1

2 Transcriptomic fingerprinting of *Pseudomonas putida* under

3 alternative physiological regimes

4

5 by

6

7 Juhyun Kim, Juan Carlos Oliveros, Pablo I. Nikel, Víctor de Lorenzo* and Rafael Silva-Rocha[§]

8

9

10

11 *Systems Biology Program, Centro Nacional de Biotecnología CSIC, Cantoblanco-Madrid, 28049 Spain*

12

13

14

15 **Running Title:** RNA-seq analysis of *P. putida's* transcriptome

16 **Keywords:** *Pseudomonas putida*, carbon metabolism, RNA-seq, regulatory network

17

18

19

20

21 _____

22

23 *Correspondence to: Víctor de Lorenzo

24 Centro Nacional de Biotecnología-CSIC

25 Campus de Cantoblanco, Madrid 28049, Spain

26 Tel.: 34- 91 585 45 36; Fax: 34- 91 585 45 06

27 E-mail: vdlorenzo@cnb.csic.es

28

29

30 [§]Current address: FMRP - University of São Paulo, Ribeirao Preto, SP, Brazil.

31 SUMMARY

1

2   *Pseudomonas putida* KT2440 is a metabolically versatile soil bacterium useful both as a model

3   biodegradative organism and as a host of catalytic activities of biotechnological interest. In this report,

4   we present the high-resolution transcriptome of *P. putida* cultured on different carbon sources as

5   revealed by deep sequencing of the corresponding RNA pools. Examination of the data from growth on

6   substrates that are processed through distinct pathways (glucose, fructose, succinate and glycerol)

7   revealed that $\geq$ 20% of the *P. putida* genome is differentially expressed depending on the ensuing

8   physiological regime. Changes affected not only metabolic genes but also a suite of global regulators

9   e.g. the *rpoS* sigma subunit of RNAP, various cold-shock proteins and the three HU histone-like proteins.

10 Specifically, the genes encoding HU subunit variants *hupA*, *hupB* and *hupN* drastically altered their

11 expression levels (and thus their ability to form heterodimeric combinations) under the diverse growth

12 conditions. Furthermore, we found that two small RNAs, *crcZ* and *crcY*, known to inhibit the Crc protein

13 that mediates catabolite repression in *P. putida,* were both down-regulated by glucose. The raw

14 transcriptomic data generated in this work is made available to the community through the Gene

15 Expression Omnibus database.

16 _____

17

18 INTRODUCTION

19

20 The KT2440 strain of the soil bacterium *Pseudomonas putida* has attracted much attention for over 30

21 years as an experimental system that embodies many of the questions and applications of

22 environmental microorganisms (Bagdasarian *et al.*, 1981). This bacterium harbours a 6.2 Mb genome

23 encoding a large number of genes related to utilization of unusual compounds as carbon sources

24 (Jimenez *et al.*, 2002; Nelson *et al.*, 2002), as well as an extensive machinery for adaptation to diverse

25 environmental conditions (dos Santos *et al.*, 2004). In order to regulate the >5000 genes encoded in its

26 genome, *P. putida* is endowed with approximately 600 transcriptional factors (TFs) and about 24

27 alternative sigma subunits of RNAP (dos Santos *et al.*, 2004; Martinez-Bueno *et al.*, 2002) which can

28 give rise to a very complex gene expression landscape. But little is known on the functionally of the vast

29 majority of these regulators because only few global factors have been investigated using

30 transcriptomics or other *wet* approaches (Amador *et al.*, 2010; Herrera *et al.*, 2010; Silva-Rocha *et al.*,

31 2012; Moreno *et al.*, 2009; Morales *et al.*, 2006). More recently, a list of candidate non-coding RNAs

1   (ncRNAs) of *P. putida* has been generated (Frank *et al.*, 2011), but again not much information is

2   available regarding the signals sensed by these and their cognate genomic targets. Fortunately, the

3   growing ease of next generation sequencing (NGS) provides massive high quality information on

4   genome-wide expression in all types or organisms (Wang *et al.*, 2009). In this way, identification and

5   cataloguing of RNAs under given environmental and metabolic conditions provides a wealth of data on

6   the regulatory and biochemical organization of the cell. In particular, the so-called RNA-seq approach

7   allows inspecting several features of transcriptional units, such as the start and termination sites, the

8   presence of 5' and 3' untranslated sequences adjacent to coding regions, the layout of operons, the

9   annotation of new genes and ncRNAs, and the refinement of the predicted start/end sites of ORFs in

10   respect to the position of the nearby transcriptional start sites (TSS; Wang *et al.*, 2009; Bar *et al.*, 2008;

11   Sittka *et al.*, 2008). Yet, the abundance of information that is encrypted in such deep RNA sequencing

12   experiments can only be decoded through a community-based effort. Typically, different teams will

13   focus on specific aspects of the same transcriptome and each will make a specialized analysis of

14   features of the bacterium at stake that may not be important to others of the same community (Guell *et*

15   *al.*, 2009).

16

17   In this report, we have inspected the complete RNA-seq transcriptome of *P. putida* KT2440 in steady

18   grown cultures of this bacterium on four different carbon sources, i.e. glucose (Glu), fructose (Fru),

19   succinate (Suc) and glycerol (Gly). As shown below, a gross analysis of the data revealed a major

20   reorganization of the genetic and biochemical network of this microorganism that involved not only

21   metabolic genes but also archetypal global regulatory factors and ncRNAs. Since a large number of

22   testable hypotheses beyond our own interests can be raised on the basis of these transcriptomes, we

23   offer the community open access to the raw data and encourage further analyses of the RNA landscape

24   of *P. putida* KT2440.

25

26   RESULTS AND DISCUSSION

27

28   *The transcriptomes of* P. putida *growing in four distinct carbon sources*

29

30   In order to gain an insight on the impact of various metabolic regimes into the genome-wide expression

31   landscape of *P. putida*, we performed four RNA-seq experiments with cells growing on carbon sources

1   known to set the bacteria in alternative biochemical modes. To this end, we chose two substrates that

2   enter the *upper* domain of carbon metabolism (Glu and Fru) and two that make it directly into the *central*

3   part (Suc and Gly). Each of these substrates engages a different set of reactions within the metabolic

4   network of *P. putida* and thus promote distinct physiological regimes, either glycolytic or gluconeogenic

5   (Nogales *et al.,* 2008; Puchalka *et al.,* 2008) which are reflected in non-identical growth rates

6   (Supplementary Table S1). RNA extracted from cells grown in batch until mid-exponential phase was

7   used for Illumina™ sequencing assay and processed as described in the Experimental procedures

8   section. A summary of the outcome of these experiments is presented in Supplementary Table S2.

9

10   The resulting sequences were aligned to the reference genome of *P. putida* KT2440 and the information

11   used for analysing differential expression of the genes under the growth conditions indicated. A simple

12   pairwise comparison of mRNA levels between each of the samples using as a reference the samples

13   grown in succinate revealed an unexpectedly large alterations in the transcriptomes of cells grown on

14   the different media (Supplementary Fig. S1). Genes were considered to be differentially expressed if

15   they presented a fold change equal or greater than 2 between compared samples and with a FDR (false

16   discovery rate) value ≤ 0.001. The most remarkable case was the divergence (~1000 genes) between

17   Fru and Glu. Although these sugars are structurally similar, have a comparable energetic value, and are

18   both metabolized through the Entner-Doudoroff route, it seems that (unlike Glu), Fru-grown cells load

19   the nitrogen-related phosphoenolpyruvate-dependent phosphotransferase system (PTS^{Ntr}) of *P. putida*

20   with an excess of high-energy phosphate (Chavarria *et al.*, 2012a; Chavarria *et al.*, 2013). This may

21   ultimately translate into changes of expression of PTS^{Ntr}-controlled genes, although the specific

22   mechanisms involved are still uncertain (Pfluger-Grau and Gorke, 2010). Many differences were

23   observed also between cells growing on the other carbon sources. While the detailed analysis of

24   transcripts in the context of each metabolic regime is beyond the scope of this article, we focused our

25   attention in a limited number of global regulators encoded in separated parts of the genome (Fig. 1A)

26   and the transcription of which conspicuously changed from one condition to the other.

