图书馆论坛
Library Tribune
ISSN 1002-1167,CN 44-1306/G2

《图书馆论坛》网络首发论文

题目：华人科学家的学术谱系特征：基于跨群体的计量比较分析
作者：侯剑华，付兰志
收稿日期：2026-04-30
网络首发日期：2026-07-27
引用格式：侯剑华，付兰志．华人科学家的学术谱系特征：基于跨群体的计量比较分析[J/OL]．图书馆论坛. https://link.cnki.net/urlid/44.1306.G2.20260727.1044.002

网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

网络首发时间：2026-07-27 11:10:23　网络首发地址：https://link.cnki.net/urlid/44.1306.G2.20260727.1044.002

华人科学家的学术谱系特征：基于跨群体的计量比较分析

侯剑华，付兰志

摘　要　学术谱系为考察科研人才培养与知识代际传承提供了结构化视角。文章整合 Elsevier 科学家数据库、学术家谱数据库与 Microsoft Academic Graph，构建华人科学家和非华人科学家的对照样本，运用六项谱系指标、非参数检验和差异化计数回归模型开展比较分析。结果显示，华人科学家的六项谱系指标均低于非华人科学家，其中繁殖力、后代规模与世代深度的群体差异较为突出，且在控制科研产出和学科分布后仍然存在；差距主要体现于具有谱系记录者的规模积累阶段，而非是否形成相关谱系特征；学术生育力计数阶段未呈现显著群体差异。境内外华人的多数指标虽达到统计显著水平，但效应量较小且中位数基本一致。文章揭示了华人科学家学术谱系的规模与代际结构特征，并为跨群体人才传承研究提供了可重复的计量方法。

关键词　华人科学家；学术谱系；跨群体比较；科学计量学

引用本文格式　侯剑华，付兰志. 华人科学家的学术谱系特征：基于跨群体的计量比较分析[J].图书馆论坛，2026

Academic Genealogy Characteristics of Chinese-Heritage Scientists: A Cross-Group Scientometric Analysis

Hou Jianhua, Fu Lanzhi

Abstract: Academic genealogy provides a distinctive quantitative lens for studying mentorship and talent cultivation in science; yet systematic cross-group genealogy comparisons involving Chinese-heritage scientists remain absent from the literature. This study integrates three large-scale databases—the Elsevier Scientist Database (ESD), Academic Family Tree (AFT), and Microsoft Academic Graph (MAG)—to construct a global comparative sample comprising 6,690 Chinese-heritage scientists and 56,417 non-Chinese-heritage scientists. Using six genealogy indicators (fecundity, fertility, descendants, cousins, generations, and relationships) alongside nonparametric tests and differentiated count regression models (Negative Binomial, Zero-Inflated Negative Binomial, and Hurdle Negative Binomial), we systematically compare the talent cultivation characteristics of the two groups. Results show that Chinese-heritage scientists score significantly lower on all six indicators, with medium-level effect sizes for core scale metrics. These differences remain highly significant after controlling for individual research output (h-index) and disciplinary distribution. Regression analysis reveals an asymmetric pattern: Chinese-heritage scientists are not less likely to assume mentorship roles (Hurdle zero-stage OR < 1), yet their genealogy scale is substantially smaller once the threshold is crossed. The fertility indicator at the count stage is consistently non-significant across all model specifications, suggesting that the genealogy gap reflects structural and institutional constraints rather than differences in mentoring quality. This study establishes a replicable cross-group genealogy comparison framework and contributes new scientometric evidence for understanding the academic lineage patterns of Chinese-heritage scientists globally.

Keywords: Chinese-heritage scientists; academic genealogy; cross-group comparison; scientometrics

0 引言

随着科研全球化进程的深入推进，科学家的跨国联系与合作日趋密切，个体科研人员已不再孤立地嵌于单一国家科研体系之中，而是被置于跨机构、跨国界的合作与传承网络之内。在这一背景下，华人科学家群体的崛起尤为引人注目。从基础科学到计算机与信息科学、工程技术，华人科学家在国际高被引学者、重要奖项获得者中的比例持续

上升[1]。本文的“华人科学家”是指在公开可检索记录中同时呈现华语文化背景特征的科研人员，而非族裔统计或法律意义上的身份认定。华人科学家不仅在知识生产方面作出了突出贡献，也通过指导学生、建设团队、参与跨国合作，在全球学术共同体的人才培养中扮演了关键角色[2]。然而，相较于关注华人科学家科研产出、论文影响力和流动路径的丰富研究，对其“如何培养人才、谱系结构有何特征”的系统性计量分析十分有限。

学术谱系（academic genealogy）为观察人才培养与知识传承提供了一种独特且结构化的视角。以导师–学生关系为基础的谱系网络，能够揭示科学家在“学术后代”规模、代际延续深度、谱系分支结构等方面的差异，并可进一步通过人才培养指标刻画科研人才培养模式。近年来，基于数学、物理、经济学等学科的学术谱系数据库与网络分析研究逐渐增多，相关工作发现导师的学术繁殖力与学生学术成就、学科发展路径等存在紧密联系。然而，已有研究具有 3 个明显局限：其一，多聚焦于单一学科或少数“名家”，缺乏大规模、跨学科、跨群体的系统比较；其二，研究对象多为欧美科学家，对华人等特定文化群体的谱系结构关注不足；其三，涉及华人科学家，也多停留在履历分析或案例研究层面，缺乏与学术谱系网络数据的深度融合。

因此，本研究整合 Elsevier 科学家数据库、学术家谱数据库与 Microsoft Academic Graph，构建华人科学家与非华人科学家的全球对照样本，重点回答 3 个问题：第一，华人科学家的学术谱系具有怎样的规模与代际结构；第二，与非华人科学家相比，其谱系特征是否存在稳定差异；第三，该差异在科研影响力、学科类型、任职地域和网络结构等不同条件下是否一致。与已有研究主要关注科研产出、国际流动和合作网络不同，学术谱系揭示的是科学知识在人际关系中的代际传递过程。相比论文、专利等成果指标，谱系指标能够更加直接反映科学家的长期人才培养贡献。因此，对华人与非华人科学家开展大规模谱系比较，不仅能够揭示不同科研文化背景下人才传承结构的差异，也能够推动科学计量学对知识传承与人才培养评价研究，为理解科学共同体形成机制提供新的分析视角。

1 文献综述

近年来，学术谱系与人才培养的研究逐渐成为揭示学科演化、知识传递与科研群体特征的重要议题。本研究主要从学术谱系与人才培养、科研人才培养模式、华人科学家的成长与培养特征等几个方面对已有研究进行述评。

1.1 学术谱系与人才培养研究

学术谱系以导师－学生关系为核心线索，揭示知识在代际间的传承轨迹与人才培养的结构特征。与合著网络或引文关系不同，谱系网络刻画的是培养关系而非协作关系，着眼于科学家作为导师在科研人才生成与学术共同体构建中的实质性角色。

国际学术谱系研究发端于精英科学家群体的传统。朱克曼在《科学界的精英——美国的诺贝尔奖获得者》中梳理了诺贝尔奖得主的师承脉络，奠定了早期范式[3]；随后多位科学家对生物[4]、物理[5]、经济学[6]等领域的学术谱系进行梳理。我国学者对学术谱系的早期研究集中于哲学、社会学、情报学等领域[7]。随后有学者对日本的诺贝尔物理学奖得主的学术谱系进行研究，探讨了其对我国科学发展的借鉴意义[8]。2011 年，王春法提出系统研究当代中国科学家学术谱系的倡议，标志着这一领域规模化发展的起点[9]，此后逐步扩展至水稻科学、物理学、力学等学科，研究视角也从个案追溯延伸至谱系结构的规律性刻画[10]。

随着 AFT 等大型谱系数据库的建立，研究重心从谁是谁的学生转向谱系结构的整体性分析与跨群体比较。在这一转变中，学术谱系的定义也逐步从文化传承意义上的师承叙事，收敛为可量化操作的导师－学生有向关系网络[11-12]。

1.2 导师－学生关系的量化研究

随着学术谱系数据库的兴起，将导师－学生关系转化为可量化指标，是近年来科学计量学的重要进展。Rossi 等系统提出繁殖力（Fecundity）、生育力（Fertility）、后代规模（Descendants）等 6 项谱系指标，将导师在人才培养上的数量和规模特征纳入可比较的度量框架[13]，为跨学科、跨群体的结构性比较提供了方法基础。后续研究显示，这些指标与导师的学术声誉、机构资源及学科组织模式之间存在系统性关联。博士生培养体系相对成熟的学科中，高被引学者往往具有更大的繁殖力与更深的世代深度[14]，而后代规模的大小则在一定程度上独立于导师自身的科研产出水平[15]。值得注意的是，Xing 等发现学生群体规模对其学术成长的影响具有非线性特征，大规模指导团队有助于提升学术成就，小规模团队则提升学术存活率[16]，这提示导师规模策略与谱系扩展逻辑之间存在复杂关联，尚待在大规模群体比较中加以检验。此外，马杰等对国内研究生师生关系的调查揭示了中国情境下导师制度的特殊性[17]，为后续分析华人科学家谱系特征时引入制度视角提供了依据。

