

Multivariate analysis of microarray data by principal component discriminant analysis: prioritizing relevant transcripts linked to the degradation of different carbohydrates in *Pseudomonas putida* S12

Mariët J. van der Werf, Bart Pieterse,[†] Nicole van Luijk, Frank Schuren, Bianca van der Werff-van der Vat, Karin Overkamp and Renger H. Jellema

TNO Quality of Life, PO Box 360, 3700 AJ Zeist, The Netherlands

Correspondence

Mariët J. van der Werf
vanderWerf@voeding.tno.nl

The value of the multivariate data analysis tools principal component analysis (PCA) and principal component discriminant analysis (PCDA) for prioritizing leads generated by microarrays was evaluated. To this end, *Pseudomonas putida* S12 was grown in independent triplicate fermentations on four different carbon sources, i.e. fructose, glucose, gluconate and succinate. RNA isolated from these samples was analysed in duplicate on an anonymous clone-based array to avoid bias during data analysis. The relevant transcripts were identified by analysing the loadings of the principal components (PC) and discriminants (D) in PCA and PCDA, respectively. Even more specifically, the relevant transcripts for a specific phenotype could also be ranked from the loadings under an angle (biplot) obtained after PCDA analysis. The leads identified in this way were compared with those identified using the commonly applied fold-difference and hierarchical clustering approaches. The different data analysis methods gave different results. The methods used were complementary and together resulted in a comprehensive picture of the processes important for the different carbon sources studied. For the more subtle, regulatory processes in a cell, the PCDA approach seemed to be the most effective. Except for glucose and gluconate dehydrogenase, all genes involved in the degradation of glucose, gluconate and fructose were identified. Moreover, the transcriptomics approach resulted in potential new insights into the physiology of the degradation of these carbon sources. Indications of iron limitation were observed with cells grown on glucose, gluconate or succinate but not with fructose-grown cells. Moreover, several cytochrome- or quinone-associated genes seemed to be specifically up- or downregulated, indicating that the composition of the electron-transport chain in *P. putida* S12 might change significantly in fructose-grown cells compared to glucose-, gluconate- or succinate-grown cells.

Received 17 June 2005

Revised 21 October 2005

Accepted 26 October 2005

INTRODUCTION

Increasingly, functional genomics tools like transcriptomics, proteomics and metabolomics are used for the identification of biological processes that are important for a specific subject of study. The most challenging aspect of functional

genomics studies is the comprehensive and accurate interpretation of the overwhelming amount of data generated.

Currently, two approaches are commonly employed in order to identify the relevant biomolecules from functional genomics data. Most frequently, analysis of the fold difference in expression (univariate data analysis) between the condition of interest and a reference condition is used to identify the biomolecules (i.e. genes, proteins or metabolites) that are potentially of interest. Subsequently, biomolecules whose response is above a certain threshold (e.g. more than twofold difference) are selected and studied in more detail. Increasingly, hierarchical cluster analysis (HCA) is employed for analysing functional genomics data. In this

[†]Present address: BioDetection Systems BV, Kruislaan 406, 1098 SM Amsterdam, The Netherlands.

Abbreviations: HCA, hierarchical cluster analysis; MVDA, multivariate data analysis; PCA, principal component analysis; PCDA, principal component discriminant analysis.

Supplementary tables are available with the online version of this paper.

case, biomolecules or experiments with similar expression profiles are grouped. HCA is only useful when more than two functional genomics datasets of more than two situations are available. Visual inspection of the hierarchical clusters results in the identification of coexpressed biomolecules that are specifically regulated under the condition of interest (e.g. Heyer *et al.*, 1999; Tefferi *et al.*, 2002).

The drawback of both these approaches is that they generally result in large numbers of leads: in the order of tens or hundreds. As it is too time-consuming to analyse all of these in more detail with molecular biological, biochemical or bioinformatic approaches in order to (experimentally) validate the targets identified, there is a need for data analysis tools that allow one to rank the potential targets. Potentially, the fold difference in expression level can be used to rank the targets. However, the fundamental basis behind such a ranking, i.e. that biomolecules that show the largest response are also the most important for the question under study, is doubtful (van der Werf, 2005).

Multivariate data analysis (MVDA) tools seem much better suited to prioritize leads from functional genomics datasets. These tools take into account the inherent interdependency of biomolecules. Principal component analysis (PCA) is the most frequently applied multivariate statistics tool. It has been applied for many decades in epidemiology, econometrics, ecology, etc., but only recently has the potential of these tools in cellular biology been recognized (Orr & Scherf, 2002; Michaud *et al.*, 2003). Although the mathematics behind PCA might seem complex to the untrained cellular biologist, the basic principle behind it is straightforward, i.e. PCA combines two, or more, correlated factors (i.e. transcripts) into one new variable, a principal component (PC) (Orr & Scherf, 2002; van der Werf *et al.*, 2005). Thus in PCA the dimensionality of the dataset is reduced by replacing the original variables by a smaller number of newly formed variables that are linear combinations of the original variables and that explain the majority of the information (variability) from the experiment. For each PC, loadings (or weights) reflect the influence of the original variables, whereas scores (coefficients of the PC) reflect the contribution of each PC in every sample. PCA and the related tool principal component discriminant analysis (PCDA; Hoogerbrugge *et al.*, 1983) are currently mostly used for the classification of samples with a similar expression pattern, e.g. related to a specific treatment or phenotype. However, these tools are not only descriptive but also allow the identification of the specific biomolecules that are most important for the differences between the groups. The most important variables are identified by analysing the strength of the correlation of every biomolecule with the biological process of interest.

The goal of the research described here was to empirically investigate the effectiveness of the multivariate data analysis tools PCA and PCDA for the ranking of important transcripts from microarray data and to compare the results with those obtained by the fold-difference and HCA approaches.

To this end, genes from *Pseudomonas putida* S12 were identified whose expression is specific for growth on one or more of four different carbon sources (i.e. glucose, gluconate, fructose and succinate), as the genes encoding enzymes of carbohydrate catabolism and their regulation are still largely unknown (Petruschka *et al.*, 2002). In order to avoid biological prejudgments during the evaluation of the data analysis tools, anonymous clone-based arrays were used. The identity of the genes only became known after completing the data analysis phase by sequencing the inserts in the clones corresponding to the relevant spots.

METHODS

Micro-organism and cultivation conditions. *P. putida* S12 was previously isolated on styrene as the sole source of carbon and energy (Hartmans *et al.*, 1990). Cultures were grown in batch fermentations at 30 °C in a Bioflow II (New Brunswick Scientific) bioreactor containing 2 litres of mineral salts medium with as carbon source 20 g l⁻¹ glucose, or the equivalent amount on a C-mol basis of fructose (20 g l⁻¹), disodium succinate.6H₂O (45 g l⁻¹) or sodium gluconate (24 g l⁻¹). The mineral salts medium contained the following (per litre of demineralized water): 1.55 g K₂HPO₄, 0.85 g NaH₂PO₄·H₂O, 8.0 g NH₄Cl, 0.5 g (NH₄)₂SO₄, 0.3 g MgCl₂·6H₂O, 40 mg EDTA, 2 mg ZnSO₄·7H₂O, 1 mg CaCl₂·2H₂O, 15 mg FeSO₄·7H₂O, 0.2 mg Na₂MoO₄·2H₂O, 2 mg CuSO₄·5H₂O, 0.4 mg CoCl₂·6H₂O and 1 mg MnCl₂·4H₂O. The culture was inoculated with 5% (v/v) of a preculture grown without shaking for 24 h on the same medium. A constant pH (pH 7.0) was maintained by automatic titration with 2 M KOH and 1 M H₂SO₄. After an overnight oxygen-limited growth phase, the dissolved oxygen concentration was maintained at 20% by manually adjusting the stirrer speed and inlet airflow. Samples were taken from the bioreactor at an OD₆₀₀ of 10, and immediately quenched in -45 °C methanol as previously described (Pieterse *et al.*, 2005). Cell pellets were stored at -45 °C until used.

RNA isolation. RNA was isolated from the cells using the hot borate method, basically as described by Wan & Wilkins (1994). RNA purity and concentrations were determined both spectrophotometrically and on agarose gel. The RNA isolates were checked for residual RNase activity by comparing samples that were incubated for 1 h at 37 °C with the initial material on an agarose gel.

Array design. The microarray used is a clone-based array. To this end, a chromosomal library of *P. putida* S12 was constructed by Baseclear (Leiden, The Netherlands). DNA fragments were obtained by shearing, and fragments of 2–3 kb were blunt-end cloned in pSMARTLC (Lucigen). The chromosomal library was ordered in 96-well microtitre plates and stored as glycerol stocks at -80 °C. Subsequently, in total 5000 genomic fragments were amplified by PCR, purified and arrayed as described previously (Pieterse *et al.*, 2005).

Fluorescent labelling and hybridization. Differential transcript levels were determined by two-colour fluorescent hybridizations of the corresponding cDNAs on the clone-array. The RNA samples were labelled by random hexamers primed *in vitro* reverse transcription with either Cy5- or Cy3-labelled dUTP. Labelling, hybridization and washing were performed as previously described (Pieterse *et al.*, 2005). In all instances, Cy3-labelled cDNA of a batch of the fructose-F3 sample was used as the reference condition.

Image analysis. The fluorescent signals from the two different labels on the hybridized arrays were quantified with a ScanArray Express scanner (Packard Bioscience) and Imagene 4.2 software

(BioDiscovery). Spots with a Flag 0 (spots the quality of which was approved by the Imagene software package) or a Flag 3 (spots which obtained a warning by the Imagene software package for a manual check on their quality) were selected. Subsequently, spots from which the difference between the mean signal of the spot and the mean signal of the background was larger than zero times the background standard deviation were excluded from further analysis. After removal of the empty spots and the spots from which the signal exceeded the detection limit of the scanner, 3676 spots (68 %) remained for further analysis.

Data preprocessing and normalization. The data from the microarray analysis were delivered as Excel files and the data were imported into Matlab (version 6.5.1; The Mathworks). Within-slide, intensity-dependent normalizations were performed with the scatter plot smoother LOWESS (polynomial order=1) using a Matlab routine (copyright 1998 by Datatool). The user-defined fraction of data used for smoothing at each point was set at 25 % for all slides. Subsequently, these preprocessed data were used as the input for significance and MVDA analysis.

