

RESEARCH

Open Access

# Genome-scale metabolic network reconstruction and *in silico* flux analysis of the thermophilic bacterium *Thermus thermophilus* HB27

Na-Rae Lee<sup>1†</sup>, Meiyappan Lakshmanan<sup>2†</sup>, Shilpi Aggarwal<sup>2</sup>, Ji-Won Song<sup>1</sup>, Iftekhhar A Karimi<sup>2</sup>, Dong-Yup Lee<sup>2,3\*</sup> and Jin-Byung Park<sup>1,4\*</sup>

## Abstract

**Background:** *Thermus thermophilus*, an extremely thermophilic bacterium, has been widely recognized as a model organism for studying how microbes can survive and adapt under high temperature environment. However, the thermotolerant mechanisms and cellular metabolism still remains mostly unravelled. Thus, it is highly required to consider systems biological approaches where *T. thermophilus* metabolic network model can be employed together with high throughput experimental data for elucidating its physiological characteristics under such harsh conditions.

**Results:** We reconstructed a genome-scale metabolic model of *T. thermophilus*, iTT548, the first ever large-scale network of a thermophilic bacterium, accounting for 548 unique genes, 796 reactions and 635 unique metabolites. Our initial comparative analysis of the model with *Escherichia coli* has revealed several distinctive metabolic reactions, mainly in amino acid metabolism and carotenoid biosynthesis, producing relevant compounds to retain the cellular membrane for withstanding high temperature. Constraints-based flux analysis was, then, applied to simulate the metabolic state in glucose minimal and amino acid rich media. Remarkably, resulting growth predictions were highly consistent with the experimental observations. The subsequent comparative flux analysis under different environmental conditions highlighted that the cells consumed branched chain amino acids preferably and utilized them directly in the relevant anabolic pathways for the fatty acid synthesis. Finally, gene essentiality study was also conducted via single gene deletion analysis, to identify the conditional essential genes in glucose minimal and complex media.

**Conclusions:** The reconstructed genome-scale metabolic model elucidates the phenotypes of *T. thermophilus*, thus allowing us to gain valuable insights into its cellular metabolism through *in silico* simulations. The information obtained from such analysis would not only shed light on the understanding of physiology of thermophiles but also helps us to devise metabolic engineering strategies to develop *T. thermophilus* as a thermostable microbial cell factory.

**Keywords:** *Thermus thermophilus*, Thermophile, Genome-scale metabolic model, Constraints-based flux analysis, Ethanol

\* Correspondence: cheld@nus.edu.sg; jbpark06@ewha.ac.kr

†Equal contributors

<sup>2</sup>Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585, Singapore

<sup>1</sup>Department of Food Science & Engineering, Ewha Womans University, 11-1 Daehyun-dong, Seodaemun-gu, Seoul 120-750, Korea

Full list of author information is available at the end of the article

## Background

*Thermus thermophilus* is a gram-negative, obligate aerobic bacterium, representing one of the best-studied thermophiles. It usually colonizes the terrestrial volcanic hot springs (grows optimally between 65 and 72°C) and was originally isolated from a Japanese thermal spa [1]. In addition to the ability of surviving at such high temperatures, *T. thermophilus* is resistant to other stress such as harsh chemical conditions [2]. These properties motivated researchers to extract or isolate numerous proteins from *T. thermophilus*, making it as a model organism in structural genomics with significant industrial potential [3-6]. For example, several thermostable proteins are already used in commercial processes, including the DNA polymerase in PCR techniques,  $\alpha$ -amylases and glucose isomerases in starch processing, esterases, lipases and proteases in organic synthesis, and xylanases in paper and pulp manufacturing [7,8]. Moreover, *T. thermophilus* is being recognized as a potential microbial cell factory for the low cost ethanol fermentation from lignocellulosic waste materials since it can grow by utilizing most of the C5/C6 carbon sources at relatively high temperatures, i.e. 70–80°C, thus reducing the energy costs: no cooling step is required following enzymatic hydrolysis, rendering it easier to distil subsequent fermentations [9].

Despite enormous potentials for biotechnological applications, the current knowledge on the unique cellular physiology of *T. thermophilus* is very limited; to date, the production of distinctive carotenoid molecules [10] and the use of adaptive protein synthesis strategies [11] are only two notable traits unravelled at the molecular level. Such limited studies are mostly due to the technical difficulties in cultivating and analysing thermophilic microbes; cell culture experiments require high amount of energy to maintain the optimal growth conditions. Hence, it is indeed required to develop more systematic approaches for improving our understanding of *T. thermophilus* cellular behaviour. In this regard, constraints-based *in silico* metabolic modeling and analysis can be considered as one of the promising techniques to characterize the physiological behaviour and metabolic states of an organism upon various environmental/genetic changes as they systematically capture the genotype-phenotype relationships from the entire genome information [12,13]. As a result, several genome-scale metabolic models (GSMMs) are now available for describing the metabolic organization of various organisms including *Escherichia coli* [14], *Bacillus subtilis* [15], *Saccharomyces cerevisiae* [16], *Pichia pastoris* [17], *Corynebacterium glutamicum* [18], *Ralstonia eutropha* [19], *Pseudomonas aeruginosa* [20], and even for multicellular eukaryotes such as *Mus musculus* [21] and *Homo sapiens* [22]. Moreover, with the availability of several conveniently accessible constraints-based modeling software tools [23], these models have been largely

utilized to postulate various strain improvement strategies [17-19,24,25]. Thus, the development of *T. thermophilus* GSMM based on the currently available biochemical and genomic information and its subsequent *in silico* analysis enables us to elucidate its unique metabolic behaviour.

In thermophilic microbes regard, there have been only a few initial attempts to model their cellular metabolisms. First, an *in silico* model of *Thermotoga maritima* was presented, covering its central metabolism along with the 3D structures of all the enzymes accounted in the network [26]. Recently, the genome-scale metabolic model of thermophilic archeon, *Sulfolobus solfataricus*, was also developed, and used to describe its autotrophic growth in bicarbonate via hydroxypropionate-hydroxybutyrate cycle under aerobic conditions [27]. However, both models are not mature enough to explain the molecular mechanisms of high temperature adaptations as they do not consider the detailed biosynthetic machinery of biomolecules which help them to retain the integrity of their cell wall membranes. Therefore, in this work, we reconstructed the genome-scale metabolic model of *T. thermophilus* based on the genome annotation of HB27 wild-type strain [28] for investigating unique characteristics of thermophilic microbes. Additionally, the model was functionally characterized by gene essentiality studies to identify essential genes for cellular growth while growing in both glucose minimal and amino acid supplemented complex media.

## Results

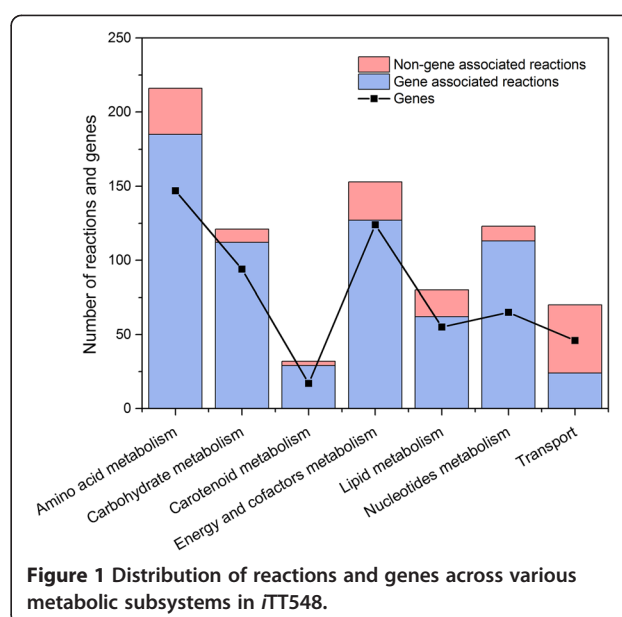
### Reconstruction of *T. thermophilus* genome-scale metabolic network