1.3 华人科学家的学术成长与谱系特征研究

现有关于华人科学家的研究主要关注教育经历、国际流动和职业发展等外部成长因素，但较少关注这些因素如何进一步影响其学术谱系结构。相关研究主要从教育背景、师承关系和职业发展 3 个方面揭示华人科学家的成长特征。

第一，华人科学家的教育经历呈现本土基础教育加国际高等教育的复合型特征。国内顶尖院校的本科培养奠定知识基础，博士及博士后阶段则多在欧美完成，美国的正向影响最为突出，且全职境外经历的效益显著优于兼职经历[18]。博士学位国别对华人早期职业成长具有可观测的影响，拥有全职境外经历者的平均成才时间较无境外经历者缩短约 0.6 年[19]。近年来，海归科学家的规模与质量均呈明显增长，学术流动不仅影响华人科学家的成长轨迹，也深刻塑造其知识网络的国际化结构[20-22]。

第二，高成就导师的引领效应显著。以院士为导师的华人学者成才时间（均值 6.83 年）短于其他导师类型，且更多拥有全职境外学习经历[19]。名师资源与顶尖院校学位的叠加是拔尖人才成长的重要结构性机制[23]。本土与海归华人学者的师承网络存在可观察的差异，本土学者的学术风格与学缘关系受第一代学者影响更深，海归学者则形成以西方华裔或对华友好学者为核心的跨国学缘网络[24]。

第三，华人科学家的职业路径日趋多元，企业成为顶尖华人科技人才的重要流向[25]。海归与本土华人学者在研究范式上存在结构性互补，前者偏重实证研究与微观议题，后者更注重宏观与思辨研究[24, 26]。然而，上述研究尚未回答不同成长背景是否会进一步塑造科学家的导师网络规模和代际结构，这正是本文试图从学术谱系视角进行分析的问题。

1.4 研究述评

现有文献在学术谱系、科研人才培养与华人科学家研究方面已经提供了重要基础，但仍存在深入探索空间：第一，学术谱系研究多聚焦于单一学科或少数精英科学家，缺乏以大规模科学家群体为单位、面向特定文化或族群的系统谱系刻画；第二，关于导师制、师承关系与人才成长路径的研究多局限于制度比较与个体履历层面，很少将导师－学生关系网络和学术谱系结构作为核心分析对象，从而难以全面揭示人才培养在代际与网络层面上的结构特征；第三，关于华人科学家的研究，更多聚焦教育背景、国际流动、

科研产出与职业发展等维度，缺乏基于学术谱系网络的量化刻画，缺乏将华人科学家与非华人科学家进行系统的谱系特征比较研究。

2 数据与方法

本研究以大规模科学家计量数据库与学术谱系数据库为基础，通过多源整合构建全球华人科学家与非华人科学家的对照样本，并综合运用描述性比较、非参数检验、计数回归与网络分析进行系统比较。

2.1 数据来源

2.1.1 科学家基准数据库

采用 Elsevier Scientist Database（以下简称 ESD）获取全球科学家的基础信息，包括：姓名、所属机构、所在国家/地区；学科分类（一级学科）；职业生涯起始年份；科研产出与引文指标，如总发文量、总被引次数、h 指数。该数据库为识别全球科学家总体、区分华人/非华人群体以及构建科研产出对比提供了基础[27]。

2.1.2 学术谱系数据库

采用 Academic Family Tree（AFT）获取科学家的导师与学生关系信息。AFT 记录了不同学科领域科学家之间的学术指导关系，包括博士阶段、部分硕士阶段以及某些长期合作关系。基于 AFT 可构建以科学家为节点、以导师–学生为有向边的学术谱系网络，为计算繁殖力、后代规模、代际深度等学术谱系特征指标提供基础[28]。

2.1.3 科研产出与学科信息数据库

采用 Microsoft Academic Graph（MAG）补充科学家的论文与引文信息，包括总被引次数及不同引文论文数量、细分学科分类等。通过官方提供的 ID 映射代码，将 AFT 中的科学家节点与 MAG 中的作者 ID 进行匹配，获得更为完整的科研产出与引文指标[29]。

表 1 数据来源概况

数据库名称	时间范围	版本信息
Elsevier Scientist Database	1788-2022	Version 6,2023-10
Academic Family Tree	end-2023	Version 2023-05
Microsoft Academic Graph	1800-2018	OAG v2,2018-07

2.2 华人科学家与对照组样本构建

2.2.1 华人科学家识别

“华人科学家”在概念层面涵盖族裔、文化认同、教育背景与公民身份等多重维度，而这些维度在具体个体上往往并不完全重合。在现有大规模科学家数据库中，亦无直接标注华人身份的字段可供使用。鉴于此，本文将华人科学家定位为一个可观测和可操作化的标签，而非族裔统计概念或个体身份判断。华人科学家的识别方法旨在构建适用于大规模计量分析的可操作性分类，不是对个体文化身份进行价值判断。

具体而言，本文所界定的华人科学家，指在公开可检索信息中同时呈现出华语文化背景特征的科研人员。主要特征线索包括：具有典型华人姓名特征（含标准汉语拼音结构及粤语、闽南语音译变体）、曾在华语地区（我国大陆、我国港澳台、新加坡等）完成教育经历，或曾在华语地区机构长期任职。然而，这一界定可能对华语背景但姓名高度英文化的科学家（如部分在他国出生的华裔后代）存在系统性漏判；对在华语地区出生但主要学术训练在他国完成的科学家，其在识别框架中的归类存在一定不确定性。因此，本文通过高、中、低 3 档置信度分层与稳健性分析，对识别误差的影响范围进行了部分量化；结论应理解为在当前识别框架下的群体比较，而非对严格定义下特定族裔群体的全量统计描述。具体步骤如下：

（1）姓名初筛。首先在 ESD 中基于姓名字段构建候选华人科学家集合。具体做法为：一是构建包含常见华人姓氏的列表（如 Chen、Zhang、Li、Wang、Liu），并补充常见复姓及其拼音/英文变体；二是对姓名中包含明显汉语拼音特征的个体（如“Xiaoming Li”“Wei Zhang”等）进行标注；三是对具有典型粤语、闽南语等音译特征的姓氏（如“Yip”“Ng”）进行扩展识别。根据上述规则，将满足至少一项条件的个体纳入“华人候选集合”；对无法明确归类且姓名形态与多族裔群体高度重合的个体暂不纳入候选集合，以减少误判风险。

（2）国别与机构校验。在姓名初筛基础上，进一步结合科学家的教育与工作信息进行校验。对于本科或博士就读机构位于我国大陆、我国港澳台或新加坡等华语区，且姓名具有华人特征的个体，认定为华人科学家；对于姓名具有华人特征，但教育及长期任职机构主要位于非华语地区，且缺乏其他可识别的华人背景线索的个体，则予以谨慎处理，对存在较大不确定性的个体予以剔除。

（3）高置信度子样本构建。为进一步控制识别误差，本研究在上述华人科学家样本之上，额外定义了一个高置信度华人科学家子样本。该子样本需同时满足以下条件中的至少两项：一是姓名字段为标准中文姓名或典型汉语拼音姓名；二是本科或博士阶段至少一个学位在华语地区高校或科研院所获得；三是职业生涯中至少有一段在华语地区长期任职经历（如高校、科研机构）。满足这些条件的科研人员被视为“华人身份高度确定”的个体，将在稳健性分析中单独作为子样本进行对比。

（4）人工抽查与识别准确率评估。为评估上述识别流程的有效性，本研究从初步识别得到的华人科学家样本中，采用分层随机抽样的方式，引入基于多源线索的一致性检验与置信度分层。按学科与所在国家/地区，抽取 120 名个体进行人工检索与核实（包括查阅个人主页、机构简介及第三方资料等）。根据人工核实结果，计算姓名规则识别的正确率（precision）与漏判率（recall），并据此对整体样本的识别误差进行粗略估计。在误差评估方面，本研究采用双向分层抽查而非单向抽查，以同时估计识别的误判风险。根据上述识别流程，将科学家划分为华人科学家与非华人科学家两类群体。后续分析均在匹配成功且具备完整谱系与科研指标的样本上展开。

2.2.2 样本筛选与可比性控制

为保证华人科学家与非华人科学家之间的基本可比性，本研究在构建样本时进行限定与筛选。第一，时间窗口控制。限定科学家职业生涯起始年份在某一合理区间内（如 1970–2010 年），剔除过早或过新的极端个案，以减少代际长度与数字化记录覆盖差异对谱系指标的影响。第二，学科范围控制。选取 ESD 中数据覆盖较为完整的主要学科（如化学、物理、数学、计算机科学、神经科学），剔除样本过少或谱系信息严重缺失的边缘学科。第三，数据完整性要求。仅保留在 AFT 中有明确导师—学生关系记录，且在 MAG/ESD 中具备基本科研产出与引文信息的科学家；对于谱系关系严重缺失或科研指标完全缺失的个体予以剔除。经过多轮清洗与匹配，最终获得有效华人科学家样本 6,690 人，非华人科学家 56,417 人，在此基础上进一步区分整体样本与具有谱系记录的子样本。后文在呈现实证结果时，将分别报告两类群体的样本规模，以便读者判断统计结果的可靠性。

