丘成桐中学科学奖 · 生物奖赛道候选课题 20 则

> 本方案面向 **2027 赛季**（2026 赛季提交截止为 2026 年 9 月 15 日，从零启动已来不及）。倒排时间表以 2026 年 9 月为第 0 周、2027 年 9 月 15 日提交为终点，共约 52 周。

学生画像与资源假设：普通重点高中生，有一定编程与数学基础；计算资源为个人笔记本、纯 CPU、无 GPU 无集群；实验资源不超过中学常规实验室；全部使用公开数据；由中学教师即可指导。凡超出这一假设的路线，都在标签里标明了资源门槛。

四条必须先知道的赛制事实（取自官方公告，2020 – 2026 全部公示文章）：指导老师不得来自培训机构或任何谋利机构，违者取消资格；总决赛全程英文（研究报告、PPT、答辩）；入围总决赛的研究报告会在 11 月全网公示；提交材料含论文查重报告与实验视频。规则每年会改，报名开放后须重新核对当年《参赛规则》与本学科《参赛指南》。

这个赛道的评委在奖励什么

统计 2020 – 2025 共 63 项生物奖获奖论文，这个赛道的口味最"宽"，也因此最容易选错。

三条并行的获奖路径，天花板都够高。

其一是**分子机制与生物信息**：单细胞转录组解析帕金森风险细胞类型、镜像神经元分子身份、RNA velocity 因果基因调控、X 染色体失活逃逸区 SNP 与免疫病易感性、灵长类神经与生殖细胞发育的进化驱动基因。2025 金奖《在细胞内设计"分子宇宙"：液-液相分离与生物凝聚体设计》属于这一族的顶点。这一族的公开数据分支（GEO/TCGA/1000 Genomes 再分析）恰好是零仪器依赖的，是本方案 B01 – B10 的主战场。

其二是**生态、行为与植物生理**：蚂蚁基于视觉线索的同巢识别（2020 金奖）、爬山虎攀爬形态发生与茎尖负向重力性（2023 金奖）、藤蔓螺旋生长、蚜虫翅发育与亚致死农药、青海湖噬菌体多样性、赤霉素延长牵牛花花期。这一族是普通高中生真正的机会窗口——它拿过两个金奖，却几乎不需要高端仪器，主要成本是时间与观测纪律。本方案 B11 – B20 的重心在此。

其三是**医学工程与 AI 辅助**：颅内出血 CT 检测、踝关节护具、帕金森餐具防抖、sEMG 跌倒风险评估、蜜蜂健康多模态评估。这一族在 2022 – 2024 席位很多，2025 明显回落。它与计算机赛道高度重叠，且"做了一个能用的系统"容易被判为演示型项目——丘奖要的是研究报告。要走这条路，必须有可否证的科学假设，不能只有系统性能指标。

这个赛道最常见的两个死法，都与时间有关。 一是生物有生长周期：拟南芥一代 6 – 8 周、果蝇 10 – 14 天、微生物培养虽快但需要重复轮次。52 周的窗口里究竟能做几代、几轮重复，必须在开题时就算清楚，本方案每条实验路线都写死了这个数字。二是季节性：野外物候、昆虫行为、植物开花都锁死在特定月份，错过一次就要等一年。凡涉及季节的路线，go/no-go 门槛都提前到了可补救的时点。

伦理与生物安全是硬约束：本方案全部路线限定 BSL-1 微生物、非侵入式动物观测、不涉及可识别个人的人类数据。涉及人类数据的路线一律使用去标识化公开汇总统计，并在方案中写明。

本赛道 20 条路线的分族逻辑

- **B01 – B10 生物信息与公开数据族**：零实验室依赖，纯 CPU 笔记本（假设 16 GB 内存）。每条路线都核实了数据是否为**开放访问**——受控访问数据（如 dbGaP、UK Biobank 个体级）中学生拿不到，这是硬否决项，已在各路线中标明。算力边界也给了具体数字（例如从 fastq 重头比对在笔记本上多半不可行，应改用已处理的表达矩阵）。
- **B11 – B20 实验、生态与行为观测族**：器材限定在中学生物实验室常规配置，总预算 ⩽ 3000 元。每条路线明确写出**可完成的世代数与重复轮数**，以及观测判读的可复现判据与双人一致性检验。

两族独立排序：B01/B11 最稳，B10/B20 风险与上限最高。

B01 · 同一疾病多个独立公开 RNA-seq 队列的差异表达可重现性：以「同队列子抽样 vs 跨队列复制」的差距量化小样本转录组结论的乐观偏倚

`优先级：最高（最稳）` · `分族：公开组学再分析` · `参赛子类：生物信息学 / 计算生物学` · `资源需求：纯 CPU 笔记本，16 GB，无需下载 fastq` · `技能取向：R + Bioconductor 统计编程`

1 · 研究问题

对同一疾病，用统一管线在 5 个及以上相互独立的公开 bulk RNA-seq 队列上做差异表达（differential expression, DE）分析，**跨队列**的显著基因列表重叠率，比在单一大队列内部做同等样本量子抽样（subsampling）所得的**队列内**重叠率低多少？这个差距是否随样本量 N 单调收敛，能否给出一条"要使跨队列重叠率达到 0.5 所需的最小 N"的经验曲线？

2 · 研究背景与空白

背景。 bulk RNA-seq 的差异表达分析是生物医学论文最常见的产出形式之一：比较病例组与对照组，报出一份"显著上调/下调基因列表"，再做富集分析（enrichment analysis）。这份列表的可信度取决于统计检验力（statistical power），而检验力由样本量、效应量（effect size）和基因表达的生物学变异共同决定。人类组织样本的个体间变异远大于细胞系实验，因此小队列（每组 3 – 10 例）的 DE 列表是否稳定，直接决定了大量已发表结论是否站得住。

已有工作到哪一步。 Gerstner 等人（PLOS Computational Biology, 2025，PMC12077797）用 18 个真实数据集做了约 18,000 次子抽样实验：从一个大数据集中反复抽出规模为 N 的小队列，两两比较其 DE 结果的一致性。结论是每组少于 5 例时可重现性极低，但"低可重现性不等于低精确率"——18 个数据集中有 10 个在 N>5 时中位精确率仍然较高。他们还给出了一套自助法（bootstrap）程序，用来从手头数据估计预期可重现性。另一条线是 Li 等人（Genome Biology, 2022）指出 DESeq2/edgeR 在人群样本上会产生远高于名义水平的假阳性。

空白在于：现有的可重现性估计几乎全部来自**同一个数据集内部的子抽样。** 同一队列的样本共享招募标准、组织取材流程、文库制备批次和测序平台，因此队列内子抽样系统性地**低估**了真实世界"另一个实验室重做一遍"的不一致程度。真正决定文献可信度的是跨队列复制率，而这个量至今没有以"同一疾病、统一管线、多队列"的形式被系统标定过。这个空白适合一年期学生课题：数据完全公开且已统一处理（无需比对原始 fastq）、方法学完全公开、**结论正负都成立**——若跨队列率接近队列内率，说明现有小样本文献比想象中稳健；若显著更低，则给出一个可直接引用的乐观偏倚系数。

3 · 可检验假设

- H1：在相同的每组样本量 N（$N \in \{3, 5, 8, 12, 20\}$）下，跨队列 DE 列表的 Jaccard 重叠率显著低于同队列子抽样重叠率，差值在 N=5 时 $\geqslant 0.10$（绝对值），且 95% 自助法置信区间不跨 0。
- H2（机制假设）：这一差距的主要来源是队列层面的技术与人群构成异质性而非随机噪声——若对每个队列先做 `limma::removeBatchEffect` 之外的**队列内标准化 + 秩变换**，再在秩空间比较，差距应缩小 $\geqslant 40\%$；若缩小不足 20%，则说明差异来自真实的人群生物学异质性，而非可校正的批次效应（batch effect）。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：选定 3 篇公开了完整差异基因表的原始论文（其数据在本项目队列集合内），用自建管线在**原文全样本、原文所述统计设定**下重跑，要求本项目得到的显著基因列表与原文发表列表的重叠率 $\geqslant 70\%$（以原文列表为分母，FDR<0.05、|log2FC|>1 的口径对齐），且 log2 倍数变化（log2 fold change）的 Spearman

相关 $\geqslant 0.85$。**任一队列不达标即视为管线未通过，该队列剔除；若 3 篇中 $\geqslant 2$ 篇不达标，全部后续结论无效，转入 go/no-go 的降级路径。**

