


METHODOLOGY

Open Access



# Parallel isotope differential modeling for instationary $^{13}\text{C}$ fluxomics at the genome scale

Zhengdong Zhang<sup>1,2†</sup>, Zhentao Liu<sup>3†</sup>, Yafei Meng<sup>1</sup>, Zhen Chen<sup>4</sup>, Jiayu Han<sup>4</sup>, Yimin Wei<sup>5</sup>, Tie Shen<sup>2\*</sup> , Yin Yi<sup>6</sup> and Xiaoyao Xie<sup>2</sup>

**Background:** A precise map of the metabolic fluxome, the closest surrogate to the physiological phenotype, is becoming progressively more important in the metabolic engineering of photosynthetic organisms for biofuel and biomass production. For photosynthetic organisms, the state-of-the-art method for this purpose is instationary  $^{13}\text{C}$  fluxomics, which has arisen as a sibling of transcriptomics or proteomics. Instationary  $^{13}\text{C}$  data processing requires solving high-dimensional nonlinear differential equations and leads to large computational and time costs when its scope is expanded to a genome-scale metabolic network.

**Result:** Here, we present a parallelized method to model instationary  $^{13}\text{C}$  labeling data. The elementary metabolite unit (EMU) framework is reorganized to allow treating individual mass isotopomers and breaking up of their networks into strongly connected components (SCCs). A variable domain parallel algorithm is introduced to process ordinary differential equations in a parallel way. 15-fold acceleration is achieved for constant-step-size modeling and ~fivefold acceleration for adaptive-step-size modeling.

**Conclusion:** This algorithm is universally applicable to isotope granules such as EMUs and cumomers and can substantially accelerate instationary  $^{13}\text{C}$  fluxomics modeling. It thus has great potential to be widely adopted in any instationary  $^{13}\text{C}$  fluxomics modeling.

**Keywords:** Instationary metabolic flux analysis, Parallel differential equations modeling, Genome-scale metabolic flux analysis,  $^{13}\text{C}$  fluxomics, Mass isotopomer network

## Background

With the arrival of the post-genome era,  $^{13}\text{C}$  fluxomics has matured as a state-of-the-art approach to derive in vivo metabolic flux information in parallel with transcriptomics, proteomics and metabolomics [1–3]. This method captures very important and unique information reflecting intracellular physiology that could never be achieved by other -omics techniques and is pivotal to metabolic engineering of microbes for biofuel and

bioproduct [4–7]. Its irreplaceability qualifies it as an extraordinarily powerful tool for exploring metabolic flux and has encouraged progressively wide application to a broad variety of organisms, such as photosynthetic organisms, fungi or mammalian cells [8–10].

$^{13}\text{C}$  fluxomics can be divided into two categories. One is stationary  $^{13}\text{C}$  fluxomics, which assumes a steady-state metabolic and isotopic labeling environment and deals with algebraic balance equations of mass and isotopes [3, 11]. Model construction and flux estimation can be performed with different isotope granules, such as isotopomers, cumomers and elementary metabolite units (EMUs) [12–15]. This capability has excited a wave

\*Correspondence: shentie@gznu.edu.cn

<sup>†</sup>Zhengdong Zhang and Zhentao Liu contributed equally to this paper

<sup>2</sup> Key Laboratory of Information and Computing Science Guizhou Province, Guizhou Normal University, Guiyang, Guizhou, China  
Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

of tool development, such as 13CFLUX2, FiatFlux, and WUFlux [16–20].

However, the isotopic steady-state assumption does not hold true for many circumstances, for example: (1) fed-batch conditions where the isotopic steady state cannot readily be reached in a short time [21, 22]. (2) Continuous cultivation conditions with a pulsed substrate supply [23–25]. In particular, steady-state 13C fluxomics becomes inefficient where the substrate is a one-carbon compound, such as CO<sub>2</sub> for photoautotrophic cultivation for biofuel and biomass conversion, since the steady-state isotopic distribution is binomial and independent of the flux values [26, 27].

To circumvent the limits of steady-state 13C fluxomics, the second type of 13C fluxomics—isotopically instationary 13C fluxomics (INST-Fluxomics) has been invented to be applied to systems that are in a metabolic steady state while being isotopically instationary [23, 24, 26–29]. This technique has been implemented in software such as INCA, OpenMeibius [30, 31], and extensively adopted to investigate the intracellular physiology of cyanobacteria, microalgae and plants [6, 32–34].

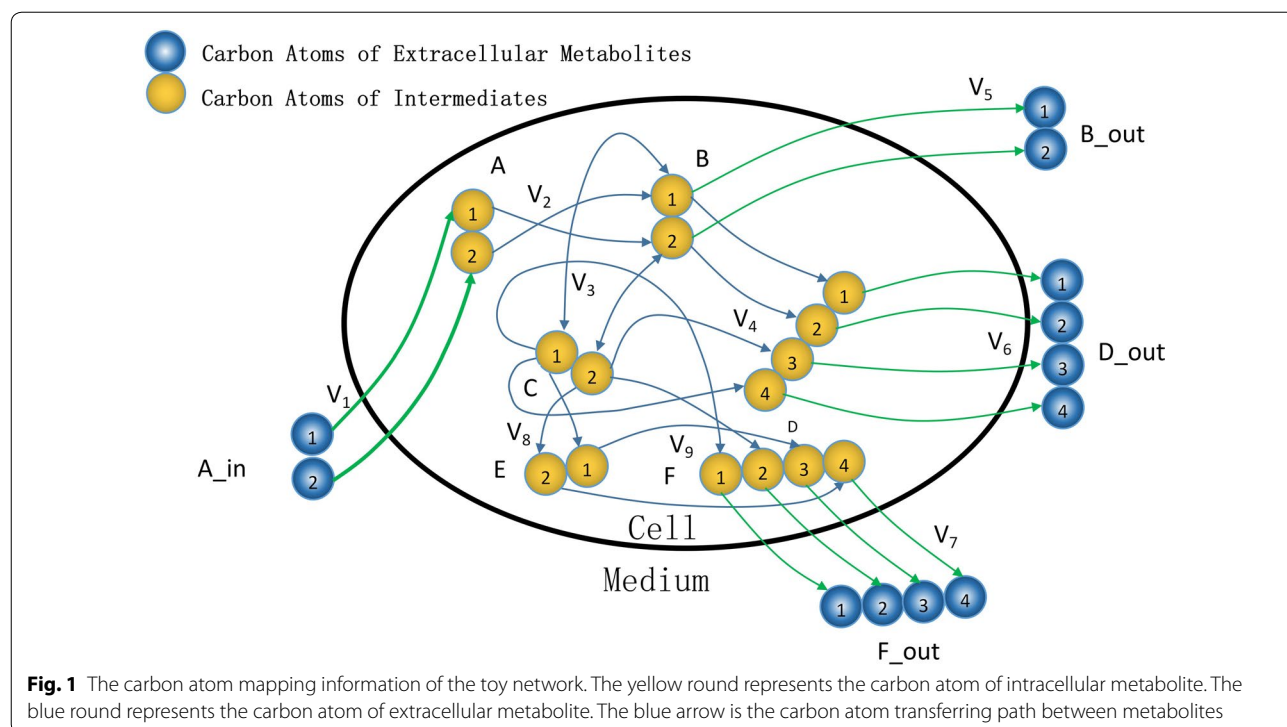
Unlike the stationary case, instationary 13C fluxomics is required to model a large set of ordinary differential equations (ODEs) instead of algebraic equations. In particular, the equations contain high-dimensional derivative variables over the flux values and pool sizes and inevitably incur large computing and time costs [10,

