The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*

Soon Ho Hong^{1,6}, Jin Sik Kim^{1,6}, Sang Yup Lee¹, Yong Ho In^{2,5}, Sun Shim Choi², Jeong-Keun Rih³, Chang Hoon Kim³, Haeyoung Jeong^{3,4}, Cheol Goo Hur³ & Jae Jong Kim⁴

The rumen represents the first section of a ruminant animal's stomach, where feed is collected and mixed with microorganisms for initial digestion. The major gas produced in the rumen is CO_2 (65.5 mol%), yet the metabolic characteristics of capnophilic (CO_2 -loving) microorganisms are not well understood. Here we report the 2,314,078 base pair genome sequence of *Mannheimia succiniciproducens* MBEL55E, a recently isolated capnophilic Gram-negative bacterium from bovine rumen, and analyze its genome contents and metabolic characteristics. The metabolism of *M. succiniciproducens* was found to be well adapted to the oxygen-free rumen by using fumarate as a major electron acceptor. Genome-scale metabolic flux analysis indicated that CO_2 is important for the carboxylation of phosphoenolpyruvate to oxaloacetate, which is converted to succinic acid by the reductive tricarboxylic acid cycle and menaquinone systems. This characteristic metabolism allows highly efficient production of succinic acid, an important four-carbon industrial chemical.

The rumen is a fermentation vat in which the feed is collected immediately after being swallowed and is digested through the action of microorganisms. The different microorganisms present in the bovine rumen are able to digest various substrates such as cellulose, hemicellulose, starch, sugar, protein and lipid. Cellulose and hemicellulose are converted to glucose, which is further converted by microbial fermentation to volatile fatty acids (VFAs) such as acetic, propionic and butyric acids¹. The VFAs generated in the rumen are absorbed through the rumen wall, and used for milk production and energy generation.

We have isolated a novel capnophilic bacterium, *Mannheimia* succiniciproducens MBEL55E, from bovine rumen². *M. succiniciprodu*cens is a non-spore-forming, mesophilic, gram-negative coccobacillus that produces VFAs under the anaerobic conditions present in the rumen. Most notably, it efficiently fixes CO₂ and produces substantial amounts of succinic acid².

Various carbohydrates produced during digestion of the feed (e.g., glucose from cellulose by cellulose-digesting microorganisms) are converted into succinic acid as well as acetic, formic and lactic acids by *M. succiniciproducens*². Acetic acid is a major VFA constituting 50%–60% of the total¹. A major gas present in the rumen is CO_2 (65.5 mol%), which is a by-product of microbial digestion of nutrients. Therefore, capnophilic bacteria may be important in regulating the overall fermentation profiles within the rumen. Succinic acid is a dicarboxylic acid produced as an intermediate of the tricarboxylic acid (TCA) cycle and also as one of the fermentation products of anaerobic metabolism. It can be used as an ion chelator, food additive,

supplement to pharmaceuticals, and as a green feedstock for the manufacture of environmentally friendly polymers, chemical intermediates and solvents³. Because succinic acid is produced by carboxylation of three-carbon (C_3)-compounds, capnophilic bacteria contribute most heavily to the production of succinic acid in the rumen. Succinic acid can be further converted to propionic acid, which constitutes about 20% of total VFAs, by other ruminal microorganisms¹. Even though the metabolic characteristics of cellulose-degrading microorganisms, acetic acid producers and methanogens have been extensively studied, those of rumen capnophilic bacteria have received little attention.

We have sequenced and analyzed the genome of a rumen capnophilic bacterium *M. succiniciproducens* MBEL55E. To understand the metabolic characteristics of *M. succiniciproducens*, we carried out genome-scale metabolic flux analyses (MFA) using the *in silico* metabolic pathways constructed directly from the genome sequence and actual fermentation profiles. Also, comparative genome analyses were carried out with several related bacteria including *Pasteurella multocida* Pm70, *Haemophilus influenzae* Rd, *Escherichia coli* K-12 and *Bacillus subtilis*. Based on these analyses, we suggest that *M. succiniciproducens* is a highly efficient succinic acid producer.

RESULTS

General genome features

The *M. succiniciproducens* genome is a single circular chromosome of 2,314,078 base pairs (bp) with no plasmid (**Fig. 1**). General features

Published online 19 September 2004; doi:10.1038/nbt1010

¹Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering, Department of BioSystems, BioProcess Engineering Research Center and Bioinformatics Research Center, Korea Advanced Institute of Science and Technology, 373–1 Guseong-dong, Yuseong-gu, Daejeon 305–701, Republic of Korea. ²Bioinfomatix, Inc., The fifth floor, Nam Chang Bldg., 748–162 Yeoksam-dong, Gangnam-gu, Seoul 135–925, Republic of Korea. ³Korea Research Institute of Bioscience and Biotechnology (KRIBB), 52 Oun-dong, Yuseong-gu, Daejeon 305–333, Republic of Korea. ⁴GenoTech Corp., 461–6 Jeonmin-dong, Yuseong-gu, Daejeon 305–390, Republic of Korea. ⁵IDREch Inc., 461–6 Jeonmin-dong, Yuseong-gu, Daejeon 305–390, Republic of Korea. ⁶These authors contributed equally to this work. Correspondence should be addressed to S.Y.L. (leesy@kaist.ac.kr).

