

SCIENTIFIC REPORTS



OPEN

Simplifying multidimensional fermentation dataset analysis and visualization: One step closer to capturing high-quality mutant strains

Received: 01 August 2016
Accepted: 28 November 2016
Published: 03 January 2017

Xiang Zhou^{1,*}, Dan Xu^{1,*} & Ting-Ting Jiang^{1,2,*}

In this study, we analyzed mutants of *Clostridium acetobutylicum*, an organism used in a broad range of industrial processes related to biofuel production, to facilitate future studies of bioreactor and bioprocess design and scale-up, which are very important research projects for industrial microbiology applications. To accomplish this, we generated 329 mutant strains and applied principal component analysis (PCA) to fermentation data gathered from these strains to identify a core set of independent features for comparison. By doing so, we were able to explain the differences in the mutant strains' fermentation expression states and simplify the analysis and visualization of the multidimensional datasets related to the strains. Our study has produced a high-efficiency PCA application based on a data analytics tool that is designed to visualize screening results and to support several hundred sets of data on fermentation interactions to assist researchers in more precisely screening and capturing high-quality mutant strains. More importantly, although this study focused on the use of PCA in microbial fermentation engineering, its results are broadly applicable.

Principal component analysis (PCA) is a statistical tool based on mathematical operations. It is widely used for high-dimensional data analysis, and it may be the most popular multivariate statistical technique used in almost all scientific disciplines. Its origin can be traced back to several mathematicians: Pearson¹, Cauchy², Jordan³, Cayley⁴, Hamilton⁵, and Boyer and Merzbach⁶. However, modern PCA was formalized by Hotelling⁷, who also suggested the 'principal component' element of its name, which was unprecedented in scientific terms. Within the fields of visualization and computer graphics alone, PCA has been used in applications including facial recognition^{8–10}, motion analysis and synthesis^{11–13}, clustering^{14–16}, and dimension reduction^{17–20}. PCA is a quantitatively rigorous method of data simplification²¹. Based on this method, a new set of variables called principal components²² is extracted, where each principal component is often a linear combination or a few or multiple of the original variables²³. All of the principal components have the characteristic of being orthogonal to each other; thus, they contain no redundant information²⁴. As a whole, the principal components of a set of data form an orthogonal basis for the data space^{25,26}.

Clostridium acetobutylicum (*C. acetobutylicum*) is an industrially valuable bacterium that is sometimes identified by the term "Weizmann Organism", named for the biochemist Chaim Weizmann. Weizmann's nationality was Israeli-British and he was born in Russia. As a senior lecturer at the University of Manchester, England, he used this bacterium in 1916 as a biochemical tool to produce acetone, ethanol, and butanol from starch. This method has since been described as the acetone-butanol-ethanol fermentation process (ABE process); it yields acetone, butanol, and ethanol at a ratio of 3:6:1. Acetone was used in the important wartime task of casting cordite during the first world war and the second world war, and the alcohols were used to produce vehicle fuels and synthetic rubber^{27,28}. Although ABE fermentation is one of the longest known large-scale biofermentation

¹Institute of Modern Physics, Chinese Academy of Sciences, 509 Nanchang Rd., Lanzhou, Gansu, 730000, P.R.China.

²University of Chinese Academy of Sciences, 19 A Yuquan Rd, Shijingshan District, Beijing, 100049, P.R.China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to X.Z. (email: syannovich@gmail.com)

processes, the depletion of fossil fuels has renewed interest in ABE fermentation^{29,30}. Between the 1890s and the 1990s, the number of research studies in this field increased considerably, with a focus on improving the overall process, including the development of alternative fermentation substrates, improved strains, improved cultivation techniques, and improved product-removal techniques^{31–33}. During the 1990s and 2000s, intensive basic research studies investigated the genetics of solvent-producing *Clostridium* sp. and strove to improve these strains through genetic manipulations^{34,35}. Rapid scale-up trials were also performed at the beginning of the 21st century to improve traditional ABE fermentation methods. However, the economic feasibility of biobutanol production via ABE fermentation suffers from product toxicity, relatively low product yields with respect to the production bacteria, multiple end products, production inhibitions, and inefficient product recovery from the produced alcohol mixtures^{36–38}.

Note that the productivity of metabolites can be improved by up to a factor of ten through suitable bacterium improvement techniques³⁹. Mutagenesis is one of the most reliable and widely used approaches for strain improvement⁴⁰. During organism breeding, mutations are induced using heavy-ion irradiation, ultraviolet (UV) rays, X-rays, γ -rays, lasers, neutrons, and thermophoresis⁴¹. Chemical physics methods based on methyl methane sulfonate (MMS), hydroxyl amine (HA), and N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) have been adopted as important research methods for inducing mutations⁴². These mutagenesis approaches tend to produce abundant mutant strains. However, because of the problems posed by the need to handle the complex multidimensional data that describe the fermentation expression states of these mutant strains, researchers have made substantial efforts to find ways to better screen and capture high-quality mutant strains. Fortunately, in these mutant strain expression datasets, which contain many variables, groups of variables often vary together. One reason for this behavior is that more than one variable might reflect the same driving principle governing the behavior of the system. In many systems, only a few such driving forces exist. However, abundant instrumentation enables the measurement of dozens of system variables. Therefore, researchers can take advantage of this redundancy of information and simplify the problem by replacing a group of variables with a single new variable.

In the previous scientific research literature dedicated to application of the approach in the analysis of datasets on fermentation processes. For example, the applicability of PCA for fermentation data analysis using eight fed-batch fermentations with a recombinant L-phenylalanine-producing *Escherichia coli* strain as a test system was investigated by Takors *et al.*⁴³; experimental determination by principal component analysis of a reaction pathway of biohydrogen production by anaerobic fermentation was determined by Aceves-Lara *et al.*⁴⁴; principal component analysis and partial least squares regression can be used to extract information from particle size distribution data and predict rheological properties was determined by Peterson *et al.*⁴⁵; principal component analysis the measurement profiles acquired during the monitoring of several fed-batch fermentations for the production of erythromycin was applied by Bicciato *et al.*⁴⁶. However, the number of PCA applications for fermentation dataset analysis and visualization are still rather limited, although this tool holds great promise, merit and interest.

A series of experiments involving the fermentation of mutant *C. acetobutylicum* have produced observations of the differential expressions of hundreds of mutant strains across multiple conditions. In this study, the application of PCA to these expression data enabled the direct comparison of a core set of independent features of the expression states of the 329 mutant strains that were investigated. Thus, we can explain the differences among the mutant strains' multidimensional fermentation datasets and move one step closer to capturing high-quality mutant strains.

Results and Discussion

Source of the mutants' fermentation multidimensional datasets. As shown in the Supplementary Information 1, the mutants all that were obtained, and they were from multiple rounds by the experiment of heavy-ion irradiation. 329 mutants were screened by MTT. It would be interesting to summarize how the mutants were generated, but this may be beyond the scope of this study since the focus of the manuscript is the PCA analysis. In addition, we are currently investigating and will dissect the top-performing mutants. All 329 mutants according to supplementary Information 1 (Materials & Methods for data procurement from mutant strains) conducted data acquisition. Table S1 (Supplementary Information 2) showed all multidimensional datasets that unlike otherwise noted, fermentation was carried out in serum bottles, this is the source of the datasets.

