Estimation of biomass composition from genomic and transcriptomic information

Sophia Santos[1] and Isabel Rocha[1,*]

[1]CEB, Center of Biological Engineering, University of Minho, Braga, Portugal,
http://ceb.uminho.pt

Summary

Given the great potential impact of the growing number of complete genome-scale metabolic network reconstructions of microorganisms, bioinformatics tools are needed to simplify and accelerate the course of knowledge in this field. One essential component of a genome-scale metabolic model is its biomass equation, whose maximization is one of the most common objective functions used in Flux Balance Analysis formulations. Some components of biomass, such as amino acids and nucleotides, can be estimated from genome information, providing reliable data without the need of performing lab experiments. In this work a java tool is proposed that estimates microbial biomass composition in amino acids and nucleotides, from genome and transcriptomic information, using as input files sequences in FASTA format and files with transcriptomic data in the csv format. This application allows to obtain the results rapidly and is also a user-friendly tool for users with any or little background in informatics (http://darwin.di.uminho.pt/biomass/). The results obtained using this tool are fairly close to experimental data, showing that the estimation of amino acid and nucleotide compositions from genome information and from transcriptomic data is a good alternative when no experimental data is available.

1 Introduction

Genome-scale metabolic models are a valuable tool for the study of metabolic systems [1] and are becoming available for an increasing number of organisms [2]. These network reconstructions are used to compute a variety of phenotypic states [3] in order to implement metabolic engineering strategies, or identify drug-targets, among other applications [4]. Flux balance analysis (FBA) is a mathematical widely used approach for genome-scale simulation of metabolic fluxes. FBA uses linear optimization to determine one steady-state reaction flux distribution in a metabolic network by maximizing an objective function. The most common objective function involves the maximization of biomass formation, which has proven to be consistent with experimental observations in several conditions [5]. The formulation of the biomass composition to be used as objective function can be performed at different levels of detail: basic level (defining the macromolecular content on the cell, i.e., percentages of protein, RNA, DNA, lipids), intermediate level (basic level plus calculating the necessary biosynthetic energy) and advanced level (further detailing the necessary vitamins, elements, and cofactors)[3].

For n biomass constituents, the biomass formation equation can be formulated as:

*To whom correspondence should be addressed. Email: irocha@deb.uminho.pt

$$\sum_{i=1}^{n} c_i X_i \rightarrow Biomass \quad (1)$$

where c_i is the coefficient of each component, X_i, considered in the biomass. The units of all the coefficients are defined in mmol per gram of dry weight (mmol/gDW) and the biomass formation units are defined per hour (h^{-1}).

If a biomass component is not accounted for in the biomass objective function, the corresponding synthesis reactions may not be required for growth, as well as the associated genes. Thus, the composition of biomass plays an important role for example for in silico predictions of essential genes [1]. In order to achieve good predictions, a detailed biomass composition of an organism needs to be experimentally determined for cells growing in log phase using available methods [6]. However, often experimental methods are laborious and time consuming or the modelled organism is difficult to grow in the lab. In many cases, when no experimental data are available, biomass composition of related organisms is included in the model [7, 8]. Also, some components such as amino acids, nucleotides (NTPs) and deoxynucleotides (dNTPs) can be estimated from genome information, as described in 2010 by Thiele and Palsson in their detailed protocol to create a genome-scale metabolic network reconstruction. Some studies indicate that this approach is more reliable than performing aproximations to closely related organisms, having an impact in the predictions of the specific growth rate and flux distributions as low as 1.5 % when compared with experimental values [9]. However, when estimating amino acids compositions directly from the genome, it is assumed that all proteins are being expressed at all times, in the same proportions, a fact that is known to be false. Indeed, some authors have already used genome information allied with transcriptomic data to estimate more accurately the biomass composition in amino acids [10]. The main goal of this work was to develop a java tool which returns the estimated biomass composition in amino acids, NTPs and dNTPs for an organism, from files with selected sequences and transcriptomic data. The obtained data can be directly included in the biomass equation of a genome-scale metabolic model. Also, the impact of using estimated versus experimental amino acid composition in genome-scale metabolic models predictions was analyzed. This java tool, in the future, will be integrated as a plug-in for the *merlin* (MEtabolic model Reconstruction using genome scaLe INformation) framework [11], that was created to assist in the processes of (re) annotation and reconstruction of genome-scale metabolic models. At the moment the tool is available as a standalone tool in the page `http://darwin.di.uminho.pt/biomass/`.