ARTICLES

Figure 1 Circular representation of the M. succiniciproducens MBEL55E genome. The outermost circle indicates the chromosomal location in bp (each tick is 100 kb). The distribution of genes is represented by colored boxes according to the functional category in the COG database and direction of transcription (outer band indicates forward direction: inner band indicates reverse direction). Red and green arrows indicate the locations and direction of rRNA and tRNA genes, respectively. The colored inner circle represents the GC skew value (blue indicates positive value and red indicates negative value). The innermost circle shows the G+C content (higher values outward).

of the genome are summarized in **Table 1**. The average G+C content is 42.5%. The first base of the *gidA* gene was assumed to be the origin of replication based on the lower bound of the cumulative GC skew and the inversion of the GC skew value⁴ (see **Supplementary Fig. 1** online). Gene prediction and annotation identified 2,384 open reading frames (ORFs) with an average length of 873 bp; we assigned putative functions to 1,471 (61.7%) of the ORFs. Of the remaining 913 ORFs, 360 (15.1%) showed similarity to ORFs encoding hypothetical proteins of unknown function annotated in other gen-

omes, whereas 553 (23.2%) had no significant similarity (using the cutoff E-value of 10^{-5}) to data in the public databases (see **Supplementary Table 1** online). The total length of the protein coding region is 2,079,969 bp, which covers 89.9% of the chromosome (see **Supplementary Fig. 1** online). There were 19 rRNAs, including six 16S, six 23S and seven 5S rRNAs, as well as 60 tRNAs covering all 20 amino acids. No pseudogenes or spliced genes were found.

Comparative genome analysis

M. succiniciproducens is a nonpathogenic capnophilic rumen bacterium, which, on the basis of 16S rRNA sequence alignment, is considered closely related to the pathogenic strain P. multocida Pm70 (ref. 2). Each of the 2,384 ORFs of M. succiniciproducens was compared with the annotated genes in the public databases using BLASTP (Fig. 2). More than one-third (38.6%, 920 ORFs) of the ORFs showed the greatest similarity to the genes of P. multocida, and 21.7% (517 ORFs) to those of H. influenzae (Fig. 2). From these results, M. succiniciproducens was categorized into group H (Pasteurellaceae) of the clusters of orthologous groups (COG) database. The ORFs of M. succiniciproducens were functionally classified based on the COG database, and were compared with the ORFs of P. multocida, E. coli and B. subtilis in more detail (see Supplementary Table 2 online). The M. succiniciproducens genome had fewer ORFs than E. coli (4,288 ORFs) and B. subtilis (4,100 ORFs), but more than P. multocida (2,014 ORFs)5-7. The distribution of ORFs based on functional categories was similar to that of the most closely related strain, P. multocida. The numbers of ORFs involved in cell envelope biosynthesis (code M), cell motility (code N), signal transduction (code T), energy production and conversion (code C), carbohydrate transport and metabolism (code G), amino acid transport and metabolism (code E), and biosynthesis and catabolism (code Q)



were similar to those in *P. multocida*, but were much lower than those in *E. coli* or *B. subtilis. M. succiniciproducens* and *P. multocida* were found to possess relatively higher percentages of genes involved in translation, ribosomal structure and biogenesis (code J). The percentage and number of genes involved in signal transduction mechanisms were considerably lower than those in *E. coli* and *B. subtilis.*

Analysis of regional G+C content variation from the average G+C content of the genome revealed that the *M. succiniciproducens* genome contains one distinct region or genomic island. This genomic island of about 40 kb having a G+C content of 30% shows as much as 12.5% of G+C content deviation from the core genome. However, the 44 ORFs found in this island did not show significant homologies (using the cutoff E-value of 10^{-5}) with the annotated genes of any

Table 1 General features of *M. succiniciproducens* MBEL55E genome

Size (base pairs, bp)	2,314,078
G+C content (%)	42.5
Protein-coding sequences (CDS)	
No. similar to known genes	1,471
No. similar to genes of general function	193
No. similar to genes of unknown function	167
No. of genes not in COG database	553
Total CDSs	2,384
Coding density (%)	89.9
Average CDS size (bp)	873
No. of 16S-23S rRNA	6
No. of 5S rRNA	7
No. of tRNA	60

