC</i> .	Upper and parts of the common purine pathway. Cluster 52 (DUC-700) combines both parts. Almost complete pathway is recovered in many MC clusters. The adenine subpathway (<i>purA, purB</i>) has only weak links with the common subpathway (upper and middle part), although <i>purB</i> is also a step in the middle part. The guanine subpathway (<i>guaA, guaB, guaC</i>) has no links with other genes, nor between the <i>gua</i> genes themselves.
52	Pyrimidine biosynthesis	6/6/6	178	<i>pyrD, pyre, pyrF</i>	<i>pyrBI, pyrC, pyrG, (ntk)</i>		Middle (dihydroorotate_UDP) part of the pyrimidine biosynthesis pathway. Overall, 9th most tightly linked pathway. Missed: <i>carAB</i> (nitrogen metabolism_carbamoyl-L-aspartate, common step with the arginine pathway; in <i>E.coli</i> co-regulated with pyrimidine and arginine genes), <i>pyrBI</i> (strong and medium genomic links, no immediate metabolic links), <i>pyrC</i> (only one weak

							link); <i>pyrG</i> and <i>ndk</i> (interconversion of pyrimidine di- and triphosphates).
54	One carbon metabolism	5/6/5	No cluster				
55	Membrane lipid biosynthesis	11/6/11	Diverse MC clusters only				
56	Fatty acid metabolism	7/3/7	20	<i>atoB</i> , <i>fadABDE</i>	<i>fadL</i>	<i>adhE</i>	<i>fadL</i> is a transporter. <i>adhE</i> is from the pyruvate metabolism
57-72	Biosynthesis of cofactors and electron carriers						
58	Ubiquinone	9/9/9	No cluster				Anonymous reaction in the middle, most links weak or absent.
59 / 60	Menaquinone / Enterochelin	Menaq.: 8/7/8. Enteroch.: 6/6/6	180	Menaquinone: <i>menE</i> , <i>menF</i> , <i>menD</i> . Enterochelin: <i>entC</i> , <i>entB</i>	<i>menE</i> , <i>menB</i> , <i>menA</i> , <i>menG</i> , <i>entA</i> , <i>entE</i> , <i>entF</i> , <i>entD</i>		Early reactions of the menaquinone and enterochelin biosynthesis. <i>menE</i> has only weak links and de-clusters highly linked <i>menB</i> . <i>entC</i> is an isoenzyme of <i>menF</i> from enterochelin pathway. Late reaction of the enterochelin pathway have almost no links. Menaquinone: 7th most tightly linked pathway.
61	Riboflavin	9/7/9	No cluster				Several strong links (not necessarily linked metabolically) and no other links.
62	Folate	11/10/11	See cluster 42 (aromatic AA)				10th most tightly linked pathway.
63	Coenzyme A	12/8/11	No cluster				Many anonymous reactions (no genes).
64	NAD	13/8/13	No cluster				
67	Tetrapyrrole	16/14/14	44	<i>hemA</i> , <i>hemL</i> , <i>hemB</i> , <i>hemC</i> , <i>hemD</i> , <i>hemX</i> , <i>cysG</i>	<i>hemM</i> , <i>gltX</i> , <i>hemE</i> , <i>hemF</i> , <i>hemG</i> , <i>hemH</i> , <i>cyoE</i>		Upper part of tetrapyrrole biosynthesis (thr glutamyl-tRNA_uroporphyrinogen-III subpathway). Overall, 12th most tightly linked pathway. <i>hemM</i> and <i>gltX</i> are glutamyl-tRNA synthetases. No heme subpathway (<i>hemE</i> , <i>hemF</i> , <i>hemG</i> , <i>hemH</i> , <i>cyoE</i>).
68	Biotin	4/4/4	24	<i>bioF</i> , <i>bioA</i> , <i>bioD</i> , <i>bioB</i>			Biotin biosynthesis (exactly). 4th most tightly linked pathway.
70	Thiamine	13/12/10	No standard cluster, only DUC-400-Neighborhood	<i>thiG</i> , <i>thiE</i> , <i>thiF</i> , <i>thiH</i> , <i>thiM</i>	<i>thiC</i> , <i>thin</i> , <i>thiD</i> , <i>thiB</i> , <i>thiK</i> , <i>thiL</i>		<i>thiG</i> , <i>thiE</i> , <i>thiF</i> , <i>thiH</i> catalyze the same reaction.
71	Pyridoxine	18/9/	13	<i>pdxH</i>	Rest of the pathway		Trivial cluster of four reactions catalyzed by <i>pdxH</i>
72-75	Cell membrane and cell wall						
73	Lipid A biosynthesis	9/8/9	182	<i>lpxA</i> , <i>lpxC</i> , <i>lpxD</i> , <i>lpxB</i>	<i>ushA</i> , <i>kdtA</i> , <i>htrB</i> , <i>msbB</i>	<i>pyrH</i>	Upper part of the Lipid-A precursor biosynthesis pathway. <i>pyrH</i> is from the salvage pathways. <i>ushA</i>

							(3.6.1.45) has no genomic links because it may be known under a different name (<i>lpxH</i>). <i>kdtA</i> has strong genomic links, but is not clustered, because the intermediate step (2.7.1.130) has no name (<i>lpxK</i>). Pathway recovered in MC clusters. 5th most tightly linked pathway.
74	LPS sugar biosynthesis	9/8/9	159	<i>lpcA, rfaED, rfaLJIGFK</i>	<i>kdsA, kdsB, galU, galF, dgkA</i>	<i>htrB, msbB</i>	Small chimeric cluster of genes from the Lipid A and LPS sugar pathways. Many anonymous (no gene assigned) reactions in the LPS pathway. LPS and Lipid A form the 4th most highly cross-linked pathway pair.
75	Murein biosynthesis	10/10/10	45	<i>murA, murB, murC, murD, murE, murF, ddlAB, mraY, murG</i>	<i>murI</i>		Almost complete murein pathway. Most tightly linked pathway. <i>murI</i> : belongs to a side branch (glutamate racemase)

References.

1. Panina, E.M., Vitreschak, A.G., Mironov, A.A., & Gelfand, M.S. (2003) *FEMS Microbiol. Lett.* **222**, 211-220.
2. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., & Gelfand, M.S. (2003) *Nucleic Acids Res.* **31**: 6748-6757.
3. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., & Gelfand, M.S. (2004) *Nucleic Acids Res.* **32**: 3340-3353.