Visualization of multidimensional datasets. Table S1 (Supplementary Information 2) shows all of the effects of an increasing butyric acid concentration on the ABE fermentation yields from the substrate, the butanol productivity of fermentation for each of the 329 mutant strains supplemented with 5.0-g/L butyric acid, and the maximal specific growth rates of the 329 mutant strains. Hence, a dataset is obtained that consists of 329 mutant strains and 8 variables. The actual measurements can be arranged in a table or a matrix with dimensions of 329×8 . The variables are butanol productivity (g/L/h), butanol yield (g/g), solvent (ABE) yield (g/g), acetone yield (g/g), ethanol yield (g/g), maximal specific growth rate (added butyrate: 5.0 g/L; μ_{Max}^A), maximal specific growth rate (added butyrate: 6.5 g/L; μ_{Max}^B), and maximal specific growth rate (added butyrate: 8.5 g/L; μ_{Max}^C). With 329 mutant strains and 8 columns (variables), obtaining an overview of the various types of information available in this multidimensional dataset is difficult. A good starting point is to plot the individual variables for the 329 mutant strains. As shown in Fig. 1, there is more variability in the values of the butanol and solvent (ABE) yields than in the values of butanol productivity and μ_{Max}^C . Normally, one would consider plotting each of the original variables, but doing so would result in 32 variograms. Thus, the advantage of PCA is that it may reduce the number of variables that must be considered. Sometimes, the original data can be used to calculate the principal components, if the same unit is appropriate for each variable. However, when different variables in different columns have different units or when the variance in value among the columns is large, the data must be standardized to improve performance.

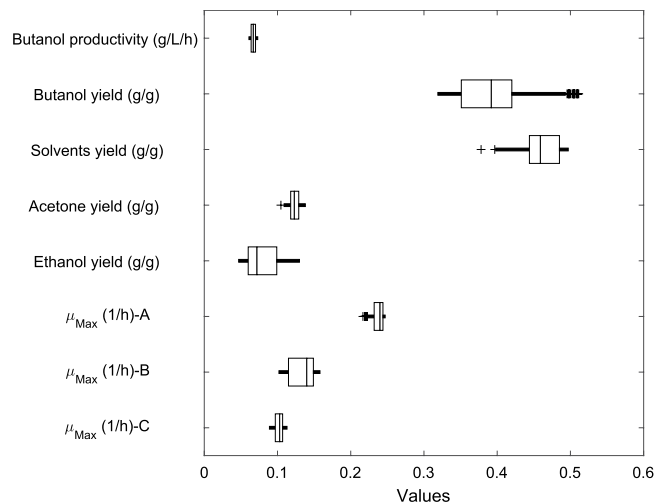


Figure 1. The distribution of the measured data matrix of 329 rows and 8 columns, corresponding to a dataset consisting of 329 mutant strains and 8 variables. These variables are butanol productivity (g/L/h), butanol yield (g/g), solvent (ABE) yield (g/g), acetone yield (g/g), ethanol yield (g/g), maximal specific growth rate (added butyrate: 5.0 g/L; μ_{Max} -A), maximal specific growth rate (added butyrate: 6.5 g/L; μ_{Max} -B), and maximal specific growth rate (added butyrate: 8.5 g/L; μ_{Max} -C).

| | Butanol productivity (g/L/h) | Butanol yield (g/g) | Solvents (ABE) yield (g/g) | Acetone yield (g/g) | Ethanol yield (g/g) | μ_{Max} -A (1/h) | μ_{Max} -B (1/h) | μ_{Max} -C (1/h) |
|------------------------------|------------------------------|---------------------|----------------------------|---------------------|---------------------|----------------------|----------------------|----------------------|
| Butanol productivity (g/L/h) | 1.000000 | 0.062521 | 0.016778 | -0.002446 | 0.047993 | 0.027159 | 0.009812 | 6.699713e-04 |
| Butanol yield (g/g) | 0.062521 | 1.000000 | 0.781528 | -0.386327 | 0.444585 | 0.643167 | -0.027757 | 0.409016 |
| Solvents (ABE) yield (g/g) | 0.016778 | 0.781528 | 1.000000 | -0.422182 | 0.513211 | 0.654756 | -0.053534 | 0.463406 |
| Acetone yield (g/g) | -0.002446 | -0.386327 | -0.422182 | 1.000000 | -0.308724 | -0.349048 | -0.040566 | -0.202345 |
| Ethanol yield (g/g) | 0.047993 | 0.444585 | 0.513211 | -0.308724 | 1.000000 | 0.588894 | 0.012975 | 0.245488 |
| μ_{Max} -A (1/h) | 0.027159 | 0.643167 | 0.654756 | -0.349048 | 0.588894 | 1.000000 | -0.084181 | 0.318853 |
| μ_{Max} -B (1/h) | 0.009812 | -0.027757 | -0.053534 | -0.040566 | 0.012975 | -0.084181 | 1.000000 | -0.029020 |
| μ_{Max} -C (1/h) | 6.699713e-04 | 0.409016 | 0.463406 | -0.202345 | 0.245488 | 0.318853 | -0.029020 | 1.000000 |

Table 1. In the correlation matrix, which has a size of 8×8 , the variables are highly correlated, with many having correlation values exceeding 0.45. *All experiments were repeated 3 times, and the average values are reported.

Analysis of multidimensional datasets. The first three principal component coefficient vectors are shown in Table 1. As mentioned above, when all variables have the same unit, it is appropriate to compute the principal components from the raw data. The correlation matrix, which has dimensions of 8×8 , reveals that the variables are highly correlated (Table 1).

When the variables are expressed in different units, or if the variance in value among the columns is substantial, as in this study, scaling or weighting the data is preferred. The correlations between select variables can be as large as 0.33. PCA serves to construct new independent variables from linear combinations of the originals. For this purpose, in PCA, the inverse variances of the measured values are used as weights. The coefficient vectors for the first three principal components, namely, butanol productivity (g/L/h), butanol yield (g/g) and solvent (ABE) yield (g/g), are shown in Table 2. It is seen that the first principal component makes the largest contributions to the first and fifth variables: butanol productivity and ethanol yield. The coefficients of this principal component are positive. The principal component variables are defined as linear combinations of the original variables. The extracted eigenvector table provides the coefficients for the following equations:

$$\begin{aligned}
 PC1 = & 0.0001 * \text{Butanol productivity (g/L/h)} + 0.0164 * \text{Butanol yield (g/g)} \\
 & + 0.0128 * \text{Solvent (ABE) yield (g/g)} - 0.0023 * \text{Acetone yield (g/g)} \\
 & + 0.0086 * \text{Ethanol yield (g/g)} + 0.0036 * \mu_{Max} - A(1/h) \\
 & - 0.0002 * \mu_{Max} - B(1/h) + 0.0021 * \mu_{Max} - C(1/h)
 \end{aligned} \quad (1)$$

$$\begin{aligned}
 PC2 = & 0.0016 * \text{Butanol productivity (g/L/h)} - 0.0000 * \text{Butanol yield (g/g)} \\
 & - 0.0012 * \text{Solvent (ABE) yield (g/g)} - 0.0014 * \text{Acetone yield (g/g)} \\
 & + 0.0023 * \text{Ethanol yield (g/g)} - 0.0005 * \mu_{Max} - A(1/h) + 0.0050 * \mu_{Max} \\
 & - B(1/h) - 0.0007 * \mu_{Max} - C(1/h)
 \end{aligned} \quad (2)$$

| | Coefficients of PC1 | Coefficients of PC2 | Coefficients of PC3 |
|------------------------------|---------------------|---------------------|---------------------|
| Butanol productivity (g/L/h) | 0.0001 | 0.0016 | 0.0034 |
| Butanol yield (g/g) | 0.0164 | 0.0000 | 0.0011 |
| Solvents (ABE) yield (g/g) | 0.0128 | -0.0012 | -0.0005 |
| Acetone yield (g/g) | -0.0023 | -0.0014 | 0.0013 |
| Ethanol yield (g/g) | 0.0086 | 0.0023 | 0.0005 |
| μ_{Max} -A (1/h) | 0.0036 | -0.0005 | 0.0003 |
| μ_{Max} -B (1/h) | -0.0002 | 0.0050 | -0.0023 |
| μ_{Max} -C (1/h) | 0.0021 | -0.0007 | -0.0004 |