27

28   *Effect of carbon sources on* rpoS *levels*

29

30   The first noticeable regulatory gene found to be differentially expressed was *rpoS*, encoding the sigma

31   subunit of the RNA Polymerase (RNAP), that is generally associated to stress and gene expression

1   under stationary phase (Ramos-Gonzalez and Molin, 1998). While the regulation of *rpoS* differs

2   between *E. coli* and *P. putida*, it involves in either case transcriptional and translational controls (Venturi,

3   2003). As shown in Fig. 1B, the *rpoS* gene was clearly upregulated in cells growing in Gly and Fru, with

4   a maximal 3-fold change between Fru and Suc conditions. Such variation of *rpoS* levels with different

5   carbon sources inversely correlated with the growth rates reported for *P. putida* KT2440 cells growing

6   on the same compounds (Holtel *et al.*, 1994). Since Gly-grown or Fru-grown cells from which RNA was

7   extracted were growing exponentially, this raises a connection between *rpoS* expression and cell

8   division rate that is not related to growth phase. The transcriptomic fingerprinting of e.g. Fru-grown

9   bacteria thus merges the effects of the sugar as such with those created by slow growth (e.g. relief of

10   some catabolite-repressed genes; Scott *et al.*, 2010) and those linked to *rpoS* variations.

11

12   *Cold-shock regulatory functions*

13

14   Another type of regulatory genes clearly influenced by carbon sources included those encoding the cold

15   shock proteins CspA-1, CspA-2 and CspD (PP1522, PP2463 and PP4010, respectively). The CspA

16   protein family encompasses a number of RNA chaperones related to adaptation to cold shock. These

17   proteins work both as destabilizers of secondary RNA structures that block translation at low

18   temperatures as well as transcriptional anti-terminators (Jiang *et al.*, 1997; Bae *et al.*, 2000). The two

19   *cspA* orthologs present in the genome of *P. putida* displayed different expression profiles under the

20   conditions assayed. As shown in Fig. 1C, *cspA-1* showed a higher expression in Fru and Suc. In

21   contrast *cspA-2* was highly expressed in cells growing in Suc, presented an intermediate level in Glu

22   and a low mRNA signal in Fru and Gly. On the other hand, CspD is a toxin that inhibits DNA replication

23   (Yamanaka *et al.*, 2001). The *cspD* homologue of *P. putida* genome was highly expressed in Fru and

24   Gly while it displayed a low mRNA level in Glu and Suc (Fig. 1C). In *E. coli*, expression of this protein is

25   inversely correlated to the growth rate, being up regulated during stationary phase, although it is not

26   dependent on RpoS (Yamanaka and Inouye, 1997). As shown in Fig. 1C, the expression profile of *cspD*

27   gene matches the differences in growth rate as discussed for the *rpoS* gene in Fig. 1B. In any case,

28   available data on regulation of this type of *csp* genes rely on growing cells at different temperatures but

29   the experiments summarized in Fig. 2C are all made at 30ºC. How can then C sources influence the

30   process? The known mechanism that controls expression of Csp proteins involves either by post-

31   translational control of mRNA translation or the stimulation of anti-termination inside the target genes

1   (Phadtare and Severinov, 2010). In order to investigate the conservation of these mechanisms in the

2   control of the Csp proteins production by the carbon sources tested we examined the RNA-seq data for

3   the three genes in order to identify their transcription start and termination sites (TSS and TTS,

4   respectively). Fig. 2 shows the high-resolution structure of the transcripts for the *csp* genes. *cspA-1* (Fig.

5   2B) and *cspD* (Fig. 2C) have two TSSs that generate 5'-UTRs of ~100 nt. In contrast, *cspA-2* (Fig. 2B),

6   has a single TSS and the 5'-UTR is shorter (~50 nt). The presence of long UTRs in the 5' regions opens

7   the possibility that some post-transcriptional regulation can ultimately control expression of these genes.

8   Still, the most remarkable feature of these transcripts was the presence of premature TTSs within each

9   of the three coding sequences. In the case of *cspA-1*, two TTSs seem to occur inside the gene while a

10   third termination (TTS-3) appears several nucleotides after the STOP codon of the gene (Fig. 2A). The

11   same situation was found for *cspA-2* (Fig. 2B), while two TTSs were found for *cspD*, one inside the

12   coding region (TTS-1) and another following the ORF (TTS-2, Fig. 2C). Such a premature termination of

13   mRNA observed in the *csp* genes was noticed in very few other cases of the transcriptome of *P. putida*

14   and it is thus likely to be a genuine phenomenon. In contrast, the large majority of the other genes

15   examined display signals and expression patterns that are maintained (at different levels) along the

16   entire coding region (see e.g. expression of the housekeeping gene *rpoD* in Supplementary Fig. S2).

17   The presence of premature TTSs inside the *csp* genes reported above thus suggest that both

18   transcriptional regulation and anti-termination play a role in the control of expression of Csp proteins in

19   *P. putida* both by cold shock (similarly to the genes of *E. coli*; Phadtare and Severinov, 2010) and by the

20   C sources tested.

21

22   *Differential expression of the three HU-coding genes of P. putida*

23

24   Various genes encoding subunits of the histone-like HU protein were also found to vary in a fashion

25   dependent on the growth regime. In *E. coli*, HU is formed by two subunits (α and β) encoded by the

26   *hupA* and *hupB* genes (Ramstein *et al.*, 2003; Broyles and Pettijohn, 1986). This protein can exist in

27   three dimeric forms, two homodimers HUα_2 and HUβ_2 and a heterodimer HU$\alpha\beta$ (Ramstein *et al.*, 2003).

28   These nucleoid-associated proteins play a major role in the regulation of genes related to stress

29   adaptation (Oberto *et al.*, 2009). Three HU-coding genes have been reported to exist in *P. putida*

30   KT2440 namely *hupB*, *hupN* (Bartels *et al.*, 2001) and *hupA* (formerly designated *hupP*; Cases and de

31   Lorenzo, 2002). Six potential dimers could in principle be formed by the combination of such three HU

1   subunits, and each dimer could have different DNA-binding specificities or properties, as is the case in *E.*

2   *coli* (Claret and Rouviere-Yaniv, 1997; Pinson *et al.*, 1999). Inspection of expression profiles of *hupA,*

3   *hupB* and *hupN* transcripts revealed a considerable change of their respective levels in each of the

4   conditions tested (Fig. 3). *hupA* had the lowest expression in Glu, while it was induced by ~2-fold in Suc

5   and Gly. However, the same gene multiplied its transcription by 10-fold when growing in Fru (Fig. 3A). In

6   the case of *hupB*, the maximal fold-change observed was ~2.5-times between Suc and Gly conditions

7   (expression in this case was higher in Glu/Suc than in Gly/Fru conditions). Finally, a comparatively lower

8   expression level was observed for *hupN* under all conditions although the pattern was also Glu/Suc $\geq$

9   Gly/Fru (**Fig. 3a**). Considering that the three *hup* genes could account for six dimeric forms, their

10   differential expression could bring about a switch in the relative abundance of each of these molecular

11   species. As sketched in **Fig. 3b**, the higher expression of *hupB* and *hupN* with Glu/Suc could favour

12   formation of the BB, NN and BN dimers of HU. In contrast, the higher expression of *hupA* and lower of

13   *hupN* with Fru could displace the equilibrium towards formation of AA, BB and AB dimers. Finally, in

14   cells growing in Gly all three genes are expressed at low level and thus the six dimeric forms of the

15   protein could be generated at similar ratios, but with a lower absolute protein level. Should these

16   variations in HU populations occur, they are likely to translate into differential regulation of subsets of

17   genes controlled by specific protein forms (Oberto *et al.*, 2009).

