I would like to thank the reviewers for their constructive comments on my manuscript "In silico identification of switching nodes in metabolic networks". In this document, I describe how the revised manuscript addresses the comments.

In summary, I have clarified several points throughout and better explained the results and the relevance of the approach. I have carried out a new procedure to address the second comment of Reviewer #2 concerning a possible bias due to the non-uniqueness of flux vectors.
Additionally, following some remarks of Reviewer #1, the order of the sections has been modified to improve the presentation of the approach. First, I propose two validations of the method on *E. coli* (the core network and the identification of metabolites involved in PTMs), followed by an application leading to the identification of a putative key metabolite for mixotrophy. The ISIS method is then compared with the reporter metabolite approach and finally, its robustness is examined.

Text changes have been highlighted in blue in the revised manuscript. I have also modified Figures 2 and 5, and added a new figure (Fig. 6) and supplementary information.

Reviewed by anonymous reviewer, 10 Jul 2023 10:11
In this paper, the author presents a constraint-based approach for the identification of key metabolites in metabolic networks. Such molecules, denoted as "switch metabolites" are identified using parsimonious flux balance analysis by comparing the flux distributions in the network obtained with multiple nutrient inputs. Fluxes are reoriented depending around these metabolites depending on the environmental conditions. The manuscript is overall well-written. The approach proposed by the author is interesting but the manuscript would be more convincing should more context, precisions and justifications be added. There is no convincing validation of the results provided by the approach. Sections are quite short, sometimes very short, more details could be added without making the manuscript excessively long. To my opinion, the conclusions on the relevance of the approach are not sufficiently supported by the results, although I believe that the work has a potential.
The manuscript has been revised, in particular to give more insights on the results and to better show the relevance of the approach. Below we explain more in details the modifications done in the manuscript.

Introduction:
The introduction is a bit short for non-experts and could better motivate the research question. Previous studies and or approaches are listed as references but they are not discussed, even briefly, making it complicated for a non-expert reader to grasp the context of the current study.
More details are now given in the introduction, as detailed below.

- Two consecutive sentences starting with "More and more"
This has been modified.

- A description of FBA that does not mention its main characteristics
FBA is now briefly explained (l. 28-30).

- Paragraph 2: no reference on the complexity of the analysis of large newtorks, or the fact that identification of key metabolites is a major concern: how will an answer to this question help tackle the complexity of metabolic networks?
Some references have been added ([15, 6, 30]). I also explain more in details that the switching nodes can be useful to understand the cellular response and to develop modeling, monitoring or control of cellular systems (l. 36-40).

- Existing methods for identifying key metabolites and the vision of the author on switch points could also be summarised in the text rather than simply being referred to.
The principles on which each method is based are now presented. I also describe my vision on switch point (l. 43-48).

Results:
A main remark on the results is the lack of validation of the switch approach. The idea of comparing to reporter metabolites is interesting but I do not see a validation of the results here, as very few metabolites are shared between the two approaches. The PTM analysis with comparison with ligand metabolites is more convincing but results do not go into details.
The approach has been developed to respond to a specific objective, and the proposed method, based on linear algebra, respond to it. Comparing the results with other approaches is not straightforward given that the objectives are not exactly the same. As explained below, I have modified the manuscript to better highlight the relevance of the approach, illustrated with the different case studies. The results are also now further discussed.

- E. coli simulations

* The names of the metabolites in Fig 2C are hardly visible.
Figure 2 has been modified.

* "This is in line with what we could expect..." L72. Any reference?
Two references have been added (l.89).

- S. cerevisiae simulations

* The three sizes of the pvalues/scores are hard to decipher in Fig3. Consider adding color?
This figure (Fig.5 in the new version) has been modified using gray scale.

* The text is brief, does not mention the number of switch metabolites wrt to reporters for instance.
More details are given now, in particular the number of switch and reporter metabolites (l.129-132).

* Few metabolites are common between reporters and switch metabolites in Fig 3.
Both methods have some drawbacks, which could explain why only some metabolites are shared between the two approaches. The differences between both approaches are now further discussed (l. 140-150).

* L89: "several reporter metabolites appears in the top of the switch metabolite list (see SI), even if they do not appear in Fig. 3": why do they not appear in Fig3? Please mention the contents of each SI (in the github repository) you refer to, somewhere in the text (after methods?)
Only the metabolites with the highest scores appear in the figure for presentational convenience, as explained in l. 131.
All the SI material mentioned in the article is now numbered, and a list is given at the end of the article.

