Computers in Biology and Medicine 61 (2015) 19–35

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Evolutionary growth of certain metabolic pathways involved in the functioning of GAD and INS genes in Type 1 *Diabetes Mellitus*: Their architecture and stability

Somnath Tagore [a,*], Rajat K. De [b]

[a] Department of Biotechnology and Bioinformatics, Padmashree Dr. D.Y. Patil University, Navi Mumbai 400614, India
[b] Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700108, India

ARTICLE INFO

Article history:
Received 1 August 2014
Accepted 13 March 2015

Keywords:
Community structure
Evolution
Hamming distance
Hasse diagram
Payoff matrix
Tanimoto coefficient

ABSTRACT

Background: Studying biochemical pathway evolution for diseases is a flourishing area of Systems Biology. Here, we study Type 1 *Diabetes Mellitus* (T1D), focusing on growth of *glutamate*, *β-alanine*, *taurine and hypotaurine*, and *butanoate* metabolisms involved in onset of *GAD* and *INS* genes in *Homo sapiens* with comparative analysis in non-obese diabetic *Mus musculus*, biobreeding Diabetes-prone *Rattus norvegicus*, *Pan troglodytes*, *Oryctolagus cuniculus*, *Danio rerio* and *Drosophila melanogaster* respectively.

Methods: We propose an algorithm for growth analysis for four metabolic pathways involved in T1D. It has three modules, *pattern* finding, *interaction* identification and *growth* detection. The first module identifies patterns using *Community structures* using *Hamming distances* and the *Tanimoto coefficient*. We have performed functional analysis by representing patterns using ODEs, and identified *Stoichiometric*, *Gradient* and *Jacobian* matrices. The second module identifies interactions among patterns using cut-sets and network-partitioning by 'Divide-and-conquer'. The third module identifies functions of patterns using interactions, thereby highlighting their nature of growth.

Results: We observed that metabolites that are genetically robust and resist alterations against stable state during evolution, account for emergence of a scale-free network.

Discussion: New modules get acquired to the fundamental cluster in a preferential manner, an instance of *micro-evolution theory*. For instance, *(S)-3-hydroxy butanoyl-CoA, acetoacetyl-CoA, acetoacetate, acetyl-CoA, (S)-3-hydroxy-3-methyl glutaryl-CoA* acts as a fundamental cluster in *butanoate* metabolism. Moreover, the interactions among metabolites are divergent in nature.

1. Introduction

The behavior and evolutionary properties of metabolic pathways have been studied for a long time using architectural and network-based methods [1]. Among various research attempts undertaken in network-based studies, an essential one is based on identifying meaningful patterns within metabolic pathways. It aims at identifying essential regions within metabolic pathways and is useful for studying evolution. Furthermore, it also identifies functionally important metabolites that help in determining alternate paths in metabolic pathways [2,3]. This also help in identifying analogous pathways that exist in different species and provide a key step in discerning essential properties of evolution in terms of interactions among metabolic pathways [4]. To analyze the importance of interactions among metabolic pathways in evolution, we have selected certain metabolic pathways involved in the prevalence of Type 1 *Diabetes Mellitus* (T1D).

Diabetes Mellitus (DM), one of the most prevalent diseases in India, is mostly due to the food habit and lifestyle. We have implemented our proposed algorithm on certain metabolic pathways involved in the emergence of Type 1 *Diabetes Mellitus* (T1D). It is a disorder characterized by high blood sugar, resulting in improper production of insulin. It results in failure and dysfunction of multiple organs. Moreover, till date around 250 genes have been studied for analyzing their role with T1D, one of which is *anti-Glutaric acid decarboxylase* (GAD) that leads to autoimmune processes for *β*-cell destruction [5]. Similarly, the *insulin* gene (INS) is the second well established susceptibility locus in *Diabetes Mellitus*, contributing to about 10% towards T1D susceptibility [6]. We have considered four metabolic pathways which have shown role in the expression of *GAD* and *INS* genes, namely, *glutamate*, *β-alanine*, *taurine and hypotaurine*

* Corresponding author.
E-mail addresses: somnathtagore@yahoo.co.in (S. Tagore), rajat@isical.ac.in (R.K. De).

http://dx.doi.org/10.1016/j.compbiomed.2015.03.012
0010-4825/

and *butanoate* metabolisms, for the purpose of studying their development from an evolutionary point-of-view. This study is very critical, as analyzing the growth history and lineage of these metabolic pathways can lead to understanding the circumstances and role which might be responsible for a diseased state.

The major hypotheses proposed till date regarding metabolic pathway evolution are based on gene duplication events. One of the earliest attempts in this regard has been the 'Retrograde' hypothesis, which suggested that the biosynthesis of metabolites requires sequential activities and sub-step conversions of precursors [7]. Alternatively, the 'Granick' hypothesis suggested that biosynthesis of metabolites are results of forward evolution from simple precursor compounds [8]. Likewise, the 'Patchwork' hypothesis suggested that metabolic pathways have evolved through the reactions of certain primitive enzymes with a variable number of chemically related metabolites, followed by duplication events, resulting in metabolite diversification [9]. Similarly, the 'micro-evolution' theory states that certain divergent features are significant among growing entities in metabolic pathways [10]. Lastly, the 'Semienzymatic' origin of metabolic pathways suggested that certain prebiotic stable metabolites, mediated by specific and non-specific enzymes resulted in their evolution.

Our proposed algorithm has three modules, *first*, 'pattern finding' identifying patterns in metabolic pathways based on structural and functional features. We have used *Hamming distance*, the *Tanimoto coefficient* and *Community structures* for detecting meaningful patterns [11] (discussed in 'Methodology' section). The *second* module, i.e., 'interaction identification' identifies interactions based on the patterns [12], wherein interaction is 'the effect that is conferred upon one biological moiety to another'. Fig. 1 illustrates interactions occurring within carbohydrate metabolism in *H. sapiens* at levels of representations [13]. We use cut-sets and a 'Divide-and-conquer' strategy for studying interactions [14]. For example, in Fig. 2, *butanoate* metabolism is partitioned into S_1 and S_2 by removing a link between *acetoacetyl CoA* and *(S)-3-hydroxybutanoyl CoA*. The *third* module, i.e., 'growth detection' identifies growth using the patterns, their interactions and certain perturbations, for which we have associated weights to metabolites and reaction links (*butanoate* metabolism in Fig. 3). But, analyzing metabolic pathways on the basis of *architectural* features only may not facilitate functional significance. For this purpose, we have identified *cycles* and *Hasse* [15] (*qualitative*), ODEs for identifying *Stoichiometric* [16], *Gradient* [17], *Jacobian* [18] and *payoff* matrices [19] (*quantitative*); and 'bi-directional best hit criterion' and nutrient availability (*functional*) features respectively. For evolutionary studies, we have studied scale-free properties [20] (discussed in 'Methodology' section) (Fig. 4), wherein the evolutionary growth resides on novel link attachment in existing metabolic network, as per the 'micro-evolution' theory. We term this evolutionary process as 'divergent root-to-leaf' growth model (discussed in detail in 'Methodology' section).

2. Methodology

In this section we describe the proposed methodology. We have integrated the pathway related information from 36 public repositories. We initiate with this integration followed by describing the proposed methodology.

2.1. Integration and construction of glutamate, β-alanine, taurine and hypotaurine and butanoate metabolic pathways

We have performed this integration step, since none of the repositories have complete information about the metabolic pathways and suffers from redundancy too. We observed that in case of *glutamate* metabolic pathway, the highest number of links (12) is predicted by KEGG, followed by Gold.db (11), HumanCyc (11), HMDB (11), INOH (11), Kappa-View4 (11), KOBAS (11), MetaCyc (11), MANET (11), PathCase (11) and Pathguide (11). Likewise, the repositories predicting lowest number of links (1) are Ambion Pathway Atlas,

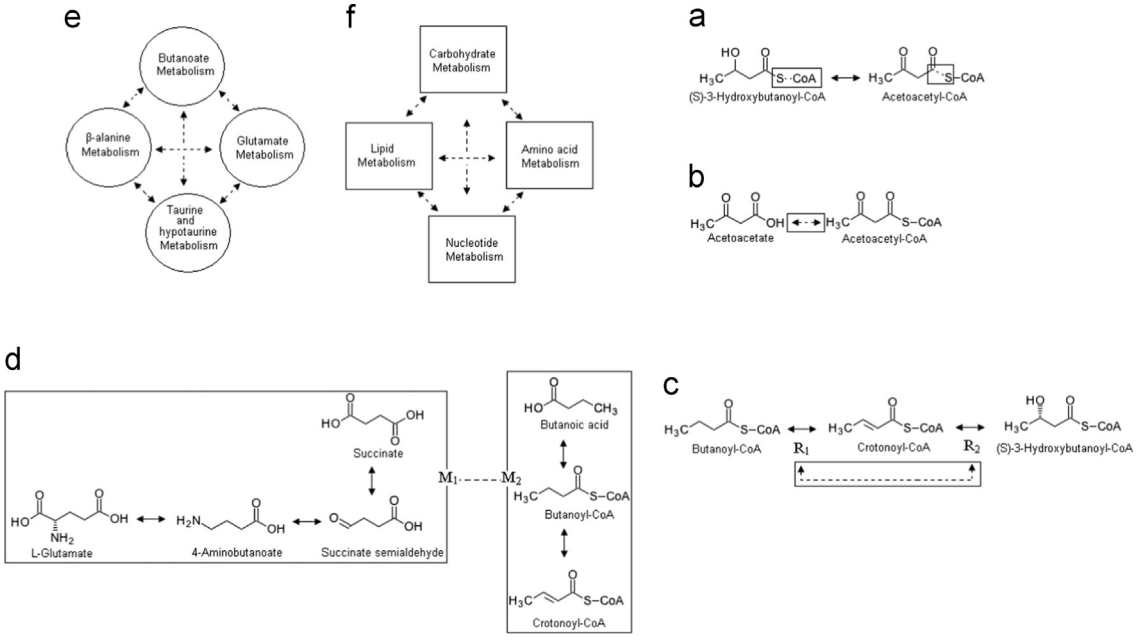

Fig. 1. Types of interactions at (a) atomic level interactions occur among atoms present in metabolites within a metabolic reaction in butanoate metabolism,viz., S-CoA, S-C, (b) metabolite level interactions exist among different metabolites within reactions in butanoate metabolism, viz., among acetoacetyl CoA and acetoacetate, (c) reaction level interactions constitute interactions among the various reactions within a network (denoted by R_i's) in butanoate metabolism, viz., between R1 and R2, (d) module level interactions are existent among different reaction blocks (groups) within pathways (denoted by M_i's) as shown between M_1 and M_2, created on the basis of some common occurring features, (e) pathway level interactions constitute interactions among various pathways present within a metabolic network, as in between butanoate metabolism and other metabolic pathways, (f) network level interactions constitute interactions within various metabolic networks in a cell as illustrated interactions within T1D metabolic network and other metabolic networks within a single cell; dotted lines denote interactions.

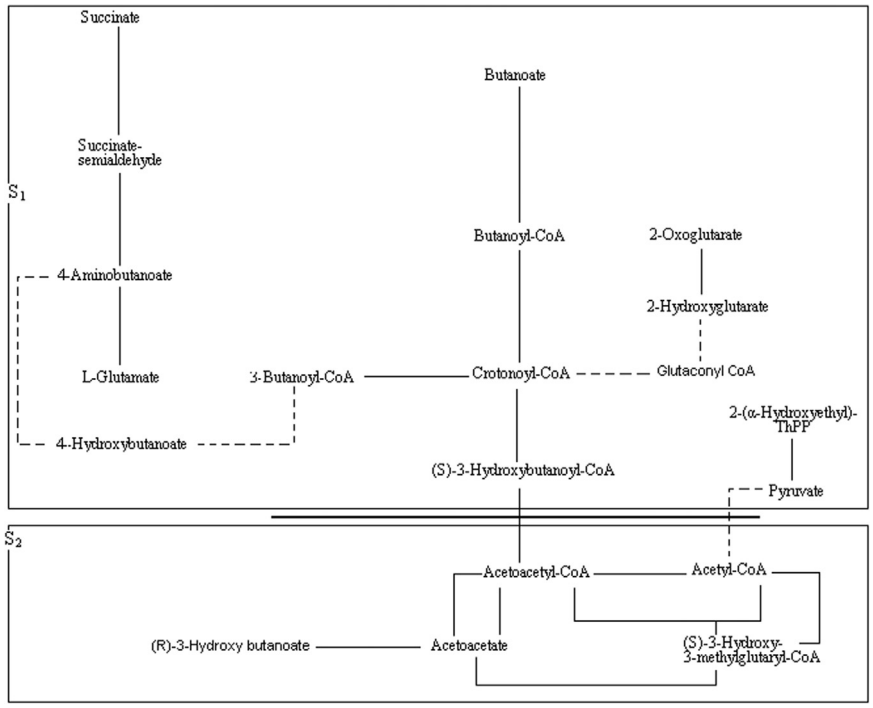

Fig. 2. Cut sets representation of *butanoate* metabolism.

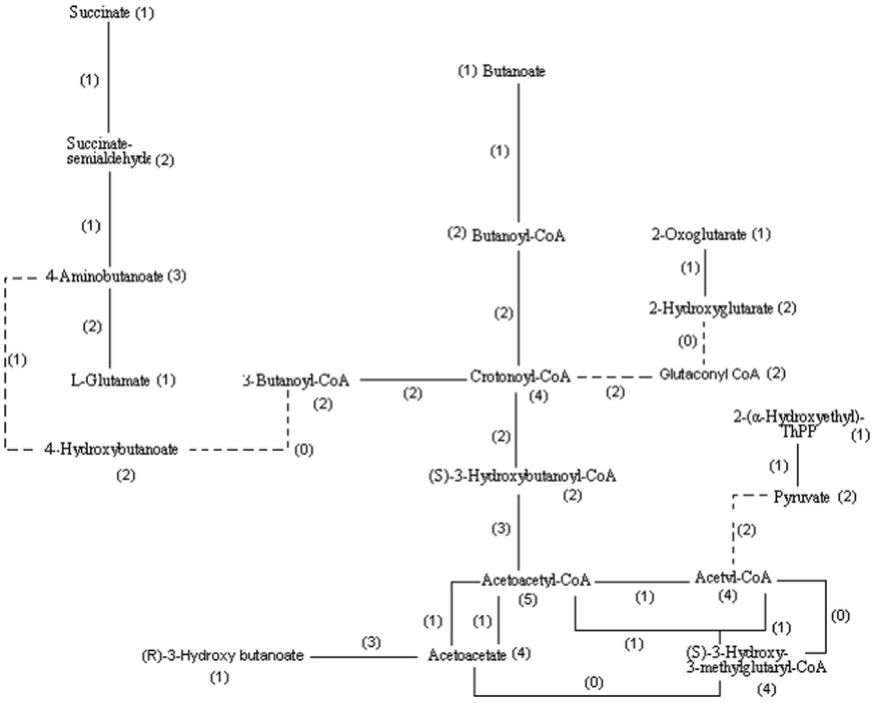

Fig. 3. Weighted graph of *butanoate* metabolism.