18

19   *Carbon source-dependent expression of* crcZ *and* crcY *ncRNAs*

20

21   Finally, comparison of *P. putida*'s transcriptomes growing with different substrates revealed also new

22   insights on the complex mechanism behind catabolite repression in this bacterium (Rojo, 2010). The

23   choice of preferred carbon sources over less favoured counterparts involves a number of molecular

24   actors among which the RNA-binding translation inhibitor Crc protein plays a key role (Moreno *et al.*,

25   2009). Since the activity of this factor is checked by small non-coding RNAs that compete for target 5'-

26   UTR regions of thereby regulated products (Moreno *et al.*, 2012), we examined the RNA expression

27   landscape of *P. putida* for small transcripts, the levels of which varied with the growth substrate. Two

28   sRNAs which were expressed at high levels in a fashion dependent on the carbon source used turned

29   out to be the previously identified *crcZ* and *crcY* transcripts of *P. putida* (Moreno *et al.*, 2012). These

30   sRNAs are the main inhibitors of the Crc and the actual carriers of the metabolic signal that brings about

31   CR or de-repression. According to the current model, expression of *crcZ* and *crcY* is completely

1	inhibited in cells growing in rich media (e.g. LB or casamino acids), while Crc is expressed at high levels

2	(Moreno *et al.*, 2012; Ruiz-Manzano *et al.*, 2005). In contrast, in a non-repressive scenario (e.g. minimal

3	medium with succinate) expression of *crcZ* and *crcY* is induced while Crc levels decrease, the effect

4	being the relief of CR (Rojo, 2010). As shown in Fig. 4, expression of *crcZ* and *crcY* were dramatically

5	dependent on the carbon source, following the hierarchy Glu<Suc<Gly<Fru. It is remarkable that *crcY*

6	displayed a fold-change (~9.7) higher than *crcZ* (~3.8), such that both sRNAs are expressed at similar

7	level in the culture with Fru. Since *crc* gene expression remain unchanged in the four carbon source

8	conditions (not shown), the thereby discovered carbon source regulation of *crcZ* and *crcY* in *P. putida*

9	adds a new element to the CR network in this organism that deserves further investigation.

10

11	*Conclusion*

12

13	As is the case with other bacteria, the transcriptomes of *P. putida* under various conditions represent a

14	fingerprinting of the complete physiological and regulatory status of the cells. Yet, not all aspects of any

15	given organism are equally important for specific purposes. *P. putida* is remarkable for its metabolic

16	versatility, its stress endurance and its potential as a genomic and biochemical chassis for engineering

17	whole cell catalysts (Poblete-Castro *et al.*, 2012). Despite similarity of a large number of metabolic

18	pathways and general physiological features with the *E. coli*, the lifestyle of *P. putida* seem to impose

19	different strategies for dealing with carbon sources e.g. by favouring the activity of routes conductive to

20	NADPH formation (Chavarria *et al.*, 2012b). These metabolic changes are necessary reflected in the

21	gene expression landscape that is revealed through the deep sequencing of the RNA pool. This report

22	provides a snapshot of such landscapes in cells growing in identical conditions except for the nature of

23	the carbon sources in the medium. Most differences between them should therefore originate in the

24	specific physiological regimes adopted for their metabolization. Interestingly, the changes involve many

25	functions beyond up or down regulation of metabolic genes. While the biological meaning of such

26	variations might be difficult to decipher (Price *et al.*, 2013), it is likely that they encode the evolutionary

27	history of the corresponding metabolic pathways for specific compounds. We encourage the *P. putida*

28	community to make use of the resource hereby released for exploring these questions with either the

29	generic software mentioned in the Experimental procedures or with other tools tailored for specific data

30	interrogation. The complete raw dataset is available at the Gene Expression Omnibus (GEO) site of the

31	NCBI (http://www.ncbi.nlm.nih.gov/geo) with accession code GSE46491. One example of utilization is

1   e.g. the recently reported metabolic and regulatory analysis of *P. putida* grown on glycerol (Nikel *et al.,*

2   2013).

3

4   EXPERIMENTAL PROCEDURES

5

6   *Strains, growth conditions and RNA preparation*

7

8   Wild-type *P. putida* strain KT2440 (Bagdasarian *et al.*, 1981) was used in all experiments. *P. putida* was

9   cultivated in M9 minimal medium (Sambrook *et al.*, 1989) supplemented with 2 mM $MgSO_4$ and one of

10   the four following carbon sources: glucose (10 mM), succinate (15 mM), glycerol (20 mM) and fructose

11   (10 mM). In all cases, cells were grown at 30ºC with vigorous orbital shaking. In order to obtain RNA for

12   deep sequencing, single colonies of *P. putida* were inoculated in M9 with succinate as the sole carbon

13   source. After overnight growth, cells were washed twice and diluted 100-fold in 20 mL of fresh medium

14   containing succinate, glucose, fructose or glycerol as growth substrate in 100 mL Erlenmeyer flasks.

15   Cultures were then allowed to grow until mid-exponential phase (OD_{600} ~0.4). Cultures samples (10 ml)

16   were immediately transferred into 15-ml Falcon tubes containing 1 mL of ice-cold phenol/ethanol

17   solution [5% (v/v) water-saturated phenol in ethanol] to protect RNA from degradation. Cells were then

18   harvested by centrifugation (3,800 rpm for 15 min at 4°C). After aspiration of the supernatant, the

19   pellets were frozen in liquid nitrogen and stored at -80°C until further processing. Total RNA was

20   extracted by using the RNeasy kit (Qiagen Inc.) with some modifications. The collected pellets were

21   resuspended in 0.3 mL Tris-HCl (pH 7.5) containing 2 mg/mL lysozyme and incubated for 10 min at 37

22   °C. Then, 0.1 mL of lysate was used according to the manufacturer's instructions. RNase-free DNase

23   (Qiagen Inc.) treatment was performed during the isolation procedure to eliminate the residual DNA.