2.3 学术谱系构建与学术谱系特征指标

2.3.1 导师与学生的有向网络构建

在 AFT 数据库中，每一条导师和学生记录可视为一条从导师指向学生的有向边。基于与 ESD 匹配成功的科学家，本研究构建学术谱系有向网络。在网络构建过程中，对于

一名科学家在 AFT 中对应多个可能节点或出现一对多匹配的情况，本研究优先依据姓名、学科领域和机构信息进行综合判断，选择一致性最高的节点；对无法可靠匹配的个体予以剔除，以保证谱系网络的准确性。在此基础上，可以将所有科学家及其导师与学生关系视为由若干连通分支构成的有向图。

2.3.2 学术谱系特征指标定义

为系统刻画科学家的学术谱系与学术谱系特征，本研究在谱系网络基础上构建 6 个核心指标（表 2）[13]。指标设计兼顾数量性特征与结构性特征。本文将 Fecundity 译为繁殖力、Fertility 译为生育力、Descendants 译为后代规模、Cousins 译为表亲数量、Generations 译为世代深度、Relationships 译为关系总数，与 Rossi 等的指标体系保持一致，旨在刻画导师在学术人才培养上的数量和规模特征[13]。

表 2 科学家学术谱系特征指标定义与计算方法

指标	英文名	指标定义（基于有向图）
繁殖力	Fecundity	科学家 i 的直接学生总数。
生育力	Fertility	科学家 i 的直接学生中，后续成为导师的人数。
后代规模	Descendants	从科学家 i 出发沿有向路径可达的后代节点总数（不含自身），含学生、学术孙辈等。
表亲数量	Cousins	与科学家 i 共享同一上游导师、直接导师不同的同辈节点数，反映同门同代网络规模。
世代深度	Generations	以科学家 i 为起点向下延伸的最长代际层数，反映谱系纵向延续深度。
关系总数	Relationships	以科学家 i 为根的子谱系中，导师—学生关系边总数，体现其传承中枢地位。

2.3.3 科研产出与引文指标

为辅助理解学术谱系特征与科学家科研影响力之间的关系，并在分组比较中控制科研表现差异，从 ESD 数据库引入两类科研计量指标：一是总引用次数，使用 1996—2022 年总引用次数，用于划分引文前 10%与后 10%样本；二是 h 指数，采集截至 2022 年底的 h 指数，用于回归模型中的科研产出控制变量。华人与非华人科学家在科研影响力分布上存在差异，后续分析分别通过引文分层和在回归模型中控制 h 指数降低其干扰。

2.4 统计分析方法

2.4.1 描述性统计分析

首先对华人科学家与非华人科学家在各项谱系指标（繁殖力、生育力、后代规模、表亲数量、世代深度、关系总数）上的分布特征进行描述性统计，包括均值、中位数、标准差与分位数等。同时，对两类群体的学科分布、科研产出与引文指标进行对比，为理解谱系差异提供背景信息。效应量采用 Cliff's δ 衡量，并依据其绝对值划分为可忽略（<0.147）、小（0.147～0.329）、中（0.330～0.473）和大（≥0.474）4 个等级。

2.4.2 组间差异的非参数检验

考虑到大部分谱系指标为非负计数变量，分布明显右偏且不符合正态分布，本研究采用 Mann–Whitney U 检验等非参数方法，对华人科学家与非华人科学家在各指标上的分布差异进行统计检验。具体做法是在总体样本以及按学科、科研影响力（如引文前 10%与后 10%）划分的子样本中，分别比较两类科学家的谱系指标分布，判断其差异是否具有统计显著性。

2.4.3 网络结构特征分析

在个体层面谱系指标外，本研究还对华人子网络与非华人子网络的整体结构特征进行比较，主要考察网络平均度与加权平均度、网络直径与平均路径长度、网络密度与平均聚类系数、节点的入度/出度中心性、介数中心性与接近中心性等。通过这些指标，可以从整体结构角度观察两类群体的谱系网络是否呈现出不同的凝聚性、层级性与中枢节点分布特征。

2.4.4 分层与异质性分析

为进一步揭示群体内部差异，本研究将华人科学家按学科类型（如基础科学 vs 应用科学）、科研影响力水平（如引文前 10%、后 10%）及所属国家或地区等进行分层，对不同子群体的谱系指标进行描述性比较，揭示群体内部的异质性。

2.4.5 回归分析与条件相关性检验

以华人科学家身份标识为核心解释变量、6 项谱系指标为因变量，构建三层递进回归模型。其中，M0 为基准模型；M1 引入 h 指数对数变换作为科研产出代理变量；M2 进一步加入六大学科大类固定效应。

针对 6 项因变量各自的分布特征，本研究采用差异化计数模型。世代深度与关系总数的零值比例可忽略，采用标准负二项回归；表亲数量存在一定零值堆积且过度离散，采用零膨胀负二项模型，以区分结构性零与抽样零；繁殖力、生育力与后代规模的零值堆积较为明显，采用 Hurdle 负二项模型，分别估计是否拥有该谱系特征与拥有者的规模期望。所有模型均采用学科层面聚类稳健标准误；针对共 18 组主模型检验与 14 组稳健性检验，对全部 p 值进行多重比较校正，以 $q<0.05$ 为显著性阈值。此外，采用熵平衡与逆概率加权两种协变量平衡策略进行加权回归，以处理样本中华人科学家比例较低且学科间分布不均的问题。

2.5 数据质量与识别误差评估

由于本研究依赖多源数据库构建华人与非华人对照样本与导师－学生谱系网络，潜在偏差主要来自两类环节：一是“华人身份”识别的误判与漏判；二是学术谱系数据库（AFT）在不同国家与学科上的记录完整度差异可能导致的系统性覆盖偏差（coverage bias）。针对第一类偏差，本研究在 2.2.1 节采用置信度分层与双向人工抽查进行评估。本节重点报告第二类偏差的量化诊断与稳健性处理方案。

AFT 并非对全球科研共同体进行均匀抽样，其导师－学生关系记录可能在学科、地区与时间维度上存在显著差异。若某一群体更容易被 AFT 记录，则该群体在可观测谱系网络中的“繁殖力、后代规模与代际深度”等指标可能被系统性地抬高，从而影响组间比较的解释有效性。为此，本研究首先在 ESD 的统一样本框架下，计算 AFT 的可匹配覆盖率并进行分层对比。具体而言，定义覆盖指示变量 Ri：若科学家 i 能够在 AFT 中匹配到唯一节点且具有至少一条有效导师-学生关系记录，则 Ri=1，否则 Ri=0。

3 结果分析

本节在前述数据与指标体系基础上，从样本基本特征、学术谱系指标整体比较、分层异质性、网络结构比较与回归分析等方面，对全球华人科学家与非华人科学家的学术谱系特征进行结果呈现。

3.1 样本基本特征与科研产出比较

基于学科分布的比较分析表明，华人科学家与非华人科学家在研究领域上存在一定差异。总体上，华人科学家更集中于化学、计算机科学、神经科学、物理学、数学等基础与前沿学科；非华人科学家则在化学、神经科学、物理学、进化科学等学科中占比较大（图 1）。该差异的产生可能与不同国家科研体系的重点布局、学生学科选择偏好以及历史发展路径相关，为后续比较两类群体的谱系指标提供了学科背景的支撑。进一步分析发现，尽管两个群体在科研产出方面的总发文量分布大体相似，但在“论文角色”与“引文影响力”上存在稳定的统计差异（图 1）。

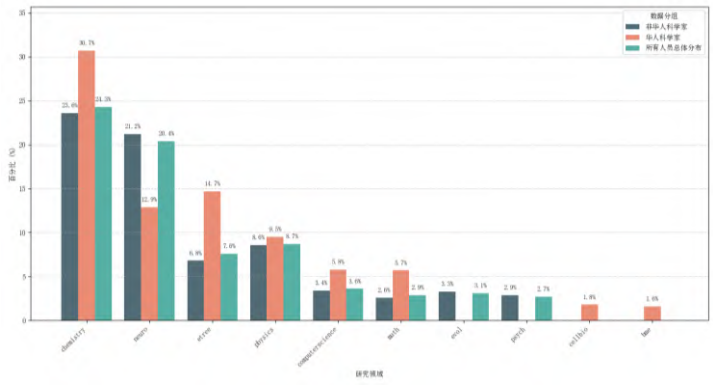


图 1 华人与非华人科学家研究领域分布

具体指标分析显示，非华人科学家的单篇、第一作者、最后作者论文数量（npsfl）低于华人科学家，但其单篇、第一作者的论文数量高于华人科学家；非华人科学家的总被引次数（nc9622）、h 指数（h22）、单篇、第一作者论文的总被引（ncsf）以及单篇、第一作者、最后作者论文的总被引（ncsfl）均高于华人科学家；不同引文论文数量指标（npciting）在非华人样本中同样表现出更高水平（图 2）。