MBEL55E ORFs with other organisms. Each of the 2,384 predicted ORFs of M. succiniciproducens was used as a query sequence against the COG database. The inset shows the results of a dot plot comparison between the genome sequences of M. succiniciproducens and P. multocida. Abbreviations are: Rpr, Rickettsia prowazekii Madrid E; Pho, Pyrococcus horikoshii OT3; Pab, Pyrococcus abyssi; Mpn, Mycoplasma pneumoniae M129; jHp, Helicobacter pylori J99; Hal(Hbs), Halobacterium sp. NRC-1; Buc, Buchnera sp. APS; Mtu, Mycobacterium tuberculosis H37Rv (lab strain); Mja, Methanococcus jannaschii DSM2661; Hpy, Helicobacter pylori 26695; Mth, Methanobacterium thermoautotrophicum deltaH; Aae, Aquifex aeolicus VF5; Dra, Deinococcus radiodurans R1; Cje, Campylobacter jejuni NCTC11168; Syn, Synechocystis sp. PCC6803; Spy, Streptococcus pyogenes M1; Xfa, Xylella fastidiosa 9a5c; Vch, Vibrio cholerae El Tor

Figure 2 Comparison of M. succiniciproducens



N16961 (serotype 01); Lla, Lactococcus lactis IL1403; Tma, Thermotoga maritima MSB8; Ccr, Caulobacter crescentus; MIo, Mesorhizobium loti MAFF303099; Nma, Neisseria meningitidis Z2491 (serogroup A); Nme, Neisseria meningitidis MC58 (serogroup B); Bsu, Bacillus subtilis 168; Pae, Pseudomonas aeruginosa PA01; Bha, Bacillus halodurans C-125; Eco, Escherichia coli K-12 MG1655; EcZ(Ece), Escherichia coli 0157:H7 EDL933; Hin, Haemophilus influenzae Rd; Man-S, no homology or M. succiniciproducens-specific; Pmu, Pasteurella multocida PM70.

microorganism in the databases and thus there is no evidence of horizontal gene transfer.

M. succiniciproducens and *P. multocida*, which had the highest number of orthologous genes, were compared for their overall genome structure conservation. Even though only 2.60% of the *M. succiniciproducens* genome sequence perfectly matched the *P. multocida* genome sequence, there were weak linear relationships between the two genomes along the two crossed diagonal directions (**Fig. 2** and **Supplementary Fig. 1** online). Therefore, it can be suggested that these two species are closely related and originated from a common ancestor, but went through different evolutionary processes. Similar dot plot analysis was also carried out with *H. influenzae*, another member of the Pasteurellaceae family, which showed substantial homologies to *M. succiniciproducens* (see **Supplementary Fig. 1** online). The two species matched in 3.55% of their genome sequences, which is greater than the similarity of *M. succiniciproducens* with *P. multocida*, but no apparent conservation of linearity was found.

Gene duplication followed by mutation is an important mechanism for generating new genes and new biological functions to adapt to a particular environment. Genome analysis revealed that 255 genes in *M. succiniciproducens* were duplicated (see **Supplementary Table 2** online). Among them, 167 genes had one duplicate, 38 genes had two duplicates and 49 genes had more than two duplicates. The high copynumber genes were found to encode global transcriptional regulators or permeases of the major facilitator superfamily (ProP). For example, the *lysR*, a transcriptional regulator gene for lysine metabolism, had the most duplicates. The *proP* gene had 11 duplicates, which were distributed over the entire genome. Many of the genes involved in transcription, amino acid transport and metabolism, and carbohydrate transport and metabolism were duplicated.

Metabolism

The genes involved in metabolic pathways were analyzed in detail using the information present in KEGG⁸ and EcoCyc⁹ (see **Supplementary Table 3** online). Based on these results, a metabolic reaction map was constructed (see **Supplementary Fig. 2** online). Most of the

central metabolic pathway reactions were present, including almost complete Embden-Meyerhof-Parnas and pentose-phosphate pathways. The TCA cycle was also present, but the succinate dehydrogenase gene was not found. This indicates that M. succiniciproducens operates an incomplete or branched TCA cycle. It is notable that the Entner-Doudoroff pathway, which is not present in P. multocida⁷ or H. influenzae¹⁰, was also detected in M. succiniciproducens. A number of reactions for anaerobic pyruvate dissimilation, organic acid production and anaplerotic reactions were identified. The glyoxylate shunt was absent. At least three enzymes that interconvert C3 compounds and C4 compounds were found; phosphoenolpyruvate (PEP) carboxykinase and PEP carboxylase interconvert PEP and oxaloacetate, whereas the NADP⁺-dependent malic enzyme interconverts pyruvate and malate. It is notable that the NADP⁺-dependent malic enzyme (maeB) is present in M. succiniciproducens instead of the NAD⁺-dependent malic enzyme (sfcA) more frequently found in other organisms.

Most of the pathways for building-block biosynthesis were present, including those for amino acids. In the case of the nucleotide biosynthesis pathway, the initial two steps for synthesizing pyrimidines from aspartate and glutamine were missing. The genes involved in these steps are aspartate carbamoyltransferase (*pyrBI*) and dihydroorotase (*pyrC*), which are also missing in *P. multocida*⁷ and *H. influenzae*¹⁰. Orotate present in the growth medium can complement pyrimidine auxotrophy. The biosynthetic pathways for some cofactors and vitamins such as coenzyme A, NAD, biotin and thiamin were found to contain too many missing links, and therefore, may not function in *M. succiniciproducens*. Virulence-related genes such as the leukotoxin and capsule biosynthesis genes were not found.

