input to compromise healthcare quality. If this happens, the effect may not be obvious within the next few years.

Factors that distinguish health economics from other areas include extensive government intervention to safeguard basic healthcare level and to secure associated research. As a result of the current global financial climate, international measures need to be taken by responsible organizations to maintain R&D. First, extra-funding needs to be allocated for R&D and the injection of funds into the system needs to be prioritized. And second, alternative funding sources need to be established by the government to identify and compensate the imminent cuts in funding that are affecting the financial sustainability of biomedical innovation in the private sector. This in turn will help secure the future of these areas against any further crises.

Kamran Ahmed & Hutan Ashrafian

Department of Biosurgery & Surgical Technology, Imperial College London.
e-mail: k.ahmed@imperial.ac.uk

1. Scangos, G.A. *Nat. Biotechnol.* **27**, 424–425 (2009).
2. Serajuddin, H.K. & Serajuddin, A.T. *J. Am. Pharm. Assoc.* **46**, 511–516 (2006).
3. Kaitin, K.I., Bryant, N.R. & Lasagna, L. *J. Clin. Pharmacol.* **33**, 412–417 (1993).
4. Mitchell, P. *Nat. Biotechnol.* **26**, 359–360 (2008).
5. Henry, D. & Lexchin, J. *Lancet* **360**, 1590–1595 (2002).
6. Maisonneuve, H. *et al. J. Contin. Educ. Health Prof.* **29**, 58–62 (2009).

Improved genome annotation for *Zymomonas mobilis*

To the Editor:

The first genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4 was reported in your journal five years ago[1]. Because of its productivity, high level of ethanol tolerance and ability to be genetically manipulated[2], *Z. mobilis* is a promising industrial bacterium for fermenting lignocellulosic biomass into ethanol, which is being advanced as an alternative to petroleum-derived transportation fuels[3]. Strains of the bacterium have been developed to ferment both hexoses and pentoses into ethanol[4], and the genome sequence will provide further opportunities for strain developments as well as providing more fundamental insights[5]. We have observed many differences between the primary annotation and one performed by the J. Craig Venter Institute (JCVI) and have detected differential gene expression in genes predicted by JCVI (http://cmr.jcvi.org/cgi-bin/CMR/GenomePage.cgi?org=ntzm01) that were absent from the primary annotation in our previous work[6]. We present here an improved version of the ZM4 genome annotation based on the original primary genome sequence, new experimental data generated by 454 pyrosequencing and mass spectrometry–based proteomics, a novel gene prediction algorithm called Prodigal (prokaryotic dynamic programming genefinding algorithm) that is incorporated into an annotation pipeline, and manual curation (**Supplementary Table 1**). Prodigal has been used to annotate all microorganisms submitted to GenBank by the Department of Energy (DOE) Joint Genome Institute (JGI; Walnut Creek, CA, USA) since November 2007 and is freely available (http://compbio.ornl.gov/prodigal/).

nature
biotechnology
The expanding promise
of cultured red blood cells
Global identification of phosphotyrosine sites
Microfluidic FACS with optical switches

Initially, we used 454-pyrosequencing data to identify those regions of DNA or single nucleotide polymorphisms that differed between the *Z. mobilis* ZM4 ATCC31821 reference genome (GenBank accession no. AE008692) and the *Z. mobilis* ZM4 ATCC31821 sequence generated in this study. We generated ~42.3-Mb sequence data that resulted in ~21-fold genome coverage. We detected 166 high-confidence sequence differences based on GSMapper (454 Roche) data and peptide hits from shotgun proteomics data. Differences were evaluated using BLASTN alignments and then were incorporated into an updated ZM4 chromosome sequence. The ZM4 sequence changes, mapped reads, their respective quality scores and the details of the software and parameters used in study are available at our website (**Supplementary Table 1**). We have also sequenced the genome of an acetate-tolerant strain derived from *Z. mobilis* ZM4 ATCC31821 that was selected in another geographically separated laboratory[7] and report 454 pyrosequencing and Sanger sequencing and peptide support for our changes to the ZM4 chromosome (**Supplementary Table 1**). In addition, the entire ZM4 pyrosequencing data set has been deposited in the National Center for Biotechnology Information (NCBI) short-read archive database (Study SRP000908). We processed the updated sequence data using the automated Oak Ridge National Laboratory (ORNL) microbial genome annotation pipeline. Finally, we examined the gene models predicted in the original GenBank annotation, the TIGR reannotation and our new reannotation and updated the ZM4 annotation in a final manual curation step. The final curation was performed in conjunction with a defined set of criteria (available with reannotation) and several proteomics data sets that showed peptide support for more than half of the theoretical proteome. An overview of the extensive changes made to the ZM4 chromosome based upon mass-spectrometry proteomics and pyrosequencing data and six illustrative examples are presented (**Table 1** and **Supplementary Fig. 1**, respectively). We have converted 61 pseudogenes in the original annotation into 43 full-length coding sequences, which include predicted genes with important metabolic and physiological functions (e.g., GenBank acc. nos. for tRNA synthetases ZMO0460, ZMO0843, ZMO0845, ZMO1508, ZMO1878 and flagella gene fliF, ZMO0633) (**Supplementary Table 2**). Several of the updated chromosomal nucleotides are consistent with earlier ZM4 fosmid DNA sequence data (e.g., GenBank acc. no. AAG29859) and we have peptide support for 6 of our 37 newly predicted chromosomal genes (**Supplementary Table 3**). We did not identify peptides corresponding to any of the putative genes that we deleted. A comprehensive comparison on a gene-by-gene basis is presented in **Supplementary Table 4**. We have provided our analysis to the authors of the primary genome annotation and they are in the process of updating their GenBank submission.

