

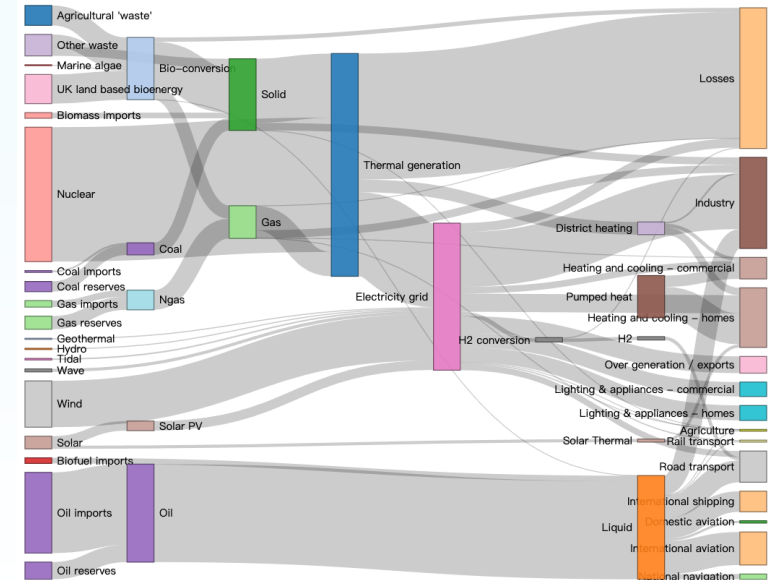


35马尔可夫链量化菌群状态转变

王金锋 中科院北京生科院

2020年12月6日

易生信，毕生缘；培训版权所有。



目录

- 状态转变及其研究方法
- 马尔可夫链介绍
- 用马尔可夫链量化菌群状态转变

易生信 毕生缘 宏基因组

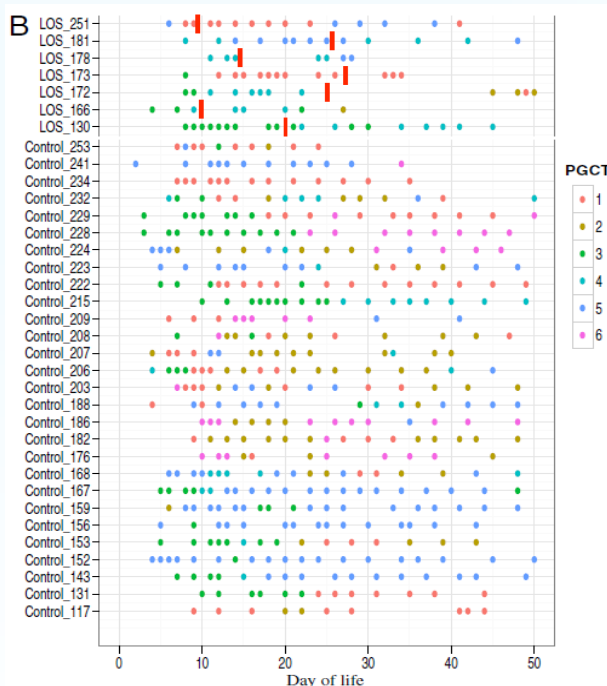


时序数据下的状态转变问题

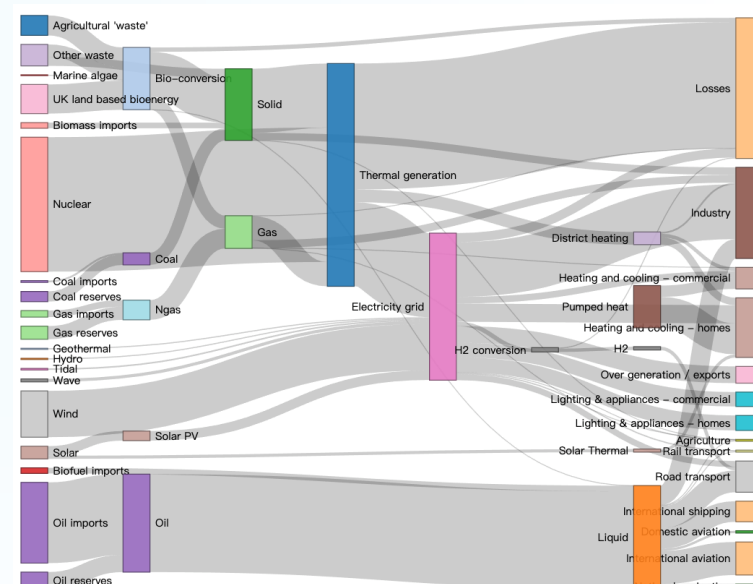
- 在对多时间点时序数据进行分析时，常常涉及到不同状态之间的转换，可以使用状态转换表和状态转变图表示不同时间点的不同状态

时间点	1	2	3	4	5	...
个体1	A	A	A	B	B	...
个体2	A	A	A	A	A	...
个体3	A	B	C	B	A	...
个体4	B	B	A	A	B	...

状态转移表显示状态转变



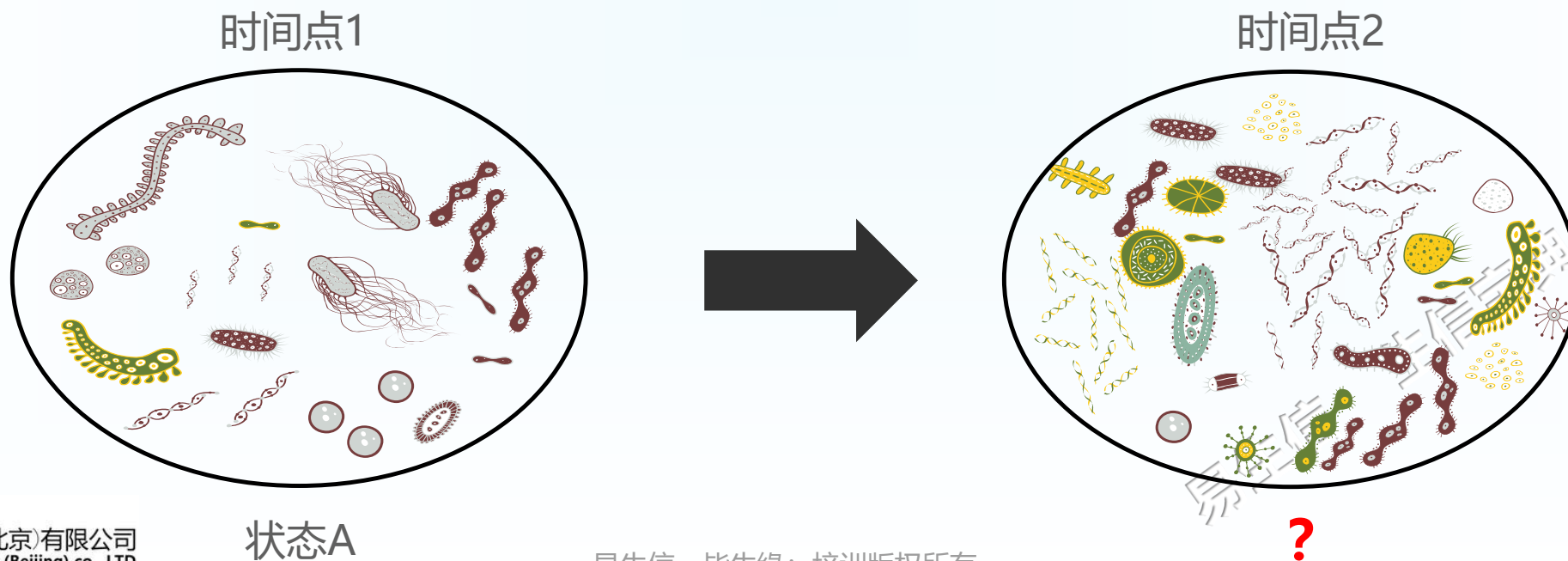
散点图显示状态转变



桑基图显示状态转变

研究状态转变的意义

- 对于不同状态的划分有助于简化分析，减少因个体间的差异过大而带来的异质性
- 通过对现有的不同状态之间的转移情况的建模和计算，还可以对未来状态的转移情况进行预测



衡量与预测状态转变的方法

- 可以选择多种模型来描述与预测状态的转移：
- 狄利克雷多重混合模型 (Dirichlet multinomial mixture modeling)
- 混合效应模型 (mixed-effect models)
- 马尔可夫模型 (Markov chain model)
-

目录

- 状态转变及其研究方法
- 马尔可夫链介绍
- 用马尔可夫链量化菌群状态转变

易生信 毕生缘 宏基因组

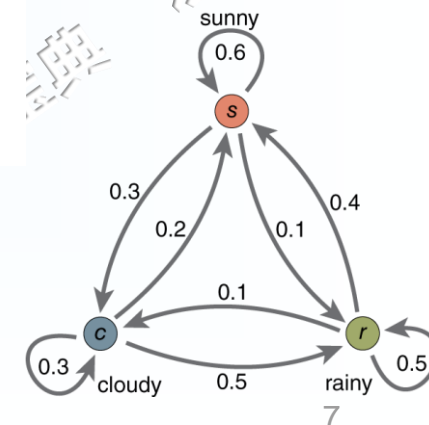
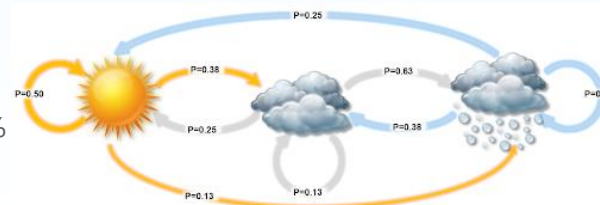


马尔可夫链

- 马尔可夫链 (Markov chain) ,
 又称离散时间马尔可夫链
 (discrete-time Markov chain) ,
 因俄国数学家安德烈·马尔可夫 (俄
 语: Андрей Андреевич
 Марков) 得名, 为状态空间中从
 一个状态转换到另一个状态的随机
 过程



安德雷·安德耶
 维齐·马尔可夫
 (1856-1922)
 俄国数学家



图片来源:
<https://zh.wikipedia.org/wiki/%E5%AE%89%E5%BE%B7%E9%9B%B7%C2%B7%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB>
<https://zhuanlan.zhihu.com/p/37847722>;
<https://blog.csdn.net/bitcarmanlee/article/details/77488620>