24   Finally, the quality of RNA samples was evaluated using an Agilent 2100 Bioanalyzer (Agilent

25   Technologies Inc.).

26

27   *Deep sequencing of retrotranscribed RNA*

28

29   The RNA library construction and sequencing was made by BGI (Shenzhen, China; http://www.cloud-

30   sequencing.com), using the Illumina mRNA sequencing sample preparation kit (cat. # RS-930-1001)

31   and the Illumina HiSeq™ 2000 system (Illumina Inc., San Diego, CA, USA). Total RNA from the

1   samples was treated with DNAse I in order to avoid any DNA contamination. Depletion of rRNA was

2   done using the Ribo-Zero™ rRNA removal kit for Gram-negative bacteria (Epicenter Biotechnologies

3   Corp.) according to the manufacturer's protocol. The isolated RNA was then treated in a reaction

4   containing 5× fragmentation buffer [200 mM Tris·CH_3COO (pH = 8.1), 500 mM CH_3COOK, and 150 mM

5    $(CH_3COO)_2Mg$] at 94°C for 5 min, and purified by ethanol/glycogen precipitation. For cDNA synthesis,

6   randomly-primed reverse transcription was performed to obtain the first strand chain, and then GEX

7   second strand buffer solution, dNTPs, RNase H and DNA polymerase I were added for synthesis of the

8   second strand cDNA, and the synthesized product was finally purified and kept at −20°C until further

9   use. The cDNA was end-repaired and phosphorylated (using a mixture of T4 DNA polymerase and

10   Klenow DNA polymerase), followed by adenylation of 3'-ends, and 5'-adapter ligation. The ligated

11   reaction was purified on a 2% (w/v) agarose gel and template cDNA in the 200 bp (± 25 bp) range was

12   selected and eluted. The purified cDNA template was enriched by PCR (using Phusion high-fidelity DNA

13   polymerase), and the purified libraries were sequenced using the Illumina HiSeq™ 2000 system to yield

14   91-paired end reads. In order to provide a statistical basis for the identification of differentially expressed

15   genes, the RNA-seq data was processed as indicated below.

16

17   *RNA-seq data analysis*

18

19   For RNA-seq analysis, the resulting sequence reads were filtered according to quality criteria. First,

20   reads containing the adaptors were removed, as well as those with more than 10% of ambiguities (N) or

21   with more than 50% of the nucleotides with quality below 5. The remaining high quality reads were

22   mapped to the database of *Pseudomonas putida* KT2440 genes (GenBank accession AE015451) using

23   SOAP 2.21 software (Li *et al.*, 2009) allowing up to 5 mismatches. In order to identify differentially

24   expressed genes, the reads per kilobase per million mapped reads (RPKM) was calculated for each

25   gene of the genome (see Supplementary Experimental Procedures). For the detection of differentially

26   expressed genes between two samples (for example, glucose *vs.* fructose condition), a Poisson

27   distribution for the small proportion of reads that uniquely maps to a given gene was assumed, and fold

28   changes and *p*-values were calculated (Audic and Claverie, 1997). Finally, to minimize the multiple

29   testing problem, *p*-values were corrected and expressed as FDR (Benjamini *et al.*, 2001). Using this

30   methodology, genes with FDR ≤ 0.001 and absolute Fold Change larger than 2 were considered as

31   differentially expressed.

1

2 *Data visualization*

3

4 Read alignments were further processed using Integrative Genomics Viewer–IGV software (Robinson *et*

5 *al.*, 2011). Analysis of the reads at each condition was performed using IGV batch tools. Plots shown in

6 Fig. 2 were generated using R package (http://www.r-project.org/).

7

8 ACKNOWLEDGEMENTS

9

14

15 REFERENCES

16

17 Amador, C. I., Canosa, I., Govantes, F. and Santero, E. (2010) Lack of CbrB in *Pseudomonas putida*

18   affects not only amino acids metabolism but also different stress responses and biofilm

19   development. *Env Microbiol* **12**: 1748-1761.

20 Audic, S. and Claverie, J. M. (1997) The significance of digital gene expression profiles. *Genome Res* **7**:

21   986-995.

22 Bae, W., Xia, B., Inouye, M. and Severinov, K. (2000) *Escherichia coli* CspA-family RNA chaperones

23   are transcription antiterminators. *Proc Natl Acad Sci USA* **97**: 7784-7789.

24 Bagdasarian, M., Lurz, R., Ruckert, B., Franklin, F. C., Bagdasarian, M. M., Frey, J. and Timmis, K. N.

25   (1981) Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number,

26   RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas. Gene* **16**:

27   237-247.

28 Bar, M., Wyman, S. K., Fritz, B. R., Qi, J., Garg, K. S., Parkin, R. K., Kroh, E. M., Bendoraite, A.,

29   Mitchell, P. S., Nelson, A. M., Ruzzo, W. L., Ware, C., Radich, J. P., Gentleman, R., Ruohola-Baker,

30   H. and Tewari, M. (2008) MicroRNA discovery and profiling in human embryonic stem cells by

31   deep sequencing of small RNA libraries. *Stem Cells* **26**: 2496-2505.

1   Bartels, F., Fernandez, S., Holtel, A., Timmis, K. N. and de Lorenzo, V. (2001) The essential HupB and
2   HupN proteins of *Pseudomonas putida* provide redundant and nonspecific DNA-bending functions.
3   *J Biol Chem* **276**: 16641-16648.

4   Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. and Golani, I. (2001) Controlling the false discovery rate in
5   behavior genetics research. *Behav Brain Res* **125**: 279-284.

6   Broyles, S. S. and Pettijohn, D. E. (1986) Interaction of the *Escherichia coli* HU protein with DNA.
7   Evidence for formation of nucleosome-like structures with altered DNA helical pitch. *J Mol Biol* **187**:
8   47-60.

9   Cases, I. and de Lorenzo, V. (2002) The genomes of *Pseudomonas* encode a third HU protein.
10   *Microbiology* **148**: 1243-1245.

11   Chavarria, M., Fuhrer, T., Sauer, U., Pfluger-Grau, K. and de Lorenzo, V. (2013) Cra regulates the
12   cross-talk between the two branches of the phosphoenolpyruvate: phosphotransferase system of
13   *Pseudomonas putida. Env Microbiol* **15**: 121-132.

14   Chavarria, M., Kleijn, R. J., Sauer, U., Pfluger-Grau, K. and de Lorenzo, V. (2012) Regulatory tasks of
15   the phosphoenolpyruvate-phosphotransferase system of *Pseudomonas putida* in central carbon
16   metabolism. *m*Bio **3**: e00028-12

17   Claret, L. and Rouviere-Yaniv, J. (1997) Variation in HU composition during growth of *Escherichia coli*:
18   the heterodimer is required for long term survival. *J Mol Biol* **273**: 93-104.

19   dos Santos, V. A., Heim, S., Moore, E. R., Stratz, M. and Timmis, K. N. (2004) Insights into the genomic
20   basis of niche specificity of *Pseudomonas putida* KT2440. *Env Microbiol* **6**: 1264-1286.

21   Frank, S., Klockgether, J., Hagendorf, P., Geffers, R., Schock, U., Pohl, T., Davenport, C. F. and
22   Tummler, B. (2011) *Pseudomonas putida* KT2440 genome update by cDNA sequencing and
23   microarray transcriptomics. *Env Microbiol* **13**: 1309-1326.

24   Guell, M., van Noort, V., Yus, E., Chen, W. H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T.,
25   Arumugam, M., Doerks, T., Kuhner, S., Rode, M., Suyama, M., Schmidt, S., Gavin, A. C., Bork, P.
26   and Serrano, L. (2009) Transcriptome complexity in a genome-reduced bacterium. *Science* **326**:
27   1268-1271.

28   Herrera, M. C., Duque, E., Rodriguez-Herva, J. J., Fernandez-Escamilla, A. M. and Ramos, J. L. (2010)
29   Identification and characterization of the PhhR regulon in *Pseudomonas putida. Env Microbiol* **12**:
30   1427-1438.

Holtel, A., Marques, S., Mohler, I., Jakubzik, U. and Timmis, K. N. (1994) Carbon source-dependent inhibition of *xyl* operon expression of the *Pseudomonas putida* TOL plasmid. *J Bacteriol* **176**: 1773-1776.

Jiang, W., Hou, Y. and Inouye, M. (1997) CspA, the major cold-shock protein of *Escherichia coli,* is an RNA chaperone. *J Biol Chem* **272**: 196-202.