2 Methods

2.1 *In silico* Biomass Determination

2.1.1 Genome Information

As Thiele and Palsson (2010) have indicated, the estimation of the composition in amino acids, NTPs and dNTPs from genome information can be performed by calculating the molar percent-

age of each monomer and converting it into mmol/gDW. Genome information is easily found in databases and can be extracted in various formats, like FASTA and GENBANK. Since the FASTA format is easier to manipulate and is also the most universal format (being the only format to export sequences in databases as the Uniprot - UniProt Consortium [12]), the java application developed requires sequences files solely in this format. Moreover, there are several online resources to convert GENBANK to FASTA format files as the GenBank Feature Extractor website (available in `http://www.bioinformatics.org/sms2/genbank_feat.html`) [13]. In order to determine the NTPs composition of the cell, the protocol described by Thiele and Palsson uses the codon usage accessed for the amino acid content. Since RNA incorporates uracil (U) instead of thymine (T), the codon usage needs to be read with every T replaced by a U. However, in this report the authors do not distinguish between the different types of RNA and, as a result, perform their calculations for messenger RNA (mRNA) only. However, in a prokaryotic cell 95% of total RNA is transfer RNA (tRNA) and ribosomal RNA (rRNA) [14]. Therefore, some changes were made to the protocol described by Thiele and Palsson regarding NTPs estimation. Genome information for mRNA, rRNA and tRNA is used in the new protocol and the NTPs are determined taking into account the percentage of each molecule in the total RNA. These percentages differ also among organisms: gram positive bacteria have on average 5% mRNA, 20% tRNA and 75% rRNA and gram negative bacteria and yeast have 5% mRNA, 15% tRNA and 80% rRNA [15, 16].

2.1.2 Transcriptomic/proteomic Information

To determine the biomass composition in amino acid, gene expression data can also be used together with genome sequencing information, as long as these data are available for a wide variety of relevant conditions. Gene expression data should be available as total abundance of expression of each gene/protein, which needs to be normalized to a ratio (referred in equation (2) as $Abundance^p$ or abundance of protein p). The composition of each protein in amino acid *i* (AA^p) is taken from the genome information (being $AA^{p,g}$ the amino acid composition obtained from the genome)and is corrected by the expression factor as shown in equation (2).

$$AA_i^p = AA_i^{p,g}(ratio) \times Abundance^p(ratio) \tag{2}$$

The total biomass content in each amino acid i is determined by the sum of values of each amino acid for all proteins (TP) divided by the sum of all amino acids (N) for all proteins:

$$AA_i^T(\%) = \frac{\sum_p^{TP} AA_i^p}{\sum_p^{TP} \sum_i^N AA_i^p} \tag{3}$$

The values obtained are expressed in molar percentage, and have to be converted to mmol/gDW to be included in the biomass equation.

3 Computational tool for the estimation of Biomass composition from the genome and transcriptome/proteome

The application developed is fully implemented in the Java language, and using BioJava packages. The main capability of the application is the estimation of biomass composition in amino acids and nucleotides from the genome and transcriptomic information, using as input files sequences in FASTA format and expression data in csv format. The application determines the frequency, in percentage, of each amino acid and nucleotide in the cell, as shown in Equations 2 and 3 and exemplified in Figure 1, and also the same variable in mmol/gDW, to directly add to the biomass equation. This application allows obtaining rapidly this kind of information and is also a user-friendly tool, facilitating its use by operators with no or little background in informatics.

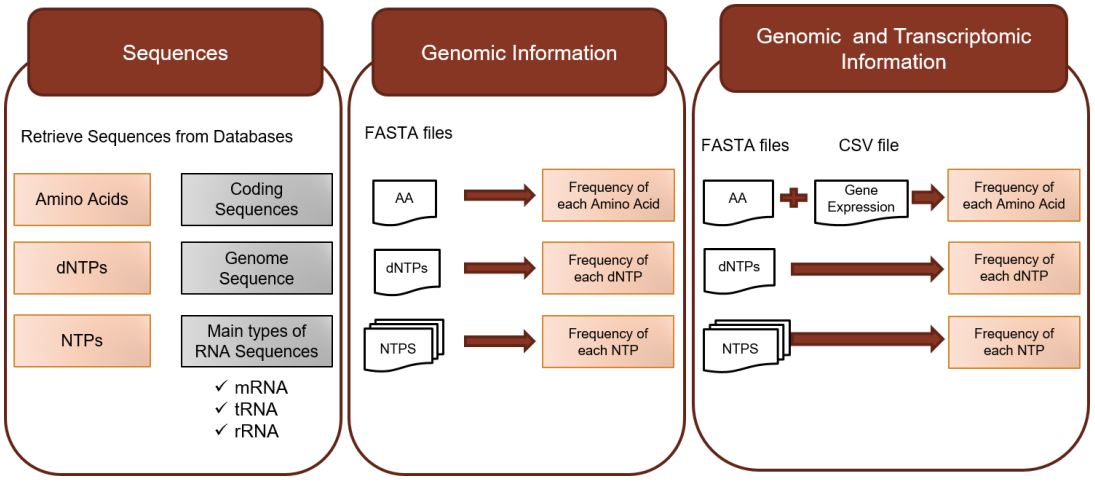

Figure 1: Process to determine the frequency of each monomer from genome and transcriptomic information

The application is separated in three tabs, one to estimate the Protein composition in each amino acid (Figure 2 A), another to estimate the DNA composition in each dNTP (Figure 2 B)and the other to estimate the RNA composition in each NTP (Figure 2 C).