马尔可夫链

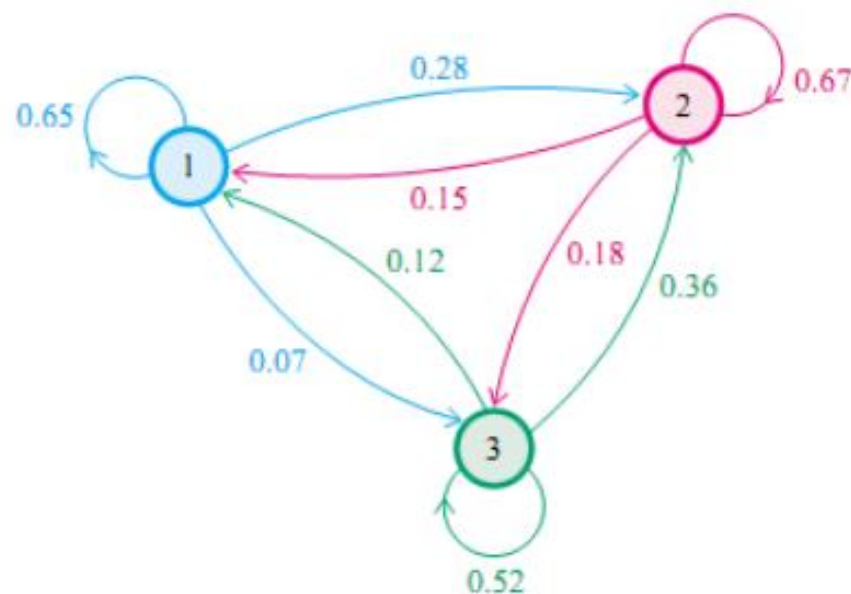
- 该过程要求具备“无记忆”的性质：**下一状态的概率分布只能由当前状态决定**，在时间序列中它**前面的事件均与之无关**。这种特定类型的“无记忆性”称作马尔可夫性质
- 在马尔可夫链的每一步，系统根据概率分布，**可以从一个状态变到另一个状态，也可以保持当前状态**。状态的改变叫做转移，与不同的状态改变相关的概率叫做转移概率
- 使用马尔可夫链可以很直观地对不同状态之间的转移进行展示、评估和预测

状态转移矩阵

- 在对马尔可夫链进行可视化之前，需要计算不同状态之间的转移概率
- 使用状态转移矩阵来表示不同状态之间的转移概率
- 依据状态转移矩阵可以计算出某一时间点任意状态的转移情况

		子代		
State		1	2	3
父代	1	0.65	0.28	0.07
	2	0.15	0.67	0.18
	3	0.12	0.36	0.52

状态转移矩阵



马尔可夫链



目录

- 状态转变及其研究方法
- 马尔可夫链介绍
- 用马尔可夫链量化菌群状态转变

易生信 信信信 宏基因组



肺部菌群随时间的变化

例1：慢性阻塞性肺病病人的肺部细菌随时间的变化

ORIGINAL ARTICLE

Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations

David Mayhew,¹ Nathalie Devos,² Christophe Lambert,² James R Brown,¹ Stuart C Clarke,^{3,4} Viktoriya L Kim,⁴ Michal Magid-Slav,¹ Bruce E Miller,⁵ Kristoffer K Ostridge,⁴ Ruchi Patel,⁶ Ganesh Sathe,⁶ Daniel F Simola,¹ Karl J Staples,^{3,4,7} Ruby Sung,⁵ Ruth Tal-Singer,⁵ Andrew C Tuck,³ Stephanie Van Horn,⁶ Vincent Weynants,² Nicholas P Williams,⁴ Jeanne-Marie Devaster,² Tom M A Wilkinson,^{3,4,7} on behalf of the AERIS Study Group

肺部菌群

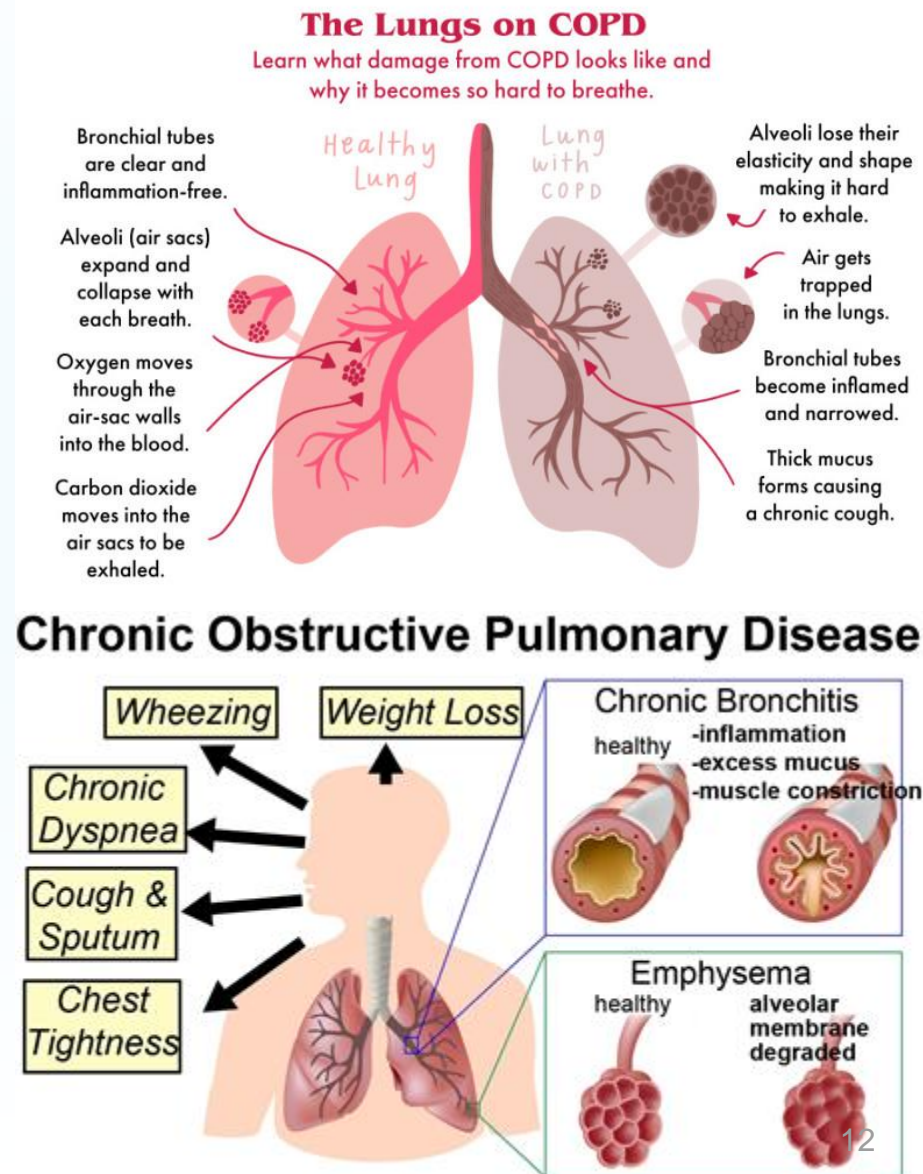
易生信

易生信



慢阻肺菌群研究背景

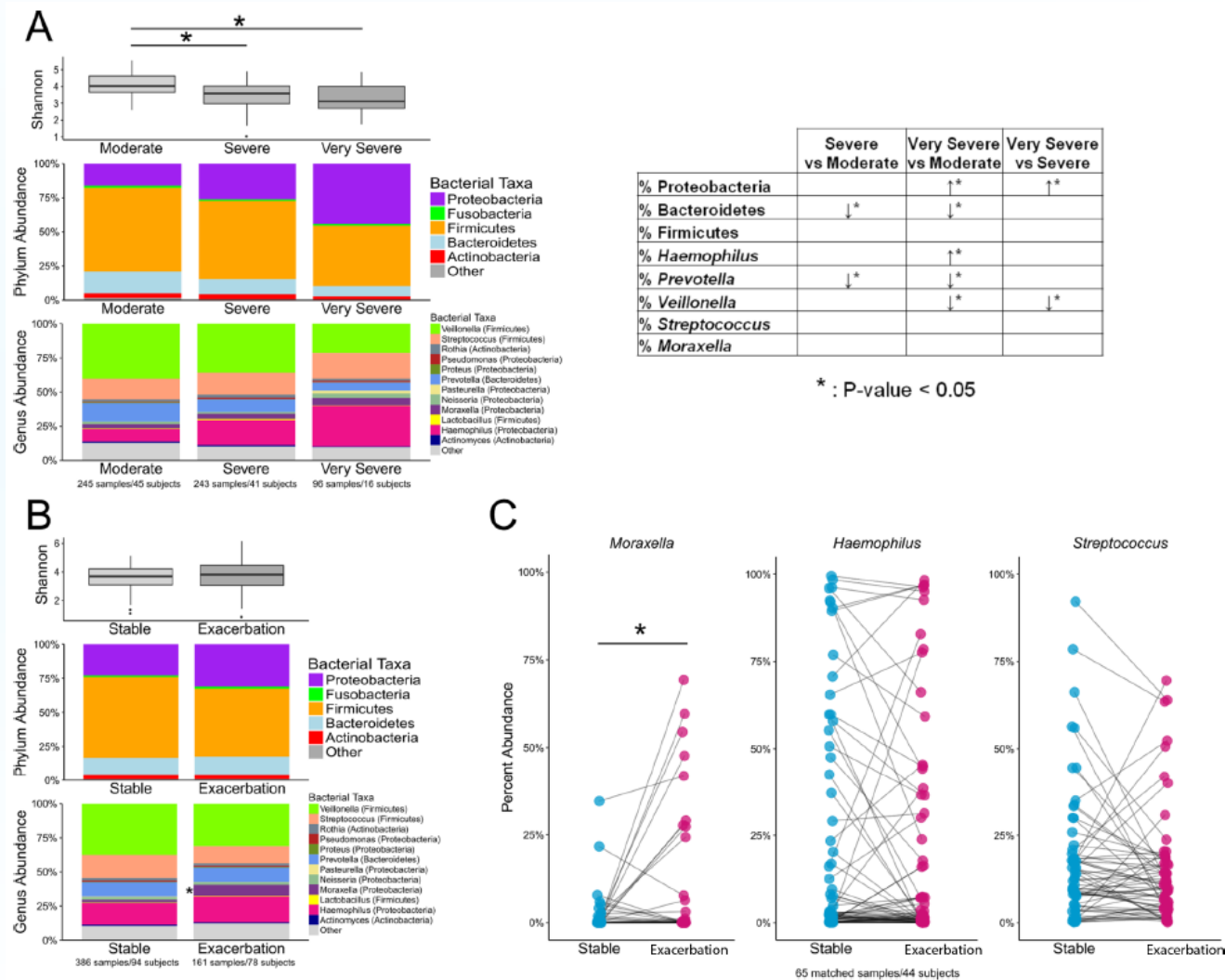
- 慢性阻塞性肺病（COPD）是一种慢性炎症性疾病，由于吸入烟草烟雾或其他刺激性物质，导致肺功能不可逆转地下降
- 根据主要致病原的不同，COPD可以分为多种恶化程度不同的亚型，包括：
细菌性感染（bacterial infection）、**病毒性感染**（viral infection）以及高水平嗜酸性粒细胞状态（high eosinophil levels）