M. succiniciproducens possesses several ABC transporters for the transport of carbohydrates, amino acids, long chain fatty acids, dipeptides/oligopeptides, Fe³⁺/cobalamin, nickel, manganese/zinc, thiamin, sulfate/molybdate and spermidine/putrescine (see **Supplementary Table 1** online). *M. succiniciproducens* also possesses the Sec machinery for the secretion of protein.

A number of genes involved in the transport of external carbon sources such as *mtlA*, *fruA*, *xylE* and *malX* were found in the genome.

ARTICLES



Figure 3 Results of metabolic flux analyses. (a,b) Graphical representation of normalized intracellular metabolic fluxes (mM/mM glucose) and fluxes not normalized [mM/gDCW/h] for CO_2 atmosphere/N_2 atmosphere (a) and $\text{CO}_2\text{-}\text{H}_2$ atmosphere/ CO_2 atmosphere (b) during the cultivation of M. succiniciproducens in MH2 medium containing 20 g/l glucose at 39 °C and pH 5.5-7.5. Red arrows indicate the fluxes that increased by more than twofold, whereas blue arrows indicate the fluxes that decreased to less than half. Orange rectangles indicate fluxes that were measured directly during the fermentations. (c) Profiles of representative intracellular fluxes of M. succiniciproducens MBEL55E obtained under N_2 , CO_2 and CO_2 -H₂ atmospheres at pH 5.5–7.5. The glucose consumption rate was not normalized (mM/gDCW/h), whereas other fluxes were normalized to the glucose consumption rate (mM/mM glucose) obtained under each condition. 6PG, 6-phosphoglucono-lactone; AC, acetic acid; AcCoA, acetyl-CoA; CIT, citrate; DHAP, dihydroxyacetone-phosphate; E4P, erythrose-4phosphate; F1,6P₂, fructose-1,6-diphosphate; F6P, fructose-6-phosphate; FOR, formic acid; FUM, fumaric acid; G3P, glyceraldehyde-3phosphate; ICT, isocitrate; α -KG, α -ketoglutarate; LAC, lactic acid; MAL, malic acid; OAA, oxaloacetate; PYR, pyruvate; Q, menaquinone; QH₂, menaguinol; R5P, ribose-5-phosphate; RL5P, ribulose-5-phosphate; SUC, succinic acid; Suc-CoA, succinyl-CoA; X5P, xylulose-5phosphate; DCW, dry cell weight.

N2 at pH 7.5 CO2 at pH 6.6 CO2 at pH 6.0 CO2 at pH 6.5 CO2 at pH 7.0 CO2 at pH 7.5 CO2 2t pH 7.5

PEP carboxylation flux

Ĕ

Ĕ

netabolic

mM alucose⁻

Mu

metabolic

nalized

mM alucose

M

(mM mM glucose⁻

at pH 6.5

at pH 6.0

÷ ~

2.0

1.5





To examine the range of carbon substrates M. succiniciproducens uses, we cultivated the bacterium on various carbon sources. As predicted from the genome sequence, M. succiniciproducens used glucose, fructose, xylose, sucrose, maltose, lactose and mannitol (see Supplementary Table 4 online). However, carbohydrate polymers such as cellulose and hemicellulose were not used by M. succiniciproducens, as predicted by the absence of the necessary enzymes. The mannitolpositive growth characteristics support the classification of this strain as a Mannheimia sp. in the Pasteurellaceae family, which also contains Acinobacillus, Pasteurella, Haemophilus and Lonepinella^{10–13}.

M. succiniciproducens was found to possess a pathway for the synthesis of menaquinone (vitamin K2), but not for ubiquinone (coenzyme Q). Ubiquinone is generally used for aerobic respiration, whereas menaquinone is used for anaerobic respiration with electron acceptors other than oxygen and nitrate, such as fumarate. The presence of a menaquinone system, operation of a branched TCA cycle and production of succinic acid even under aerobic conditions all suggest that M. succiniciproducens uses fumarate as an electron acceptor and possesses oxygen-independent metabolic characteristics.