Jimenez, J. I., Minambres, B., Garcia, J. L. and Diaz, E. (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Env Microbiol* **4**: 824-841.

Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K. and Wang, J. (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**: 1966-1967.

Martinez-Bueno, M. A., Tobes, R., Rey, M. and Ramos, J. L. (2002) Detection of multiple extracytoplasmic function (ECF) sigma factors in the genome of *Pseudomonas putida* KT2440 and their counterparts in *Pseudomonas aeruginosa* PA01. *Env Microbiol* **4**: 842-855.

Morales, G., Ugidos, A. and Rojo, F. (2006) Inactivation of the *Pseudomonas putida* cytochrome o-ubiquinol oxidase leads to a significant change in the transcriptome and to increased expression of the CIO and cbb3-1 terminal oxidases. *Env Microbiol* **8**: 1764-1774.

Moreno, R., Fonseca, P. and Rojo, F. (2012) Two small RNAs, CrcY and CrcZ, act in concert to sequester the Crc global regulator in *Pseudomonas putida*, modulating catabolite repression. *Mol Microbiol* **83**: 24-40.

Moreno, R., Martinez-Gomariz, M., Yuste, L., Gil, C. and Rojo, F. (2009) The *Pseudomonas putida* Crc global regulator controls the hierarchical assimilation of amino acids in a complete medium: evidence from proteomic and genomic analyses. *Proteomics* **9**: 2910-2928.

Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A., Fouts, D. E., Gill, S. R., Pop, M., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R. T., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Chris Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzez, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Stjepandic, D., Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J. A., Timmis, K. N., Dusterhoft, A., Tummler, B. and Fraser, C. M. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Env Microbiol* **4**: 799-808.

Nikel, P. I., Kim, J., and de Lorenzo V. (2013) Metabolic and regulatory rearrangements underlying glycerol metabolism in *Pseudomonas putida* KT2440. *Env Microbiol* (In Press).

1   Nogales, J., Palsson, B.Ø., and Thiele, I. (2008) A genome-scale metabolic reconstruction of

2   *Pseudomonas putida* KT2440: *i*JN746 as a cell factory. *BMC Syst Biol* **2:** 79.

3   Oberto, J., Nabti, S., Jooste, V., Mignot, H. and Rouviere-Yaniv, J. (2009) The HU regulon is composed

4   of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PLoS One* **4**:

5   e4367.

6   Pfluger-Grau, K. and Gorke, B. (2010) Regulatory roles of the bacterial nitrogen-related

7   phosphotransferase system. *Trends Microbiol* **18**: 205-214.

8   Phadtare, S. and Severinov, K. (2010) RNA remodeling and gene regulation by cold shock proteins.

9   *RNA Biol* **7**: 788-795.

10   Pinson, V., Takahashi, M. and Rouviere-Yaniv, J. (1999) Differential binding of the *Escherichia coli* HU,

11   homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. *J Mol Biol* **287**:

12   485-497.

13   Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V. M. and Wittmann, C. (2012) Industrial

14   biotechnology of *Pseudomonas putida* and related species. *Appl Microbiol Biotechnol* **93**: 2279-

15   2290.

16   Price, M. N., Deutschbauer, A. M., Skerker, J. M., Wetmore, K. M., Ruths, T., Mar, J. S., Kuehl, J. V.,

17   Shao, W. and Arkin, A. P. (2013) Indirect and suboptimal control of gene expression is widespread

18   in bacteria. *Mol Syst Biol* **9**: 660

19   Puchalka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K.N., *et al.* (2008)

20   Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network

21   facilitates applications in biotechnology. *PLoS Comput Biol* **4:** e1000210.

22   Ramos-Gonzalez, M. I. and Molin, S. (1998) Cloning, sequencing, and phenotypic characterization of

23   the *rpoS* gene from *Pseudomonas putida* KT2440. *J Bacteriol* **180**: 3421-3431.

24   Ramstein, J., Hervouet, N., Coste, F., Zelwer, C., Oberto, J. and Castaing, B. (2003) Evidence of a

25   thermal unfolding dimeric intermediate for the *Escherichia coli* histone-like HU proteins:

26   thermodynamics and structure. *J Mol Biol* **331**: 101-121.

27   Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J.

28   P. (2011) Integrative genomics viewer. *Nature Biotech* **29**: 24-26.

29   Rojo, F. (2010) Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and

30   interactions with the environment. *FEMS Microbiol Rev* **34**: 658-684.

1 Ruiz-Manzano, A., Yuste, L. and Rojo, F. (2005) Levels and activity of the *Pseudomonas putida* global
2   regulatory protein Crc vary according to growth conditions. *J Bacteriol* **187**: 3678-3686.

3 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular cloning: A laboratory manual.* Cold Spring
4   Harbor, New York.

5 Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. and Hwa, T. (2010) Interdependence of cell
6   growth and gene expression: origins and consequences. *Science* **330**: 1099-1102.

7 Silva-Rocha, R., Chavarria, M., Kleijn, R. J., Sauer, U. and de Lorenzo, V. (2012) The IHF regulon of
8   exponentially growing *Pseudomonas putida* cells. *Env Microbiol* **15:** 49-63

9 Sittka, A., Lucchini, S., Papenfort, K., Sharma, C. M., Rolle, K., Binnewies, T. T., Hinton, J. C. and Vogel,
10   J. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-
11   transcriptional regulator, Hfq. *PLoS Genet* **4**: e1000163.

12 Stothard, P. and Wishart, D. S. (2005) Circular genome visualization and exploration using CGView.
13   *Bioinformatics* **21**: 537-539.

14 Venturi, V. (2003) Control of *rpoS* transcription in *Escherichia coli* and *Pseudomonas*: why so different?
15   *Mol Microbiol* **49**: 1-9.

16 Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature*
17   *Rev Genet* **10**: 57-63.

18 Yamanaka, K. and Inouye, M. (1997) Growth-phase-dependent expression of *cspD*, encoding a
19   member of the CspA family in *Escherichia coli. J Bacteriol* **179**: 5126-5130.

20 Yamanaka, K., Zheng, W., Crooke, E., Wang, Y. H. and Inouye, M. (2001) CspD, a novel DNA
21   replication inhibitor induced during the stationary phase in *Escherichia coli. Mol Microbiol* **39**: 1572-
22   1584.

23

24

25

1   FIGURES

2

3   **Figure 1**. Analysis of RNA-seq data for *P. putida*.