- A rank of switch metabolites is mentioned L91. Are all metabolites of the network switches but with different scores, or are a subset of the nodes switches based on a threshold? Out of the total of 2241 nodes, many of them have a score of 0, so shouldn't be described as switches I guess.
All the metabolites are ranked by their score and the ones with the highest scores correspond to switch nodes. There is not a fixed threshold for all the cases as the score depends on several factors (size of the network, number of environmental conditions…). The manuscript has been modified to clarify this point (l.130-131 and l.255-257).

- P. tricornatum simulations

* is the role of erythrose-4-phosphate validated in the literature? There are no references in the text. Also this is presented in the abstract as a highlight but I do not see validation of the role of the metabolite. Please rephrase the sentence of the abstract with the mention "highlighting the (somehow overlooked) importance" if there is no validation.
To the best of my knowledge, the role of erythrose-4-phosphate has never been depicted, showing that ISIS can bring new knowledge. The abstract has been modified as I agree that it should deserve further study: “highlighting” has been replaced by ”suggesting”.

Discussion:
- "The methods gives sound results" L126. The validation of the relevance of the approach is only addressed in the PTM analysis
The discussion has been modified to emphasize the relevance of ISIS (l.173-183). First, I recall the mathematical principle for the identification of switch, which shows the soundness of the approach. I also recall the different cases with the corresponding results that show the relevance of the approach.

- The results on the reporter analysis on the yeast are presented positively: "One of the closer definition is that of reporter metabolites [18], and the results on a case study have shown some similarities (see the case study with S. cerevisiae). The main advantage of ISIS is that it does not require experimental data.". But one could see the glass half empty: a lot of the reporter metabolites were not identified as switches.
The difference between both approaches is further explained in the Result section (l. 140-150), as stated above, and this point is now also discussed in l.194-197.

- A justification for the absence of comparison with other methods is the difference in the definition of a key metabolite (L130): why would the definition advocated for in the manuscript be (more) biologically relevant?
Each definition corresponds to a specific objective. Several corrections have been made to better highlight the relevance of my approach, as explained above in my previous responses.

Methods:
Additional precisions should be added to facilitate the understanding what has been performed in the work.
More details are given in the new version of the manuscript, as explained below.

- A general comment on the section describing ISIS principle is that it lacks justifications on the mathematical choices, making it hard to follow for a non expert. This is the core of the approach presented in the manuscript and the underlying reasons of the modelling choices that were done are not provided. Here, pFBA has been used, which is understandable because it provides a unique optimal solution. Yet, the author mentions several times that other approaches could be used: would it have an impact on the results or are the switch metabolites robust to the optimisation choice?
The method relies on the estimation of the dimension of a vector space (as explained in l. 243-248) which is carried out as is usually the case using Singular Value Decomposition. The flux vectors are the input and are not the core of the method. They have been computed using pFBA or, for the case study on mixotrophy, they are taken from an article.
As the results are dependent on these fluxes, a section has been added to better show the influence they can have (l.152-170, 281-287 and Fig.6). The robustness of the results is evaluated using flux vectors obtained by sampling. Some switch metabolites are always identified. On the other hands, some metabolites can appear as switches only in some cases, showing that a good estimation of the flux vectors are needed, in particular for large networks.

- Is there a notion of threshold that is used to determine whether the metabolite is a switch or not, according to its score r_i? How/why were the thresholds of 0.1, 0.3, 0.5 chosen in the results?
As explained before, the score depends on several factors (in particular the number of conditions) so the idea is to depict the metabolites with the highest scores (l. 247, 254-256). The thresholds were chosen to distinguish between the highest scores.

- The definitions of FBA and pFBA could be more precisely described.
Done (see l. 229-237)

* The FBA problem has two types of constraints: the steady state assumption (SSA), and the inequality constraints of flux bounds. The latter is not presented as a main constraint of the problem at the same level as the SSA. In addition, the objective function in FBA does not have to be a maximisation. Theoretically, it can also be a minimisation.
This has been modified (l. 229-231)

* "Then, we determine the solution with the minimum overall flux that have (almost) the same objective value" L169 misses precision as well. The sum of fluxes is minimised. "Almost" in imprecise.
This has been modified (l. 237).

- L174: Could you give more details on the normalisation? By the sum of fluxes in the model?
Done (l. 242).