To evaluate these hypotheses, we analyzed the respiratory system of M. succiniciproducens and compared it with those of other microorganisms. Respiratory networks of various microorganisms were constructed based on the COG database and were hierarchically clustered for comparison (see Supplementary Fig. 3 online). The respiratory system of M. succiniciproducens was most similar to those of P. multocida, H. influenzae, Vibrio cholerae, E. coli and Campylobacter jejuni. M. succiniciproducens lacks genes for nitrate (narI and narY) and oxygen (cyoA) respiration; it also lacks the genes for aerobic fumarase (fumA) and succinate dehydrogenase (sdhA) of the TCA cycle, which convert succinic acid to malic acid under aerobic conditions. The NADH dehydrogenase genes (nuoB, nuoK and ndh) were present in M. succiniciproducens, but not in P. multocida (see Supplementary Fig. 3 online). These results suggest that M. succiniciproducens evolved the respiratory system to adapt to the CO2-rich and O2-deficient rumen condition. As a capnophilic bacterium, M. succiniciproducens efficiently carboxylates PEP to form oxaloacetate, which is converted to fumarate via malate. Fumarate acts as a major electron acceptor as mentioned above, and is converted to succinic acid. This is strongly supported by the finding that M. succiniciproducens possesses anaerobic fumarase, fumarate reductase and NADH dehydrogenase, which form an efficient respiration network, whereas it lacks the key oxygen respiratory enzymes, succinate dehydrogenase and cytochrome oxidase bo3.

MFA

To determine the overall genome-scale metabolic characteristics of M. succiniciproducens, we carried out MFA using the in silico metabolic network directly constructed from the annotated genome. The in silico metabolic network of M. succiniciproducens consisted of 373 reactions (121 reversible and 252 irreversible) and 352 metabolites (see Supplementary Table 3 online). The rumen condition and content are dependent on rumen ecology and fermentation balance, both of which are affected by feeding. The gas composition is also an important factor that can influence the microbial population and activities in the rumen¹. The average composition of rumen gas is: CO₂, 65.5 mol%; CH₄, 26.8 mol%; N₂, 7 mol%; O₂, 0.5 mol%; and H₂, 0.2 mol%. Therefore, we examined the effects of changes in gas composition on the metabolism of M. succiniciproducens by using MFA constrained with actual data derived from cultivating cells under N₂, CO₂ and CO₂-H₂ atmospheres. We did not examine the effects of O2 and CH4 because M. succiniciproducens does not use them.

Similarly, the effects of pH were examined by cultivating cells at pH values of 5.5 to 7.5. We obtained the metabolic flux distributions under various conditions by linear optimization using these fermentation data as constraints (see Supplementary Fig. 4 online), and examined them by Bayesian network analysis.

As predicted above, M. succiniciproducens was found to operate a branched TCA cycle (Fig. 3a), which is usually observed in bacteria that favor anaerobic conditions. The effect of CO₂ on the metabolism of M. succiniciproducens was examined by comparing the MFA results obtained under CO₂ versus N₂ growth conditions (Fig. 3a). Under N₂ atmosphere, cell growth and glucose consumption rates were much lower, suggesting that M. succiniciproducens, like other capnophilic bacteria, relies on a CO2-dependent metabolism. The glycolytic flux under CO₂ was four times greater than that obtained under N₂. It is notable that the fluxes of PEP carboxylation and the reductive branch of the TCA cycle considerably increased under CO2. NAD+ was regenerated mostly through the reductive branch of the TCA cycle under CO2 atmosphere, suggesting that M. succiniciproducens uses a



Figure 4 Effects of various metabolic fluxes on the succinic acid fluxes examined by Bayesian network analysis. Intracellular flux distributions obtained under N_2 , CO_2 and CO_2 -H₂ atmospheres at pH 5.5, 6.0, 6.5, 7.0 and 7.5 were used as input data, and the data were separated by two parts (high and low) according to the frequency. Red and blue dotted lines indicate positive and negative effects, respectively. The values in tables are conditional probabilities indicating the probability that X is true while Y is true, which can be defined as:

$\mathsf{P}(\mathsf{X}|\mathsf{Y}) = \mathsf{P}(\mathsf{X},\mathsf{Y})/\mathsf{P}(\mathsf{Y})$

where P(X,Y) is the joint probability inferred from the product of individual conditional probabilities P(X) and P(Y), which represent probabilities that X and Y are true, respectively.

succinic acid pathway for NAD⁺ regeneration in the rumen where CO_2 is abundant.

The effect of H₂ on the intracellular flux distribution was also evaluated by comparing the MFA results obtained under CO2-H2 (molar ratio of 1:1) versus CO₂ (Fig. 3b). When H₂ was present, cell growth rate and the glycolytic flux were lower. However, the relative succinic acid flux increased by 30% under CO₂-H₂, and the relative formation rates of ATP and NADH were higher in the presence of H₂. The MFA results indicated that M. succiniciproducens consumes 2.8 moles of CO₂ and 0.16 moles of H₂ for the production of 4.0 moles of succinic acid. The succinic acid yield obtained under CO₂-H₂ was as high as 1.3 moles succinic acid/mole glucose, which is 70% of the theoretical succinic acid yield. Until now, it has been suggested that methanogens are responsible for the consumption of H_2 to form methane from CO_2 in the rumen¹⁴. However, the MFA results suggest that M. succiniciproducens, and possibly other capnophilic bacteria, may be important in controlling the rumen atmosphere through consumption of CO2 and H2 in conjunction with production of organic acids.

