

学校编码: 10384

分类号_____密级_____

学号: 33120131152852

UDC_____

厦 门 大 学

硕 士 学 位 论 文

代谢组学数据尺度缩放新算法及其应用

New Variable Scaling Method for Metabolomic Data Analysis
and its Application

刘新卓

指导教师姓名: 董继扬教授

专 业 名 称: 电子与通信工程

论文提交日期: 2016 年 月

论文答辩时间: 2016 年 月

学位授予日期: 2016 年 月

答辩委员会主席: _____

评阅人: _____

2016 年 月

厦门大学博硕士学位论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学博硕士学位论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1.经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年月日

厦门大学博硕士学位论文摘要库

部分缩写专业名词英汉对照表

NMR	Nuclear Magnetic Resonance	核磁共振
HRMAS	High Resolution Magic-Angle-Spinning	高分辨魔角旋转
FID	Free Induction Decay	自由感应衰减
PCA	Principal Component Analysis	主成分分析
PLS	Partial Least Squares	偏最小二乘
PLS-DA	Partial Least Squares-Discriminant Analysis	偏最小二乘判别分析
OPLS-DA	Orthogonal Partial Least Squares-Discriminant Analysis	正交偏最小二乘判别分析
MAS	Model Adaptive Scaling	模型自适应尺度缩放
UV	Unit Variance	单位方差法
Par	Pareto	帕莱托法
VAST	Variable Stability	变量稳定性法
ANOVA	Analysis Of Variance	方差分析
PQN	Probabilistic Quotient Normalization	概率商归一化
MCCV	Monte Carlo cross-validation	蒙特卡洛交叉验证
AUC	Area Under a Receiver Operating Characteristic Curve	受试者特性曲线下面积

厦门大学博硕士学位论文摘要库

目录

中文摘要.....	I
英文摘要.....	III
第一章 绪论	1
1.1 代谢组学简介	1
1.2 代谢组学数据分析.....	2
1.2.1 代谢组学数据预处理	2
1.2.2 统计分析与建模.....	3
1.2.3 生物学解析	4
1.3 本文研究内容和论文结构安排.....	5
第二章 代谢组学数据的自适应尺度缩放新算法	9
2.1 引言	9
2.2 常用的尺度缩放方法	9
2.3 模型自适应的尺度缩放新算法.....	12
2.3.1 理论与方法	12
2.3.2 实验数据.....	13
2.3.3 结果分析与讨论.....	14
2.4 本章小结	20
第三章 牛磺酸水平对罗非鱼生长的代谢组学研究	24
3.1 研究背景	24
3.2 实验与方法.....	26
3.2.1 实验设计与样品采集.....	26
3.2.2 样品制备及其 ^1H NMR 谱采集	27
3.2.3 数据预处理	27
3.3 结果与分析.....	28
3.3.1 不同牛磺酸水平下的代谢轮廓差异.....	28

3.3.2 不同生长时期的代谢轮廓差异.....	34
3.4 本章小结	39
第四章 总结与展望	45
4.1 本文总结	45
4.2 展望.....	46
攻读硕士期间发表的论文.....	48
致谢.....	49

厦门大学博硕士学位论文摘要库

CONTENTS

Abstract in Chinese	I
Abstract in English	III
Chapter 1 Introduction	1
1.1 Biref introduction of metabolomics	1
1.2 Data analysis in metabolomics	2
1.2.1 Data preprocessing	2
1.2.2 Statistical analysis and modeling	3
1.2.3 Biological analysis	4
1.3 Structure of this dissertation	5
Chapter 2 New scaling method for metabolomics data analysis	9
2.1 Introduction	9
2.2 Commonly used scaling methods	9
2.3 Model adaptive scaling method	12
2.3.1 Theory and method	12
2.3.2 Grouper plasma data	13
2.3.3 Results and discussions	14
2.4 Summary	20
Chapter 3 Metabonomics studies on the effect of taurine in tilapia	24
3.1 Research background	24
3.2 Materials and Methods	26
3.2.1 Experimental design and sample collection	26
3.2.2 NMR experiment	27
3.2.3 Data preprocessing	27
3.3 Results and discussions	28
3.3.1 The effect of taurine on ¹ H NMR metabolic profiling of tilapia plasma	28
3.3.2 The effect of growth on ¹ H NMR metabolic profiling of tilapia plasma	34