Table 2. The principal component variables are defined as linear combinations of the original variables. The table of the extracted eigenvectors provides the coefficients for the equations.

$$\begin{aligned}
 PC3 = & 0.0037 * \text{Butanol productivity (g/L/h)} + 0.0011 * \text{Butanol yield (g/g)} \\
 & - 0.0005 * \text{Solvent (ABE) yield (g/g)} + 0.0013 * \text{Acetone yield (g/g)} \\
 & + 0.0005 * \text{Ethanol yield (g/g)} + 0.0003 * \mu_{Max} - A(1/h) - 0.0023 * \mu_{Max} \\
 & - B(1/h) - 0.0004 * \mu_{Max} - C(1/h)
 \end{aligned} \quad (3)$$

The coefficients for our data are weighted, and as a result, the coefficient matrix is not orthonormal. Thus, the coefficients were transformed to become orthonormal (Table 2). Based on the data presented in Table 2, this was accomplished using MATLAB code, with the following

$$\begin{aligned}
 \text{coefforth} &= \text{inv}(\text{diag}(\text{std}(\text{ratings})) * \text{wcoeff}); \\
 I &= \text{c3}' * \text{c3} \\
 I &= \\
 &1.0e - 03 * \\
 \text{results:} & \\
 &\begin{pmatrix} 0.5294 & 0.0034 & 0.0136 \\ 0.0034 & 0.0366 & -0.0061 \\ 0.0136 & -0.0061 & 0.0202 \end{pmatrix}
 \end{aligned} \quad (4)$$

As shown above, the transformed coefficients are orthonormal.

Interpreting the latent vector and visualizing the results. The latent vector describes the variability in the data that is explained by each principal component. Specifically, each independent column of the obtained score matrix has a variance equal to the value in the corresponding row of the latent vector. Based on the data presented in Tables S1, 1 and 2, using MATLAB, it is easy to calculate the elements of the latent vector: 3.3327, 1.0272, 0.9993, 0.8170, 0.7320, 0.5534, 0.3310, and 0.2074, respectively. The scree plot presented in Fig. 2 shows seven out of the eight components, which together account for 98% of the total variance. There is a large gap between the variances of the first and second components. Nevertheless, the first component alone accounts for less than 41.6582% of the variance, whereas the second component explains less than 12.8395% of the variance, and the third component explains less than 12.4917% of the variance. Consequently, multiple components might be needed to adequately describe the data. Figure 2 shows that the first three principal components together account for approximately 66% of the total variability in the standardized data values and thus may serve as a reasonable foundation for reducing the dimensionality of the data.

In Figs 3 and 4, all eight variables are plotted as vectors, where the direction and length of each vector indicate the contribution of the corresponding variable to each principal component. The first principal component, which is shown on the horizontal axis, has positive coefficients for six variables: butanol productivity (g/L/h), butanol yield (g/g), solvent (ABE) yield (g/g), ethanol yield (g/g), maximal specific growth rate (added butyrate: 5.0 g/L; μ_{Max} -A), and maximal specific growth rate (added butyrate: 8.5 g/L; μ_{Max} -C). Thus, the six corresponding vectors lie on the right side of the graph. The second and third elements of the coefficient vector for the first principal component, corresponding to the butanol yield and the solvent (ABE) yield, respectively, have the largest values (Fig. 3). The second principal component, which is shown on the vertical axis, has positive coefficients for the variables representing maximal specific growth rate (added butyrate: 6.5 g/L; μ_{Max} -B), butanol yield (g/g), and ethanol yield (g/g) and negative coefficients for the remaining five variables (Fig. 3). Thus, the second component distinguishes between clusters of the 329 mutant strains that have high values for the first set of variables and low values for the second and clusters for which the opposite is true. Note that Fig. 4 is helpful for cases in which the first two principal components do not account for a sufficient amount of the variance in the multidimensional fermentation datasets. Here, the data points have been scaled with respect to the maximum score and the number of coefficients; therefore, only their relative locations can be found by using the graph.

Capturing high-quality mutant strains. The multivariate distance of each observation from the center of a dataset can be measured using many methods; one common strategy is to use Hotelling's T^2 test, which was first introduced in 1931⁴⁷. This is an analytical way to identify the most extreme points in a dataset⁴⁸⁻⁵⁰. Hotelling's

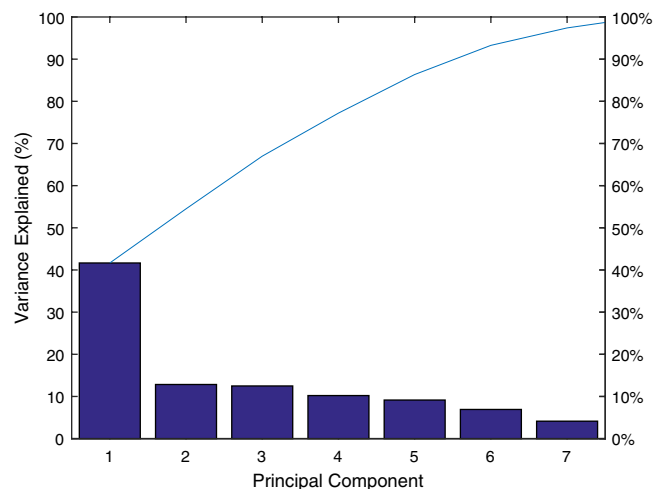


Figure 2. A scree plot of the percent variability explained by each principal component. This scree plot shows only the first seven (instead of all eight) components, which together explain 98% of the total variance. The only marked gap in the amount of variance accounted for by each component is between the first and second components. The first component alone explains less than 42% of the variance; thus, more components might be needed to adequately describe the data.

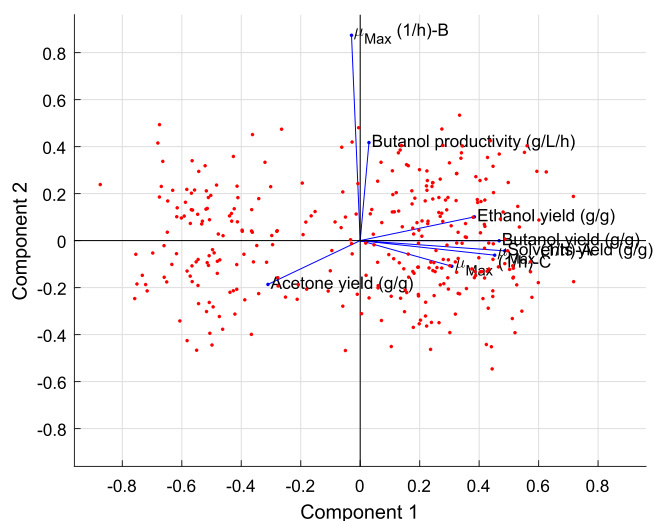


Figure 3. The 2D visualization, which includes one point for each of the 329 observations, with the coordinates indicating the scores of each observation for the two principal components represented in the plot. All eight variables are represented by vectors in this bi-plot, where the direction and length of each vector indicate how each variable contributes to the two principal components represented in the plot.