To use the application, it is indispensable to input files with sequences of Proteins, DNA and RNA, exclusive in the FASTA format. If transcriptomic data are available for the organism in study, they can be added in the csv format, with two columns separated by semicolon: the first column should contain gene identifiers and the second the expression factor (or Abundance in Equation 1) in percentage. In this case, the FASTA file with protein sequences should have the same gene identifiers at the beginning of the sequence header. To obtain the results it is only necessary to click in the **Determine** button. It is also possible to export the obtained data to a file in csv format, by clicking in the **Export** button. The application requires some additional obligatory inputs: percentage of each type of RNA (mRNA, rRNA and tRNA), that is specific for each organism and can be obtained in the literature, and also the value of the cellular content in each macromolecule (Protein, DNA and RNA) in mass percentage, in order to calculate the corresponding biomass composition in mmol/gDW to be included directly in the biomass equation. These data can be either obtained in the lab or from literature.

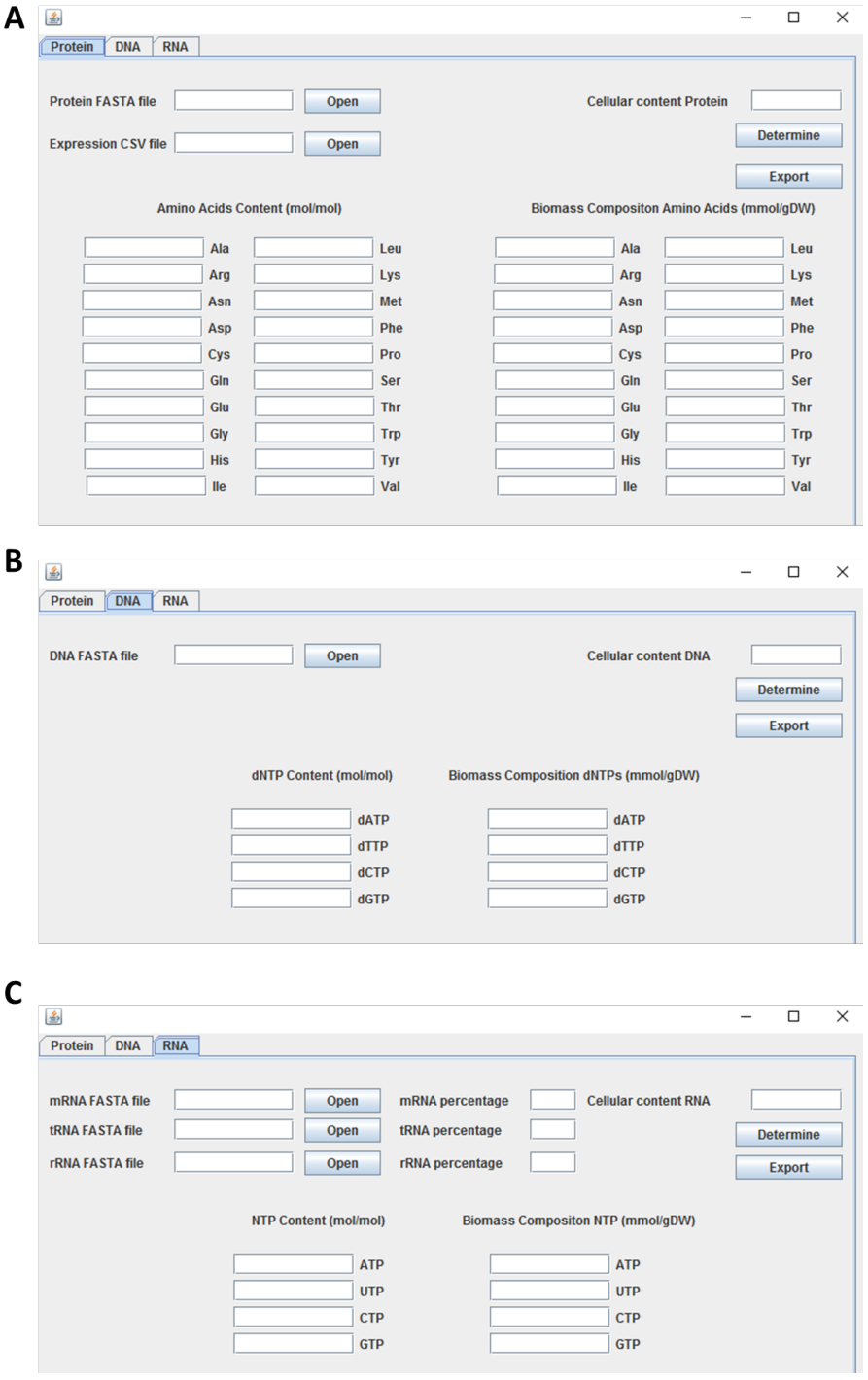

Figure 2: Screen-shots from the java application developed. Tabs to estimate and provide results for (A) Protein composition in each amino acid, (B) DNA composition in each dNTP, and (C) RNA composition in each NTP.

3.1 Impact of Biomass Composition in Model Predictions

3.1.1 *In silico* Simulations

In order to evaluate the impact of the use of different coefficients in the biomass equation, several *in silico* analyses have been performed using Optflux 3.2.8 [17]. For that, 3 genome-scale metabolic models of bacteria and yeast (given in Table 1) for which the biomass composition has been experimentally determined, and with genomic and transcriptomic data available, were used for performing simulations using parsimonious FBA (pFBA). The genome-scale metabolic models used were all available in the Systems Biology Markup Language (SBML) [18].

3.1.2 Specific growth rate determination

Wild type simulations were performed for each organism listed in Table 1, with the original experimental and altered biomass compositions. The differences calculated for the specific growth rate values obtained using the different biomass compositions were expressed in percentage, according to the following expression

$$\frac{|Exp - X|}{Exp} \times 100 \tag{4}$$

Where *Exp* represents the specific growth rate for the original biomass composition and *X* represents the specific growth rate for the biomass composition that was altered.

3.1.3 Flux distribution analysis

