
gutsplash.docs

Release 0.0.1

Jun 01, 2021

Contents

1	Using gutSMASH	1
2	Content	3
2.1	About	3
2.2	Glossary	4
2.3	Help	5

CHAPTER 1

Using gutSMASH

gutSMASH is a tool that indentifies bacterial metabolic gene clusters involved in specialized primary metabolism.

2.1 About

gutSMASH is an algorithm that enables to systematically evaluate the metabolic potential of anaerobic bacteria by predicting both known and putative Metabolic Gene Clusters (MGCs). This tool is specifically designed to uncover MGCs that are involved in specialized primary metabolism from the gut microbiome that have been associated to microbe-microbe and host-microbe interactions. Moreover, gutSMASH also predicts energy-capturing-related MGCs, to evaluate the mechanisms bacteria employ to harvest energy. To be able to detect these MGCs, gutSMASH uses detection rules logic as antiSMASH, that are Pfam combination defining of a specific reaction. For more information on how to use gutSMASH you can also visit our online documentation in: <https://gutsmash.bioinformatics.nl/help.html>

2.1.1 How to run gutSMASH

- You can try out gutSMASH by using our online webserver: <https://gutsmash.bioinformatics.nl/>
- or download and install the gutSMASH command line version from here: <https://github.com/victoriapascal/antismash/tree/gutsmash>

2.1.2 How to cite

If you have found gutSMASH useful, please cite our preprint on the algorithm: <https://www.biorxiv.org/content/10.1101/2021.02.25.432841v1>, and/or the article describing the gutSMASH web server: <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab353/6279837>

2.2 Glossary

2.2.1 Gene Cluster Classes

The anaerobic metabolic pathways that gutSMASH predicts can be divided into different gene cluster classes based on their product:

- **Aliphatic amine:** ammonia derivatives where at least one H has been replaced by alkyl substituents (e.g.: Arginine to putrescine, Putrescine to spermidine pathways)
- **npAA:** non-proteinogenic amino acids (e.g.: Proline to 5-aminovalerate)
- **Aromatic:** derivatives of benzene (e.g.: P-cresol pathway, Caffeate respiration)
- **SCFA:** fatty acids with 5 carbon atoms maximum (e.g.: Acetate to butyrate pathway, Threonine to propionate)
- **SCFA-other:** a SCFA is produced in combination with another molecule. For instance, the *pdu* operon, releases propanol and propionate
- **Other:** pathways that don't fit in any other categories are classified in here (e.g.: *bai* operon)
- **E-MGC:** related to energy-capturing mechanisms (e.g.: Ech complex, Rnf complex)
- **Putative:** gene clusters of unknown function

2.2.2 Gene Cluster Types

gutSMASH uses some abbreviations internally to refer to the different types of primary metabolic clusters:

Pathway	Reference
<i>pdu</i> operon	https://pubmed.ncbi.nlm.nih.gov/27242734/
Arginine to putrescine	https://pubmed.ncbi.nlm.nih.gov/30183487/
Putrescine to spermidine	https://pubmed.ncbi.nlm.nih.gov/30183487/
<i>tma</i> operon	https://pubmed.ncbi.nlm.nih.gov/25873372/
AAA reductive branch	https://pubmed.ncbi.nlm.nih.gov/29168502/
Caffeate respiration	https://pubmed.ncbi.nlm.nih.gov/23315745/
Gallic acid metabolism	https://pubmed.ncbi.nlm.nih.gov/30054365/
Indoleacetate to scatole	https://pubmed.ncbi.nlm.nih.gov/30310076/
P-cresol synthesis gene cluster	https://pubmed.ncbi.nlm.nih.gov/11231288/
Phenylacetate to toluene	https://pubmed.ncbi.nlm.nih.gov/29556105/
hydroxybenzoate to phenol	https://academic.oup.com/femsec/article/94/9/fiy125/5042942
Carnitine <i>cai</i> TABCDE gene cluster	https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.1994.tb00470.x
Proline to 5-aminovalerate gene cluster	https://pubmed.ncbi.nlm.nih.gov/20937090/
Threonine to propionate	https://pubmed.ncbi.nlm.nih.gov/9484901/
Acetate to butyrate gene cluster	https://pubmed.ncbi.nlm.nih.gov/17241242/
Acetyl-CoA pathway	https://pubmed.ncbi.nlm.nih.gov/27733845/
Acrylate to propionate gene cluster	https://pubmed.ncbi.nlm.nih.gov/24553467/
Aminobutyrate to butyrate	https://pubmed.ncbi.nlm.nih.gov/27994578/
<i>eut</i> operon	https://pubmed.ncbi.nlm.nih.gov/20234377/
Fumarate to succinate	https://pubmed.ncbi.nlm.nih.gov/28049145/
Glutamate to butyrate gene cluster	https://pubmed.ncbi.nlm.nih.gov/27994578/
Glycine reductase	https://pubmed.ncbi.nlm.nih.gov/11271425/

Continued on next page

Table 1 – continued from previous page

Pathway	Reference
Leucine reductive branch	https://pubmed.ncbi.nlm.nih.gov/15654892/
Lysine degradation gene cluster	https://pubmed.ncbi.nlm.nih.gov/26620920/
PFOR-II pathway	https://pubmed.ncbi.nlm.nih.gov/32301184/
Pyruvate to acetate-formate	https://pubmed.ncbi.nlm.nih.gov/20622067/
Succinate to propionate	https://pubmed.ncbi.nlm.nih.gov/24553467/
porA pathway	https://pubmed.ncbi.nlm.nih.gov/31831639/
Arginine to Hcarbonate	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5300872/
Glycine cleavage	https://pubmed.ncbi.nlm.nih.gov/8375392/
Histidine to glutamate	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3429618/
Hydroxy-L-proline to proline	https://pubmed.ncbi.nlm.nih.gov/32180548/
<i>bai</i> operon	https://www.nature.com/articles/s41586-020-2396-4
sulfate to sulfide	https://pubmed.ncbi.nlm.nih.gov/10792666/

Energy-capturing-related gene clusters (E-MGCs):

Pathway	Reference
Rnf complex	https://pubmed.ncbi.nlm.nih.gov/23269825/
Tetrathionate to thiosulfate	https://pubmed.ncbi.nlm.nih.gov/10231485/
DMSO/TMAO reductase	https://pubmed.ncbi.nlm.nih.gov/1917829/
Nitrate reductase	https://pubmed.ncbi.nlm.nih.gov/2674654/
Ech complex	https://pubmed.ncbi.nlm.nih.gov/30850546/
Formate dehydrogenase	https://pubmed.ncbi.nlm.nih.gov/26443736/
Glycerol-3-P dehydrogenase	https://pubmed.ncbi.nlm.nih.gov/26443736/
NADH dehydrogenase I	https://pubmed.ncbi.nlm.nih.gov/26443736/

2.3 Help

You can find more technical information in our help or github page:

- <https://gutmash.bioinformatics.nl/help.html>
- <https://github.com/victoriapascal/gutmash>

You can contact us for any further question at gutmash.contact@gmail.com