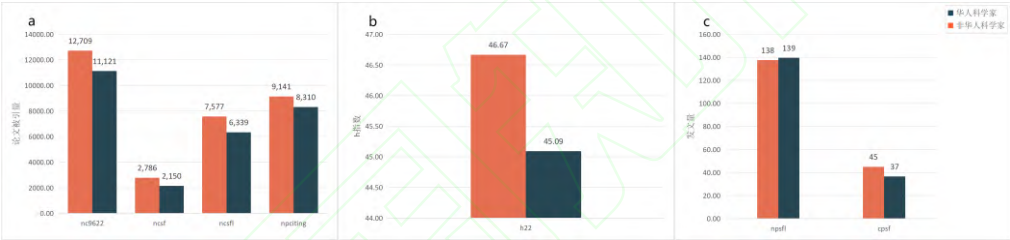


图 2 华人与非华人科研产出对比

为验证上述组间差异的统计显著性，研究采用曼-惠特尼 U 检验对两类群体的引用与产出指标进行比较，分析结果表明：在 0.05 的显著性水平下，除单篇、第一作者、最后作者论文数量（npsfl）外，其余指标的分布差异均具有统计显著性，即华人科学家与非华人科学家在科研产出结构与引文影响力方面存在系统性差异。这一结果提示，在解读后文的学术谱系指标时，需要考虑两类群体在科研表现上的整体差异背景。

3.2 学术谱系指标的整体比较

本小节基于聚合网络视角，从繁殖力、生育力、后代规模、表亲数量、世代深度与关系总数这 6 个指标，对华人科学家与非华人科学家的学术谱系特征进行整体比较（表 3）。表 3 报告了两类群体在聚合学术谱系网络中的学术谱系指标的描述性统计量。图 3 展示了两类群体学术谱系指标的分布。在繁殖力（Fecundity）上，非华人科学家的平均繁殖力约为 1.09，华人科学家为 0.34，即非华人科学家平均培养的直系学生数量更多。在生育力（Fertility）上，非华人科学家的平均生育力约为 0.24，华人科学家约为 0.04，表明在导师–学生的关系链中，非华人科学家的学生更有可能继续成为导师并延续学术传承。在后代规模（Descendants）上，非华人科学家的平均后代规模约为 5.77，华人科学家约为 0.74，图 5 进一步显示，非华人科学家的学术谱系在纵向和横向扩展方面表现出更大的规模；同时，非华人科学家的尾部分布明显更厚，在 x≥10 区域的占比显著高于华人科学家，说明其谱系外延具有更强的长尾特性（图 4）。在表亲数量（Cousins）

上，非华人科学家的表亲数量平均为 4.44，华人科学家约为 3.88，两者差距相对较小，但非华人科学家仍略高，显示非华人的同辈网络规模稍大。在世代深度（Generations）上，非华人科学家平均约为 1.97 代，华人科学家约为 1.26 代，表明非华人科学家的学术谱系具有更深的代际传承层次。在关系总数（Relationships）上，非华人科学家的关系总数均值约为 11.99，华人科学家约为 7.07，这表明华人所在的谱系网络规模偏小。

表 3 学术谱系指标描述统计

	非华人科学家			华人科学家		
	平均值	**标准差**	**方差**	**平均值**	**标准差**	**方差**
繁殖力	1.09	2.17	4.709	0.34	1.136	1.29
生育力	0.24	0.549	0.301	0.04	0.232	0.054
后代规模	5.77	40.228	1618.305	0.74	7.665	58.752
表亲数量	4.44	8.577	73.573	3.88	8.874	78.754
世代深度	1.97	1.883	3.546	1.26	0.837	0.701
关系总数	**11.99**	**44.084**	**1943.442**	**7.07**	**16.656**	**277.413**
			有效个案数：25,218			有效个案数：3,037

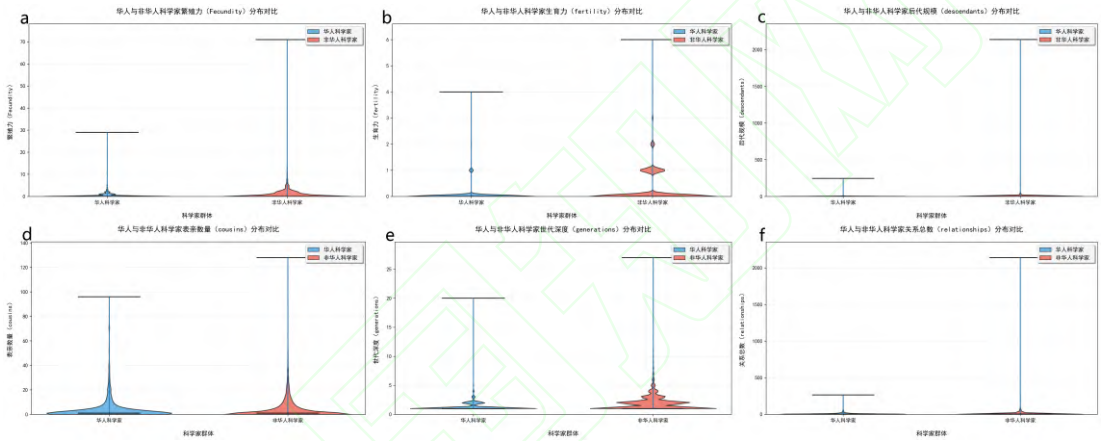


图 3 华人与非华人科学家学术谱系指标分布对比

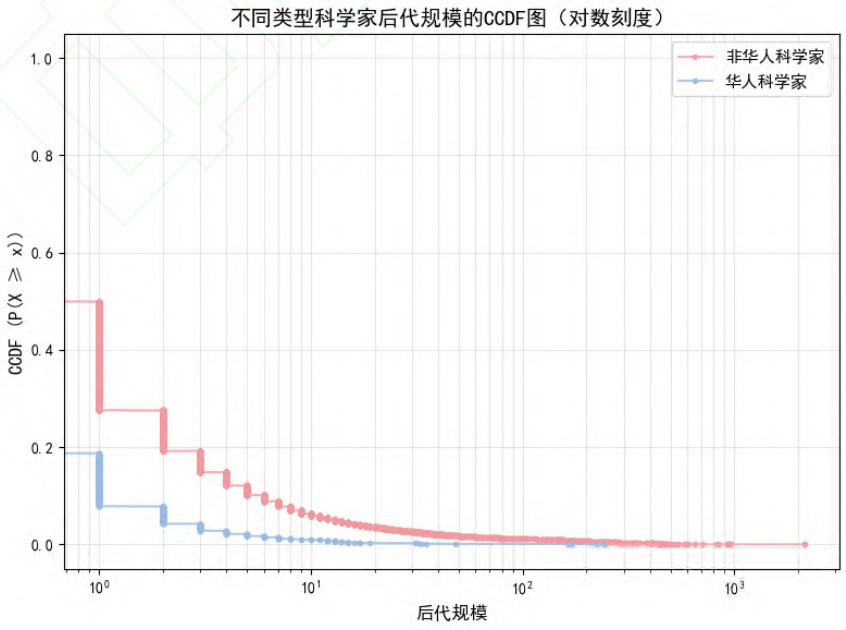


图 4 后代规模互补累积分布对比

在此基础上，采用曼-惠特尼 U 检验对两类群体各谱系指标的分布进行比较（表 4）。检验结果显示，6 项指标的组间差异均在 0.05 水平下具有统计显著性。效应量分析表明，繁殖力（δ=0.3203）、后代规模（δ=0.3290）与世代深度（δ=0.3258）达到中等效应量，生育力（δ=0.1624）与关系总数（δ=0.1220）居小至中等效应，表亲数量（δ=0.0401）效应最小，属边际性差异。核心谱系规模与代际深度指标的效应量表明，组间差异具有实质性意义，并非大样本量的统计假象。基于高置信度华人子样本的重复分析，差异方向与统计显著性与主分析一致，效应量略有收缩，主要结论对身份识别误差具有一定稳健性。

表 4 华人与非华人科学家谱系指标效应量

指标名称	δ值	量级
繁殖力	0.3203	小
生育力	0.1624	小
后代规模	0.3290	小
表亲数量	0.0401	可忽略
世代深度	0.3258	小
关系总数	0.1220	可忽略

3.3 分层与异质性分析

为进一步刻画群体内部差异，本研究从网络视角、引文层级、学科类型、华人国籍、地域分布五个维度，对学术谱系指标进行了分层比较。

3.3.1 分散网络视角下的谱系比较

在分散网络（即仅包含华人和非华人科学家的网络分别独立研究，形成规模较小或相对独立的谱系子网络）的框架下，两类群体的各谱系指标均较聚合网络有所收缩，但组间相对差异仍然存在（表 5）。表 5 的描述性统计结果显示，非华人科学家的平均繁殖力约为 1.01，华人科学家约为 0.64，与聚合网络中的对应指标相比，华人科学家的平均繁殖力有较大幅度的增加；但非华人科学家的生育力、后代规模与表亲数量等指标仍普遍高于华人科学家；在分散网络中，华人科学家的表亲数量有显著的下降；世代深度方面，非华人科学家平均约为 1.92 代，华人科学家约为 1.46 代；华人科学家的关系总数也呈现出明显的下降趋势。

表 5 分散网络学术谱系指标对比

	非华人科学家			华人科学家		
	平均值	标准差	方差	平均值	标准差	方差
繁殖力	1.01	2.04	4.17	0.64	1.06	1.11
生育力	0.23	0.53	0.28	0.04	0.19	0.04
后代规模	5.19	35.89	1287.76	0.71	1.32	1.75
表亲数量	4.16	7.78	60.53	0.88	2.05	4.19
世代深度	1.92	1.85	3.44	1.46	0.57	0.33
关系总数	11.37	40.03	1602.07	2.43	1.16	1.35
	有效个案数（成列）：24394			有效个案数（成列）：874		