2. **统计口径预先写死**：所有 DE 检验一律报 Benjamini – Hochberg FDR，不报裸 p 值；主分析阈值固定为 FDR < 0.05 且 $|\log2FC| > 1$，并另做 FDR < 0.1 与仅 FDR 无 FC 阈值两组敏感性分析。重叠率同时报 Jaccard 指数与"前 200 基因的秩相关"两个口径（前者对列表长度敏感，后者不敏感），两者都必须报。
3. **批次效应必须显式检查**：每个队列先做主成分分析（PCA）与 `sva::num.sv` 估计潜在因子数，报告病例/对照分组在前 3 个主成分上的解释比例；若某队列的病例/对照与测序批次完全混杂（confounded，即批次可完全预测分组），该队列标注为"不可去混杂"并单独列出，不进入主分析。
4. **检验力与效应量预先估计**：用 `RNASeqPower` 或基于队列内离散度（dispersion）的模拟，给出每个 N 下检测 log2FC=1、基线表达中位数的基因所需的检验力，写入结果表。样本量不足以支撑的结论一律标注。
5. **规模下限**：至少 5 个独立队列，每个队列病例与对照各 $\geqslant 15$ 例（保证可抽到 N=12）；每个 (队列, N) 组合做 $\geqslant 100$ 次子抽样重复；总计 $\geqslant 3{,}000$ 次 DE 运行。
6. **可复现性**：全部脚本（R + Snakemake 或纯 R 驱动脚本）、随机种子、会话信息（`sessionInfo()`）与中间结果的校验和开源于 GitHub，第三方 `git clone` 后可一键重跑降规模版本（$\leqslant 4$ 小时）。

5 · 数据与工具

用途	来源 / 工具
统一处理的基因水平计数矩阵（首选，**避开比对算力墙**）	recount3（https://rna.recount.bio/ ，R 包 `recount3` ）。覆盖 SRA/GTEx/TCGA 共约 75 万条人鼠 RNA-seq run，全部用 Monorail 统一处理到 Gencode v26 基因水平。单个研究的 `RangedSummarizedExperiment` 约 60k 基因 × 样本数，300 样本时对象约 150–250 MB，16 GB 内存充裕
备用统一计数源	ARCHS4（https://maayanlab.cloud/archs4/ ，HDF5 单文件，人类版约 30–40 GB，**需核实当前版本体积**；可用 `rhdf5` 按样本切片读取，不必整文件载入）
原始 GEO 队列与元数据	GEO（https://www.ncbi.nlm.nih.gov/geo/ ），用 `GEOquery` 抓取 GSE 的样本表型表。候选疾病与队列须在第 3 周锁定；候选之一为特发性肺纤维化（IPF）肺组织，已知候选 GSE150910、GSE134692、GSE52463——**三者的确切样本量与是否提供计数矩阵需核实**
受控访问核查	recount3 中 TCGA 与 GTEx 的**基因水平计数为开放数据**；dbGaP 中的个体水平基因型、BAM/CRAM 与部分 GTEx 层级为**受控访问**，中学生无法申请，本项目**一律不使用**。所有选入队列必须能匿名下载，无需 DAC 审批
差异表达	`DESeq2` 、`edgeR` （quasi-likelihood）、`limma-voom` 。主分析用 limma-voom（单次运行秒级），DESeq2 用于校验与部分复核
批次与潜变量	`sva` 、`RUVSeq` 、`limma::removeBatchEffect`
富集分析（次要终点）	`clusterProfiler` + MSigDB Hallmark（50 条基因集，规模可控）
对照基准	Gerstner 等 2025 公布的队列内可重现性曲线；各原始论文发表的差异基因表。**仅用于校验与对比，不计入本项目的数据贡献**
算力口径	纯 CPU。limma-voom 单次（60k 基因 × 24 样本）< 5 s；DESeq2 单次 20–60 s。3,000 次 limma-voom ≈ 4 小时；若全用 DESeq2 ≈ 25–50 小时（可后台分夜跑）。**从 fastq 重头比对为超出范围**：300 样本 × 约 5 GB fastq ≈ 1.5 TB 下载，STAR 人类索引需约 32 GB 内存，HISAT2 单样本 4 核约 30–60 分钟，合计 150–300 机时——16 GB 笔记本不可行，方案中不设此路径

6 · 方法路径

1. 装环境（R 4.4 + Bioconductor），跑通 `recount3` 与 `DESeq2` 内置示例，**先完成验收标准第 1 条的三篇论文复现**，把重叠率与相关系数写入 `validation/` 目录。
2. 按预先写死的筛选规则锁定疾病与 5 – 7 个队列（规则：同一组织、同为 polyA 或同为 rRNA-depleted 文库、病例/对照各 ⩾ 15、计数矩阵可从 recount3 或 GEO 补充文件直接取得）。规则写在方案里，**不得看到结果后再改**。
3. 核实每个队列的批次-分组混杂情况与平台差异，出具 PCA 图与 `num.sv` 表；剔除不可去混杂队列，剔除理由逐条记录。
4. 建立两条并行的重抽样实验：(a) 队列内子抽样——在最大的单个队列内抽 N 例/组，重复 100 次；(b) 跨队列复制——在队列 i 抽 N 例/组、在队列 j 抽 N 例/组，对所有队列对重复 100 次。两条用完全相同的 DE 设定。
5. 计算每对结果的 Jaccard 指数与前 200 基因秩相关，用自助法给 95% 置信区间，画出重叠率 – N 曲线（两条并列），并拟合达到 0.5 所需的 N。
6. 检验 H2：把队列内秩变换/批次校正加入管线，重跑跨队列分支，量化差距缩小幅度。
7. 独立交叉校验：换一个完全不同的统计框架（edgeR quasi-likelihood 与非参数的 Wilcoxon 秩和 + BH）重算主曲线，验证结论不依赖于 limma-voom 的模型假设。

7 · 新颖性边界

> 本课题**不**声称提出新的差异表达方法，**不**声称发现新的疾病生物学，**不**把任何单个队列的差异基因列表当作本项目的发现。"小样本 RNA-seq 可重现性低"这一定性结论已由 Gerstner 等（2025）、Li 等（2022）发表，学生不得声称为自己的发现。
>
> 已有工作具体完成了什么：Gerstner 等在 18 个数据集上做了约 18,000 次**同一数据集内部**的子抽样配对比较，得到队列内可重现性随 N 的曲线与一套自助法估计程序；他们的比较对象始终是同源样本。
>
> 本项目的贡献（且是主结论）：把评价维度从"队列内可重现性"换成"**跨独立队列复制率**"，并首次给出同一疾病下这两条曲线的**并列差距**，即量化"同队列子抽样对真实复制率的高估幅度"。这与研究手册里"随机划分 vs 时间划分对照量化高估幅度"是同一类贡献。
>
> 为什么有价值：读者引用小样本 RNA-seq 结论时，真正需要的是"另一个实验室能否重复"，而现有可重现性估计给的是同源样本的上界。给出这个差距系数，等于给现有文献配一把折扣尺。
>
> 风险声明：**"两条曲线差异不显著"同样是有效结论**，但必须给出自助法误差棒，证明本设计在 N=5 时有能力分辨 0.10 的重叠率差异（预实验需先做这个功效检查）。

8 · 决策门槛（go / no-go）

- **第 4 周末**：确认能否找到满足筛选规则的 ⩾ 5 个独立队列。若合格队列 < 5，**立即降级**：其一，把疾病换成公开队列最多的方向（结直肠癌肿瘤/癌旁配对，recount3 中 SRA 项目数量级更大）；其二，把"独立队列"的定义放宽为"同一疾病的不同组织采样部位"，并在论文中显式标注这一降级及其对结论解释的限制。两条降级路径都保留"两条曲线并列 + 差距量化"的主结论框架。
- **第 8 周末（硬门槛）**：验收标准第 1 条必须通过。若 3 篇复现中 ⩾ 2 篇重叠率 < 70%，**立即降级**：改用作者已公开完整分析代码的队列（如带 GitHub 仓库的论文），把校验目标从"复现论文结果"改为"复现论文代码的输出"，阈值提高到 ⩾ 90%。主结论框架不变。
- **第 20 周末**：若跨队列曲线的自助法置信区间宽度 > 0.20（即无法分辨假设中的 0.10 差异），**立即降级**：把队列对数从"所有两两组合"扩到"含重复抽样的更大配对池"，或把 N 网格收缩到 {5, 12, 20} 三点以把重复次数从 100

提到 300。

- **第 36 周**：结果冻结，进入英文写作与查重。第 44 周完成英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) 候选 GSE 是否提供基因计数矩阵而非仅提供受控访问的原始数据；(b) recount3 对目标 SRA 项目的覆盖情况（并非所有 GEO 研究都在 recount3 中）；(c) 各队列的文库类型是否可比。三项必须在第 3 周前查清。
- **选择前提**：适合已能独立写 R 脚本、且愿意接受"结论可能是'差距不显著'"的学生。本路线几乎不可能完全失败——即使 H1 与 H2 都被否证，两条并列曲线本身就是可发表的交付物。
- **预算裁剪顺序**：先砍富集分析分支（次要终点），再砍 edgeR 交叉校验（保留 Wilcoxon），最后砍 N 网格点数。砍到只剩 limma-voom + 3 个 N 点时，主结论仍成立。

B02 · AlphaFold DB 的 pLDDT 作为内在无序预测器：按分类群与同源序列深度分层，检验"低置信度"混淆了真无序与预测不确定

`优先级：高` · `分族：结构与序列分析` · `参赛子类：结构生物信息学` · `资源需求：纯 CPU 笔记本，数据 < 10 GB，不跑任何结构预测推理` · `技能取向：Python + 统计评估`