Examination of the overall metabolic flux profiles (**Fig. 3c**) obtained under various culture conditions suggests the following conclusions on the metabolic characteristics of *M. succiniciproducens*. The normalized glycolytic fluxes did not change much under varying culture conditions, suggesting the essential role of glycolytic fluxes on the metabolic housekeeping activities of *M. succiniciproducens* by producing 50%–70% of total ATP and 77%–85% of NADH. The normalized fluxes of the TCA cycle and amino acid biosynthesis under CO_2 were lower than those under N_2 . On the other hand, the normalized succinic acid flux considerably increased under CO_2 , which coincides with the increase of PEP carboxylation flux.

Bayesian network analysis showed that the CO_2 consumption rate, PEP carboxylation flux and succinic acid flux showed a direct relationship with one another, suggesting that PEP carbolxylation is a major CO_2 -consuming, C_3 - C_4 link in *M. succiniciproducens* leading to NAD⁺ regeneration and succinic acid production (**Fig. 4**). Besides CO_2 , the glucose consumption rate and NADH formation rate also exerted a positive effect on succinic acid production. Malic enzyme flux did not affect succinic acid flux.

DISCUSSION

More than 145 microbial genome projects have been completed. They have mainly focused on the characterization of virulence-related genes and specific pathways or enzymes. Only recently, genome-scale in silico models are beginning to appear for E. coli^{15,16} and several other microorganisms¹⁷. In this study, we were able to decipher through the genome-scale MFA how the metabolism of M. succiniciproducens evolved and how it is regulated in response to variations in the conditions in the rumen. M. succiniciproducens possesses an efficient system for the carboxylation of PEP to oxaloacetate, which is further converted to succinic acid through the reductive TCA cycle and the menaquinone system. It was predicted by MFA that M. succiniciproducens can produce as much as 1.71 and 1.86 moles of succinic acid from 1 mole of glucose under CO2 and CO2-H2 atmospheres, respectively. Therefore, it will be possible to use this bacterium for the efficient production of succinic acid, an important industrial chemical that can be used as a green feedstock for the manufacture of biodegradable polymers, synthetic resins and various chemical intermediates and additives³. In conclusion, we were able to decipher the metabolic characteristics of M. succiniciproducens directly from the genome sequence by genome-scale MFA based on actual fermentation data. Based on these findings, we can now design metabolic engineering strategies for the enhanced production of succinic acid; one such strategy will be increasing the PEP carboxylation flux while decreasing the fluxes to acetic, formic and lactic acid. This approach will prove useful in developing metabolic engineering strategies for the production of various bioproducts by other sequenced organisms.

METHODS

Genome sequencing strategy. The M. succiniciproducens MBEL55E (KCTC 0769BP; the Korean Collection for Type Cultures) genome was shotgun sequenced¹⁸. E. coli DH5 α (supE44, Δ lacU169 (ϕ 80 lacZ Δ M15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1) was used for genomic DNA library construction. M. succiniciproducens was grown in Luria-Bertani medium¹⁹. Nine plasmid libraries of small inserts (3-4 kb) were constructed using pGEM-3Zf(+) (Applied Biosystems) and pUC18 (Amersham Biosciences) after random mechanical shearing of genomic DNA. A large insert library (~38.5 kb) was constructed using the fosmid pEpiFOS-5 (Epicentre) after size selection of genomic DNA without additional shearing. Plasmid DNA was extracted from the cells grown in 96-well plates by the alkaline lysis method followed by purification using glass fiber-based filter plates. Sequencing was carried out using an ABI prism 3700 DNA analyzer (Applied Biosystems). A total of 36,820 reads with an average length of 614 bp (\sim 9.7-fold coverage) were obtained by bidirectional sequencing. A total of 530 end reads from 275 fosmid clones served as genome scaffolds. The sequences were assembled using PHRAP (http://www.phrap.org/). Visual inspection of assembly results, primer design for gap closing and sequence editing were done using CONSED²⁰. Possible misassemblies were corrected by validity checking on the directions and distances between the mate reads. Sequence gaps were closed by primer walking on plasmid or fosmid clones (1,875 primer walk reads), and physical gaps were closed by direct sequencing of genomic DNA or PCR products. The accuracy of the final consensus sequence was 99.987%, including 1.8% edited bases.

ORF prediction and annotation. ORF prediction and annotation were done using the JAVA-based airBASE software package (KAIST-Bioinfomatix) installed on an IBM server, which integrated GLIMMER²¹, ARTEMIS²², BLAST²³ and Hidden Markov Models (HMM) modules²⁴. Automatic analysis of each gene candidate was done using BLAST with COG data²⁵ and HMM modules. The cut-off E-values of 10⁻⁵ and 10⁻¹ were used in the BLAST and HMMpfam analyses, respectively. The tRNAs were predicted by tRNAscan-SE software²⁶. Genes assigned to multiple COGs were further analyzed to determine their primary functional categories.

Dot matrix plot analysis. The dot matrix plots comparing two genomes were generated by Genalysis software (http://www.genetix.com/productpages/Software/Genalysis.htm). The output displays a 2-D plot representing matched regions between two genomes by colored dots. The minimum size of matched sequences was set to 20 bp.