Significance analysis. Prior to significance analysis, a data transformation was applied to the normalized ratios in order to obtain a normal distribution of the data (Pieterse *et al.*, 2005). Significance analysis was performed by means of a 1-way ANOVA. Subsequently a Tukey HSD test was performed to determine whether a significant differential expression level (99 % confidence interval) was observed under a specific condition.

Multivariate data analysis. Datasets were scaled [$x/(x_{\max}-x_{\min})$] per variable prior to MVDA analysis. Two different MVDA tools were applied in Matlab (The Mathworks): (i) principal component analysis (PCA) (PLS Toolbox for Matlab, version 3.0.2, Eigenvector Research), and (ii) principal component discriminant analysis (PCDA) [algorithm reproduced from Hoogerbrugge *et al.* (1983) and programmed into Matlab]. In PCDA, the centre of a group was established by taking the mean score on D1 and D2. Subsequently, the loadings were determined under the angle present between the fructose group and that of one of the other three carbon sources.

For the hierarchical clustering and visualization of the results, the programs CLUSTER and TREEVIEW were applied (Eisen *et al.*, 1998). Only those genes or operons were included that fell within the 99 % confidence interval of the Tukey HSD test. Average linkage clustering was performed on the natural logarithm of the ratios of the spots taken into account.

Nucleotide sequencing and sequence analysis. Clones containing the inserts selected by MVDA as being relevant were traced back in the 96-well plates. Approximately 500 bp of both the 3'- and the 5'-end of the inserts in these clones was subsequently sequenced using universal primers based on pSMART by Baseclear (Leiden, the Netherlands), and the complete gene content of these inserts was inferred from the *P. putida* KT2440 genome sequence (Nelson *et al.*, 2002). DNA sequences were identified by similarity searches against the TIGR (www.tigr.org/) and NCBI (www.ncbi.nlm.nih.gov/) database libraries using BLAST. Gene numbers used in this study (PP numbers) are based on the gene numbering of the *P. putida* KT2440 genome (Nelson *et al.*, 2002).

RESULTS

Experimental design

As the full genome sequence of *P. putida* S12 is not known and in order to unbiasedly select the relevant clones using PC(D)A (see below), an anonymous clone-based array of

P. putida S12 was constructed. It was assumed that the presence of 200 nucleotides of a specific gene was sufficient to detect a hybridization signal with a spot. Using the formula of Akopyants (Akopyants *et al.*, 2001), and based on an estimated genome size of 6.2 Mb for *P. putida* S12 (see also Nelson *et al.*, 2002) and the fact that 5000 spots could be spotted on the microarray slides, chromosomal fragments of, on average, 2.5 kb were generated by shearing, resulting in a full genome coverage of 95 %.

In order to address the biological and technical (array) variability, *P. putida* S12 was grown on the four carbon sources, i.e. D-fructose [F] ($\mu=0.18\pm0.02\text{ h}^{-1}$), D-glucose [G] ($\mu=0.28\pm0.02\text{ h}^{-1}$), D-gluconate [N] ($\mu=0.21\text{ h}^{-1}$) and succinate [S] ($\mu=0.21\pm0.01\text{ h}^{-1}$), in triplicate in independent batch cultures. The cells were harvested and immediately quenched in order to prevent alterations in the mRNA composition of the samples (Pieterse *et al.*, 2005). In one instance, two samples were harvested from the same fermenter (fermentation 3 of glucose-grown cells). Subsequently, mRNA was isolated from these samples. In all instances, RNA isolated from the third fructose fermentation was used as the reference. Two independent microarray hybridizations of every sample were performed.

PCA analysis of transcriptomes

In order to identify the transcripts that are the most important for the differences between the cells grown on the different carbon sources, the multivariate data analysis tool PCA was applied. In Fig. 1(a), the results of the PCA analysis of the transcription datasets are visualized in a two-dimensional plot. A cloud of points is observed, with each point representing the transcriptome of the different samples (van der Werf *et al.*, 2005). Transcriptomes that end up close together are overall more similar, while more dissimilar transcriptomes are further apart. It can clearly be seen that, with the exception of transcriptomes of gluconate- and succinate-grown cells, the transcriptomes of cells grown on the same carbon source are more similar than the transcriptomes of cells grown on different carbon sources. Also in plots of PC1 versus PC3 and of PC2 versus PC3, the transcriptome datasets of gluconate- and succinate-grown cells slightly overlapped (results not shown). The separation of the different groups of transcriptomes originating from the same carbon sources by PCA indicates that the overall variation in the datasets due to biological and technical variation is less than the differences introduced by changing the growth substrate of *P. putida* S12. PC1 explains 49 % of the total variance in these datasets, while PC2 explains only 6 % of the variance.

PCDA analysis

A supervised variant of PCA is PCDA (Hoogerbrugge *et al.*, 1983). In contrast to PCA, PCDA also takes into account information about external variables (group information) when reducing the dimensionality of the datasets. Therefore, it is better suited than PCA for clustering analysis. The result

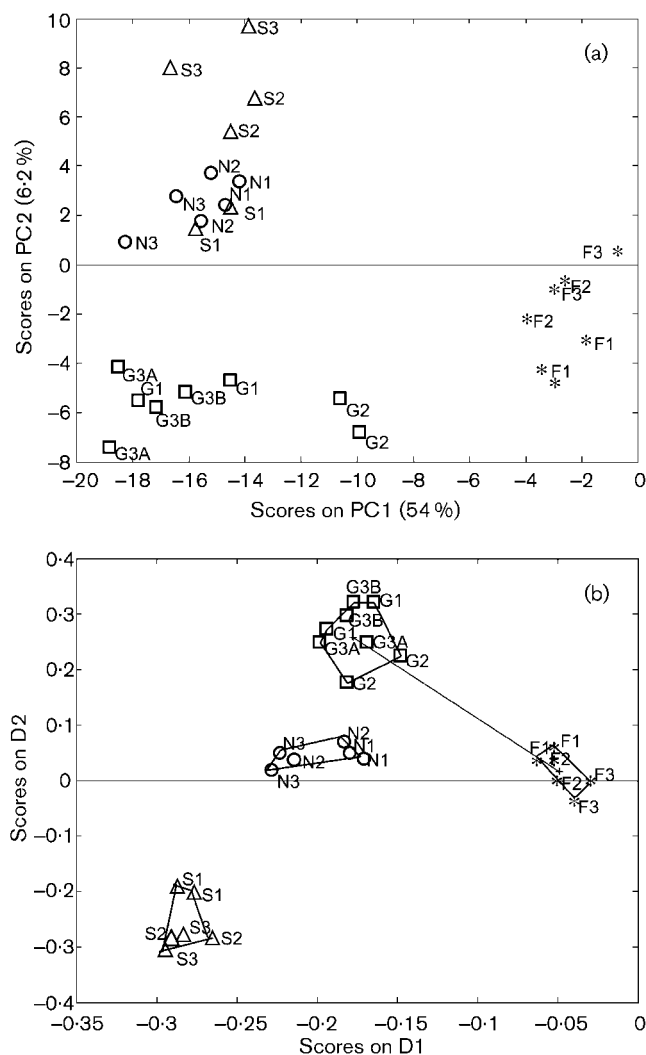


Fig. 1. PCA (a) and PCDA (b) plots of the transcription profiles. The letters F (fructose), G (glucose), N (gluconate) and S (succinate) refer to the carbon sources from which these transcription profiles were obtained. The line between the fructose and the glucose group in (b) is a new discriminant axis through the centre of the fructose group and the centre of the glucose group.

of analysing the transcriptome data by PCDA is shown in Fig. 1(b). The transcriptomes originating from cells grown on the same carbon source form much compacter groups compared to PCA analysis (Fig. 1a) and, moreover, the distances between the groups are larger, indicating a more optimal data analysis with respect to the carbon sources studied. The fact that non-overlapping groups are observed after PCDA analyses indicates that there is information present in the transcriptome datasets that is specific for all the four different carbon sources.

As with PCA, in PCDA the strongly correlating variables are combined into one new variable that is now called a discriminant (D). The discriminants are linear combinations

of the original variables, i.e. the transcripts. When the (absolute values of the) loadings for each of the transcripts in the different D's are studied, transcripts can be identified that are the most important for the variance explained by that D. D1 is mainly responsible for explaining the difference between fructose and succinate. This can most easily be seen by projecting all transcriptomes in Fig. 1(b) on D1 (the x-axis). In a similar way, transcripts with a high absolute loading in D2 are important for explaining the difference between glucose and succinate. The 3'- and the 5'-ends of the inserts in the clones belonging to the spots with the highest absolute value for the loadings on D1 and D2 were subsequently sequenced. The identities of the genes present on these inserts (Table 1) were identified by performing a homology search using BLAST (Altschul *et al.*, 1990) based mainly on the annotated *P. putida* KT2440 genome sequence (Nelson *et al.*, 2002).

Several of the spots that are the most important in D1 contain the genes encoding fructose-specific phosphotransferase and phosphofructokinase (Fig. 2). This operon was present on two separate clones with high loadings on D1, indicating that it is statistically relevant for the variance explained by D1, and that it is not a false positive. The presence of these genes is in agreement with the transcriptome groups originating from the same carbon sources separated on D1 (mainly fructose and succinate). Another group of genes that is present more than once amongst the 30 clones with the highest loading on D1 is that encoding the outer-membrane protein H1 and the transcriptional regulator PhoP/sensor protein PhoQ operon. For D2, (i) the gluconokinase/gluconate transporter, (ii) the ferric siderophore receptor, (iii) the 2-ketogluconate/2-ketogluconate kinase/epimerase/regulator, (iv) the genes involved in the biosynthesis of flagella (chemotaxis) and (v) the glucose-6-phosphate dehydrogenase/6-phosphogluconolactonase/2-dehydro-3-deoxyphosphogluconate aldolase gene clusters are present more than once amongst the variables that have the highest loading in D2 (see also Fig. 2). Also in this case, the fact that specifically these genes are important for D2 is not surprising in view of the fact that the difference between the groups of glucose- and succinate-grown cells is mainly explained by D2 (Fig. 1b).