图片来源: <https://www.healthcentral.com/article/how-copd-affects-lungs>

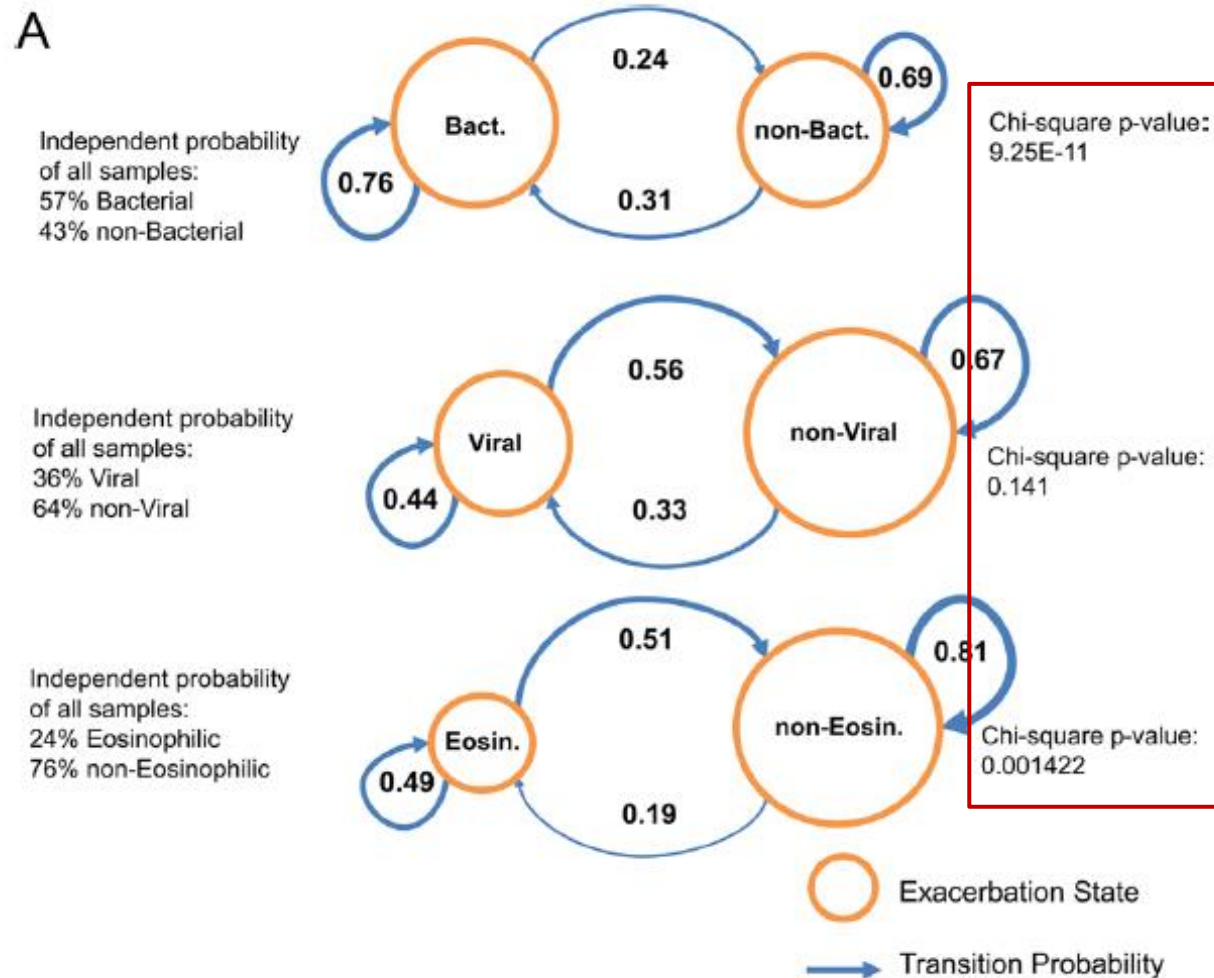
慢阻肺菌群的基本情况

- 不同恶化程度病人的肺部菌群在群落结构和丰度上都具有显著差异
- 但在病情稳定和恶化的病人之间，菌群差异并不明显
- 对于同一病人，病人在病情稳定和恶化前后，只有摩拉克氏菌属 (*Moraxella*) 出现显著改变



马尔可夫链对恶化程度的衡量

- 使用**马尔可夫链**来分别描述三种不同**恶化程度**的慢性阻塞性肺病的转变情况
- 每个圆圈大小代表该状态下的人群数量，每条边的粗细代表不同状态之间的转移概率
- 细菌感染**和**嗜酸性恶化**状态的患者更有可能**重复自身的恶化状态**，而病毒感染的患者恶化方向出随机性



马尔可夫链衡量不同类型感染

- 流感嗜血杆菌 (*Haemophilus influenzae*, *Hi*) 会感染呼吸道, 是 COPD 患者肺部的主要菌属
- 使用马尔可夫链分析 *Hi* 阳性和 *Hi* 阴性细菌感染的患者以及非细菌恶化状态的患者
- 与 *Hi* 阴性相比, *Hi* 阳性患者更有可能重复进行随后的 *Hi* 阳性恶化, 转移为非细菌性恶化状态的可能性较小

B

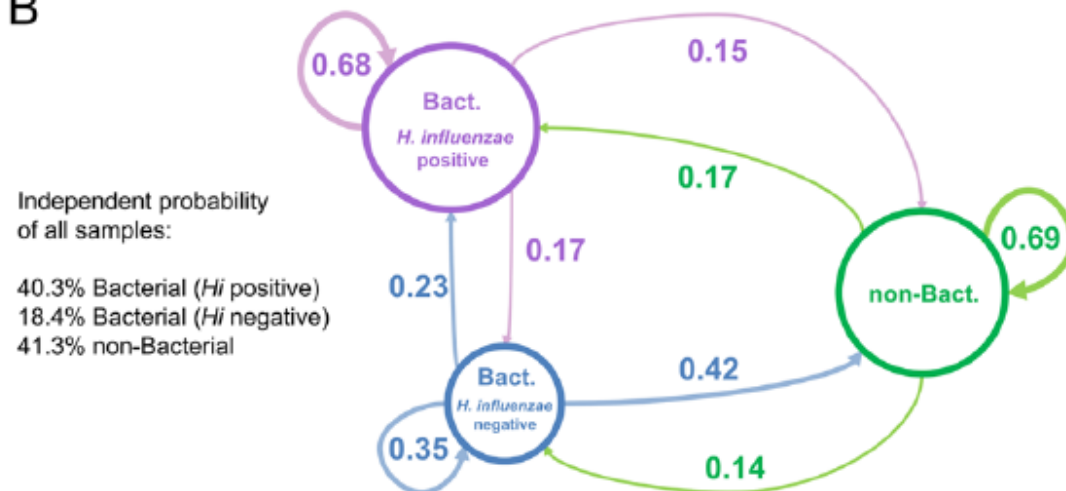


Figure 4 Markov chain analysis of transitions between exacerbation states. (A) Markov chain analysis from longitudinal exacerbation sampling within individuals identifies non-random transition probabilities for bacterial and eosinophilic exacerbations, but not viral. The size of each node is proportional to abundance of that exacerbation type, and the width of the edges is proportional to the transition probabilities. (B) Markov chain analysis of the bacterial exacerbation identifies significantly different transition probabilities for bacterial exacerbations that were positive or negative for the presence of *Haemophilus influenzae* (HI).

应用马尔可夫链的意义

- 使用马尔可夫链对COPD的恶化方向进行描绘和预测
- 根据不同亚型的COPD恶化状态的转变方向，可以为COPD的诊断进行有效的预测并在后期进行积极的干预治疗
- 摩拉克氏菌属 (*Moraxella*) 与嗜血杆菌 (*Haemophilus*) 可以作为COPD的标志菌 (marker bacteria) 对不同程度的恶化状态进行区分

- 例2：孕期妇女不同身体位点的菌群变化

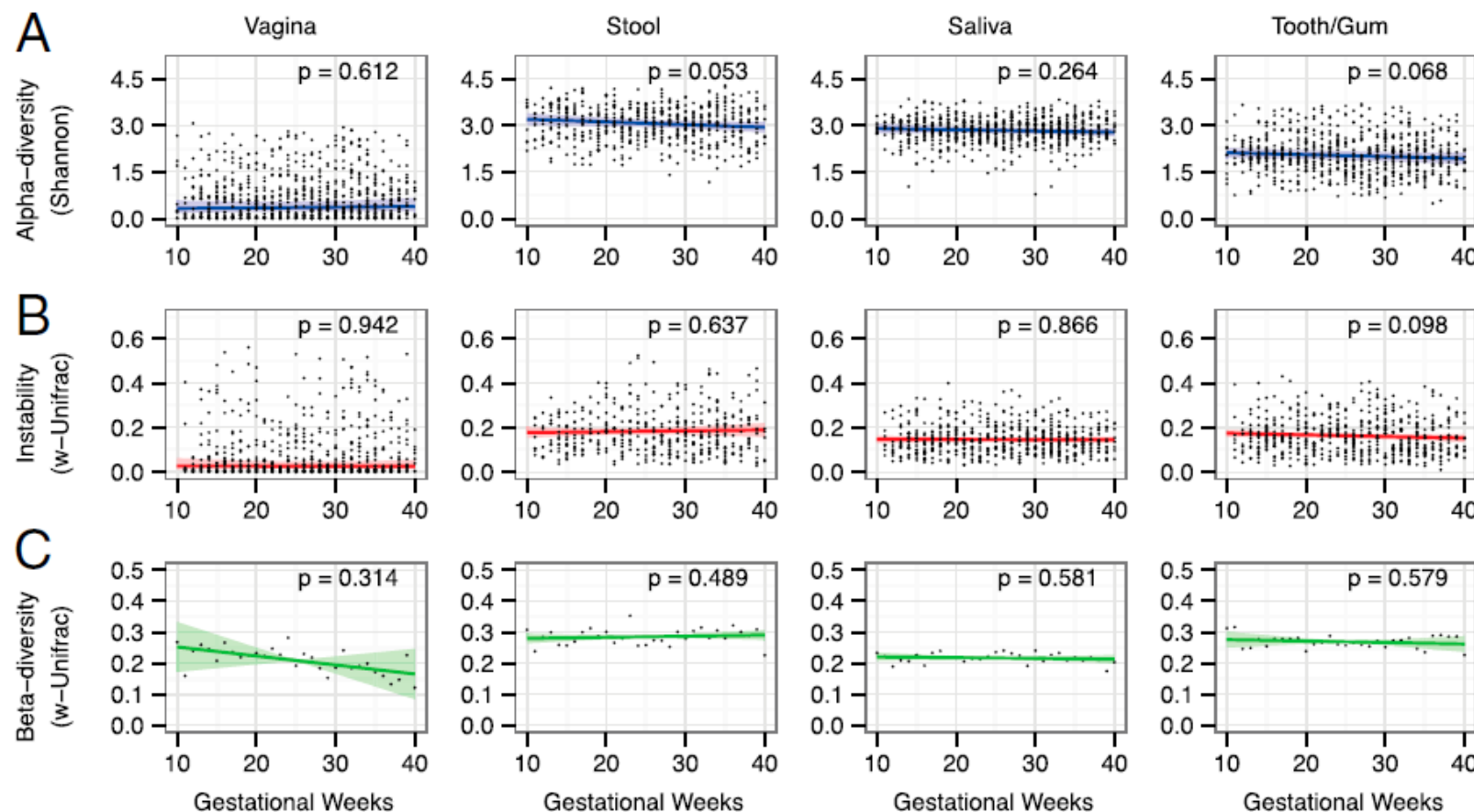
Temporal and spatial variation of the human microbiota during pregnancy