BioCarta, BioCyc, BioSilico, Caleydo, CellML, ChemProt, MSEA, Pathway Commons, Reactome and TRMP.

In case of *β-alanine* metabolic pathway, the highest number of links (16) is predicted by BMBP and KEGG, followed by Brenda (15), Pathguide (15) and David (11), whereas the lowest number of links is predicted by PMAP (1), CellML (2), E-Cell (2), Gold.db (2), Modomics (2), PathExpress (2) and Peroxisome DB (2). Similarly, in *taurine and hypotaurine* metabolic pathway, the highest number of links is predicted by Brenda (6), Gold.db (6), HumanCyc (6), HMDB (6), INOH (6), Kappa-View4 (6), KEGG (6), MetaCyc (6), MANET (6), PathCase (6), Pathguide (6) and SMPDB (6), whereas the lowest number of links (1) is predicted by E-Cell and PMAP. Lastly, in *butanoate* metabolic pathway, the highest number of links (15) is predicted by Gold.db, INOH and KEGG, whereas the lowest number of links (2) is predicted by Bionemo and ChemProt.

Moreover, all these repositories are different from one another in terms of the information that is stored, offered, visualized and maintained in them. Thus, in addition to quality, there exists a problem of lacking specific biological or experimental facts listed in the repositories [21]. Thus, for providing the most correct input, we chose

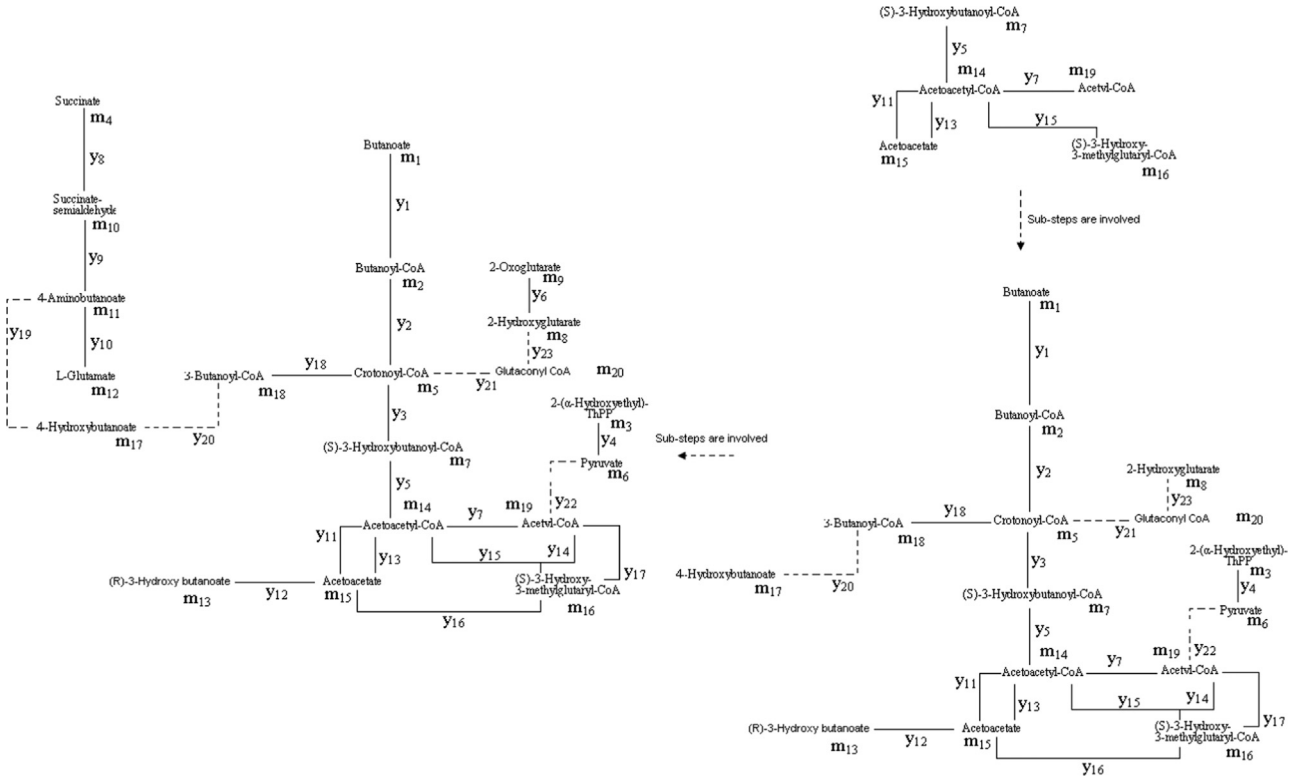

Fig. 4. Tracing the path of growth in *butanoate* metabolism (only a section of the metabolic pathways is displayed).

to integrate and construct the above four metabolic pathways. The constructed metabolic pathways are given in Fig. 5. Moreover, since our algorithm accepts SBML file format as input, we have constructed the SBML format of all the four constructed metabolic pathways using XML Editor [22]. As discussed in the 'Introduction' section, our algorithm consists of three modules, namely, *pattern finding*, *interaction identification* and *growth detection* (Fig. 6).

2.2. Finding patterns

This module deals with identifying patterns from metabolic pathways based on commonly occurring functionally related sequence of metabolites. The community structure-based approach detecting patterns provides information about how metabolic pathways are different topologically. Certain approaches for community finding are already existent which includes minimum-cut method, hierarchical clustering and modularity maximization [8]. Here, we have used Hamming distance as a basis to identify and group community structures.

We describe our algorithm taking into consideration *butanoate* metabolism in *Homo sapiens*. This module (finding patterns) has nine sub-steps, namely, building adjacency matrix, Hamming distance calculation, community structure identification, pattern identification, cycle detection, ODE conversion, stoichiometric matrix buildup, Jacobian matrix buildup and eigenvalue calculation. The first sub-step is creating an adjacency matrix between all metabolites involved within the pathway with a '1' placed in the matrix if a reaction occurs between the corresponding metabolites. The second sub-step calculates the Hamming distance, where for two rows in the adjacency matrix, it is 'the number of positions where values (1 or 0) are different' [23]. For instance, if the two rows are 1101001 and 1001011, then Hamming distance is 2. The third sub-step deals with grouping metabolites into community structures using similar Hamming distance values [24] (Pseudocode 1). For example, metabolite pairs having Hamming distance 3 are grouped into one community structure. The fourth sub-step module identifies patterns for metabolites which are already grouped based on community structures.

Pseudocode 1. Community (*adj*).

input adjacency matrix *adj*
output communities for metabolites m_j
  for each m_j in *adj* **do**
    compare connectivities within m_j pairs
    predict dissimilarities for all values
    store number of dissimilarities in dissimilarity array *dis*
    push m_j in array *cstr* using identical values of *dis*
endfor

Each pattern consists of two segments, *first*, the longest branch and *second*, the sub-branches [25]. The longest branch consists of 1st line of each pattern, followed by other sub-branches denoted with prefix '∼'. The character '– –' is used to represent whether there is a direct connectivity among neighboring metabolites. For instance, in *butanoate* metabolism the patterns found for Hamming distance 4 are [*butanoyl coA*]–[*crotonoyl coA*]–[*S-3-hydroxybutanyl coA*], suggesting that there are direct links occurring between *butanoyl coA* and *crotonoyl coA* as well as *crotonoyl coA* and *S-3-hydroxybutanyl coA* (Table 1).

The patterns so identified are further validated by identification of cycles (fifth sub-step). For every visited metabolite *M*(*usedindex*), if there is an adjacent metabolite *M*(*unusedindex*) such that *M* (*unusedindex*) is already visited and *M*(*unusedindex*) is not predecessor of *M*(*usedindex*), then a cycle *cy* is existent in pattern *pat*. If such an adjacent *M*(*unusedindex*) is non-existent for any metabolite, a cycle does not exist (Pseudocode 2). For example, we identified two cycles in *butanoate* metabolism, namely, *acetoacetyl CoA–acetoacetate–(S)-3-hydroxy-3-methyl glutaryl CoA* (cycle 1) and *acetoacetyl CoA–acetoacetate–acetyl CoA–(S)-3-hydroxy-3-methyl glutaryl CoA* (cycle 2) respectively (Fig. 7) [26].

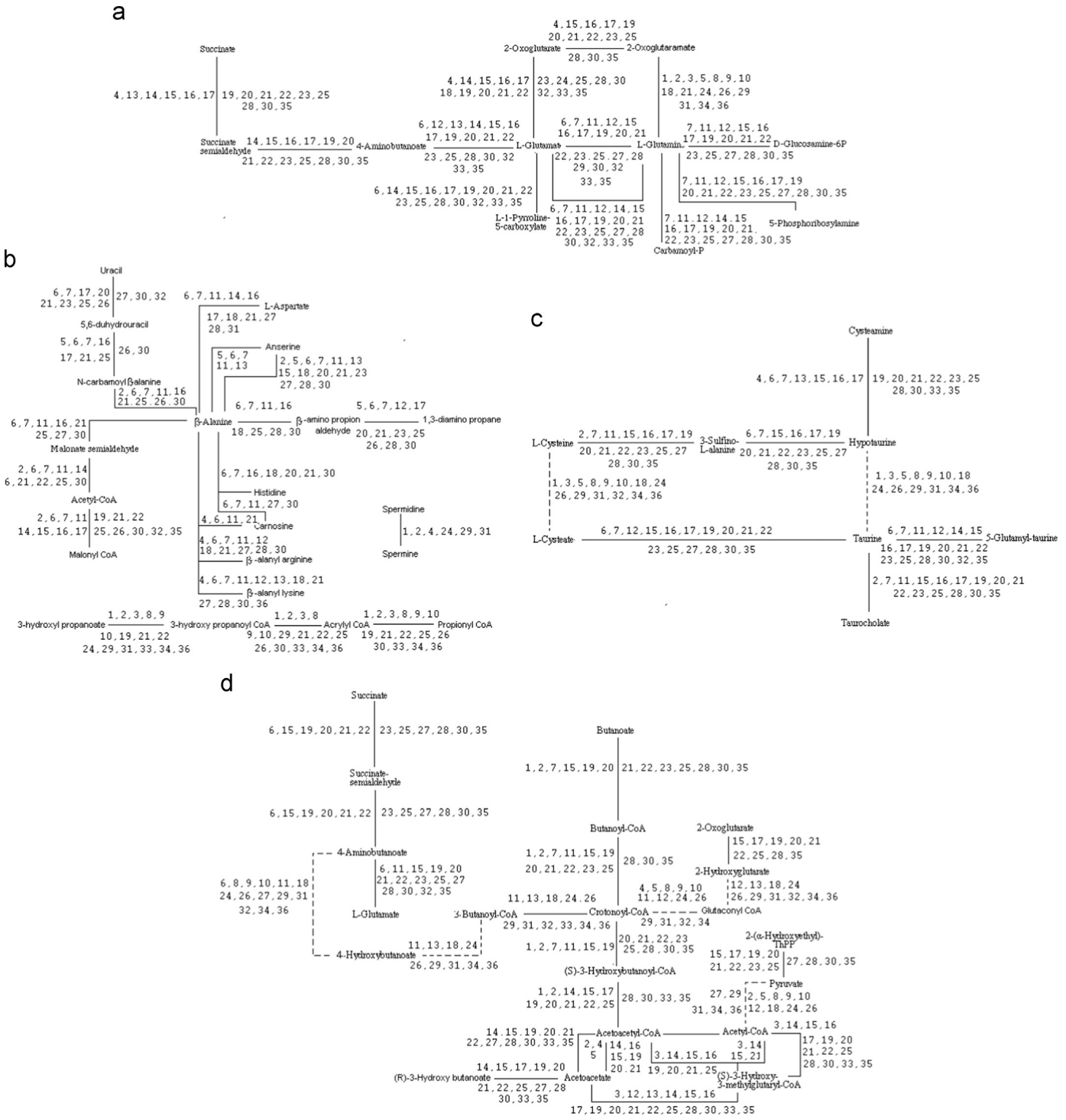

Fig. 5. Integrated metabolic pathways using our algorithm (a) *Glutamate* metabolism, (b) *β-alanine* metabolism, (c) *Taurine* and *hypotaurine* metabolism, and (d) *Butanoate* metabolism (numbers denote the repositories as listed in Table 20 in Supplementary Information).

Pseudocode 2. Cycle (*pat*).

input pattern *pat*
output cycle *cy*
$M(index) \leftarrow index$
$M(unusedindex) \leftarrow index$
$index \leftarrow index + 1$
$cy \leftarrow m_j$
for each $(m_j, m_{j'})$ in *pat* **do**
  if $m_j(index)$ is undefined **then**
    push m_j into new cycle *newcy*
    $M(usedindex) \leftarrow \min(M(usedindex), M^l(usedindex))$
  elseif $m_{j'} cy$ **then**
    $M(usedindex) \leftarrow \min(M(usedindex), M^l(usedindex))$
  endif
endfor **if** ($M(unusedindex)$ equals $M(index)$) **then**
  initiate *newcy* **Repeat**
  $M \leftarrow (usedindex)$ **pop** *cy*
  $newcy \leftarrow m_{j'}$
  endif
endfor

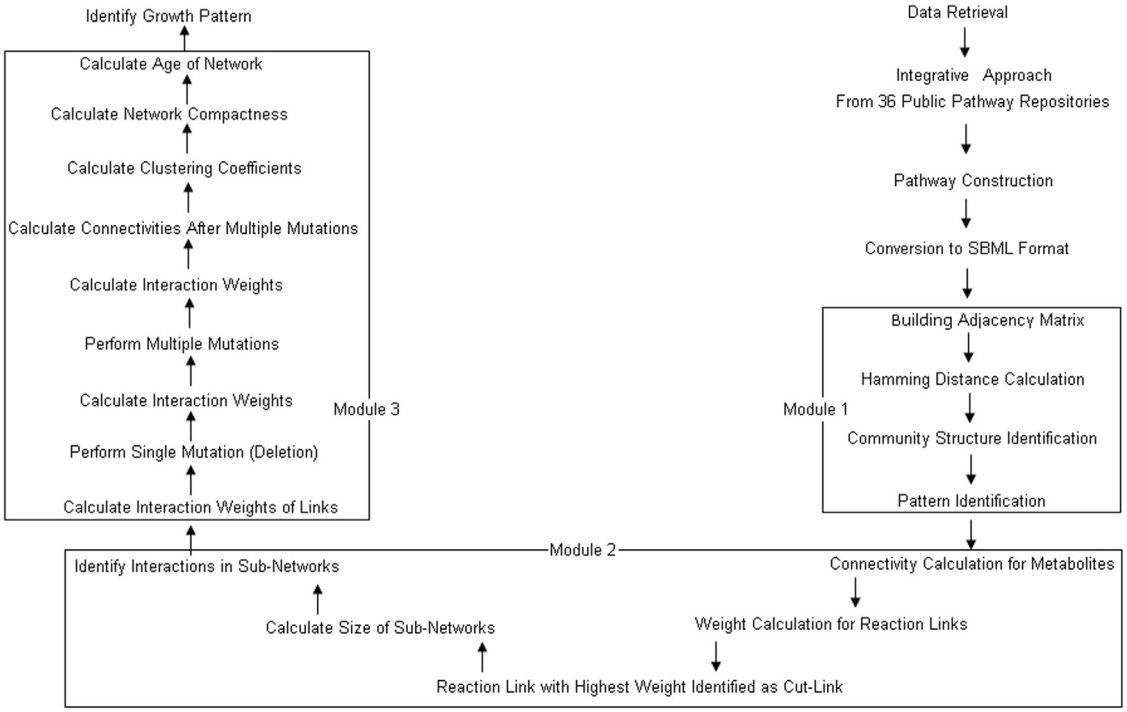

Fig. 6. Flow chart of our strategy.