Identification of transcripts important for the different carbon sources

Although the identification of the genes that have a high loading in the different D's is helpful in identifying genes that are important for the growth on a specific carbon source, it is not always possible to identify genes that are important for one of the different carbon sources: compare, for example, the glucose and gluconate group information on D1 and the gluconate and fructose group information on D2 (Fig. 1b). Therefore, the loadings under an angle were determined. Now, three new discriminants were defined through the centre of the fructose group and the centres of one of the three other groups (Fig. 1b). Subsequently, transcripts important for the difference between the fructose

Table 1. Identity of the genes present on the 30 clones with the highest absolute loadings on the discriminants

| Rank | Clone name | Loading | ORF no. | Identity of genes on clone |
|------------------------|------------|---------|---------------|---|
| Discriminant D1 | | | | |
| 1 | Pp.017D9 | 3.83 | PP0793–PP0795 | Phosphotransferase system, fructose-specific EI/HPr/EIIA components; 1-Phosphofructokinase; Phosphotransferase system, fructose-specific IIBC component |
| 2 | Pp.010H3 | 3.56 | PP0902–PP0904 | Sensor histidine kinase; Conserved hypothetical protein; InaA protein |
| 3 | Pp.015G9 | 3.52 | – | No hits |
| 4 | Pp.008F6 | –3.50 | PP1185–PP1187 | Outer-membrane protein H1; Transcriptional regulatory protein PhoP; Sensor protein PhoQ |
| 5 | Pp.029H8 | –3.48 | PP2146–PP2147 | Helicase, SNF2/RAD54 family; Conserved hypothetical protein |
| 6 | Pp.052C12 | 3.41 | PP3992–PP3997 | Xanthine/uracil permease family protein; DsrE protein; Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Glycosyl transferase, putative |
| 7 | Pp.054E8 | 3.37 | PP0795–PP0798 | Phosphotransferase system, fructose-specific IIBC component; Conserved hypothetical protein; Conserved hypothetical protein; GGDEF domain protein |
| 8 | Pp.051C4 | –3.30 | PP1185–PP1187 | Outer-membrane protein H1; Transcriptional regulatory protein PhoP; Sensor protein PhoQ |
| 9 | Pp.045D9 | 3.17 | – | No hits |
| 10 | Pp.111G3 | 3.16 | PP0072 | Quinone oxidoreductase |
| 11 | Pp.051G11 | 3.15 | PP1839–PP1842 | Hypothetical protein; Conserved hypothetical protein; Cytochrome <i>c</i> family protein; Glutamine amidotransferase, class I |
| 12 | Pp.045G3 | 3.06 | PP1872–PP1875 | Aminotransferase, class I; PilB-related protein; Glutathione peroxidase; Sensor histidine kinase/response regulator |
| 13 | Pp.008A7 | –3.04 | PP0865–PP0867 | RNA polymerase σ^{70} factor, putative; Transmembrane sensor, putative; FecA-like outer-membrane receptor |
| 14 | Pp.006C4 | 3.02 | – | <i>P. putida</i> S12 isolate KL54C2 toluene-induced genomic sequence |
| 15 | Pp.032D9 | 3.02 | PP0897–PP0900 | Fumarate hydratase, class I; Hypothetical protein; Ferredoxin reductase, putative; PAP2 family protein |
| 16 | Pp.061D7 | –2.99 | PP3797–PP3799 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein |
| 17 | Pp.008C3 | –2.98 | – | No hits |
| 18 | Pp.020C7 | 2.97 | PP0709–PP0712 | Transporter, NCS1 nucleoside transporter family; Conserved domain protein hydrolase, isochorismatase family; Conserved hypothetical protein |
| 19 | Pp.028C2 | 2.94 | PP0071, – | Hypothetical protein; No hits |
| 20 | Pp.059C7 | 2.94 | PP0982–PP0985 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Cold-shock domain family protein |
| 21 | Pp.051G1 | 2.94 | PP1297–PP1298 | General amino acid ABC transporter, periplasmic binding protein; General amino acid ABC transporter, permease protein |
| 22 | Pp.110A6 | –2.94 | PP0151–PP0154 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Acetyl-CoA hydrolase/transferase family protein |
| 23 | Pp.031H6 | 2.91 | – | <i>Pseudomonas</i> sp. Y2 DNA phenylacetate-CoA-ligase; <i>Pseudomonas fluorescens</i> hybrid histidine kinase homologue; <i>Pseudomonas</i> sp. Y2 DNA styrene sensor kinase |
| 24 | Pp.045D8 | 2.90 | PP0021–PP0024 | Hypothetical protein; Hypothetical protein; Hypothetical protein; Membrane protein, putative |
| 25 | Pp.044G5 | 2.89 | PP0790–PP0792 | Inner membrane protein AmpE; Deoxyribonuclease, TatD family; Fructose transport system repressor FruR |
| 26 | Pp.008G6 | –2.88 | PP3606–PP3607 | Quinone oxidoreductase; Conserved hypothetical protein |
| 27 | Pp.060D9 | 2.88 | PP3461–PP3462 | Hypothetical protein; Hypothetical protein |
| 28 | Pp.060B4 | 2.87 | PP0982–PP0985 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Cold-shock domain family protein |
| 29 | Pp.047E11 | 2.86 | PP3159, – | <i>benABC</i> operon transcriptional activator BenR; No hits |
| 30 | Pp.020E3 | –2.86 | PP3800–PP3802 | Conserved hypothetical protein; Cation ABC transporter, periplasmic cation-binding protein, putative; Cation ABC transporter, ATP-binding protein, putative |

Table 1. cont.

| Rank | Clone name | Loading | ORF no. | Identity of genes on clone |
|------------------------|------------|---------|---------------|---|
| Discriminant D2 | | | | |
| 1 | Pp.053G5 | 1·87 | PP3416–PP3418 | Gluconokinase; Gluconate transporter; Hypothetical protein |
| 2 | Pp.028C8 | 1·83 | PP3415–PP3417 | Transcriptional regulator, LacI family; Gluconokinase; Gluconate transporter |
| 3 | Pp.113C1 | 1·81 | PP0168 | Surface adhesion protein, putative |
| 4 | Pp.044B3 | 1·79 | PP3416–PP3417 | Gluconokinase; Gluconate transporter |
| 5 | Pp.013F11 | 1·63 | PP3414–PP3415 | Methyl-accepting chemotaxis transducer/sensory box protein; Transcriptional regulator, LacI family |
| 6 | Pp.008A7 | 1·57 | PP0865–PP0867 | RNA polymerase σ^{70} factor, putative; Transmembrane sensor, putative; FecA-like outer-membrane receptor |
| 7 | Pp.112B12 | 1·57 | PP3155 | Outer-membrane ferric siderophore receptor, putative |
| 8 | Pp.112C2 | 1·56 | PP3154–PP3155 | Conserved hypothetical protein; Outer-membrane ferric siderophore receptor, putative |
| 9 | Pp.111H3 | 1·55 | PP3414–PP3416 | Methyl-accepting chemotaxis transducer/sensory box protein; Transcriptional regulator, LacI family; Gluconokinase |
| 10 | Pp.038H2 | 1·54 | PP3378–PP3380 | 2-Ketogluconate kinase; Epimerase KguE, putative; Transcriptional regulator PtxS |
| 11 | Pp.028E3 | 1·52 | PP3376–PP3377 | 2-Ketogluconate 6-phosphate reductase; 2-Ketogluconate transporter, putative |
| 12 | Pp.053D11 | 1·48 | PP3377–PP3380 | 2-Ketogluconate transporter, putative; 2-Ketogluconate kinase; Epimerase KguE, putative; Transcriptional regulator PtxS |
| 13 | Pp.012E11 | –1·47 | PP4383–PP4386 | Flagellar P-ring protein precursor FlgI; Flagellar L-ring protein precursor FlgH; Flagellar basal-body rod protein FlgG; Flagellar basal-body rod protein FlgF |
| 14 | Pp.009G4 | 1·44 | PP0535 | Outer-membrane ferric siderophore receptor |
| 15 | Pp.111F2 | 1·44 | PP3418–PP3419 | Hypothetical protein; σ^{54} -dependent transcriptional regulator/response regulator |
| 16 | Pp.054B10 | 1·42 | PP1020–PP1022 | Conserved hypothetical protein; Transcriptional regulator HexR; Glucose-6-phosphate 1-dehydrogenase |
| 17 | Pp.051A4 | 1·41 | PP4669–PP1016 | OmpA family protein; GGDEF domain protein; Sugar ABC transporter, permease protein |
| 18 | Pp.014E3 | –1·41 | PP4389–PP4392 | Flagellar basal-body rod modification protein FlgD; Flagellar basal-body rod protein FlgC; Flagellar basal-body rod protein FlgB; Chemotaxis protein methyltransferase CheR |
| 19 | Pp.036E9 | 1·40 | PP1023–PP1025 | 6-Phosphogluconolactonase; 2-Dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase; 2-Isopropylmalate synthase |
| 20 | Pp.044A3 | –1·39 | PP4388–PP4390 | Flagellar hook protein FlgE; Flagellar basal-body rod modification protein FlgD; Flagellar basal-body rod protein FlgC |
| 21 | Pp.060C9 | –1·39 | PP0257–PP0259 | Formate dehydrogenase accessory protein FdhD; LysM domain protein; Hydrolase, haloacid dehalogenase-like family |
| 22 | Pp.022E4 | –1·39 | PP0613–PP0614 | Amidase family protein; <i>N</i> -Carbamoyl- β -alanine amidohydrolase, putative |
| 23 | Pp.036C12 | –1·39 | PP4386–PP4389 | Flagellar basal-body rod protein FlgF; Hypothetical protein; Flagellar hook protein FlgE; Flagellar basal-body rod modification protein FlgD |
| 24 | Pp.057B10 | 1·38 | PP1021–PP1023 | Transcriptional regulator HexR; Glucose-6-phosphate 1-dehydrogenase; 6-Phosphogluconolactonase |
| 25 | Pp.047A12 | 1·38 | PP1015 | Sugar ABC transporter, periplasmic sugar-binding protein |
| 26 | Pp.028E6 | –1·37 | PP2453 | L-Asparaginase II |
| 27 | Pp.036C7 | –1·35 | PP5338–PP5340 | Aspartate ammonia-lyase; Transcriptional regulator, AraC family; Acetylpolymine aminohydrolase |
| 28 | Pp.032G9 | –1·34 | PP0308–PP0310 | Dipeptidase, putative; Conserved hypothetical protein; Oxidoreductase, FMN-binding |
| 29 | Pp.045A7 | 1·34 | PP1023–PP1025 | 6-Phosphogluconolactonase; 2-Dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase; 2-Isopropylmalate synthase |
| 30 | Pp.014C6 | 1·34 | PP1018–PP1019 | Sugar ABC transporter, ATP-binding subunit; Porin B |