Daniel B. DiGiulio^{a,b,c,1}, Benjamin J. Callahan^{a,d,1}, Paul J. McMurdie^{a,d}, Elizabeth K. Costello^{a,e}, Deirdre J. Lyell^{a,f}, Anna Robaczewska^{a,b,c}, Christine L. Sun^{a,e}, Daniela S. A. Goltsman^{a,e}, Ronald J. Wong^{a,g}, Gary Shaw^{a,g}, David K. Stevenson^{a,g}, Susan P. Holmes^{a,d}, and David A. Relman^{a,b,c,e,2}

^aMarch of Dimes Prematurity Research Center, Stanford University School of Medicine, Stanford, CA 94305; ^bDepartment of Medicine, Stanford University School of Medicine, Stanford, CA 94305; ^cVeterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304; ^dDepartment of Statistics, Stanford University, Stanford, CA 94305; ^eDepartment of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305; ^fDepartment of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA 94305; and ^gDepartment of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305

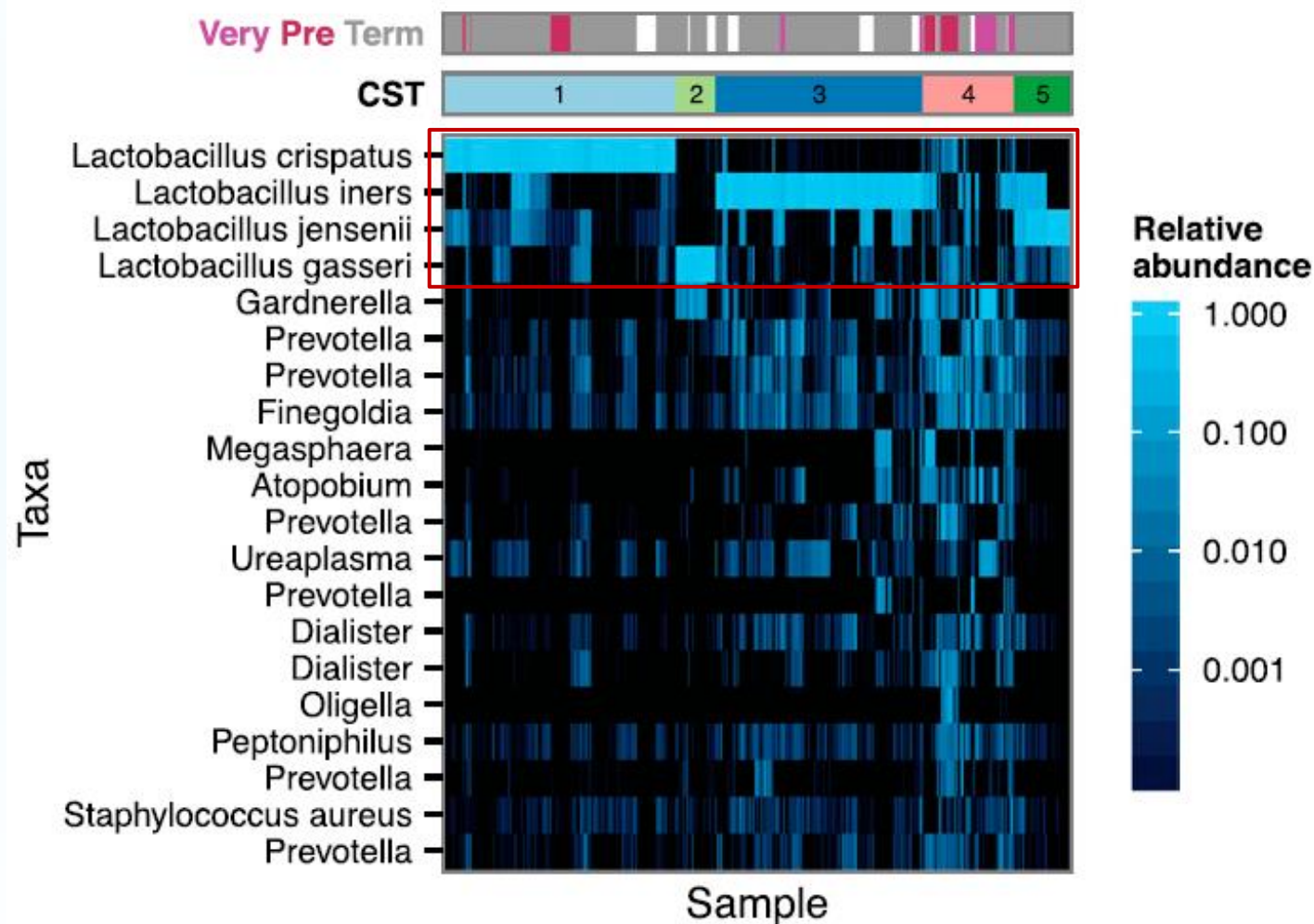
数据类型及时间梯度

- 收集了49名妇女在怀孕期间的阴道、肠道、唾液以及牙龈菌群时序数据
- 在怀孕期间，4个不同位点的菌群结构（ α 多样性、 β 多样性）基本保持稳定



生殖道菌群的聚类分型

- 对生殖道菌群进行聚类分析，在OTU水平上能够被划分为5个不同的群落结构类型（community state type, CST）
- CST1-3, 5被不同的乳酸杆菌主导
- CST4具有较高的菌群多样性



生殖道菌群状态类型转变

- 尽管个体的阴道菌群在整体上保持稳定，但存在不同CSTs之间的转换
- 使用马尔可夫链对群落间的转换进行描述，点的大小代表属于该群落的人群数量，线的粗细代表不同群落间的转移概率
- 4个乳酸杆菌主导的CSTs具有更高的自转移概率（self-transition probability），在孕期维持稳定

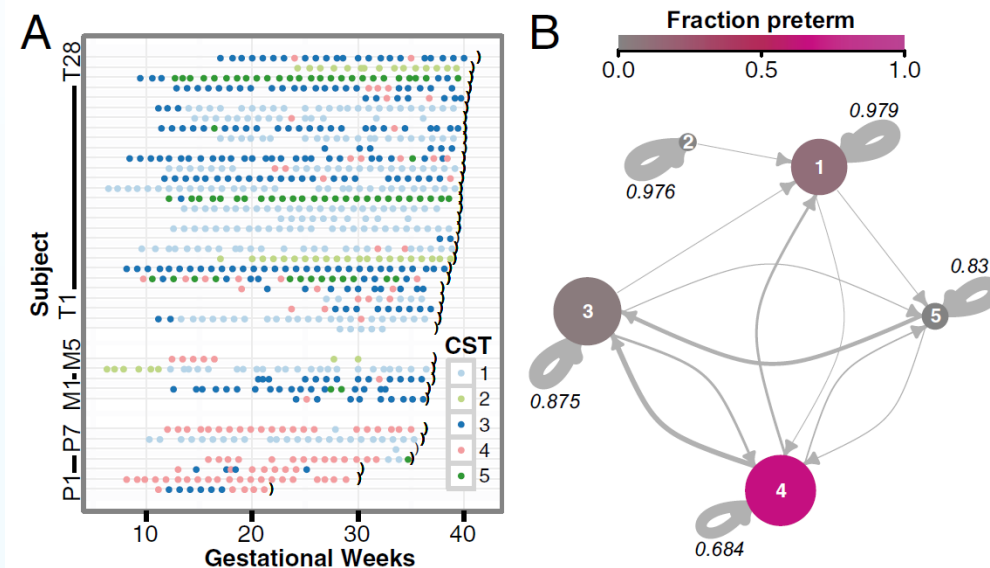
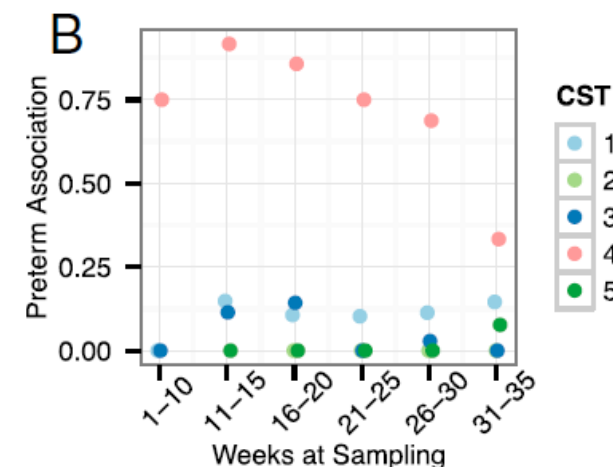
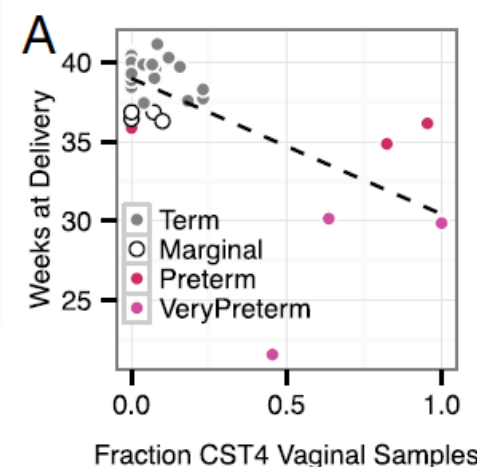
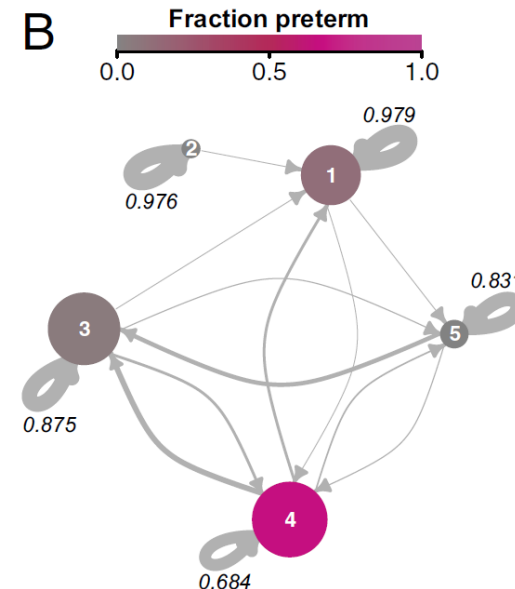
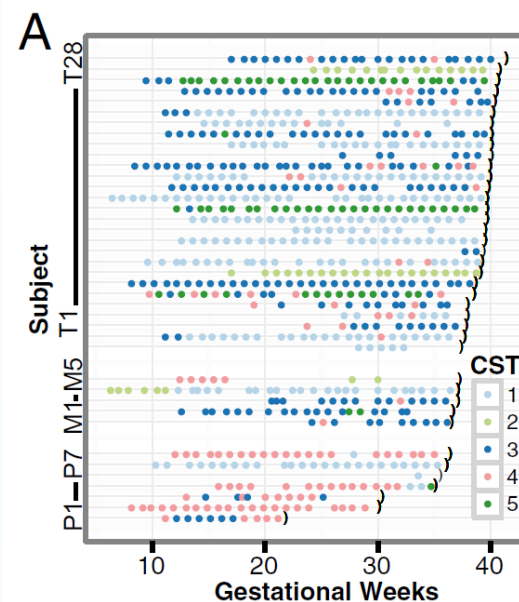