The genome-scale metabolic network of *T. thermophilus* HB27 was reconstructed through a three step procedure: (1) construction of draft network via compilation of genes, reactions and pathway information from biochemical databases based on the genome annotation of *T. thermophilus* HB27, (2) manual curation of the draft model by verifying the elemental balances in reactions and assigning proper gene-reaction relationships, and (3) gap filling using organism specific knowledge (see Methods). During the reconstruction process, significant efforts were highly required to identify and resolve the network gaps across various metabolic pathways. Such gaps exist due to the incomplete genome annotations which result in missing biochemical reactions and dead ends. These gaps can be appropriately filled by the addition of new reactions based on information obtained from the literature or inferred by the genome annotation of other organisms. For example, the initial model contained several metabolic gaps in the synthetic pathway of thermozeaxanthin and thermobiszeaxanthin, the unique type of carotenoids that are found only in thermophiles, enabling their cellular membrane to retain its fluidity even at very high temperatures [29-31]. Therefore, in order to fill such gaps, we added the reactions

corresponding to glycosyltransferase and acyltransferase enzymes from *Staphylococcus aureus* subsp. *aureus* 11819–97 and *Halobacillus halophilus* DSM 2266 for the sequential glycosylation and esterification of zeaxanthin with glucose and branched-chain fatty acids, producing thermozeaxanthin. Similarly, the draft model also had several gaps in the carbon assimilatory pathways, and thus was unable to consume six carbon sources, namely, trehalose, palatinose, isomaltose, cellobiose, glutamate and mannose. However, earlier studies have reported that *T. thermophilus* can grow on all these carbon sources [32,33]. Such discrepancies were again resolved by adding new reactions corresponding to the  $\alpha$ -glucosidase and mannokinase enzymes based on the information available from KEGG and MetaCyc databases. Overall, we included 74 new reactions for 63 enzymes, thereby improving its network connectivity (see Additional file 1 for complete list of reactions added). The gap filling of draft model was followed by the identification of genetic evidence for the newly added enzymes via sequence-based homology searches. For this purpose, a BLASTp search was performed in NCBI database for the enzymes that could resolve the network gaps using their amino acid sequences collected from various other organisms against the non-redundant protein sequences of *T. thermophilus* HB27 genome. In such a way, of the 63 enzymes added, we could assign a putative locus for 10 of them, thus providing possible new annotations (see Additional file 1 for the list of new annotations). The final genome-scale metabolic network of *T. thermophilus*, *i*TT548, contains 548 unique genes (ORF coverage – 24%), 796 reactions and 635 unique metabolites. In *i*TT548, all the 796 reactions were classified into seven major metabolic subsystems: carbohydrates, amino acid, energy and cofactors, lipids, nucleotides, carotenoids and transport. Among them, amino acid metabolism has largest number of reactions and genes, followed by carbohydrates and energy and cofactors metabolism (Figure 1). The detailed list of completely curated *T. thermophilus* HB27 metabolic network containing the various genes, reactions, and metabolites can be obtained from Additional file 1, and also available as Systems Biology Markup Language (SBML) file (level 2, version 1, <http://sbml.org/>) (Additional file 2).

#### Network characteristics of *i*TT548 and its comparison with *E. coli* GSMM

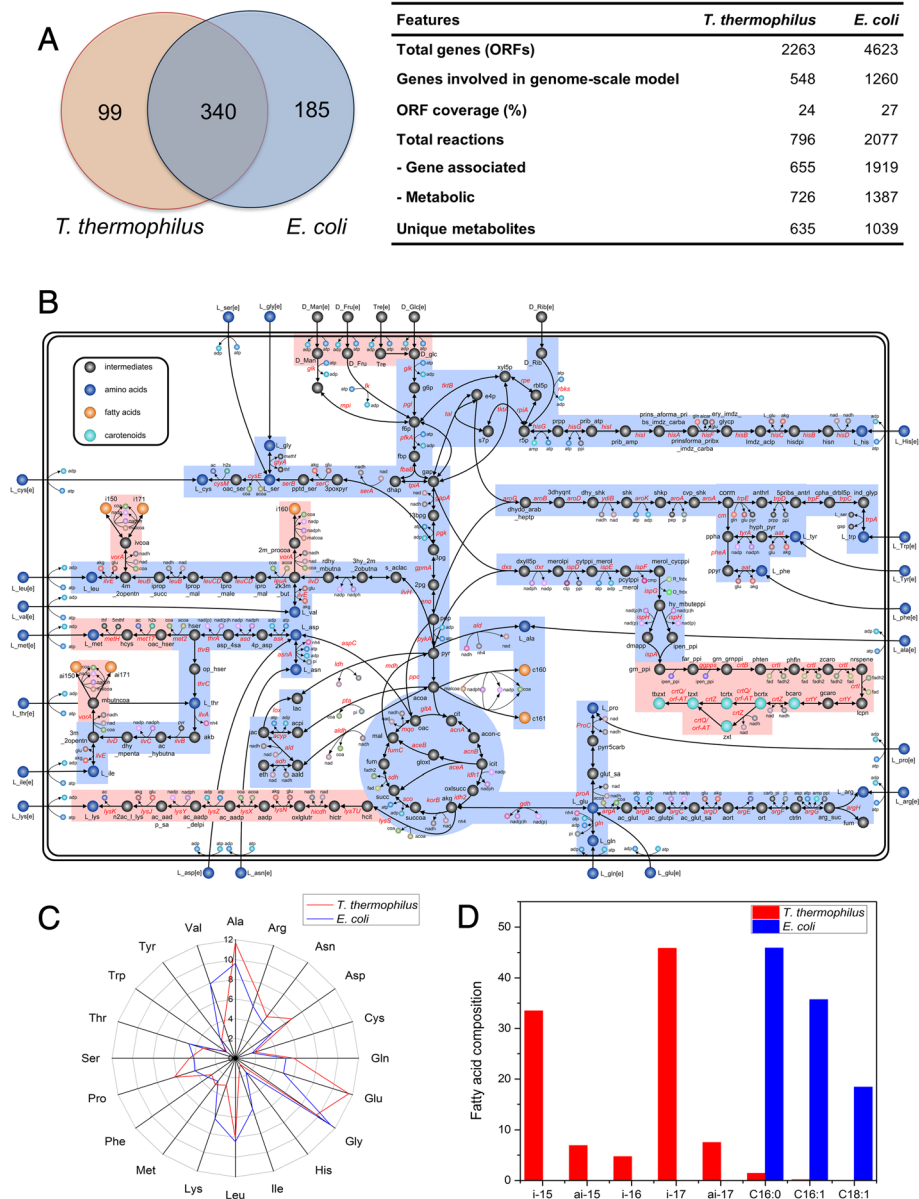
Figure 2A presents the overall network features of *i*TT548 and its comparison with *E. coli* using the metabolic model (*i*AF1260) [14] in terms of the EC numbers. Here, it should be noted that we were unable to fairly compare the metabolic characteristics of *T. thermophilus* with other thermophiles since the *T. maritima* model is only limited to the central metabolism while *S. solfataricus* GSMM was not accessible and no link was provided. From the



**Figure 1** Distribution of reactions and genes across various metabolic subsystems in *i*TT548.

comparison, there are 363 enzymes/genes are common between *E. coli* and *T. thermophilus*, mostly belonging to the central metabolic pathways such as glycolysis, pentose phosphate pathway and the TCA cycle. However, certain notable differences were observed in the amino acid synthetic pathways of *T. thermophilus*. For instance, lysine is synthesized via the alpha-amino adipate pathway instead of diaminopimelate pathway. Similarly, the upstream of methionine synthetic pathway is not conserved with *E. coli*: the precursor molecule, homocysteine, is produced from O-acetyl-L-homoserine and hydrogen sulphide via O-acetyl-L-(homo) serine sulphydrylase in *T. thermophilus* and from O-succinyl-L-homoserine and cysteine through O-succinyl-homoserine lyase in *E. coli*. The comparison between transport reactions further revealed that *T. thermophilus* lacks the phosphoenolpyruvate-dependent phosphotransferase system (PTS), a typical bacterial transport system, and thus consumes most of the carbohydrates including glucose via ATP-binding cassette (ABC) transporters.

Interestingly, as a unique feature of *T. thermophilus*, *i*TT548 contains the necessary biosynthetic machinery for synthesizing several molecules which help them in habituating high temperatures. Unlike many other gram-negative bacteria, *T. thermophilus* does not contain lipopolysaccharides in cell outer membrane [34]. Instead, it embeds complex carotenoid glucoside esters with various branched chain fatty acids, known as thermozeaxanthins and thermobiszeaxanthins, in the lipid bilayers. Such an arrangement offers multiple advantages including the retention of membrane fluidity at high temperatures and reduction of oxygen diffusion through the membrane for preventing oxidation damage [10,35]. Furthermore,



**Figure 2** Metabolic organization and biomass composition of *T. thermophilus* and *E. coli*. **(A)** General features of the *iTT548* in comparison with *E. coli* *iAF1260* GSMM (Feist et al. 2007), **(B)** Central metabolic network of *T. thermophilus*, **(C)** amino acid composition (mol%) and **(D)** fatty acid composition (mol%). The numbers in the Venn diagram represents the enzymes in each organism. The common and unique pathways of *T. thermophilus* are highlighted with blue and red backgrounds, respectively. The number of unique and common enzymes was identified using the EC numbers. The biomass data for *E. coli* was obtained from *iAF1260* GSMM. See supplementary 1 for metabolite and enzyme abbreviations used in the network diagram.

