

·运动人体科学·

高水平男子中短距离游泳成绩预测的代谢组学模型

李江华¹, 刘承宜², 沙海燕¹, 丁伟祥¹, 殷晓旺¹, 刘建坤¹

(1.江西师范大学 体育学院, 江西 南昌 330027; 2.华南师范大学 激光运动医学实验室, 广东 广州 510006)

摘 要: 为了探索利用代谢组学方法预测高水平男子中短距离游泳成绩的可能性。第 15 届亚运会赛前 1 个月, 每周 1 次, 利用基于核磁共振(NMR)的代谢组学方法对 18 名准备参加亚运会的男子运动员晨尿进行分析, 然后采用有监督的模式识别方法——Fisher 逐步判别分析法, 根据从一维核磁共振氢谱获取的标准化数据, 建立高水平男子中短距离游泳成绩预测代谢组学模型。结果显示: 模型的预测性能检验结果, 对比赛中奖牌选手(进入前 3 名)与非奖牌选手(未进入前 3 名)、决赛选手(进入前 8 名)与非决赛选手(未进入前 8 名)的样本预测准确率可达到 90%以上。初步证明了利用代谢组学模型预测高水平男子中短距离游泳比赛成绩的可行性。

关 键 词: 运动生理学; 代谢组学模型; 成绩预测; 男子游泳运动员; 晨尿

中图分类号: G804.2; G861.1 **文献标识码:** A **文章编号:** 1006-7116(2010)04-0103-04

Metabonomical models for predicting the performances of high performance male medium and long distance swimmers

LI Jiang-hua¹, LIU Cheng-yi², SHA Hai-yan¹, DING Wei-xiang¹, YIN Xiao-wang¹, LIU Jian-kun¹

(1.School of Physical Education, Jiangxi Normal University, Nanchang 330027, China;

2.Laser Sports Medicine Laboratory, South China Normal University, Guangzhou 510006, China)

Abstract: In order to probe into the possibility to predict the performances of high performance male medium and long distance swimmers by utilizing metabonomical methods, the authors analyzed the morning urine of 18 male swimmers preparing for participating in the Asian Games once a week, 1 month before the 15th Asian Games, by utilizing metabonomical methods based on nuclear magnetic resonance (NMR), then established metabonomical models for predicting the performances of high performance male medium and long distance swimmer based on standardized data obtained from the one dimensional NMR spectrum of hydrogen by using a monitored mode identification method – the gradual discriminant analysis method created by Fisher, and revealed the following test result of the prediction performance of the models: the sample prediction accuracy rates of the medalists (top 3), non medalists (beyond top 3), finalists (top 8) and non finalists (beyond top 8) were up to 90%, which preliminarily proved the feasibility to predict the competition performances of high performance male medium and long distance swimmers by utilizing metabonomical methods.

Key words: sports physiology; metabonomical model; performance prediction; male swimmer; morning urine

运动成绩是运动员竞技能力的体现, 除了技术、战术等要求外, 与运动员身体机能水平有密切的关系。因此, 运用某些生理生化指标来预测运动成绩, 已逐渐应用于体育科学研究和运动训练中^[1]。例如, 根据运动员的最大摄氧量来预测中长跑的运动成绩, 根据两

级负荷运动后血乳酸值和跑速来预测 400 m 跑的运动成绩; 根据尿肌酐排泄量来预测举重成绩等, 都有相应的文献报道^[1]。这些指标部分反映了运动员的机能状态, 确实与运动员的运动成绩相关。然而, 这类文献的研究结果几乎都不是来源于优秀运动员群体。