1 · 研究问题

把 AlphaFold DB 中的 pLDDT（predicted local distance difference test）当作内在无序区（intrinsically disordered region, IDR）预测器时，其性能在不同分类群（taxon）与不同同源序列深度（以 UniRef50 簇大小为代理）之间是否存在系统性偏置？若存在，"低 pLDDT" 在多大程度上是在报告**真实无序**，又在多大程度上只是在报告**多序列比对（MSA）太浅导致的预测不确定**？

2 · 研究背景与空白

背景。 AlphaFold2 对每个残基输出一个 0 – 100 的 pLDDT 置信分。作者在原文中即指出低 pLDDT 区域与内在无序区高度相关，随后 `1 − pLDDT/100` 被广泛当作免费的无序预测器使用，并被整合进 AlphaFold-disorder 等工具（同时使用 pLDDT 与相对溶剂可及性 RSA）。内在无序蛋白在信号转导、相分离、转录调控中至关重要，因此这个"免费预测器"被大量下游研究直接采信。

已有工作到哪一步。 Piovesan / Necci 等的 AlphaFold-disorder 工作报告 pLDDT 在 CAID（Critical Assessment of protein Intrinsic Disorder）的 PDB-DisProt 参照集上表现优异，加入 RSA 后成为该基准上准确率最高的方法之一；但同一批评估明确指出，在 Disorder-NOX 参照集上 pLDDT 与 RSA 都**无法正确优先排序无序位点**。CSBJ 2023 的系统比较进一步发现，用于识别"完全无序蛋白"时 pLDDT **系统性欠预测**，总是预测出一些残余结构。CAID 第二轮（Proteins, 2023）给出了跨方法的完整排名。

空白在于：所有这些评估都把 pLDDT 当成一个整体性能数字来报，没有沿"这条蛋白在序列数据库里有多少同源序列"这一维度做分层。 而 AlphaFold2 的置信度在机制上依赖 MSA 深度——同一条蛋白若同源序列稀少，pLDDT 会因信息不足而降低，这与"该区域真的无序"是两件事，但在 `1 − pLDDT` 的用法里被混为一谈。CAID 的参照集又以人类与模式生物的热门蛋白为主，同源序列普遍很深，因此这个混淆在现有基准上**被系统性掩盖**。这个空白适合学生课题：所有数据可匿名下载、算力需求近乎为零、结论正负都成立。

3 · 可检验假设

- H1：把评估蛋白按 UniRef50 簇大小分成四分位组后，pLDDT 作为无序预测器的 ROC-AUC 在最深四分位与最浅四分位之间的差值 $\geqslant 0.05$，且经 1,000 次自助法重抽样后 95% 置信区间不跨 0；方向为"同源越浅、AUC 越低"。
- H2（机制/备择假设）：这一偏置主要作用于**假阳性**而非假阴性——在浅同源组中，被标注为有序（ordered）却被 pLDDT 判为无序的残基比例，比深同源组高 $\geqslant 50\%$（相对值）；而无序残基的召回率在两组间差异 < 0.05。若 H2 成立，说明可用一个依赖 MSA 深度的阈值校正来改善；若 H2 被否证（偏置对称出现在两类错误上），则说明 pLDDT 在浅同源蛋白上整体不可用，这是更强的警示结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建评估脚本，在 CAID2 的 Disorder-PDB 与 Disorder-NOX 两个参照集上重算 `1 − pLDDT` 的 ROC-AUC 与 APS（average precision score），要求与 AlphaFold-disorder 原文/CAID2 官方发布的对应数值偏差 $\leqslant 0.02$（AUC 绝对值）。同时必须重现已发表的定性反差——Disorder-PDB 上 AUC 明显高于 Disorder-NOX。**此项不过关（偏差 > 0.02 或反差方向不符），后续全部分层结论无效。**

2. **统计口径预先写死**：无序残基在多数参照集中是少数类，属不平衡分类，因此**主指标报 PR-AUC（precision–recall AUC）与在固定召回 0.8 处的精确率，不报 accuracy**——原因是负类（有序残基）占比常达 70 – 90%，全判有序即可得到虚高的 accuracy，无法反映方法价值。ROC-AUC 仅作为与文献对齐的辅助指标报出。所有分层间比较用 1,000 次自助法（按蛋白整条重抽样，不按残基，避免同一蛋白内残基的非独立性）给 95% 置信区间；多组比较报 BH 校正后的 FDR。
3. **混杂控制**：分层比较必须同时报"未调整"与"按蛋白长度、无序内容占比、是否有实验结构覆盖三项做分层/回归调整"两套结果。若调整后效应量衰减 > 60%，判定 H1 的表观效应主要由混杂驱动，须在结论中明确写出。
4. **样本量与检验力**：评估集必须 ⩾ 500 条有 AlphaFold 模型且有残基级无序标注的蛋白，每个四分位组 ⩾ 100 条。开工前先做检验力估计：在给定组内蛋白数下，能分辨的最小 AUC 差值是多少（用自助法模拟），写入方案。若最小可分辨差值 > 0.05，必须先扩充评估集。
5. **可复现性**：全部脚本、UniProt/AFDB 的下载时间戳与版本号、每条蛋白的中间特征表开源；提供一份 `≤ 30 分钟` 可重跑的降规模复现包。

5 · 数据与工具

用途	来源 / 工具
残基级无序金标准	DisProt（https://disprot.org/ ，可整库 JSON/TSV 下载，约 2,900 条蛋白、数万条区域标注，**确切数量需核实当前版本**）；CAID2 官方参照集与官方预测结果（https://caid.idpcentral.org/ ）
独立的第二套标签（NMR 化学位移推断）	CheZOD Z-score 数据集（约 1,300 条蛋白，公开于文献补充材料/GitHub）。用作"标签来源不同"的稳健性检验
pLDDT 与结构	AlphaFold Protein Structure Database（https://alphafold.ebi.ac.uk/ ）。按 UniProt accession 逐条取 mmCIF/PDB，B-factor 字段即 pLDDT。3,000 条 × 约 0.3–1 MB ≈ 1–3 GB，REST 逐条下载约 1–3 小时（须加限速与断点续传）。也可用官方 FTP 的按物种打包文件
分类群与序列元数据	UniProt REST API（https://rest.uniprot.org/ ），按 accession 批量取 taxonomy lineage、序列长度、是否有 PDB 交叉引用
同源深度代理变量	UniRef50 簇成员数，经 UniProt REST 的 UniRef 端点按 accession 查询（数千次调用，纯网络开销）。**代理变量的有效性需核实**：应额外用一个小子集（⩽ 200 条）跑 MMseqs2 对 UniRef50 的实际检索得到 Neff，检验两者相关系数 ⩾ 0.6，否则该代理不可用
相对溶剂可及性（复现 AlphaFold-disorder 所需）	DSSP（`mkdssp`）或 `biotite` 的 SASA 计算，纯 CPU，每条蛋白 < 1 s
统计与绘图	Python + `numpy` / `pandas` / `scikit-learn`（PR-AUC、ROC）/ `statsmodels`（BH 校正、回归调整）/ `matplotlib`
对照基准	AlphaFold-disorder、flDPnn、SPOT-Disorder 等在 CAID2 上的**官方发布预测与得分。仅用于校验与对比，不计入本项目的数据贡献**
明确排除的路径	**自行运行 AlphaFold2/3 推理属超出范围**：完整 AlphaFold2 需约 2.2 TB 遗传数据库（BFD 约 1.7 TB）、MSA 检索阶段单条蛋白数小时 CPU，推理阶段需 16 GB 以上显存 GPU。ColabFold 在纯 CPU 上对 300 残基蛋白通常需数小时至十余小时/条，不可用于数千条规模。本项目**只消费 AlphaFold DB 已发布的模型**，此点须在论文方法部分明写

6 · 方法路径

1. 装环境（Python 3.11 + biotite/DSSP），下载 CAID2 参照集与官方得分文件，**先完成验收标准第 1 条的 AUC/APS 复现**，把复现表与文献值并列写入 `validation/` 。

2. 构建评估集：交叉 DisProt/CAID2 蛋白列表与 AlphaFold DB 覆盖，逐条抓取 mmCIF、提取残基级 pLDDT，抓取 UniProt 元数据与 UniRef50 簇大小；剔除无 AF 模型、序列不一致（AFDB 与 DisProt 序列须完全一致）的条目，剔除规则与数量逐条记录。
3. **核实同源深度代理变量的有效性**：在 ⩽ 200 条子集上用 MMseqs2 检索 UniRef50 得到真实 Neff，与簇大小求相关；相关不足则改用 Pfam 家族大小或直接放弃分层维度（见 go/no-go）。
4. 按 UniRef50 簇大小四分位与分类群（人 / 其他真核 / 原核 / 病毒）双向分层，计算每层的 PR-AUC、固定召回 0.8 处精确率与两类错误率，自助法给区间。
5. 做混杂调整：以残基级 logistic 回归（无序标签 ~ pLDDT + 蛋白长度 + 无序内容 + 有无 PDB 覆盖 + 同源深度分组 + pLDDT×同源深度交互项），报交互项系数与 FDR；交互项显著即为 H1 的模型化证据。
6. 用 CheZOD 这套**标注方式完全不同**的标签重跑第 4、5 步，检验结论不依赖 DisProt 的标注习惯。
7. 独立交叉校验：换一个不依赖 AlphaFold 的经典无序预测器（如 IUPred3，纯 CPU、秒级）做同样的分层评估。若 IUPred3 不表现出同样的同源深度依赖，则该依赖确属 AlphaFold 的 MSA 机制而非评估集人为构造——这是本课题最关键的一次对照。