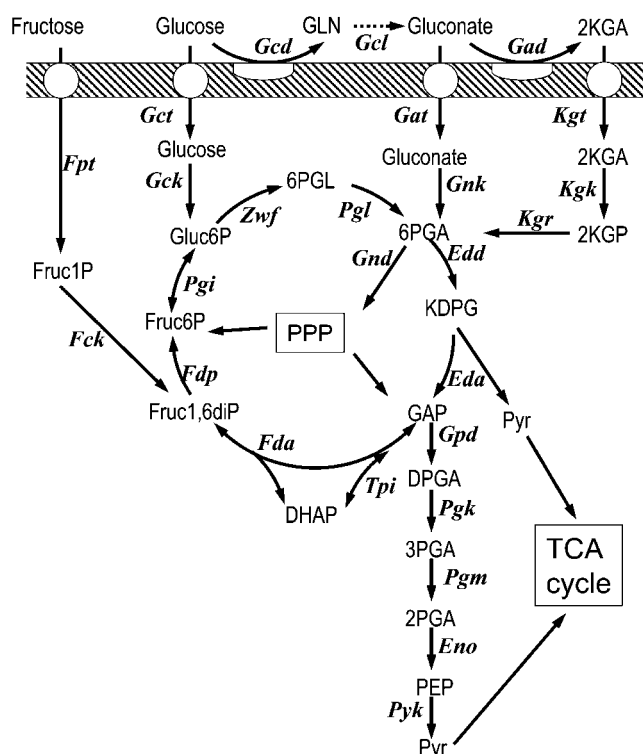


Fig. 2. Degradation of fructose, glucose and gluconate by the cyclic Entner–Doudoroff pathway in pseudomonads (adapted from Lessie & Phibbs, 1984). GLN, gluconolactone; 2KGA, 2-ketogluconate; 2KGP, 2-keto-6-phosphogluconate; 6PGA, 6-phosphogluconate; KDPG, 2-keto-3-deoxy-6-phosphogluconate; GAP, glyceraldehyde-3-phosphate; Pyr, pyruvate; DHAP, dihydroxyacetone phosphate; Fruc1,6diP, fructose 1,6-diphosphate; Fruc6P, fructose 6-phosphate; Fruc1P, fructose 1-phosphate; Gluc6P, glucose 6-phosphate; 6PGL, 6-phosphogluconolactone; DPGA, 1,3-diphosphoglycerate; 3PGA, 3-phosphoglycerate; 2PGA, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PPP, pentose phosphate pathway; TCA cycle, tricarboxylic acid cycle. *Gcd*, glucose dehydrogenase; *Gcl*, gluconolactonase; *Gad*, gluconate dehydrogenase; *Fpt*, fructose phosphotransferase; *Gct*, glucose transporter; *Gat*, gluconate transporter; *Kgt*, 2KGA transporter; *Fck*, Fruc1P kinase; *Gck*, glucose kinase; *Gnk*, gluconate kinase; *Kgk*, 2KGA kinase; *Kgr*, 2KGP reductase; *Zwf*, Gluc-6P dehydrogenase; *Pgl*, 6-phosphogluconolactonase; *Gnd*, 6PGA dehydrogenase; *Edd*, 6PGA dehydratase; *Eda*, KDPG aldolase; *Pgi*, Gluc6P dehydrogenase; *Fdp*, fructose-1,6-diphosphatase; *Fda*, Fruc1,6diP aldolase; *Tpi*, triosephosphate isomerase; *Gpd*, GAP dehydrogenase; *Pgk*, 3PGA kinase; *Pgm*, phosphoglyceromutase; *Eno*, enolase; *Pyk*, Pyr kinase.

group and one of the other three groups were identified by selecting the original variables with the highest absolute values for the loadings (under an angle) on these newly defined discriminants. Table 2 lists the results of this biplot analysis.

For the glucose group, (i) the ferric siderophore receptor, (ii) the PP0897–PP0904 gene cluster, (iii) the fructose

transport (fructose-specific phosphotransferase)/phosphofructokinase, (iv) the gluconokinase/gluconate transporter, (v) the hypothetical protein PP3459–PP3462 and (vi) the 2-ketogluconate/2-ketogluconate kinase/epimerase/regulator gene clusters are present more than once amongst the variables that have the highest absolute loading under an angle in the direction of the glucose group. For the gluconate group, (i) the fructose transport (fructose-specific phosphotransferase)/phosphofructokinase, (ii) the PP0897–PP0904 gene cluster, (iii) the outer-membrane protein H1/transcriptional regulator PhoP/sensor protein PhoQ and (iv) the conserved hypothetical protein PP3797–PP3800 gene clusters are present more than once. For succinate, (i) the glycine betaine/carnitine/choline ABC transporter, (ii) the C₄-dicarboxylate transport protein, (iii) the two-component regulator PhoP/PhoQ and (iv) the cold-shock domain family protein gene clusters are present more than once. The glucose and gluconate group have the fructose transport (fructose-specific phosphotransferase)/phosphofructokinase gene cluster in common. Moreover, this cluster is also highly ranked amongst the succinate-specific transcripts (results not shown), indicating that this cluster is not so important for glucose, gluconate and succinate, but in contrast is important for fructose-grown cells (see also Fig. 2), the reference condition for this differential transcript profiling study. Moreover, the genes of this cluster have a negative loading under an angle, indicating a negative correlation with the glucose, gluconate or succinate group, while most of the other important genes for these three carbon sources have a positive loading under an angle (Table 2). A similar phenomenon is observed for the spots containing genes PP0897–PP0904 (Table 2).

Comparing PCDA analysis with the fold-difference and hierarchical clustering approaches

Currently, two approaches other than the above-described PCA and PCDA analysis are commonly used for identifying the important transcripts from microarray studies: the fold-difference approach and HCA. We also applied the fold-difference approach to rank the transcripts based on the ratio or, when the ratio was < 1, i.e. in the case of down-regulated genes, on 1/ratio. Again, inserts of the 30 clones belonging to the spots with the highest value for (1/ratio) were sequenced and the genes present on these inserts identified (see Table S1, available as supplementary data with the online version of this paper). By and large, completely different transcripts were identified compared to the PCDA approach: only 30 % of the spots were the same.

The datasets were also analysed by HCA (Fig. 3) and subsequently clusters of transcripts whose expression was specifically affected by (one of) the carbon sources were identified (yellow boxes in Fig. 3). All inserts of the spots in these clusters were sequenced and the genes present on these inserts identified (see supplementary Table S2). Most of the genes identified by HCA were also identified by the PCDA and/or the fold-difference approach (see also Table 3).

Table 2. Identity of the genes present on the 30 genome fragments present on the spots with the highest absolute loadings under an angle, after PCDA biplot analysis, in the direction from the centre of the fructose group to the centre of the glucose, gluconate and succinate groups, respectively

| Rank | Clone name | Loading under an angle | ORF no. | Identity of genes on clone |
|----------------|------------|------------------------|---------------|---|
| Glucose | | | | |
| 1 | Pp.008A7 | 2.80 | PP0865–PP0867 | RNA polymerase σ^{70} factor, putative; Transmembrane sensor, putative; FecA-like outer-membrane receptor |
| 2 | Pp.112B12 | 2.67 | PP3155 | Outer-membrane ferric siderophore receptor, putative |
| 3 | Pp.010H3 | −2.64 | PP0902–PP0904 | Sensor histidine kinase; Conserved hypothetical protein; InaA protein |
| 4 | Pp.112C2 | 2.59 | PP3154–PP3155 | Conserved hypothetical protein; Outer-membrane ferric siderophore receptor, putative |
| 5 | Pp.111G3 | −2.56 | PP0072 | Quinone oxidoreductase |
| 6 | Pp.017D9 | −2.44 | PP0793–PP0795 | Phosphotransferase system, fructose-specific EI/HPr/EIIA components; 1-Phosphofructokinase; Phosphotransferase system, fructose-specific IIBC component |
| 7 | Pp.052C12 | −2.38 | PP3992–PP3997 | Xanthine/uracil permease family protein; DsrE protein; Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Glycosyl transferase, putative |
| 8 | Pp.110A6 | 2.37 | PP0151–PP0154 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Acetyl-CoA hydrolase/transferase family protein |
| 9 | Pp.051G1 | −2.37 | PP1297–PP1298 | General amino acid ABC transporter, periplasmic binding protein; General amino acid ABC transporter, permease protein |
| 10 | Pp.032D9 | −2.33 | PP0897–PP0900 | Fumarate hydratase, class I; Hypothetical protein; Ferredoxin reductase, putative; PAP2 family protein |
| 11 | Pp.009G4 | 2.32 | PP0535 | Outer-membrane ferric siderophore receptor |
| 12 | Pp.054E8 | −2.32 | PP0795–PP0798 | Phosphotransferase system, fructose-specific IIBC component; Conserved hypothetical protein; Conserved hypothetical protein; GGDEF domain protein |
| 13 | Pp.011A1 | 2.31 | PP3612–PP3613 | TonB-dependent receptor, putative; L-Sorbose dehydrogenase |
| 14 | Pp.060C9 | −2.31 | PP0257–PP0259 | Formate dehydrogenase accessory protein FdhD; LysM domain protein; Hydrolase, haloacid dehalogenase-like family |
| 15 | Pp.047E11 | 2.31 | PP3159 | <i>benABC</i> operon transcriptional activator BenR |
| 16 | Pp.053G5 | 2.30 | PP3416–PP3418 | Gluconokinase; Gluconate transporter; Hypothetical protein |
| 17 | Pp.051A1 | −2.29 | PP3461–PP3462 | Hypothetical protein; Hypothetical protein |
| 18 | Pp.030B12 | 2.29 | PP3823–PP3827 | Cytochrome <i>c</i> -type protein; Hypothetical protein; Conserved hypothetical protein; Hypothetical protein; Conserved hypothetical protein |
| 19 | Pp.113C1 | 2.28 | PP0168 | Surface adhesion protein, putative |
| 20 | Pp.044B3 | 2.27 | PP3416–PP3417 | Gluconokinase; Gluconate transporter |
| 21 | Pp.028C8 | 2.25 | PP3415–PP3417 | Transcriptional regulator, LacI family; Gluconokinase; Gluconate transporter |
| 22 | Pp.001A4 | −2.24 | PP1725–PP1726 | Conserved hypothetical protein; ABC transporter, periplasmic binding protein |
| 23 | Pp.051G11 | −2.24 | PP1839–PP1842 | Hypothetical protein; Conserved hypothetical protein; Cytochrome <i>c</i> family protein; Glutamine amidotransferase, class I |
| 24 | Pp.012G2 | −2.23 | – | No hits |
| 25 | Pp.054C8 | −2.22 | PP3458–PP3459 | Long-chain-fatty-acid-CoA ligase, putative; Hypothetical protein |
| 26 | Pp.051E7 | 2.18 | PP5209–PP5212 | Conserved hypothetical protein; Alcohol dehydrogenase, zinc-containing; ChaC-related protein; Oxidoreductase, iron-sulfur-binding |
| 27 | Pp.113G5 | 2.17 | – | No hits |
| 28 | Pp.017F1 | 2.13 | PP4217–PP4219 | Outer-membrane ferripyoverdine receptor; Lipase/esterase family protein; Non-ribosomal siderophore peptide synthetase |