24]. For example, modeling a realistic central carbon model of *E. coli* with constant time integration method requires several hours [30]. This problem grows significantly when a genome-scale metabolic network is considered [35, 36]. Such models usually have hundreds of metabolites and reactions, resulting in a steep increase in the number of equations [35, 36]. The motivation of this paper is to find a way to more quickly model a large set of isotope ordinary differential equations, especially those for a genome scale metabolic network. Here, we present the first parallel method for instationary fluxomics analysis. We reorganized the mass isotopomers of the EMU to model them individually. This treatment facilitates downstream isotopic network decomposition and simplification. An algorithm for parallelization in the variable domain is used to model the ODE systems in a parallel way. Implementing this method has achieved tens of fold improvement with constant-step-size ODEs and ~five-fold improvement for adaptive-step-size ODEs. Since the method is universally applicable to labeling frameworks such as EMUs and cumomers, we expect that this method will benefit all 13C fluxomics communities.

## Results

### SCC decomposition and EMU reorganization in a mass isotopomer network

Here, we use a toy network previously reported to show the framework (Fig. 1 and Table 1) [13].



**Table 1** The stoichiometry and carbon transition of toy model

Reaction name	Reaction stoichiometry	Carbon transition
V <sub>1</sub>	A → B	#AB <sup>a</sup> → #BA
V <sub>2</sub>	B → C	#AB → #AB
V <sub>3</sub>	B + C → D	#AB + #CD → #ABCD
V <sub>4</sub>	C → E	#AB → #BA
V <sub>5</sub>	C + E → F	#AB + #CD → #ABCD
V <sub>6</sub>	B → B_OUT	#AB → #AB
V <sub>7</sub>	D → D_OUT	#ABCD → #ABCD
V <sub>8</sub>	F → F_OUT	#ABCD → #ABCD

<sup>a</sup> Represents the carbon atom at different positions

According to the generation relationship, we can draw the mass isotopomer labeling network in a directed graph  $G_m$ . Then, the unnecessary part that could not contribute to the measured mass isotopomer distribution (MID) is removed from this mass isotopomer labeling network. Thus, the computation complexity can be significantly reduced.

If two mass isotopomers are interdependent, they should be modeled together and simultaneously. The interdependency is that the nodes belong to one strongly connected component (SCC). To model the mass isotopomers separately and in parallel, the classic Tarjan algorithm is employed to decompose the whole network  $G_m$  into small pieces of SCCs (Fig. 2a) [37, 38].

The SCCs are organized by their topological sort to reveal their dependency, as shown in Fig. 2a (see Method description section). As such, the mass isotopomers are transformed into different SCCs, which still contain all original mass isotopomers, as shown in Fig. 2b. The motivation to perform such reorganization is that the whole network can be split into SCCs with smaller scale and large quantity and similar manipulation is universally applicable to other isotope granules like cumomer and isotopomer [12, 14]. In addition to the toy network, this study also uses an *E. coli* metabolic network and a genome-scale metabolic network of *Synechococcus* 2973, which is modified from *imSyn593* reconstruction [36]. The number of SCCs is 14, 45 and 98 for the 3 networks, respectively.

Additionally, the adjacent SCCs with the same weight can be combined in a head-to-tail way, to form a larger unit that could be modeled simultaneously similarly. We also call such a unit as SCC. To assess the impact of such combination, we set a parameter  $\lambda$  representing the minimum mass isotopomer quantity of an SCC with each of the new SCCs being larger than  $\lambda$ . As  $\lambda$  goes up, the number of SCCs declines.

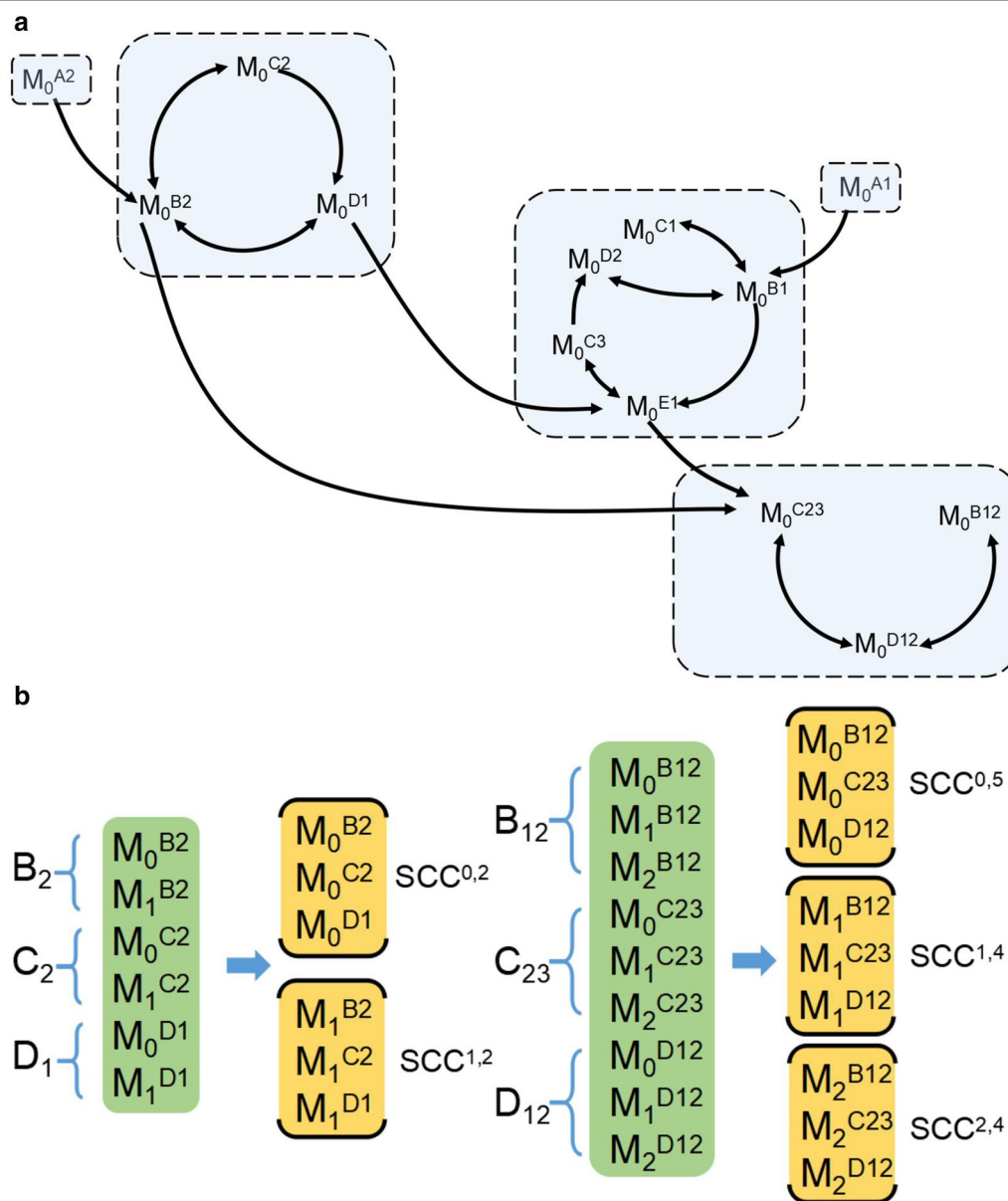
#### Parallelization for a constant-step ODE at the genome scale