*T. thermophilus* synthesizes several unique polyamines such as thermine, spermine, thermospermine and caldo-pentamine using a distinct pathway from L-arginine via aminopropyl agmatine [36]. These polyamines are essential for high temperature protein synthesis by ensuring the proper structure formation of the initiation complex among 30S ribosomal subunit, the messenger, and the initial aminoacyl-tRNA. As a notable exception to gram-negative bacterium, *T. thermophilus* also synthesizes

branched chain fatty acids from amino acids such as valine, leucine and isoleucine via ketoisovalerate oxidoreductase (*vorA*) as in gram-positive bacteria such as *Bacillus*. Table 1 list the unique carotenoid and polyamine molecules, and relevant genes of the corresponding synthetic pathways accounted in *iTT548*; Figure 2 illustrates the central metabolic network of *T. thermophilus* where the branched chain fatty acids and carotenoid synthetic pathways are highlighted.

**Table 1 Biosynthetic machinery of unique molecules in *T. thermophilus***

Molecules	No. of reactions	Genes involved	Major metabolic precursors
<b>Branched fatty acids</b> (i-15:0, ai-15:0, i-16:0, i-17:0, ai-17:1)	19	<i>ilvE</i> (TTC1870), <i>vorA</i> (TTC1756) and <i>vorB</i> (TTC1757), <i>FabF</i> (TTC0049 or TTC0045), <i>FabG</i> (TTC0047 or TTC0394), <i>FabZ</i> (TTC1463), <i>FabI</i> (TTC0343)	Leucine, valine, isoleucine and acetyl-coA
<b>Carotenoids</b> (thermocytochrome, thermozeaxanthin and thermobiszeaxanthin)	18	<i>CrtE</i> (TTC1986), <i>CrtB</i> (TT_P0057), <i>CrtI</i> (TT_P0066), <i>CrtY</i> (TT_P0060), <i>CrtZ</i> (TT_P0059), <i>CruC</i> (TT_P0062), <i>CruD</i> (TT_P0061)	Isopentenyl diphosphate (IPP), UDP-glucose and branched fatty acids
<b>Polyamines</b> (1,3-diaminopropane, norspermidine, sym-homospermidine, thermine, thermospermine, homospermine, caldopentamine, thermopentamine, homocaldopentamine, caldohexamine, homocaldohexamine, tris(3-aminopropyl)amine, tetrakis(3-aminopropyl)ammonium)	16	<i>SpeA</i> (TTC1277), <i>SpeB</i> (TTC0764 or TTC1132), <i>SpeD1</i> (TTC0473), <i>SpeD2</i> (TTC1093), <i>SpeE</i> (TTC0472)	Putrescine, spermidine and spermine

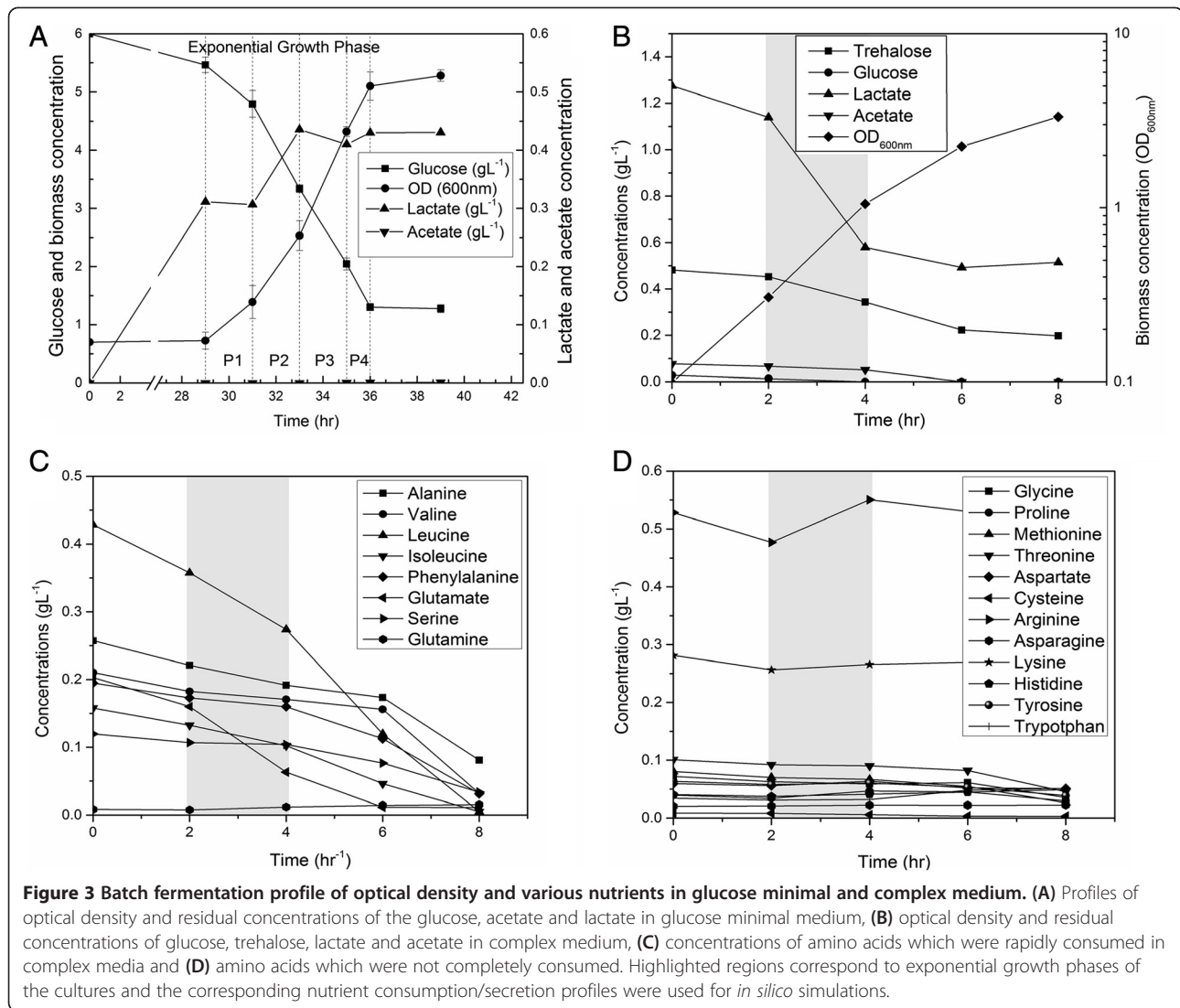
In *iTT548*, we have also included a biomass equation based on our amino acid compositional analysis and the data obtained from literature. Importantly, such biomass composition must be carefully formulated to avoid any erroneous conclusions from the flux balance analysis [37]. It should be noted that the earlier thermophile models, *T. maritima* and *S. solfataricus*, adopted the biomass equation from *E. coli* and *M. bakeri*, respectively, both grow at 37°C which is well below the optimal growing temperature range of thermophiles (50 ~ 80°C). In this regard, the compositions of some amino acids, valine, lysine, threonine, lysine, glutamine and isoleucine, are very distinctive between *T. thermophilus* and *E. coli* (Figure 2C). Similarly, the fatty acid compositions of *T. thermophilus* were also different from *E. coli*; the linear chain fatty acids composition are almost negligible in *T. thermophilus* while the branched-chain fatty acids such as iso-C17:0 and iso-C15:0, which are not present in *E. coli*, contribute the bulk of total lipid compositions (Figure 2D). Collectively, these results highlight the need for the careful estimation of biomass equation while modeling thermophiles.

#### Model validation using minimal and complex media during batch cultures

We validated *iTT548* using data from the batch cultures of *T. thermophilus* growing in glucose minimal and complex media. In case of glucose minimal medium, cells were cultured in a DMM containing 0.6% (w/v) glucose at 70°C. The residual concentration of glucose and the cell density were monitored (Figure 3A). Initially, a prolonged lag phase of 2 d was observed, followed by an exponential growth phase of 8 h before the stationary phase was reached. The cell cultures also indicated the presence of acetate, lactate and ethanol in trace amounts during lag and exponential growth phases. In order to analyse the growth behaviour of *T. thermophilus* in a rich medium, the cells were grown in a complex TM medium

at 70°C. It should be noted that this medium was supplemented with all 20 amino acids. The nutrient consumption profiles indicated that glucose was consumed first. Subsequently, other carbohydrates such as trehalose and amino acids were assimilated (Figure 3B). Notably, cells did not consume all the amino acid supplemented in the medium and preferred branched chain amino acids ahead of other amino acids, possibly to synthesize the branched chain fatty acids (Figures 3C and D). Furthermore, the complex medium did not show any appreciable lag phase and the cells grew almost twice as fast as the minimal media.