4
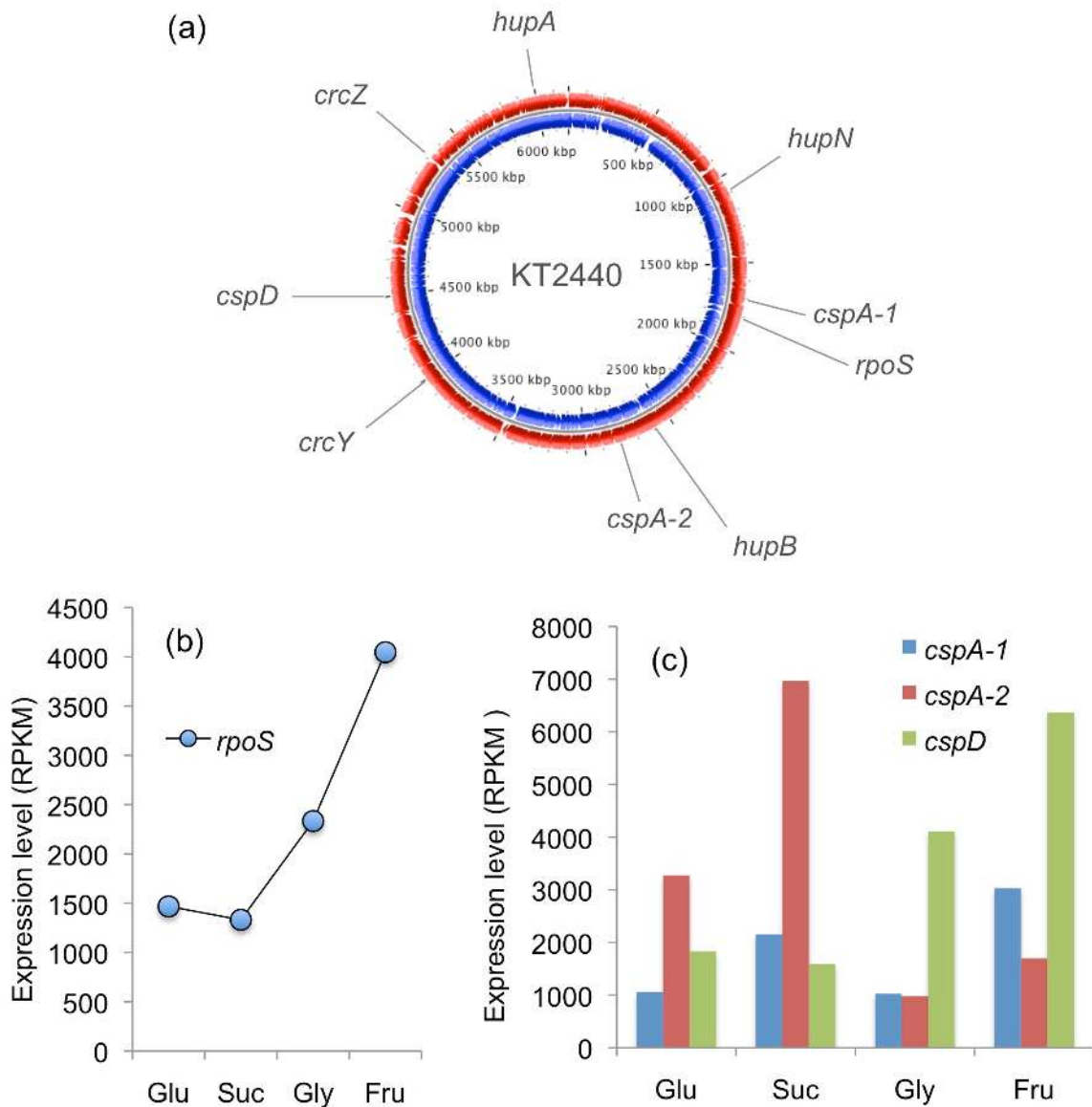
5

6

7   **(a)** Genomic location of the genes and ncRNAs analyzed in this report. Circular genomic representation

8   was generated using cgview software (Stothard and Wishart, 2005). **(b)** Expression level of *rpoS* gene

9   (reads-per-kb-per-million read, RPKM) in the four growth conditions analyzed. (**c**) Expression profiles of

10   the genes encoding cold shock related proteins *cspA-1*, *cspA-2* and *cspD*.

11

12

Figure 2. Anatomy of mRNAs for cold shock proteins of *P. putida*.
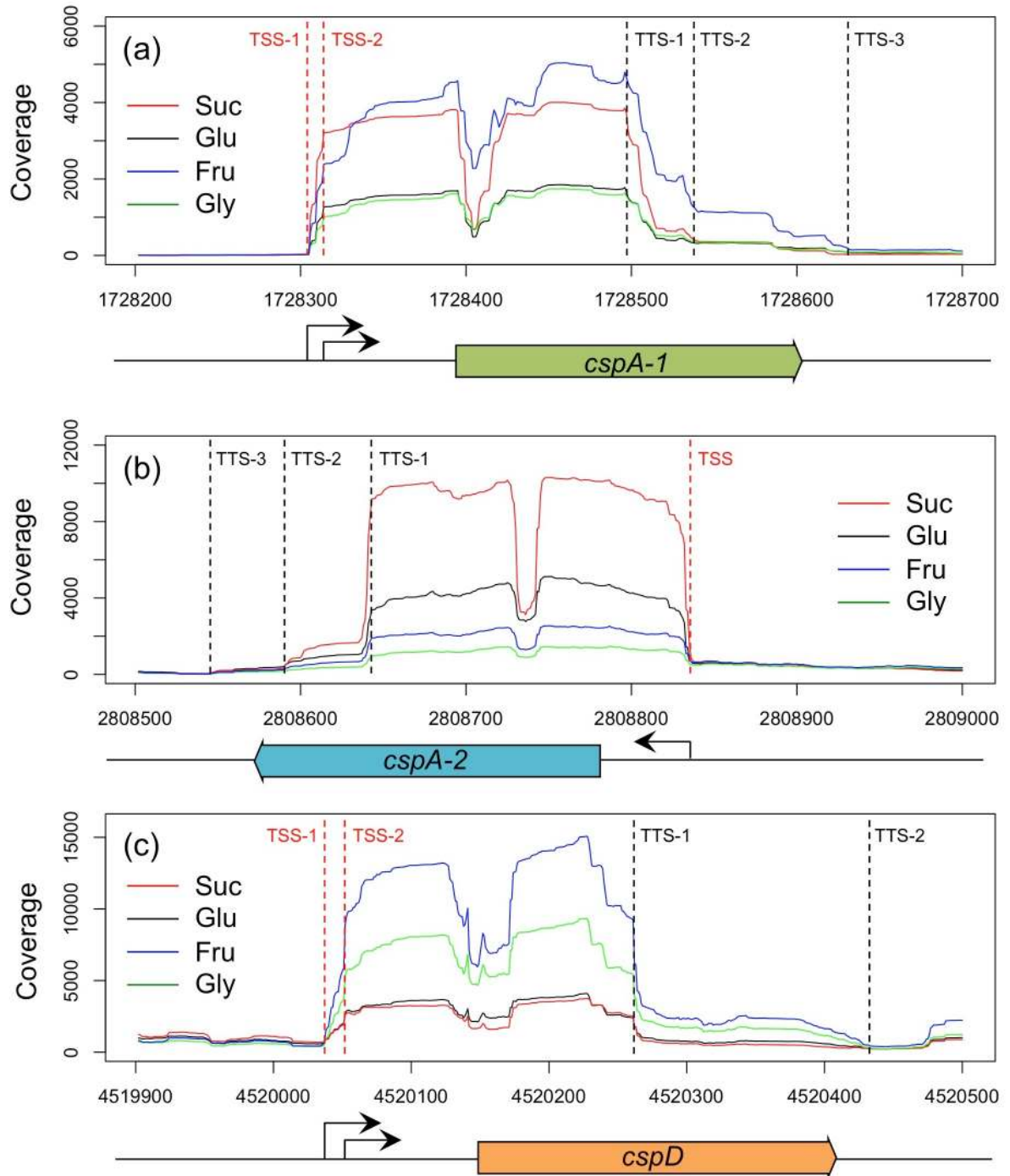


The genomic coordinates are given in the *x*-axes. The coverage (i.e. the count of 91-nt reads that align at each base position) is represented at the *y*-axes. The transcription start sites (TSSs) and transcription termination sites (TTSs) are presented for each gene. The colour of the lines represents the expression profiles of the cells growing in succinate (red), glucose (black), fructose (blue) and glycerol (green). **(a)** *cspA-1*. **(b)** *cspA-2*. **(c)** *cspD*.