For all the simulations described above, the flux values for each reaction were obtained. Flux distribution data were analyzed using standard statistical techniques available in R software (version 3.3.1) using RStudio IDE, version 3.1.3 [19]. To evaluate significance in comparing paired samples the 2-tailed dependent Students t test was used, while the Pearson correlation coefficient was computed to evaluate the degree of linear dependence between the fluxes. The differences between the flux distributions obtained using the original and the altered biomass compositions were evaluated using the distance measure sum of squared differences (SSD), according with the following expression:

$$SSD = \sum_{i}^{\mathbb{N}} (x_i - y_i)^2 \tag{5}$$

Where x_i represents the flux value of reaction i in the simulation performed with experimental biomass composition and y_i represents the flux value for the same reaction for the altered biomass reaction. $\mathbb{N}$ represents the total number of metabolic reactions.

The flux difference was also evaluated by computing the Jaccard distance that evaluates which reactions change from having zero flux to a greater than zero flux from one condition to the other, according with the following expression:

$$J_\delta = \frac{p}{p+q+d} \tag{6}$$

Where p represents the number of reactions with a nonzero flux in both the experimental and altered biomass composition models, q represents the number of reactions with nonzero flux in the experimental biomass composition model and with zero flux in altered biomass compositions models and d represents the number of reactions with no flux in the experimental biomass composition model and with nonzero flux in the altered biomass composition model.

4 Results and Discussion

To illustrate some of the main features of the developed procedure and the java tool, the genome sequences and transcriptomic data for the organisms *Bacillus subtilis*, *Escherichia coli* and *Saccharomyces cerevisiae* were used. The results obtained were then compared with experimental data for biomass compositions. All FASTA files with genome information were retrieved from the National Center for Biotechnology Information (NCBI) database (`http://www.ncbi.nlm.nih.gov/`). The data and models used and their respective references are listed in Table 1.

Table 1: Organisms data and models used to estimate biomass composition in amino acids and nucleotides and to perform simulations.

Organism	Abbreviations	Transcriptomics Data	Biomass Experimental data	Model ID	Model
Bacillus subtilis	Bsu	[20]	[21]	*i*Bsu1103	[22]
Escherichia coli	Eco	[23]	[15]	*i*AF1260	[24]
Saccharomyces cerevisiae	Sce	[25]	[26]	yeast 7.0	[27]

4.1 Composition in Amino Acids

Figure 3 illustrates examples of the two input files used to estimate the biomass composition in amino acids. In this example, locus tag identifiers are used in both files. The results obtained for amino acid composition were compared to experimental data and are represented in Figure 4.