Table 2. cont.

| Rank | Clone name | Loading under an angle | ORF no. | Identity of genes on clone |
|------------------|------------|------------------------|---------------|---|
| 29 | Pp.038H2 | 2·13 | PP3378–PP3380 | 2-Ketogluconate kinase; Epimerase KguE, putative; Transcriptional regulator PtxS |
| 30 | Pp.028E3 | 2·12 | PP3376–PP3377 | 2-Ketogluconate 6-phosphate reductase; 2-Ketogluconate transporter, putative |
| Gluconate | | | | |
| 1 | Pp.017D9 | −3·90 | PP0793–PP0795 | Phosphotransferase system, fructose-specific EI/HPr/EIIA components; 1-Phosphofructokinase; Phosphotransferase system, fructose-specific IIBC component |
| 2 | Pp.010H3 | −3·71 | PP0902–PP0904 | Sensor histidine kinase; Conserved hypothetical protein; InaA protein |
| 3 | Pp.015G9 | −3·54 | – | No hits |
| 4 | Pp.052C12 | −3·51 | PP3992–PP3997 | Xanthine/uracil permease family protein; DsrE protein; Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Glycosyl transferase, putative |
| 5 | Pp.008F6 | 3·50 | PP1185–PP1187 | Outer-membrane protein H1; Transcriptional regulatory protein PhoP; Sensor protein PhoQ |
| 6 | Pp.029H8 | 3·49 | PP2146–PP2147 | Helicase, SNF2/RAD54 family; Conserved hypothetical protein |
| 7 | Pp.054E8 | −3·46 | PP0795–PP0798 | Phosphotransferase system, fructose-specific IIBC component; Conserved hypothetical protein; Conserved hypothetical protein; GGDEF domain protein |
| 8 | Pp.111G3 | −3·33 | PP0072 | Quinone oxidoreductase |
| 9 | Pp.051C4 | 3·32 | PP1185–PP1187 | Outer-membrane protein H1; Transcriptional regulatory protein PhoP; Sensor protein PhoQ |
| 10 | Pp.008A7 | 3·26 | PP0865–PP0867 | RNA polymerase σ^{70} factor, putative; Transmembrane sensor, putative; FecA-like outer-membrane receptor |
| 11 | Pp.051G11 | −3·25 | PP1839–PP1842 | Hypothetical protein; Conserved hypothetical protein; Cytochrome <i>c</i> family protein; Glutamine amidotransferase, class I |
| 12 | Pp.045D9 | −3·16 | – | No hits |
| 13 | Pp.032D9 | −3·15 | PP0897–PP0900 | Fumarate hydratase, class I; Hypothetical protein; Ferredoxin reductase, putative; PAP2 family protein |
| 14 | Pp.045G3 | −3·13 | PP1872–PP1875 | Aminotransferase, class I; PilB-related protein; Glutathione peroxidase; Sensor histidine kinase/response regulator |
| 15 | Pp.051G1 | −3·09 | PP1297–PP1298 | General amino acid ABC transporter, periplasmic binding protein; General amino acid ABC transporter, permease protein |
| 16 | Pp.110A6 | 3·09 | PP0151–PP0154 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Acetyl-CoA hydrolase/transferase family protein |
| 17 | Pp.061D7 | 3·07 | PP3797–PP3799 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein |
| 18 | Pp.020C7 | −3·06 | PP0709–PP0712 | Transporter, NCS1 nucleoside transporter family; Conserved domain protein; Hydrolase, isochorismatase family; Conserved hypothetical protein |
| 19 | Pp.112B12 | 3·04 | PP3155 | Outer-membrane ferric siderophore receptor, putative |
| 20 | Pp.006C4 | −3·03 | – | <i>P. putida</i> S12 isolate KL54C2 toluene-induced genomic sequence |
| 21 | Pp.028C2 | −3·02 | PP0071, – | Hypothetical protein; no hits |
| 22 | Pp.047E11 | −3·01 | PP3159, – | <i>benABC</i> operon transcriptional activator BenR; no hits |
| 23 | Pp.008C3 | 3·00 | – | No hits |
| 24 | Pp.044G5 | −2·98 | PP0790–PP0792 | Inner-membrane protein AmpE; Deoxyribonuclease, TatD family; Fructose transport system repressor FruR |
| 25 | Pp.060D9 | −2·97 | PP3461–PP3462 | Hypothetical protein; Hypothetical protein |
| 26 | Pp.046H4 | −2·96 | PP2664–PP2667 | Sensory box histidine kinase/response regulator; DNA-binding response regulator AgmR; Hypothetical protein; ABC efflux transporter, permease protein |

Table 2. cont.

| Rank | Clone name | Loading under an angle | ORF no. | Identity of genes on clone |
|------------------|------------|------------------------|---------------|--|
| 27 | Pp.031H6 | -2.91 | - | <i>Pseudomonas</i> sp. Y2 DNA phenylacetate-CoA-ligase; <i>P. fluorescens</i> hybrid histidine kinase homologue; <i>Pseudomonas</i> sp. Y2 DNA styrene sensor kinase |
| 28 | Pp.020E3 | 2.91 | PP3800-PP3802 | Conserved hypothetical protein; Cation ABC transporter, periplasmic cation-binding protein, putative; Cation ABC transporter, ATP-binding protein, putative |
| 29 | Pp.045D8 | -2.90 | PP0021-PP0024 | Hypothetical protein; Hypothetical protein; Hypothetical protein; Membrane protein, putative |
| 30 | Pp.023A1 | -2.90 | PP0553-PP0555 | Acetoin dehydrogenase, dihydrolipoamide acetyltransferase component; Acetoin dehydrogenase, β subunit; Acetoin dehydrogenase, α subunit |
| Succinate | | | | |
| 1 | Pp.061A5 | 2.20 | PP0870-PP0872 | Glycine betaine/carnitine/choline ABC transporter, periplasmic binding protein, putative; Glycine betaine/carnitine/choline ABC transporter, permease protein, putative; Peptide chain release factor 3 |
| 2 | Pp.038F4 | 2.11 | PP1188-PP1191 | C ₄ -dicarboxylate transport protein; Conserved hypothetical protein; Conserved hypothetical protein; S4 domain protein |
| 3 | Pp.031H2 | 2.10 | PP1750-PP1751 | Aparagine synthetase; D-amino acid oxidase family protein |
| 4 | Pp.008F6 | 2.03 | PP1185-PP1187 | Outer-membrane protein H1; Transcriptional regulatory protein PhoP; Sensor protein PhoQ |
| 5 | Pp.045D3 | 2.00 | PP1187-PP1188 | Sensor protein PhoQ; C ₄ -dicarboxylate transport protein |
| 6 | Pp.029H8 | 1.98 | PP2146-PP2147 | Helicase, SNF2/RAD54 family; Conserved hypothetical protein |
| 7 | Pp.111F10 | 1.98 | PP0869-PP0871 | Glycine betaine/carnitine/choline ABC transporter, permease protein, putative; Glycine betaine/carnitine/choline ABC transporter, periplasmic binding protein, putative; Glycine betaine/carnitine/choline ABC transporter, permease protein, putative |
| 8 | Pp.032F2 | 1.97 | PP2837-PP2838 | Major facilitator family transporter; Conserved hypothetical protein |
| 9 | Pp.059C7 | -1.96 | PP0982-PP0985 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Cold-shock domain family protein |
| 10 | Pp.110E1 | 1.95 | - | <i>P. putida</i> solvent transporter gene:inner-membrane transporter protein SrpB; <i>P. putida</i> solvent transporter gene:outer-membrane channel protein SrpC |
| 11 | Pp.015G9 | -1.94 | - | No hits |
| 12 | Pp.045B3 | 1.93 | PP1187-PP1188 | Sensor protein PhoQ; C ₄ -dicarboxylate transport protein |
| 13 | Pp.017D9 | -1.92 | PP0793-PP0795 | Phosphotransferase system, fructose-specific EI/HPtr/EIIA components; 1-Phosphofructokinase; Phosphotransferase system, fructose-specific IIBC component |
| 14 | Pp.045D9 | -1.92 | - | No hits |
| 15 | Pp.046H3 | 1.89 | PP1187-PP1188 | Sensor protein PhoQ; C ₄ -dicarboxylate transport protein |
| 16 | Pp.011A2 | -1.88 | PP0801-PP0803 | Hypothetical protein; Chemotaxis protein CheV; Protein secretion ABC efflux system, membrane fusion protein |
| 17 | Pp.019C7 | -1.87 | PP0983-PP0985 | Conserved hypothetical protein; Conserved hypothetical protein; Cold-shock domain family protein |
| 18 | Pp.052G1 | 1.86 | PP1418-PP1419 | Tricarboxylate transport protein TctC, putative; Porin, putative |
| 19 | Pp.060B4 | -1.86 | PP0982-PP0985 | Conserved hypothetical protein; Conserved hypothetical protein; Conserved hypothetical protein; Cold-shock domain family protein |
| 20 | Pp.051C4 | 1.85 | PP1185-PP1187 | Outer-membrane protein H1; Transcriptional regulatory protein PhoP; Sensor protein PhoQ |
| 21 | Pp.109F10 | 1.84 | - | No hits |
| 22 | Pp.008B7 | -1.83 | PP0985-PP0987 | Cold-shock domain family protein; Glycine cleavage system T protein; L-serine dehydratase, iron-sulfur-dependent, single-chain form |

Table 2. cont.