The parallelization strategy is shown in Fig. 3. In simulating the ODE, the value of each time step of one SCC depends on not only its own value in the previous step, but also the values of its parent SCCs in the previous step. Once the values of one SCC for time  $T_i$  have been computed, these values are delivered immediately to the threads where their downstream SCCs are waiting. The delivery paths of these values (Fig. 3a) are the same as the cascade relationship in the mass isotopomer network. With this strategy, all SCCs can be computed simultaneously step by step, and the corresponding time is significantly shorter than the sum of times for each SCC individually (Fig. 3b).

There are at least two kinds of ODE methods applicable according to the step-size choice: a constant-step-size method and an adaptive-step-size method [39, 40]. The constant-step-size method is well suited for parallelization. A 4th-order Runge–Kutta method with a constant step size is implemented with a nonparallel technique and a parallel technique on the genome-scale carbon mapping models (Fig. 4) [36, 39]. The evolution modeling is carried out with tensor-based and vector-based method. The following data are from the vector-based method since it got a significantly faster speed.

The model encompasses 78 free fluxes in 174 total fluxes and 189 free pool size in 266 metabolites, resulting in 407 carbon transition reactions (Additional file 1). The bicarbonate uptake rate was set as 10 mmol/gDW/h before scaling, while the growth-related dilution was not considered. The pool sizes were randomly set within the physiological range from micromole per liter to millimole per liter. The MID of whole molecules of 14 metabolites was set as the measured MID for network simplification as previously reported (2-phospho-D-glycerate, 6-phospho-D-gluconate, ribose-5-phosphate, D-glucose-6-phosphate, phosphoenolpyruvate, pyruvate, sedoheptulose-7-phosphate, succinate, malate, 3-phospho-D-glycerate, D-fructose-6-phosphate, citrate, sedoheptulose-1,7-bisphosphate and D-fructose-1,6-bisphosphate, 5-methyltetrahydrofolate, and shikimate) [36].

Then, a total of 819 mass isotopomers were identified in the transitive closure of 14 metabolites. They were transformed into 98 pieces of SCCs with up to  $2.16 \times 10^5$  ODE equations. The ordinary equations were simulated from 0 to 10 s. The step size for constant-step-size method is set to be 0.005 s. The absolute and relative tolerance was set as  $10^{-9}$  and  $10^{-7}$ . 3 different sets of the flux distribution were used for the *S.2973* network. The ODEs parameters are set in Table 2 and flux values are documented in the Additional file 2. The speed of the nonparallel and parallel methods is compared in Fig. 5. The parallel method



**Fig. 2** The framework for SCC of mass isotopomer. **a** A SCC decomposition for the  $m_0$ s of the toy network. The mass isotopomer network was decoupled based on mass weight and network connectivity. **b** The reorganization of mass isotopomer into SCC. Green background is corresponding to the EMU vector and yellow background is corresponding to SCC of mass isotopomer. Subscripts refer to the mass weight and superscripts refer to the code of EMU

obtains an average 15-fold acceleration over the nonparallel method, as shown in Fig. 5. The  $\lambda$  is 10, generated by a grid search for best speed-up.

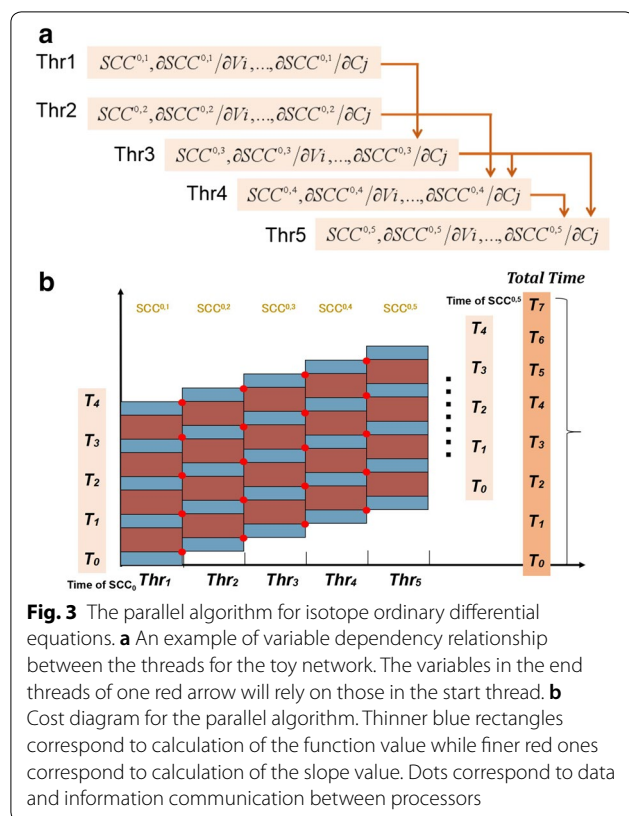
#### Parallelization for the adaptive-step-size ODE at the genome scale

In contrast, the adaptive-step-size method is potentially hard to parallelize since the step size depends on the current value and may differ for different SCCs [40].

Therefore, a universal step size requires computation on all SCCs simultaneously.

Fortunately, the labeling curve of any mass isotopomer is not complicated and can be mimicked by a cubic spline [41]. In particular, the sum of the mass isotopomers of the same EMU equals one while the sum of the mass isotopomers' derivatives relative to particular flux or pool size equals zero. So, the dynamic curves of the mass isotopomers and the derivatives belonging to the same





metabolite have similar steepness and require similar sampling density, even though there is a wide distribution in the concentrations of metabolites. We calculated the trajectory of all mass isotopomers and the derivatives of EMUs of different sizes in a medium-scale carbon metabolic network of *E. coli* (the Additional file 3). The calculation was repeated for 3 sets of flux rate and pool size. The result indicates the evolution curve of different mass isotopomers and their derivatives from the same EMU have the similar changing amplitude in the same time-scale. This phenomenon is exemplified by the curves of all EMUs of AcCoA and isocitrate in Fig. 6a–c, from one dataset. More data are provided in the Additional file 4.

Figure 6d–f shows the mass isotopomers of mass 0 ( $m_0$ ) and their derivatives of different EMUs of the same metabolite for 18 metabolites, according to above parameter settings. The curves belonging to the same metabolites are indeed similar to each other, implying that the step size calculated for  $m_0$  of one EMU from one metabolite is suitable for other mass isotopomers of any EMU of the same metabolite. More data are provided in the Additional file 4.