Fig. 3. Dynamics of the vaginal community during pregnancy. (A) Vaginal CST time course of the 40 subjects from the first subject group. Color indicates CST as shown in the key; the black parenthesis indicates delivery. Subjects P1-P7 delivered preterm (before gestational week 37); subjects M1-M5 were considered marginal (gestational week 37); subjects T1-T28 delivered at term (>37 gestational weeks). Marginal subjects were excluded when calculating associations with preterm birth. (B) Dynamics of the vaginal communities were approximated as a Markov chain with subject-independent transition probabilities between CSTs. Arrow weights are proportional to the maximum-likelihood-estimate of the week-to-week transition probabilities between states. Node sizes scale with the number of subjects in which the CST was seen. Color indicates the strength of the association with preterm birth (i.e., the proportion of the specimens from the CST that came from subjects who delivered preterm). The self-transition probabilities are shown numerically.

生殖道菌群状态类型转变

- 缺乏主导菌的CST4在孕期经历频繁转换，具有相对较低的自转移概率
- 在CST4停留时间较长的妇女，其胎儿的早产概率显著增加
- 在孕期的任何时间点转变为CST4都与后期的早产相关
- CST4的菌落特征可以作为胎儿早产的预测指标



皮肤菌群随时间的变化

例3：慢性伤口愈合过程中的菌群变化



Journal of Investigative
Dermatology

Volume 137, Issue 1, January 2017, Pages 237-244



Original Article

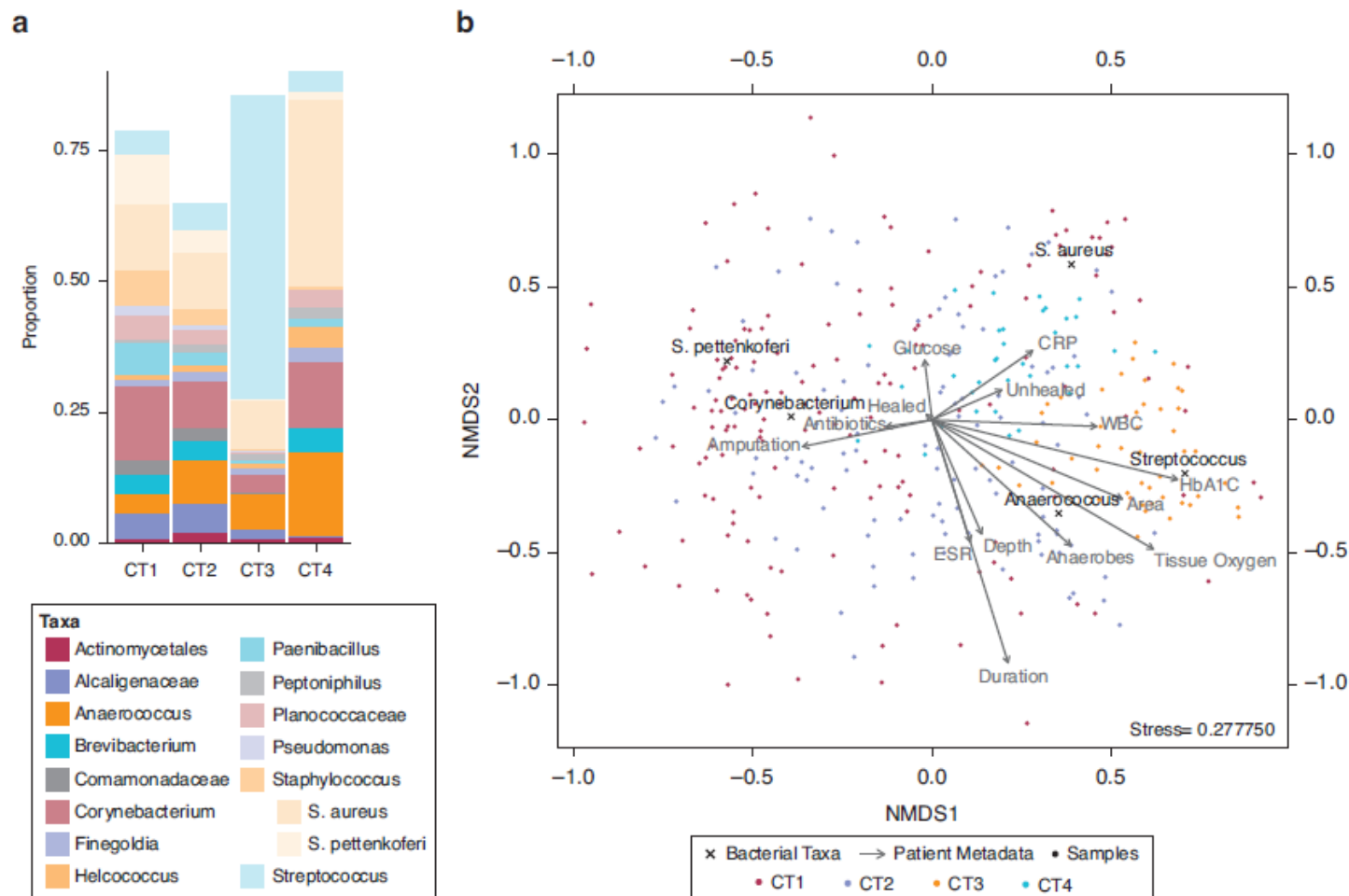
Wound Healing

Temporal Stability in Chronic Wound Microbiota Is Associated With Poor Healing

Michael Loesche^{1, 8}, Sue E. Gardner² ⁸ , Lindsay Kalan¹, Joseph Horwinski¹, Qi Zheng¹, Brendan P. Hodkinson¹, Amanda S. Tyldsley¹, Carrie L. Franciscus³, Stephen L. Hillis⁴, Samir Mehta⁵, David J. Margolis^{1, 6}, Elizabeth A. Grice^{1, 7} ⁷ 

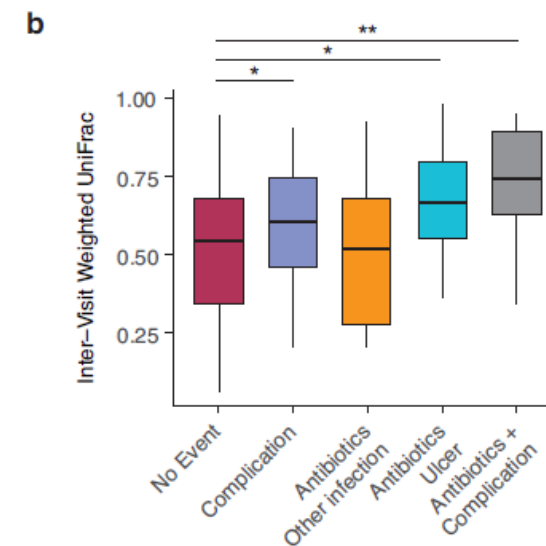
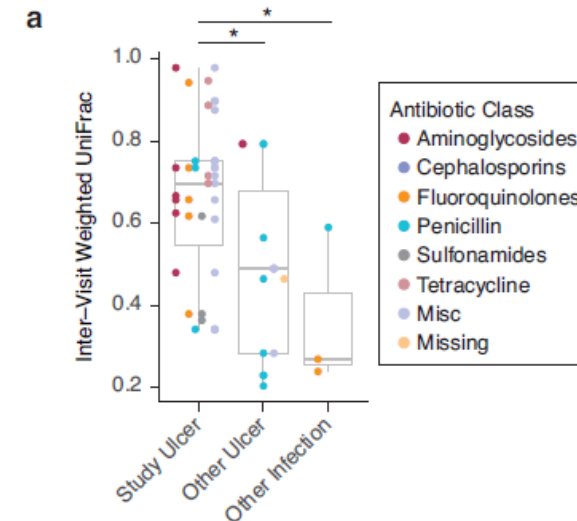
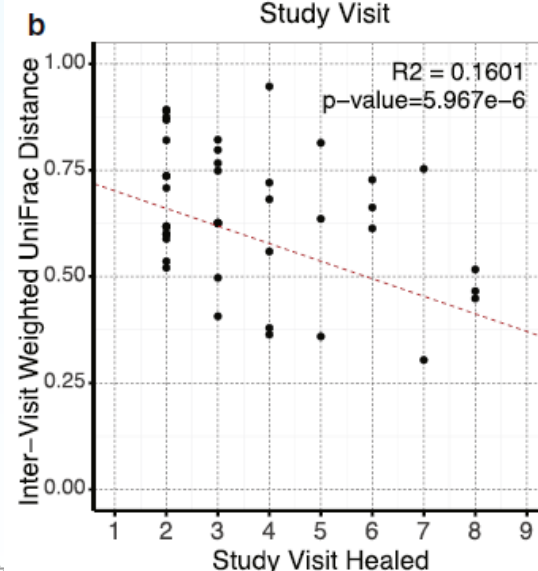
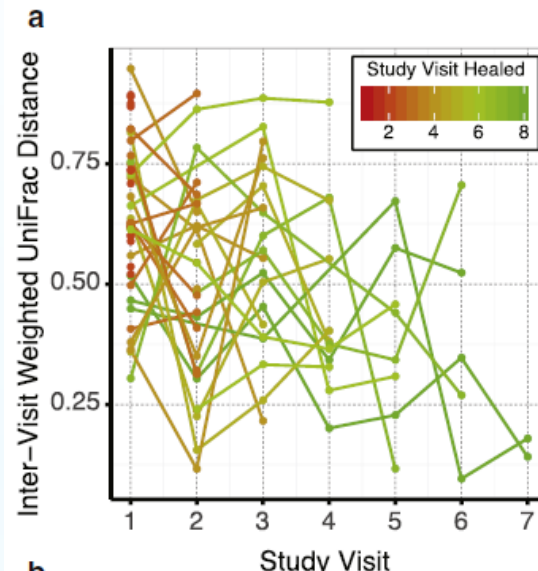
伤口愈合相关的菌群研究背景