7 · 新颖性边界

> 本课题**不**声称提出新的无序预测方法，**不**运行任何结构预测推理，**不**声称"pLDDT 与无序相关"这一已被 AlphaFold 原文与 Piovesan 等发表的事实为本项目发现；**也不**声称"pLDDT 在 Disorder-NOX 上表现差"为本项目发现（CSBJ 2023 与 CAID2 已报）。
>
> 已有工作具体完成了什么：AlphaFold-disorder 在 CAID/DisProt-PDB 上报出了 pLDDT 与 pLDDT+RSA 的总体 AUC 排名；CSBJ 2023 报出了在识别完全无序蛋白时的欠预测；CAID2 给出了跨方法的整体排名与两套参照集之间的性能反差。**这三项工作都只报整体数字，评估集以人类与模式生物的深同源蛋白为主。**
>
> 本项目的贡献（且是主结论）：把评价维度从"整体性能"换成"**沿同源序列深度与分类群的分层性能**"，并检验一个可否证的机制命题——低 pLDDT 同时编码"真无序"与"MSA 浅导致的不确定"，两者的混淆程度随同源深度系统变化。第 7 步的 IUPred3 对照使这个命题真正可被否证。
>
> 为什么有价值：大量下游研究（尤其是对非模式生物、环境宏基因组来源蛋白的研究）把 `1 − pLDDT` 当无序注释直接用，而这些蛋白恰恰处在同源最浅的一端。若偏置存在，本项目给出一条"何时不该这么用"的可操作判据；若不存在，则为这种廉价用法提供一次独立背书。
>
> 风险声明：**"分层间无显著差异"同样是有效结论**，但必须由第 4 条的检验力估计证明本设计有能力分辨 0.05 的 AUC 差异，否则该结论无效。

8 · 决策门槛（go / no-go）

- **第 3 周末**：确认 DisProt/CAID2 与 AlphaFold DB 的可交叉条目数。若序列完全一致且有残基级标注的蛋白 < 500，**立即降级**：其一，把标签来源扩到 MobiDB 的"共识无序"注释（规模大一个数量级，代价是标签质量下降，须在论文中标注为弱标签）；其二，把残基级评估改为**蛋白级**评估（是否为"高无序含量蛋白"的二分类），此时可用蛋白数上万。两条降级都保留"分层 + 混杂调整 + IUPred3 对照"的主结论框架。
- **第 6 周末（硬门槛）**：验收标准第 1 条必须通过。若自建脚本的 AUC 与官方值偏差 > 0.02，先排查残基映射与缺失值处理；两周内仍不通过则**降级**为直接使用 CAID2 官方发布的 pLDDT 预测得分文件（放弃自建特征提取，把方法学校验改为"官方得分复算一致性 ⩾ 0.99"），分层分析照常进行。
- **第 10 周末**：核实同源深度代理变量。若 UniRef50 簇大小与 MMseqs2 实测 Neff 的相关 < 0.6，**立即降级**：改用"是否属于模式生物 / 是否有 PDB 实验结构覆盖 / Pfam 家族大小"三个可直接查得的分层变量替代，主结论从"同源深度依赖"改述为"**注释充分度依赖**"，框架与全部分析步骤不变。

- **第 30 周**：若 H1 与 H2 都未达阈值，按风险声明写"无显著偏置"结论，并把检验力估计作为核心证据放进正文。第 36 周结果冻结，第 44 周英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) AlphaFold DB 的 REST 逐条下载是否有速率限制与配额；(b) DisProt 序列版本与 AFDB 所用 UniProt 版本是否一致（版本漂移会造成残基编号错位，这是本课题最隐蔽的坑）；(c) CAID2 官方得分文件的格式与缺失值约定。三项必须在第 5 周前查清。
- **选择前提**：适合数学与统计感觉较好、能耐心处理序列坐标映射的学生。本路线不需要任何湿实验与大算力，但对"评估设计的严谨性"要求高——这也正是它在评委面前的说服力来源。
- **预算裁剪顺序**：先砍 CheZOD 第二套标签（保留 DisProt 主线），再砍分类群维度（只保留同源深度维度），最后砍混杂回归（只留分层比较）。砍到只剩"同源深度四分位 + PR-AUC + IUPred3 对照"时，主结论仍成立。

B03 · 单基因条形码物种界定方法间的不一致度能否被数据集属性预测：跨 40 个公开 COI 数据集的元分析与一条"何时可信"的可操作判据

优先级：高 · 分族：系统发育与进化基因组学 · 参赛子类：分子系统学 / 生物信息学 · 资源需求：纯 CPU 笔记本，每数据集 ≤ 3,000 条序列 · 技能取向：R + 命令行系统发育工具链

1 · 研究问题

在 40 个及以上从 BOLD/GenBank 公开获取的 COI（细胞色素 c 氧化酶亚基 I）条形码数据集上，用统一管线跑四种主流物种界定（species delimitation）方法所得的分子操作分类单元（MOTU）数量之间的不一致度，能否由数据集本身在分析前就可测得的属性（每形态种序列数、序列长度覆盖、地理跨度、种内/种间距离间隙的宽度）线性预测？能否据此给出一条"当属性 X 低于阈值时四方法必然分歧、不应据单基因下分类学结论"的可操作判据？

2 · 研究背景与空白

背景。 DNA 条形码用一段标准化基因（动物用 COI 约 658 bp）把未知个体归入物种。当形态学信息不足时，研究者用物种界定算法直接从序列推断"有几个物种"。主流方法分两类：基于遗传距离的 ABGD / ASAP（Assemble Species by Automatic Partitioning）、BIN，以及基于系统树的 GMYC（Generalized Mixed Yule Coalescent）、PTP/bPTP/mPTP。它们的理论前提不同，因此在同一数据上给出不同答案。

已有工作到哪一步。 大量单类群论文报告了这种分歧的量级：中国螳螂目（2025）用 ASAP / jMOTU / bPTP / GMYC 分别得到 32 / 58 / 68 / 60 个 MOTU；螽斯属 Gampsocleis 分别得到 6 / 13 / 10 / 23 个；斯里兰卡 Sericini 金龟（2022）明确报告"多种界定方法彼此拟合很差，且与形态种拟合也差"；而石首鱼科 Stelliferinae（2023）中 ABGD/GMYC/bPTP 却高度一致（30 – 32 个 MOTU，多数与有效种吻合）。**也就是说，分歧有时极大、有时极小，这一点在文献里是公开的。**

空白在于：没有人把这几十个各自独立的单类群结果放到统一管线下重做，并检验"分歧大小是否可由数据集属性预先预测"。现有文献的结论停在"结果依类群而异，需结合形态"，这是一句无法执行的建议。研究者真正需要的是：拿到一个新数据集时，能否在跑方法之前就判断结果可不可信。这个空白适合学生课题：数据全部匿名公开、每个数据集算力在分钟级、方法完全公开、**结论正负都成立**（"不可预测"本身就是对该领域的重要警示）。

3 · 可检验假设

- H1：以"四方法 MOTU 数的变异系数（coefficient of variation, CV）"为不一致度指标，用数据集属性做多元回归，跨 40 个数据集的留一交叉验证 $R^2 \geqslant 0.35$；其中"每形态种平均序列数"与"种内/种间距离间隙宽度"两个变量的回归系数在 BH 校正后 FDR < 0.05。
- H2（机制/备择假设）：不一致主要来自**采样密度**而非真实的分类学复杂度——若对每个数据集把每形态种的序列数随机稀释到 3 条，四方法的 CV 应普遍上升且跨数据集差异缩小（组间方差下降 ⩾ 30%）。若稀释后组间差异不缩小，说明不一致由类群固有的进化速率异质性驱动，采样量再大也无法消除，这是更强的结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：从上述文献中选 3 个已公开完整序列登录号与 MOTU 结果的数据集（候选：中国螳螂目、Gampsocleis、Stelliferinae——**三者补充材料是否给出可复原的登录号列表需核实**），用自建管线在**原文所述参数**下重跑，要求每个方法的 MOTU 数与原文报告值的相对偏差 ⩽ 15%，且系统发育树上原文强调的关键分支（原文标注支持率 ⩾ 0.95 的节点）在本项目重建树中的超快自展（UFBoot）支持率 ⩾ 90%。**3 个数据集中 ⩾ 2 个不达标，则管线不可信，后续全部元分析结论无效。**