This property enables us to carry out parallelization even in the adaptive-step-size method. First, a set of  $m_0$  values is constructed from the first few SCCs, which contain at least one EMU of each metabolite. This set is used

to calculate the step size with an adaptive method, which is simultaneously delivered to the SCCs to carry out the parallelized constant-step-size modeling.

The speed-up in the adaptive-step-size method is shown in Fig. 5. The parallel method obtains an average speed-up of 5 times.  $\lambda$  does not exhibit an observable effect because the time-limiting step is the process of calculating the step size for the systems, which is not correlated with  $\lambda$ . The adaptive-step-size method can identify and use the maximal step size in each step while producing an error of an allowed magnitude. This results in a significant reduction in the total number of steps and a decrease in the running time.

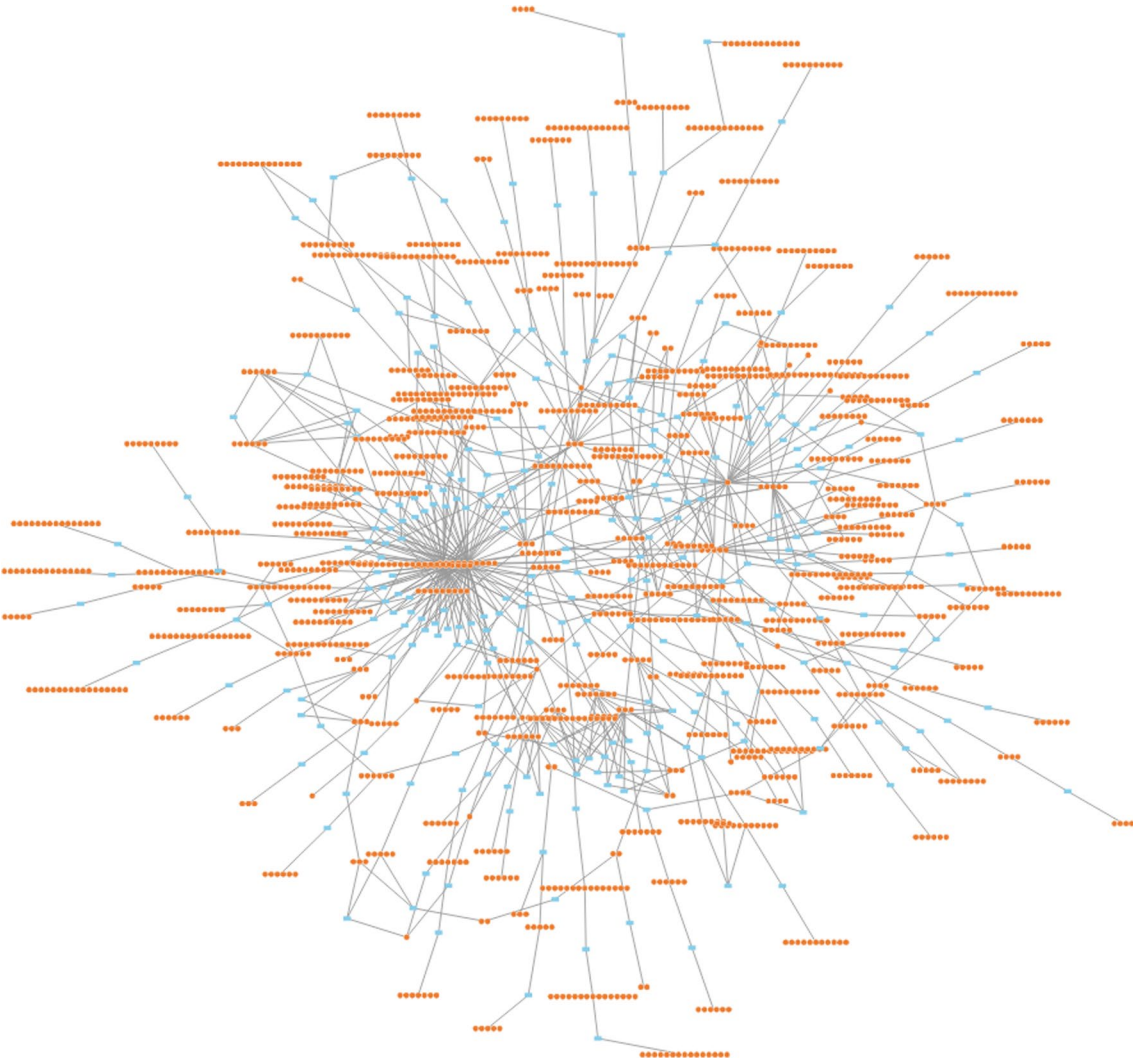
Stiffness problem is an interesting problem for instantaneous  $^{13}\text{C}$  fluxomics. As shown in Eq (7), the ODEs for each SCC are one order and their derivatives are determined by the value of SCC at right side and the SCCs with lower topological sort. Since the value of the SCCs with lower topological sort are known, the derivatives from these components are constant over time. The truncate error will be fixed once the step-size fixed. As such, the stiffness of  $SCC_i$  is major resulted from the square matrix  $M_i$  in Eq (7), which is multiplying with  $SCC_i$ . The stiffness is characterized by this matrix's stiffness ratio, which is the ratio between maximal value and minimal value of the real components of the eigenvalues of the matrix.

We have calculated the eigenvalues of the corresponding matrix of different SCCs with 3 sets of flux distribution on *E. coli* network. All the eigenvalues are negative real number, which shows that the isotope differential equation is stable. The maximal stiffness ratio is less than  $10^4$  for all the SCCs with  $\lambda$  ranging from 5 to 1000, which means the stiffness of these SCCs is not a serious problem.

## Discussion

Modelling a large set of isotope ordinary differential equations in a parallel and faster way have a potentiality to be highly used by the researchers who develop new  $^{13}\text{C}$  fluxomics method and will help to utilizes  $^{13}\text{C}$  fluxomics to explore more species. To this end, a new framework for modeling the isotope is generated. A parallel strategy is realized for isotope ordinary equations modeling and achieved observable speed-up. Some factors affecting the speed of isotope ordinary equations modeling are discussed.