T^2 test can be thought of as a supplement to the t-test; it can be applied to the scores obtained for a PCA model as follows:

$$T_i^2 = \frac{t_i^T (T^T T)^{-1} t_i}{I - 1} \quad (5)$$

where the matrix of scores ($I \times R$) obtained from the calibration samples is represented by T and t_i is the $R \times 1$ vector representing the R scores for the i^{th} sample. Under the assumption that the scores are normally distributed, the confidence limits for T_i^2 can be assigned as follows:

$$T_{i(I,R)}^2 = \frac{R(I-1)}{1-R} F_{R,I-R,\alpha} \quad (6)$$

Thus, one of the highest-quality mutant strains can be quickly and accurately captured from the multidimensional fermentation datasets representing the 329 mutant strains. As mentioned earlier, this was accomplished using MATLAB code. Based on all previous data analyses, the most extreme point in the datasets was identified as follows:

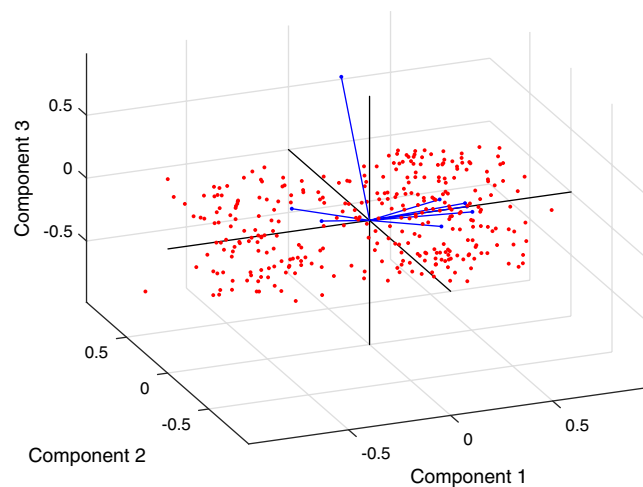


Figure 4. The 3D visualization, which includes one point for each of the 329 observations, with the coordinates indicating the scores of each observation for the three principal components represented in the plot. All eight variables are represented by vectors in this bi-plot, where the direction and length of each vector indicate how each variable contributes to the three principal components represented in the plot.

```
[st2, index] = sort(tsquared, 'descend'); % sort in descending order
extreme = index(1);
names(extreme, :)
ans=
FS - JWS - SN - 79 - 6
```

(7)

This extreme point is not surprising because the measured values for mutant FS-JWS-79-6 are the farthest from the average of the multidimensional fermentation datasets for the 329 mutant strains. In addition, mutants FS-SXT-AP-R631, FS-SXT-LE-BH-A7, FS-SXT-GC-V9-77, FS-ZKS-TF-MT18 and FS-ZKS-TF-ZT637 also appear to be more extreme than the other data.

Influence of addition of butyric acid on extreme mutants. Most industrial processes are described to not use addition of external butyric acid as costs are prohibiting, e.g. the South African process²⁷, the former Soviet Union process⁵¹ or the historic and current processes in China⁵². Yet, the presence of butyric acid in the fermentation broth has been shown to activate solvent production by *C. acetobutylicum*^{53–55}. Additionally, increased yields and increased production of solvents have been reported following the addition of butyric and acetic acid to cultures of *Clostridium beijerinckii* and *C. acetobutylicum*^{56–58}. The production of solvents, especially butanol, is clearly influenced by butyric acid. Adding butyric acid shifts the metabolic stage of the culture by decreasing the pH of the medium, and it can also be utilized by the cells as a co-substrate for the formation of butanol^{59–63}. Therefore, controlling the amount of butyrate in the culture medium is of great industrial importance. However, butyric acid has previously been reported to inhibit cell growth^{64,65}. The fermentation results presented in Table 3 show that cells of the mutant FS-JWS-SN-79-6 showed a better “apparent” tolerance at butyric acid concentrations between 5.0 and 11.5 g/L, although this mutant did exhibit a gradual inhibition of cell growth; no growth was observed for butyric acid concentrations above 11.5 g/L, and the effect of increasing the butyric acid concentration on the batch efficiency parameters (yield from the substrate and productivity) was investigated for this mutant. Often times strains that have a high tolerance (to butyric acid) do not have the highest production rates (of butanol). It is very interesting that the wild-type *C. acetobutylicum* ATCC 824 showed (low) butanol production levels in contrast to extreme mutants FS-JWS-SN-79-6 which produced under absence of butyrate supplementation. The results presented in Table 4 shows the mutant FS-JWS-SN-79-6 that produces more butanol without supplemented butyrate (0.262 g/g) than the wild-type strain. The cells unaffected by butyric acid addition consume *Clostridium* growth medium (CGM) mainly for growth producing simultaneously butyrate as primary metabolite. As butyrate builds up in the system the medium pH drops with the cells shifting their metabolic state from acidogenesis to solventogenesis. It is obvious that extreme mutants FS-JWS-SN-79-6 have strong ability of metabolic. This resulted in a 2.2-fold increase in butanol yield from substrate coupled with 3.1-times more butanol productivity. In addition, the maximal specific growth rates results presented in Table 3 show that butyric acid (5.0 g/L) was slightly added in the media before the optimal point was reached, solvents production was stimulated at the cost of lower specific cell growth rates but with moderate biomass levels. The outcome was a significant increase in the solvents yields from biomass for all strains. The fact that these bacteria were metabolically inactive for solvent production due to low butyric acid levels and suboptimal pHs necessary for solvents production, reflected into high cell growth rates with resultant high biomass levels in the system. As showing in Table 3, the results demonstrate that butyric acid has effectively a prominent inhibitory effect on cell growth with all specific

| Bacterium | Butanol productivity (g/L/h) | Butanol yield (g/g) | Solvents (ABE) yield (g/g) | Acetone yield (g/g) | Ethanol yield (g/g) | Maximal specific growth rate (added butyrate 5.0 g/L) μ_{Max} (1/h) | Maximal specific growth rate (added butyrate 6.5 g/L) μ_{Max} (1/h) | Maximal specific growth rate (added butyrate 8.5 g/L) μ_{Max} (1/h) | Maximal specific growth rate (added butyrate 11.5 g/L) μ_{Max} (1/h) |
|-----------------------------------|------------------------------|---------------------|----------------------------|---------------------|---------------------|---|---|---|--|
| <i>C. acetobutylicum</i> ATCC 824 | 0.022 | 0.183 | 0.235 | 0.076 | 0.024 | 0.183 | 0.087 | 0.069 | — |
| Mutant FS-JWS-SN-79-6 | 0.068 | 0.418 | 0.494 | 0.113 | 0.060 | 0.224 | 0.146 | 0.093 | 0.057 |
| Mutant FS-SXT-AP-R631 | 0.070 | 0.401 | 0.528 | 0.117 | 0.126 | 0.247 | 0.159 | 0.108 | 0.032 |
| Mutant FS-SXT-LE-BH-A7 | 0.073 | 0.405 | 0.53 | 0.118 | 0.128 | 0.240 | 0.157 | 0.111 | 0.041 |
| Mutant FS-SXT-GC-V9-77 | 0.068 | 0.435 | 0.533 | 0.113 | 0.117 | 0.243 | 0.157 | 0.110 | 0.029 |
| Mutant FS-ZKS-TF-MT18 | 0.072 | 0.435 | 0.524 | 0.113 | 0.091 | 0.241 | 0.152 | 0.113 | 0.036 |
| Mutant FS-ZKS-TF-ZT637 | 0.067 | 0.426 | 0.532 | 0.109 | 0.101 | 0.245 | 0.147 | 0.112 | 0.038 |