- 微生物是导致伤口愈合缓慢最主要的原因之一
- 对糖尿病足溃疡 (diabetic foot ulcers, DFU) 病人进行微生物分析, 可聚集成4个具有显著差异的类群 (community types, CT)



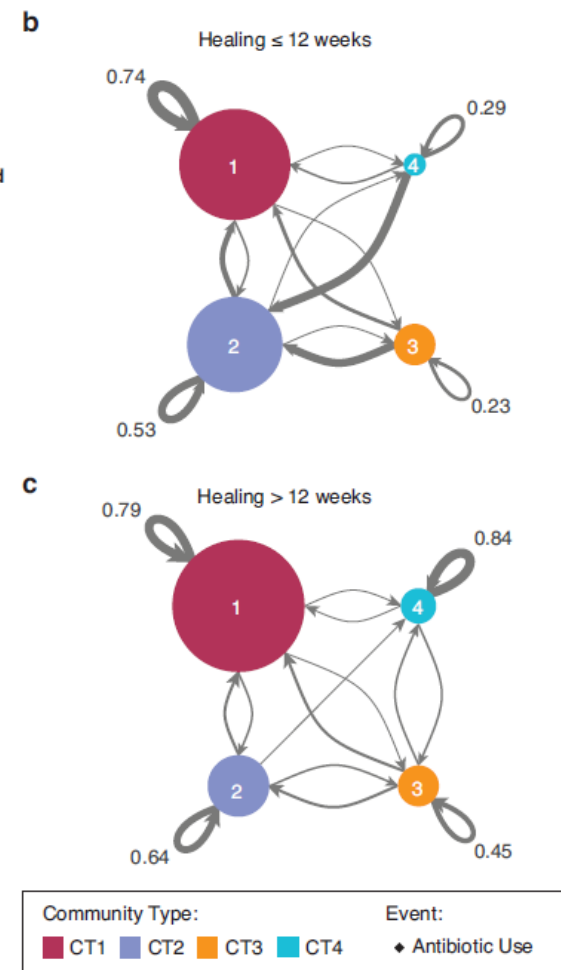
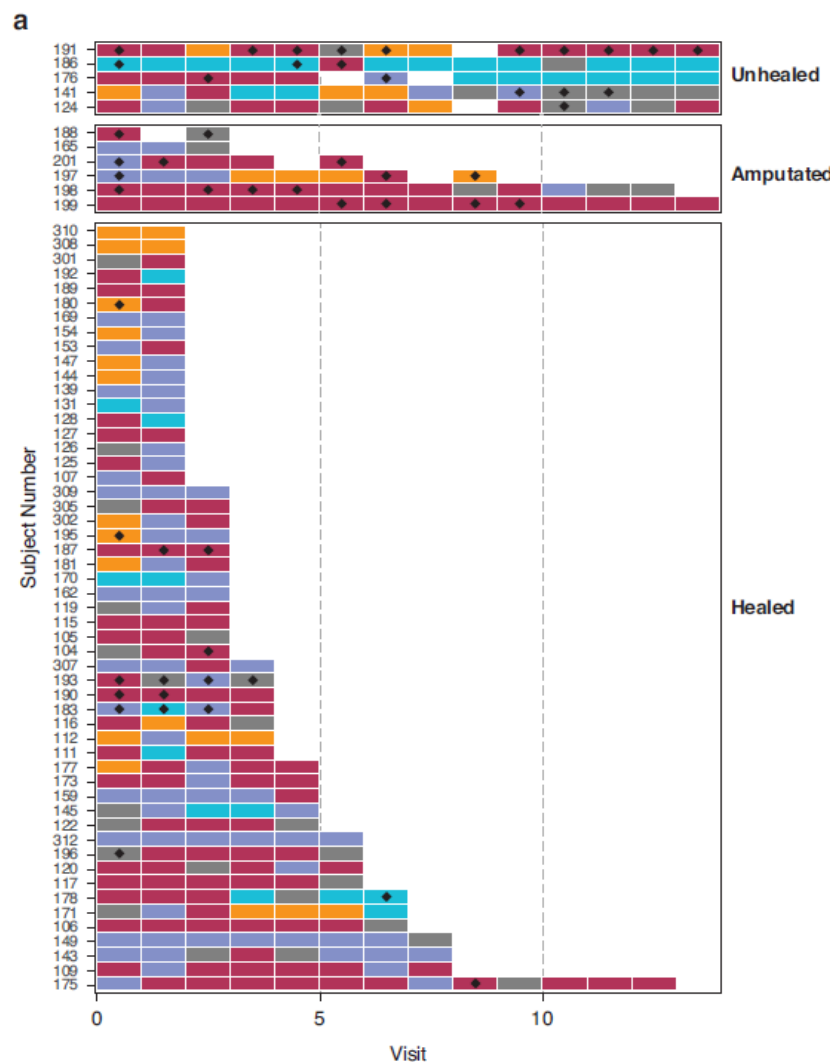
群落类型的稳定性与病程

- 用加权UniFrac距离 (weighted UniFrac distance) 衡量不同群落共有OTUs的比例、系统发生距离以及相对分布，其值越接近1，代表群落越不稳定
- 多数痊愈较快的病人（随访次数少）群落之间的UniFrac距离更大，表明其微生物经历更大的扰动
- 针对DFU的抗生素会更大破坏菌群的稳定性



群落类型的转化率与愈合效果

- 大多数病人在CT1和CT2停留时间最长，并且此两个类群具有最高的转化率和自我转化率，提示其可能与疾病联系更紧密
- 对于伤口愈合周期大于12周的病人，其在CT3和CT4的自我转化率显著上升，表明微生物的动态变化可能会加快伤口愈合



衡量皮肤菌群动态变化的意义

- 之前研究表明，身体其他位点的微生物菌群在多数时间处于**稳定状态**，微生物菌群的**非常规扰动**会导致疾病
- 在DFU研究中，微生物菌群的**动态变化反而促进伤口愈合**。这可能是因为DFU菌群多数属于致病菌，其本身就处于“病理”状态，因此菌群结构的稳定性反而会加重溃疡和感染的程度
- DFU病人**频繁变化的菌群**破坏了致病菌的生存环境，进而加快伤口愈合的速度



呼吸道菌群随时间的变化

例4：呼吸道菌群的差异是否与不同流感易感性相关

RESEARCH ARTICLE

The respiratory microbiome and susceptibility to influenza virus infection

Kyu Han Lee¹, Aubree Gordon¹, Kerby Shedden^{2,3}, Guillermina Kuan^{4,5}, Sophia Ng¹, Angel Balmaseda^{5,6}, Betsy Foxman^{1*}

1 Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, United States of America, 2 Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan, United States of America, 3 Department of Statistics, University of Michigan, Ann Arbor, Michigan, United States of America, 4 Centro de Salud Sócrates Flores Vivas, Ministry of Health, Managua, Nicaragua, 5 Sustainable Sciences Institute, Managua, Nicaragua, 6 National Virology Laboratory, Centro Nacional de Diagnóstico y Referencia, Ministry of Health, Managua, Nicaragua