The results obtained for the amino acid composition are fairly close to the experimental data. As expected, the composition in amino acids obtained using the genome and transcriptomic data is closer to experimental data than the one obtained using only genome information, for all organisms. Differences are bigger in the case of *S. cerevisiae*, probably caused by the fact that

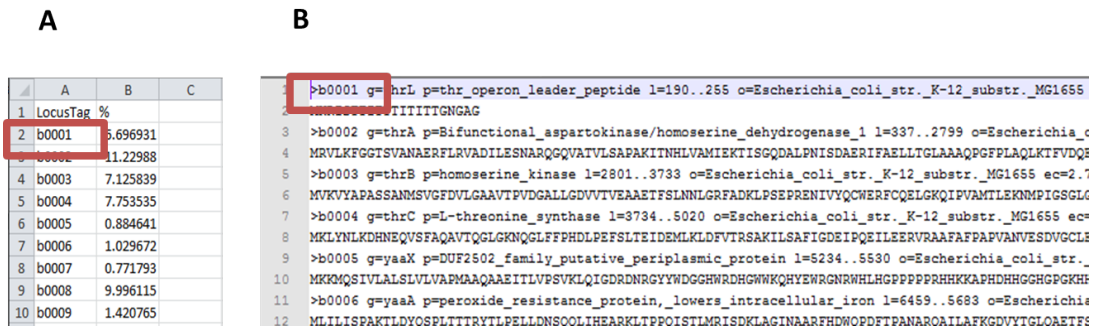

Figure 3: Screen-shots from the files used to estimate the biomass composition in amino acids for *E. coli*. (A) Csv file with gene/protein identifier and respective expression factor and (B) FASTA file with all protein sequences, where the same identifier is used for each sequence.

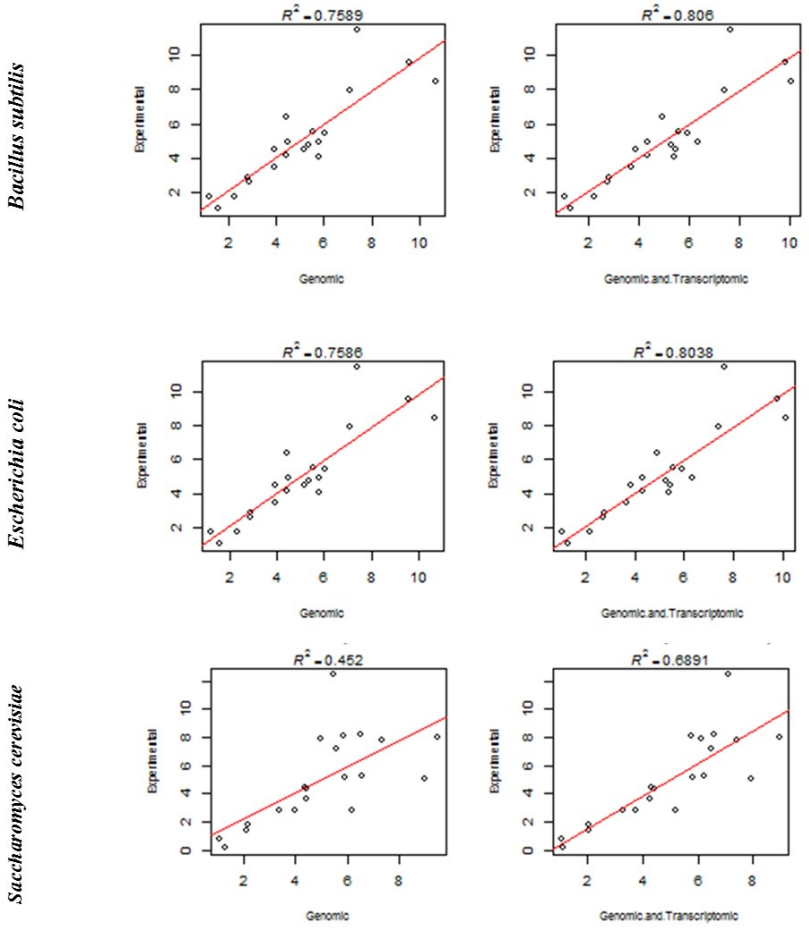

Figure 4: Comparison of the biomass composition in amino acids obtained using the java tool (either with only genome information or using genome and transcriptomic information) with experimental data from the literature. Correlation between the data is represented by the R^2. Amino acid composition is represented in molar percentage.