| Rank | Clone name | Loading under an angle | ORF no. | Identity of genes on clone |
|------|------------|------------------------|---------------|--|
| 23 | Pp.032F3 | −1.82 | – | No hits |
| 24 | Pp.112D3 | −1.80 | PP2855–PP2856 | Conserved hypothetical protein; Conserved hypothetical protein |
| 25 | Pp.039C10 | 1.79 | PP2906–PP2907 | Sensor histidine kinase; DNA-binding response regulator |
| 26 | Pp.020H3 | −1.79 | PP1007–PP1009 | Transmembrane sensor, putative; RNA polymerase σ^{70} factor, ECF subfamily; Glyceraldehyde-3-phosphate dehydrogenase |
| 27 | Pp.002B4 | 1.79 | PP2375 | 5-Methyltetrahydrofolate–homocysteine methyltransferase |
| 28 | Pp.052F9 | −1.73 | PP4567–PP4570 | Hypothetical protein; Membrane protein, putative; Hypothetical protein; Conserved hypothetical protein |
| 29 | Pp.054G2 | 1.73 | PP2049–PP2052 | Alcohol dehydrogenase, iron-containing; Conserved hypothetical protein; Acetyl-CoA acetyltransferase; Hydrolase, haloacid dehalogenase-like family |
| 30 | Pp.051C5 | 1.72 | PP5328–PP5329 | Phosphate ABC transporter, permease protein; Phosphate ABC transporter, periplasmic phosphate-binding protein |

However, several gene clusters, encoding proteins such as surface adhesion protein, formate dehydrogenase and flagellar proteins, were not identified in one of the other two approaches. Remarkable is the fact that also the gene clusters encoding proteins such as 6-phosphogluconate dehydratase/glucokinase (PP1010–PP1011 – downregulated in S), glucose-6-phosphate dehydrogenase/6-phosphogluconolactonase/2-dehydro-3-deoxyphosphogluconate aldolase (PP1022–PP1024 – slightly upregulated in G and N, downregulated in S) and cytochrome *o* ubiquinol oxidase (PP0812–PP0814 – upregulated in G, N and S), which are of key importance in the degradation of (one of the) carbon sources studied (Fig. 2), are clearly visible as clusters in the HCA plot (Fig. 3), but do not end up in the top 30 by either PCDA or ratio analysis.

The effectiveness of the different data analysis tools for identifying specific groups of genes relevant for processes that proved to be important for growth on (one of the) different carbon sources used (see Discussion) is shown in Table 3. It can clearly be seen that the genes involved in specific processes of importance for growth on (one of the) different carbon sources are not always identified amongst the 30 most important spots by the different data analysis methods tested in this study. In contrast, the different data analysis approaches seem to be complementary.

DISCUSSION

The true challenge in functional genomics is the translation of the avalanche of data generated by these analytical tools into information. So far, no gold standard for analysis of microarray data to accomplish this goal has emerged. Currently, many sophisticated data analysis tools for functional genomics data are being developed. This has led to the analysis of biological data tending to become the field for statisticians, where the development of new, ‘fancy’, data analysis tools seems to be more of an issue than their applicability for the extraction of relevant

biological information from functional genomics data. As the method used has a profound influence on the interpretation of the results (Quackenbush, 2001), the back-up of the results of data analysis by biological studies is becoming a critical issue. There is, therefore, a need for comparing, preferably readily available and straightforward, data analysis tools in order to evaluate which tool is ‘the best’ for obtaining biologically relevant information (Carpentier *et al.*, 2004). In this paper we compare four different data analysis approaches for selecting and ranking transcripts (genes) important for growth of *P. putida* on four carbon sources – fructose, glucose, gluconate and succinate – for which the degradation pathways have been well established in this micro-organism (Lessie & Phibbs, 1984; Temple *et al.*, 1998).

In order to avoid bias during the data analysis, caused by a perceived immediate understanding of the importance of a specific transcript (or transcripts) being detected as relevant, and in that way directing the data analysis process, an anonymous clone-based array was used. Genome fragments of *P. putida* were cloned, and the inserts were amplified and spotted on glass slides. Only when spots were identified to be relevant by one of the data analysis methods were the inserts sequenced, and the identity of the transcripts unravelled. In many instances genes or operons of relevance proved to be present on multiple clones, indicating that clone-based arrays are a reliable means of identifying genes that are relevant for a specific biological process. This generic approach seems therefore very suitable for studying the comprehensive transcript response of micro-organisms whose full genome sequence is not available.

This study clearly shows that the data analysis method chosen has a profound effect on the transcripts identified as being the ‘most’ relevant; large differences were observed between the transcripts that ranked the highest based on the PCA approach (results not shown), the PCDA approach or the fold-difference approach. The most frequently used

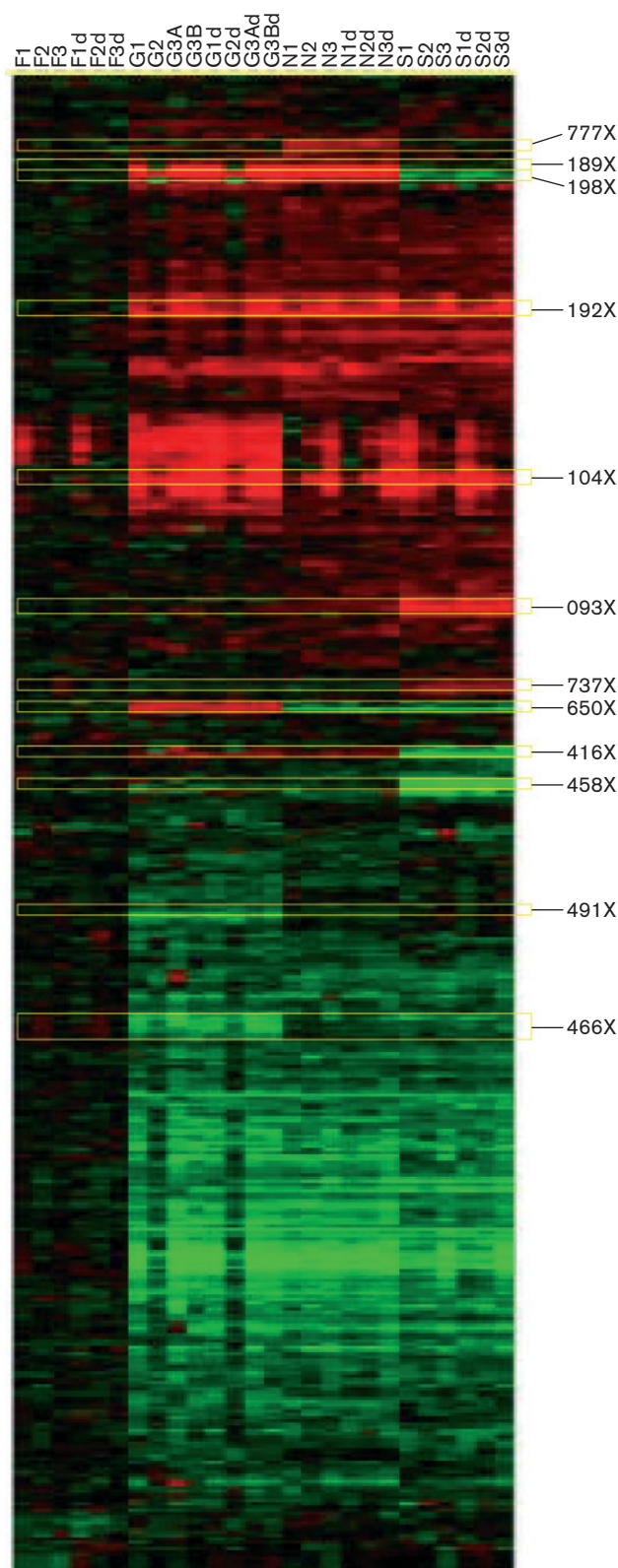


Fig. 3. Hierarchical cluster of the transcription profiles of *P. putida* S12 grown on the four different carbon sources. Red indicates increased expression, green decreased expression and black unchanged expression.

method for selecting data from transcriptomics experiments, the fold-difference or ratio approach, has the disadvantage that genes that inherently show a low response, like constitutively expressed genes, are overlooked (van der Werf, 2005; Wu, 2001; Slonim, 2002). In this study, this is clearly illustrated by the 6-phosphogluconate dehydratase/glucokinase (PP1010–PP1011) and glucose-6-phosphate dehydrogenase/6-phosphogluconolactonase/2-dehydro-3-deoxyphosphogluconate aldolase (PP1022–PP1024) operons that are only slightly up- or downregulated (at most a factor of 2; Table 3) compared to the responses of the 2-ketogluconate 6-phosphate reductase/2-ketogluconate transporter/2-ketogluconate kinase/epimerase KguE/transcriptional regulator PtxS (PP3376–PP3380) and the fructose degradation (PP0792–PP0795) gene clusters that are up- and downregulated by a factor of 7 to 24 (Table 3). The other frequently used tool for analysing transcriptome data, HCA, has the disadvantage that only gene clusters that show a specific expression profile are identified as relevant. For instance, in this study, the fructose degradation (PP0792–PP0795) gene cluster, which is strongly downregulated in cells grown on glucose, gluconate and succinate, was not identified by HCA as it ended up in the large bulk of downregulated genes (lower third of the hierarchical cluster; Fig. 3), and was therefore not specifically identified.

So far, there have only been a few isolated studies in which PC(D)A biplots, i.e. making use of the loadings under an angle, have been applied for analysing transcriptome data (Chapman *et al.*, 2001). PCDA is particularly well suited for the analysis of functional genomics datasets derived from samples originating from more than two different biological groups. This study clearly demonstrates that the loadings under an angle resulting from PCDA analysis are an appropriate quantitative statistical parameter with which relevant transcripts for a specific phenotype can be ranked, as illustrated by the fact that many genes encoding enzymes known to be involved in the degradation of the carbon sources studied [Fig. 2 – i.e. the fructose degradation operon, the gluconokinase/gluconate transporter (PP3415–PP3417) gene cluster, the 2-ketogluconate 6-phosphate reductase/2-ketogluconate transporter/2-ketogluconate kinase/epimerase KguE/transcriptional regulator PtxS (PP3376–PP3380) gene cluster, and the C₄-dicarboxylate transporter (PP1188)], were identified in this way. Moreover, in many instances, a transcript identified to be relevant for a specific carbon source by PCDA analysis was present on genome fragments of several other spots that were the most important, again indicating that this is relevant information, and that it is not chance correlations that identified these genes as being relevant. In this respect, the many regulatory genes on the inserts of spots identified by PCDA to strongly correlate with one of the different carbon sources are of special interest.