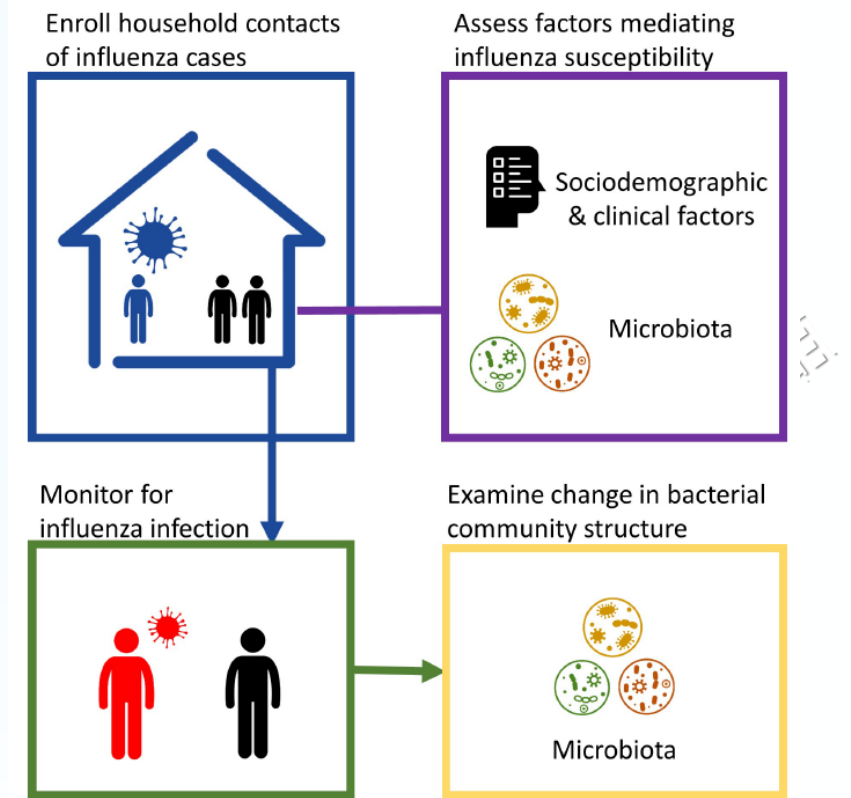
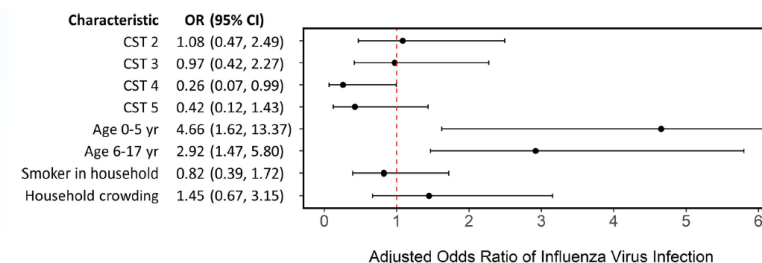
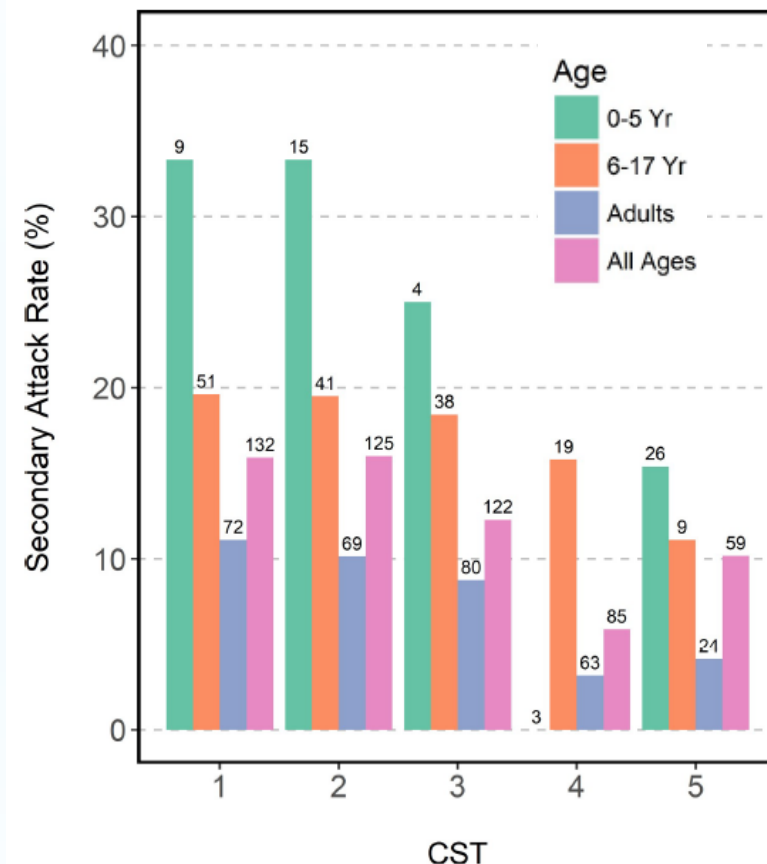


Fig 1. Graphical abstract.

Plos One, 2019

数据类别及群落状态分型

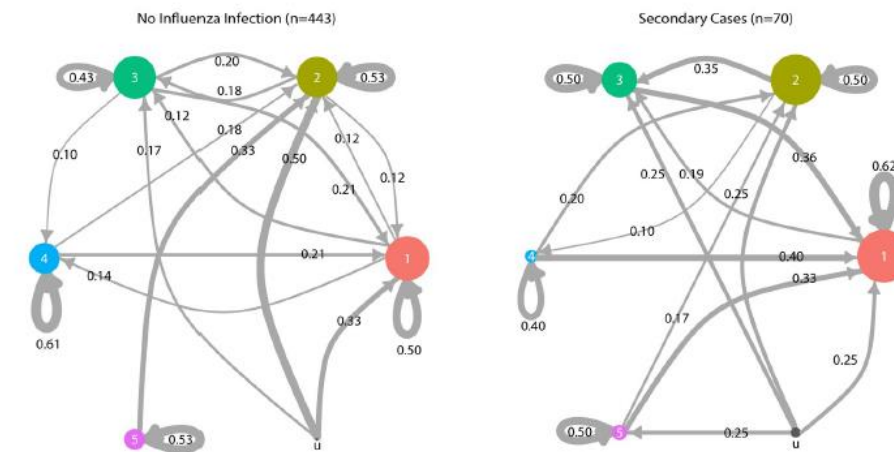
- 分别对鼻腔和咽喉的微生物取样
- 鼻咽部微生物分为5个有差异的群落结构 (community state type, CST), 且与不同年龄组成和不同流感易感性显著相关
- CST4具有较低的流感易感性, 而家庭中的年幼儿童 (0-5岁) 具有较高的概率感染流感



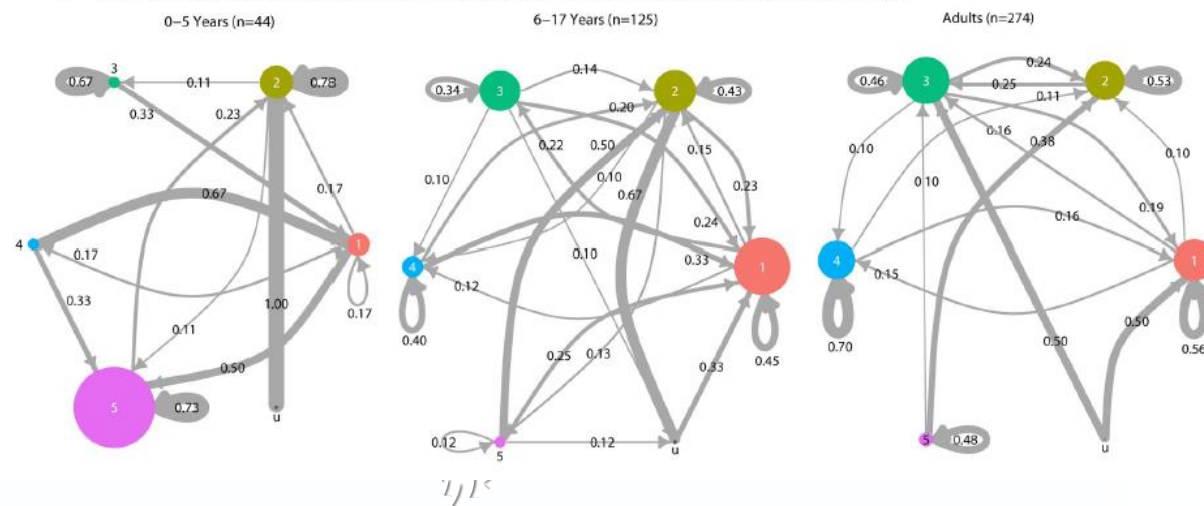
病毒感染后的呼吸道菌群转变

- 为探究流感病毒感染过程中呼吸菌群的稳定性，使用马尔可夫链对不同细菌群落结构的变化进行特征分析
- 不同群落之间出现频繁转换，大约一半人群在与家庭成员接触之后其群落类型出现转变，但与流感易感性无关

A Among all household contacts, by influenza case status



B Among household contacts with no influenza virus infection during follow up, by age group



肠道菌群稳定性的干扰因素

例5：不同影响因素下微生物群落的稳定性

MDPbiome: microbiome engineering through prescriptive perturbations

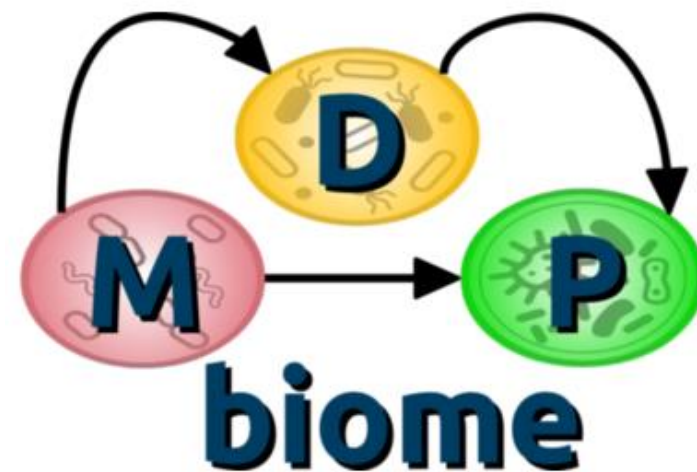
Beatriz García-Jiménez^{1,†}, Tomás de la Rosa² and Mark D. Wilkinson^{1,*}

¹Center for Plant Biotechnology and Genomics UPM - INIA, Universidad Politécnica de Madrid, Madrid 28223, Spain and ²Department of Computer Science, Universidad Carlos III de Madrid, Madrid 28911, Spain

*To whom correspondence should be addressed.