Table 3. Based on supplementation with 5.0-g/L butyric acid, the effects of increasing the butyric acid concentration (6.5, 8.5 and 11.5 g/L) on the ABE fermentation parameters from the substrate, the butanol productivity, and the maximal specific growth rates for *C. acetobutylicum* ATCC 824 and mutant strains were determined as shown below. *All the experiments were repeated 3 times and the average value was taken.

| Bacterium | Butanol Production (g/L) | Butanol productivity (g/L/h) | Butanol yield (g/g) | Solvents (ABE) yield (g/g) | Acetone yield (g/g) | Ethanol yield (g/g) | Maximal specific growth rate μ_{Max} (1/h) |
|-----------------------------------|--------------------------|------------------------------|---------------------|----------------------------|---------------------|---------------------|--|
| <i>C. acetobutylicum</i> ATCC 824 | 7.73 | 0.017 | 0.164 | 0.318 | 0.083 | 0.031 | 0.236 |
| Mutant FS-JWS-SN-79-6 | 18.43 | 0.048 | 0.262 | 0.413 | 0.126 | 0.068 | 0.242 |

Table 4. The performance of the best mutants against the wild-type in absence of butyrate supplementation on the ABE fermentation parameters from the substrate, the butanol productivity, and the maximal specific growth rates for *C. acetobutylicum* ATCC 824 and mutant FS-JWS-SN-79-6 were determined as shown below.

growth rates declining with increasing butyrate concentrations. This finding confirms previous results obtained with *Clostridium butyricum* grown in a glucose-limited chemostat culture⁶⁶. While cells from *C. beijerinckii* BA 101 could be considered the most resistant ones to critical concentrations of butyrate (10 g/L), cells of *C. beijerinckii* ATCC 55025 evidenced a better “apparent” tolerance in the butyric acid region between 2~8 g/L. In all cases one can see that butyrate feeding favors ABE-solvents production over the control cultures. Above the optimal butyrate feeding concentration (5.0 g/L) cell growth is strongly inhibited lessening butanol yields and productivities for all strains. As showing in Tables 3 and 4, in contrast, as the butanol production pathway of extreme mutants FS-JWS-SN-79-6 is induced by the addition of external butyric acid (5.0 g/L), the levels of *butyryl-CoA* are increased from *acetoacetyl-CoA* instead of forming acetoacetate. This results in a lower acetone production. This corresponded to a 1.1-fold decrease in the acetone yield from 0.126 g/g to 0.113 g/g. Likewise, the wild-type *C. acetobutylicum* ATCC 824 showed (low) acetone yield levels which produced under butyrate supplementation (5.0 g/L). Further details on the metabolic pathways for butanol and acid production can be found elsewhere^{34,67}.

Evaluation of the mutants from the ABE fermentation perspective. In all cases, 5.0 g/L was found to be the optimal concentration of butyric acid for maximizing the yield for all ABE solvents and the butanol productivity. Adding butyric acid to the medium significantly increased the production of butanol, resulting in a global maximum productivity of 0.068 g/L/h in the fermentation broth for this mutant. This corresponded to a 2.28-fold increase in the butanol yield from the substrate (0.183 g/g), coupled with a 3.09-fold increase in productivity (0.022 g/L/h) (Table 3). Moreover, high intracellular concentrations of this acid activated the enzymes to produce neutral products. Thus, the mutant FS-JWS-SN-79-6 will synthesize the enzymes for butanol production at pH 7 as the butyrate concentration in the medium increases. In similar experiments, cells have been routinely observed to continue to grow when supplemented with butyric acid at concentrations of 5.0~8.5 g/L and, in some cases, above 11.5 g/L. Above the optimal level of butyric acid concentration, the yield values decrease as a direct consequence of gradual cell growth inhibition by the co-substrate with concomitant low biomass concentration coupled by low levels of butanol produced. ABE formation of ethanol and acetone limits the amount of metabolic precursors available for butanol production⁶⁸. As showing in Tables 3 and 4, the results demonstrate that acetone and ethanol production levels were not significantly affected. Based on supplementation with 5.0-g/L butyric acid compared to the wild-type strain, the conversion of CGM to butanol yield by the mutant FS-JWS-SN-79-6 had increased with 43.8%, and total ABE solvent yields from CGM were up with 47.6%. In absence of butyrate supplementation compared to the wild-type strain, the conversion of CGM to butanol yield by the mutant FS-JWS-SN-79-6 had increased with a 1.6-fold, and total ABE solvent yields from CGM were up with a 1.3-fold. Above the results correspond to the previous investigations, to restore solvent productivity, Nair and Papoutsakis

(1994) expressed the alcohol dehydrogenase gene (*adhE*), normally located on *pSOL1*, in strain M5 from a replicative plasmid⁶⁹. Butanol yield was restored without acetone formation, but at reduced levels compared to the wild-type strain, while large amounts of acetate and butyrate accumulated in the growth medium.

High-production of biobutanol by *Clostridium* have been reported detailedly in the following research literature. Such as hyper-butanol producing strains from various mutagenesis strategy and mutants, including acetone-butanol-ethanol production with high productivity using *Clostridium acetobutylicum* BKM19⁷⁰, acetone-butanol-ethanol production from cane molasses using *Clostridium beijerinckii* mutant obtained by combined low-energy ion beam implantation and N-methyl-N-nitro-N-nitrosoguanidine induction⁷¹, comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in *Clostridium acetobutylicum* EA 2018⁷² and recent advances in ABE fermentation: hyper-butanol producing *Clostridium beijerinckii* BA101⁷³. Yields from biomass up to 17.6 g/L of butanol and the maximum butanol and ABE productivities of 9.6 and 20.0 g/L/h from 85.2 g/L glucose in the non-mutant *C. acetobutylicum* batch fermentation were obtained⁷⁰. Supplementing the fermentation medium (MP2) with sodium acetate enhances solvent production to 33 g/L by in the non-mutant *Clostridium beijerinckii* BA101⁷³. The production of butanol was 15.8 ± 0.7 g/L by *Clostridium beijerinckii* L175 after mutagenesis technique with N⁺ ion implantation⁷¹. In contrast, our results presented in Table 4 show that mutant *FS-JWS-SN-79-6* was also enhanced to improve butanol production from 7.73 to 18.43 g/L after ¹²C⁶⁺ heavy ion irradiation. Clearly, Mutant *FS-JWS-SN-79-6* produces more butanol without supplemented butyrate (18.43 g/L) than the *C. acetobutylicum* ATCC 824. Our findings suggested using ¹²C⁶⁺ heavy ion irradiation favors ABE-solvents production over the non-radiated strain and other mutagenesis strategy. The strategy reported here may contribute to develop a cost-effective butanol fermentation process, making it competitive compared with similar fermentation processes.

Conclusion

PCA is a multivariate method that is used to examine datasets in which the observations can be expressed in terms of many inter-correlated quantitative dependent variables. Furthermore, PCA can be simplified to a correspondence analysis that handles qualitative variables. It can also be regarded as a multiple factor analysis that handles heterogeneous sets of variables. The purpose of PCA is to determine relevant information from a dataset, characterize it in terms of a set of new orthogonal variables (principal components), and visualize the patterns of similarity in the variables and observations as specific locations on a map. In this work, data on a large number of mutant strains of *C. acetobutylicum* secreted after mutagenesis were collected. For industrial ABE fermentation, distinguishing between high-quality and mediocre producing mutants is highly important. Because these organisms' multidimensional fermentation datasets contain many correlated variables, PCA can serve as an inexpensive, efficient and reliable approach for identifying high-quality mutants. Through data acquisition, normalization, analysis simplification, and visualization, the proposed interactive approach helps users to understand and rapidly apply PCA by creating a visual model in their minds. Finally, the most extreme points are clearly evident, allowing high-quality mutant strains to be easily captured. In conclusion, the information obtained in this research will support further studies of bioreactor and bioprocess design and scale-up, which are very important topics for ABE industrial applications.