However, PCDA did not identify all the genes that are involved in the degradation of the different carbon sources; the complete set of genes involved in the degradation of the

Table 3. Overview of the data analysis methods by which specific groups of genes relevant for processes that proved to be important for growth on (one of the) different carbon sources were identified

The letters G, N and S in parentheses refer to the data analysis of the carbon source (see legend Fig. 1) with which these transcripts were identified. The ratios are the mean fold difference in expression of the different inserts on which the gene of interest was present.

| HCA | Ratio | PCDA | Ratio G | Ratio N | Ratio S | ORF no. | Identity of gene |
|---|-------------|-------------|------------|------------|------------|------------|--|
| Carbon metabolism | | | | | | | |
| | | + (N) | 0.59 | 0.63 | 0.64 | PP0792 | Fructose transport system repressor FruR |
| | + (N) | + (G, N, S) | 0.13 | 0.10 | 0.13 | PP0793 | Phosphotransferase system, fructose-specific EI/HPr/EIIA components |
| | + (N) | + (G, N, S) | 0.13 | 0.10 | 0.13 | PP0794 | 1-Phosphofructokinase |
| | + (N) | + (G, N, S) | 0.15 | 0.15 | 0.16 | PP0795 | Phosphotransferase system, fructose-specific IIBC component |
| + | | | 0.78 | 1.05 | 0.55 | PP1010 | 6-Phosphogluconate dehydratase |
| + | | | 0.71 | 0.96 | 0.49 | PP1011 | Glucokinase |
| + | | | 0.70 | 0.91 | 0.60 | PP1012 | DNA-binding response regulator GltR |
| + | + (G) | | 14.4 | 0.38 | 0.36 | PP1015 | Sugar ABC transporter, periplasmic sugar-binding protein |
| + | + (G) | | 11.5 | 0.21 | 0.57 | PP1016 | Sugar ABC transporter, permease protein |
| + | + (G) | | 12.1 | 0.61 | 0.58 | PP1017 | Sugar ABC transporter, permease protein |
| + | + (G) | | 13.4 | 0.56 | 0.54 | PP1018 | Sugar ABC transporter, ATP-binding subunit |
| + | + (G) | | 13.4 | 0.56 | 0.54 | PP1019 | Porin B |
| + | | | 1.87 | 1.00 | 0.45 | PP1021 | Transcriptional regulator HexR |
| + | | | 1.25 | 1.00 | 0.45 | PP1022 | Glucose-6-phosphate 1-dehydrogenase |
| + | | | 2.04 | 2.15 | 0.32 | PP1023 | 6-Phosphogluconolactonase |
| + | | | 2.32 | 2.44 | 0.24 | PP1024 | 2-Dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase |
| | + (G, N) | + (G) | 8.76 | 13.5 | 0.90 | PP3376 | 2-Ketogluconate-6-phosphate reductase |
| | + (G, N) | + (G) | 11.7 | 6.73 | 0.84 | PP3377 | 2-Ketogluconate transporter, putative |
| | + (G, N) | + (G) | 14.0 | 23.6 | 0.80 | PP3378 | 2-Ketogluconate kinase |
| | + (G, N) | + (G) | 14.0 | 23.6 | 0.80 | PP3379 | Epimerase KguE, putative |
| | + (G, N) | + (G) | 14.0 | 23.6 | 0.80 | PP3380 | Transcriptional regulator PtxS |
| + | + (G, N) | + (G) | 7.15 | 4.87 | 0.66 | PP3415 | Transcriptional regulator, LacI family |
| + | + (G, N) | + (G) | 14.1 | 8.94 | 0.62 | PP3416 | Gluconokinase |
| + | + (G, N) | + (G) | 17.6 | 11.0 | 0.56 | PP3417 | Gluconate transporter |
| + | + (S) | + (S) | 1.63 | 1.85 | 13.2 | PP1188 | C ₄ -dicarboxylate transport protein |
| Electron transport | | | | | | | |
| | | + (G, N) | 0.62 | 0.64 | 0.73 | PP0072 | Quinone oxidoreductase |
| + | | | 2.50 | 2.78 | 3.42 | PP0812 | Cytochrome <i>o</i> ubiquinol oxidase, subunit II |
| + | | | 2.44 | 2.69 | 2.93 | PP0813 | Cytochrome <i>o</i> ubiquinol oxidase, subunit I |
| + | | | 2.35 | 2.57 | 2.28 | PP0814 | Cytochrome <i>o</i> ubiquinol oxidase, subunit III |
| + | | | 2.19 | 2.30 | 2.12 | PP0815 | Cytochrome <i>o</i> ubiquinol oxidase, protein CyoD |
| | | + (G, N) | 0.49 | 0.38 | 0.52 | PP1841* | Cytochrome <i>c</i> family protein |
| + | | | 0.37 | 0.67 | 0.58 | PP2010 | Cytochrome <i>b</i> ₅₆₁ |
| + | | | 2.95 | 3.69 | 4.34 | PP3606 | Quinone oxidoreductase |
| | + (N) | + (G) | 0.16 | 0.07 | 0.31 | PP3823 | Cytochrome <i>c</i> -type protein |
| Iron limitation responsive genes | | | | | | | |
| | + (G, S) | | 22.1 | 4.54 | 9.17 | PP0943* | FagA protein |
| | + (G, N, S) | + (G) | 26.1 | 6.26 | 10.2 | PP3612 | TonB-dependent receptor, putative |
| + | | | 2.63 | 0.93 | 0.83 | PP0267 | Outer-membrane ferric siderophore receptor, putative |
| | | + (G) | 4.17 | 1.66 | 1.91 | PP0535 | Outer-membrane ferric siderophore receptor |
| | + (G, N, S) | | 51.9 | 7.54 | 6.39 | PP0861 | Outer-membrane ferric siderophore receptor |
| | + (G, S) | + (G, N) | 27.4 | 10.4 | 15.1 | PP3155 | Outer-membrane ferric siderophore receptor, putative |

Table 3. cont.

| HCA | Ratio | PCDA | Ratio G | Ratio N | Ratio S | ORF no. | Identity of gene |
|--------------------|-------------|----------|------------|------------|------------|------------|---|
| Regulators/sensors | + (G, N, S) | + (G) | 42.2 | 15.2 | 15.6 | PP4217 | Outer-membrane ferripyoverdine receptor |
| | | + (G, N) | 2.63 | 1.84 | 1.70 | PP0867 | FecA-like outer-membrane receptor |
| | | + (N) | 0.46 | 0.45 | 0.53 | PP2664 | Sensory box histidine kinase/response regulator |
| | | + (N) | 0.46 | 0.45 | 0.53 | PP2665 | DNA-binding response regulator AgmR |
| | | + (S) | 1.32 | 2.24 | 2.82 | PP2906 | Sensor histidine kinase |
| | | + (S) | 1.32 | 2.24 | 2.82 | PP2907 | DNA-binding response regulator |
| | | + (N) | 0.68 | 0.73 | 0.68 | PP1875 | Sensor histidine kinase/response regulator |
| | | | 1.18 | 0.33 | 5.83 | PP5339 | Transcriptional regulator, AraC family |
| | | + (G, N) | 0.50 | 0.61 | 0.63 | PP3159 | <i>benABC</i> operon transcriptional activator BenR |
| | | | | | | | |

carbon sources studied (Fig. 2) was only obtained by combining the results of PCDA analysis, the fold-difference and the HCA approach. In this respect, the complementary nature of the three approaches (HCA, fold-difference and PCDA; Table 3) is very notable.

Besides the complementary nature of the different data analysis tools studied, this paper also demonstrates the strength of the clone-based array approach for the identification of relevant transcripts. It resulted in the identification of all except two of the genes involved in the degradation of the different carbohydrates studied (Fig. 2; Lessie & Phibbs, 1984; Temple *et al.*, 1998): the fructose utilization gene cluster (PP0792–PP0795), the glucose utilization gene cluster (PP1010–PP1012; Sage *et al.*, 1996), the *zwf-pgl-eda* gene cluster (PP1021–PP1024; Petruschka *et al.*, 2002; Hager *et al.*, 2000), the 2-ketogluconate utilization gene cluster (PP3376–PP3380; Swanson *et al.*, 2000) and the gluconate utilization gene cluster (PP3415–PP3417). Only the genes encoding glucose dehydrogenase and gluconate dehydrogenase were not positively identified. Unfortunately, the genes encoding these enzymes have so far not been isolated from a *Pseudomonas* species. Although these genes were annotated in the *P. putida* KT2240 genome (PP1444 and PP3383, respectively) a BLAST study showed no significant homology between these genes and any of the functionally characterized glucose and gluconate dehydrogenase genes. Therefore, it is possible that one of the many genes encoding hypothetical proteins identified in this study encodes one of these two enzymes. Also a sugar ABC transporter gene cluster (PP1015–PP1019; Wylie & Worobec, 1994, 1995) was identified that was specifically induced upon growth on glucose (Table 3). This gene cluster encodes a glucose porin and a sugar transporter, of which the sugar-binding protein (PP1015) is very likely the previously purified glucose-specific glucose binding protein (Stinson *et al.*, 1977), as this gene encodes a protein of a similar size as the purified protein (44.5 kDa) and has a similar amino acid composition.

The clone-based-array approach, in combination with the different data analysis tools, not only resulted in the

identification of genes encoding the enzymes known to be involved in the degradation of the carbon sources studied, but also gave new insights into the physiology of the degradation of the carbon sources studied. Most remarkable was an upregulation of a large number of genes that respond to iron limitation in glucose-, gluconate- or succinate-grown cells in comparison with fructose-grown cells (Table 3). This includes six different iron chelate receptors – the siderophore receptors (PP0267, PP0535, PP0861, PP3155 and PP4217) and the ferric citrate receptor FecA (PP0867; Enz *et al.*, 2003) – TonB, which is involved in the translocation of the iron chelate bound to the siderophore and ferric citrate receptors across the outer membrane (PP3612; Moeck & Coulton, 1998), and the iron-responsive transcript FagA (PP0943; Hassett *et al.*, 1997). This coincides with an upregulation of two RNA polymerase σ^{70} factors of the ECF subfamily (PP0162 and PP0704) that are involved in the regulation of siderophore biosynthesis (Redly & Poole, 2003). All these transcripts are under control of the Fur repressor protein. The Fur repressor (PP4730) was not amongst the 200 clones that were sequenced in this study.