Specifically, we have put forward a new framework for modeling the isotope fate within the atom transition network. This framework is in parallel with current isotope granule EMU, cumomer and bondmer and easy to be implemented. The modeling is based upon a new combination of mass isotopomer. The mass isotopomer



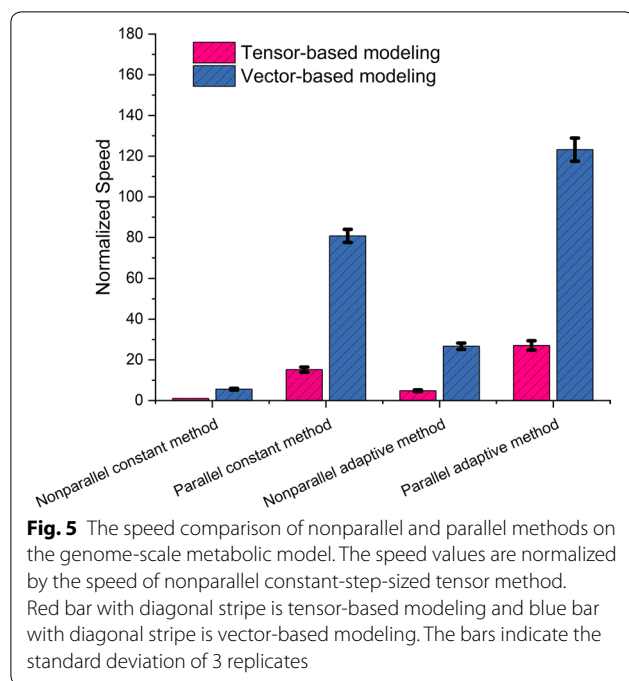
**Fig. 4** The genome-scale atom mapping network modified from *imSyn593*. The light blue rectangles represent the enzymes. The orange circles with white bound represent the carbon atom of a metabolite

**Table 2** ODEs parameters

Parameter	Value	Comment
$\lambda$	10	SCC aggregation parameter, i.e., the minimum mass isotopomer quantity of an SCC
Tn	10	ODEs end time point
Step	0.005	Step size
Tolerance_scaling_factor	$10^{-9}$	Tolerance scaling factor for adaptive method
Tolerance_addition_factor	$10^{-7}$	Tolerance addition factor for adaptive method

generation relationship has been abstracted as a mass isotopomer network. Mass isotopomer network is an important concept here. The equations can be significantly simplified by preserving the transitive closures

of the measured mass isotopomers and cutting out the rest parts. This is like what have been done for cumomer and EMU. The left mass isotopomer network is then decomposed into different SCCs, which is the basic



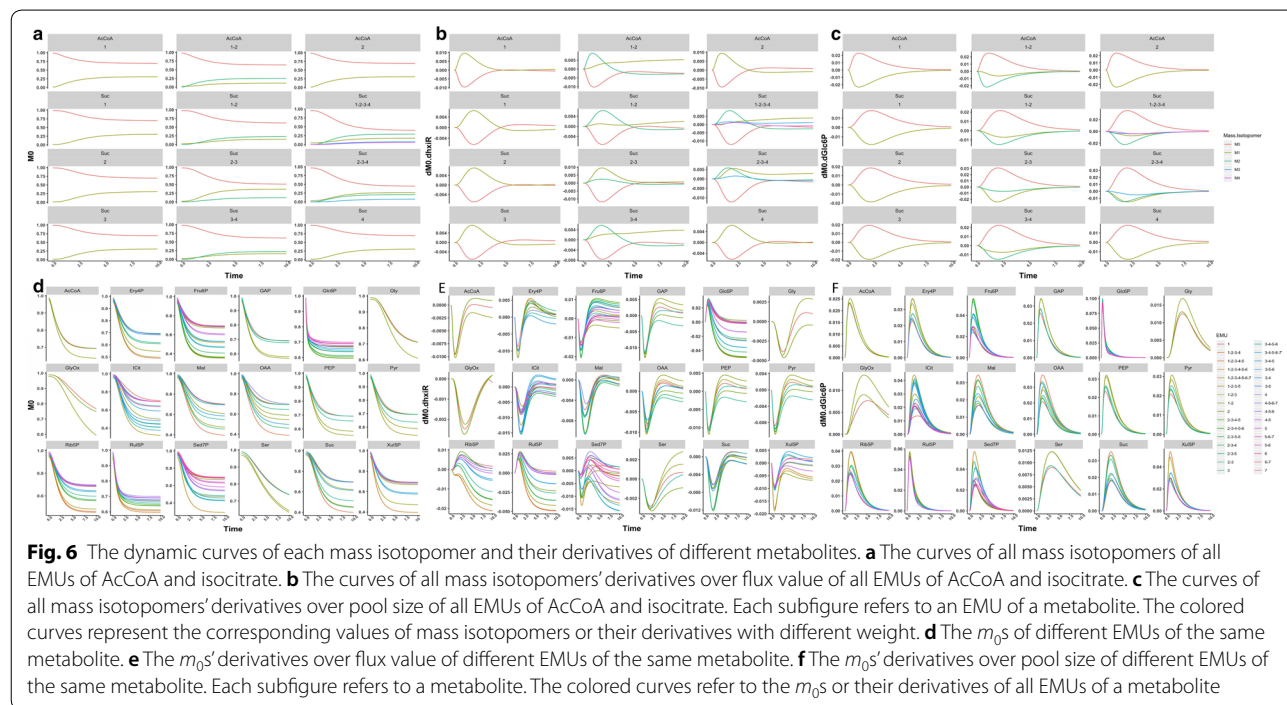
unit for modeling. The modeling can be implemented in two ways, one of which is tensor-based and the other vector-based. The tensor-based method is analogue to the cumomer algorithm from Weitzel's work [38]. The vector-based method is analogue to the EMU block algorithm from Young's work [26].

A parallel strategy of modeling isotope ordinary equations has been realized with the constant-step-size and adaptive-step-size integration methods. Their comparison is as shown in Fig. 5. This strategy is essentially universally applicable and can be adapted in different isotope modeling method and benefit to the  $^{13}\text{C}$  fluxomics community. For the test cases, the vector-based method shows systems advantage in modeling speed over the tensor-based method. This is reasonable because the vector-based method requires only the product of two vectors, while the tensor algorithm requires an additional component. These results confirm the wide applicability of our parallel method.

Many other factors also impact the modeling speed. In addition to the hardware performance such as the number of CPU cores, the speed of the modeling is also affected by the number of EMU reactions and the flux distribution value. When the difference between the reaction values within one flux distribution goes down, the size of time point calculated by the adaptive step size method become smaller and the modeling can be quickly ended. This information is also instructive for modeling based upon EMU and cumomer.

## Conclusion

Isotopically instationary  $^{13}\text{C}$  fluxomics has emerged as the gold standard to obtain a precise picture of the fluxome for photosynthetic microorganisms utilizing a one-carbon substrate. The increased computational



and storage demands may hinder easy adoption when a genome-scale fluxome is required. The parallel algorithm for isotope ODEs here is found to be a successful strategy to promote the speed of instationary  $^{13}\text{C}$  fluxomics, regardless of whether constant-step-size or adaptive-step-size modeling is used. This parallel strategy utilizes the cascade relationship of the isotope granules balance equations. This property also holds true for any other isotope granules, such as EMUs or cumomers. It is essentially universally applicable and will benefit to the  $^{13}\text{C}$  fluxomics community.

## Method description

### Reorganization of EMU mass isotopomers

An EMU can be defined as a specific subset of metabolite atoms [13]. A mass isotopomer is a group of isotopomers that is classified according to the number of heavy isotopes rather than the position [42]. The mass isotopomers here are considered individually, unlike those previously treated in one EMU as a whole. One mass isotopomer is treated as one node in the graph. When a mass isotopomer A is a precursor of mass isotopomer B, we define a directed arrow starting from node A and ending on node B.