- For all simulations, I assume the biomass reaction was chosen as an objective function? This is not mentioned for any of the case studies.
This is now mentioned (l. 252).

- S. cerevisiae simulations.

* L195 "The switch metabolites are computed two by two". Could you precise the meaning (also used in results)?
The nitrogen sources are compared two by two. It has been clarified (l. 278).

* L194 mention some details about the remaining compounds of the medium?
Done (l. 276).

* L196: give access to the list of currency metabolites used, or maybe the raw results? Why removing them as you state that they are mentioned as switch metabolites and seem important in E. coli core?
The raw results are available on the Github repertory (SI4). The currency metabolites are removed because they were also removed in the reporter metabolite study. This has now been clarified (l. 280).

- E. coli simulations

- L200 what do you mean "with all the conditions at once"?
I compare all the conditions at the same time, and not two by two as in the *S. cerevisiae* case. This is now clarified (l. 266).

- P. tricornatum simulations

- Are the flux simulations obtained with pFBA as well? Context on the study from which the data was taken is missing.
The flux vectors, taken from [12], were computed by FBA coupled with Euclidean norm minimization. This is clarified in the Method section (l. 272).

- There is no section on statistics but tests are used in results (eg L84)
A statistics section has been added (l. 289-296).

Implementation: The implementation of the tool in Python 2.7 is a bit surprising as the support for this version of Python has ended 3 years ago. For ensuring the usage of the method by the community, the author could consider upgrading the code to a more recent version. First tests on the notebooks provided seem to show compatibility. It is not clear in the paper whether the implementation of the method is a main contribution of the work or not. At the moment, based on the repository, it would seem it is not: no installation procedure (library dependencies), no information on the versions used (eg cobrapy L182, Escher L192)... If the implementation is a contribution, the author might consider building a package out of the code, and using the notebooks as a demonstration support.
All the codes have been run in Python 3. A package will be proposed in the future.

A more general remark on the writing, that is maybe personal. The author uses a lot of parentheses to give additional information in the text, sometimes several occurrences in a single sentence (L175,176), or to mention rather important information (L92). I personally find that it disrupts the reading and makes sentences harder to understand.
Several parentheses have been removed and some sentences have been reworded to make them easier to read. For example, the information that was given l.92 in parentheses is now given in a new sentence (l. 149).

Reviewed by anonymous reviewer 2, 07 Jul 2023 14:35
The author proposes a method for determining the metabolic hubs around which metabolism rewiring occurs under different physiological conditions. The technique identifies the switching nodes by testing the algebraic collinearity of flux vectors corresponding to the reactions involving the studied metabolites. The author showcases the method through four case studies. The text is well-structured and relatively clearly written. However, many details are missing, and it would be challenging to replicate the study results.
More details are now given in the method section. The results are also further discussed (cf. comments to the first reviewer). Finally, all the codes are available on github, so that the results should be replicated now more easily and I hope that the approach could also be applied on other cases with minimum effort.

Whereas the author claims the generality of the approach, i.e., that it can use any set of flux solutions from different conditions, pFBA is used in all studies with the assumption that pFBA is giving a unique solution – which allows the author to use 1 "representative" flux vector per condition. However, except for specific and relatively small metabolic networks, pFBA will not give a unique solution, i.e., it will also allow multiple flux vectors satisfying both the objective (e.g., optimal growth) and the minimal sum of the fluxes. In my experience, it will even enable some reactions to operate in both forward and reverse direction (i.e., specific fluxes could take both positive and negative values). And this will give a significant bias in the provided results. In general, variability in the fluxes within each condition case will also affect the proposed measure. This should be studied and clarified.
This drawback is now further studied and discussed (l. 152-170, 282-288). To assess the robustness of the results, the identification of switch nodes has been performed using flux sampling. These results show that some key metabolites appear in almost all cases. On the other hand, some switch nodes are dependent of the flux sampling, pointing out that erroneous flux vectors can bias the approach. This is a common pitfall for metabolic modelling studies, calling for an improvement of flux estimations, in particular for large scale networks.

Since the proposed method identifies the key metabolites by studying fluxes and their dependencies, it reminds of the flux coupling (FC) methods. Can one use FC methods or inspire from them in the

proposed approach? The author should discuss how this is different/ how it improves over the flux coupling approaches.

FC analysis can be used to eliminate candidate switch metabolites, but does not provide a sufficient condition for identifying switches. This point is now discussed (l. 185-190).