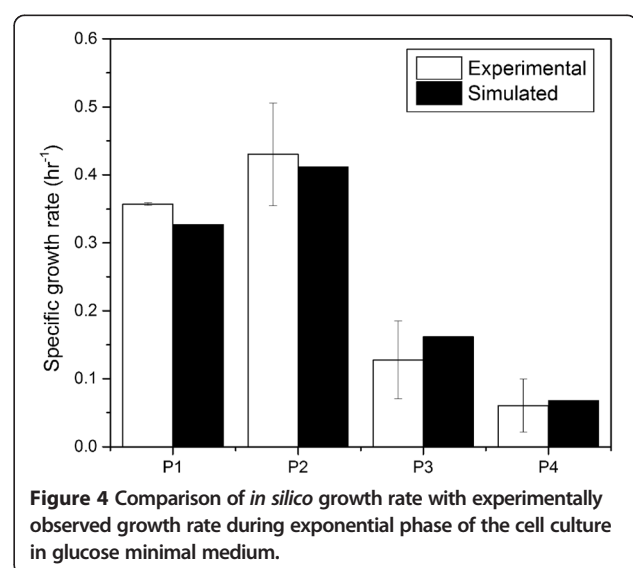
To simulate the cellular growth in the minimal medium, the biomass equation was maximized in the flux analysis simulations while simultaneously constraining the glucose uptake rates measured during the exponential phase based on the assumption that wild-type organisms typically evolve towards the maximization of cellular growth during exponential phase [38]. The exchange fluxes of NH<sub>3</sub>, phosphate, sulphite, H<sub>2</sub>O, Fe<sup>2+</sup>, Mg and H<sup>+</sup> were left unconstrained to provide basic nutrients and minerals for cell growth. The oxygen uptake rate was constrained at the average specific uptake rate of 10 mmol g<sup>-1</sup> DCW hr<sup>-1</sup> based on previous publication [39]. Additionally, the lactate exchange flux was also constrained at the measured uptake/secretion rates in each phase. A growth associated maintenance (GAM) value of 58.34 mmol g<sup>-1</sup> DCW h<sup>-1</sup>, from *E. coli* GSMM, and a NGAM requirement of 14 mmol g<sup>-1</sup> DCW h<sup>-1</sup>, calculated based on established methods [14], were also used for the simulations (see Additional file 3 for detailed calculation of NGAM calculations). Here, it should be noted that while comparing the NGAM requirement of *T. thermophilus* with *E. coli* [14], it possess a very high value (14 compared to that of 8 mmol g<sup>-1</sup> DCW h<sup>-1</sup>) possibly due to the differences in H<sup>+</sup> permeability of their cytoplasmic membrane at optimal growth temperature; *T. thermophilus* cytoplasmic membrane is highly permeable to H<sup>+</sup>, and thus, leaks protons



without ATP synthesis via ATPase [40]. Notably, simulation results were highly consistent with observed growth rates (Figure 4). For complex media, once again the biomass equation was maximized while simultaneously constraining the uptake/secretion rates of all nutrients (glucose, trehalose, lactate, acetate and amino acids) during the exponential growth phase. The *in silico* predicted cell growth of  $0.66 \text{ h}^{-1}$  was very close (within the acceptable error range of 10%) to the experimentally observed specific growth rate of  $0.64 \text{ h}^{-1}$ , thus clearly indicating the high predictive ability of *iTT548* even in complex media.

#### *In silico* comparative metabolic flux analysis of minimal and complex media

When comparing the cellular growth rates of *T. thermophilus* in minimal and complex media, not surprisingly, it was significantly higher in the latter owing to the availability of



rich carbon and nitrogen sources in the form of various amino acids. Since the microbe can take up some of the amino acids directly from the medium, it could be possible that the protein biosynthetic demand can be partially fulfilled. In order to confirm such hypothesis, we compared the simulated metabolic fluxes between minimal (phase P1) and complex media, and observed much lower fluxes through the relevant biosynthetic reactions of amino acids in the complex medium (Table 2). Herein, it should be noted that we also conducted the flux variability analysis [41] to re-assure the confidence of the simulated metabolic states in each environmental condition.

Interestingly, although all 20 amino acids were supplemented in the complex medium, the cells consumed only a few of them preferably. In order to understand this cellular behavior, we compared the uptake rate of individual amino acids with their actual biosynthetic demands. Methionine and cysteine were consumed from the medium according to their biosynthetic demands while alanine, valine, leucine, isoleucine and glutamate were consumed in excess of their individual biosynthetic requirements (Table 3). Further analysis revealed that the surplus valine/leucine/isoleucine and alanine/glutamate were utilized to synthesize the branched chain fatty acids and other amino acids, respectively. Notably, leucine was consumed much more than its biosynthetic demand (almost 7 times higher) mainly to synthesize iso-15 and iso-17 fatty acids, which constitute the major composition in *T. thermophilus*. Remarkably, this amino acid utilization pattern is completely different from *E. coli* grown in complex media, which consumed serine, aspartate, glycine and threonine ahead of other amino acids [42]. The subsequent *in silico* analysis in *E. coli* revealed that the rapidly depleted serine was primarily converted into pyruvate via serine deaminase, and then to acetate for producing more ATP as similarly observed in several other microbes such as *C. glutamicum* and *Lactococcus lactis*. In this regard, it is interesting to note that *T. thermophilus* possess unique nutrient consumption pattern, preferably synthesizing branched chain fatty acids rather than improving energy production further. Overall, as the *in silico* analysis highlights that all consumed amino acids contribute either directly or indirectly towards biomass synthesis, strategies to increase the uptake of non-consumed amino acids such as tyrosine, lysine, tryptophan and histidine can be postulated to enhance the cellular growth in complex medium.

#### **Analysis of essential genes *T. thermophilus***

We analysed the essentiality of individual genes of the *T. thermophilus* under glucose minimal and complex TT medium conditions using *iTT548* model by deleting every gene one at a time (see Methods). The genes were then categorized into three classes: (i) completely essential – genes which are required for cellular growth in both the

media, (ii) conditionally essential – required only in one of the media and (iii) non-essential – dispensable in both the media. The gene essentiality study revealed that 23.5% and 19.5% of the total 548 genes in *iTT548* are essential for cell growth in minimal and complex media, respectively (see Additional file 4 for complete list of essential genes). A total of 107 genes were essential in both conditions and an additional 21 genes were essential only in minimal media. In order to further understand the knock-out of which functional category of genes are more crucial for cell viability in either condition, we identified the distribution of essential genes across various metabolic processes (Figure 5). Interestingly, the carotenoids metabolism contained most of the lethal genes (82%), suggesting that this unique pathway synthesizing the thermozeaxanthin and thermobiszeaxanthin does not have many alternative routes and is quite rigid in *T. thermophilus*. Following carotenoids metabolism, the nucleotides and lipids metabolism has second and third highest number of completely essential genes (32.7% and 24.6%, respectively). On the other hand, examination of conditionally essential genes, i.e. genes which are essential only in minimal media, revealed that amino acid metabolism contains almost all such genes (Figure 5). Since the complex media is supplemented with all the amino acids, some of them were directly consumed from the media without utilizing their biosynthetic pathways, thus classifying the genes from those pathways as non-essential under such conditions. However, at the same time, distribution of certain completely essential genes in amino acid metabolism also indicates that biosynthetic pathways of tryptophan, proline and tyrosine are crucial for the cell growth since they should be synthesized within the cell albeit their availability in the complex medium. Interestingly, our gene deletion analyses also showed a high number of completely essential genes within the oxidative phosphorylation and TCA cycle. Oxygen is the key electron acceptor in cellular metabolism which can accept electrons from other redox cofactors involved in the TCA cycle and generates energy through oxidative phosphorylation. If oxygen is devoid or any of the oxidative phosphorylation and TCA cycle reactions are perturbed, generally, most of the bacteria regenerate the redox cofactors by switching to fermentative growth with the help of substrate level phosphorylation. In this regard, since *T. thermophilus* lacks the PTS and utilizes ABC transporters even for glucose uptake, it could be possible that the microbe cannot switch to fermentative metabolism completely and thus, requires oxygen to generate sufficient energy for cell growth via TCA cycle and oxidative phosphorylation.