收稿日期: 2010-01-19

基金项目: 江西省教育科学“十一五”规划课题(09YB210); 江西师范大学博士启动基金(2863)。

作者简介: 李江华 (1972-), 男, 博士, 研究方向: 运动代谢组学。通讯作者: 刘承宜教授。

Myburgh^[2]曾指出,在水平相对接近的一流运动员(elite athletes)中,甚至二流运动员(sub-elite athletes)群体中,这些指标并不足以用来预测运动员的运动成绩。以最大吸氧量(VO_{2max})为例,虽然至今仍然被认为是评价运动员耐力水平的最佳指标^[3],但是众多的研究结果显示,在训练良好的中长跑优秀运动员群体中,运动成绩与 VO_{2max} 的相关系数较低^[4-6];Abe等^[7]的研究显示即使是日本的二流中长跑青年运动员,运动成绩与 VO_{2max} 相关系数也仅仅是0.39。因此,有识之士早已指出,优秀运动员的运动成绩并不是由单因素或少数几个生理、生化指标所决定的,只有多学科交差、整体的生物学方法才可能对优秀运动员的运动成绩进行预测^[2,8]。

代谢组学是一种基于整体思路的科研方法,它试图通过测量样本中所有的代谢产物来反映生物体对不同刺激的反应^[9]。与传统生物学将生命现象与少数几个生物学指标进行相关研究的方法相反,代谢组学利用整体的分析方法对生命系统进行研究^[10]。李江华等^[11]曾经利用基于核磁共振(NMR)的代谢组学方法对参加15届亚洲运动会的游泳运动员进行了研究,通过无监督的模式识别方法——主成分分析,证明了运动员尿液中包含了运动员竞技水平的信息。本研究进一步采用有监督的模式识别方法——Fisher逐步判别分析法,利用从一维核磁共振氢谱获取的标准化数据,探索利用代谢组学方法预测高水平男子中短距离游泳成绩的可能性。

1 研究对象与方法

1.1 对象

参加多哈亚洲运动会的18名男游泳运动员,分别参与了4种不同游泳姿势和3种不同距离游泳项目的比赛:蛙泳(50、100、200 m)、仰泳(50、100、200 m)、自由泳(50、100、200 m)、蝶泳(50、100、200 m)和混合泳(200 m)。进入决赛的项目包括蛙泳(50 m)、仰泳(50、100、200 m)、自由泳(100、200 m)、蝶泳(50、100 m),其中7人进入了前8名,4人进入了前3名(其中2人次获得金牌2枚,3人次获得银牌4枚,2人次获得铜牌2枚)。本研究将运动员按亚运会的比赛成绩分为决赛组(进入前8名)和非决赛组(未进入前8名)、奖牌组(进入前3名)和非奖牌组(未进入前3名)。

1.2 样品收集和处理

亚运会赛前1个月内,每周1次,连续收集运动员晨尿3次。18名参与实验的运动员中,有1名运动员只收集到1次尿液,3名运动员收集了2次,其余14名运动员收集3次。尿液收集后,10 mL离心管分装,-20℃保存,直至进行核磁共振信号采集^[11]。

1.3 一维核磁共振氢谱

尿液样品于室温下解冻,每个样品取400 μL,加入200 μL磷酸缓冲液(pH7.4,0.2 mol/L NaH_2PO_4 和0.8 mol/L Na_2HPO_4),质量分数10%重水(D_2O ,锁场),质量分数0.05%TSP(内标)。静置10 min,再于13 000 g离心10 min,取上清,用Bruker DRX500型超导傅立叶变换核磁共振(NMR)仪在500.13 MHz共振频率下测量。经过傅立叶变换,手动相位和基线校正,以TSP共振峰为参考调零,得到一维氢谱。

1.4 数据处理

为了消除核磁共振采集信号过程中压水峰所造成的影响,去除了水峰和尿素峰附近6.2~4.6区段。然后对10~0.2区段进行分段积分,每段为0.02,结果从每张氢谱获得了409个相应的积分数据。在建立数学模型之前,每段的积分数据相对于每张氢谱进行了标准化处理,计算公式如下:

$$x_{ik}^* = x_{ik} / \sum_{i=1}^n x_{ik}$$

式中, x_{ik} 为第*k*个样本,*i*区段的原始积分数据; x_{ik}^* 为标准化以后的数据。

1.5 数学模型的建立

利用从一维核磁共振氢谱获取的标准化数据,采用有监督的模式识别方法——Fisher逐步判别分析,建立中短距离男子游泳成绩预测数学模型。模型预测的准确性与可靠性分别用两种方法进行检验。第1种是“留一法”,即每次留1个样本做为未知样本(预测集),其它样本做为“训练集”建模,待模型建好后,再把该未知样本的数据输入,让计算机自动识别该样本的类别,有*n*个样本,则此过程重复*n*次,保证每个样本的类别都是计算机预测出来的。第2种是“独立检验法,即将每一类别的所有实验对象平均分成两半,其中来自一半实验对象的样本做为训练集,计算机经过训练之后,对来自另一半实验对象的样本类别进行预测。