Moreover, also several cytochrome- or quinone-associated genes were specifically upregulated (i.e. PP0812, PP0813, PP0814, PP0815, PP3606) or downregulated (i.e. PP0071, PP1841, PP2010 and PP3823). This indicates that the composition of the electron-transport chain in *P. putida* S12 is different in fructose-grown cells compared to glucose-, gluconate- or succinate-grown cells. This is in agreement with the distinct degradation pathway for fructose compared to that of glucose and gluconate in pseudomonads (Fig. 2; Lessie & Phibbs, 1984; Temple *et al.*, 1998). Potentially, the different composition of the electron-transport chain reflects its greater importance in cells grown on glucose or gluconate, which are initially degraded extracellularly in a sequence from glucose to gluconate and subsequently 2-ketogluconate by a PQQ-dependent glucose dehydrogenase (Matsushita & Ameyama, 1982) and a cytochrome-containing gluconate dehydrogenase (Matsushita *et al.*, 1979), respectively. Both glucose and gluconate dehydrogenase are directly linked to the electron-transport chain. The observed iron limitation when cells

were grown on glucose or gluconate suggests that there is a larger demand for iron as the prosthetic groups in proteins, such as cytochrome-containing enzymes, when *P. putida* S12 is cultivated on either of these carbon sources. Gluconate dehydrogenase is known to contain dihaem cytochrome *c* as a prosthetic group (Matsushita *et al.*, 1979). Also succinate dehydrogenase is directly linked to the electron-transport chain and is a cytochrome-containing enzyme (<http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=gpp>).

The multivariate data analysis tools that are currently used in functional genomics research originate from other research fields. Although specific adaptations have been made to these tools in order to optimize them for biological purposes (Eisen *et al.*, 1998; Heyer *et al.*, 1999; Tamayo *et al.*, 1999; Tavazoie *et al.*, 1999), this paper demonstrates an important role for multiple complementary approaches. In the near future further improvement of multivariate data analysis methods for analysing functional genomics datasets is to be expected using mathematical considerations that are based on a molecular biological rationale. Further improvements are also expected to overcome the problem of having far more variables than samples available for statistical analysis. This can lead to both false positives and false negatives with the existing multivariate data analysis tools when applied to functional genomics datasets.

In conclusion, this paper clearly demonstrates that the data analysis method used has a large effect on the ranking of the transcripts that are relevant for a specific phenotype. The methods used in this study were complementary: only when the results of the transcripts that were ranked the highest were combined did a complete picture of the processes important for the catabolism of the different carbon sources studied become apparent. For the more subtle, regulatory processes in a cell, especially the multivariate data analysis tool PCDA seemed to be very effective, as relatively more regulator genes were identified by this method. Moreover, this study showed that anonymous cloned-based arrays provide a reliable means of identifying relevant genes from micro-organisms whose full genome sequence is not available.

ACKNOWLEDGEMENTS

We thank Roelie Bijl, Evelyn Wesseling, Martien Caspers, Mieke Havekes and Alie Kat Angelino for technical assistance, Ted van der Lende for fruitful discussions, Carina Rubingh and Sabina Bijlsma for critically reading the manuscript, and Baseclear (Leiden, The Netherlands) for constructing the chromosomal library of *P. putida* S12 and sequence analysis.

REFERENCES

Akopyants, N. S., Clifton, S. W., Martin, J., Pape, D., Wylie, T., Li, L., Kissing, J. C., Roos, D. S. & Beverley, S. M. (2001). A survey of the *Leishmania major* Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling. *Mol Biochem Parasitol* **113**, 337–340.

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Carpentier, A.-S., Riva, A., Tisseur, P., Didier, G. & Henaut, A. (2004). The operons, a criterion to compare the reliability of transcriptome analysis tools: ICA is more reliable than ANOVA, PLS and PCA. *Comput Biol Chem* **28**, 3–10.

Chapman, S., Schenk, P., Kazan, K. & Manners, J. (2001). Using biplots to interpret gene expression patterns in plants. *Bioinformatics* **18**, 202–204.

Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863–14868.

Enz, S., Brand, H., Orellana, C., Mahren, S. & Braun, V. (2003). Sites of interaction between the FecA and FecR signal transduction proteins of ferric citrate transport in *Escherichia coli* K-12. *J Bacteriol* **185**, 3745–3752.

Hager, P. W., Calfee, M. W. & Phibbs, P. V. (2000). The *Pseudomonas aeruginosa* devB/SOL homolog, *pgl*, is a member of the *hex* regulon and encodes 6-phosphogluconolactonase. *J Bacteriol* **182**, 3934–3941.

Hartmans, S., van der Werf, M. J. & de Bont, J. A. M. (1990). Bacterial degradation of styrene involving a novel flavin adenine dinucleotide-dependent styrene monooxygenase. *Appl Environ Microbiol* **56**, 1347–1351.

Hassett, D. J., Howell, M. L., Ochsner, U. A., Vasil, M. L., Johnson, Z. & Dean, G. E. (1997). An operon containing *fumC* and *sodA* encoding fumarate C and manganese superoxide dismutase is controlled by the ferric uptake regulator in *Pseudomonas aeruginosa*: *fur* mutants produce elevated alginate levels. *J Bacteriol* **179**, 1452–1459.

Heyer, L. J., Kruglyak, S. & Yooseph, S. (1999). Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* **9**, 1106–1115.

Hoogerbrugge, R., Willig, S. J. & Kistemaker, P. G. (1983). Discriminant analysis by double stage principal component analysis. *Anal Chem* **55**, 1710–1712.

Lessie, T. G. & Phibbs, P. V. (1984). Alternative pathways of carbohydrate utilization in *Pseudomonas*. *Annu Rev Microbiol* **38**, 359–387.

Matsushita, K. & Ameyama, M. (1982). D-Glucose dehydrogenase from *Pseudomonas fluorescens*, membrane-bound. *Methods Enzymol* **89**, 149–154.

Matsushita, K., Shinagawa, E., Adachi, O. & Ameyama, M. (1979). Membrane-bound D-gluconate dehydrogenase from *Pseudomonas aeruginosa*. *J Biochem* **85**, 1173–1181.

Michaud, D. J., Marsh, A. G. & Dhurjati, P. S. (2003). eXPatGen: generating dynamic expression patterns for the systematic evaluation of analytical methods. *Bioinformatics* **19**, 1140–1146.

Moeck, G. S. & Coulton, J. W. (1998). TonB-dependent iron acquisition: mechanisms of siderophore-mediated active transport. *Mol Microbiol* **28**, 675–681.

Nelson, K. E., Weinel, C., Paulsen, I. T. & 40 other authors (2002). Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**, 799–808.

Orr, M. S. & Scherf, U. (2002). Large-scale gene expression analysis in molecular target discovery. *Leukemia* **16**, 473–477.

Petruschka, L., Adlf, K., Burchardt, G., Denedde, J., Jurgensen, J. & Hermann, H. (2002). Analysis of the *zwf-pgl-eda* operon in *Pseudomonas putida* strains H and KT2440. *FEMS Microbiol Lett* **215**, 89–95.

Pieterse, B., Jellema, R. H. & van der Werf, M. J. (2005). Quenching of microbial samples for increased reliability of microarray data. *J Microbiol Methods* doi:10.1016/j.mimet.2005.04.035

- Quackenbush, J. (2001).** Computational analysis of microarray data. *Nat Rev Genet* **2**, 418–427.
- Redly, G. A. & Poole, K. (2003).** Pyoverdine-mediated regulation of FpvA synthesis in *Pseudomonas aeruginosa*: involvement of a probable extracytoplasmic-function sigma factor, FpvI. *J Bacteriol* **185**, 1261–1265.
- Sage, A. E., Proctor, W. D. & Phibbs, P. V. (1996).** A two-component response regulator, *gltR*, is required for glucose transport activity in *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **178**, 6064–6066.
- Slonim, D. K. (2002).** From patterns to pathways: gene expression data analysis comes of age. *Nat Genet Suppl* **32**, S502–S508.
- Stinson, M. W., Cohen, M. A. & Merrick, J. M. (1977).** Purification and properties of the periplasmic glucose-binding protein of *Pseudomonas aeruginosa*. *J Bacteriol* **131**, 672–681.
- Swanson, B. L., Hager, P., Phibbs, P., Ochsner, U., Vasil, M. & Hamood, A. N. (2000).** Characterization of the 2-ketogluconate utilization operon in *Pseudomonas aeruginosa* PAO1. *Mol Microbiol* **37**, 561–573.
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S. & Golub, T. R. (1999).** Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A* **96**, 2907–2912.
- Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. & Church, G. M. (1999).** Systematic determination of genetic network architecture. *Nat Genet* **22**, 281–285.
- Tefferi, A., Bolander, M. E., Ansell, S. M., Wieben, E. D. & Spelsberg, T. C. (2002).** Primer on medical genomics. Part III: microarray experiments and data analysis. *Mayo Clin Proc* **77**, 927–940.
- Temple, L. M., Sage, A. E., Schweizer, H. P. & Phibbs, P. V. (1998).** Carbohydrate catabolism in *Pseudomonas aeruginosa*. In *Pseudomonas*, pp. 35–72. Edited by T. C. Montie. New York: Plenum.
- van der Werf, M. J. (2005).** Towards replacing closed with open target selection strategies. *Trends Biotechnol* **23**, 11–16.
- van der Werf, M. J., Jellema, R. H. & Hankemeier, T. (2005).** Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. *J Ind Microbiol Biotechnol* **32**, 234–252.
- Wan, C.-Y. & Wilkins, T. A. (1994).** A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Anal Biochem* **223**, 7–12.
- Wu, T. D. (2001).** Analysing gene expression data from DNA microarrays to identify candidate genes. *J Pathol* **195**, 53–65.
- Wylie, J. L. & Worobec, E. A. (1994).** Cloning and nucleotide sequence of the *Pseudomonas aeruginosa* glucose-selective OrpB porin gene and distribution of *oprB* within the family *Pseudomonaceae*. *Eur J Biochem* **220**, 505–512.
- Wylie, J. L. & Worobec, E. A. (1995).** The OprB porin plays a central role in carbohydrate uptake in *Pseudomonas aeruginosa*. *J Bacteriol* **177**, 3021–3026.