1

Figure 3. Expression of HU-encoding genes.

3

4

5
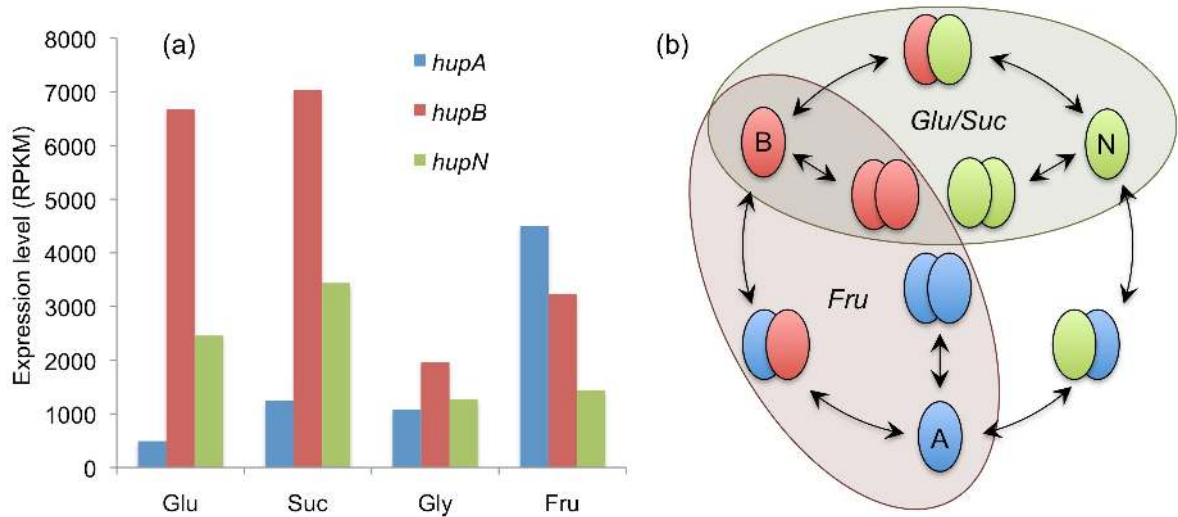
(a) Expression profiles of *hupA*, *hupB* and *hupN* under the four growth conditions assayed. Expression levels in reads-per-kb-per-million read (RPKM). **(b)** Representation of the potential dimers of HU protein in relation to the three subunits encoded by *hupA* (labeled as A), *hupB* (B) and *hupN* (N). Shaded ovals highlights the forms of the protein that are expected to be favoured under the particular growth condition.

10

11

1   **Figure 4.** Carbon source influences *crcZ* and *crcY* expression.

2
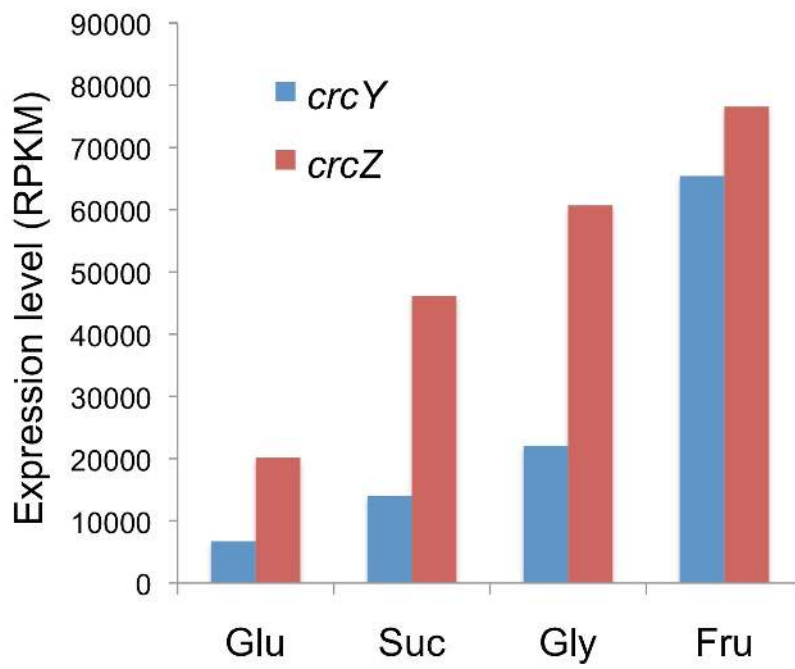
3
4
5   Expression levels represent the average number of reads over the length of the sRNAs (RPKM, see
6   above).

7
8

1 SUPPORTING INFORMATION

2

3 **Supplementary Table S1.** Growth parameters for batch cultures of *Pseudomonas putida* KT2440

4 growing on different C sources.

5

6 **Supplementary Table S2.** Summary of the RNA sequencing coverage data.

7

8 **Supplementary Figure S1.** Summary of genes differentially expressed in Glu, Fru or Gly, in relation to

9 the reference Suc condition.

10

11 **Supplementary Figure S2.** Anatomy of the mRNAs for the housekeeping gene *rpoD* of *P. putida*.

12

13 **Supplementary Experimental Procedures**

14

1 SUPPORTING INFORMATION

2

3

4

5

6 **Supplementary Table S1.** Growth parameters[a] for batch cultures of *Pseudomonas putida* KT2440

7 growing on different C sources.

8

C source[b]	μ^c (h^{-1})
Glucose	0.68 ± 0.05[d]
Succinate	0.72 ± 0.11[d]
Glycerol	0.46 ± 0.02[d]
Fructose	0.21 ± 0.07

9

10 [a] Values shown represent the mean of the corresponding parameter $\pm$ SD of triplicate measurements

11   from at least five independent experiments.

12 [b] Each C source was amended in order to provide an equivalent of 60 mM C atoms (*i.e.*, 10 mM glucose,

13   15 mM succinate, 20 mM glycerol, and 10 mM fructose).

14 [c] The specific growth rate (μ) was determined during exponential growth.

15 [d] Figures taken from Nikel *et al.* (2013)

16

17

18

1   **Supplementary Table S2.** Summary of the RNA sequencing coverage data.

2

Statistics	Carbon source[a]			
	Glucose	Succinate	Glycerol	Fructose
Total number of reads	13,781,166	13,297,244	12,917,984	13,781,166
Total reads (nt)[b]	1,240,304,940	1,196,751,960	1,162,618,560	1,240,304,940
Genome coverage (fold)	201	194	188	201
Mapped reads	13,604,897	13,133,739	12,781,572	13,596,425
Perfect matches	11,060,161	10,704,651	10,509,614	11,098,500
< 5 nt mismatches	2,544,736	2,429,088	2,271,958	2,497,925
Unique matches	13,329,826	12,859,198	12,568,093	13,279,345
Multi-position matches[c]	275,071	274,541	213,479	317,080
Percentage of the genome represented	98.72%	98.77%	98.94%	98.66%
Total unmapped reads rRNA reads	176,269	163,505	136,412	184,741
Reads that aligned to rRNA	70,514	76,221	43,385	76,682
Percentage of rRNA reads	1.02%	1.15%	0.67%	1.11%

3

4   [a] Cells were grown in M9 minimal medium containing the corresponding carbon source and total RNA

5   was extracted when the culture reached the mid-exponential phase of growth (OD_{600} *ca.* 0.4) as