1.6 主要仪器与软件

信号采集使用德国Bruker公司ARX-500型NMR波谱仪,NMR波谱处理使用MestReC4.9.9.6,数据处理及数学建模使用SPSS11.0。

2 结果

2.1 Fisher线性判别分析

Fisher判别分析方法的基本思想是投影,即将*K*类*p*维投影到某个方向,使得不同类别间的样本距离最大,相同类别中的样本各自的距离最小,也即采用方差分析的方法判断分类^[12]。根据SPSS输出的Fisher线性判别函数(Fisher's linear discriminant functions)做

图,投影效果如图1所示。从图1a可以看出,投影取得了良好的效果,比赛中进入了前8名的选手(决赛选手, finalists)和没有进入前8名的选手(非决赛组)的样本被明确地分开了,同时相同类别中的样本则出现了聚集;图1b的效果虽然相对稍差,但比赛中进入了前3名的运动员(奖牌组)和没有进入前3名的运动员(非奖牌组)的样本还是出现了明显的距离。至此,对于任一个样本 X ,判断其属于哪个总体(类别),可以按马氏距离最近准则对 X 进行判别分类。如图1a所示,图1a中只有决赛选手与非决赛选手两个类别,首先计算样本 X 到决赛选手样本总体的马氏距离 $d_1^2(X, G_1)$,然后计算样本 X 到非决赛选手样本总体的马氏距离 $d_2^2(X, G_2)$,通过比较 d_1^2 与 d_2^2 的大小,把样本 X 判归距离较小的那个类别。图1b中样本类别的判别分析过程与此一致。