#### **Discussion**

Thermophilic microbes represent a unique class of organisms with a distinct cell wall assembly that enables them

**Table 2 Comparison of metabolic reaction fluxes of amino acids biosynthetic reactions between minimal and complex media**

Reaction ID	Enzyme name (Gene)	Reaction details	Complex media (LB < L < UB)	Minimal media (LB < L < UB)
<b>R192</b>	<b>Alanine dehydrogenase (<i>ald</i>)</b>	<b>H[c] + NADH[c] + PYR[c] + NH4[c] &lt;=&gt; L_ALA[c] + NAD[c] + H2O[c]</b>	<b>-0.031 &lt; -0.031 &lt; -0.031</b>	<b>0.234 &lt; 0.234 &lt; 2.416</b>
R521	Asparagine synthase ( <i>asnB</i> )	L_ASP[c] + L_GLN[c] + ATP[c] + H2O[c] <=> L_GLU[c] + L_ASN[c] + PPi[c] + H[c] + AMP[c]	-1.808 < -1.808 < -1.783	-0.537 < -0.537 < -0.522
R198	Aspartate aminotransferase ( <i>aspC</i> )	L_ASP[c] + AKG[c] <=> L_GLU[c] + OAC[c]	-1.782 < -1.782 < -1.782	-0.416 < -0.416 < -0.416
R205	Glutamine synthetase ( <i>gln</i> )	L_GLU[c] + ATP[c] + NH4[c] -> L_GLN[c] + ADP[c] + Pi[c] + H[c]	-1.793 < -1.793 < -1.793	-0.474 < -0.474 < -0.474
R202	Glutamate dehydrogenase ( <i>gdh</i> )	L_GLU[c] + NADP[c] + H2O[c] <=> H[c] + NADPH[c] + AKG[c] + NH4[c]	-9.830 < -9.145 < 4.713	-9.288 < -8.813 < 2.977
<b>R227</b>	<b>Cysteine synthase (<i>cysM</i>)</b>	<b>OAC_SER[c] + H2S[c] -&gt; L_CYS[c] + H[c] + AC[c]</b>	<b>0 &lt; 0 &lt; 0</b>	<b>0 &lt; 0.005 &lt; 0.005</b>
<b>R235</b>	<b>Methionine synthase (<i>metH</i>)</b>	<b>L_HCYS[c] + 5MTHF[c] -&gt; L_MET[c] + H[c] + THF[c]</b>	<b>0 &lt; 0 &lt; 0</b>	<b>0 &lt; 0.009 &lt; 0.009</b>
R272	N2-acetyl-L-lysine deacetylase ( <i>lysK</i> )	H2O[c] + N2AC_L_LYS[c] -> L_LYS[c] + AC[c]	0.229 < 0.229 < 0.229	0.05 < 0.05 < 0.05
R651	3-phosphoserine phosphatase ( <i>serB</i> )	H2O[c] + PPTD_SER[c] -> L_SER[c] + Pi[c]	0 < 0 < 0	0 < 1.072 < 2.545
R655	Serine hydroxymethyltransferase ( <i>glyA</i> )	L_SER[c] + THF[c] -> L_GLY[c] + H2O[c] + METHF[c]	0 < 0 < 0	0 < 0.512 < 6.42
R213	Serine hydroxymethyltransferase ( <i>glyA</i> )	L_GLY[c] + H2O[c] + 5_10_MNTHF[c] <=> L_SER[c] + THF[c]	0.135 < 0.135 < 0.135	-1.787 < -0.442 < 3.993
R212	Tryptophan synthase ( <i>trpA</i> )	L_SER[c] + IND_GLYP[c] <=> L_TRP[c] + H2O[c] + GAP[c]	0.054 < 0.054 < 0.054	0.026 < 0.026 < 0.026
R221	Threonine synthase ( <i>thrC</i> )	H2O[c] + OP_HSER[c] -> L_THR[c] + Pi[c]	0.0714 < 0.0714 < 0.0714	0.063 < 0.063 < 0.063
<b>R245</b>	<b>Branched-chain amino acid aminotransferase (<i>ilvE</i>)</b>	<b>L_VAL[c] + AKG[c] &lt;=&gt; L_GLU[c] + 2K3M_BUT[c]</b>	<b>0.113 &lt; 0.113 &lt; 0.113</b>	<b>-0.068 &lt; -0.068 &lt; -0.068</b>
<b>R244</b>	<b>Branched-chain amino acid aminotransferase (<i>ilvE</i>)</b>	<b>L_GLU[c] + 4M_2OPENTN[c] &lt;=&gt; L_LEU[c] + AKG[c]</b>	<b>-1.179 &lt; -1.179 &lt; -1.179</b>	<b>0.154 &lt; 0.154 &lt; 0.154</b>
<b>R246</b>	<b>Branched-chain amino acid aminotransferase (<i>ilvE</i>)</b>	<b>L_ILE[c] + AKG[c] &lt;=&gt; L_GLU[c] + 3M_2OPENTN[c]</b>	<b>0.478 &lt; 0.478 &lt; 0.478</b>	<b>-0.024 &lt; -0.024 &lt; -0.024</b>
R341	Proline-5-carboxylate reductase ( <i>ProC</i> )	L_PRO[c] + NADP[c] <=> 2 H[c] + NADPH[c] + PYRR5CARB[c]	-0.432 < -0.432 < -0.432	-0.119 < -0.119 < -0.119
R195	Argininosuccinate lyase ( <i>argH</i> )	L_ARG_SUCC[c] -> L_ARG[c] + FUM[c]	1.144 < 1.144 < 1.144	0.132 < 0.132 < 0.132
<b>R297</b>	<b>Aspartate aminotransferase (<i>aat</i>)</b>	<b>L_PHE[c] + AKG[c] &lt;=&gt; L_GLU[c] + PPYR[c]</b>	<b>0 &lt; 0 &lt; 0</b>	<b>-0.054 &lt; -0.054 &lt; -0.054</b>
<b>R284</b>	<b>Histidinol dehydrogenase (<i>hisD</i>)</b>	<b>2 NAD[c] + H2O[c] + L_HISN[c] -&gt; L_HIS[c] + 3 H[c] + 2 NADH[c]</b>	<b>0.025 &lt; 0.025 &lt; 0.025</b>	<b>0.015 &lt; 0.015 &lt; 0.015</b>

The reactions which are highlighted in bold possess less flux in complex media. FVA was conducted to re-assure the simulated fluxes and the results are provided with the attainable lower and upper bounds identified from this analysis.

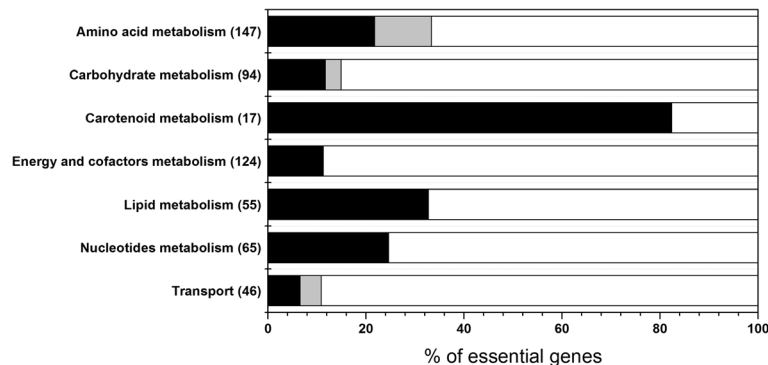


**Table 3 Consumption or production pattern of amino acids in complex media**

Amino acids	Supply from medium	Biosynthetic demand	Contribution		Remarks
			To others	From others	
Alanine	0.7399	0.355	0.3853		Surplus alanine contributes to glycine synthesis
Glycine	0.116	0.308		0.1918	Some are produced from alanine
Valine	0.224	0.109	0.1151		Excess valine is utilized in synthesis of anteiso-17:1 and anteiso-15:0 fatty acids
Leucine	1.427	0.244	1.1827		Excess leucine is utilized in iso-15:0 and iso-17:0 fatty acids synthesis
Isoleucine	0.5181	0.039	0.4794		Excess isoleucine is utilized in synthesis of anteiso fatty acid synthesis
Proline	-0.24	0.189		0.4291	Synthesized from glutamate
Methionine	0.015	0.015			Utilized from media as per biosynthetic demand
Serine	0.057	0.135		0.0785	Synthesized from glycine
Threonine	0.03	0.100		0.0699	Synthesized from aspartate
Phenylalanine	0.086	0.086			Utilized from media as per biosynthetic demand
Aspartate	-0.12	0.103		0.2232	Synthesized from oxaloacetate and glutamate
Cysteine	0.0087	0.009			Utilized from media as per biosynthetic demand
Glutamate	1.48	0.176	1.3041		Excess glutamate is utilized in several other amino acid synthesis
Arginine	-0.93	0.211		1.1409	Synthesized from aspartate
Asparagine	-0.023	0.103		0.1262	Synthesized from aspartate
Lysine	-0.141	0.087		0.2283	Synthesized from oxoglutarate
Glutamine	0	0.177		0.1771	Synthesized from glutamate
Histidine	0	0.024		0.0245	Synthesized from PRPP
Tyrosine	-0.03204	0.069		0.1007	Synthesized from PEP, E4P
Tryptophan	-0.01116	0.043		0.0538	Synthesized from serine