纳入部分非华人节点后，得到较为完整的华人学者及其直接关系学术谱系图（图 5）。该子图聚焦于华人学者在学术谱系中的嵌入位置与代际连接模式，子图中相当比例的华人学者处于以非华人导师为根的学术谱系分支之中。华人学者群体在学术谱系中并非均匀散布，而是高度凝聚于若干学术分支周围，形成明显的聚类格局。图中清晰可见若干度值远超平均水平的超级节点（如 H1~H5），这些节点构成了学术传承的关键枢纽，对知识流动与人才扩散发挥着支配作用。

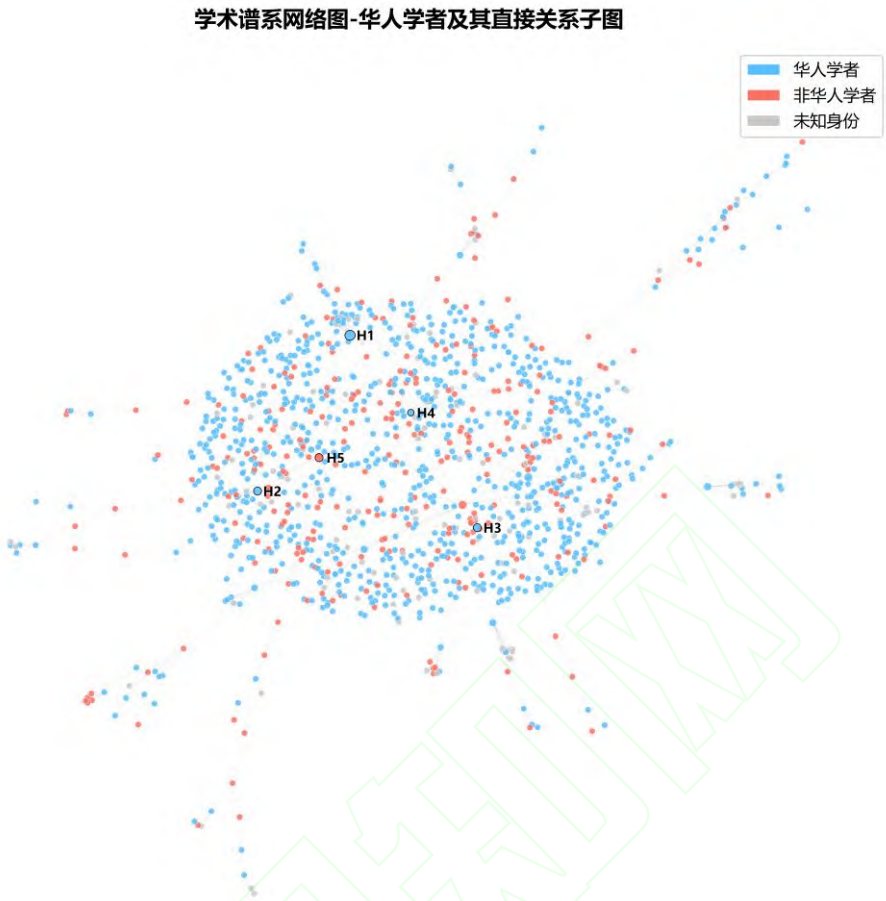


图 5 华人学者学术谱系网络子图

与聚合网络相比，分散网络中的两类群体的世代深度指标差异相对缩小，且统计检验结果显示两类群体在代际深度上的差异不再具有统计学显著性。这一结果提示，在局部、规模较小的谱系结构中，华人科学家与非华人科学家在纵向延续方面较为接近，这一情况在繁殖力指标的变化上也有所反映。效应量分析结果显示，分散网络下部分指标虽仍显著，但效应量数值明显下降（如繁殖力 δ=0.0801），表明两类群体在学术谱系上的部分差异可能受网络规模与结构差异的影响（表 6）。与此同时，非华人科学家在繁殖力、生育力与后代规模上依然保持明显优势，说明两类群体在培养学生数量和谱系外延方面的差异具有一定稳定性。其次，华人科学家表亲数量和关系总数的下降，以及网络子图中大量存在非华人根节点的事实也反映了非华人科学家在学术谱系中占据主导地位。

表 6 分散网络学术谱系指标效应量

指标名称	δ值	量级
繁殖力	0.0801	可忽略
生育力	0.1538	小
后代规模	0.1088	可忽略
表亲数量	0.3253	小
世代深度	0.1079	可忽略
关系总数	**0.5155**	**大**

3.3.2 按科研影响力分层的谱系特征

为考察科研影响力不同层级科学家的谱系特征，本研究进一步选择各自群体中被引量前 10%与后 10%的科学家，分别计算其学术谱系指标的均值（表 7-8）。对高被引科学家群体（被引频次前 10%）的分析显示，非华人科学家的平均繁殖力为 2.71，华人科学家为 1.03；生育力分别为 0.48 与 0.13；后代规模分别为 12.11 与 3.24；表亲数量分别为 6.14 与 2.47；世代深度分别为 2.6 与 1.72，关系总数分别为 18.78 与 9.57。

表 7 学术谱系指标被引分层对比

	引用量前 10%		引用量后 10%	
	非华人科学家	华人科学家	非华人科学家	华人科学家
繁殖力	2.71	1.03	0.53	0.08
生育力	0.48	0.13	0.15	0.01
后代规模	12.11	3.24	2.69	0.1
表亲数量	6.14	7.3	2.47	2.14
世代深度	2.6	1.72	1.65	1.07
关系总数	18.78	9.57	7.01	5.5
有效个案数	2,521	303	2,521	303

表 8 学术谱系指标被引分层效应量

	引文分层前 10%		**引文分层后 10%**	
指标名称	**δ值**	**量级**	**δ值**	**量级**
繁殖力	0.3591	中	0.2860	小
生育力	0.2457	小	0.1220	可忽略
后代规模	0.3747	中	0.2881	小
表亲数量	0.0166	可忽略	0.0245	可忽略
世代深度	0.3621	中	0.2880	小
关系总数	**0.1626**	**小**	**0.0578**	**可忽略**

在引文后 10%的低被引科学家中，两类群体的谱系指标整体水平显著下降，但仍保持非华人科学家更高、华人科学家较低的相对格局。虽然各组间差异均达到统计学显著水平，但效应量分析进一步揭示，其实际的差异程度在不同层级中存在明显异质性。例如，在高被引群体中，繁殖力的 δ 值达到 0.3591（中），而在低被引群体中 δ 降至 0.2860（小），说明学术影响力与谱系规模之间的关联在不同群体中表现出不同的特征（图 6）。

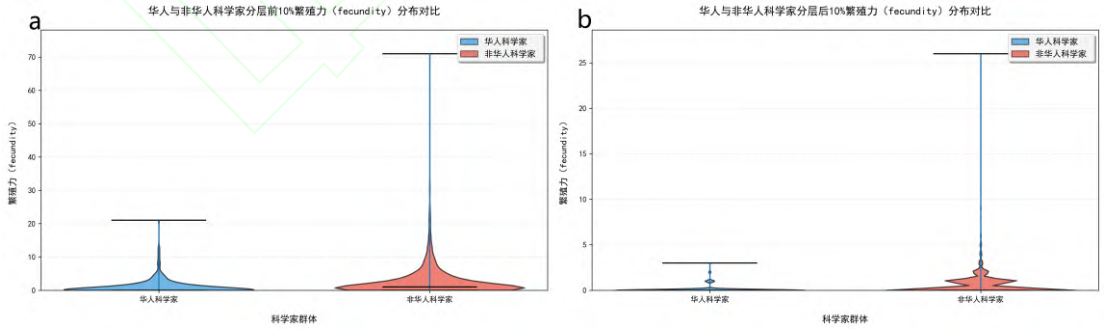


图 6 不同影响力分层下的繁殖力分布

3.3.3 学科类型内部的华人科学家谱系差异

考虑到不同学科的博士培养规模、导师与学生比以及学位制度存在差异，进一步按学科类型对华人科学家内部进行了划分（表 9-10）。将物理学、数学、化学、哲学等归为基础学科，将工程、材料等归为应用学科，并比较两类学科中的华人科学家谱系指标。

在华人科学家内部，基础学科的繁殖力高于应用学科，后代规模与表亲数量的差异方向相同；世代深度在两类学科间几乎无差别；应用学科的关系总数略高。效应量分析显示，除关系总数外，其余指标的学科间差异效应量均较小，提示华人科学家内部的学科分化主要体现在谱系组织结构规模上，而非代际深度。

表 9 学科分组谱系指标描述统计

谱系指标	基础科学	应用科学
繁殖力	0.38	0.27
生育力	0.05	0.05
后代规模	0.90	0.63
表亲数量	5.49	3.90
世代深度	1.28	1.28
关系总数	6.68	15.47

表 10 学科分组谱系指标效应量

指标名称	δ值	量级
繁殖力	0.0241	可忽略
生育力	0.0116	可忽略
后代规模	0.0212	可忽略
表亲数量	0.0199	可忽略
世代深度	0.0214	可忽略
关系总数	**0.1983**	**小**