Subsequently, according to EMU equations, a directed graph  $G_m (V_m, E_m)$  of the mass isotopomer network can be drawn [38]. The vortex  $V_m$  represents the mass isotopomers, while the directed edge  $E_m$  connects the reactant isotopomers to the product mass isotopomers in an EMU reaction. To remove the unnecessary mass isotopomers, the graph  $G_m$  is first transformed into its transposed graph  $G_m^T$  by reversing each edge of  $G_m$ . Then, the transitive closures of the measured mass isotopomers are identified by a Floyd–Warshall algorithm [43].

Then, the Tarjan algorithm has been implemented to decompose  $G_m$  into different SCCs, which is a subgraph where there exists at least one directed path for any pair

contains a node with an edge directed toward another node in SCC B, then SCC A has a prioritized topological sort over B, and SCC A is called the parent SCC of SCC B. Thus, for all edges from A to B, node A appears before node B. The SCCs with higher sort rely on the SCCs with lower sort, while the latter do not rely on the former.

As a heavier mass isotopomer will never contribute to a lower mass isotopomer in terms of an EMU, one SCC contains mass isotopomers of the same weight. If one directed edge connects SCC A to SCC B, then we define SCC A as having a topological sort less than that of SCC B [38]. The content of an SCC is as follows:

$$SCC = (SCC^{1,1} SCC^{1,2} \dots SCC^{i,j} \dots), \quad (1)$$

where  $i$  represents the weight of the mass isotopomer, while  $j$  represents the topological sort of the SCC for the same weight.

### Tensor-based modeling of the mass isotopomers of SCCs

The instationary framework presented here is based upon the mass isotopomers of SCCs. Generally, any matrix manipulation of isotopes can be represented by a combination of SCC reactions. One SCC reaction with  $n$  reactants and one product has the following form:

$$SCC_1 + SCC_2 + \dots + SCC_n \Rightarrow SCC_p. \quad (2)$$

The generation of the MID of  $SCC_p$  can be calculated as the product of the transition tensor of SCCs of all reactants, such as

$$SCC_p = Q \otimes SCC_{R_1} \otimes SCC_{R_2} \dots \otimes SCC_{R_n}. \quad (3)$$

Here,  $SCC_p \dots SCC_{R_n}$  is the vector of mass isotopomers of product and reactant SCCs sorted by their weights and topological sort.  $Q$  is a transition tensor with order equal to the number of reactant and product SCCs and dimensions equal to the dimensions of the SCCs. The precise definition of the transition tensor can be described as follows:

$$Q_{r,(i_1,i_2,\dots,i_n,j)} = \begin{cases} 1, & \text{if the } i_1 \text{th mass isotopomer of } SCC_{R_1}, i_2 \text{th mass isotopomer of } SCC_{R_2}, \dots, \\ & \text{join together to make the } j \text{th mass isotopomer of } SCC_p \text{ in the } r \text{th reaction} \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

of its vertices [37]. A solitary mass isotopomer shall be treated as an SCC consisting of a single node. If SCC A

The consumption of the mass isotopomer can be formulated by an eliminating matrix. Its content can be defined as follows:

$$E_{k(i,j)} = \begin{cases} -1, & \text{if } i = j \text{ and the } k \text{th reaction consumes the } i \text{th mass isotopomer.} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$



The dimension is equal to the number of mass isotopomers for the corresponding SCC. The ODEs of mass isotopomer SCCs have the following form:

$$\begin{aligned} C_{i_k,j_k} \frac{dSCC^{i_k,j_k}}{dt} = & \sum_p v_p Q_p^{(i_1,j_1),(i_2,j_2),\dots,(i_p,j_p)} \otimes SCC^{i_1,j_1} \otimes SCC^{i_2,j_2} \\ & \dots \otimes SCC^{i_p,j_p} + \sum_c v_c E_c SCC^{i_k,j_k} \\ & + \sum_{inp} v_{inp} Q_{inp} SCC^{inp}. \end{aligned} \quad (6)$$

$SCC^{i_k,j_k}$  is an SCC with the  $j_k$ th topological sort and the  $i_k$ th weight.  $v_p$  is the flux value of reactions producing certain mass isotopomers of  $SCC^{i_k,j_k}$ .  $v_c$  is the flux value of reactions consuming certain mass isotopomers of  $SCC^{i_k,j_k}$ .  $v_{inp}$  is the flux value of the input reactions.  $SCC^{i_1,j_1}$  and so on are the SCCs whose element is involved in  $v_p$  to produce  $SCC^{i_k,j_k}$ .  $i_p$  is the number of reactants of  $v_p$ , while  $j_p$  is the topological sort.  $Q_p^{(i_1,j_1),(i_2,j_2),\dots,(i_p,j_p)}$  is the transition tensor of  $v_p$  as defined in Eq. (6) to adapt the SCC instead of the EMU, which produces  $SCC^{i_k,j_k}$  from  $SCC^{i_1,j_1}$  and so on.  $E_c$  is the eliminating matrix of  $v_c$ .  $Q_{inp}$  is the transition tensor of input reactions.  $C_{i_k,j_k}$  is a diagonal matrix whose diagonal elements are the metabolite pool size associated with  $SCC^{i_k,j_k}$ .

Implicit differentiation of Eq. (3) with respect to the free fluxes and pool size generates the differential equations of first-order derivatives of the measured mass isotopomer equations. The initial conditions for these derivatives are set as zero to solve these equations.

#### Vector-based modeling of mass isotopomers of SCCs

Like EMU [13], mass isotopomer SCCs can also be modeled directly in a vector way as Eq. (7)

$$[M_{j_1} M_{j_2} \dots M_i \dots M_{j_n}] \begin{bmatrix} SCC_{j_1} \\ SCC_{j_2} \\ \dots \\ SCC_i \\ \dots \\ SCC_{j_{k1}} \otimes SCC_{j_{k2}} \end{bmatrix} = C_i \frac{dSCC_i}{dt}, \quad (7)$$

where  $SCC_{j_1}$  and  $SCC_{j_2}$  are the SCCs which contribute to  $SCC_i$  through a single molecule transition. The single molecule transition from certain SCCs to target  $SCC_i$  can be characterized in the matrix  $M$ .

$$M_{j_1}(r_1, \dots, r_2) = \begin{cases} v_l * 1, & \text{if the } r_1 \text{th element of } SCC_{j_1} \text{ give birth to } r_2 \text{th element of } SCC_i \text{ via } v_l \\ 0, & \text{otherwise} \end{cases}, \quad (8)$$

$SCC_{j_{k1}}$  and  $SCC_{j_{k2}}$  are the SCCs which contribute to  $SCC_i$  through a double molecule transition.  $\otimes$  is a user-defined vector product specific to the set of  $SCC_{j_{k1}}$ ,  $SCC_{j_{k2}}$  and  $SCC_i$ . For different set of  $SCC_{j_{k1}}$ ,  $SCC_{j_{k2}}$  and  $SCC_i$ , the content of  $\otimes$  is different.  $M_{j_n}$  is set as a unit matrix to be compatible with such configuration. The sensitivities equations can be organized in a similar way.

#### Parallel algorithm for the ODEs

The decomposed SCCs are a cascade system, and heavier SCCs rely on the value of lower SCCs. Generally, the ODEs are solved sequentially to obtain the time-series data of mass isotopomers. Therefore, parallel implementation of the SCC differential equations substantially accelerates the forward simulation. The parallel strategy is schematically shown in Fig. 4.