2. **统计口径预先写死**：回归的显著性一律报 BH 校正后的 FDR，不报裸 p 值（40 个数据集 × 多个候选属性构成多重检验）。R^2 必须报**留一交叉验证 R^2**而非拟合 R^2，并明确写出两者差值（过拟合幅度）。所有数据集级统计量给 1,000 次自助法（按数据集整体重抽样）95% 置信区间。样本量仅 40，因此在开工前先做效应量与检验力估计：在 n=40、5 个预测变量下，能以 80% 检验力检出的最小 R^2 是多少（预期约 0.30，**须自行用模拟核算并写入方案**）；若最小可检出 R^2 高于 H1 阈值，必须先把数据集数扩到 60。
3. **分类学标签的地位**：BOLD/GenBank 中的形态种名**仅作为对照标签使用，不计入本项目的数据贡献**；且必须报告标签本身的不确定性（BOLD 记录的鉴定者与鉴定等级字段），把"无鉴定者信息"的记录比例写入每个数据集的元数据表。
4. **规模下限**：≥ 40 个数据集，每个数据集 ≥ 8 个形态种且 ≥ 60 条序列；覆盖 ≥ 4 个动物门/纲，避免结论只适用于昆虫。
5. **主观判断的可复现化**：数据集的纳入/排除规则、序列修剪规则（比对两端缺失比例阈值）、异常序列（可能的 NUMT 或污染）的剔除判据，必须全部写成可执行的数值规则并随代码发布，**不得人工逐条挑选**。
6. **可复现性**：Snakemake 或 Makefile 驱动的完整管线、全部登录号清单、随机种子、各工具版本号开源；提供一个 5 数据集的降规模版本，第三方在 ≤ 2 小时内可重跑。

5 · 数据与工具

用途	来源 / 工具
COI 条形码序列与形态种标签	BOLD Systems 公开数据门户（https://www.boldsystems.org/ ），支持按分类群批量下载 TSV/FASTA（含 BIN、鉴定者、采集经纬度）。**BOLD v5 的批量下载接口与配额需核实**；备用入口为 GenBank/NCBI Nucleotide 按 `COI[Gene] AND <taxon>[Organism]` 检索 + `efetch`
数据集属性中的地理跨度	BOLD 记录自带经纬度字段；无坐标记录单独统计比例
序列比对	MAFFT（`--auto`）。1,000 条 × 658 bp 约 10–30 s；3,000 条约 2–5 min。COI 为编码基因，另用 MACSE 或按密码子比对做一次一致性检查
遗传距离	R 包 `ape`（K2P 与 p-distance 两种，均报）
距离法界定	ASAP（网页服务与本地二进制均有，1,000 条秒级）、ABGD（同级）
树法界定	IQ-TREE 2 + ModelFinder + 1,000 次 UFBoot。500 条 × 658 bp、4 线程约 2–8 min；3,000 条约 1–3 小时。mPTP（比 bPTP 快 1–2 个数量级，1,000 条分钟级）；GMYC 用 R 包 `splits`，需**超度量树（ultrametric tree）**
超度量化	`ape::chronos`（惩罚似然，秒–分钟级）。**BEAST2 严格钟属超出范围**：300 条序列、10^7 步 MCMC 在 4 核 CPU 上通常需 1–3 天/数据集，40 个数据集不可行；方案中改用 chronos 并在论文中明确标注这一近似及其对 GMYC 结果的可能影响（须用 3 个小数据集做 chronos vs BEAST2 的对照，量化 MOTU 数差异）
元分析统计	R：`lm` / `glmnet`（含正则化以防 n=40 过拟合）、`boot`、`p.adjust(method="BH")`
对照基准	上述四篇已发表论文的 MOTU 计数与树拓扑。**仅用于校验与对比，不计入本项目的数据贡献**
伦理	全部为已公开的序列数据库记录，无人体数据、无活体动物操作、无需伦理审批

6 · 方法路径

1. 装环境（MAFFT / IQ-TREE 2 / mPTP / ASAP / R+ape+splits），跑通各工具自带示例，**先完成验收标准第 1 条的三数据集复现**，把 MOTU 数与关键分支支持率的复现表写入 `validation/`。
2. 按预先写死的数值规则从 BOLD 抓取候选数据集（≥ 8 形态种、≥ 60 序列、≥ 500 bp 有效长度），生成 ≥ 40 个数据集的清单与元数据表；纳入/排除全过程留日志。

3. 对每个数据集计算 6 – 8 个**分析前可得**的属性：每形态种平均与最小序列数、有效比对长度、缺失率、地理跨度（最大成对大圆距离）、种内最大 K2P 距离的中位数、种内-种间距离间隙宽度（barcode gap）、形态种数、单序列种（singleton）比例。
4. 统一管线跑四种界定方法，记录每数据集的四个 MOTU 数、CV，以及与形态种划分的匹配指标（调整兰德指数 ARI 与匹配比例）。
5. **核实工具能力边界**：在 3 个小数据集上做 chronos vs BEAST2 超度量树的 GMYC 结果对照（BEAST2 只跑这 3 个，成本可控），量化近似带来的 MOTU 数偏差并写入限制条款。
6. 做元回归（属性 → CV），报留一交叉验证 R^2、系数与 BH-FDR、自助法区间；按 H1 阈值判定，并把回归转成一条阈值判据（例如"每形态种序列数 < 4 且 barcode gap < 2% 时，四方法 CV 中位数 > 0.4，不应据单基因下结论"）。
7. 检验 H2 的稀释实验；最后做独立交叉校验——用一套**完全不参与建模**的 10 个新数据集做外部验证，报判据在外部集上的命中率。

7 · 新颖性边界

> 本课题**不**提出新的物种界定算法，**不**声称发现任何新物种，**不**对任何类群做分类学修订，**不**声称"不同界定方法结果不一致"为本项目发现——这一点已由斯里兰卡 Sericini（2022）、中国螳螂目（2025）等多篇论文明确报告。丘奖 2021 年金奖论文《A molecular phylogeny of cavernicolous Oniscidea…》做的是**具体类群的系统发育重建与新种描述**，本课题与之路径完全不同：本项目不描述新种、不做形态学工作，交付物是跨类群的方法学判据。
>
> 已有工作具体完成了什么：各单类群论文在自己的数据上并列跑 2 – 5 种方法，报出各自的 MOTU 数与和形态种的吻合度，结论止于"依类群而异，建议结合形态学"。**它们没有统一管线、没有跨数据集比较、没有把不一致度当成因变量去建模**。
>
> 本项目的贡献（且是主结论）：把评价维度从"某类群应界定出几个物种"换成"**方法间不一致度本身是否可预测**"，并交付一条可在分析前使用的定量判据 + 外部验证命中率。这属于研究手册所列的"评价维度转换 + 方法层面的可复现性贡献"，在计算生物学中是被承认的贡献类型，但定位必须讲清楚，否则会被误读为重复劳动。
>
> 为什么有价值：条形码数据每年以百万条增长，绝大多数新数据集不会有配套形态学工作。一条"什么时候单基因结论不可信"的先验判据，直接影响这些数据的使用方式。
>
> 风险声明："**属性无法预测不一致度**"（R^2 接近 0）**同样是有效结论**，它意味着不一致源于不可事前观测的因素，从而否定"用简单指标筛数据集"的做法。但必须由第 2 条的检验力估计证明本设计（n=40）有能力检出 R^2=0.35，否则该结论无效。

8 · 决策门槛（go / no-go）

- **第 4 周末**：确认能否从 BOLD/GenBank 稳定批量抓到 ⩾ 40 个合格数据集。若合格数据集 < 25，**立即降级**：其一，放宽形态种下限到 6 并把类群范围扩到植物 rbcL/matK 与真菌 ITS（代价是需为每个标记重设距离阈值，须在论文中分标记报结果）；其二，把因变量从"四方法 CV"改为"两方法（ASAP vs mPTP）的 ARI 不一致度"，此时可用数据集数量大幅增加。两条降级都保留"属性 → 不一致度回归 + 外部验证判据"的主结论框架。
- **第 8 周末（硬门槛）**：验收标准第 1 条必须通过。若 ⩾ 2 个复现数据集的 MOTU 数偏差 > 15%，先排查比对与修剪参数；四周内仍不通过则**降级**：把校验目标改为"复现原文**树拓扑的关键分支**与 barcode gap 分布"（这两项对参数远不敏感），并在论文中明确写出 MOTU 数未能复现的事实——**这个失败本身就是研究内容**，它直接支持本课题"方法结果不稳定"的动机。
- **第 14 周末**：完成 chronos vs BEAST2 对照。若 GMYC 的 MOTU 数在两种超度量化下相对偏差 > 25%，**立即降级**：把 GMYC 从四方法中剔除，主分析改为 ASAP / ABGD / mPTP 三方法，并把"GMYC 对超度量化方式高度