3.3.4 境内与境外华人科学家的谱系差异

为考察不同学术制度环境下华人科学家的谱系特征，本研究依据科学家当前主要任职机构所在地，将华人样本划分为境内华人与境外华人两个群体。境内组包括主要任职于中国大陆、中国香港、中国澳门和中国台湾机构的华人科学家，境外组包括主要任职于其他国家和地区机构的华人科学家。采用曼-惠特尼 U 检验和克利夫 δ 效应量比较两组在 6 项谱系指标上的差异，结果见表 11。

表 11 境内外华人科学家谱系指标比较

指标	境内华人均值	境内华人中位数	境外华人均值	境外华人中位数	M-WU 检验 p 值	Cliff'sδ
繁殖力	0.1726	0	0.4771	0	1.88E-27	-0.1551
生育力	0.0222	0	0.0579	0	1.78E-06	-0.0325
后代规模	0.2479	0	1.1645	0	2.91E-27	-0.1547
表亲数量	3.0315	1	4.5935	1	2.78E-03	-0.0604
世代深度	1.1261	1	1.3772	1	1.31E-27	-0.1552
关系总数	**6.4348**	**3**	**7.6051**	**3**	**6.02E-02**	**-0.0378**
有效样本数	1,396		1,641			

境外华人科学家的 6 项谱系指标均值均高于境内华人，其中繁殖力、后代规模和世代深度的差异达到统计显著水平，效应量分别为−0.1551、−0.1547 和−0.1552，均属小效应。生育力与表亲数量的组间差异虽然具有统计显著性，但效应量较小，实际差异有限；关系总数差异未达到统计显著水平（p=0.0602）。同时，两组在各项指标上的中位数基本一致，说明其差异主要表现为境外华人群体中部分科学家具有更大的谱系规模和更深的代际延展，而非整体分布位置的明显分离。这一结果表明，境内与境外华人科学家的谱系结构存在一定差异，但任职地域的解释力总体有限。华人科学家的谱系形成可能同时受到职业阶段、学科构成、博士生培养制度、科研组织环境以及数据库记录完整度等因素的共同影响，不能将组间差异简单归因于地域环境。

3.3.5 不同区域的谱系比较

考虑到不同国家和地区的学术制度、博士生培养规模存在显著差异，进一步将非华人科学家按所属国家和地区划分为 3 个主要子群：美国组、欧洲组（涵盖英国、德国、瑞士、法国、荷兰等 30 个主要欧洲国家）和日本、韩国组，进行比较，结果如表 12-13。

表 12 按国家和地区分组的谱系指标描述统计

指标名称	华人科学家	非华人—美	非华人—欧	非华人—日本、韩
繁殖力	0.34	1.23	0.88	0.47
生育力	0.04	0.26	0.23	0.11
后代规模	0.74	6.17	5.63	1.52
表亲数量	3.88	4.63	4.18	5.50
世代深度	1.26	2.04	1.88	1.43
关系总数	**7.07**	**12.23**	**12.30**	**8.89**

表 13 非华人-美国组与华人科学家谱系指标效应量

	非华人—美		非华人—欧		非华人—日本、韩	
指标名称	δ值	量级	δ值	量级	δ值	量级
繁殖力	0.3966	中	0.3561	中	0.1193	可忽略
生育力	0.4100	中	0.4369	中	0.2723	小
后代规模	0.1595	小	0.1160	可忽略	0.0964	可忽略
表亲数量	0.0842	可忽略	0.0372	可忽略	0.1808	小
世代深度	0.4205	中	0.4278	中	0.1961	小
关系总数	**0.1348**	**可忽略**	**0.1155**	**可忽略**	**0.1066**	**可忽略**

繁殖力、生育力、后代规模与世代深度 4 项指标均呈现清晰梯度，依次为美国、欧洲、日韩、华人。华人繁殖力（0.34）为美国组（1.23）的 28%、欧洲组（0.88）的 39%；生育力差距在所有指标中最为悬殊；表亲数量的 4 组差距相对最小。效应量分析（表 13）显示，美国组与欧洲组在繁殖力、生育力、世代深度 3 项上效应量达中水平，其余多为可忽略；日本和韩国组与华人在全部指标上的效应量均为小或可忽略，日本、韩国组与华人科学家的效应量总体小于美国和欧洲组，表明东亚样本之间的谱系指标更为接近；这种接近是否与博士培养制度、科研组织结构或文化传统有关，仍需结合制度变量进一步检验。

3.4 网络结构特征比较

除个体层面的谱系指标外，本研究还从整体网络结构角度比较了华人科学家与非华人科学家的学术谱系网络特征。表 14 报告了主要网络指标。

表 14 谱系网络主要网络指标

	非华人科学家网络结构特征	华人科学家网络结构特征	总网络结构特征
平均度	1.007	0.638	1.007
平均加权度	1.015	0.644	1.015
网络直径	22	3	22
图密度	0.00	0.001	0.00
平均聚类系数	0.009	0.003	0.009
特征向量中心度	3.257	0.156	3.257
平均路径长度	4.74	1.111	4.74

从整体网络结构看，非华人谱系网络的平均度与加权平均度均高于华人网络；网络直径与平均路径长度的差异主要反映两类网络规模的差异。华人网络图密度略高于非华人，但非华人网络的平均聚类系数更高，在更大规模下呈现出更多局部聚类结构。非华人网络的特征向量中心度远高于华人网络，表明非华人群体中存在影响力更为突出的核心节点。上述指标受 AFT 数据库在各学科与地区记录完整度的影响，更适合作为两类群体相对结构差异的参照，而非对各自学术共同体整体结构的完整刻画。

3.5 学术谱系指标的回归分析结果

回归分析基于三层递进模型（M0/M1/M2），针对各因变量的分布特征分别采用标准负二项回归（Negative Binomial, NB）、零膨胀负二项回归（ZINB）或 Hurdle 负二项模型（Hurdle-NB），主要结果见表 15。

表 15 华人科学家身份对学术谱系指标的回归分析结果

谱系指标	模型	模型类型	零膨胀 OR	计数 IRR	95% CI	p（FDR 校正）	显著性
	M0（基准）	NB	—	0.642	[0.596, 0.691]	<0.001	***
世代深度	M1（+ln_h22）	NB	—	0.646	[0.606, 0.688]	<0.001	***
	M2（+学科 FE）	NB	—	0.656	[0.618, 0.697]	<0.001	***
	M0（基准）	NB	—	0.589	[0.479, 0.725]	<0.001	***
关系总数	M1（+ln_h22）	NB	—	0.600	[0.497, 0.723]	<0.001	***
	M2（+学科 FE）	NB	—	0.661	[0.560, 0.781]	<0.001	***
	M0（基准）	ZINB	0.144	0.873	—	>0.05	
表亲数量	M1（+ln_h22）	ZINB	0.339	0.868	—	>0.05	
	M2（+学科 FE）	ZINB	0.272	0.791	—	<0.01	**
	M0（基准）	Hurdle-NB	0.232	0.824	[0.743, 0.915]	<0.001	***
学术繁殖力	M1（+ln_h22）	Hurdle-NB	0.224	0.793	[0.713, 0.881]	<0.001	***
	M2（+学科 FE）	Hurdle-NB	0.219	0.805	[0.724, 0.895]	<0.001	***
	M0（基准）	Hurdle-NB	0.152	0.936	[0.723, 1.212]	>0.05（计数）	
学术生育力	M1(+ln_h22)	Hurdle-NB	0.153	0.927	[0.715, 1.200]	>0.05（计数）	
	M2（+学科 FE）	Hurdle-NB	0.158	0.927	[0.715, 1.201]	>0.05（计数）	
	M0（基准）	Hurdle-NB	0.232	0.343	[0.312, 0.376]	<0.001	***
后代规模	M1（+ln_h22）	Hurdle-NB	0.224	0.332	[0.302, 0.365]	<0.001	***
	M2（+学科 FE）	Hurdle-NB	0.219	0.462	[0.421, 0.508]	<0.001	***

注：OR<1 表示华人科学家更不可能属于零膨胀群体（即更可能拥有该谱系特征）；IRR<1 表示华人科学家在该指标计数水平上低于非华人科学家。*** p<0.001, ** p<0.01, * p<0.05（FDR 校正后）。标准误均为学科层面聚类标准误。

3.5.1 原始组间比较（M0）

从基准模型来看，华人科学家的世代深度发生率比率（Incidence Rate Ratio，IRR）为 0.642（p<0.001），较非华人低约 35.8%；关系总数发生率比率为 0.589（p<0.001），差距约 41.1%。Hurdle 模型中，繁殖力计数部分发生率比率为 0.824，后代规模计数部分发生率比率仅为 0.343，后代总量约低 65.7%，为所有指标中效应量最大者。生育力计数部分发生率比率为 0.936 且不显著，表明原始状态下华人科学家在学生出师率上与非华人无显著差异。

障碍模型零阶段的 OR 均小于 1，表示在本研究设定下，华人科学家落入零值组的概率较低，即其拥有直接学生或可观测学术后代的概率并未低于非华人科学家。群体差距主要出现在非零样本的计数阶段，即一旦形成可观测谱系，华人科学家的直接学生数量、后代规模和代际延展仍相对较小。