to maintain the cellular membrane integrity even at very high temperatures. It is reported that many thermophilic organisms, including *T. maritima* and *S. solfataricus*, synthesize ether lipids from long chain dicarboxylic fatty acids and fatty alcohols [43]. However, *Thermus* sp. do not synthesize ether lipids but produce unique carotenoid molecules such as thermozeaxanthin and thermobizeaxanthin, and embed them in the lipid bi-layer to attain the required cellular membrane fluidity at high

temperatures [10]. In this regard, *i*TT548 completely captures all the biosynthetic pathways of thermozeaxanthins, in addition to the metabolic routes of other biomass precursors such as amino acids, nucleotides and lipids. Similarly, *i*TT548 also contains the unique biosynthetic pathways of several unusual polyamines which help in stabilizing the nucleotide strands and proteins synthesis at high temperatures. It has been earlier reported that *T. thermophilus* is unique in polyamine synthesis: even



**Figure 5 Distribution of essential genes in *T. thermophilus* metabolic subsystems.** Black, grey and white colors indicate the completely-, conditionally- and non-essential genes, respectively. The numbers within the parenthesis represent the number of genes in each subsystem.

the extreme thermophiles such as *S. solfataricus* produces relatively shorter polyamines [36]. Collectively, these results clearly show the detailed metabolic coverage of *i*TT548 of thermophiles when comparing with its preceding GSMMs.

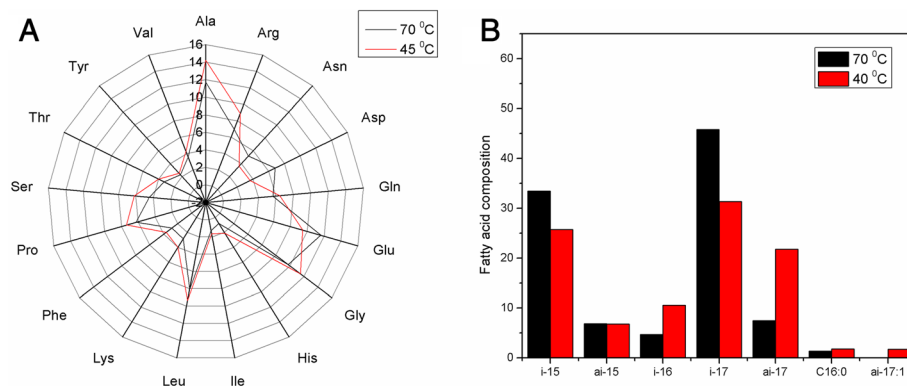
Furthermore, this work includes a prudently drafted biomass equation that is specific to thermophiles, especially *Thermus* sp. As mentioned earlier, the comparative analysis of *T. thermophilus* and *E. coli* biomass compositions have highlighted significant differences between amino acid and fatty acid compositions. Noticeably, the *T. thermophilus* biomass analysis revealed that the concentration of some of the thermolabile amino acids such as threonine and histidine are substantially lesser than *E. coli* whereas the proline concentration is much higher. It should be highlighted that these observations are in good agreement with earlier reports which suggested the selective usage of amino acid residues as one of the key adaptive strategy employed by thermophiles [44,45]. Arguably, the cellular compositions in thermophilic microbes may change depending on growth temperature; the current biomass equation was derived based on compositional analysis of *T. thermophilus* grown at 70°C. In order to clarify the temperature dependent compositional change in biomass, we measured amino acid compositions in *T. thermophilus* at 45°C. Their comparison with compositional data at 70°C clearly indicated that there is no significant difference in both individual and overall amino acid concentrations (Figure 6A). Similarly, we also compared the fatty acid compositions between 40°C and 70°C using the data from literature [30]. Very interestingly, unlike amino acid comparison, fatty acid compositions, both overall and individual were much lower at 40°C (Figure 6B). Although we were not able to make a complete comparison between low and high growth temperatures since no data was available on other cellular constituents such as peptidoglycans and thermotolerant carotenoids at

low temperature range, we can still hypothesize that thermophiles are most likely to adjust their biomass composition selectively to better adapt to the growth environment. Therefore, the use of appropriate biomass equations for simulating the cellular growth in corresponding temperature ranges is crucial for reliable prediction.

The gene deletion analyses of *i*TT548 have revealed several interesting traits about the function of deleted genes with respect to overall cellular metabolism of *T. thermophilus*. Among them, the most notable is the relatively high percentage of essential genes when compared to *E. coli* (23% to that of 13%), possibly due to the smaller Open Reading Frames (ORF) content (only 2,263 as compared to 4,623) despite possessing all the necessary modules for the cell to be viable at high temperatures. Furthermore, this observation also highlights the fact that since *T. thermophilus* thrives at higher temperatures than most other microbes, its fitness might be relatively less competitive with more rigid network organization. Interestingly, the gene essentiality analyses also indicated that the carotenoids metabolism is functionally quite fragile since almost all of its genes are essential for cellular growth. However, it should be noted that the gene deletion analysis results are sensitive to several parameters such as *in silico* medium setup and biomass composition. In this regard, the current biomass composition is obtain from *T. thermophilus* at optimum growth temperatures, i.e. 70°C, and thus the gene deletion results of the current study are only applicable to this condition.

## Conclusions

We presented the genome-scale metabolic network of *T. thermophilus*, *i*TT548, the first ever representing thermophiles, containing 548 unique genes, 796 reactions and 635 unique metabolites. As a unique feature of *T. thermophilus*, *i*TT548 contains necessary metabolic pathways for synthesizing several unique carotenoids



**Figure 6** Influence of temperature on *T. thermophilus* biomass composition. (A) amino acid composition (mol%) at 70°C and 45°C and (B) fatty acid composition (mol%) at 70°C and 40°C.

and polyamines which help them in habituating high temperatures. The reconstructed metabolic model was subsequently validated with the batch culture experiments on glucose minimal and complex medium where the *in silico* growth predictions of the *i*TT548 were in good agreement with the observed experimental results. The comparative flux analysis between minimal and complex media highlighted that the consumption and utilization of branched chain amino acids directly in the relevant fatty acids anabolic pathways, thus resulting in higher growth rates in the rich medium. A gene essentiality study was also conducted through *in silico* simulation studies in both minimal and complex media, highlighting a very high percentage of lethal genes in comparison with *E. coli*, suggesting that the metabolic backbone of *T. thermophilus* could be quite rigid. Overall, the metabolic network presented in the current study is expected to be a significant contribution towards systems analysis of thermophiles where the metabolic model can be utilized along with high throughput datasets for the better understanding of organism.

## Methods

### Microorganism and culture conditions

*T. thermophilus* HB27 strain was used as a model organism. For fermentation in complex medium, a single colony was cultivated overnight at 70°C with 150 rpm in 5 mL of the TM medium [46], and the culture was transferred to a 500 mL baffled-flask containing 100 mL of TM broth. In case of cultivation in defined glucose minimal medium, the overnight seed grown at 70°C with 150 rpm in 5 mL of the TM medium was then transferred to a 500 mL baffled-flask containing 100 mL of defined minimal medium (DMM) with 0.6% (w/v) glucose and cultivated for 24 hours. Then, 5 mL of flask culture in DMM was inoculated to 100 mL of fresh DMM. During fermentations, cell growth was monitored by measuring the optical density at 600 nm. The dry cell weight (DCW) was then estimated by a predetermined conversion factor of 0.34.