1. Each SCC evolution is carried out in one individual thread.
2. For the first thread, a 4th-order Cash–Karp Runge–Kutta scheme is employed to solve the equations about the first  $SCC^{L,1}$ .

2.1 Each time step of  $h_k$  and each value of  $SCC_k^{L,1}$  of this process are preserved and communicated to the downstream threads.

3. For the thread of  $SCC^{i,j}$ , the current value  $SCC_k^{i,j}$  is calculated based upon  $SCC_{k-1}^{i,j}$  together with  $h_k$  and  $SCC_k^{im,jm}$  received from previous threads where  $i_m$  and  $j_m$  are less than  $i$  and  $j$ , respectively. The algorithm is a typical explicit 4th-order Runge–Kutta method as described in Eq. (9) [36, 39].

3.1 Each time step of  $h_k$  and each value of  $SCC_k^{i,j}$  of this process are preserved and communicated to the downstream threads:

$$\begin{cases} SCC_k^{i,j} = SCC_{k-1}^{i,j} + h_k (g + 2g_2 + 3g_3 + g_4) \\ g_1 = f(t_k, SCC_{k-1}^{i,j}, \dots, SCC_k^{im,jm}, \dots) \\ g_2 = f(t_k, SCC_{k-1}^{i,j} + g_1 h_k / 2, \dots, SCC_k^{im,jm} + g_1 h_k / 2, \dots) \\ g_3 = f(t_k, SCC_{k-1}^{i,j} + g_2 h_k / 2, \dots, SCC_k^{im,jm} + g_2 h_k / 2, \dots) \\ g_4 = f(t_k, SCC_{k-1}^{i,j} + g_3 h_k / 2, \dots, SCC_k^{im,jm} + g_3 h_k / 2, \dots) \end{cases} \quad (9)$$

4. Repeat (3).

### Algorithm implementation and parameter setting

The parallel algorithm is performed via ExecutorService. The data communication between threads is executed efficiently through ConcurrentHashMap. Constraint-compatible initial flux distributions are generated by the Python-based sampler OptGPSampler in COBRAPy [44]. A clear protocol to guide the method development has been included in the Additional file 5. The toy network is adopted from a previous report [13]. The central carbon metabolism network of *E. coli* is described in the Additional file 3. The genome-scale metabolic model of *S.2973* was modified from *imSyn593* from a previous study [36]. The computer is equipped with 2 Xeon E7 4820 V2 2.2G CPUs with 8 cores and 512G memory card, which can support 32 threads at most.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13068-020-01737-5>.

**Additional file 1. A genome-scale metabolic network of *Synechococcus* 2973 modified from *imSyn593*.**

**Additional file 2. 3 different sets of the flux distribution of the *S.2973* network.**

**Additional file 3. A medium-scale carbon metabolic network of *E.coli*.**

**Additional file 4. The trajectory of all mass isotopomers and the derivatives of EMUs of different sizes in *E.coli* for 3 set of flux distribution.**

**Additional file 5. Instruction for isotope differential equations parallelization.**

### Abbreviations

AKG: Alpha-ketoglutarate; AcCoA: Acetyl-coA; Ery4P: Erythrulose-4-phosphate; EMU: Elementary metabolite unit; Fru6P: Fructose-6-phosphatase; Glc6P: Glucose-6-phosphate; GAP: Glyceraldehyde 3-phosphate; ICit: Isocitrate; Mal: Malate; OAA: Oxaloacetate; ODE: Ordinary differential equation; SCC: Strongly connected component; PEP: Phosphoenolpyruvate; Pyr: Pyruvate; Rib5P: Ribose 5-phosphate; Ru5P: Ribulose-5-bisphosphate; Ser: Serine; Suc: Succinate; Xul5P: Xylose-5-phosphate; MID: Mass isotopomer distribution.

### Acknowledgements

Not applicable.

### Authors' contributions

TS conceived and designed the study. TS, ZL, YW designed the algorithm. ZZ, ZL wrote code. TS, ZZ wrote the manuscript. YM, ZZ, ZC, JH prepared the figures. XX, YY supervised the study. All authors read and approved the final manuscript.

### Funding

This work was supported by the National Science Foundation of China NSFC (Grant 31760254), the Joint Fund of the Natural Science Foundation of China and the Karst Science Research Center of Guizhou Province (Grant No. U1812401) and the Doctoral Scientific Research Foundation of Guiyang University (Grant GYU-ZRD [2018]-018).

### Availability of data and materials

All data generated or analyzed during this study are included in this published article and its additional files.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors consented on the publication of this work.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup> College of Mathematics and Information Science, Guiyang University, Guiyang, Guizhou, China. <sup>2</sup> Key Laboratory of Information and Computing Science Guizhou Province, Guizhou Normal University, Guiyang, Guizhou, China. <sup>3</sup> College of Computer Science and Technology, Guizhou University, Guiyang, Guizhou, China. <sup>4</sup> School of Mathematics and Sciences, Guizhou Normal University, Guiyang, Guizhou, China. <sup>5</sup> School of Mathematics Sciences and Key Laboratory of Mathematics for Nonlinear Sciences, Fudan University, Shanghai, China. <sup>6</sup> College of Life Science, Guizhou Normal University, Guiyang, Guizhou, China.

Received: 21 November 2019 Accepted: 22 May 2020

Published online: 08 June 2020

### References

- Zamboni N, Sauer U. Novel biological insights through metabolomics and 13C-flux analysis. *Curr Opin Microbiol*. 2009;12:553–8.
- Crown SB, Antoniewicz MR. Publishing 13C metabolic flux analysis studies: a review and future perspectives. *Metab Eng*. 2013;20:42–8.
- Niederführ S, Wiechert W, Nöh K. How to measure metabolic fluxes: a taxonomic guide for 13C fluxomics. *Curr Opin Biotechnol*. 2015;34:82–90.
- Hollinshead WD, Rodriguez S, Martin HG, Wang G, Baidoo EE, Sale KL, Keasling JD, Mukhopadhyay A, Tang YJ. Examining *Escherichia coli* glycolytic pathways, catabolite repression, and metabolite channeling using Deltapfk mutants. *Biotechnol Biofuels*. 2016;9:212.
- Liu N, Qiao K, Stephanopoulos G. (13)C metabolic flux analysis of acetate conversion to lipids by *Yarrowia lipolytica*. *Metab Eng*. 2016;38:86–97.
- Wu C, Xiong W, Dai J, Wu Q. Genome-based metabolic mapping and 13C flux analysis reveal systematic properties of an oleaginous microalga *Chlorella protothecoides*. *Plant Physiol*. 2015;167:586–99.
- Zhao L, Zhang H, Wang L, Chen H, Chen YQ, Chen W, Song Y. (13)C-metabolic flux analysis of lipid accumulation in the oleaginous fungus *Mucor circinelloides*. *Bioresour Technol*. 2015;197:23–9.
- Christen S, Sauer U. Intracellular characterization of aerobic glucose metabolism in seven yeast species by 13C flux analysis and metabolomics. *FEMS Yeast Res*. 2011;11:263–72.
- Metallo CM, Walther JL, Stephanopoulos G. Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. *J Biotechnol*. 2009;144:167–74.
- Qian X, Zhang Y, Lun DS, Dismukes GC. Rerouting of metabolism into desired cellular products by nutrient stress: fluxes reveal the selected pathways in cyanobacterial photosynthesis. *ACS Synth Biol*. 2018;7:1465–76.
- Heux S, Berges C, Millard P, Portais JC, Letisse F. Recent advances in high-throughput (13)C-fluxomics. *Curr Opin Biotechnol*. 2017;43:104–9.
- Wiechert W, Mollney M, Isermann N, Wurzel M, de Graaf AA. Bidirectional reaction steps in metabolic networks: III. explicit solution and analysis of isotopomer labeling systems. *Biotechnol Bioeng*. 1999;66:69–85.
- Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab Eng*. 2007;9:68–86.
- Schmidt K, Carlsen M, Nielsen J, Villadsen J. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol Bioeng*. 1997;55:831–40.
- Zupke C, Stephanopoulos G. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. *Biotechnol Prog*. 1994;10:489–98.
- Zamboni N, Fischer E, Sauer U. FiatFlux—a software for metabolic flux analysis from 13C-glucose experiments. *BMC Bioinform*. 2005;6:209.