3.5.2 控制科研产出后 h 指数的解释力（M1）

模型 M1 在 M0 基础上引入 ln_h22 后，多数指标的核心系数变化极为有限：世代深度发生率比率从 0.642 仅微升至 0.646，关系总数从 0.589 升至 0.600，表明华人科学家的谱系深度与规模差距几乎独立于个体科研产出。繁殖力计数部分发生率比率反而从 0.824 降至 0.793。在控制 h 指数后，华人科学家相对非华人的培养学生数量劣势进一步扩大，控制 h 指数后系数未明显缩小，说明科研产出差异不能充分解释两组的谱系差距，其他可能因素仍需进一步检验。

3.5.3 控制学科分布后的估计结果（M2）

模型 M2 在 M1 基础上进一步加入六大学科大类固定效应，引入后带来明显的系数回升。后代规模发生率比率从 0.332 大幅升至 0.462，关系总数从 0.600 升至 0.661，表亲

数量计数部分首次达到显著（IRR=0.791，p<0.01），说明学科分布结构能够解释相当部分的原始差距，华人科学家在谱系规模天然较小的学科中代表性更高。然而，全部显著指标的华人科学家系数在 M2 中仍显著低于 1。世代深度 IRR=0.656，后代规模 IRR=0.462，繁殖力 IRR=0.805，表明学科分布只能部分解释而远非完全说明族群间的谱系规模差距。学术生育力计数部分在 M2 中依然不显著。

3.5.4 稳健性检验

为验证上述结论的可靠性，本研究以世代深度和关系总数两项代表性指标为对象，进行稳健性检验（表 16）。（R1）将样本限制为 NC9622 排名体系内的科学家，以控制排名体系的可比性；（R2）剔除观测数不足 50 的稀疏学科，以保证学科层面估计的稳定性；（R3）改用学科聚类标准误（与主回归一致）；（R4）采用熵平衡再加权以提升协变量均衡性；（R5）采用逆概率处理加权（IPTW）以校正选择偏差；（R6）对因变量进行 1%缩尾处理以降低极端值的扰动；（R7a）与（R7b）分别采用 OLS 对数线性回归与中位数分位数回归，以检验结论对模型分布假设的敏感性。

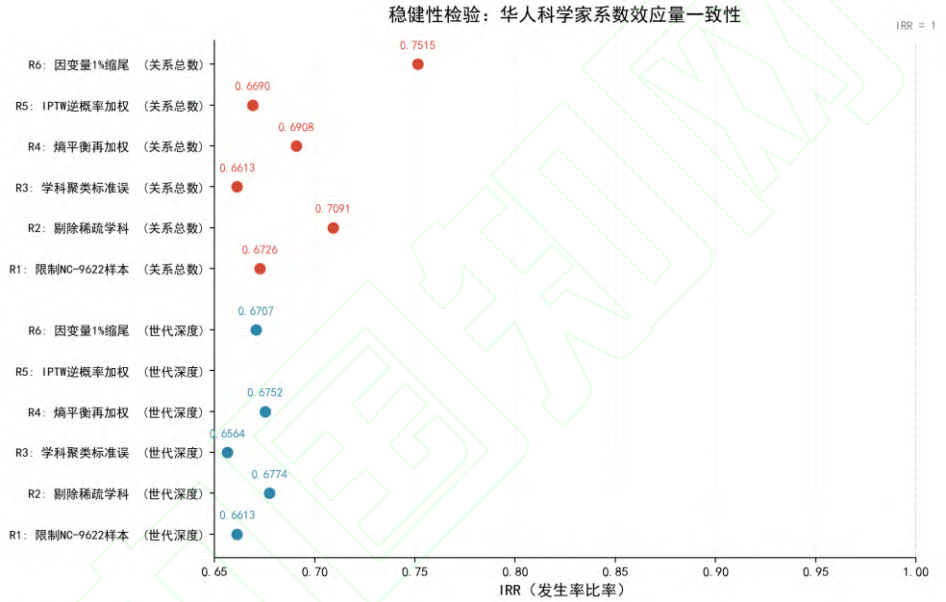


图 7 稳健性检验的效应估计结果

表 16 稳健性检验结果汇总（世代深度与关系总数）

检验方案	世代深度 IRR	世代深度 q 值	关系总数 IRR	关系总数 q 值
R1 限制 NC9622 样本	0.661	<0.001	0.673	<0.001
R2 剔除稀疏学科	0.677	<0.001	0.709	<0.001
R3 学科聚类标准误	0.656	<0.001	0.661	<0.001
R4 熵平衡再加权	0.675	<0.001	0.691	<0.001
R5 IPTW 加权	0.645	<0.001	0.669	<0.001
R6 因变量 1%缩尾	0.671	<0.001	0.752	<0.001
R7a OLS 对数线性回归	—(β=-0.200)	<0.001	—（β=-0.115）	<0.001
R7b 中位数分位数回归	—（β=-0.408）	<0.001	—（β=-0.416）	<0.001

所有 7 类稳健性检验结果均一致支持主回归的结论：华人科学家在世代深度与关系总数上显著低于非华人科学家，且各项检验的错误发现率（False Discovery Rate，FDR）校正 q 值均小于 0.001。其中，世代深度的 IRR 估计值在所有稳健性检验中介于 0.645 至 0.677 之间，关系总数的发生率比率估计值介于 0.661 至 0.752 之间，波动范围有限且方向完全一致。R7a 与 R7b 的结果进一步表明，即使放松计数模型的分布假设，改用 OLS

对数变换和中位数分位数回归等替代模型设定，华人科学家系数仍为负且高度显著。具体而言，世代深度的 OLS 对数系数为-0.200，中位数分位数回归系数为-0.408；关系总数的对应系数分别为-0.115 与-0.416，所有替代设定均指向稳健的负向效应。图 7 以一致性散点图呈现了稳健性检验中各方案效应估计值的分布，所有数据点紧密聚集且方向一致，进一步印证了结论的可靠性。

4 研究讨论

本节先将核心实证发现与既有文献对话，明确研究结论在已有研究积累中的位置；其次提出可能的解释线索。本研究揭示了两类群体在学术传承模式上的系统性差异，为理解全球学术人才培养的不均衡性提供了经验证据，也为后续机制探索指明了方向。

首先，研究发现华人科学家在繁殖力、生育力、后代规模、代际深度等学术谱系指标上的表现均显著低于非华人科学家，这一发现印证了既有学术谱系研究中导师资源与培养能力存在显著异质性的结论[3,30]。现有的国际谱系资料显示，大型学科（如物理、化学）往往形成多代延续的长链式谱系[14]，而在数据覆盖程度较低或导师与学生比不稳定的领域，谱系网类似呈现出短链条、低密度的特征。本研究的结果显示，华人科学家的学术谱系更可能相似于后者，这一结果支持了相关文献指出的学科组织结构与制度环境对导师与学生关系的影响的结论[16-17]。此外，境内与境外华人科学家的比较显示，多数谱系指标差异有限，说明华人科学家的谱系结构具有较强群体共性，任职地域可能影响学术资源获取，但并非解释谱系规模差异的唯一因素。

其次，本研究在以引文层级和不同学科类型进行分组的前提下，均观察到了稳定的组间差异；即使在高被引科学家中，华人群体的各个谱系指标仍显著低于非华人群体。这支持了科研产出与学术团队规模之间不存在简单线性对应关系与高科研产出不必然意味着更大的学生群体或更深的代际延续的研究结论[15]。学术繁殖力与后代规模在很大程度上会受团队规模、导师制度、博士生招收能力等结构性因素的影响，但这些重要的因素在现阶段的数据体系中仍难以直接观察或量化。

最后，在华人科学家内部的学科差异分析中，基础科学领域的学术繁殖力和表亲网络规模普遍高于应用科学领域。这一点印证了已有研究“基础学科博士培养规模更大、导师与学生传承链更稳定”的描述性证据，也反映了不同学科知识生产模式的内在差异[31]。

本研究尚不能直接识别群体差异的形成机制，但有 3 条值得检验的解释线索：一是博士生招生规模、导师资格制度和科研项目组织方式可能影响直接学生数量与谱系扩展；二是科学家的职业生涯长度可能影响其形成多代谱系的时间；三是 AFT 在不同地区、学科和历史时期的记录完整度可能造成可观测谱系差异。分散网络分析和境内外华人比较表明，华人科学家多遵循“本土基础教育－海外博士/博士后训练”的学术路径，其学术关系网络可能更具跨国特征[18-19]。网络构造与任职环境确实与部分指标变化相关，但其作用尚不足以从现有数据中作因果判断。后续研究需要引入职业阶段、机构资源和博士培养制度等变量，对上述解释进行直接检验。

5 研究结论与展望

5.1 研究结论

本研究整合 ESD、AFT 与 MAG 等 3 类数据库，在全球统一框架下首次系统构建华人与非华人科学家的可比谱系样本，综合运用描述性比较、非参数检验与差异化计数回归，对两类群体的学术谱系特征进行了系统比较。主要结论如下。