Table 1
The Patterns identified in *glutamate*, *β-alanine*, *taurine* and *hypotaurine* and *butanoate* metabolisms in *Homo sapiens*.

Metabolic pathway	Hamming distance	Pattern
Glutamate metabolism	3	[L-*glutamate*]–[L-*glutamine*] [L-*glutamate*]–[2-oxoglutarate] [L-*glutamate*]–[4-aminobutanoate] [L-*glutamate*]–[L-*1-pyroline 5-carboxylate*]
β-Alanine metabolism	3	[L-*aspartate*]–[L-*asparagine*] [L-*aspartate*]–[*Oxaloacetate*] [L-*aspartate*]–[L-*arginino succinate*] [L-*aspartate*]–[L-*alanine*]
	4	[*Oxaloacetate*]–[*Pyruvate*] [*Oxaloacetate*]–[*N-acetyl*-L-*asparate*] [*Oxaloacetate*]–[*D-aspartate*]
Taurine and hypotaurine metabolism	3	[*Taurine*]–[L-*cysteate*] [*Taurine*]–[*taurocholate*] [*Taurine*]–[*hypotaurine*] [*Taurine*]–[5-glutamyl-taurine]
Butanoate metabolism	3	[*butanoate*]–[*butanoyl coA*]–[*crotonyl coA*]–[*(S)-3-hydroxybutanoyl coA*]–[*acetoacetyl coA*]–[*acetoacetate*]–[*HAS-3-hydroxybutanoate*] [*acetoacetyl coA*]–[*acetyl coA*]
	3	[L-*glutamate*]–[4-aminobutanoate]–[*succinate semialdehyde*]–[*succinate*]
	3	[2-acetolactate]–[2-α hydroxymethyl ThiPP]–[*pyruvate*]
	4	[*butanoyl coA*]–[*crotonoyl coA*]–[*S-3-hydroxybutanyl coA*]

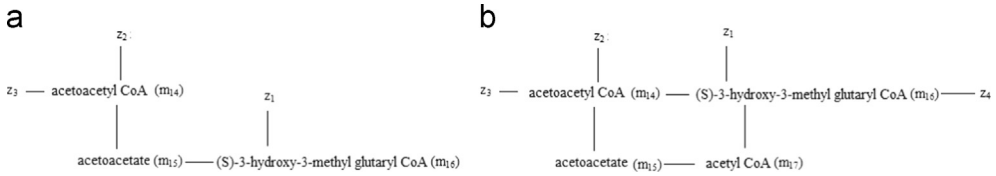

Fig. 7. Cycles (1 and 2) (a) and (b) as identified in *butanoate* metabolism.

The cycles thus identified are subjected to next set of functional validations by conversion into sets of ordinary differential equations (ODEs) (sixth sub-step). In this regard, if concentration of a metabolite m_j is described by C, its change over time is represented by $dC(m_j)/dt$, where t represents time [17]. For instance, in case of *butanoate* metabolism, the series of ODEs for *acetoacetyl CoA* (m_{14}), *acetoacetate* (m_{15}) and *(S)-3-hydroxy-3 methyl glutaryl CoA* (m_{16}) are

$$m_{14} = r_1 z_{15} - m_{15}(r_2 + r_4) \rightarrow \frac{d(m_{14})}{dt} = r_1 \frac{d(z_{15})}{dt} - (r_1 + r_2)\frac{d(m_{15})}{dt} \quad (1)$$

$$m_{15} = r_2 m_{14} + m_{16} r_3 \rightarrow \frac{d(m_{15})}{dt} = r_2 \frac{d(m_{14})}{dt} + r_3 \frac{d(m_{16})}{dt} \quad (2)$$

$$m_{16} = r_4 m_{14} - r_3 m_{16} \rightarrow \frac{d(m_{16})}{dt} = r_4 \frac{d(m_{14})}{dt} - r_3 \frac{d(m_{16})}{dt} \quad (3)$$

respectively.

The seventh sub-step is calculating the stoichiometric matrix for the corresponding cycles and ODEs. We first identify the stoichiometric matrix for the complete metabolic pathways and use them as reference maps for validation. For instance, in *butanoate* metabolism, the stoichiometric matrix (SM) for the

Table 2
SM for *butanoate* metabolism pattern (Cycle 1).

Metabolites	r_1	r_2	r_3	r_4
m_{14}	1	−1	0	−1
m_{15}	0	1	1	0
m_{16}	0	0	−1	1
z_{15}	−1	0	0	0

Table 3
SM for *butanoate* metabolism pattern (Cycle 2).

Metabolites	r_1	r_2	r_3	r_5	r_6	r_7	r_8
m_{14}	1	−1	0	−1	0	0	0
m_{15}	0	1	1	0	0	1	0
m_{16}	0	0	−1	0	1	0	0
m_{17}	0	0	0	1	−1	0	1
z_{15}	−1	0	0	0	0	0	0
z_{16}	0	0	0	0	0	−1	0
x_{17}	0	0	0	0	0	0	−1

Table 4
JM (based on concentration) for *butanoate* metabolism.

	Acetoacetyl-CoA (m14)	Acetoacetate (m15)	(S)-3-Hydroxy-3-methylglutaryl-CoA (m16)	Acetyl-CoA (m17)
L-Glutamate	0	0	0	0
3-Butyn-1-al	0	0	0	0
4-Aminobutanoate	0	0	0	0
3-Butynoate	0	0	0	0
Succinate	0	0	0	0
Succinate semialdehyde	0	0	0	0
(R)-3-Hydroxybutanoate	0	0.1	0	0
Acetoacetate	0.1	−0.3	0.1	0
1-Butanol	0	0	0	0
Butanal	0	0	0	0
Butanoic acid	0	0	0	0
Butanoyl-CoA	0	0	0	0
Crotonoyl-CoA	0	0	0	0
(S)-3-Hydroxybutanoyl-CoA	0.1	0	0	0
Acetoacetyl-CoA	−0.3	0.2	0	0.1
2-Acetolactate	0	0	0	0
(S)-3-Hydroxy-3-methylglutaryl-CoA	0	0	−0.2	0
2-(alpha-Hydroxyethyl) thiamine diphosphate	0	0	0	0
Thiamin diphosphate	0	0	0	0
2-Hydroxyglutarate	0	0	0	0
2-Oxoglutarate	0	0	0	0
Pyruvate	0	0	0	0
Acetyl-CoA	0.1	0	0.2	−0.1

corresponding cycles (1 and 2) as well as ODEs are represented in Tables 2 and 3 respectively. The eighth sub-step deals with converting *SM* to Jacobian matix (*JM*). The *JM* is composed of both *SM* and gradient matrix (*GM*), i.e.,

$$(JM) = (SM) \cdot (GM) \tag{4}$$

We can say that

$$\frac{d(C(t))}{dt} = (SM) \cdot f(C, p) \tag{5}$$

where C is an a-dimensional concentration vector of metabolites, f is a b-dimensional flux vector, and p represents a set of rate parameters [19]. Thus, if we linearize the expression of $d(C(t))/dt$,

Table 5
Eigenvalues for patterns identified in the metabolic pathways.

Eigenvalues	Real	Imaginary
Glutamate metabolism		
m_1	−0.06799	0
m_2	−0.1	0
m_3	−0.17503	0
m_4	−1.00449	0
β-alanine metabolism		
m_5	−0.01487	0
m_6	−0.06496	0
m_7	−0.1	0
m_8	−0.144	0
Taurine and hypotaurine metabolism		
m_9	−0.05	−0.0866
m_{10}	−0.05	0.086603
m_{11}	−0.1	0
m_{12}	−0.1	0
m_{13}	−0.2	0
Butanoate metabolism		
m_{14}	−0.01154	−0.05897
m_{15}	−0.01154	0.058974
m_{16}	−0.1	0
m_{17}	−0.48852	0

we can easily identify the Taylor series expansion of reaction rates,

$$f(C) = f(C_0) + \frac{1}{1!}\frac{df}{dC}|_{C_0} \cdot (C - C_0) + \frac{1}{2!}\frac{d^2f}{dC^2}|_{C_0} \cdot (C - C_0)^2 + \frac{1}{3!}\frac{d^3f}{dC^3}|_{C_0} \cdot (C - C_0)^3 + \cdots \tag{6}$$

i.e.,

$$f(C) \approx f(C_0) + \frac{df}{dC}|_{C_0} \cdot (C - C_0) \tag{7}$$

Now, for steady state we have

$$(SM) \cdot f(C_0) = 0 \tag{8}$$

i.e.,

$$\frac{d(C - C_0)}{dt} = (SM) \cdot \frac{df}{dC}|_{C_0} \cdot (C - C_0) \tag{9}$$

$$\therefore\ (GM) = \frac{df}{dC} \tag{10}$$

$$(JM) = (SM) \cdot \frac{df}{dC} \tag{11}$$

For instance, in case of *butanoate* metabolism *JM* for the *SM* is represented in Table 4.

The ninth sub-step identifies eigenvalues from *JM* using Newton's finite difference method. Considering a synthetic system, $\mathrm{Func}(x) = 0$, where $\mathrm{Func} = 0$ is a function of a variable x (in steady state), we implement Newton's method by initial approximation of x_0 for obtaining a solution. Thus, we solve the expression

$$\mathrm{Func}_x(x^{param}) q_1^{param} = -\mathrm{Func}(x^{param}) \tag{12}$$

where *param* is a constant and q_1 is a variable, by performing an approximation operation like

$$x^{param+1} = x^{param} + q_1^{param} \tag{13}$$

Furthermore, for applying Newton's method, *JM* (in this case, $\mathrm{Func}_x(x^{param})$) needs to be nonsingular. Moreover, if Jacobian at the root is nonsingular, then the Jacobian will be nonsingular for a x^{param} close to the root. In our case, for solving $\mathrm{Func}(x) = 0$, we can have two set of equations,

$$a_1 q_1 - tb_1 q_1 - lc_2 q_1 = 0 \tag{14}$$

and

$$a_2q_2 - tb_2q_2 - lc_2q_2 = 0 \quad (15)$$

Thus,

$$\text{Func}(x) = \begin{bmatrix} a_1q_1 - tb_1q_1 - lc_1q_1 - \frac{1}{2}q_1^Tq_1 + \frac{1}{2} \\ a_2q_2 - tb_2q_2 - lc_2q_2 - \frac{1}{2}q_2^Tq_2 + \frac{1}{2} \end{bmatrix}$$

and $x = [q_1;\ t;\ q_2;\ l]^T$ is a root of Func and q_2 is a variable [19]. For instance, in case of *butanoate* metabolism, the eigenvalues for the complete pathway as well as the patterns are shown in Table 5.

2.3. Finding interactions among patterns

The second module identifies interactions using *Hasse* diagrams and *cut-sets*. Hasse diagrams identify smaller reaction sets from metabolic pathways and provide a hierarchical view with smaller sets at the bottom (Fig. 8). Likewise, cut-set detection initiates with calculating the *degree* of metabolites, followed by assigning weights (w) to all reaction links (Pseudocode 3).

Pseudocode 3. Cuts (pat, rn).

input pattern *pat*, weight vector for reaction links *rn*
output cut links *cutedge*
  for each m_j in *pat* **do**
    calculate degree(m_j) in *pat*
    add highest degree of metabolites hd(m_j) to *mat*
    identify edges of hd(m_j)
    push highest metabolic reaction link *highestrn* of *rn* in cut-set edge *cutedge*
    partition *pat* based on *cutedge*
  endfor

This is followed by applying the 'Divide-and-conquer' (DC) method for back-validating the cuts and group metabolites based on similarity features using *HD* and *TC* [39] (Pseudocode 4). We first create an 'incidence matrix' (*IM*) for detecting metabolite similarities (Tables 7a–d and 8 in Supplementary Information). For instance, in case of *butanoate* metabolism *HD* is found to be 1 for all, whereas *TC* is found for (*acetyl-CoA, acetoacetyl-CoA*)=0.79, (L-*glutamate, 4-butanoate*)=0.52. Next, we we find 'Strong' links (+1), those which are most essential for *GAD* and *INS* genes functionality, and 'weak' links (−1), those which are of lesser importance as compared to strong links. For *butanoate* metabolism considering *acetyl-CoA* as 'basis node' creates a weak links with *acetoacetyl CoA,* L-*glutamate* and 4-butanoate. Similarly, considering *acetoacetyl CoA* as 'basis node' creates a weak links with all, whereas for L-*glutamate* to be considered as 'basis node' creates strong link with 4-butanoate (Fig. 10 in Supplementary Information). The metabolite groups are, namely, Group1={L-*glutamate,* L-*glutamine,* L-*1-pyrroline-5-carboxylate*}, Group 2={*β-alanine,* L-*aspartate, oxaloacetate,* L-*arginino-succinate*}, Group 3={*taurine, hypotaurine,* L-*cysteate, 5-glutamyl-taurine*}, and Group 4={L-*glutamate, 4-butanoate, acetyl-CoA*}. Furthermore, we have done a cross-validation for *DC* by identifying the 'strongly connected components' for the metabolic pathways, consisting of functionally significant metabolite groups having strong intra-interactions leading to essential gene functionality and vice-versa. In case of *butanoate* metabolism, only 1 strong component exists, namely, *butanoate,* L-*glutamate; acetoacetyl CoA, acetyl-CoA* (Fig. 10 in Supplementary Information). Thus, we observe that the most essential metabolites, namely, *acetyl-CoA, acetoacetyl CoA,* L-*glutamate, 4-butanoate* (in *butanoate* metabolism), that are actively responsible for the functioning of *GAD* and *INS* genes are present in strong components.

Pseudocode 4. DC (pat).

input pattern *pat*
output cut links *cutedge*
  for each m_j in *pat* **do**
    calculate $TC(m_j)$ for all pairs in *pat*
    denote *basis*(m_j)
    identify *highestTC*(m_j)
    denote *highestTC*(m_j) as *stronglink*
    denote other $TC(m_j)$ as *weaklink*
  endfor

Next, we arrange the metabolites having highest to lowest weights followed by identifying the reaction link between the highest and next highest weighted metabolite. Now, we calculate size (φ) of the sub-pathways, using the formula, $\varphi = \Sigma w_j y_j$. Size of the metabolic pathway denotes the boundary assignment of the metabolites with higher the size lower the robustness and more the chance of metabolic pathways being targeted for external perturbation [27].