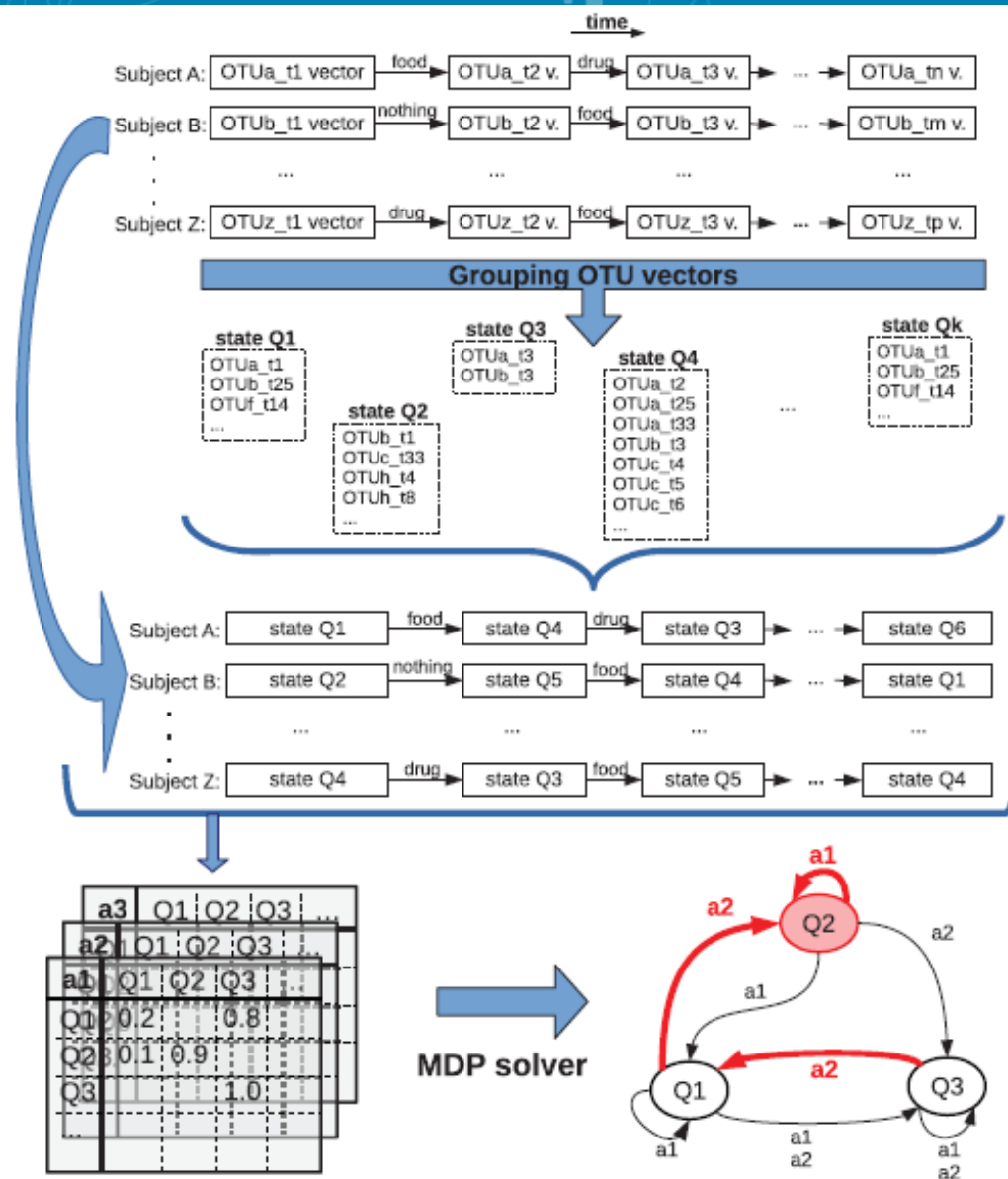
[†]Present address: Department of Systems Biology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CNB-CSIC), Madrid 28049, Spain

MDPbiome



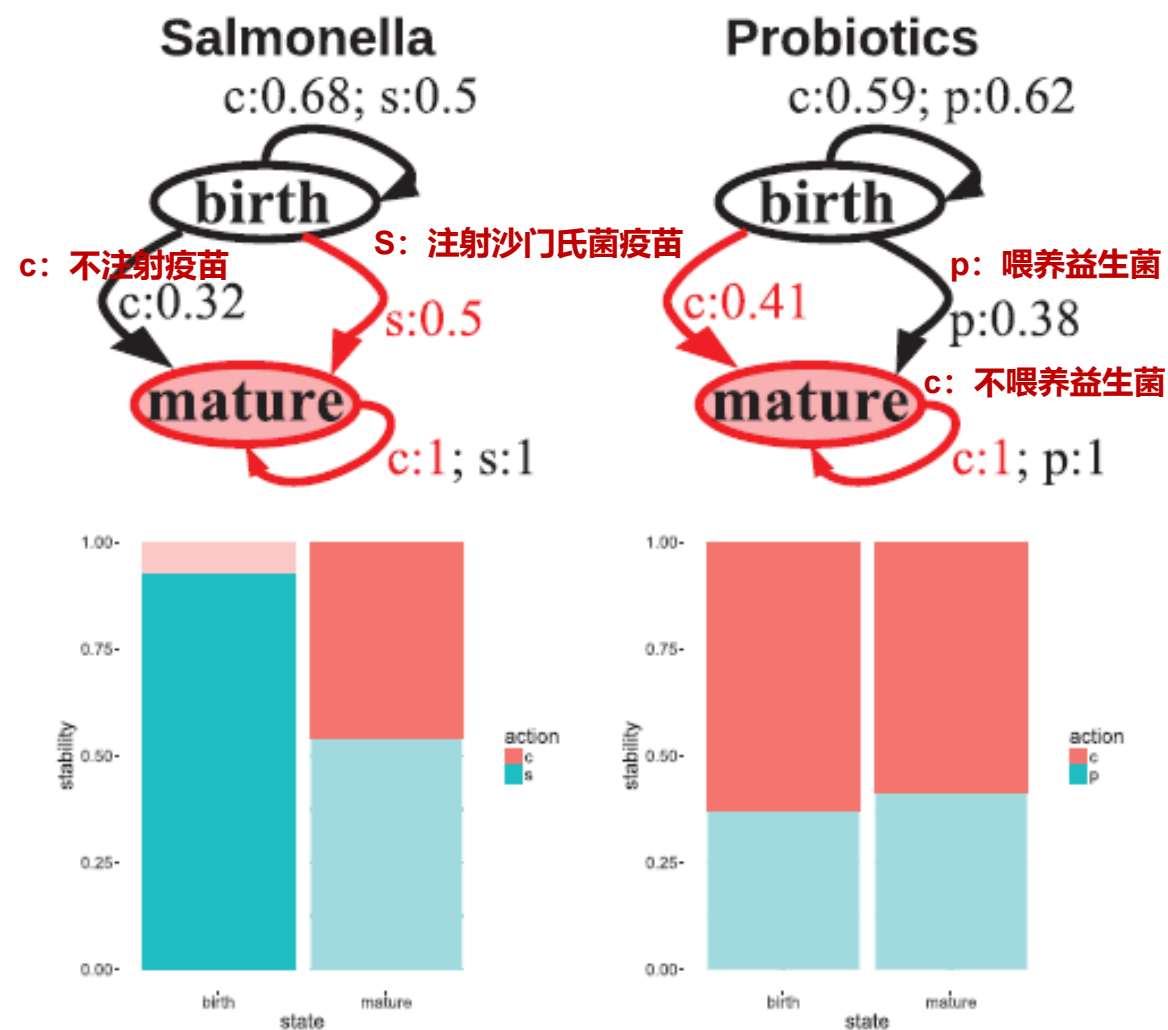
基于马尔可夫模型的工具

- 基于**马尔可夫决策过程** (Markov Decision Process, MDP) 开发了一款统计工具——MDPbiome
- 输入初始微生物菌群组成，并设定目标状态，MDPbiome能够给出达到目标状态的**最优扰动策略** (optimal perturbation policy)，从而指导医疗或工业发展



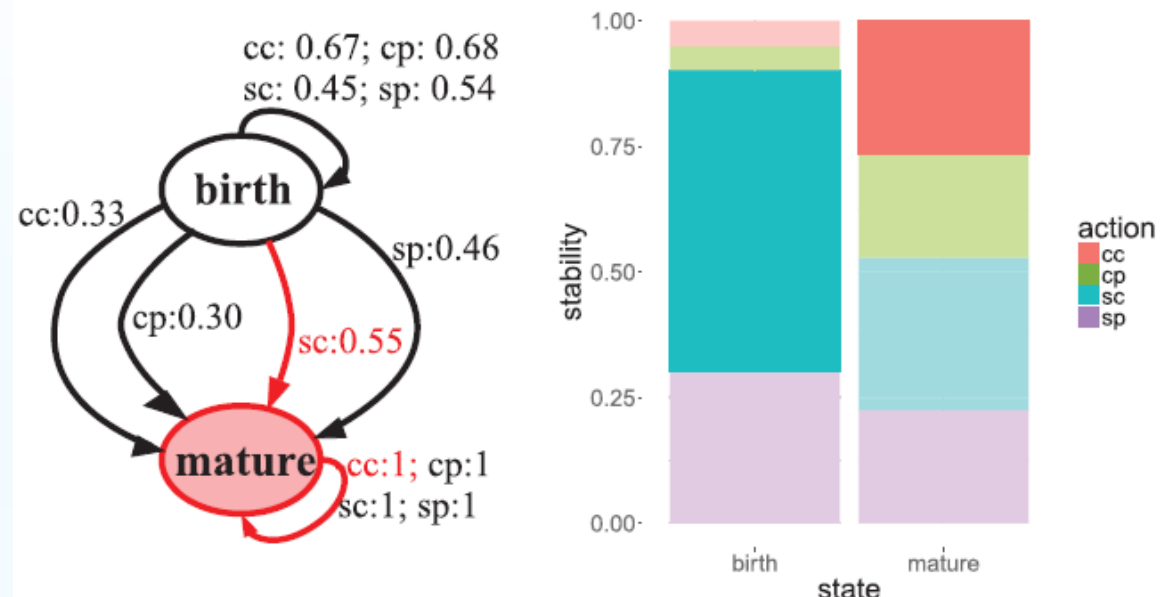
鸡肠道菌群成熟的影响因素

- 根据出生年龄将鸡分为birth（0-3天）和mature（14-28天）两个状态，后者菌群具有更高的多样性
- 红色箭头和条形图中更深的颜色代表最优决策。将多样性更高的成熟状态设置为目标状态
- 使用沙门氏菌疫苗注射会加速鸡肠道菌群的成熟，不推荐使用益生菌喂养



鸡肠道菌群成熟的影响因素

- 对是否注射沙门氏菌疫苗以及是否喂养益生菌进行协同分析
- 将多样性更高的成熟状态设置为目标状态
- MDPbiome判定**注射沙门氏菌疫苗同时不喂养益生菌**为从出生转变为成熟的最优决策
- 达到成熟状态后，**不注射疫苗并且不喂养益生菌**是最优决策



cc: 不注射疫苗并且不喂养益生菌

sc: 注射沙门氏菌疫苗但不喂养益生菌

cp: 不注射疫苗但是喂养益生菌

sp: 既注射沙门氏菌疫苗又喂养益生菌

何时适合使用马尔可夫链：

- 1.时序数据
- 2.个体之间**存在较大异质性**，需要将不同个体聚集为不同的状态
- 3.个体在不同时间经历**频繁的状态转变**
- 4.在整体层面**无法观测到显著的改变**，需要划分为更小单位



扫码关注生信宝典，学习更多生信知识



扫码关注宏基因组，获取专业学习资料

易生信，没有难学的生信知识