Methods

Cultures and medium. To test the production of butanol by various strains, a rich P2 medium containing 60 g/L glucose, 3.6 g/L yeast extract, 2.7 g/L peptone, 3.2 g/L K₂HPO₄, 3.2 g/L KH₂PO₄, 0.2 g/L MgSO₄, 0.2 g/L MnSO₄, 0.02 g/L FeSO₄, 0.02 g/L NaCl, 1.5 g/L yeast extract (Difco, USA), 2.5 g/L ammonium acetate, 0.0005 g/L p-aminobenzoate, 0.0005 g/L thiamin, 0.00005 g/L biotin, and 35 µg/mL thiamphenicol was used. To test the production of butanol from different substrates, the same rich P2 medium with 30 g/L instead of 60 g/L glucose was used^{73,74}. Unless otherwise noted, the fermentation was conducted in serum bottles, each of which contained 40 mL of the medium and was inoculated with 1% (v/v) of an overnight culture in Reinforced *Clostridial* Medium (RCM; Difco, Detroit, MI, USA) at 37 °C and 250 rpm. The pH was maintained between 5.0 and 6.5 by adding NaOH solution twice a day⁷⁵.

Microorganisms and breeding. *Clostridium acetobutylicum* ATCC 824 was obtained from the Drug R & D Center of Institute of Modern Physics, Chinese Academy of Sciences, China. All bacteria were maintained in P2-medium at 4 °C as stock cultures. To prepare inocula of all 4-strains in totally anoxic conditions the following procedure was employed: serum tubes containing 5.0 ml of P2-medium were first purged with sterile nitrogen gas for 5-min. To prevent caramelization of sugar, a browning reaction, a separate 50 g-l-1 dextrose solution in distilled water was prepared in a 100-mL serum bottle and purged with nitrogen gas for 15-min again to attain perfect anaerobic conditions^{76,77}. Both vessels were tightly sealed with rubber stoppers and aluminum crimps to prevent ingress of air and contamination with oxygen. Both liquids were sterilized by autoclaving at 121 °C, 15 psig for 20-min after which they were left at room temperature for cooling. 1.0 ml of dextrose solution was then added to the first tube followed by cell inoculation with 1/30 volume of each original stock culture. Anaerobic stock cultures for all strains were taken from an original serum tubes stored at 4 °C. Prior to inoculation the stock culture tubes were left resting at room temperature for 30-min in order to pre-activate the cells. The pre-culture was incubated at 37 °C during 16-hours for cell growth followed by another inoculation around in order to obtain final fresh cell culture inocula.

Experimental setup and heavy-ion beam irradiation. The experiment was performed at the Cancer Therapy Terminal of the Heavy Ion Research Facility at Lanzhou (HIRFL). The upgraded accelerator system of HIRFL consists of a Sector Focus Cyclotron (SFC), a Separated Sector Cyclotron (SSC), the main Cooling Storage

Ring (CSRm), and the experimental Cooling Storage Ring (CSRe). High-energy $^{12}\text{C}^{6+}$ -ions with an energy of 196 AMeV were extracted by CSRm. Energies of 117 AMeV was obtained by adding the absorbers (water) and calibrating using the LISE program, and the corresponding uncertainty of the energies is not higher than 0.27%⁷⁸. The extraction time of the carbon ions (approximately 10^6 – 10^8 ions/pulse) was approximately 3 s, and the priming dose was 80 Gy. The dose rates were up to 10 Gy/min. The temperature of the $^{12}\text{C}^{6+}$ heavy-ion beams was $<35^\circ\text{C}$ under these conditions⁷⁹. For irradiation experiments, strains cells were grown in microcentrifuge tube (5 mL) to reach 90% confluence and they were completely filled with Dulbecco's modified Eagle's medium to avoid artifacts by irradiation through air layers.

Totally anoxic conditions. Serum tubes containing 7.0 mL of P2 medium were first purged with sterile 80% N_2 , 10% CO_2 , and 10% H_2 for 9 min. To prevent caramelization of the sugar, which is a browning reaction, a separate 60 g/L dextrose solution in distilled water was prepared in a 120-mL serum bottle and purged with nitrogen gas for 18 min to attain completely anaerobic conditions. Both vessels were tightly sealed with rubber stoppers and aluminum crimps to prevent the ingress of air and contamination with oxygen. Both liquids were sterilized by autoclaving at 121°C and 15 psig for 25-min and were then cooled at room temperature. After 1.5 mL of dextrose solution was added to the first tube, the tube was inoculated with a 0.04 volume of each original stock culture. The anaerobic stock cultures of all of the strains were collected from the original serum tubes stored at 4°C . Prior to inoculation, the stock culture tubes were incubated at room temperature for 25 min to pre-activate the cells. The pre-culture was incubated at 37°C for 36 h to allow cell growth and then inoculated to obtain the final fresh cell culture inocula⁸⁰.

The source of the multidimensional fermentation dataset. The generation of mutant strains, the fermentation screening, the measurements and analytical methods accompanies this paper at Supplementary Information 1.

Simplifying multidimensional fermentation dataset analysis and visualization. PCA is performed by determining the eigenvalues and eigenvectors of a covariance matrix. This covariance matrix is then utilized to determine the variation in the values of each dimension with respect to the mean. The dimensions of the data considered in our study can be described as random variables and often vary together. Such a relationship can be described as follows:

$$\text{Cov}(X, Y) = E[E[X] - X] \cdot E[E[Y] - Y] \quad (8)$$

where $E[X]$ and $E[Y]$ are the expected values of X and Y , respectively. This can be further written as follows for a sampled dataset:

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N} \quad (9)$$

where \bar{x} and \bar{y} are the mean values of X and Y , respectively, and N is the number of dimensions of the dataset. The covariance matrix is then defined as $A_{i,j} = \text{Cov}(i, j)$, where the data have been mean centered.

For an element of the covariance matrix, the sign is more important than the magnitude. For example, if the sign is positive, it indicates that both of the corresponding dimensions (X and Y) increase simultaneously. Conversely, if the sign of a matrix element is negative, it indicates that when one of the corresponding dimensions increases, the other decreases. When the covariance is zero, the two dimensions are independent of each other. According to the commutative property, $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

The eigenvalues and eigenvectors of interest are computed using the covariance matrix. Then, the eigenvalues are arranged in descending order, creating an order of significance. The eigenvector with the largest eigenvalue is considered to be the most dominant principle component (PC1), which describes the most significant relationship. The principal components are calculated through multiplication of the eigenvectors by the stratified eigenvalues.

PCA can be used as a dimension-reduction method through determination of the principal components of the input data. However, for the transformation of a high-dimensional dataset into a lower-dimensional space, the ideal low-dimensional space must be found from the eigenvectors of the covariance matrix. The ideal low-dimensional space minimizes the error between the input dataset and the PCA results based on

$$\left. \frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^K \lambda_i} \right\} \theta \quad (10)$$

where K is the number of dimensions chosen from among the original N dimensions of the matrix, θ is the threshold criterion (typically 0.9 or 0.95), and λ is an eigenvalue. Using this information, the $N \times N$ matrix is linearly transformed into an $N \times K$ matrix. Although the number of dimensions decreases with the application of PCA, the difference between the input and output matrices is small. Common values of K are 2 and 3, which correspond to the mapping of a dataset into 2D and 3D coordinate systems, respectively.