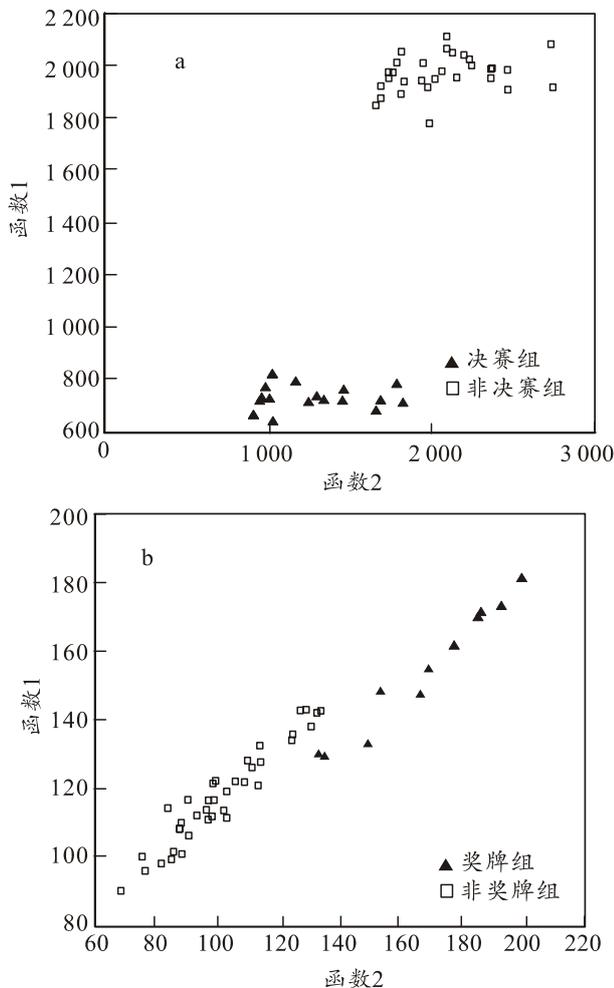


图1 Fisher 线性判别函数投影效果图

2.2 预测结果及模型的预测性能检验

用 Fisher 判别分析法确立模型后,计算机的初始识别正确率达到了百分之百,其中 19 个决赛选手样本

与 30 个非决赛选手样本、11 个奖牌选手样本与 38 个非奖牌选手样本全部被正确识别。同样,在“留一法”检验中,计算机也正确地预测了所有样本的类别。在用“独立检验法”检验时,也只有 3 个非决赛选手的样本和 1 个奖牌选手的样本的类别被计算机预测错误,其中 91.5% 的待测样本预测正确。

3 讨论

3.1 用代谢组学方法预测优秀运动员运动成绩的可行性

代谢组学(metabonomics)是继基因组学、转录组学和蛋白质组学之后新近发展起来的一门学科,是系统生物学的重要组成部分^[13],相对于基因组学和蛋白质组学,在人体研究中,代谢组学,尤其是以尿液为样本的基于核磁共振(NMR)的代谢组学研究有其独特的魅力,也正是由于代谢组学研究方法的优越性,为建立高水平中短距离男子游泳成绩预测模型奠定了方法学基础。

(1)时效性。基因组学和蛋白质组学分别从基因和蛋白质层面探寻生命的活动,但是基因组的变化不一定能够得到表达,某些蛋白的浓度变化也可能因为这个蛋白不具备活性,而对系统没有产生影响。并且由于基因或蛋白的功能补偿作用,某个基因或蛋白的变化会由于其它基因或蛋白的变化而得到补偿,最后反应的净结果为零,而小分子的产生和代谢才是这一系列事件的最终结果^[14]。因此有人认为,基因组学和蛋白质组学告诉你什么可能会发生,而代谢组学则告诉你什么确实发生了^[15]。

(2)整体性。首先人体是由细胞构成的,细胞的任何活动都会伴随相应的代谢反应,产生相应的小分子代谢产物。其次,人类的整体代谢活动应该包括人体本身的代谢、寄生菌群的代谢、两者的共代谢以及两者代谢物质交换引起的变化^[16-17],建立这些生命活动的相互联系才可能完成所谓“系统水平的认识”。尿液的代谢包含了人体内每一个细胞的代谢信息(包括宿主和菌群),也包含了宿主和菌群代谢活动的相互作用^[16-17]。因此,以尿液为研究对象的代谢组学研究能够从整体的角度反映人体这一复杂系统的状况。

(3)灵敏性。代谢产物作为生物系统的生理终端,对生理信号具有放大作用^[18]。人体内各种代谢变化快速而又协调,几乎都由酶催化或参与调节^[19],酶具有巨大的催化能力,因此,人体内各种代谢反应中,任何微小的生理调节,都会引起代谢产物的剧烈变化^[20]。所以代谢组学方法相对其他研究方法来说,具有高度的灵敏性。

3.2 模式识别方法与模型预测性能检验

模式识别方法可分为有监督方法和无监督方法^[21]。有监督的方法需要有一训练集,例如对于两类的情况,在训练集中,有一些样本属于 A 类,另外一些样本属于 B 类,然后教给计算机,计算机经过训练之后则可识别未知样本,因此被广泛地用来建立各种预测模型。并且由于“训练集”的存在,有监督的模式识别方法往往更具有实用性和更高的预测精度。本文采用有监督的模式识别方法——Fisher 逐步判别分析,建立了中短距离男子游泳成绩预测代谢组学模型,并使用不同的交叉验证(cross validation, CV)方法对模型的预测性能进行了检验。

在“留一法”验证中,所有样本的类别都得到了正确的预测,即使用最严苛模型预测性能检验方法“独立检验法”,进行检验时,待测样本预测正确率还是达到了 91.5%。CV 是著名统计学家 Seymour Geisser 提出来的数学模型预测性能检验方法,这一方法的推出,对于科研建模工作者来说,无疑具有里程碑式的意义。尤其是当样本量较小,而进一步收集样本又很难,或者根本不可能的情况下,CV 能够很好地对模型预测性能进行检验。其核心思想是将测试的样本人为地分为“训练集”和“预测集”,其中“训练集”用来训练建模,“预测集”则用来检验所建模型的预测能力。验证结果表明用 Fisher 逐步判别分析法建立的预测模型的预测性能良好,初步证明了利用代谢组学方法预测高水平男子中短距离游泳比赛成绩的可行性。