### Analytical methods

Concentrations of glucose, organic acids and ethanol in the culture broth were measured by high performance liquid chromatography (HPLC) (Waters, Milford, MA) equipped with an HPX-87H column (Bio-Rad, Hercules, CA), a dual  $\lambda$  absorbance detector. The collected samples were centrifuged at 14,000 *g* and 4°C for 5 min and the supernatant was analyzed with the column using 5 mM sulfuric acid as a mobile phase at 0.6 mL<sup>-1</sup> min. Concentrations of amino acids were determined by gas chromatograph/mass spectrophotometer (GC/MS) (Agilent, Santa Clara, CA) equipped with an HP-5MS column (Agilent), as previously reported [47]. In brief, the samples were centrifuged, dried and derivatized with methyl-N-t-

butyldimethylsilyl-trifluoro-acetamide (MBDSTFA) in DMF at 80°C for 30 min. After centrifugation at 14,000 *g* for 5 min, the supernatant was injected to GC/MS in split injection mode (1:10 split ratio).

### Metabolic network reconstruction

The genome-scale metabolic network of *T. thermophilus* HB27 was reconstructed using the published genome annotation [28] and the information collected from various biological and genomic databases on the basis of the established procedure [48]. First, an initial draft model was constructed by compiling the annotated metabolic genes and their corresponding biochemical reactions from KEGG [49] and MetaCyc [50]. Then, these reactions were corrected for any elemental imbalances and mapped with appropriate genes to devise proper gene-protein-reaction (GPR) relationships. Additionally, some spontaneous as well as non-gene-associated reactions including metabolite transport were also incorporated into the model based on the physiological evidence from literature and databases. The connectivity of the draft network was then checked using the GapFind algorithm to find the gaps [51]. The identified missing links were filled either by introduction of sink reactions to allow for material exchange between the cell and its surrounding environment or by adding reactions from other similar microbes to close the knowledge gaps.

### Biomass composition

Cellular biomass composition is an important prerequisite for the *in silico* flux analysis, especially during the exponential growth phase, where the primary cellular objective is to maximize growth. Amino acid composition of *T. thermophilus* HB27 was estimated by hydrolyzing the cell pellets with 6 N HCl for 24 h at 130°C, and subsequently analysing the hydrolysates using HPLC equipped with UV-detector and C18 column. Cell wall and lipid compositions were obtained from previous publications on *Thermus* sp. [30,31,35,52]. The overall DNA and RNA composition was assumed to be same as *E. coli* [14] since no data was available on *Thermus* sp. The individual weights of nucleotides in the DNA and RNA were calculated based on the reported G + C content of 69.4% [28]. Detailed information on biomass composition calculations could be found in Additional file 3.

### Constraints-based flux analysis

We implemented constraints-based flux analysis to simulate the *T. thermophilus* metabolism under varying environmental conditions. The biomass reaction was maximized to simulate the exponential growth phase as described elsewhere [53-55]. Mathematically, the optimization problem, i.e. maximization of biomass subjected

to stoichiometric and capacity constraints, can be formulated as follows:

$$\begin{aligned} \max Z &= \sum_j c_j v_j \\ \text{s.t. } \sum_j S_{ij} v_j &= 0 \quad \forall \text{ metabolite } i \\ v_j^{\min} &\leq v_j \leq v_j^{\max} \quad \forall \text{ reaction } j \end{aligned}$$

where  $S_{ij}$  refers to the stoichiometric coefficient of metabolite  $i$  involved in reaction  $j$ ,  $v_j$  denotes to the flux or specific rate of metabolic reaction  $j$ ,  $v_j^{\min}$  and  $v_j^{\max}$  represent the lower and upper limits on the flux of reaction  $j$ , respectively; and  $Z$  corresponds to the cellular objective as a linear function of all the metabolic reactions where the relative weights are determined by the coefficient  $c_j$ . In this study, the constraints-based flux analysis problems were solved using COBRA toolbox [56].

### Flux variability analysis

As constraints-based flux analysis is an optimization based technique, it is often possible to have multiple flux distributions attaining the same physiological state. Therefore, in order to confirm the plausibility of internal metabolic fluxes simulated in minimal and complex media by flux analysis, we performed the flux variability analysis (FVA) to identify the possible range of all fluxes while simulating a particular phenotypic state. Mathematically, the optimization problem specific to FVA can be represented as follows:

$$\begin{aligned} \max / \min \quad & v_j \\ \text{s.t. } \quad & \sum_j S_{ij} v_j = 0 \\ & \sum_j c_j v_j = Z_{obj} \\ & v_j^{\min} \leq v_j \leq v_j^{\max} \quad \text{for } j = 1, \dots, n \end{aligned}$$

where  $Z_{obj}$  denotes the value of objective calculated by flux analysis and  $n$  is the number of fluxes. The upper range of fluxes is identified by maximizing the objective whereas the lower range is obtained by minimizing the same. In this study, the FVA was implemented using COBRA toolbox.

### Gene deletion analysis

Gene deletion simulations were performed by maximizing the cellular biomass while constraining flux through the corresponding reaction(s) to be zero via the GPR relationships under defined nutrient uptake rates. In case of glucose minimal medium, only glucose was fueled as carbon source. On the other hand, glucose, trehalose and amino acids such as valine, leucine and isoleucine

were supplied as carbon source based on the complex media based on nutrient consumption profile. The simulation results were subsequently analyzed to identify the essential genes where a gene is classified to be essential if the resulting cell growth prediction for the corresponding mutant is less than or equal to 5% of wild-type. Note that all the gene deletion analysis in this study was performed using COBRA toolbox.

### Additional files

**Additional file 1: Details of *IT548* containing all genes, reactions, metabolites.** A list of reactions added during gap-filling and possible new annotations identified in this study are also provided.

**Additional file 2: SBML file of *IT548*.**

**Additional file 3: Biomass composition of *T. thermophilus* HB27 and NGAM calculations.**

**Additional file 4: List of essential genes in glucose minimal and complex media.**

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

NRL, ML, JBP and DYL conceived and designed the study. NRL and JWS performed the batch culture experiments. NRL and SA created the draft model. ML refined the model and performed simulations. NRL, ML, IAK, JBP and DYL wrote the manuscript. JBP and DYL coordinated and directed the project. All authors have read and approved the final manuscript.

### Acknowledgements

This work was supported by the National University of Singapore, Biomedical Research Council of A\*STAR (Agency for Science, Technology and Research), Singapore, Korea Research Foundation (KRF-2010-0009169), Republic of Korea, and a grant from the Next-Generation BioGreen 21 Program (SSAC, No. PJ009520), Rural Development Administration, Republic of Korea.

### Author details

<sup>1</sup>Department of Food Science & Engineering, Ewha Womans University, 11-1 Daehyun-dong, Seodaemun-gu, Seoul 120-750, Korea. <sup>2</sup>Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585, Singapore. <sup>3</sup>Bioprocessing Technology Institute, Agency for Science, Technology and Research (A\*STAR), 20 Biopolis Way, #06-01, Centros, Singapore 138668, Singapore. <sup>4</sup>Global Top5 Research Program, Ewha Womans University, 11-1 Daehyeon-dong, Seodaemun-gu, Seoul 120750, Korea.

Received: 14 February 2014 Accepted: 17 April 2014

Published: 28 April 2014

### References

- Oshima T, Imahori K: Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulation thermophilic bacterium from a Japanese thermal spa. *Int J Syst Bacteriol* 1974, **24**:102–112.
- Cava F, Hidalgo A, Berenguer J: *Thermus thermophilus* as biological model. *Extremophiles* 2009, **13**:213–231.
- Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V: Structure of the 30S ribosomal subunit. *Nature* 2000, **407**:327–339.
- Sazanov LA, Hinchliffe P: Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science* 2006, **311**:1430–1436.
- Selmer M, Dunham CM, Murphy FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V: Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 2006, **313**:1935–1942.