References

1. Pearson, K. Principal components analysis. *The London, Edinburgh and Dublin Philosophical Magazine and Journal*. **6**, 566 (1901).
2. Grattan-Guinness, I. *The Norton history of the mathematical sciences: The rainbow of mathematics*. (WW Norton & Company, 1997).
3. Jordan, C. Mémoire sur les formes bilinéaires. *J. Math. Pure. Appl.* **19**, 35–54 (1874).

4. Cayley, P. On the analytical forms called trees. *Am. J. Math.* **4**, 266–268 (1881).
5. Hamilton, J. D. *Time series analysis*. Vol. 2 (Princeton university press Princeton, 1994).
6. Boyer, C. B. & Merzbach, U. C. *A history of mathematics*. (John Wiley & Sons, 2011).
7. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417 (1933).
8. Turk, M. A. & Pentland, A. P. Face recognition using eigenfaces. *Computer Vision and Pattern Recognition*, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on. 586–591 (IEEE).
9. Chellappa, R., Wilson, C. L. & Sirohey, S. Human and machine recognition of faces: A survey. *P. IEEE.* **83**, 705–741 (1995).
10. Belhumeur, P. N., Hespanha, J. P. & Kriegman, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE T. Pattern. Anal.* **19**, 711–720 (1997).
11. Safonova, A., Hodgins, J. K. & Pollard, N. S. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM. T. Graphic.* **23**, 514–521 (2004).
12. Chai, J. & Hodgins, J. K. Performance animation from low-dimensional control signals. *ACM. T. Graphic.* **24**, 686–696 (2005).
13. Liu, C. K., Hertzmann, A. & Popović, Z. Learning physics-based motion style with nonlinear inverse optimization. *ACM. T. Graphic.* **24**, 1071–1081 (2005).
14. Koren, Y. & Carmel, L. Visualization of labeled data using linear transformations. *IEEE Symposium on Information Visualization 2003 Institute of Electrical & Electronics Engineers (IEEE)* (2003).
15. Pang, S., Ozawa, S. & Kasabov, N. Incremental linear discriminant analysis for classification of data streams. *IEEE T. Syst. Man. Cy. B.* **35**, 905–914 (2005).
16. Yu, J. & Turk, G. Reconstructing surfaces of particle-based fluids using anisotropic kernels. *ACM. T. Graphic.* **32**, 5 (2013).
17. Roberts, J. C. State of the art: Coordinated and Multiple Views in Exploratory Visualization, CMV'07. Fifth International Conference on. 61–71 (IEEE) (2007).
18. Cockburn, A., Karlson, A. & Bederson, B. B. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM. Comput. Surv.* **41**, 2 (2009).
19. McLachlan, P., Munzner, T., Koutsofios, E. & North, S. LIVeRac: interactive visual exploration of system management time-series data. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1483–1492 (ACM).
20. Gleicher, M. *et al.* Visual comparison for information visualization. *Inform. Visual.* **10**, 289–309 (2011).
21. Hyvärinen, A. & Oja, E. Independent component analysis: algorithms and applications. *Neural. Networks.* **13**, 411–430 (2000).
22. Tropsha, A., Gramatica, P. & Gombar, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSAR models. *QSAR Comb. Sci.* **22**, 69–77 (2003).
23. Hastie, T., Tibshirani, R. & Friedman, J. *Unsupervised learning. The elements of statistical learning* 485–585 (Springer, 2009).
24. Moore, B. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Automat. Contr.* **26**, 17–32 (1981).
25. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemometr. Intell. Lab.* **2**, 37–52 (1987).
26. Jolliffe, I. *Principal component analysis*. (Wiley Online Library, 2002).
27. Jones, D. T. & Woods, D. R. Acetone-butanol fermentation revisited. *Microbiol. Rev.* **50**, 484 (1986).
28. Köpke, M. & Dürre, P. Biochemical production of biobutanol. *Handbook of biofuels production: processes and technologies* 221–257 (2011).
29. Nigam, P. S. & Singh, A. Production of liquid biofuels from renewable resources. *Prog. Energ. Combust.* **37**, 52–68 (2011).
30. Bankar, S. B., Survase, S. A., Ojamo, H. & Granström, T. Biobutanol: the outlook of an academic and industrialist. *RSC. Adv.* **3**, 24734–24757 (2013).
31. Nimcevic, D. & Gapes, J. R. The acetone-butanol fermentation in pilot plant and pre-industrial scale. *J. Mol. Microb. Biotech.* **2**, 15–20 (2000).
32. Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409**, 258–268 (2001).
33. Ezeji, T. C., Qureshi, N. & Blaschek, H. P. Butanol fermentation research: upstream and downstream manipulations. *Chem. Rec.* **4**, 305–314 (2004).
34. Nölling, J. *et al.* Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**, 4823–4838 (2001).
35. Papoutsakis, E. T. Engineering solventogenic clostridia. *Curr. Opin. Biotech.* **19**, 420–429 (2008).
36. Jiang, Y. *et al.* Disruption of the acetoacetate decarboxylase gene in solvent-producing *Clostridium acetobutylicum* increases the butanol ratio. *Metab. Eng.* **11**, 284–291 (2009).
37. Green, E. M. Fermentative production of butanol—the industrial perspective. *Curr. Opin. Biotech.* **22**, 337–343 (2011).
38. Jang, Y. S. *et al.* Butanol production from renewable biomass: rediscovery of metabolic pathways and metabolic engineering. *Biotech. J.* **7**, 186–198 (2012).
39. Demain, A. L. & Adrio, J. L. Contributions of microorganisms to industrial biology. *Mol. Biotechnol.* **38**, 41–55 (2008).
40. Baltz, R. H., Demain, A. L. & Davies, J. E. *Manual of industrial microbiology and biotechnology*. (American Society for Microbiology Press, 2010).
41. Donadio, S. *et al.* Antibiotic discovery in the twenty-first century: current trends and future perspectives. *J. Antibiot.* **63**, 423–430 (2010).
42. Kornilowicz-Kowalska, T. & Bohacz, J. Biodegradation of keratin waste: theory and practical aspects. *Waste. Manage.* **31**, 1689–1701 (2011).
43. Takors, R., Gerigk, M., Paschold, H. *et al.* Principal-component analysis for microbial L-phenylalanine production. *Bioproc. Biosyst. Eng.* **24**, 93–99 (2001).
44. Aceves-Lara, C. A. *et al.* Experimental determination by principal component analysis of a reaction pathway of biohydrogen production by anaerobic fermentation. *Chem. Eng. Process.* **47**, 1968–1975 (2008).
45. Petersen, N., Stocks, S. & Gernaey, K. V. Multivariate models for prediction of rheological characteristics of filamentous fermentation broth from the size distribution. *Biotechnol. Bioeng.* **100**, 61–71 (2008).
46. Biccato, S. *et al.* Fermentation diagnosis by multivariate statistical analysis. *Appl. Biochem. Biotech.* **102**, 49–62 (2002).
47. Hotelling, H. The generalization of student's ratio. *Breakthroughs in Statistics* 54–65 (Springer, 1992).
48. Reinganum, M. R. A new empirical perspective on the CAPM. *J. Financ. Quant. Anal.* **16**, 439–462 (1981).
49. Mason, R. L. & Young, J. C. Improving the sensitivity of the T2 statistic in multivariate process control. *J. Qual. Technol.* **31**, 155 (1999).
50. Girma, S. & Görg, H. Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. *J. Int. Econ.* **72**, 97–112 (2007).
51. Zverlov, V. V., Berezina, O., Velikodvorskaya, G. A. *et al.* Bacterial acetone and butanol production by industrial fermentation in the Soviet Union: use of hydrolyzed agricultural waste for biorefinery. *Appl. Microbiol. Biot.* **71**, 587–597 (2006).
52. Ni, Y. & Sun, Z. Recent progress on industrial fermentative production of acetone-butanol-ethanol by *Clostridium acetobutylicum* in China. *Appl. Microbiol. Biot.* **83**, 415–423 (2009).
53. Bowles, L. K. & Ellefson, W. L. Effects of butanol on *Clostridium acetobutylicum*. *Appl. Environ. Microb.* **50**, 1165–1170 (1985).
54. Tashiro, Y. *et al.* High butanol production by *Clostridium saccharoperbutylacetonicum* N1-4 in fed-batch culture with pH-stat continuous butyric acid and glucose feeding method. *J. Biosci. Bioeng.* **98**, 263–268 (2004).