第一，华人科学家的六项谱系指标整体低于非华人科学家，其中繁殖力、后代规模和世代深度的差异较为突出。分散网络、引文分层与地域子群分析表明，差距方向在各分层情境中高度一致，即便在高被引华人科学家中亦然。华人群体的境内—境外比较显示多数谱系指标差异有限，进一步揭示了华人科学家的谱系特征更多体现为群体共同模式，而非简单由任职地域决定；地域比较呈现清晰梯度，日本、韩国组与华人科学家的谱系指标相对接近，呈现出值得进一步解释的区域差异。华人科学家内部亦存在明显的学科差异，基础学科的繁殖力与后代规模高于应用学科，谱系结构更为集中稳定；网络结构分析进一步表明，两类群体在谱系传承形态上存在整体性差异。

第二，回归结果显示，华人科学家形成可观测谱系的概率并不低，但在非零样本中，其直接学生数量、后代规模和代际深度相对较小。控制 h 指数（M1）后，各指标核心系数几乎未发生实质衰减，表明谱系差距独立于个体科研生产力；控制学科大类（M2）后，后代规模与关系总数的系数出现明显回升，而世代深度等指标的系数几乎不变，揭示出两类渠道的不同性质。后代规模差距有相当部分来自华人科学家集中于谱系规模天然较小的学科，但剩余差距仍高度显著；世代深度等指标的差距则与科研产出和学科归属均无关，可能指向更深层的制度性因素或学术网络嵌入方式。学术生育力计数阶段在全部模型设定中均不显著，表明华人导师学生的出师率与非华人无系统性差异。

第三，多源数据库整合与差异化计数模型能够区分谱系特征的形成概率和非零规模，为跨群体比较科研人才传承结构提供了可复用的方法。在数据整合层面，通过 ESD、AFT 和 MAG 三库联动构建大规模华人科学家识别与对照样本，突破了已有谱系研究局限于单一学科或少数科学家的约束；在方法层面，系统引入差异化计数模型，区分是否拥有谱系特征与拥有者的规模期望两种机制，为跨群体谱系比较的定量分析提供参考。

5.2 不足与展望

本研究仍存在若干局限。首先，使用的 AFT 等数据库存在地域与学科覆盖偏差，尤其是早期华人科学家的导师–学生关系可能存在系统性缺失。其次，未对职业阶段、机构层级、学科制度等潜在重要变量进行严格控制，因此呈现的群体差异仅为相关性，而不能被视为严格的因果关系。最后，聚焦于学术谱系的师承关系结构，尚未触及导师–学生之间在研究内容、学术风格等方面的知识传承，这也限制了谱系分析的解释深度。

本文华人科学家的操作化识别框架存在固有局限，其误差方向涉及两类情形：一是对华语文化背景但姓名高度英文化的科学家（如部分在他国出生的华裔后代）存在漏判风险，可能导致真实华人群体规模被低估；二是对在华语地区出生但主要在他国完成学术训练的科学家，其识别归类存在边界模糊。本文通过置信度分层（高、中、低）与稳健性分析对误差范围进行了部分量化，但识别误差对实证结论的潜在影响尚难完全排除。

未来研究可从 3 个方向推进：一是结合博士论文和机构档案改善谱系数据的地区与历史覆盖；二是引入职业阶段、机构资源和博士培养制度等变量，检验群体差异的形成条件；三是结合全文内容分析，追踪导师与学生之间的研究主题传承。

参考文献

[1] 中国科学技术信息研究所. 2025 中国科技论文统计报告[R]. 北京: 科学技术文献出版社.

[2] LIU M, HU X. Movers' advantages: The effect of mobility on scientists' productivity and collaboration[J]. Journal of Informetrics, 2022, 16(3): 101311. DOI: 10.1016/j.joi.2022.101311.

[3] 朱克曼. 科学界的精英——美国的诺贝尔奖奖金获得者[M]. 周叶谦, 冯世则, 译.北京: 商务印书馆, 1979: 102.

［4］ TYLER V M, TYLER V E. The academic genealogy of Arthur E. Schwarting, pharmacognosist [J]. Journal of Natural Products, 2004, 67(5): 833-844.

［5］ CHANG S. Academic genealogy of physicists[M]. Seoul: Seoul National University Press, 2005: 87-92.

［6］ PIERRE F. Self-interest before Adam Smith: a genealogy of economic science[M]. Cambridge: Cambridge University Press, 2003: 133-136.

［7］ 周积明, 雷平. 清代浙东学派学术谱系的构建[J]. 学术月刊, 2004(6): 40-46.

［8］ 乌云其其格, 袁江洋. 谱系与传统: 从日本诺贝尔奖获奖谱系看一流科学传统的构建[J]. 自然辩证法研究, 2009, 25(7): 57-63.

［9］ 刘莉, 贾婧. 我国启动当代中国科学家学术谱系研究[N/OL]. 科技日报, 2011-09-25（7）.

［10］杨晓薇, 高继平. 学术谱系的起源、现状和发展[J]. 科技管理研究, 2022, 42(11): 236-242.

［11］韩天琪, 樊小龙, 袁江洋. 唐敖庆谱系与福井谦一谱系比较研究[J]. 科学与社会, 2013, 3(1): 110-123. DOI: 10.19524/j.cnki.10-1009/g3.2013.01.011.

［12］杨晓薇, 雷鸣, 高继平. 杰出科学家学术谱系的指标构建与实证研究——以学者马费成为例[J]. 图书情报工作, 2022, 66(23): 115-124.

［13］ROSSI L, FREIRE I L, MENA-CHALCO J P. Genealogical index: A metric to analyze advisor–advisee relationships[J]. Journal of Informetrics, 2017, 11(2): 564-582.

［14］冯靖雯, 赵勇. 学术谱系视角下杰出科学家的师承关系特征研究——以诺贝尔化学奖得主 Lipscomb 为例[J]. 情报工程, 2020, 6(6): 22-32.

［15］WU L, WANG D, EVANS J A. Large teams develop and small teams disrupt science and technology[J]. Nature, 2019, 566(7744): 378-382. DOI: 10.1038/s41586-019-0941-9.

［16］XING Y, MA Y, FAN Y, et al. Academic mentees thrive in big groups, but survive in small groups[J]. Nature Human Behaviour, 2025, 9(5): 902-916.

［17］马杰, 别敦荣. 我国研究生教育师生关系调查研究[J]. 华东师范大学学报(教育科学版), 2021, 39(12): 81-98.

［18］乌云其其格. 我国“高被引科学家”的成才路径[J]. 中国人才, 2016(5): 26-29.

［19］王佳匀, 于璇, 汪雪锋, 等. 优秀青年科技人才成长典型特征分析：以优秀青年科学基金获资助者为例[J]. 中国科学基金, 2021, 35(2): 290-297.

［20］刘进， 于宜田， 尹璐。海归科学家的学术表现——基于 12.7 万名美国返华科学家大数据的分析[J]. 复旦教育论坛, 2024, 22(6): 75-87.

［21］孙康, 司月芳. 基于高被引华人科学家流动的城市网络空间演化分析[J]. 世界地理研究, 2023, 32(9): 109-119.

［22］司月芳, 孙康, 朱贻文, 等. 高被引华人科学家知识网络的空间结构及影响因素[J]. 地理研究, 2020, 39(12): 2731-2742.

［23］李晓轩, 牛珩, 冯俊新. 科研拔尖人才的成才规律与启示[J]. 科学学研究, 2004(3): 273-277.

［24］公钦正, 秦惠民. 中国高等教育学自主知识体系建构的主体自觉研究[J]. 华东师范大学学报(教育科学版), 2025, 43(12): 97-110.

［25］尹志欣, 王宏广. 顶尖科学人才现状及发展趋势研究[J]. 科学学与科学技术管理, 2017, 38(6): 23-30.

［26］李琳琳. 走向世界的中国教育研究:华人学者的视角[J]. 现代大学教育, 2020, 36(5): 45-52.

［27］IOANNIDIS J P A. October 2023 data-update for "Updated science-wide author databases of standardized citation indicators"[DS/OL]. Amsterdam: Elsevier, 2023. DOI: 10.17632/btchxktzyw.6.

［28］DAVID S V. Academic Family Tree Data Export (2023-05)[DS/OL]. Geneva: Zenodo, 2021. DOI: 10.5281/zenodo.7921789.

［29］SINHA A, SHEN Z, SONG Y, et al. An overview of Microsoft Academic Service (MAS) and applications[C]//Proceedings of the 24th International Conference on World Wide Web Companion. New York: ACM, 2015: 243-246.

［30］LIÉNARD J F, ACHAKULVISUT T, ACUNA D E, et al. Intellectual synthesis in mentorship determines success in academic careers[J]. Nature Communications, 2018, 9(1): 4840. DOI: 10.1038/s41467-018-07034-y.

[31] KAKU R, BITO M, NISHIMOTO K, et al. Asymmetric impact of basic scientists during applied shift [P/OL]. 2025-09-01. arXiv: 2509.01530. https://arxiv.org/abs/2509.01530.

作者简介 侯剑华（通信作者，houjh5@mail.sysu.edu.cn），中山大学信息管理学院教授，博士生导师；付兰志，中山大学信息管理学院图书馆学专业本科生。

收稿日期 2026-04-30

（责任编辑：刘洪）