敏感"作为一条独立结论报出。主结论框架不变。

- **第 26 周**：完成主回归。若留一 $R^2 < 0.15$，按风险声明写"不可预测"结论，并把检验力估计与外部验证的失败率作为核心证据。
- **第 36 周**结果冻结，第 44 周英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) BOLD 批量下载的接口、配额与许可条款（能否在论文中再分发序列 ID 列表）；(b) 候选复现论文的补充材料是否给出可复原的登录号；(c) IQ-TREE 2 在 3,000 条序列上的内存峰值（须实测，16 GB 是否够）。三项在第 6 周前查清。
- **选择前提**：适合愿意花时间打磨命令行管线、并接受"主结论可能是一条否定判据"的学生。这条路线的最大风险是数据清洗工作量（污染序列、NUMT、错误鉴定），须把清洗规则当成研究内容而非杂活来写。
- **预算裁剪顺序**：先砍地理跨度属性（BOLD 坐标缺失率高），再砍稀释实验（H2），再砍外部验证集（从 10 个降到 5 个）。砍到只剩"30 个数据集 + 三方法 + 五个属性回归"时，主结论仍成立。

B04 · 随机基因签名的预后显著性在 TCGA 泛癌中的分布：以增殖元基因校正为轴，检验已发表预后签名相对于随机零模型的增量价值

优先级：高 · 分族：公开组学再分析 · 参赛子类：癌症基因组学 / 生物统计 · 资源需求：纯 CPU 笔记本，主表 ≤ 1 GB，全部为开放访问数据 · 技能取向：Python/R 生存分析 + 大规模重抽样

1 · 研究问题

在 TCGA 泛癌（Pan-Cancer Atlas）的 ⩾ 15 种癌型中，一个随机抽取的 100 基因签名与患者生存显著相关的概率各是多少？这个"随机显著率"是否可由该癌型中增殖元基因（meta-PCNA）自身的预后强度预测？在此零模型下，已发表的预后基因签名中有多大比例的预后价值超出随机水平，且在扣除增殖信号后仍然成立？

2 · 研究背景与空白

背景。 "发现一组基因签名可预测某癌症预后"是肿瘤基因组学产出量最大的论文体裁之一，每年数以千计。其统计基础是：把签名基因的表达汇总成一个分数，按中位数分高低两组做 log-rank 检验或 Cox 比例风险回归。问题在于，肿瘤转录组存在一个占据大量方差的主轴——细胞增殖，它本身就与预后强相关，因此任何与增殖相关的基因组合都会"显著"。

已有工作到哪一步。 Venet、Dumont 与 Detours（PLOS Computational Biology, 2011）在乳腺癌中给出了决定性结果：**任意随机抽取的 ⩾ 100 个基因，有 90% 的概率与预后显著相关**；47 个已发表乳腺癌签名中有 28 个（60%）的预后强度不强于随机预期；他们用"meta-PCNA"增殖元基因做校正后，多数签名基因的预后价值消失。另一条线是 Ramaker 等（Oncotarget, 2017）在 19 种癌症上做增殖分析，识别出一批"增殖信息型（proliferation-informative）"癌症并给出共同生存签名。

空白在于：Venet 的随机零模型只在乳腺癌一种癌型上建立过，而它的核心量——"随机签名显著率"——从未在泛癌尺度上被系统标定。 Ramaker 等虽然做了泛癌增殖分析，但其目标是**构建**签名，不是**证伪**签名；他们没有做随机零模型。于是今天研究者在肺癌、肝癌、胶质瘤上发表签名时，手边并没有对应癌型的"随机基线"可比。这个空白适合学生课题：TCGA 表达与临床结局均为开放访问数据、算力在小时级、方法学完全公开、**结论正负都成立**。

3 · 可检验假设

- H1：随机 100 基因签名在各 TCGA 癌型中的显著率（Cox 单变量、FDR < 0.05）跨癌型跨度极大，最高与最低癌型之差 ⩾ 40 个百分点；且该显著率与该癌型 meta-PCNA 分数自身的 Cox 风险比（|log HR|）的 Spearman 相关 ⩾ 0.6（FDR < 0.05）。
- H2（备择/机制假设）：在按 meta-PCNA 分数校正后，随机签名的显著率在"增殖信息型"癌型中下降 ⩾ 50%（相对值），但在非增殖信息型癌型中下降 < 20%——即增殖并非所有癌型预后信号的共同来源。若 H2 被否证（所有癌型下降幅度相近），则说明存在一个跨癌型通用的混杂主轴，需要另找解释变量（如免疫浸润或肿瘤纯度），这本身是更有价值的发现方向。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建管线在 Venet 等 2011 使用的乳腺癌设定下复现其三项核心已发表结果：(a) 随机 100 基因签名的显著率 ⩾ 85%（原文报 90%）；(b) 复现原文点名的至少 5 个已发表乳腺签名的预后 p 值方向与量级（log10 p 的偏差 ⩽ 0.5）；(c) 复现 meta-PCNA 校正后显著性消失的定性结论。**三项中任一不达标，管线不可信，全部泛癌结论无效。**（若原文所用的 NKI/公开乳腺队列不可得，则在 TCGA-BRCA 上复现，并明确标注这是"设定迁移"而非严格复现。）

2. **统计口径预先写死**：(a) 生存终点严格按 Liu 等（Cell, 2018）TCGA 泛癌临床数据资源的**官方推荐**逐癌型选择——多数癌型用 PFI（progression-free interval）而非 OS（overall survival），事件数少的癌型不得用 OS，选择理由逐癌型写入表格；(b) 所有 p 值报 Benjamini – Hochberg FDR，不报裸 p 值；(c) 每个癌型必须报事件数与按 10 events-per-variable 规则计算的可容纳协变量上限，事件数 < 30 的癌型一律剔除并写明；(d) Cox 模型必须做比例风险假设检验（`cox.zph` / Schoenfeld 残差），违反者改用限制平均生存时间（RMST）差作为备用终点并两者都报；(e) 分组比较若涉及不平衡二分类的判别性能，报 PR-AUC 与 C-index，**不报 accuracy**，原因是事件率在多数癌型低于 30%，全判"无事件"即可得虚高 accuracy。
3. **批次效应必须检查并写明**：使用 Pan-Cancer Atlas 的 EB++ 批次校正版表达矩阵；同时用未校正版做一次敏感性分析，报告两版之间随机显著率的差值。肿瘤纯度（来自公开的 ABSOLUTE/共识纯度表）作为协变量做一次调整分析。
4. **规模下限**：⩾ 15 个癌型，每个癌型 ⩾ 200 例且事件数 ⩾ 30；每个癌型 ⩾ 5,000 次随机签名抽样（签名长度网格 {10, 50, 100, 250, 500}）；已发表签名 ⩾ 60 条，覆盖 ⩾ 6 个癌型。
5. **签名来源的地位**：所有已发表签名的基因列表**仅作为对照与评估对象，不计入本项目的数据贡献**；每条签名必须记录原文出处、癌型、原报告的效应量，形成可核查的附表。
6. **可复现性**：随机种子、癌型-终点对照表、全部脚本与结果 CSV 开源；提供 3 癌型的降规模复现包（⩽ 1 小时可跑完）。

5 · 数据与工具

用途	来源 / 工具
泛癌表达矩阵	UCSC Xena，TCGA Pan-Cancer Atlas 的 EB++AdjustPANCAN 基因表达（RSEM 归一化，约 20,500 基因 × 11,000 样本，压缩包数百 MB 至 1 GB 级；https://xenabrowser.net/datapages/ ）。float32 存 20.5k×11k 约 0.9 GB，16 GB 内存下必须**按癌型切片处理**，不整表载入 pandas
生存终点与临床变量	TCGA-CDR（Liu et al., Cell 2018，"An Integrated TCGA Pan-Cancer Clinical Data Resource"）的补充表，含 OS/DSS/DFI/PFI 四个终点及**逐癌型的终点使用推荐**。此表是本课题统计口径的权威依据
肿瘤纯度	TCGA 公开的共识纯度估计表（Aran et al. 2015 等），用于协变量调整
已发表预后签名	(a) MSigDB C2:CGP 基因集合中的预后相关集合；(b) 手工整理近五年针对目标癌型的"novel prognostic signature"论文（每条记录出处、基因列表、原报告 HR），目标 ⩾ 60 条
meta-PCNA 增殖元基因	Venet 等 2011 补充材料给出的 meta-PCNA 基因列表；备用为 MSigDB Hallmark 的 `E2F_TARGETS` 与 `G2M_CHECKPOINT`
受控访问核查	TCGA 的**基因表达（Level 3）与 TCGA-CDR 临床结局表为开放访问**，可直接下载，无需 dbGaP DAC 审批；**个体水平体细胞/胚系变异 VCF、BAM/CRAM 与部分详细临床字段为受控访问，中学生无法申请，本项目一律不使用**。数据已去标识化，年龄 > 89 岁被掩码，不涉及可识别个人信息，无需伦理审批
生存分析	Python `lifelines`（CoxPHFitter、`proportional_hazard_test`）或 R `survival` + `survminer`。单变量 Cox（n≈500）单次拟合约 3–10 ms
算力口径	纯 CPU。15 癌型 × 5 长度 × 5,000 次 = 375,000 次 Cox 拟合，按 8 ms/次约 50 分钟（单核），多核或用 R `coxph` 的向量化更快。表达矩阵按癌型切片后每片约 20.5k × 500，内存 < 100 MB。**全流程可在一个通宵内跑完，无算力风险**
对照基准	Venet 等 2011 的乳腺癌结果、Ramaker 等 2017 的 19 癌型增殖分析。**仅用于校验与对比，不计入本项目的数据贡献**