Next, we identify the interactions present in *butanoate* metabolism. Here, we have categorized the interactions into *neighbored*, *serial*, *parallel*, *convergent* and *divergent*. An interaction is said to be *neighbored* if the number of metabolites in one sub group equals that of the other sub groups. For instance in Fig. 11a (in Supplementary Information), sub groups s_1 and s_2 have three metabolites each. An interaction is said to be *serial* if the number of sub groups is atleast two and there is one to one interaction between the metabolites. For instance in Fig. 11b (in Supplementary Information), there are only two possible sub groups, s_1 and s_2. An interaction is said to be *parallel* if the number of sub groups is greater than two. For instance in Fig. 11c (in Supplementary Information), there are only three possible sub groups, s_1, s_2 and s_3 [28]. An interaction is said to be *convergent* if average degree (avgd) of metabolites of one sub group is less than highest degree (hd) as well as the avgd of second sub group. For instance in Fig. 11d (in Supplementary Information) s_1 has an avgd=2.8 whereas for s_2 avgd=3.1, avgd (s_1) < hd(s_1)=3 and avgd (s_1) < avgd(s_2). An

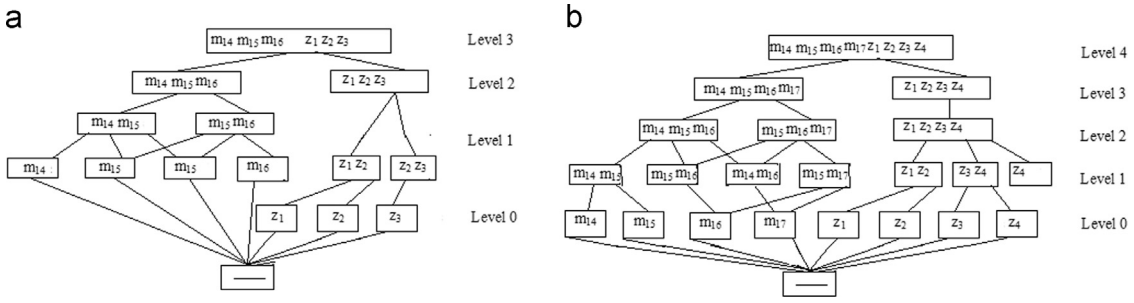

Fig. 8. *Hasse* diagram representation of Cycles (1 and 2) in *butanoate* metabolism.

interaction is said to be *divergent* if avgd of metabolites of one sub group is less than hd and is greater than avgd of the other sub group. For instance in Fig. 11e (in Supplementary Information), s_1 has an avgd=3 whereas for s_2 avgd=2.6, avgd $(s_1) < \text{hd}(s_1)=4$ and avgd $(s_1) > \text{avgd}(s_2)$. In case of butanoate metabolism, only two sub groups are possible and the number of metabolites in each of the sub groups is unequal. Similarly, avgd $(s_1)=1.3$ whereas avgd $(s_2)=3.2$, thus avgd $(s_1) < \text{hd}(s_1)$ and avgd $(s_1) < \text{avgd}(s_1)$, showing convergent interaction. In case of butanoate metabolism, avgd $(s_1)=1.36$ and avgd $(s_2)=3.2$, hd $(s_1)=2$, hd $(s_2)=4$. Thus, avg $(s_1) < \text{hd}(s_1)$ and avg $(s_2) < \text{hd}(s_2)$, avg $(s_1) <$ avg(s_2) for $\varphi(s_2)=71$, elucidating the fact that the evolution was divergent in nature [29].

We perform interaction validation using 'bootstrapping' [30] (Pseudocode 5).

Pseudocode 5. Bootstrap(*hasse*).

```
input pattern hasse
output tree linktree
number of boots numboots ← 100
pattern length patlen ← length(pattern of hasse)
boots ← cell(numboots,1)
  for n ← 1:numboots do
    reorderindex ← random sample function randsample
   (patlen,patlen,true)
      for i ← length of hasse-1 to 1 do
        header of bootseq ← header of hasse
        pattern of bootseq ← strrep(Pattern(reorderindex) of
   hasse with '-')
      endfor
    bootsn ← bootseq
  endfor
fun ← linkage function linkage(x,'average',[header of hasse])
boottrees ← cell(numboots,1)
  for i ← length of hasse-1 to 1 do
    branchpointer ← i + length of hasse
    sub-tree graph subtree ← subtree(orighassetree,
   branchpointer)
    origpointers[i] ← getcanonical(subtree)
    origdata[i] ← sort(get(subtree))
  endfor
  for j ← numboots−1 to 1 do
    for i ← length of hasse-1 to 1 do
      branchptr ← i + length of hasse
      subtree ← subtree(boottreesj,branchptr)
      clusterspointersi,j ← getcanonical(subtree)
      clustersdatai,j ← sort(get(subtree))
    endfor
  endfor
count ← zeros(length of hasse-1,1)
Pc ← count · /numboots
  for i ← 1 to length of hasse -1 do
    branchptr ← i + length of hasse
    names[branchptr] ← [names[branchptr]]
  endfor
```

2.4. Identifying evolutionary growth of metabolic pathways

The third module identifies lineage of metabolic pathways by analyzing mutations as well as interactions in both healthy and diseased states. We use weighted graphs for identifying interactions which change due to the occurrence of mutations where weights, $w_j = r_j/(o_j + n_j)$. We calculate the interaction weights before and after the mutational event and analyze whether there are any changes in the weights. These weights highlight the compactness of the structure and thus, lower the weight more compact the structure and thus the metabolites are more robust. The reason being, compact structures are conserved during evolution and remain as it is through time span [31].

For instance, for *butanoate* metabolism (Fig. 12 in Supplementary Information), interaction weights, $w_j(y_j)$, ranged from 4.6 to 11.5. For single mutation the typical metabolic network is given in Fig. 8b. In case when the modifier $p = y_{17}$, the corresponding interaction weights are recalculated to be from 2.05 to 2.8 respectively. In case of multiple random mutations, $pt = y_3, y_5, y_{12}, y_{13}, y_{17}$, modified interaction weights are from 0 to 0.63 (Fig. 12c in Supplementary Information). The identified interactions and their weights are validated by constructing *payoff* matrices. For instance, in case of two metabolites m_1 and m_2, if m_1 is interacting with m_2 using an interaction payoff (reaction link value) ip_{r_1}, m_2 is interacting with m_1 using an interaction payoff ip_{r_2}, m_1 is self-interacting with an interaction payoff ip_{r_3} and m_2 is self-interacting with an interaction payoff ip_{r_4}, the payoff matrix is [19]

$$\begin{bmatrix} ip_{r_3} & ip_{r_1} \\ ip_{r_2} & ip_{r_4} \end{bmatrix}.$$

Likewise, for *butanoate* metabolism the payoff matrices are represented as

butanoate cycle1

	m_{14}	m_{15}	m_{16}	z_1	z_2	z_3
m_{14}	X	0.03	0.06	0.0	X	X
m_{15}	0.03	X	0.03	X	X	X
m_{16}	0.06	0.03	X	X	X	X
z_1	0.0	X	X	X	X	X
z_2	X	X	X	X	X	X
z_3	X	X	X	X	X	X

(for cycle 1) and

butanoate cycle 2

	m_{14}	m_{15}	m_{16}	m_{17}	z_1	z_2	z_3	z_4
m_{14}	X	0.03	X	0.06	0.03	0.03	0.03	X
m_{15}	0.03	X	0.03	X	X	X	X	X
m_{16}	X	0.03	X	0.03	X	X	0.03	X
m_{17}	0.03	X	0.03	X	X	X	X	X
z_1	0.0	X	X	X	X	X	X	X
z_2	0.03	X	X	X	X	X	X	X
z_3	X	X	0.03	X	X	X	X	X
z_4	X	X	X	X	X	X	X	X

(for cycle 2). Here, *X* stands for no interaction.

Next, we identify the most robust and dense metabolites, in terms of connectivity patterns within the metabolic pathway. For this reason we target the non-pool metabolites only. We checked the occurrence nature of these non-pool metabolites before and after the mutational event. The occurrence for all the metabolites involved in butanoate metabolism ranged from 1 to 5. After single mutation, i.e., $pt = y_{17}$, no metabolite is removed as there were parallel reactions, and for multiple mutations, i.e., $pt = y_3, y_5, y_{12}, y_{13}, y_{17}$, the metabolites removed are m_7, m_{14} for which occurrences ranged from 0 to 5 respectively. Next, we check the compactness of metabolites using clustering coefficient,

$$cc(m_j) = \frac{\text{conn}(m_j)}{\frac{\text{degree}(\text{degree}-1)}{2}},$$

where conn(m_j) is the number of connectivities between neighbors of metabolite m_j and *degree* is the degree of metabolite m_j. Here, clustering coefficients ranged from 0 to 0.33, suggesting that metabolites m_{14}, m_{15}, m_{16}, m_{17} are more compact than the others among which m_{19} acts as central metabolite having the highest connectivity value. After a single mutation none of the metabolites are removed, whereas for multiple mutations, when m_{13}, m_{15}, m_{16} are removed, clustering coefficients ranged from 0 to 1.0 respectively. Thus, we

found that with a deletion mutation of *(S)-3-hydroxybutanoyl CoA* (m_7) and *(R)-3-hydroxybutanoate* (m_{14}), the overall structure of the metabolic network broke suggesting that the compact structure was more dense and robust than the other metabolites.

The last step is to analyze the growth pattern of metabolic pathway. For this purpose, we considered two parameters, namely, $cc(m_j)$ of metabolites and weight of interactions. We suggest that metabolites which are likely to be the most robust and dense should have high $cc(m_j)$ and their interaction weights with other neighboring metabolites should be less. Thus, the growth of metabolic pathways initiates from the densest metabolite and proceed to the least dense one. With decrease in connectivity, $cc(m_j)$ decrease whereas their interaction weight with neighboring metabolites increases suggesting less participation in the metabolic pathway [32,40].

For instance, considering the healthy metabolic network in Fig. 13a (in Supplementary Information), the densest metabolite in terms of connectivity is *acetoacetyl CoA* (compactness value, C.V.=0.1). Thus, the next metabolites produced from it are *(S)-3-hydroxybutanoyl CoA* (C.V.=0), *acetyl CoA* (C.V.=0.16), *(S)-3-hydroxy-3-methylglutaryl-CoA* (C.V.=0.33), and *acetoacetate* (C.V.=0.16). Similarly, the next groups of metabolites arisen from these are m_{14} (C.V.=0), m_5 (C.V.=0). The growth pattern of other metabolites like m_1 (C.V.=0), m_9 (C.V.=0), m_8 (C.V.=0), m_{20} (C.V.=0), m_3 (C.V.=0), m_6 (C.V.=0), m_4 (C.V.=0), m_{10} (C.V.=0), m_{11} (C.V.=0), m_{12} (C.V.=0) and m_{17} (C.V.=0) are not dependent on the previously discussed metabolites and they grew simultaneously along with the others, due to low compactness and high interaction weights (Fig. 13b–f in Supplementary Information). Another interesting feature that we observed from this study was that with an increase in mutational events the compactness of the densest metabolite and its neighbors increases. But this phenomenon occurs only when mutations occur in its outermost neighbors. If the same events occur in the nearest neighbors of the earliest metabolites, the complete metabolic network collapses. The evolutionary theory proposed by us in this work is an extension to the classical 'micro-evolution' theory, which states that certain divergent features are significant among growing entities [10]. Moreover, we saw that the nature of this 'micro-evolution' is based on the presence of divergent interactions as well as the growth patterns, which is observed to initiate from a root or the densest metabolite to the least dense metabolite in terms of occurrence [10].

We validate the growth results using 'Hurwicz's index' [41] (Pseudocode 6).

Pseudocode 6. Hurwicz's index

for ∀payoff matrices **do**
  select a value ν in between 0 to 1
  choose largest and smallest ν for each action
  multiply largest payoff by ν and smallest by $\nu-1$
  select action with largest sum
  endfor

3. Results and analysis

In this section we discuss the results of our methodology on the four metabolic pathways, namely, *glutamate*, *β-alanine*, *taurine* and *hypotaurine* and *butanoate* metabolisms, for elucidating the changes adhered by them during evolution, for identifying their possible role in Type 1 *Diabetes Mellitus* (T1D).

3.1. Analyzing the evolutionary significance of pathways belonging to Homo sapiens and other eukaryotes

We performed a comparison among metabolic pathways belonging to *Homo sapiens* with non-obese diabetic *Mus musculus*, biobreeding Diabetes-prone *Rattus norvegicus*, *Pan troglodytes*, *Oryctolagus cuniculus*, *Danio rerio* and *Drosophila melanogaster*. For selecting the species that are closely related to *H. sapiens*, we performed a phylogenetic analysis by comparing the mitochondrial DNA sequences for all. Fig. 9 illustrates our results in this regard. The comparison was on the basis of similarity among pathway structure, patterns, interaction and growth nature. The four metabolic pathways were integrated from the 36 databases and validated using literature resources as well as a method called SAGPAR [33]. Structure-wise the pathways displayed similarities among results obtained. For instance, in case of *glutamate* metabolic pathway 12 links were predicted by KEGG, *β-alanine* metabolic pathway 16 links were predicted by KEGG, *taurine* and *hypotaurine* metabolic pathway 6 links were predicted by Brenda whereas *butanoate* metabolic pathway 15 links were predicted by KEGG (in *M. musculus*, *R. norvegicus*, *P. troglodytes* and *O. cuniculus*).

The essential pattern in *glutamate* metabolism involves L-*glutamine, 2-oxoglutarate, 4-aminobutanoate,* L-*1-pyroline 5-carboxylate* and L-*glutamate* with Hamming distance 3 and the eigenvalues for each *JM* have the largest real part as $1.49 * 10^{-17}$ where all negative real part (in *M. musculus*, *R. norvegicus*, *P. troglodytes* and *D. rerio*). Similarly, in case of *β-alanine* metabolism, the pattern found was L-*asparagine*, *oxaloacetate*, L-*arginino succinate*, L-*alanine* and L-*asparate* with Hamming distance 3 and the eigenvalues for each *JM* have the largest real part as 0 where all negative real part (in *M. musculus* and *D. melanogaster*). In case of *taurine* and *hypotaurine* metabolism the pattern was L-*cysteate*, *taurocholate*, *hypotaurine*, 5-*glutamyl-taurine* and *taurine* having Hamming distance 3 and the eigenvalues for each *JM* have the largest real part as 0 where all negative real part (in *R. norvegicus), P. troglodytes, and O. cuniculus*. Lastly, in case of *butanoate* metabolism the pattern found was *butanoate*, L-*glutamate*, 2-*acetolactate* and *butanoyl coA* with Hamming distances 3 and the eigenvalues for each *JM* have the largest real part as 0 where all negative real part (in both *M. musculus*, *R. norvegicus* and *O. cuniculus*). Moreover, the interactions for all the four metabolic pathways belong to divergent category.