Metabolic flux and computational analyses. MFA was carried out for the calculation of volumetric rates of intracellular metabolite formation using the *in silico* metabolic network constructed directly from the genome sequence. Because the number of reactions was greater than the number of metabolites, the system was underdetermined. Therefore, MFA was carried out by optimization using linear programming with an objective function of achieving the maximum growth rate²⁷:

 $\begin{array}{l} \mbox{Maximize (shown in wt\%): } 0.495 \times \mbox{protein} + 0.25 \times \mbox{RNA} + 0.041 \times \mbox{DNA} \\ + 0.091 \times \mbox{lipids} + 0.034 \times \mbox{lipopolysaccharides} + 0.025 \times \mbox{petidoglycan} + \\ 0.025 \times \mbox{glycogen} + 0.004 \times \mbox{polyamines} + 0.035 \times \mbox{cofactors.} \end{array}$

In the above equation, the contents of protein, RNA and DNA in *M. succiniciproducens* were measured values, whereas those of other macromolecules were based on average *E. coli* cellular contents²⁸. MFA was carried out using the program package MetaFluxNet²⁹. The glucose uptake rate and the production rates of lactic, acetic, formic and succinic acids experimentally measured were used as the constraints during the MFA.

Even though flux analysis using linear programming is effective to describe the overall metabolic characteristics, there may be several flux scenarios when the objective function and/or applied constraints are altered. To examine if the metabolic characteristics deciphered by flux analysis above are reliable, we repeated flux analyses using several suboptimal growth rates as objective functions; we varied the objective values from 100% to 90% of the maximum objective value and monitored changes in flux distributions. Most of anabolic fluxes decreased with decreasing objective value, whereas the central metabolic fluxes in glycolysis and C3-C4 linkages were less affected by the changes in the objective value. Most importantly, the activation of alternative fluxes and reversal of flux direction were not observed during the suboptimal flux analysis, which suggests that the results obtained by flux analysis based on linear programming represent consistent and reliable information on the intracellular metabolism of *M. succiniciproducens*.

For the comparative analysis of respiratory systems in various microorganisms, hierarchical clustering was performed using the Cluster program package³⁰ with the average-linkage hierarchical clustering algorithm. The results were visualized using the TreeView software package³⁰.

Culture conditions. Cells were routinely grown in a CO₂ atmosphere in sealed anaerobic bottles containing 100 ml of MH medium² plus 10 g/l of glucose. For flask cultures, exponentially growing cells, washed anaerobically with the MH medium, were used to inoculate sealed anaerobic bottles containing 100 ml of MH medium plus 10 g/l of glucose. Batch cultures were kept at 39 °C in a jar fermentor (2.5l; KoBioTech) containing 1 liter of MH2 medium plus 20 g/l of glucose and 1 mg/l of Na₂S·9H₂O. MH2 medium contains per liter: 5 g yeast extract, 3 g K₂HPO₄, 1 g NaCl, 1 g (NH₄)₂SO₄, 0.2 g CaCl₂·2H₂O and 0.2 g MgCl₂·6H₂O. To examine the effect of pH on metabolism, the pH was controlled at a desired value between 5.5 and 7.5 using 5 N NaOH. Cells were cultured under three different atmospheric conditions: CO2, CO2-H2 mixture (molar ratio of 1:1) or N2 (ref. 2). Gases, scrubbed free of oxygen by a gas purifier (P.J. Cobert Associates), were sparged at 0.25 liter min⁻¹ during the cultivation. When bacteria were cultured under CO2 and CO2-H2 atmospheres, 7 g of Na₂CO₃ was added to prevent culture acidification. The agitation speed was 200 r.p.m.

Analytical procedure. The concentrations of glucose and fermentation products were determined by HPLC (Hitachi L–3300 RI detector, D2500 chromato–integrator) equipped with an ion exchange column (Aminex HPX–87H, 300 mm \times 7.8 mm) using 0.012 N H₂SO₄ as a mobile phase². Cell growth was monitored by measuring the absorbance at 660 nm (OD₆₆₀; Ultrospec3000, Pharmacia Biotech). The protein concentration was measured by the Bradford method³¹. The DNA and RNA contents were determined using the Burton and orcinol methods, respectively³².

Nucleotide sequence accession number. The sequence reported in this paper has been deposited in GenBank under the accession number AE016827.

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

ACKNOWLEDGMENTS

We thank Sung Ho Goh and Heyung Ju Shin at the KRIBB, Young Ho Moon, Kang Ryul Choi, Sun Ho Cha, Sung Soo Kim, Soo Hyun Jeong, Eun Mi Chung and Sun Rye Jung at GenoTech, and Jinyoung Park, Sujin Chae and Hee Sun Chung at Bioinfomatix for their contributions during genome sequencing and annotation. This work was supported by the Korean Systems Biology Research Program (M10309020000-03B5002-00000) of the Ministry of Science and Technology (MOST), Bioinfomatix and by the Brain Korea 21 Project. Further support through the LG Chem Chair Professorship and IBM SUR program is appreciated.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 7 June; accepted 3 August 2004 Published online at http://www.nature.com/naturebiotechnology/

 Sniffen, C.J. & Herdt, H.H. The Veterinary Clinics of North America: Food Animal Practice (W.B. Saunders Company, Pennsylvania, USA, 1991).