参考文献:

- [1] 冯炜权,谢敏豪,王香生,等. 运动生物化学研究进展[M]. 北京:北京体育大学出版社,2006.
- [2] Myburgh K H. What makes an endurance athlete world-class? Not simply a physiological conundrum[J]. *Comp Biochem Phys: Part A*, 2003, 136: 171-190.
- [3] Wilmore J H, Costill DL Costill. *Physiology of sport and exercise*[M]. 3rd ed, Champaign (IL): Human Kinetics, 2004.
- [4] Conley D L, Krahenbuhl G S. Running economy and distance running performance of highly trained athletes[J]. *Med Sci Sports Exerc*, 1980, 12: 357-360.
- [5] Morgan D W, Baldini F D, Martin P E, et al. Ten kilometer performance and predicted velocity at $\dot{V}O_{2max}$ among well-trained male runners[J]. *Med Sci Sports Exerc*, 1989, 21: 78-83.
- [6] Daniels J, Daniels N. Running economy of elite male and elite female runners[J]. *Med Sci Sports Exerc*, 1992, 24: 483-489.
- [7] Abe D, Yanagawa K, Yamanobe K, et al. Assessment of middle-distance running performance in sub-elite young runners using energy cost of running[J]. *Eur J Appl Physiol Occup Physiol*, 1998, 77: 320-325.
- [8] Collins J P. May you live in interesting times: using multidisciplinary and interdisciplinary programs to cope with change in the life sciences[J]. *Bioscience*, 2002, 52: 75-84.
- [9] Bertram H C, Hoppe C, Petersen B O, et al. An NMR-based metabolomic investigation on effects of milk and meat protein diets given to 8-year-old boys[J]. *Br J Nutr*, 2007, 97(4): 758-763.
- [10] Jonsson P, Johansson A I, Gullberg J, et al. High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses[J]. *Anal Chem*, 2005, 77: 5635-5642.
- [11] 李江华,刘承宜,徐晓阳,等. 2006 多哈亚运会短距离游泳男运动员志愿者代谢组学研究[J]. *体育科学*, 2008, 28(2): 42-46.
- [12] 鲁晓春,施先亮. 应用 Fisher 线性判别分析法评估物流规划项目[J]. *物流技术*, 2007, 26(8): 76-80.
- [13] Goodacre R. Making sense of the metabolome using evolutionary computation: seeing the wood with the trees[J]. *J Exp Bot*, 2004, 10: 431-435.
- [14] Henry C M. New “ome” in town[J]. *Chem Eng-News*, 2002, 80(48): 66-70.
- [15] German J B, Bauman D E, Burrin D G, et al. Metabolomics in the opening decade of the 21st century: building the roads to individualized health[J]. *J Nutr*, 2004, 134(10): 27-29.
- [16] Nicholson J K. Global systems biology, personalized medicine and molecular epidemiology[J]. *Mol Systems Biol*, 2006, 2: 52-56.
- [17] Nicholson J K, Wilson I D. Understanding ‘global’ systems biology: metabolomics and the continuum of metabolism[J]. *Nat Rev Drug Discov*, 2003, 2: 668-676.
- [18] Oresic M, Vidal-Puig A, Hänninen V. Metabolomic approaches to phenotype characterization and applications to complex diseases[J]. *Expert Rev Mol Diagn*, 2006, 6(4): 575-585.
- [19] 冯美云. 运动生物化学[M]. 北京:人民体育出版社,1999: 162-163.
- [20] Cascante M. Metabolic control analysis in drug discovery and disease[J]. *Nat Biotechnol*, 2002, 20: 243-249.
- [21] 许禄,化学计量学[M]. 北京:科学出版社,2005.