3.4 Summary	39
Chapter 4 Conclusions and prospects	45
4.1 Conclusions of this thesis	45
4.2 Prospects	46
Publication List.....	48
Acknowledgements.....	49

厦门大学博硕士学位论文摘要库

摘要

代谢组学，即利用高通量、高灵敏度与高精确定度的现代分析技术，结合模式识别方法研究生物体系在受到外界刺激或干扰因素后，其内源性小分子代谢产物的动态变化，探究代谢模式变化与生命现象之间的关系，已成为当今系统生物学的一个研究热点领域。随着研究的深入，数据分析的问题越来越突出，并已成为代谢组学研究的关键技术和瓶颈之一。本文以代谢组学数据尺度缩放为主要研究内容，发展新的代谢组学数据分析方法，并将其应用于鱼类营养代谢组学研究中。主要的研究内容包括：

一、提出了模型自适应尺度缩放方法。该方法根据后续的多元统计模型，自适应地优化尺度缩放系数，使后续的统计模型能够更鲁棒、更准确的获取数据中与研究问题相关信息。将新方法 with UV、Pareto 和 VAST 等常用方法比较，采用真实的代谢组学数据进行验证，结果表明，新方法能够减小噪声和无关代谢物信号的影响，提高重要代谢物的信号强度，有效保留原始谱图的结构信息，提高多元统计分析模型对特征信息提取的准确性。

二、将新的数据分析方法应用于罗非鱼营养代谢研究中。研究牛磺酸水平和生长时期对罗非鱼代谢的影响。分析结果表明：随着饲料中牛磺酸含量升高，多种氨基酸、O-乙酰糖蛋白、肌醇和柠檬酸含量有所降低，而牛磺酸、极低密度脂蛋白和乳酸含量显著升高；随着生长时期变化，血浆中多种氨基酸、糖类、脂类等物质含量发生了复杂的变化。

本文提出的模型自适应尺度缩放方法具有良好的普适性，应用新算法研究牛磺酸和生长时期对罗非鱼生长的影响，研究结果可为牛磺酸作用机制以及鱼类科学养殖提供参考依据。

关键词： 代谢组学；核磁共振；尺度缩放；罗非鱼；牛磺酸。

厦门大学博硕士学位论文摘要库

ABSTRACT

Metabolomics is a new discipline developed rapidly since the end of last century, which studies the dynamic changes of biological systems' endogenous small molecular metabolites caused by external stimulations or disturbances and explores the relationship between metabolism mode change and life phenomenon. Metabolomics use modern analytical techniques with high-throughput, high sensitivity and high precision and pattern recognition method to analyze complex metabolic data. Data analysis technique has become one of the key bottlenecks in metabolomics research. Variable scaling can adjust variables' weight and is of great significance to subsequent multivariate statistical analysis and biological interpretation. In this paper, the main research are the following two aspects:

1. In this paper, we proposed a model adaptive scaling(MAS) method, which updates the scaling coefficients of the variables by upgrading the preselected statistical model. A real-world nuclear magnetic resonance based metabolomics data set was used to evaluate the proposed method and compare with three commonly used scaling methods, *i.e.*, unit variance scaling, Pareto scaling and variable stability scaling. We analyze spectra profiling, predicative ability of the preselected multivariate model. Experimental results showed that the proposed method can effectively enhance important variables, preserve molecular information and promote the predicative ability of the preselected multivariate model.

2. In this part, we apply the new variable scaling method to the preprocessing of tilapia plasma data and study the effects of taurine levels and growth periods of the performance of tilapia. Results showed that the content of various amino acids, O-acetylglycoprotein, myo-inositol and citrate have decreased and taurine, VLDL and lactate have increased as the level of taurine upgrade. Besides, the content of amino acids, carbohydrate and lipid all change complexly at different growth periods.

In conclusion, the new variable scaling method proposed in this paper is flexible, effective and has good universality. The study on the effects of taurine levels and

growth periods in tilapia performance can provide reference basis for taurine mechanism and feed science about tilapia.

Keywords: Metabolomics; NMR; variable scaling; tilapia; taurine.

厦门大学博硕士学位论文摘要库

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.