- Lee, P.C., Lee, S.Y., Hong, S.H. & Chang, H.N. Isolation and characterization of a new succinic acid-producing bacterium, *Mannheimia succiniciproducens* MBEL55E, from bovine rumen. *Appl. Microbiol. Biotechnol.* 58, 663–668 (2002).
- Lee, S.Y., Hong, S.H., Lee, S.H. & Park, S.J. Fermentative production of chemicals that can be used for polymer synthesis. *Macromol. Biosci.* 4, 157–164 (2004).
- Salzberg, S.L., Salzberg, A.J., Kerlavage, A.R. & Tomb, J.F. Skewed oligomers and origins of replication. *Gene* **217**, 57–67 (1998).
- Blattner, F.R. et al. The complete genome sequence of Escherichia coli K–12. Science 277, 1453–1462 (1997).
- Kunst, F. et al. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature 390, 249–256 (1997).
- May, B.J. et al. Complete genomic sequence of Pasteurella multocida, Pm70. Proc. Natl. Acad. Sci. USA 98, 3460–3465 (2001).
- Kanehisa, M., Goto, S., Kawashima, S. & Nakaya, A. The KEGG databases at GenomeNet. Nucleic Acids Res. 30, 42–46 (2002).
- Karp, P.D. Pathway databases: a case study in computational symbolic theories. Science 293, 2040–2044 (2001).
- Fleischmann, R.D. et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269, 496–512 (1995).
- 11. Angen, Ø., Mutters, R., Caugant, D.A., Olsen, J.E. & Bisgaard, M. Taxonomic relationships of the [*Pasteurella*] haemolytica complex as evaluated by DNA-DNA hybridizations and 16S rRNA sequencing with proposal of Mannheimia haemolytica gen. nov., comb. nov., Mannheimia granulomatis comb. nov., Mannheimia glucosida sp. nov., Mannheimia ruminalis sp. nov. and Mannheimia varigena sp. nov. Int. J. Syst. Bacteriol. 49, 67–86 (1999).
- Osawa, R. et al. Lonepinella koalarum gen. nov., sp. nov., a new tannin-protein complex degrading bacterium. Syst. Appl. Microbiol. 18, 368–373 (1995).
- Guettler, M.V., Rumler, D. & Jain, M.K. Actinobacillus succinogenes sp. nov., a novel succinic-acid-producing strain from the bovine rumen. Int. J. Syst. Bacteriol. 49, 207– 216 (1999).
- Church, D.C. The Ruminant Animal: Digestive Physiology and Nutrition (Prentice Hall, New Jersey, USA, 1988).
- Ibarra, R.U., Edwards, J.S. & Palsson, B.O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silco* predicted optimal growth. *Nature* **420**, 186–189 (2002).
- Edwards, J.S., Ibarra, R.U. & Palsso, B.O. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 19, 125–130 (2001).
- Price, N.D., Papin, J.A., Schilling, C.H. & Palsson, B.O. Genome-scale microbial *in silico* models: the constraints-based approach. *Trends Biotechnol.* **21**, 162–169 (2003).
- Favello, A., Hillier, L. & Wilson, R.K. Genomic DNA sequencing methods. *Methods Cell Biol.* 48, 551–569 (1995).
- Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
- Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. Genome Res. 8, 195–202 (1998).
- Delcher, A.L., Harmon, D., Kasif, S., White, O. & Salzberg, S.L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27, 4636–4641 (1999).
- Rutherford, K. et al. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945 (2000).
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990).
- 24. Eddy, S.R. Profile hidden Markov models. Bioinformatics 14, 755-763 (1998).
- Tatusov, R.L. et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29, 22–28 (2001).
- Lowe, T.M. & Eddy, S.R. tRNAscan–SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964 (1997).
- Varma, A. & Palsson, B.O. Metabolic flux balancing—basic concepts, scientific and practical use. *Bio/Technology* 12, 994–998 (1994).
- Neidhardt, F.C. & Umbarger, H.E. in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edn. 2 (ed. Neidhardt, F.C.) 13–16 (ASM Press, Washington, DC, 1996).
- Lee, D.Y., Yun, H., Park, S. & Lee, S.Y. MetaFluxNet: the management of metabolic reaction information and quantitative metabolic flux analysis. *Bioinformatics* 19, 2144–2146 (2003).
- Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95, 14863–14868 (1998).
- Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254 (1976).
- Burton, K. A study of the conditions and mechanism of the diphenylamine reaction for the colorimetric estimation of deoxyribonucleic acid. *Biochem. J.* 62, 314–323 (1956).