6 · 方法路径

1. 装环境，下载 Xena 表达矩阵与 TCGA-CDR，**先完成验收标准第 1 条的 Venet 复现**，把三项复现结果与原文数值并列写入 `validation/`。
2. 按 TCGA-CDR 官方推荐为每个癌型锁定生存终点，剔除事件数 < 30 的癌型，生成"癌型 – 终点 – 样本数 – 事件数"对照表（这张表必须在跑任何随机抽样之前冻结）。
3. 在每个癌型上建立随机零模型：按长度网格抽取随机基因集，签名分数取标准化表达的均值，单变量 Cox 拟合，统计 FDR < 0.05 的比例；输出各癌型的随机显著率曲线（显著率 vs 签名长度）。
4. 计算各癌型的 meta-PCNA 分数及其自身 Cox 结果，检验 H1 中的跨癌型相关。
5. 把 ⩾ 60 条已发表签名逐条放入对应癌型，计算其效应量在该癌型随机零分布中的**分位数**（这是本课题的核心量），并报出"超过随机 95 分位"的签名比例。
6. 做 meta-PCNA 校正（把 meta-PCNA 分数作为协变量放入 Cox），重算第 3、5 步，检验 H2；同时做肿瘤纯度调整与未批次校正版的敏感性分析。
7. 独立交叉校验：换一套完全独立的生存框架——非参数 log-rank（按中位数二分）与 RMST 差——重算主结论；再用一个 TCGA 之外的公开队列（如 GEO 上带随访信息的肺腺癌或肝癌表达数据集）复算至少 2 个癌型的随机显著率，验证结论不是 TCGA 特有的技术产物。

7 · 新颖性边界

> 本课题**不**提出新的预后模型，**不**声称发现任何新的癌症生物标志物，**不**声称"随机签名常常显著"与"增殖是主要混杂"为本项目发现——这两点由 Venet、Dumont、Detours（2011）在乳腺癌中确立，学生必须在正文首段明确归属。
>
> 已有工作具体完成了什么：Venet 等在**乳腺癌一种癌型**上建立随机零模型，量化 47 条已发表签名相对随机的增量，并提出 meta-PCNA 校正；Ramaker 等（2017）在 19 种癌症上做增殖分析并**构建**了一个共同生存签名，但没有随机零模型、没有对已发表签名做证伪。
>
> 本项目的贡献（且是主结论）：把 Venet 的零模型从单癌型**扩展到泛癌尺度并转成一张可查的基线表**——给出每个癌型的"随机显著率 vs 签名长度"曲线，使任何新签名都能被放进对应癌型的随机分布里读出分位数。附带的第二个主结论是 H2 所检验的机制命题：增殖是否是所有癌型的共同混杂轴。这属于"新样本的独立检验 + 评价维度转换"的双重定位。
>
> 为什么有价值：过去十五年的预后签名文献几乎全部缺少癌型特异的随机基线。一张可直接查用的基线表把"这条签名是否值得相信"从主观判断变成可计算的分位数。
>
> 风险声明：**"泛癌显著率高度一致、且与增殖强度无关"同样是有效结论**（它会否定 Venet 结论的可推广性，同样重要），但必须给出自助法误差棒，证明本设计有能力分辨假设中的 40 个百分点差异与 0.6 的相关系数。

8 · 决策门槛（go / no-go）

- **第 5 周末（硬门槛）**：验收标准第 1 条必须通过。若 Venet 复现的随机显著率 < 85%，先排查表达值标准化方式与生存终点定义（这两项最常出错）；三周内仍不通过则**降级**：把校验基准换成 Ramaker 等 2017 公布的癌型级增殖 – 生存关联（其补充材料给出逐癌型数值），校验阈值定为 Spearman ⩾ 0.8。主结论框架不变。
- **第 8 周末**：确认满足"⩾ 200 例且事件数 ⩾ 30"的癌型数量。若 < 12 个，**立即降级**：其一，把终点统一放宽到 PFI 并接受随访较短带来的检验力损失，在论文中量化这一损失；其二，把癌型合并为器官系统级分组（消化道 / 泌尿生殖 / 胸部 / 神经系统等 6 组），基线表由"逐癌型"变为"逐系统"。两条降级都保留"随机零模型 + 已发表签名分位数 + 增殖校正"的主结论框架。

- **第 16 周末**：确认已发表签名的收集量。若可完整复原基因列表的签名 < 40 条，**立即降级**：改用 MSigDB C2:CGP 中的策展基因集作为"已发表基因集"的代理，明确标注这是代理而非真实预后签名，并把结论范围相应收窄。
- **第 24 周末**：若 H1 的癌型间跨度 < 20 个百分点且相关系数不显著，按风险声明写"随机显著率跨癌型一致"结论，并把该结论转为一张统一基线表（对使用者同样有价值）。
- **第 36 周**结果冻结，第 44 周英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) Xena 的 EB++ 矩阵与 TCGA-CDR 的样本 ID 命名规则（TCGA barcode 前 12 位 vs 前 15 位）对齐方式，错位会导致全盘错误；(b) Venet 2011 补充材料中的 meta-PCNA 基因列表是否仍可下载；(c) 目标癌型是否存在同一患者多样本（需去重规则）。三项在第 6 周前查清。
- **选择前提**：适合统计基础扎实、能接受"本课题的价值在于给他人的工作降温而非做出新发现"这一定位的学生。答辩时必须能清楚说明为什么"证伪型工作"是正当的科研贡献——这一点须提前准备英文表述。
- **预算裁剪顺序**：先砍 GEO 外部队列复算（第 7 步后半），再砍肿瘤纯度调整，再砍签名长度网格（只保留 100 与 500 两点），最后把癌型数从 15 降到 10。砍到只剩"10 癌型 × 100 基因随机零模型 + 40 条签名分位数 + meta-PCNA 校正"时，主结论仍成立。

B05 · 经典正选择信号在 1000 Genomes 30× 高覆盖重测序数据上的独立复现：以数据版本更替为自变量，量化低覆盖填补对 iHS/XP-EHH 扫描的影响

`优先级：中高` · `分族：群体遗传学` · `参赛子类：群体基因组学 / 计算生物学` · `资源需求：纯 CPU 笔记本，按染色体切片下载 ≤ 40 GB 磁盘，单群体单体型矩阵 < 1 GB 内存` · `技能取向：命令行基因组学（bcftools/selscan）+ Python`

1 · 研究问题

基于 1000 Genomes Project phase 3 低覆盖（平均 7.4×）数据发表的经典正选择（positive selection）信号，在同一批个体的 30× 高覆盖重测序（NYGC，3,202 例）call set 上重跑相同的单体型扫描统计量后，能否重现？重现率与信号强度、等位基因频率谱、填补（imputation）依赖程度之间是什么关系——即"低覆盖 + 统计填补"这一数据生成方式给单体型类选择扫描带来了多大的系统偏差？

2 · 研究背景与空白

背景。 检测人类基因组上的近期正选择，主流方法依赖单体型长度：iHS（integrated haplotype score）、nSL、XP-EHH（cross-population extended haplotype homozygosity）都通过比较携带不同等位基因的单体型衰减速度来识别"频率升高但单体型尚未被重组打断"的位点。这类统计量对**单体型定相（phasing）质量**极其敏感——而定相质量又直接取决于测序覆盖度与填补算法。

已有工作到哪一步。 1000 Genomes phase 3（2015）的 2,504 例基于低覆盖 WGS（平均 7.4×）+ 高覆盖外显子组 + 芯片基因型的组合调用，是过去十年绝大多数人类选择扫描的底层数据；LCT/MCM6（欧洲乳糖耐受）、SLC24A5（欧洲肤色）、EDAR（东亚毛发与汗腺）等经典位点的信号强度都是在这套数据上标定的。2022 年 Byrska-Bishop 等（Cell）发布了同一批个体的 30× 高覆盖重测序（3,202 例，含 602 个三口之家），明确报告相对 phase 3 在**罕见 SNV、INDEL 与结构变异**上灵敏度与精确度显著提升。此外 2019 年已有工作把 phase 3 重新比对/调用到 GRCh38 上，使两版可在同一坐标系比较。

空白在于：数据版本更新了，但没有人系统检验过基于旧版数据的选择扫描结论在新版上是否成立。 Byrska-Bishop 等评估的是变异调用层面的灵敏度/精确度，不是下游群体遗传推断的稳定性；而选择扫描依赖的是单体型结构，恰恰是低覆盖填补最可能引入系统性伪迹的地方。这是研究手册中最稳的差异化轴——**换样本（此处是换同一样本的新数据版本）做独立检验**：方法完全公开可复现、算力门槛可控、**结论正负都成立**（重现则为经典结论提供一次独立支持；不重现则指出一整类文献的数据依赖性）。