17. Quek L, Wittmann C, Nielsen LK, Kromer JO. OpenFLUX: efficient modeling software for  $^{13}\text{C}$ -based metabolic flux analysis. *Microb Cell Fact*. 2009;8:25.
18. Weitzel M, Nöh K, Dalman T, Niedenführ S, Stute B, Wiechert W. 13CFLUX2—high-performance software suite for  $^{13}\text{C}$ -metabolic flux analysis. *Bioinformatics*. 2012;29:143–5.
19. He L, Wu SG, Zhang M, Chen Y, Tang YJ. WUFlux: an open-source platform for  $^{13}\text{C}$  metabolic flux analysis of bacterial metabolism. *BMC Bioinform*. 2016;17:444.
20. Birkel GW, Ghosh A, Kumar VS, Weaver D, Ando D, Backman TW, Arkin AP, Keasling JD, Martín HG. The JBEI quantitative metabolic modeling library (jqMM): a python library for modeling microbial metabolism. *BMC Bioinform*. 2017;18:205.
21. Antoniewicz MR, Kraynie DF, Laffend LA, Gonzalezlegier J, Kelleher JK, Stephanopoulos G. Metabolic flux analysis in a nonstationary system: Fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. *Metab Eng*. 2007;9:277–92.
22. Ahn WS, Antoniewicz MR. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. *Metab Eng*. 2011;13:598–609.
23. Noh K, Wahl A, Wiechert W. Computational tools for isotopically instationary  $^{13}\text{C}$  labeling experiments under metabolic steady state conditions. *Metab Eng*. 2006;8:554–77.
24. Noh K, Wiechert W. Experimental design principles for isotopically instationary  $^{13}\text{C}$  labeling experiments. *Biotechnol Bioeng*. 2006;94:234–51.
25. Toya Y, Ishii N, Nakahigashi K, Hirasawa T, Soga T, Tomita M, Shimizu K.  $^{13}\text{C}$ -metabolic flux analysis for batch culture of *Escherichia coli* and its pyk and pgi gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnol Prog*. 2010;26:975–92.
26. Young JD, Walther JL, Antoniewicz MR, Yoo H, Stephanopoulos G. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnol Bioeng*. 2008;99:686–99.
27. Young JD, Shastri AA, Stephanopoulos G, Morgan JA. Mapping photoautotrophic metabolism with isotopically nonstationary ( $^{13}\text{C}$ ) flux analysis. *Metab Eng*. 2011;13:656–65.
28. Nakajima T, Yoshikawa K, Toya Y, Matsuda F, Shimizu H. Metabolic flux analysis of the *Synechocystis* sp. PCC 6803  $\Delta$  nrtABCD mutant reveals a mechanism for metabolic adaptation to nitrogen-limited conditions. *Plant Cell Physiol*. 2017;58:537–45.
29. Abernathy MH, Yu J, Ma F, Liberton M, Ungerer J, Hollinshead WD, Gopalakrishnan S, He L, Maranas CD, Pakrasi HB. Deciphering cyanobacterial phenotypes for fast photoautotrophic growth via isotopically nonstationary metabolic flux analysis. *Biotechnol Biofuels*. 2017;10:273.
30. Kajihata S, Furusawa C, Matsuda F, Shimizu H. OpenMebius: an open source software for isotopically nonstationary  $^{13}\text{C}$ -based metabolic flux analysis. *Biomed Res Int*. 2014;2014:627014.
31. Young JD. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics*. 2014;30:1333–5.
32. Jazmin LJ, O'Grady JP, Ma F, Allen DK, Morgan JA, Young JD. Isotopically nonstationary MFA (INST-MFA) of autotrophic metabolism. *Methods Mol Biol*. 2014;1090:181–210.
33. Sake CL, Metcalf AJ, Boyle NR. The challenge and potential of photosynthesis: unique considerations for metabolic flux measurements in photosynthetic microorganisms. *Biotechnol Lett*. 2019;41:35–45.
34. Ma F, Jazmin LJ, Young JD, Allen DK. Isotopically nonstationary  $^{13}\text{C}$  flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. *Proc Natl Acad Sci USA*. 2014;111:16967–72.
35. Gopalakrishnan S, Pakrasi HB, Maranas CD. Elucidation of photoautotrophic carbon flux topology in *Synechocystis* PCC 6803 using genome-scale carbon mapping models. *Metab Eng*. 2018;47:190–9.
36. Hendry JJ, Gopalakrishnan S, Ungerer J, Pakrasi HB, Tang YJ, Maranas CD. Genome-scale fluxome of *Synechococcus elongatus* UTEX 2973 using transient ( $^{13}\text{C}$ ) labeling data. *Plant Physiol*. 2019;179:761–9.
37. Tarjan RE. Depth-first search and linear graph algorithms. *SIAM J Comput*. 1972;1:146–60.
38. Weitzel M, Wiechert W, Noh K. The topology of metabolic isotope labeling networks. *BMC Bioinform*. 2007;8:315.
39. Dormand JR, Prince PJ. A family of embedded Runge-Kutta formulae. *J Comput Appl Math*. 1980;6:19–26.
40. Lourakis MI. A brief description of the Levenberg-Marquardt algorithm implemented by levmar. *Foundation Res Technol*. 2005;4:1–6.
41. Horl M, Schnidder J, Sauer U, Zamboni N. Non-stationary ( $^{13}\text{C}$ )-metabolic flux ratio analysis. *Biotechnol Bioeng*. 2013;110:3164–76.
42. Christensen BB, Nielsen J. Isotopomer analysis using GC-MS. *Metab Eng*. 1999;1:282–90.
43. Hougard S. The Floyd-Warshall algorithm on graphs with negative cycles. *Inform Process Lett*. 2010;110:279–81.
44. Megchelenbrink W, Huynen MA, Marchiori E. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLOS ONE*. 2014;9:e86587.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