S. cerevisiae has a more complex apparatus and a bigger genome that *E. coli* and *B. subtilis*. Consistent differences have been observed between experimental and estimated data for some specific amino acids, which probably is due to some amino acids being more sensitive to the experimental methods than others. Some amino acids are very sensitive to hydrolysis, particularly cysteine, tryptophan and methionine [28], while asparagine is transformed in aspartic acid and glutamine is transformed in glutamic acid before the measurements. Other amino acids are more sensitive to the derivatization step, producing more than one derivative, such as glycine and lysine, or losing detectable response, such as leucine [29]. Thus, the differences observed do not necessarily mean that experimental data are more accurate than the simulated ones.

4.2 Nucleotide and Deoxynucleotide compositions

The results obtained for NTPs and dNTPs compositions are summarized in Table 2 and are also compared with reference data.

Table 2: Biomass composition in nucleotides and deoxynucleotides estimated from genome information and obtained from reference. Nucleotide and deoxynucleotide compositions are presented in molar percentages.

Deoxynucleotide or Nucleotide	*Bacilus* Reference [21]	*subtilis* Genome	*Escherichia* Reference [15]	*coli* Genome	*Saccharomyces* Reference [26]	*cerevisiae* Genome
dATP	28.5	28.2	24.6	24.6	29.8	31.0
dCTP	21.6	21.8	25.4	25.4	20.2	19.1
dGTP	21.6	21.7	25.4	25.4	20.2	19.1
dTTP	28.4	28.3	24.6	24.6	29.8	30.9
ATP	26.6	25.1	20.0	23.5	23.3	28.8
CTP	18.8	23.2	32.2	25.7	22.8	17.7
GTP	34.3	29.9	21.6	29.0	23.3	24.8
UTP	20.4	21.8	26.2	21.8	30.6	28.7

The values estimated for the biomass composition in dNTPs are very similar to the reference data, as expected. However, the same does not hold for NTPs. This fact can be explained by the use of different methods to calculate the composition in NTPs. Some authors account only open reading frames sequences for the NTPs composition [30]. Other authors use genomic DNA sequences to calculate the percentage of messenger RNA (mRNA) [31]. These differences can also be caused by the use of different databases to retrieve the sequences of each organism and the existence of differences between these sequences.

4.3 Impact of biomass composition in genome-scale metabolic model predictions

The estimated biomass compositions in amino acids obtained using genomic and transcriptomic data were incorporated in the biomass equation of each genome-scale metabolic model present in Table 1. Several *in silico* experiments were performed using Optflux 3.2.8, where specific

growth rates and flux distributions were analyzed. For each organism, simulations were performed using the biomass composition computed using either genomic or both genomic and transcriptomic information. The obtained results (specific growth rate and flux distribution) were then compared with simulations performed with the original experimentally determined biomass equation (called experimental data in Figure 5 for simplicity). Analyzing the simulation results in Figure 5 it can be concluded that when genomic and transcriptomic information are used to estimate biomass composition, the differences in the simulation results to data obtained with experimental biomass composition are lower. These differences appear with more emphasis in the flux distribution than in the specific growth rate. Surprisingly, the coefficients for amino acids have more impact in *E. coli* simulations than in the other organisms. The fact that a good correlation between estimated amino acids and original experimental data was obtained for this organism (see Figure 4) suggests that the large differences found for specific amino acids might be affecting genome-scale metabolic predictions.