Likewise, in case of *glutamate* metabolism the payoff matrix has stable interactions existing within the modules ranging from 0.01 to 0.05 with L-*glutamate* having clustering coefficient 0.06. Similarly, in *β-alanine* metabolism the payoff matrix has stable interactions existing within the modules ranging from 0.01 to 0.06 with *β-alanine* having clustering coefficient 0.05. Furthermore, in case of *taurine* and *hypotaurine* metabolism the payoff matrix has stable interactions existing within the modules ranging from 0.01 to 0.08 with *taurine* having clustering coefficient 0. Lastly, in case of *butanoate* metabolism the payoff matrix has stable interactions existing within the modules ranging from 0.01 to 0.03 with *butanoate* having clustering coefficient 0.03. Table 6 provides a comparative study of our results.

3.2. Pattern identification for metabolic pathways

We implement the pattern identification strategy to identify suitable patterns that remain conserved, are of primary importance

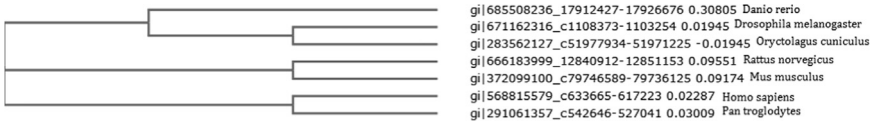

Fig. 9. Evolutionary tree for the species considering mitochondral DNA sequences.

Table 6
Pathway comparison studies among multiple species.

	Links	HD	Eigenvaluesfor JM	Stable interaction range	Metabolite in stable interaction	Average CC Average CC
Glutamate metabolism						
H. sapiens	12	3	1.47×10^{-17}	0.01–0.05	L-*glutamate*	0.06
M. musculus	12	3	1.47×10^{-17}	0.01–0.05	L-*glutamate*	0.06
R. norvegicus	12	3	1.47×10^{-17}	0.01–0.05	L-*glutamate*	0.06
P. troglodytes	12	3	1.47×10^{-17}	0.01–0.05	L-*glutamate*	0.06
O. cuniculus	12	3	1.47×10^{-17}	0.01–0.05	L-*glutamate*	0.06
D. rerio	12	3	1.47×10^{-17}	0.01–0.05	L-*glutamate*	0.06
D. melanogaster	12	3	1.47×10^{-17}	0.01-0.05	L-*glutamate*	0.06
β-alanine metabolism						
H. sapiens	16	3	0	0.01–0.06	*β-alanine*	0.05
M. musculus	14	3	0	0.01–0.06	*β-alanine*	0.05
R. norvegicus	14	3	0	0.01–0.06	*β-alanine*	0.05
P. troglodytes	13	3	0	0.01–0.06	*β-alanine*	0.05
O. cuniculus	15	3	0	0.01–0.06	*β-alanine*	0.05
D. rerio	15	3	0	0.01–0.06	*β-alanine*	0.05
D. melanogaster	14	3	0	0.01–0.06	*β-alanine*	0.05
Taurine and hypotaurine metabolism						
H. sapiens	6	3	0	0.01–0.08	*Taurine*	0
M. musculus	4	3	0	0.01–0.08	*Taurine*	0
R. norvegicus	4	3	0	0.01–0.08	*Taurine*	0
P. troglodytes	3	3	0	0.01–0.08	*Taurine*	0
O. cuniculus	5	3	0	0.01–0.08	*Taurine*	0
D. rerio	5	3	0	0.01–0.08	*Taurine*	0
D. melanogaster	4	3	0	0.01–0.08	*Taurine*	0
*Butanoate*metabolism						
H. sapiens	15	3	0	0.01–0.03	*Butanoate*	0.03
M. musculus	15	3	0	0.01–0.03	*Butanoate*	0.03
R. norvegicus	15	3	0	0.01–0.03	*Butanoate*	0.03
P. troglodytes	15	3	0	0.01–0.03	*Butanoate*	0.03
O. cuniculus	15	3	0	0.01–0.03	*Butanoate*	0.03
D. rerio	15	3	0	0.01–0.03	*Butanoate*	0.03
D. melanogaster	15	3	0	0.01–0.03	*Butanoate*	0.03

during evolution for the above given metabolic pathways and establish their role in Type 1 *Diabetes Mellitus*. As already discussed in methodology section, we use an approach based on detecting community structures, providing an insight into how network function and topology affected each other. We identify Hamming distances as a basis to group community structures.

From Table 1 we observe that in case of *glutamate* metabolism, L-*glutamate* is involved in four reactions leading the formation of a pattern having Hamming distance 3. This pattern consists of metabolites L-*glutamine*, 2-oxoglutarate, 4-aminobutanoate, L-1-*pyroline* 5-*carboxylate* forming links with L-*glutamate*. Similarly, in case of *β-alanine* metabolism, L-*aspartate* and *oxaloacetate* are involved in four and three reactions and had Hamming distances 3 and 4 respectively. The patterns consist of metabolites L-*asparagine*, *oxaloacetate*, L-*arginino succinate*, L-*alanine* (forming links with L-*asparate*) and *pyruvate*, *N-acetyl*-L-*aspartate*, *D-aspartate* (forming links with *oxaloacetate*). Also, in *taurine* and *hypotaurine* metabolism *taurine* was involved in four reactions having a pattern with *taurine* having Hamming distance 3. The metabolites involved in this pattern are L-*cysteate*, *taurocholate*, *hypotaurine*, 5-glutamyl-taurine. Furthermore, in case of *butanoate* metabolism for Hamming distances 3 and 4, three and four patterns are found having *butanoate*, L-*glutamate*, 2-acetolactate and *butanoyl coA* as maximally found metabolites respectively (already discussed in 'Methodology' section).

Hamming distance helps in signifying a biological meaning to the patterns. Moreover, the community structures found on the basis of Hamming distances consist of metabolites separated by the same number of substitutions. We observe that the patterns found above are separated by least number of substitutions and remain conserved through evolution. Thus, we could say that these patterns are significant as they are least mutated through evolution and needs to be further studied for analyzing any possible interactions, to predict their role in evolution [24].

3.3. Stability studies of patterns using cycles and quantitative methods

We further perform functional validation of the patterns by representing them using cycles and ODEs. *Glutamate* metabolism has a single cycle, L-*glutamate*–L-*glutamine*–2-oxoglutaramate–2-oxoglutarate represented by $m_6 - m_{11} - m_9 - m_3$ with five artificial entities z_1, z_2, z_3, z_4, z_5 respectively (Fig. 14 in Supplementary Information). The ODEs for the same are represented as

$$m_6 = r_1 z_1 - m_6(r_3 + r_2) + z_2 r_9 \rightarrow \frac{d(m_6)}{dt} = r_1 \frac{d(z_1)}{dt} - (r_3 + r_2)\frac{d(m_6)}{dt} + (r_9)\frac{d(z_2)}{dt} \tag{16}$$

$$m_{11} = r_2 m_6 + r_8 z_3 + r_7 z_4 + r_6 z_5 - r_5 m_9 \rightarrow \frac{d(m_{11})}{dt} = r_2 \frac{d(m_6)}{dt} + r_8 \frac{d(z_3)}{dt} + r_7 \frac{d(z_4)}{dt} + r_6 \frac{d(z_5)}{dt} - r_5 \frac{d(m_9)}{dt} \tag{17}$$

$$m_3 = r_3 m_6 - m_3 r_4 \rightarrow \frac{d(m_3)}{dt} = r_3 \frac{d(m_6)}{dt} - r_3 \frac{d(m_3)}{dt} \tag{18}$$

$$m_9 = r_4 m_3 + m_{11} r_5 \rightarrow \frac{d(m_9)}{dt} = r_4 \frac{d(m_3)}{dt} + r_5 \frac{d(m_{11})}{dt} \tag{19}$$

$$z_1 = -r_1 m_6 \rightarrow \frac{d(z_1)}{dt} = -r_1 \frac{d(m_6)}{dt} \tag{20}$$

$$z_2 = -r_9 m_6 \rightarrow \frac{d(z_2)}{dt} = -r_9 \frac{d(m_6)}{dt} \tag{21}$$

$$z_3 = -r_8 m_{11} \rightarrow \frac{d(z_3)}{dt} = -r_8 \frac{d(m_{11})}{dt} \tag{22}$$

$$z_4 = -r_7 m_{11} \rightarrow \frac{d(z_4)}{dt} = -r_7 \frac{d(m_{11})}{dt} \tag{23}$$

$$z_5 = -r_6 m_{11} \rightarrow \frac{d(z_5)}{dt} = -r_6 \frac{d(m_{11})}{dt} \tag{24}$$

respectively. Meanwhile, the *SM* and *JM* for the cycles are illustrated in Tables 9 and 10 (in Supplementary Information). By calculating the eigenvalues for each *JM* (Table 5), we observe that the largest real part is $1.50481 * 10^{-17}$, eleven are purely real, nine are equal to zero and *eleven* have negative real part, suggesting that the cycles so formed are stable [34].

Similarly, for *β-alanine* metabolism too one cycle is identified, namely, *β-alanine–β-aminopropion aldehyde–carnosine–histidine* represented by $m_3 - m_4 - m_7 - m_6$ with six artificial entities $z_6, z_7, z_8, z_9, z_{10}, z_{11}$ respectively (Fig. 15 in Supplementary Information). The ODEs for the same are represented as

$$m_3 = r_1 z_6 + z_7 r_2 + z_8 r_3 + z_9 r_4 - m_3(r_9 + r_7 + r_5) \rightarrow \frac{d(m_1)}{dt} = r_1 \frac{d(z_6)}{dt} + r_2 \frac{d(z_7)}{dt} + r_3 \frac{d(z_8)}{dt} + r_4 \frac{d(z_9)}{dt} - (r_9 + r_7 + r_5) \frac{d(m_3)}{dt} \tag{25}$$

$$m_4 = r_5 m_3 - r_6 m_4 \rightarrow \frac{d(m_4)}{dt} = r_5 \frac{d(m_4)}{dt} - r_6 \frac{d(m_4)}{dt} \tag{26}$$

$$m_7 = r_9 m_3 - m_7(r_8 + r_{10}) \rightarrow \frac{d(m_7)}{dt} = r_9 \frac{d(m_3)}{dt} - (r_8 + r_{10}) \frac{d(m_7)}{dt} \tag{27}$$

$$m_6 = -r_7 m_3 + m_7 r_8 \rightarrow \frac{d(m_6)}{dt} = -r_7 \frac{d(m_3)}{dt} + r_8 \frac{d(m_7)}{dt} \tag{28}$$

$$z_6 = -r_1 m_3 \rightarrow \frac{d(z_6)}{dt} = -r_1 \frac{d(m_3)}{dt} \tag{29}$$

$$z_7 = -r_2 m_3 \rightarrow \frac{d(z_7)}{dt} = -r_2 \frac{d(m_3)}{dt} \tag{30}$$

$$z_8 = -r_3 m_3 \rightarrow \frac{d(z_8)}{dt} = -r_3 \frac{d(m_3)}{dt} \tag{31}$$

$$z_9 = -r_4 m_3 \rightarrow \frac{d(z_9)}{dt} = -r_4 \frac{d(m_3)}{dt} \tag{32}$$

$$z_{10} = -r_{10} m_7 \rightarrow \frac{d(z_{10})}{dt} = -r_{10} \frac{d(m_7)}{dt} \tag{33}$$

$$z_{11} = -r_6 m_4 \rightarrow \frac{d(z_{11})}{dt} = -r_6 \frac{d(m_4)}{dt} \tag{34}$$

Moreover, the *SM* and *JM* for the cycles are illustrated in Tables 11 and 12 (in Supplementary Information). By calculating the eigenvalues for each *JM* (Table 5), we observe that the largest real part is 0, thirteen are purely real, one is equal to zero and *thirteen* have negative real part, again suggesting that the cycles so formed are stable [34].

Likewise, for *taurine* and *hypotaurine* metabolism too one cycle is identified, namely, L-*cysteine*–L-*cysteate*–3-sulfino L-alanine–*hypotaurine*–*taurine* represented by $m_6 - m_5 - m_8 - m_2 - m_4$ with three artificial entities z_{12}, z_{13}, z_{14} respectively (Fig. 16 in Supplementary Information). The ODEs for the same are represented as

$$m_6 = -(r_1 + r_2) m_6 \rightarrow \frac{d(m_6)}{dt} = -(r_1 + r_2) \frac{d(z_9)}{dt} \tag{35}$$

$$m_5 = r_1 m_6 - r_3 m_5 \rightarrow \frac{d(m_5)}{dt} = r_1 \frac{d(m_6)}{dt} - r_3 \frac{d(m_5)}{dt} \tag{36}$$

$$m_8 = r_2 m_6 - m_8 r_4 \rightarrow \frac{d(m_8)}{dt} = r_2 \frac{d(m_6)}{dt} - r_4 \frac{d(m_8)}{dt} \tag{37}$$

$$m_2 = r_4 m_8 + z_{14} r_6 - m_2 r_5 \rightarrow \frac{d(m_2)}{dt} = r_4 \frac{d(m_8)}{dt} + r_6 \frac{d(z_{14})}{dt} - r_5 \frac{d(m_2)}{dt} \tag{38}$$

$$m_4 = r_5 m_2 + r_3 m_5 + r_7 z_{13} - r_8 m_4 \rightarrow \frac{d(m_4)}{dt} = r_5 \frac{d(m_2)}{dt} + r_3 \frac{d(m_5)}{dt} + r_7 \frac{d(z_{13})}{dt} - r_8 \frac{d(m_4)}{dt} \tag{39}$$

$$z_{12} = r_8 m_4 \rightarrow \frac{d(z_{12})}{dt} = r_8 \frac{d(m_4)}{dt} \tag{40}$$

$$z_{13} = -r_7 m_4 \rightarrow \frac{d(z_{13})}{dt} = -r_7 \frac{d(m_4)}{dt} \tag{41}$$

$$z_{14} = -r_6 m_2 \rightarrow \frac{d(z_{14})}{dt} = -r_6 \frac{d(m_2)}{dt} \tag{42}$$

Lastly, the *SM* and *JM* for the cycles are illustrated in Tables 13 and 14 (in Supplementary Information). By calculating the eigenvalues for each *JM* (Table 5), we observe that the largest real part is 0, three are purely real, one is equal to zero and *five* have negative real part, again suggesting that the cycles so formed are stable.