6. Yokoyama K, Ohkuma S, Taguchi H, Yasunaga T, Wakabayashi T, Yoshida M: **V-type H<sup>+</sup> -ATPase/synthase from a thermophilic eubacterium, *Thermus thermophilus* - Subunit structure and operon.** *J Biol Chem* 2000, **275**:13955-13961.
7. Pantazaki AA, Pritsa AA, Kyriakidis DA: **Biotechnologically relevant enzymes from *Thermus thermophilus*.** *Appl Microbiol Biotechnol* 2002, **58**:1-12.
8. Niehaus F, Bertoldo C, Kahler M, Antranikian G: **Extremophiles as a source of novel enzymes for industrial application.** *Appl Microbiol Biotechnol* 1999, **51**:711-729.
9. Riyanti EI: **Genetic manipulation of thermophiles for ethanol production.** In *PhD Thesis*. The University of New South Wales: School of biotechnology and biomolecular sciences; 2007.
10. Tian B, Hua Y: **Carotenoid biosynthesis in extremophilic *Deinococcus-Thermus* bacteria.** *Trends Microbiol* 2010, **18**:512-520.
11. Shigi N, Suzuki T, Tamakoshi M, Oshima T, Watanabe K: **Conserved bases in the TPsi C loop of tRNA are determinants for thermophile-specific 2-thiouridylation at position 54.** *J Biol Chem* 2002, **277**:39128-39135.
12. Lewis NE, Nagarajan H, Palsson BO: **Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods.** *Nat Rev Microbiol* 2012, **10**:291-305.
13. Liu L, Agren R, Bordel S, Nielsen J: **Use of genome-scale metabolic models for understanding microbial physiology.** *FEBS Letters* 2010, **584**:2556-2564.
14. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO: **A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.** *Mol Syst Biol* 2007, **3**:121.
15. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R: **Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data.** *J Biol Chem* 2007, **282**:28791-28799.
16. Mo ML, Palsson BO, Herrgard MJ: **Connecting extracellular metabolomic measurements to intracellular flux states in yeast.** *BMC Syst Biol* 2009, **3**:37.
17. Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee DY: **Genome-scale metabolic reconstruction and *in silico* analysis of methylotrophic yeast *Pichia pastoris* for strain improvement.** *Microb Cell Fact* 2010, **9**:50.
18. Kjeldsen KR, Nielsen J: ***In silico* genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network.** *Biotechnol Bioeng* 2009, **102**:583-597.
19. Park JM, Kim TY, Lee SY: **Genome-scale reconstruction and *in silico* analysis of the *Ralstonia eutropha* H16 for polyhydroxyalkanoate synthesis, lithoautotrophic growth, and 2-methyl citric acid production.** *BMC Syst Biol* 2011, **5**:101.
20. Oberhardt MA, Puchalka J, Fryer KE, Martins dosSantos VA, Papin JA: **Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1.** *J Bacteriol* 2008, **190**:2790-2803.
21. Selvarasu S, Karimi IA, Ghim GH, Lee DY: **Genome-scale modeling and *in silico* analysis of mouse cell metabolic network.** *Mol bioSyst* 2010, **6**:152-161.
22. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO: **Global reconstruction of the human metabolic network based on genomic and bibliomic data.** *Proc Natl Acad Sci U S A* 2007, **104**:1777-1782.
23. Lakshmanan M, Koh G, Chung BK, Lee DY: **Software applications for flux balance analysis.** *Brief Bioinform* 2014, **15**:108-122.
24. Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY: **Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation.** *Appl Environ Microbiol* 2005, **71**:7880-7887.
25. Matsuda F, Furusawa C, Kondo T, Ishii J, Shimizu H, Kondo A: **Engineering strategy of yeast metabolism for higher alcohol production.** *Microb Cell Fact* 2011, **10**:70.
26. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, Deacon AM, Wooley J, Lesley SA, Wilson IA, Palsson BO, Osterman A, Godzik A: **Three-dimensional structure view of the central metabolic network of *Thermotoga maritima*.** *Science* 2009, **325**:1544-1549.
27. Ulas T, Riemer SA, Zaparty M, Siebers B, Schomburg D: **Genome-scale reconstruction and analysis of the metabolic network in the hyperthermophilic archaeon *Sulfolobus solfataricus*.** *PLoS One* 2012, **7**:e43401.
28. Henne A, Bruggemann H, Raasch C, Wierer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martinez-Arias R, Jacobi C, Starkuviene V, Schlenczek S, Dencker S, Huber R, Klenk HP, Kramer W, Merkl R, Gottschalk G, Fritz HJ: **The genome sequence of the extreme thermophile *Thermus thermophilus*.** *Nat Biotechnol* 2004, **22**:547-553.
29. Kaneda T: **Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance.** *Microbiological Rev* 1991, **55**:288-302.
30. Nordstrom KM, Laakso SV: **Effect of growth temperature on fatty acid composition of ten *Thermus* strains.** *Appl Environ Microbiol* 1992, **58**:1656-1660.
31. Pask-Hughes RA, Shaw N: **Glycolipids from some extreme thermophilic bacteria belonging to the genus *Thermus*.** *J Bacteriol* 1982, **149**:54-58.
32. Alarico S, da Costa MS, Empadinhas N: **Molecular and physiological role of the trehalose-hydrolyzing alpha-glucosidase from *Thermus thermophilus* HB27.** *J Bacteriol* 2008, **190**:2298-2305.
33. Lengsfeld C, Schonert S, Dippel R, Boos W: **Glucose- and glucokinase-controlled mal gene expression in *Escherichia coli*.** *J Bacteriol* 2009, **191**:701-712.
34. Holst O: **Structure of the Lipopolysaccharide Core Region.** In *Bacterial Lipopolysaccharides*. Edited by Knirel YA, Valvano MA. Wien: Springer-Verlag; 2011:21-39.
35. Mandelli F, Yamashita F, Pereira JL, Mercadante AZ: **Evaluation of biomass production, carotenoid level and antioxidant capacity produced by *Thermus filiformis* using fractional factorial design.** *Braz J Microbiol* 2012, **43**:126-134.
36. Oshima T: **Unique polyamines produced by an extreme thermophile, *Thermus thermophilus*.** *Amino Acids* 2007, **33**:367-372.
37. Raghunathan AU, Perez-Correa JR, Bieger LT: **Data reconciliation and parameter estimation in flux-balance analysis.** *Biotechnol Bioeng* 2003, **84**:700-709.
38. Schuster S, Pfeiffer T, Fell DA: **Is maximization of molar yield in metabolic networks favoured by evolution?** *J Theor Biol* 2008, **252**:497-504.
39. Demirtas MU, Kolhatkar A, Kilbane JJ 2nd: **Effect of aeration and agitation on growth rate of *Thermus thermophilus* in batch mode.** *J Biosci Bioeng* 2003, **95**:113-117.
40. McKay A, Quilter J, Jones CW: **Energy conservation in the extreme thermophile *Thermus Thermophilus* HB8.** *Arch Microbiol* 1982, **131**:43-50.
41. Mahadevan R, Schilling CH: **The effects of alternate optimal solutions in constraint-based genome-scale metabolic model.** *Metab Eng* 2003, **5**:264-276.
42. Selvarasu S, Ow DS-W, Lee SY, Lee MM, Oh SK-W, Karimi IA, Lee D-Y: **Characterizing *Escherichia coli* DH5a growth and metabolism in a complex medium using genome-scale flux analysis.** *Biotechnol Bioeng* 2009, **102**:923-934.
43. Koga Y: **Thermal adaptation of the archaeal and bacterial lipid membranes.** *Archaea* 2012, **2012**:789652.
44. Kumar S, Tsai CJ, Nussinov R: **Factors enhancing protein thermostability.** *Protein Eng* 2000, **13**:179-191.
45. Mallik S, Kundu S: **A comparison of structural and evolutionary attributes of *Escherichia coli* and *Thermus thermophilus* small ribosomal subunits: signatures of thermal adaptation.** *PLoS One* 2013, **8**:e69898.
46. Koyama Y, Hoshino T, Tomizuka N, Furukawa K: **Genetic transformation of the extreme thermophile *Thermus thermophilus* and of other *Thermus* spp.** *J Bacteriol* 1986, **166**:338-340.
47. Yang KM, Lee NR, Woo JM, Choi W, Zimmermann M, Blank LM, Park JB: **Ethanol reduces mitochondrial membrane integrity and thereby impacts carbon metabolism of *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2012, **12**:675-684.
48. Thiele I, Palsson BO: **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nat Protocol* 2010, **5**:93-121.
49. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Res* 2006, **34**:D354-D357.
50. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.** *Nucleic Acids Res* 2012, **40**:D742-D753.
51. Satish Kumar V, Dasika MS, Maranas CD: **Optimization based automated curation of metabolic reconstructions.** *BMC Bioinform* 2007, **8**:212.

52. Ray PH, White DC, Brock TD: **Effect of growth temperature on the lipid composition of *Thermus aquaticus*.** *J Bacteriol* 1971, **108**:227–235.
53. Lee JM, Gianchandani EP, Papin JA: **Flux balance analysis in the era of metabolomics.** *Brief Bioinform* 2006, **7**:140–150.
54. Oberhardt MA, Chavali AK, Papin JA: **Flux balance analysis: interrogating genome-scale metabolic networks.** *Meth Mol Biol* 2009, **500**:61–80.
55. Orth JD, Thiele I, Palsson BO: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245–248.
56. Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BO: **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.** *Nat Protocol* 2011, **6**:1290–1307.

doi:10.1186/1475-2859-13-61

**Cite this article as:** Lee et al.: Genome-scale metabolic network reconstruction and *in silico* flux analysis of the thermophilic bacterium *Thermus thermophilus* HB27. *Microbial Cell Factories* 2014 **13**:61.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