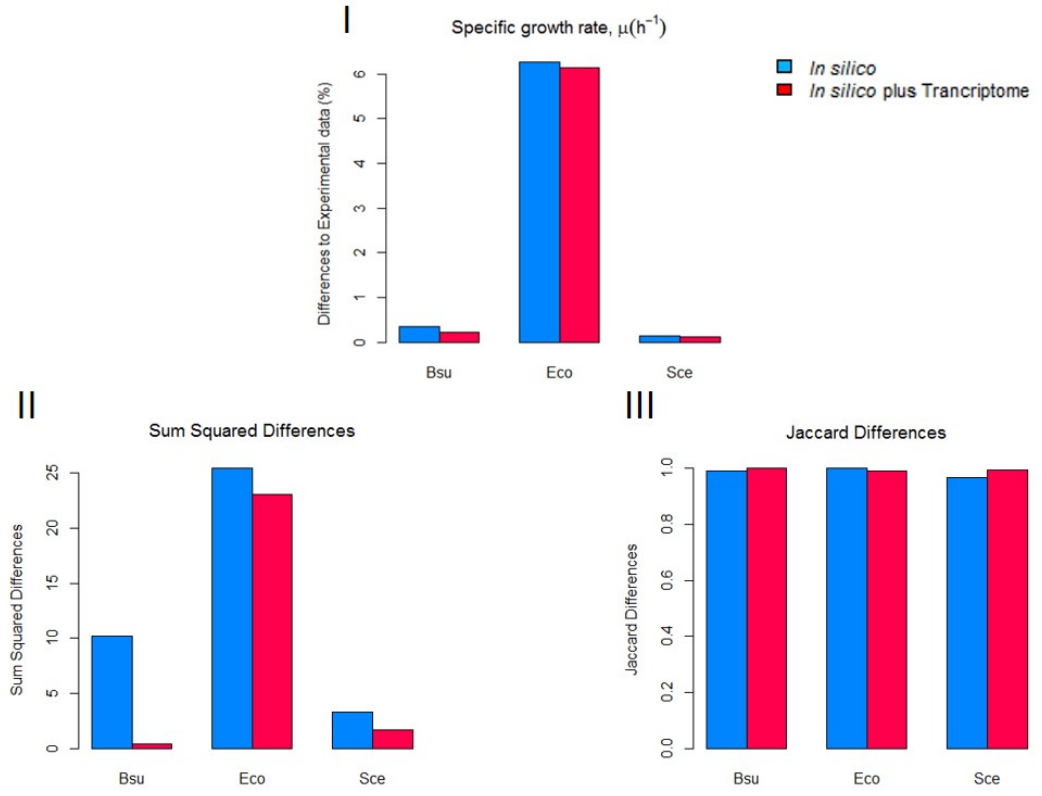

Figure 5: Differences in specific growth rate and flux distribution predictions when changing the biomass coefficients for amino acids in genome-scale metabolic models from the ones experimentally determined to the ones determined in silico from genomic and genomic + transcriptomic information. (I) Differences in the specific growth rate, (h^{-1}) obtained, (II) Sum of squared differences of flux distribution and (III) Jaccard coefficient for flux distributions. Abbreviations: Bsu *B. subtilis*; Eco *E. coli*; Sce *S. cerevisiae*.

The evaluation of the significance of the differences found was performed for paired samples using the 2-tailed dependent Students t test, and the Pearson correlation coefficient was used to evaluate the degree of linear dependence between the fluxes (comparing the altered biomass

compositions with the original experimentally determined biomass composition). Despite of a slight improvement when using transcriptomic information allied to genome information, the obtained results are not statistically significant as can be seen in Table 3.

Table 3: Statistical results for the 2-tailed dependent Students t test (p-value) and Pearson Correlation (PC) test when comparing flux distributions with altered biomass compositions with the original experimentally determined biomass composition.

Organisms	*Bacilus*	*subtilis*	*Escherichia*	*coli*	*Saccharomyces*	*cerevisiae*
	In silico	*In silico* and Trascriptome	*In silico*	*In silico* and Trascriptome	*In silico*	*In silico* and Trascriptome
p-value	0.74765	0.66848	0.22775	0.21664	0.68891	0.74899
PC	0.99977	0.99999	0.99956	0.99960	0.99999	0.99999

5 Conclusions and Future work

The java application created is a tool that provides the estimation of biomass composition in nucleotides and amino acids, with input files containing sequences from DNA, RNA and protein, in the FASTA format. When expression data are available, it can also be used, provided in a csv file containing percentages of each gene/protein. All data obtained can be easily exported to a csv file. The results obtained using that tool and the described procedure are fairly close to experimental data, showing that the estimation of amino acid and nucleotide compositions from genome information and from transcriptomic data is a good alternative when no experimental data is available. In the future, this java tool will be integrated as a plug-in for the *merlin* framework. At the moment the tool is available as a standalone tool in the page `http://darwin.di.uminho.pt/biomass/`.

Acknowledgments

The authors thank the project DeYeastLibrary - Designer yeast strain library optimized for metabolic engineering applications, Ref. ERA-IB-2/0003/2013, funded by national funds through FCT/MCTES.

This study was supported by the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UID/BIO/04469/2013 unit and COMPETE 2020 (POCI-01-0145-FEDER-006684) and BioTecNorte operation (NORTE-01-0145-FEDER-000004) funded by European Regional Development Fund under the scope of Norte2020 - Programa Operacional Regional do Norte.

References

[1] I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5(1):93–121, 2010.

[2] E. Simeonidis and N. D. Price. Genome-scale modeling for metabolic engineering. *Journal of industrial microbiology & biotechnology*, 42(3):327–338, 2015.

[3] A. M. Feist and B. O. Palsson. The biomass objective function. *Current opinion in microbiology*, 13(3):344–349, 2010.

[4] K. Raman and N. Chandra. Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, 10(4):435–449, 2009.

[5] J. S. Edwards, R. U. Ibarra and B. O. Palsson. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature biotechnology*, 19(2):125–130, 2001.

[6] M. Carnicer, K. Baumann, I. Töplitz, F. Sánchez-Ferrando, D. Mattanovich, P. Ferrer and J. Albiol. Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microbial cell factories*, 8(1):1, 2009.