3.4. Validation of stability results

We perform the validation of stability results using 'Gerschgorin circle theorem' based on eigenvalues [35]. Let X be a complex square matrix $(n \times n)$, having x_{ij} values. For $i \in \{1, \ldots, n\}$, $V_i = \sum_{j \neq i} |v_{ij}|$ is the sum of absolute values of non-diagonal entries in the i^{th} row and $G(a_{ii}, V_i)$ is a closed disc centered at a_{ii} having radius V_i, known as 'gerschgorin disc'. Assuming that e is an eigenvalue of X and $v = (v_j)$ is the corresponding eigenvector, $|v_i| = \max_j |v_j|$. Then $|v_i| > 0$, else $v = 0$. As v is an eigenvector, $Xv = ev$ and $\sum_j x_{ij} v_j = e v_i \quad \forall i \in \{1, \ldots, n\}$. Thus, $\sum_{j \neq i} x_{ij} v_j = e v_i - x_{ii} v_i$. Furthermore, dividing both sides by v_i and taking the absolute value, $|e - x_{ii}| = \left|\sum_{j \neq i} x_{ij} v_j / v_i\right| \leq \sum_{j \neq i} \left|x_{ij} v_j / v_i\right| \leq \sum_{j \neq i} |x_{ij}| = V_i$ as $|v_j / v_i| \leq 1$ for $j \neq i$. Thus, each eigenvalue of X lies within at least one of the $G(a_{ii}, V_i)$.

For *butanoate* metabolism, we have the Jacobian $-0.3, 0.0, -0.3, -0.1$ corresponding to metabolites *acetoacetate*, *acetyl−CoA*, (*R* -3-*hydroxyl butanoate* and (*S*) -3-*hydroxy* 3-*methyl glutaryl CoA* respectively. Likewise, $G_1(-0.3, 0.2(= 0.1 + 0.1 + 0.0))$, $G_2(0.0, 0.1 (= 0.0 + 0.1 + 0.0))$, $G_3(-0.3, 0.3(= 0.2 + 0.0 + 0.1))$ and $G_4(-0.1, 0.3(= 0.0 + 0.2 + 0.1))$ respectively. Thus, we observe that the eigenvalues for *acetoacetate*, *acetyl−CoA*, (*R*)-3-*hydroxyl butanoate* and (*S*)-3-*hydroxy* 3-*methyl glutaryl CoA* are with the 'gerschgorin discs' G_1, G_2, G_3, G_4 respectively. Similarly, for *glutamate* metabolism, the Jacobian $0.0, -0.7, -0.3, -0.1$ corresponds to 2-*oxoglutarate*, 2-*oxoglutaramate*, L-*glutamate* and L-*glutamine* respectively. Thus, $G_1(0.0, 1.3(= 0.2 + 0.3 + 0.8))$, $G_2(-0.7, 0.1(= 0.0 + 0.0 + 0.1))$, $G_3 (-0.3, 0.3(= 0.0 + 0.0 + 0.3))$ and $G_4(-0.1, 1.2(= 0.8 + 0.1 + 0.3))$ corresponding to 2-*oxoglutarate*, 2-*oxoglutaramate*, L-*glutamate* and L-*glutamine* respectively. Likewise, for *β*-alanine metabolism the Jacobian $-0.4, -0.1, -0.05, -0.01$ corresponds to *β-alanine*, *β-amino propion aldehyde*, *carnosine* and *histidine* respectively. Thus, $G_1(-0.4, 0.1(= 0.0 + 0.1 + 0.0))$, $G_2(-0.1, 0.1(= 0.1 + 0.0 + 0.0))$, $G_3(-0.05, 0.13(= 0.1 + 0.0 + 0.03))$ and $G_4(-0.01, 0.02(= 0.0 + 0.0 + 0.02))$ corresponding to *β*-alanine, *β*-amino propion aldehyde, *carnosine* and *histidine* respectively. Lastly, for *taurine and hypotaurine* metabolism the Jacobian $-0.1, -0.1, -0.1, 0.0, -0.02$ for L-*cysteine*, L-*cysteate*, 3-sulfino L-alanine, *hypotaurine* and *taurine* respectively. The $G_1((-0.1, 0.0(= 0.0 + 0.0 + 0.0 + 0.0))$, $G_2((-0.1, 0.0(= 0.0 + 0.0 + 0.0 + 0.0))$, $G_3(-0.1, 0.1(= 0.1 + 0.0 + 0.0 + 0.0))$, $G_4(0.0, 0.1 (= 0.0 + 0.0 + 0.1 + 0.0))$ and $G_5(-0.02, 0.1(= 0.0 + 0.1 + 0.0 + 0.0))$ corresponding to L-*cysteine*, L-*cysteate*, 3-sulfino L-alanine, *hypotaurine* and *taurine* respectively. Furthermore, identifying the largest real part of eigenvalues specifies the slowest timescale of the metabolic system

which is stable if all the eigenvalues have a negative real part. For all these four metabolic pathways, the metabolite groups correspond to the cycles predicted previously which are stable as they correspond to eigenvalues with negative real part [34].

3.5. Interaction analysis of metabolic pathways

Interaction analysis on the four metabolic pathways is performed for analyzing if any interaction is existing among the patterns already detected using pattern identification strategy. We found interactions among the already generated patterns as detected in the previous section as we are specifically interested in finding the type of interactions among patterns, which could further be used to study evolution.

From Table 15 (in Supplementary Information), it is evident that the cut set L-*glutamate* … L-*glutamine* and L-*glutamate* … *Carbamoyl-P* in *glutamate* metabolism, one of the patterns already found before, had an avgd (1.7 (A), 1.2 (B)), hd (4.0 (A), 2.0 (B)) and φ (24.0 (A), 6.0 (B)), making the interaction belonging to divergent category. Similarly, in β-*alanine* metabolism L-*aspartate* … *oxaloacetate* and L-*aspartate* … L-*alanine*, two of the patterns found before, having avgd (2.0 (A), 1.8 (B)), hd (3.0 (A), 4.0 (B)) and φ (22.0 (A), 53.0 (B)), making the interaction divergent. In *taurine* and *hypotaurine* metabolism, L-*cysteate* … L-*cysteine* and *taurine* … *hypotaurine* pattern present in cut set had an avgd (1.7 (A), 2.0 (B)), hd (3.0 (A), 2.0 (B)) and φ (3.0 (A), 0.0 (B)), making the interaction belonging to divergent category. Furthermore, patterns found in *butanoate* metabolism also have divergent type interactions (already discussed in 'Methodology' section).

From the above discussion we found that the interactions present in patterns behave in divergent manner. However, a 'divergent' interaction might not suggest that the metabolites involved in them behave in divergent manner during evolution. For this reason it is evident that we further perform the analysis on these patterns and their interactions for studying their growth pattern and evolutionary significance [36].

3.6. Implementing Hasse diagrams, cuts and DC for interactions analysis

The Hasse diagram for *glutamate* metabolism has five layers (hierarchically) with individual metabolites residing in the bottom-most layer. The top layer (Layer 4) consists of 9 entities, namely, m_6, m_{11}, m_9, m_3, z_1, z_2, z_3, z_4, z_5, layer 3 has two modules with 4 and 5 entities, namely, m_6, m_{11}, m_9, m_3 (module 1), z_2, z_3, z_4, z_5 (module 2), layer 2 has three modules with 3, 3, 5 entities, namely, m_6, m_{11}, m_9, (module 1), m_3, m_{11}, m_9 (module 2), z_2, z_3, z_4, z_5 (module 3), layer 1 has seven modules, namely, m_6, m_{11} (module 1), m_6, m_9 (module 2), m_{11}, m_3 (module 3), m_9, m_3 (module 4), z_1, z_2 (module 5), z_3, z_4 (module 6), z_4, z_5 (module 7), layer 0 has nine modules, namely, m_6 (module 1), m_{11} (module 2), m_9 (module 3), m_3 (module 4), z_1 (module 5), z_2 (module 6), z_3 (module 7), z_4 (module 8), z_5(*module*9) (Fig. 17 in Supplementary Information).

Similarly, Hasse diagram for β-*alanine* metabolism too has five layers layer 4 consisting of 10 entities, namely, m_3, m_4, m_7, m_6, z_6, z_7, z_8, z_9, z_{10}, z_{11}, layer 3 has three modules with 4, 4 and 2 entities, namely, m_3, m_4, m_7, m_6 (module 1), z_6, z_7, z_8, z_9 (module 2), z_{10}, z_{11} (module 3), layer 2 has seven modules with 3 entities each except one which has two entities, namely, m_3, m_4, m_7 (module 1), m_3, m_7, m_6 (module 2), m_3, m_4, m_6 (module 3), z_7, z_8, z_9 (module 4), z_7, z_8, z_6 (module 5), z_7, z_6, z_9 (module 6), z_{10}, z_{11} (module 7), layer 1 has eleven modules, namely, m_3, m_4 (module 1), m_3, m_7 (module 2), m_7, m_6 (module 3), m_3, m_6 (module 4), z_7, z_8 (module 5), z_7, z_9 (module 6), z_8, z_9 (module 7), z_6, z_7 (module 8), z_8, z_6 (module 9), z_6, z_9 (module 10), z_{10}, z_{11} (module 11), layer 0 has ten modules, namely, m_3 (module 1), m_4 (module 2), m_7 (module 3), m_6 (module 4), z_6 (module 5), z_7 (module 6), z_8 (module 7), z_9 (module 8), z_{10} (module 9), z_{11} (module 10) respectively (Fig. 18 in Supplementary Information) [36].

Lastly, Hasse diagram for *taurine and hypotaurine* metabolism has six layers with layer 5 consisting of 8 entities, namely, m_7, m_8, m_2, m_4, m_5, z_{12}, z_{13}, z_{14}, layer 4 has two modules with 5, 3 entities, namely, m_7, m_8, m_2, m_4, m_5 (module 1), z_{12}, z_{13}, z_{14} (module 2), layer 3 has three modules with 4, 4, 3 entities, namely, m_7, m_8, m_2, m_4 (module 1), m_8, m_2, m_4, m_5 (module 2), z_{12}, z_{13}, z_{14} (module 3), layer 2 has three modules with three entities each, namely, m_8, m_4, m_5 (module 1), m_8, m_7, m_2 (module 2), z_{12}, z_{13}, z_{14} (module 3), layer 1 has six modules, namely, m_8, m_5 (module 2), m_4, m_5 (module 2), m_8, m_7 (module 3), m_7, m_2 (module 4), m_2, m_4 (module 5), z_{12}, z_{13}, z_{14} (module 6), layer 0 has eight modules, namely, m_5 (module 1), m_8 (module 2), m_4 (module 3), m_7 (module 4), m_2 (module 5), z_{14} (module 6), z_{13} (module 7), z_{12} (module 8) (Fig. 19 in Supplementary Information).

Continuing our analysis with cuts and *DC*, we observe that for *glutamate* metabolism, high *TC* is found for L-*glutamate*, L-*glutamine* = 0.78, (L-*glutamate*, L-1-*pyrroline* 5-*carboxylate*) = 0.64, (L-*glutamate*, 4-*aminobutanoate*) = 0.61. In case of β-*alanine* higher *TC* is found for (L-*aspartate*, L-*asparagine*) = 0.75, (L-*aspartate*, β-*alanine*) = 0.64. Similarly, for *taurine* and *hypotaurine* metabolism *TC* is found for (*taurine*, *hypotaurine*) = 0.82, (*taurine*, L-*cysteate*) = 0.75. Likewise, implementing *DC* for *glutamate* metabolism, we consider L-*glutamate* to be the 'basis node' initially. We assume that all metabolites that are paired with *glutamate* have 'weak' links initially. On the basis of higher values for *TC* a weak link is converted to a strong link, whereas a lower *TC* value result in a weak link. We observe that L-*glutamine* has a strong link with L-*glutamate*, whereas L-1-*pyrroline*-5-*carboxylate* is weakly linked to L-*glutamate*. Similarly, considering L-*glutamine*, 2-*oxoglutarate* and 4-*aminobutanoate* as the 'basis node' creates all weak links with their corresponding pairs (Fig. 20 in Supplementary Information). Likewise, for β-*alanine* metabolism considering L-*aspartate* as 'basis node' creates strong links with L-*asparagine* as well as β-*alanine* and weak links with *oxaloacetate* as well as L-*argininosuccinate* respectively. Similarly, considering L-*asparagine*, *oxaloacetate* and L-*arginino succinate* as 'basis node' creates all weak links with their corresponding pairs (Fig. 21 in Supplementary Information). In case of *taurine* and *hypotaurine* metabolism considering *taurine* as the 'basis node' creates strong links with *hypotaurine* and L-*cysteate* whereas weak links with 5-*glutamyl*−*taurine*. Similarly, considering L-*cysteate*, *taurocholate* and *hypotaurine* as 'basis node' creates all weak links with their corresponding pairs (Fig. 22 in Supplementary Information). Thus, on the basis of link type metabolite groups were formed, namely, $Group1 = \{L-glutamate, L-glutamine, L-1-pyrroline-5-carboxylate\}$, $Group2 = \{\beta-alanine, L-aspartate, oxaloacetate, L-arginino-succinate\}$, $Group3 = \{taurine, hypotaurine, L-cysteate, 5-glutamyl-taurine\}$, and $Group4 = \{L-glutamate, 4-butanoate, acetyl-CoA\}$ [36].

The metabolites having highest degree in each group is identified, namely, L-*glutamate* = 4(*Group*1), L-*aspartate* = 4(*Group*2), *taurine* = 4(*Group*3), *acetylCoA* = 1, L − *glutamate* = 1(*Group*4). We have already defined 'cut' as the minimum number of links and/or metabolites that need to be removed to segment a metabolic pathway into two or more sub-pathway having similar size. Thus, for *Group*1 the cuts are identified between $L-1-pyrroline5-carboxylate \rightarrow L-glutamate$, $L-glutamine \rightarrow L-glutamate$ whose removal causes a strong change in gene functionality and $2-oxoglutarate \rightarrow L-glutamate$, $4-aminobutanoate \rightarrow L-glutamate$ whose removal causes a weak change in gene functionality (Fig. 23a in Supplementary Information). Thus, for *Group*2 the cuts are identified between $\beta-alanine \rightarrow L-aspartate$, $L-asparagine \rightarrow L-aspartate$ whose removal causes a strong change in gene functionality and $oxaloacetate \rightarrow L-aspartate$, $L-arginino-succinate \rightarrow L-aspartate$ whose removal causes a weak change in gene functionality (Fig. 23b in Supplementary Information). Likewise, in *Group*3 the cuts are identified between

taurocholate → taurine, *hypotaurine → taurine* whose removal causes a strong change in gene functionality and *5−glutamyl−taurine → taurine*, *L−cysteate → taurine* whose removal causes a weak change in gene functionality (Fig. 23c in Supplementary Information). Finally, in *Group*4 the cuts are identified between *acetylCoA → acetoacetylCoA*, *L−glutamate → 4−butanoate* whose removal causes a strong change in gene functionality (Fig. 23d in Supplementary Information). We continue with validating the *DC* by identifying the 'strongly connected components' for the metabolic pathways. A strong component in a metabolic pathway consists of functionally significant metabolite groups, which have strong intra-interactions leading to essential gene functionality whereas a weak interaction leads to non-essential gene functionality. In case of *glutamate* metabolism, a single strong component exists, i.e., *D−glutamate*, *D−glutamine*; *L−glutamine*, *2−oxoglutarate*; *L−glutamate*, *L−glutamine* (Fig. 24a in Supplementary Information). In case of *β−alanine* metabolism, 5 strong components exist, namely, *4−aminobutanoate*, *4−aminobutanal*, *acetyl CoA, malonyl CoA*, *spermindine*, *spermine*, *β−alanine*, *L−aspartate*; *β−alanine*, *3−oxopropionate*, and *L-asparagine, L-aspartate; 4-aminobutanyl CoA, propenoyl CoA* (Fig. 24b in Supplementary Information). In case of *taurine and hypotaurine* metabolism, 5 strong components exist, namely, *succinate, succinate semialdehyde; 4-aminobutanoate, L-glutamate*, 2-oxoglutarate, 2-hydroxyglutarate, *taurocholate*, *taurine*, *hypotaurine*, *taurine*, and 2-acetolactate, thiamine diphosphate (Fig. 24c in Supplementary Information). In case of *butanoate* metabolism, a single strong component exist, namely, *acetyl-CoA, acetoacetyl-CoA; butanoate, L-glutamate* (Fig. 23d in Supplementary Information) [36].