Plasmid DNA was also identified in our 454-pyrosequencing data, which was consistent with earlier reports on *Z. mobilis* plasmids[8] and with plasmid sequence deposited earlier in GenBank for strain ZM4 (plasmid 1, GenBank acc. no. AY057845). Plasmid DNA sequence data was not reported in the primary ZM4 genome annotation[1]. We present a collaborative analysis of extensive changes to the ZM4 chromosome gene models and five new ZM4 plasmids to provide a comprehensive update to the ZM4 genome. Purified plasmid DNA was prepared as described previously[9] and sequenced at JGI using a 3 kb DNA library, which provided tenfold coverage for the draft assemblies. Gaps between contigs were closed by editing in Consed, custom primer walk or PCR amplification and a total of five additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The finished ZM4 plasmids were between 30,952 bp and 37,066 bp. The plasmid sequence data contained 2,261 reads, achieving an average of tenfold sequence coverage per base with an error rate <1 in 100,000. The 156 new plasmid gene models represent coding sequences for hypothetical genes (71) with unknown functions, genes for transport, regulation and metabolism (e.g., an L-fucose isomerase, ZZM4_0056; an iron-containing alcohol dehydrogenase, ZZM4_0085), genes for plasmid replication and/or maintenance, genes belonging to restriction-modification systems, as well as phage-related genes. These genes can be found via http://genome.ornl.gov/microbial/zmobORNL/paper/ (under “Sequence and annotation of ZM4 plasmids” and “Gene Model Summary” (in tab-delimited files)) or at http://maple.lsd.ornl.gov/cgi-bin/JGI_microbial/display_page.cgi?page=summary&org=zmob_ZM4&chr=10jul09. We have improved the *Z. mobilis* ZM4 genome sequence, which is an essential component for successful future systems biology studies in this and other important ethanologenic microorganisms. Our sequence and annotation update will immediately enable systems biology studies (e.g., microarray and proteomics) that rely on the best and most up-to-date genome annotation.

An understanding of fermentative microorganisms at the systems level has the potential to assist in strain development, which is an important component in displacing a substantial amount of petroleum-derived fuels with biofuels at an industrial scale. One of the most important implications of our study is, however, the value of the experimental validation of gene model predictions. We submit when microbial DNA is prepared for future genome sequencing studies that baseline experimental validation by high-throughput methodologies, such as mass spectrometry–based proteomics or transcript profiling via next-generation sequencing, should be considered as a component for the assessment and development of gene models and annotation.

Table 1 Summary of the ZM4 genome reannotation

Features changed	Number
Genes not changed	1,180
Sequence change/no annotation change	127
Predicted start changed/longer gene	140
Sequence change and predicted start/longer gene	22
Predicted start changed/shorter gene	188
Sequence change and predicted start/shorter gene	1
Genes deleted	271
New genes added to the chromosome	37
New plasmid genes	156
Number of genes corrected from pseudogenes	43
Genes changed to pseudogenes and new pseudogenes	11
Miscellaneous RNA binding features added	3
tRNAs changed to the correct strand	28
Edges of 16S RNA gene changed	3
CRISPR repeats annotated	3

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS
We thank A. Savidor for making his six-frame translation proteomics search scripts available, L. Goodwin (US DOE JGI) for management of plasmid sequencing and O. Chertkov (JGI–Los Alamos National Laboratory) for plasmid sequence improvement and finishing. Research sponsored in part by the Laboratory Directed Research and Development Program of ORNL, managed by UT-Battelle, LLC for the US Department of Energy under contract no. DE-AC05-00OR22725; the BioEnergy Science Center, a US DOE Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science; and the DOE JGI managed by the University of California, Lawrence Berkeley National Laboratory under contract no. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract no. DE-AC52-07NA27344 and Los Alamos National Laboratory under contract no. DE-AC02-06NA25396.

Shihui Yang[1,2], Katherine M Pappas[3], Loren J Hauser[1,2], Miriam L Land[1,2], Gwo-Liang Chen[1], Gregory B Hurst[4], Chongle Pan[2,5], Vassili N Kouvelis[3], Milton A Typas[3], Dale A Pelletier[1], Dawn M Klingeman[1,2], Yun-Juan Chang[1], Nagiza F Samatova[5] & Steven D Brown[1,2]

[1]Biosciences Division and [2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA. [3]Department of Genetics & Biotechnology, University of Athens, Athens, Greece. [4]Chemical Sciences Division and [5]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA. e-mail: brownsd@ornl.gov, kmpappas@biol.uoa.gr

1. Seo, J.S. *et al. Nat. Biotechnol.* **23**, 63–68 (2005).
2. Rogers, P.L., Jeon, Y.J., Lee, K.J. & Lawford, H.G. *Biofuels* **108** 263–288 (2007).
3. Farrell, A.E. *et al. Science* **311**, 506–508 (2006).
4. Zhang, M., Eddy, C., Deanda, K., Finkestein, M. & Picataggio, S. *Science* **267**, 240–243 (1995).
5. Jeffries, T.W. *Nat. Biotechnol.* **23**, 40–41 (2005).
6. Yang, S. *et al. BMC Genomics* **10**, 34 (2009).
7. Joachimsthal, E., Haggett, K.D., Jang, J.H. & Rogers, P.L. *Biotechnol. Lett.* **20**, 137–142 (1998).
8. Yablonski, M.D. *et al.* J. *Biotechnol.* **9**, 71–79 (1988).
9. Pappas, K.M., Galani, I. & Typas, M.A. *J. Appl. Microbiol.* **82**, 379–388 (1997).