[7] I. Thiele, T. D. Vo, N. D. Price and B. Ø. Palsson. Expanded metabolic reconstruction of *Helicobacter pylori* (*i*IT341 GSM/GPR): an *in silico* genome-scale characterization of single-and double-deletion mutants. *Journal of bacteriology*, 187(16):5818–5830, 2005.

[8] C. Risso, J. Sun, K. Zhuang et al. Genome-scale comparison and constraint-based metabolic reconstruction of the facultative anaerobic Fe (III)-reducer *Rhodoferax ferrireducens*. *BMC genomics*, 10(1):447, 2009.

[9] S. T. Santos. *Development of computational methods for the determination of biomass composition and evaluation of its impact in genome-scale models predictions*. Master's thesis, Universidade do Minho, 2013.

[10] R. Saha, A. T. Verseput, B. M. Berla, T. J. Mueller, H. B. Pakrasi and C. D. Maranas. Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece sp.* ATCC 51142 and *Synechocystis sp.* PCC 6803. *PloS one*, 7(10):e48285, 2012.

[11] O. Dias, M. Rocha, E. C. Ferreira and I. Rocha. Reconstructing genome-scale metabolic models with *merlin*. *Nucleic acids research*, page gkv294, 2015.

[12] U. Consortium et al. Uniprot: a hub for protein information. *Nucleic acids research*, 2014.

[13] C. Plot. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28(6), 2000.

[14] C. Rosenow, R. M. Saxena, M. Durst and T. R. Gingeras. Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. *Nucleic acids research*, 29(22):e112–e112, 2001.

[15] F. C. Neidhardt and R. Curtiss. *Escherichia coli and Salmonella : cellular and molecular biology*, volume 2. ASM Press, Washington, D.C., 2 edition, 1996.

[16] J. R. Warner. The economics of ribosome biosynthesis in yeast. *Trends in biochemical sciences*, 24(11):437–440, 1999.

[17] I. Rocha, P. Maia, P. Evangelista, P. Vilaça, S. Soares, J. P. Pinto, J. Nielsen, K. R. Patil, E. C. Ferreira and M. Rocha. Optflux: an open-source software platform for *in silico* metabolic engineering. *BMC systems biology*, 4(1):1, 2010.

[18] M. Hucka, A. Finney, H. M. Sauro et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[19] J. S. Racine. Rstudio: A platform-independent IDE for R and Sweave. *Journal of Applied Econometrics*, 27(1):167–172, 2012.

[20] P. Nicolas, U. Mäder, E. Dervyn et al. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science*, 335(6072):1103–1106, 2012.

[21] M. Dauner and U. Sauer. Stoichiometric growth model for riboflavin-producing *Bacillus subtilis*. *Biotechnology and bioengineering*, 76(2):132–143, 2001.

[22] C. S. Henry, J. F. Zinner, M. P. Cohoon and R. L. Stevens. *i*Bsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome biology*, 10(6):1, 2009.

[23] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard and B. O. Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987):92–96, 2004.

[24] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis and B. Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology*, 3(1):121, 2007.

[25] L. Treu, S. Campanaro, C. Nadai, C. Toniolo, T. Nardi, A. Giacomini, G. Valle, B. Blondin and V. Corich. Oxidative stress response and nitrogen utilization are strongly variable in *Saccharomyces cerevisiae* wine strains with different fermentation performances. *Applied microbiology and biotechnology*, 98(9):4119–4135, 2014.

[26] T. L. Nissen, U. Schulze, J. Nielsen and J. Villadsen. Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology*, 143(1):203–218, 1997.

[27] H. W. Aung, S. A. Henry and L. P. Walker. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Industrial Biotechnology*, 9(4):215–228, 2013.

[28] M. Fountoulakis and H.-W. Lahm. Hydrolysis and amino acid composition analysis of proteins. *Journal of Chromatography A*, 826(2):109–134, 1998.

[29] I. Molnar-Perl. Derivatization and chromatographic behavior of the o-phthaldialdehyde amino acid derivatives obtained with various SH-group-containing additives. *Journal of Chromatography A*, 913(12):283 – 302, 2001.

[30] A. Raghunathan, J. Reed, S. Shin, B. Palsson and S. Daefler. Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. *BMC systems biology*, 3(1):38, 2009.

[31] K. R. Kjeldsen and J. Nielsen. *In silico* genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnology and bioengineering*, 102(2):583–597, 2009.

[32] S. B. Sohn, T. Y. Kim, J. M. Park and S. Y. Lee. *In silico* genome-scale metabolic analysis of *Pseudomonas putida* KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. *Biotechnology journal*, 5(7):739–750, 2010.