3.7. Validation of interaction results

For 'bootstrapping' [30], we assign accuracy measure in terms of confidence score for estimating the interaction correlation among metabolites (Pseudocode 5). We perform various combinations of the metabolites and estimate their correlation tendency to identify the most probable combination, as predicted by the *Hasse*. For each metabolic pathway, we perform 7 distinct combinations of the metabolites, estimate the confidence score and compared with the already generated *Hasse*.

For *butanoate* metabolism we represent the metabolites as $acetoacetylCoA \rightarrow a$, $acetoacetate \rightarrow b$, $(S)-3-hydroxyl-3-methylglutarylCoA \rightarrow c$, $z_1 \rightarrow e, z_2 \rightarrow f$, $z_3 \rightarrow g, z_4 \rightarrow h$ (cycle 1). The combinations are *abcefgh*, *bcaefgh*, *cabefgh*, *abcefgh*, *cabefgh*, *bcaefgh*, and *cbaefgh* with confidence scores 0.61, 0.48, 0.50, 0.17, 0.15, 0.15 respectively. The best possible combinations after bootstrapping are *Acetoacetyl CoA, acetoacetate, (S)-3-hydroxyl-3-methyl glutaryl CoA, Acetoacetate, (S)-3-hydroxyl-3-methyl glutaryl CoA, Acetoacetyl CoA* and $(S)-3-hydroxyl-3-methylglutarylCoA$, *AcetoacetylCoA, acetoacetate* having 61%, 48% and 50% score. Likewise for cycle 2, the metabolites are represented as $AcetoacetylCoA \rightarrow a$, $acetoacetate \rightarrow b$, $(S)-3-hydroxyl-3-methylglutarylCoA \rightarrow c$, $acetylCoA \rightarrow d$, $z_1 \rightarrow e, z_2 \rightarrow f$, $z_3 \rightarrow g, z_4 \rightarrow h$. The combinations are *abcdefgh*, *bcadefgh*, *dcabefgh*, *dabcefgh*, *cabdefgh*, *bdcaefgh*, and *dcbaefgh* with confidence scores 0.44, 0.41, 0.38, 0.30, 0.04, 0.04 respectively.

The best possible combinations after bootstrapping are *Acetoacetyl CoA, acetoacetate, (S)-3-hydroxyl-3-methyl glutaryl CoA, acetyl CoA, Acetoacetate, acetyl CoA, (S)-3-hydroxyl-3-methyl glutaryl CoA, Acetoacetyl CoA* and *(S)-3-hydroxyl-3-methyl glutaryl CoA, Acetoacetyl CoA, acetyl CoA, acetoacetate* having 44%, 41% and 38% score, similar to the predicted *Hasse*. For *glutamate* metabolism we represent the metabolites as *L-glutamate* $\rightarrow a$, *L-glutamine* $\rightarrow b$, $2-oxoglutarate \rightarrow c$, $2-oxoglutaramate \rightarrow d$, $z_1 \rightarrow e, z_2 \rightarrow f$, $z_3 \rightarrow g$, $z_4 \rightarrow h, z_5 \rightarrow i$. The combinations are *abcdefghi*, *bacdefghi*, *bcadefghi*, *cabdefghi*, *dabcefghi*, *cdabefghi*, and *cdbaefghi* with confidence scores 0.67, 0.33, 0.34, 0.28, 0.06, 0.03 respectively. The best possible combinations after bootstrapping are *L-glutamate, L-glutamine, 2-oxoglutarate, 2-oxoglutaramate, 2-oxoglutarate, L-glutamine, L-glutamate, 2-oxoglutaramate* and *L-glutamine, 2-oxoglutarate, L-glutamate, 2-oxoglutaramate* having 67%, 33% and 34% score, similar to the predicted *Hasse* [37].

Likewise, for *β-alanine* metabolism we represent the metabolites as *β-alanine* $\rightarrow a$, *β-aminopropionaldehyde* $\rightarrow b$, *carnosine* $\rightarrow c$, *histidine* $\rightarrow d$, $z_6 \rightarrow e, z_7 \rightarrow f$, $z_8 \rightarrow g, z_9 \rightarrow h$, $z_{10} \rightarrow i, z_{11} \rightarrow j$. The combinations are *acbdefghij*, *bcadefghij*, *dcabefghij*, *cbadefghij*, *cdbaefghij*, *dabcefghij*, and *dcbaefghij* with confidence scores 0.37, 0.34, 0.14, 0.23, 0.13, 0.09 respectively. The best possible combinations after bootstrapping are *β-alanine, carnosine, β-amino propion aldehyde, histidine and carnosine, β-amino propion aldehyde, β-alanine, histidine* having 37% and 34% score, similar to the predicted *Hasse*. Similarly, for *taurine* and *hypotaurine* metabolism we represent the metabolites as *L-cysteine* $\rightarrow a$, *L-cysteate* $\rightarrow b$, $3-sulfinoL-alanine \rightarrow c$, $taurine \rightarrow d$, $hypotaurine \rightarrow e$, $z_{12} \rightarrow f$, $z_{13} \rightarrow g$, $z_{14} \rightarrow h$. The combinations are *acbdefgh, bcaedfgh, bcadefgh, adbcefgh, ecdbafgh, debcafgh*, and *aedbcfgh* with confidence scores 0.61, 0.48, 0.50, 0.17, 0.15, 0.15 respectively. The best possible combinations after bootstrapping are *L-cysteine, 3-sulfino L-alanine, L-cysteate, taurine, hypotaurine, L-cysteate, 3-sulfino L-alanine, L-cysteine, hypotaurine, taurine* and *L-cysteate, 3-sulfino L-alanine, L-cysteine, taurine, hypotaurine* having 61%, 48% and 50% score, similar to the predicted *Hasse* [37].

3.8. Growth detection in metabolic pathways, payoffs and mutational significance

After performing interaction analysis on the metabolic pathways, next step is analyzing the evolutionary growth history of metabolic pathways with special emphasis on patterns and their interactions. In case of *glutamate* metabolism (Fig. 25a and Table 16 in Supplementary Information), the interaction weights ranged from 2.0 to 12.0 for 12 reaction links. For single mutation the typical metabolic network is given in Fig. 25b (in Supplementary Information). In case when the modifier $pt = y_{12}$, the corresponding interaction weights range from 0.0 to 2.0. In case of multiple random mutations, $pt = y_{12}, y_{10}, y_8, y_6, y_2$, modified interaction weights ranges from 0.0 to 0.75 (Fig. 25c and Table 17 in Supplementary Information).

Now, we checked the occurrence nature of non-pool metabolites before and after the mutational event. The occurrence, degree (m_j) for all the metabolites involved in *glutamate* metabolism are found to be from 1 to 6. After single mutation, i.e., $pt = y_{12}$, metabolite m_5 was removed, and for multiple mutations, i.e., $p = y_{12}, y_{10}, y_8, y_6, y_2$, the metabolites removed were m_2, m_5 for which the connectivities ranged from 0 to 4 respectively. Next, we proceed for checking the compactness of the metabolites using clustering coefficient which ranged from 0 to 0.06, suggesting that metabolite m_6 was more compact than the others. After a single mutation where y_7 is removed and the corresponding clustering coefficient values from 0 to 0.06, whereas for multiple mutations when $y_{12}, y_{10}, y_8, y_6, y_2$ are removed, clustering coefficient values are 0. This suggested that deletion mutation of reaction links 12, 10, 8, 6, 2 result in removal of 2, 11, 5. It was evident that the compact structure had *L-glutamate* (m_6) with six reaction links. Likewise, the payoff matrix also suggests that the interactions existing within the modules thus found are stable and have a

significant value, thus validating our results [31]

glutamate payoff	m_6	m_{11}	m_7	m_3	z_1	z_2	z_3	z_4	z_5
m_6	X	0.02	0.06	X	0.01	0.01	X	X	X
m_{11}	0.02	X	X	0.08	X	X	0.01	0.01	0.01
m_7	0.06	X	X	0.01	X	X	X	X	X
m_3	X	0.08	X	X	X	X	X	X	X
z_1	0.01	X	X	X	X	X	X	X	X
z_2	0.01	X	X	X	X	X	X	X	X
z_3	X	0.01	X	X	X	X	X	X	X
z_4	X	0.01	X	X	X	X	X	X	X
z_5	X	X	0.01	X	X	X	X	X	X

Similarly, for analyzing the growth pattern of *glutamate* metabolism, as none of the metabolites had any significant C.V., we propose that the growth is based on the interaction weights and connectivity value. For instance, considering the healthy metabolic network in Fig. 25 (in Supplementary Information), the densest metabolite in terms of age is L-*glutamate* having clustering coeffiecient 0.06. The next metabolites which were produced from m_{13} were L-*glutamine*, *carbomoyl-P*, L-*1-pyrroline-5-carboxylate*, 4-aminobutanoate, L-*glutamyl-tRNA*)=0.0), 2-oxoglutarate. Similarly, the next groups of metabolites arising from this was m_8. The growth pattern of the other metabolites does not give rise to meaningful result as clustering coefficients were 0 (Fig. 26a and Tables 16 and 17 in Supplementary Information) [38].

In case of *β-alanine* metabolism (Fig. 13a in Supplementary Information), the interaction weights ranges from 0 to 21.0. For single mutation the typical metabolic network is given in Fig. 13b (in Supplementary Information). In case when the modifier was $pt=y_7$, the corresponding interaction weights are found to be within 0–21.0, with changes in most of the links, especially for 7th, and 19th links. In case of multiple random mutations, $pt=y_7, y_4, y_3, y_{11}, y_{21}$, modified interaction weights 0–21.0, respectively (Fig. 26c and Table 17 in Supplementary Information). Now, we check the occurrence nature of non-pool metabolites before and after the mutational event. The occurrence for all the metabolites involved in *β-alanine* metabolism ranged from 1 to 9. After single mutation, i.e., $pt=y_7$, none of the metabolites is removed, and for multiple mutations, i.e., $pt=y_7, y_4, y_3, y_{11}, y_{21}$ the metabolites removed are m_6, m_7 for which occurrences ranged from 1 to 6 displaying a slight change. Next, we proceeded for checking the compactness of the metabolites using clustering coefficient, which ranged from 0 to 0.05 (for m_3), suggesting that it is more compact than the others (Table 17 in Supplementary Information).

After a single mutation when y_7 was removed, the clustering coefficients range from 0 to 0.05, whereas for multiple mutations when y_7, y_4, y_3, y_{11}, y_{21} are removed, clustering coefficients ranges from 0 to 0.06, respectively. Thus, we find that with a deletion mutation of the link having L-*aspartate* made the overall structure of the metabolic network unstable suggesting that the links having L-*aspartate* as one of the metabolite is more densed than the other metabolites [38]. Likewise, the payoff matrix again suggest that the interactions existing within the modules have a significant value, thus validating our results

β−alanine payoff	m_3	m_4	m_7	m_6	z_6	z_7	z_8	z_8	z_{10}	z_{11}
m_3	X	0.09	0.09	0.11	0.01	0.01	0.01	0.01	X	X
m_4	0.09	X	X	X	X	X	X	X	X	0.01
m_7	0.09	X	X	0.01	X	X	X	X	0.01	X
m_6	0.11	X	0.01	X	X	X	X	X	X	X
z_6	0.01	X	X	X	X	X	X	X	X	X
z_7	0.01	X	X	X	X	X	X	X	X	X
z_8	0.01	X	X	X	X	X	X	X	X	X
z_9	0.01	X	X	X	X	X	X	X	X	X
z_{10}	X	X	0.01	X	X	X	X	X	X	X
z_{11}	X	0.01	X	X	X	X	X	X	X	X

Similarly, for analyzing the growth pattern of *β*-alanine metabolism, as none of the metabolites had any significant C.V., we propose that the growth is based on the interaction weights and connectivity value. For instance, considering the healthy metabolic network in Fig. 26 (in Supplementary Information), the densest metabolite in terms of connectivity was L-*aspartate* having C.V. of 0.05. Thus, the next metabolites which were produced from it were L-*alanine* (C.V.=0), *malonate semialdehyde* (C.V.=0), L-*aspartyl tRNA* (C.V.=0), *N-carbamoyl l-aspartate* (C.V.=0), *D-aspartate* (C.V.=0), *N-acetyl* L-*aspartate* (C.V.=0), L-*asparagine* (C.V.=0), L-*arginino succinate* (C.V.=0), *adenylo succinate* (C.V.=0).

In case of *taurine* and *hypotaurine* metabolism (Fig. 27a in Supplementary Information), the interaction weights ranged from 0 to 9. For single mutation the typical metabolic network is given in Fig. 27b (in Supplementary Information). In case when the modifier $pt=y_1$, the corresponding interaction weights are found to be within 0–3, respectively. In case of multiple random mutations, $pt=y_1, y_8, y_5, y_4, y_6$, modified interaction weights were from 0 to 1, respectively (Fig. 27c and Table 19 in Supplementary Information). Now, we check the occurrence nature of non-pool metabolites before and after the mutational event. The occurrences for all the metabolites involved in *taurine* and *hypotaurine* metabolism range from 1 to 4 respectively. After single mutation, i.e., $pt=y_1$, the metabolite removed was m_1, and for multiple mutations, i.e., $pt=y_1, y_8, y_5, y_4, y_6$, the metabolites removed were m_1 and m_6 for which occurrences ranges from 0 to 1 respectively. Furthermore, the payoff matrix also suggests that the interactions existing within the modules have a positive value

Taurine and hypotaurine payoff	m_7	m_8	m_5	m_4	m_2	z_{12}	z_{13}	z_{14}
m_7	X	0.0	0.0	X	X	X	X	X
m_8	0.0	X	X	X	0.16	X	X	X
m_5	0.0	X	X	0.08	X	X	X	X
m_4	X	X	0.08	X	0.08	X	X	0.01
m_2	X	0.16	X	0.08	X	0.01	0.01	X
z_{12}	X	X	X	X	0.01	X	X	X
z_{13}	X	X	X	X	0.01	X	X	X
z_{14}	X	X	X	0.01	X	X	X	X

Next, the clustering coefficient values are found to be 0 for all metabolites. After a single mutation when y_1 was removed, clustering coefficient values were still found to be 0, whereas for multiple mutations where y_1, y_8, y_5, y_4, y_6 were removed, still no change in clustering coefficients took place (Table 19 in Supplementary Information). This suggests that even after mutational event/deletion, no significant change in compactness is evident. The reason for this was the small size of the metabolic pathway [38]. For growth pattern analysis, considering the healthy metabolic network in Fig. 27 (in Supplementary Information), the densest metabolite in terms of connectivity was *taurine* (C.V.=0). Thus, the next metabolites which were produced from it were *taurocholate* (C.V.=0), 5-glutaryl taurine (C.V.=0), *hypotaurine* (C.V.=0), L-*cysteate* (C.V.=0) (Fig. 28a–c in Supplementary Information). The public repositories that we considered for our pathway construction are listed in Table 19 in Supplementary Information.