55. Ezeji, T., Milne, C., Price, N. D. & Blaschek, H. P. Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms. *Appl. Microbiol. Biot.* **85**, 1697–1712 (2010).
56. Zhu, Y. & Yang, S.-T. Effect of pH on metabolic pathway shift in fermentation of xylose by *Clostridium tyrobutyricum*. *J. Biotechnol.* **110**, 143–157 (2004).
57. Lee, S. Y. *et al.* Fermentative butanol production by Clostridia. *Biotechnol. Bioeng.* **101**, 209–228 (2008).
58. Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B. & Keasling, J. D. Microbial engineering for the production of advanced biofuels. *Nature* **488**, 320–328 (2012).
59. Gehrels, N. *et al.* The Swift gamma-ray burst mission. *Astrophys. J.* **611**, 1005 (2004).
60. Ramey, D. & Yang, S.-T. Production of butyric acid and butanol from biomass. final report to the US Department of Energy, Contract No.: DE-F-G02-00ER86106 (2004).
61. Tashiro, Y., Takeda, K., Kobayashi, G. & Sonomoto, K. High production of acetone–butanol–ethanol with high cell density culture by cell-recycling and bleeding. *J. Biotechnol.* **120**, 197–206 (2005).
62. Antoni, D., Zverlov, V. V. & Schwarz, W. H. Biofuels from microbes. *App. Microbiol. Biot.* **77**, 23–35 (2007).
63. Feng, L., Chen, Y. & Zheng, X. Enhancement of waste activated sludge protein conversion and volatile fatty acids accumulation during waste activated sludge anaerobic fermentation by carbohydrate substrate addition: the effect of pH. *Environ. Sci. Technol.* **43**, 4373–4380 (2009).
64. Hague, A. *et al.* Sodium butyrate induces apoptosis in human colonic tumour cell lines in a p53-independent pathway: implications for the possible role of dietary fibre in the prevention of large-bowel cancer. *Inter. J. Cancer.* **55**, 498–505 (1993).
65. Hague, A., Elder, D. J., Hicks, D. J. & Paraskeva, C. Apoptosis in colorectal tumour cells: induction by the short chain fatty acids butyrate, propionate and acetate and by the bile salt deoxycholate. *Int. J. Cancer.* **60**, 400–406 (1995).
66. de Mattos, C. A. *et al.* Heterogeneity of the L2 gene of field isolates of bluetongue virus serotype 17 from the San Joaquin Valley of California. *Virus. Res.* **31**, 67–87 (1994).
67. Desai, A. *et al.* Kin I kinesins are microtubule-destabilizing enzymes. *Cell.* **96**, 69–78 (1999).
68. Ezeji, T. *et al.* Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms. *Appl. Microbiol. Biot.* **85**, 1697–1712 (2010).
69. Nair, R. V. & Papoutsakis, E. T. Expression of plasmid-encoded aad in *Clostridium acetobutylicum* M5 restores vigorous butanol production. *J. Bacteriol.* **176**, 5843–5846 (1994).
70. Jang, Y. S., Malaviya, A. & Lee, S. Y. Acetone–butanol–ethanol production with high productivity using *Clostridium acetobutylicum* BKM19. *Biotechnol. Bioeng.* **110**, 1646–1653 (2013).
71. Li, H. *et al.* Acetone, butanol, and ethanol production from cane molasses using *Clostridium beijerinckii* mutant obtained by combined low-energy ion beam implantation and N-methyl-N-nitro-N-nitrosoguanidine induction. *Bioresource. Technol.* **137**, 254–260 (2013).
72. Hu, S. *et al.* Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in *Clostridium acetobutylicum* EA 2018. *BMC genomics.* **12**, 1 (2011).
73. Qureshi, N. & Blaschek, H. P. Recent advances in ABE fermentation: hyper-butanol producing *Clostridium beijerinckii* BA101. *J. Ind. Microbiol. Biot.* **27**, 287–291 (2001).
74. Groot, W. J., Van der Lans, R. & Luyben, K. C. A. M. Technologies for butanol recovery integrated with fermentations. *Process. Biochem.* **27**, 61–75 (1992).
75. Huang, W. C., Ramey, D. E. & Yang, S. T. Continuous production of butanol by *Clostridium acetobutylicum* immobilized in a fibrous bed bioreactor. *Appl. Biochem. Biotechnol.* **115**, 887–898 (2004).
76. Chang, J. J. *et al.* Syntrophic coculture of aerobic *Bacillus* and anaerobic *Clostridium* for bio-fuels and bio-hydrogen production. *Int. J. Hydrogen. Energ.* **33**, 5137–5146 (2008).
77. Oshiro, M. *et al.* Efficient conversion of lactic acid to butanol with pH-stat continuous lactic acid and glucose feeding method by *Clostridium saccharoperbutylacetonicum*. *Appl. Microbiol. Biot.* **87**, 1177–1185 (2010).
78. Lei, Y. *et al.* Fluorescence emission from CsI(Tl) crystal induced by high-energy carbon ions. *Opt. Mater.* **35**, 1179–1183 (2013).
79. Zhou, X. *et al.* Radiation induces acid tolerance of *Clostridium tyrobutyricum* and enhances bioproduction of butyric acid through a metabolic switch. *Biotechnol. Biofuels.* **7**, 1 (2014).
80. Bogin, O. *et al.* Enhanced thermal stability of *Clostridium beijerinckii* alcohol dehydrogenase after strategic substitution of amino acid residues with prolines from the homologous thermophilic *Thermoanaerobacter brockii* alcohol dehydrogenase. *Protein. Science.* **7**, 1156–1163 (1998).

Acknowledgments

We sincerely thank the National Laboratory of the Heavy Ion Research Facility in Lanzhou (HIRFL), the National Natural Science Foundation of China and the Chinese Academy of Science (CAS) Light of West China Program for giving us the opportunity to perform this project. This work was supported by grants from the National Natural Science Foundation of China (Grant No.11105193), the CAS Light of West China talent training Program (Ke-Fa-Ren-Zi [2015] No. 77) and the Natural Science Foundation of Gansu provincial Sci. and Tech. Department (Grant No. 1506RJZA293). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

X.Z. conceived, designed and supervised the study. X.Z., D.X., and T.T.J. performed the experiments. X.Z., D.X., and T.T.J. analyzed the data. X.Z., D.X., and T.T.J. contributed reagents/materials/analysis tools. X.Z. wrote the paper. X.Z. critically revised the manuscript. X.Z. final approval of the version to be published. All authors contributed to the discussion and comments on the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhou, X. *et al.* Simplifying multidimensional fermentation dataset analysis and visualization: One step closer to capturing high-quality mutant strains. *Sci. Rep.* **7**, 39875; doi: 10.1038/srep39875 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017