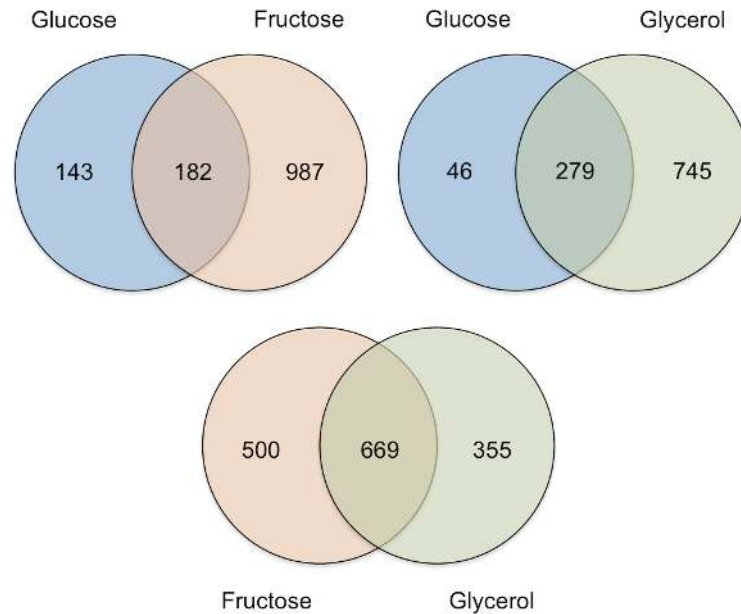6   described in the Experimental Procedures section of the article.

7   [b] nt, nucleotide.

8   [c] These matches were discarded from the analysis.

9

1 **Supplementary Figure S1.** Summary of genes differentially expressed in Glu, Fru or Gly, in relation to

2 the reference Suc condition.

3

4

5

6 For elaboration of the Venn diagrams displayed in the figure, we first computed the genes with a fold

7 change of 2 or more (either up of down regulated) between each of the three conditions (Glu, Fru and

8 Gly) in comparison with Suc. Next, we compared the list of genes differently regulated for each data set

9 to find genes common or specific in each case. As shown comparison between Glu and Fru exposes

10 182 genes differentially regulated in both conditions compared to Suc, while 143 were specific of Glu

11 and 987 were specific to Fru condition. Similarly, comparison between Glu and Gly shows that most of

12 the genes differentially expressed in Glu varied also in Gly (279 out of 325), with only 46 genes found to

13 be specific to Glu. In contrast, most of the genes identified in the Gly condition were not found in Glu

14 (745 genes in total). Finally, the case of Fru *vs* Gly shows the more extensive overlapping. In this case,

15 669 genes were differentially expressed in both conditions in respect to Suc, while 500 were Fru specific

16 and 355 were Gly specific.

17

1   **Supplementary Figure S2.** Anatomy of the mRNAs for the housekeeping gene *rpoD* of *P. putida*.

2
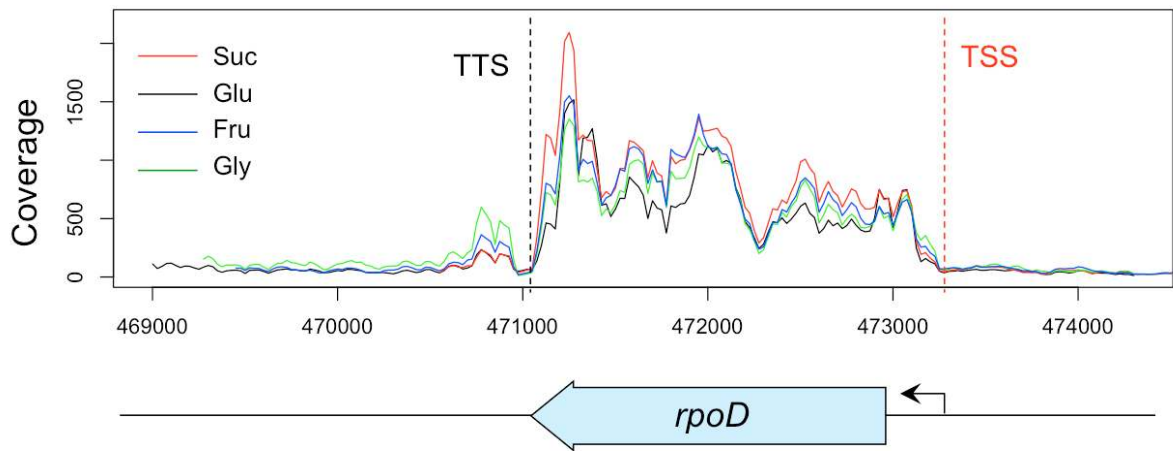
3

4

5   The genomic coordinates are given in the *x*-axes. The count of reads at each nt position are

6   represented at the *y*-axes. The transcription start site (TSS) and transcription termination site (TTS) are

7   presented for each gene. The colour of the lines represents the expression profiles of the cells growing

8   in succinate, glucose, fructose and glycerol as indicated. *rpoD* displays a typical expression signature

9   where the RNAseq signal is observed along the entire gene.

10

11

12

1 **Supplementary Experimental Procedures**

2

3 *Calculation of gene expression*

4

5 We resorted to the reads-per-kb-per-million read (RPKM) method as described by Mortazavi *et al.*

6 (2008) to calculate gene expression levels as deduced from RNA reads. RPKM values in each case

7 were obtained according to:

$$RPKM(A) = \frac{10^6 \times C}{\frac{NL}{10^3}}$$

8

9

10 were RPKM(*A*) is the expression level of gene *A*, *C* is the number of reads uniquely aligned to gene *A*,

11 *N* is the total number of reads uniquely aligned to all genes, and *L* is the base number in the CDS of

12 gene A. The RPKM method is able to eliminate the influence of different gene lengths and sequencing

13 discrepancies on the calculation of gene expression. Therefore, the calculated gene expression can be

14 directly used to compare differences in gene expression among different samples. If there were more

15 than one transcript for a given gene, the longest one was used to calculate its expression level and

16 coverage.

17

18 *Analysis of differentially expressed genes*

19

20 This analysis aims at predicting genes with different expression levels. The algorithm used to identify

21 differentially expressed genes between two samples was based on the work by Audic and Claverie

22 (1997). The null hypothesis and alternative hypothesis used when performing hypothesis testing were:

23 H_0, a given gene has the same expression level in two samples; and H_a, some gene has different

24 expression levels in two experimental samples. If *x* is the number of reads that can uniquely map to

25 gene *A*, for each transcript representing a small fraction of the library, *p(x)* will closely follow the Poisson

26 distribution according to λ, the actual transcript number for gene *A*.

27

$$p(x) = \frac{e^{-\lambda} \times \lambda^x}{x!}$$

28

29

30 If thousands of hypothesis tests are conducted simultaneously, the suitable *p*-value for individual tests is

31 not enough to guarantee low false discovery rate (FDR) within the same sample, and multiple testing

correction has to be applied in order to decrease the *p*-value for each individual hypothesis being tested to meet the lowest FDR value among samples.

The FDR control is a statistical method used in multiple hypothesis testing to correct the raw comparison results for *p*-values. The procedure suggested by Benjamini *et al.* (2001) was followed in this study. The smaller the FDR and the larger the RPKMs ratio between two technical samples, the larger the difference of the expression level between the samples. In the present analysis, we choose those samples with FDR ≤ 0.001 and a RPKMs ratio larger than 2.

REFERENCES

Audic, S., and Claverie, J. M. (1997) The significance of digital gene expression profiles. *Genome Res* **7:** 986-995.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125:** 279-284.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621-628.

Nikel, P. I., Kim, J., and de Lorenzo V. (2013) Metabolic and regulatory rearrangements underlying glycerol metabolism in *Pseudomonas putida* KT2440. *Env Microbiol* (In Press).