3.9. Validation of growth results

The 'Hurwicz's index' [41] defines a coefficient of optimism (Pseudocode 6) by identifying the best possible Payoff value, suitable to analyze the interactions during metabolic pathway growth. It initiates with identifying a random value between 0 and 1 (0 is pessimistic whereas 1 is optimistic). For each of the payoff matrices, the largest and smallest payoff score is selected corresponding to individual metabolites. The largest score is multiplied with the random value whereas the smallest one is multiplied with 1-random value, followed by summation. For each of the payoff matrices, the number of summed-up value is equal to the number of metabolites selected. For payoff validation, the highest

summed-up value is considered and all the payoff scores corresponding to the payoff matrix should be less than or equal to it.

In case of each payoff matrix, v is selected to be 0.9. Initiating with *butanoate* metabolism (cycle 1), for *acetoacetyl CoA* the highest and lowest payoff scores are 0.06 and 0.03 respectively. Thus, the summed-up value is $0.03 \times (1-0.9)+0.06 \times 0.9 = 0.057$. Similarly, for metabolites *acetoacetate* and *(S)-3-hydroxy-3-methyl glutaryl CoA* the summed-up values are 0.054 and 0.057 respectively. Similarly, for cycle 2 the summed-up values for metabolites *acetoacetyl CoA*, *acetyl CoA*, *acetoacetate* and *(S)*-3-*hydroxy*-3-*methyl glutaryl CoA* are 0.06, 0.054, 0.081 and 0.111 respectively. The highest summed value in both the cycles are 0.057 and 0.111 whereas the highest payoff scores for both the cycles are 0.06, thus satisfying the criteria. Likewise, in case of *glutamate* metabolism summed-up values for metabolites L-*glutamate*, L-*glutamine*, 2-*oxoglutarate*, 2-*oxoglutaramate* are 0.056, 0.074, 0.054 and 0.072 respectively. The highest summed value is 0.074 whereas the highest payoff score is 0.06. Similarly, in case of *β-alanine* metabolism the summed-up values for metabolites *β-alanine*, *β-amino propion aldehyde*, *carnosine*, *histidine* are 0.117, 0.0081, 0.0091 and 0.0091 respectively. The highest summed value is 0.117 whereas the highest payoff score is 0.11. Lastly, for *taurine* and *hypotaurine* metabolism the summed-up values for metabolites L-*cysteine*, L-*cysteate*, 3-*sulfino* L-*alanine*, *taurine*, *hypotaurine* are 0.0, 0.144, 0.072, 0.144 and 0.072 respectively. The highest summed value is 0.144 whereas the highest payoff score is 0.16 in this case [41].

4. Conclusions

Among various investigations undertaken in systems biology, an important task is identifying meaningful patterns within metabolic pathways. This approach aims at identifying essential regions within metabolic pathways of various organisms and is useful for studying evolution as well. Here, we developed an algorithm that has three modules, of which the *first* one was used for pattern identification on the basis of structural, functional features and topologies. Commonly occurring patterns identified similar patterns of conversion and mutation, and thus could provide key insight essential for drug development in relation to rogue pathways. We analyzed the onset of T1D in *H. sapiens* by studying the role of *GAD* and *INS* genes in metabolic pathways, namely, *glutamate*, *β-alanine*, *taurine* and *hypotaurine* and *butanoate* metabolisms, involving in T1D for the purpose of studying their growth in evolutionary time scale.

In our algorithm we used the concept of *Hamming distance* and *Tanimoto coefficient* for identifying certain graph-based *Community structures* in metabolic pathways, which were further used to detect meaningful patterns. We further used functional validations representing the patterns using cycles and ODEs. We then tested the stability of these cycles by calculating eigenvalues from the *Jacobian* matrix that we formulated from the ODEs. Likewise, metabolic networks consisted of large number of interactions that occur among its components. These interactions also gave an insight into the functional, structural, topological and evolutionary organization of the network. The *second* module of our algorithm identified interactions based on the patterns found by the first module. Study of interactions among the components of a network is usually carried out by partitioning it into sub-networks and analyzing the types of interactions occurring among them. In this work, we developed a technique for network-partitioning based on cut sets. We analyzed the size of metabolic pathways, which denoted the boundary assignment of the metabolites involved in the metabolic pathways. We further used *hasse* diagrams and implemented *DC* method to test for cuts and metabolite groups. Furthermore, higher the size of metabolic pathway, lower was the robustness, as more was the chance that the metabolic pathways could be targeted for external perturbation. In this work we categorized the interactions into *neighbored*, *serial*, *parallel*, *convergent* and *divergent*. The *third* module of our algorithm was to identify the growth of metabolic network based on the already known patterns, their interactions and some other parameters like performing mutational events. We further tested the various interactions by plotting *Payoff* matrices. These matrices identified those interactions which are of functional significance to the metabolic pathway. For evolutionary studies, we used the concept of identifying the densest metabolites in the pathway along with their reactions. We suggest that the densest metabolite in the metabolic pathway was the earliest one in terms of evolutionary occurrence. This is an extension of the classical 'micro-evolution' hypothesis. We use a combinatorial approach for the same. This approach began with comparing the metabolic pathways of *glutamate*, *β-alanine*, *taurine* and *hypotaurine* and *butanoate* metabolisms involved in *T1D* ongrowth in *H. sapiens* from the available public repositories.

An interesting feature that we could observe from this study was that with increase in mutational events the compactness of the earliest metabolite and its neighbors increased. However, this phenomenon occurred only when mutations occurred in its outermost neighbors. But, if the same events occurred in the nearest neighbors of the oldest metabolites, the complete metabolic network collapsed. Thus, in terms of functionality, the oldest metabolite was of the topmost priority. We concluded that *Hamming* distance and *TC* helped in signifying a biological meaning to the patterns. Moreover, the community structures found on the basis of Hamming distances consisted of metabolites separated by the same number of substitutions. We observed that the patterns found above were separated by the least number of substitutions and remain conserved through evolution. We could say that these patterns were significant as they were the least mutated through evolution, and needed to be further studied for analyzing any possible interactions, to predict their role in evolution. Furthermore, we found that the interactions found in the pattern behaved in divergent manner. But, a 'divergent' interaction might not suggest that the metabolites involved in them behaved in divergent manner during evolution. For this reason, it was evident that we further perform the analysis on these patterns and their interactions for studying their growth pattern and evolutionary significance. We found that the most significant interactions existing among the patterns found in these metabolic pathways were divergent in nature. We found that the densest metabolite, in terms of occurrence, in *glutamate*, *β-alanine*, *taurine* and *hypotaurine* and *butanoate* metabolisms were L-*glutamate*, L-*aspartate*, *taurine* and *acetoacetyl CoA*, respectively. We finally conclude that these four metabolic pathways grow in a manner which was very similar to divergent growth, i.e., the earliest metabolite evolves first, followed by the growth of its reaction links and branches, leading to a method, we termed as '*divergent root-to-leaf*' model, an extension to the 'micro-evolution' theory. Thus, the metabolites and associated reaction links other than the root or the most dense metabolites are thought to have evolved from these dense metabolites.

Conflict of interest statement

None declared.

Appendix A. Supplementary data

Supplementary data associated with this paper can be found in the online version at http://dx.doi.org/10.1016/j.compbiomed.2015.03.012.

References

[1] S.C.G. Risson, J.M. Thornton, Pathway evolution, structurally speaking, Curr. Opin. Struct. Biol. 12 (2002) 374–382.
[2] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.L. Barabási, Hierarchical organization of modularity in metabolic networks, Science 297 (5586) (2002) 1551–1555.
[3] M.E.J. Newman, Detecting community structure in networks, Eur. Phys. J. B 38 (2) (2002) 321–330.
[4] R.Y. Pinter, O. Rokhlenko, E. Yeger-Lotem, M. Ziv-Ukelson, Comparison of metabolic pathways, Bioinformatics 21 (16) (2005) 3401–3408.
[5] G.K. Holmes, Coeliac disease and Type 1 diabetes mellitus – the case for screening, Diabet. Med. 18 (3) (2001) 169–177.
[6] A.R.G. Humphrey, D.J. McCarty, I.R. Mackay, M.J. Rowley, T. Dwyer, P. Zimmet, Autoantibodies to glutamic acid decarboxylase and phenotypic features associated with early insulin treatment in individuals with adult-onset diabetes mellitus, Diabet. Med. 15 (2) (1998) 113–119.
[7] N.H. Horowitz, On the evolution of biochemical syntheses, Proc. Natl. Acad. Sci. USA 31 (6) (1945) 153–157.
[8] S. Granick, Speculations on the origins and evolution of photosynthesis, Ann. N. Y. Acad. Sci. 69 (2) (1957) 292–308.
[9] M. Ycas, On earlier states of the biochemical system, J. Theor. Biol. 44 (1) (1974) 145–160.
[10] L. Demetrius, Directionality theory and the evolution of body size, Proc. Biol. Sci. 267 (1460) (2000) 2385–2391.
[11] P. Wegner, A technique for counting ones in a binary computer, Commun. ACM 3 (5) (1960) 322.
[12] A. Clauset, M.E. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E. Stat. Nonlin. Soft Matter Phys. 70 (6 Pt 2) (2004) 066111.
[13] E. Klipp, Systems Biology in Practice: Concepts, Implementation And Application, John Wiley & Sons Inc., New York, 2005.
[14] S. Arora, S. Rao, U. Vazirani, Expander flows, geometric embeddings and graph partitioning, J. ACM 56 (2) (2009) 1–37.
[15] L.-E. Quek, L.K. Nielson, A depth-first search algorithm to compute elementary flux modes by linear programming, BMC Syst. Biol. 8 (2014) 94.
[16] J. Behre, L.F. de Figueiredo, S. Schuster, C. Kaleta, Detecting structural invariants in biological reaction networks, Methods Mol. Biol. 804 (2012) 377–407.
[17] N. Jamshidi, B.Ø. Palsson, Flux-concentration duality in dynamic nonequilibrium biological networks, Biophys J. 97 (5) (2009) L11–L13.
[18] M. Domijan, E. Pècou, The interaction graph structure of mass-action reaction networks, J. Math. Biol. 65 (2) (2012) 375–402.
[19] S. Schuster, J.-U. Kreft, A. Schroeter, T. Pfeiffer, Use of game-theoretical methods in biochemistry and biophysics, J. Biol. Phys. 34 (1–2) (2008) 1–17.
[20] J.B. Bard, The next evolutionary synthesis: from Lamarck and Darwin to genomic variation and systems biology, Cell Commun. Signal 9 (1) (2011) 30.
[21] I. Toby, Biological databases as research tools in the post-genomic era, Aviat. Space Environ. Med. 83 (4) (2012) 452–453.
[22] M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, Bioinformatics 19 (4) (2003) 524–531.
[23] R.W. Hamming, Error detecting and error correcting codes, Bell Syst. Techn. J. 26 (2) (1950) 147–160.
[24] W. Ndifon, J.B. Plotkin, J. Dushoff, On the accessibility of adaptive phenotypes of a bacterial metabolic network, PLoS Comput. Biol. 5 (8) (2009) e1000472.
[25] B.S. Chen, Y.P. Lin, On the interplay between the evolvability and network robustness in an evolutionary biological network: a systems biology approach, Evol. Bioinform. Online 7 (2011) 201–233.
[26] P.M. Gleiss, P.F. Stadler, A. Wagner, D.A. Fell, Relevant cycles in chemical reaction networks, Adv. Complex Syst. 1 (2010) 1–18.
[27] S. Tagore, R.K. De, Detecting breakdown points in metabolic networks, Comput. Biol. Chem. 35 (6) (2011) 371–380.
[28] M.J. Davis, C.J. Shin, N. Jing, M.A. Ragan, Rewiring the dynamic interactome, Mol. Biosyst. 8 (8) (2012) 2054–2066.
[29] D.J. Conway, C. Roper, Micro-evolution and emergence of pathogens, In. J. Parasitol. 30 (12–13) (2000) 1423–1430.
[30] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability), Springer, Dordrecht, 1993.
[31] R.M. Nesse, C.T. Bergstrom, P.T. Ellison, J.S. Flier, P. Gluckman, Evolution in health and medicine Sackler colloquium: making evolutionary biology a basic science for medicine, Proc. Natl. Acad. Sci. USA 107 (1) (2010) 1800–1807.
[32] J. Gu, Evolutionary systems biology, Curr. Genomics 12 (6) (2011) 379.
[33] S. Tagore, R.K. De, SAGPAR: structural grammar-based automated pathway reconstruction, Interdiscip. Sci. 4(2), 116–127.
[34] S. Grimbs, J. Selbig, S. Bulik, H.-G. Holzhütter, R. Steuer, The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks, Mol. Syst. Biol. 3 (2007) 146.
[35] S. Gerschgorin, Über die Abgrenzung der Eigenwerte einer Matrix, Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk. 6 (1931) 749–754.
[36] B. Papp, B. Szappanos, R.A. Notebaart, Use of genome-scale metabolic models in evolutionary systems biology, Methods Mol. Biol. 759 (2011) 483–497.
[37] Y. Goldstein, A. Bockmayr, A lattice-theoretic framework for metabolic pathway analysis, in: Computational Methods in Systems Biology, CMSB 2013, Klosterneuburg, Springer, LNBI, vol. 8130, 2013, pp. 178–191.
[38] L. Loewe, A framework for evolutionary systems biology, BMC Syst. Biol. 3 (2009) 27.
[39] D. Anastasiou, Metformin: a case of divide and conquer, Breast Cancer Res. 15 (2) (2013) 306.
[40] B.S. Chen, C.Y. Hsu, J.J. Liou, Robust design of biological circuits: evolutionary systems biology approach, J. Biomed. Biotechnol. 2011 (2011) 304236.
[41] B.F. Arnold, I. Größl, P. Stahlecker, The minimax, the minimin, and the Hurwicz adjustment principle, Theory Decis. 52 (3) (2002) 233–260.