3 · 可检验假设

- **H1**：在 30× 数据上重跑 iHS 后，phase 3 中排名前 100 的候选窗口（每群体）有 ⩾ 80% 仍落在 30× 结果的前 1% 窗口内；但重现率随信号强度单调下降——phase 3 排名 500 – 1000 的窗口重现率 ⩽ 50%，两档差值的 95% 自助法置信区间不跨 0。
- **H2**（机制假设）：不重现的窗口在两版之间的**定相开关错误率（switch error）代理指标**（用 602 个三口之家的孟德尔一致性/单体型长度分布推得）显著高于重现窗口；且不重现窗口富集于低重组率区域与低等位基因频率区间（0.05 < MAF < 0.15）。若 H2 被否证（不重现窗口无系统特征），说明差异是随机噪声，经典结论的稳健性反而更强，这同样是有效结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：先在 **phase 3 数据本身**上重跑扫描，要求复现已发表的经典结果：(a) CEU（欧洲裔）中 LCT/MCM6 区域与 SLC24A5 区域落入 iHS |score| 的全基因组前 0.1% 窗口；(b) CHB/CHS（东亚）中 EDAR 区

域落入前 0.1%；(c) 与已发表的 1000 Genomes Selection Browser / Voight 等公布的 iHS 值在共有 SNP 上的 Spearman 相关 $\geqslant 0.85$。**三项任一不达标，管线不可信，后续版本对比全部无效**。

2. **统计口径预先写死**：(a) 扫描以 100 kb 非重叠窗口汇总（同时报 50 kb 与 200 kb 作敏感性分析）；(b) 全基因组多重检验一律报经验分位数与 BH-FDR，**不报裸 p 值**；(c) "重现率"这类不平衡二分类判断（真正的选择窗口占全基因组比例 < 1%）报 **PR-AUC 与固定召回处的精确率，不报 accuracy**，原因是负类占 99% 以上，全判"非候选"即得 99% accuracy 而毫无信息；(d) 所有重现率给 1,000 次自助法（按染色体块重抽样，保留连锁不平衡结构）95% 置信区间；(e) 两版比较必须限制在**两版共有的双等位 SNP 且 MAF $\geqslant 0.05$** 的子集上，共有位点数与各版特有位点数逐染色体报出。
3. **混杂因素必须显式处理并写明**：(a) 坐标系统——两版必须在 GRCh38 上比较（使用 1000G phase 3 的 GRCh38 重调用版本，而非自行 liftOver；若必须 liftOver，须报告映射失败率）；(b) 遗传图谱——两版使用**同一份** GRCh38 重组图谱，图谱来源与版本写入方法；(c) 样本集——30× 版含 698 个亲属样本，主分析必须限制到与 phase 3 完全相同的 2,504 例无关个体，亲属样本只用于 H2 的定相质量评估。
4. **规模下限**：$\geqslant 5$ 条染色体（须含 chr2、chr4、chr11、chr15 这些携带经典靶点的染色体，外加 chr22 或 chr20 作为"无已知强信号"的阴性对照染色体）；$\geqslant 4$ 个群体（至少覆盖 EUR、EAS、AFR、SAS 各一）；每染色体每群体的分析 SNP 数 $\geqslant 200{,}000$。
5. **可复现性**：bcftools 过滤命令、selscan 参数、图谱文件校验和、全部脚本与结果表开源；提供 chr22 单染色体的降规模复现包（$\leqslant$ 2 小时可跑完）。

5 · 数据与工具

用途	来源 / 工具
30× 高覆盖定相 call set	1000 Genomes FTP：`http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/`（2,504 无关个体的定相 VCF，GRCh38）与 `.../working/20220422_3202_phased_SNV_INDEL_SV/`（3,202 例含亲属）。**完全开放访问，无需任何 DAC 审批**——1000 Genomes 的知情同意即为全球公开发布。单染色体压缩 VCF 约 1–6 GB，5 条染色体约 15–35 GB 磁盘
phase 3 低覆盖定相 call set（GRCh38 重调用版）	1000 Genomes FTP 的 GRCh38 phase 3 目录（2019 年 remapping 发布）。**确切路径与文件命名需核实**；若 GRCh38 版不可用，退回 GRCh37 版 + 官方 chain 文件 liftOver，并报告映射失败率
样本–群体对应表	1000 Genomes `integrated_call_samples_v3.20130502.ALL.panel` 及 30× 版对应的 pedigree 文件
GRCh38 重组图谱	Beagle 发布的 GRCh38 genetic maps（公开下载）或 HapMap GRCh37 图谱经官方 liftOver。**两版分析必须用同一份图谱**，来源写入方法
VCF 处理	`bcftools`（view/filter/annotate/isec），流式处理，内存占用与文件大小无关，MB 级
选择扫描	`selscan` v2（C++，多线程；iHS/nSL/XP-EHH 均支持）；交叉校验用 Python `scikit-allel`（`allel.ihs`）。单群体单体型矩阵：CEU n=99 → 198 单体型 × 1.5M SNP 的 int8 数组约 0.3 GB，16 GB 内存充裕；**若一次性载入全部 2,504 例（5,008 单体型 × 1.5M）则约 7.5 GB，接近上限，因此管线一律按群体切片**
算力口径	纯 CPU。selscan iHS：200 单体型 × 1M SNP、4 线程约 15–60 分钟/染色体/群体。5 染色体 × 4 群体 × 2 版本 = 40 次运行 ≈ 15–40 小时，可后台分夜跑。**全基因组 26 群体两版全跑属超出范围**（约 25 倍工作量 + 200 GB 以上磁盘），方案中不设此路径
已发表选择信号对照	Voight 等 2006 与 1000 Genomes Selection Browser 公布的 iHS 结果表；Byrska-Bishop 等 2022 的变异调用评估表。**仅用于校验与对比，不计入本项目的数据贡献**
伦理	1000 Genomes 为完全去标识化、以群体为单位公开发布的开放数据，不涉及可识别个人、不涉及人体实验，无需伦理审批。方案中**不得**做任何个体层面的表型推断

6 · 方法路径

1. 装环境（bcftools + selscan + scikit-allel），用 chr22 小片段跑通两条管线，**先完成验收标准第 1 条的三项经典信号复现**，把复现结果与文献值并列写入 `validation/`。
2. **核实数据版本与坐标**：确认 phase 3 GRCh38 重调用版的可用性与路径；用 `bcftools isec` 求两版共有 SNP 集合，逐染色体报共有/特有位点数与 MAF 分布。此步是全课题最关键的前置核实，**不得边做边发现**。
3. 按预先写死的规则做过滤（双等位 SNP、MAF $\geqslant 0.05$、非缺失、共有位点集），按群体切片生成单体型输入；过滤前后位点数逐步记录。
4. 在两版数据上分别跑 iHS 与 nSL（4 群体 × 5 染色体），汇总到 100 kb 窗口，生成排名表。
5. 计算重现率：按 phase 3 排名分档（1 – 100、101 – 500、501 – 1000）统计落入 30× 前 1% 的比例，自助法给区间；同时报 PR-AUC 与固定召回处精确率。检验 H1。
6. 检验 H2：用 602 个三口之家计算区域级定相一致性代理指标，把不重现窗口与重现窗口在重组率、MAF 谱、定相一致性上做对比（BH 校正）。
7. 独立交叉校验：换一类**不依赖单体型定相**的统计量（群体分化 F_ST 与 PBS）在同两版数据上重算同一批窗口。若 F_ST/PBS 的两版一致性远高于 iHS，则差异确由定相/填补引起而非位点集差异——这是本课题最关键的一

次对照。

7 · 新颖性边界

> 本课题**不**提出新的选择检测统计量，**不**声称发现任何新的选择位点，**不**声称 LCT/SLC24A5/EDAR 处于正选择为本项目发现（这些是二十年来的经典结论），**不**声称 30× 数据优于 phase 3（Byrska-Bishop 等已在变异调用层面报告）。丘奖 2025 年铜奖论文《Association Study of SNPs in X-Chromosome Inactivation Escape Regions…》做的是关联分析，与本课题的选择扫描/数据版本对比路径不同。
>
> 已有工作具体完成了什么：Voight 等与后续大量论文在 phase 3（及更早的 HapMap）上标定了 iHS/XP-EHH 的候选位点清单；Byrska-Bishop 等（Cell, 2022）报告了 30× 相对 phase 3 在 SNV/INDEL/SV 调用灵敏度与精确度上的提升，并发布了定相 panel。**没有人把这两件事接起来，检验下游选择推断是否随数据版本改变**。
>
> 本项目的贡献（且是主结论）：把评价维度从"变异调用质量"换成"**下游群体遗传推断的版本稳健性**"，给出分信号强度档的重现率曲线，以及"哪一类窗口最不稳健"的可操作特征刻画（H2）。这是研究手册所列最稳的差异化轴的一个变体：换的不是新群体，而是同一群体的新数据生成方式。
>
> 为什么有价值：过去十年数以百计的人类适应性进化结论建立在 phase 3 之上。若高排名信号重现率高，这些结论获得一次真正独立的技术支持；若中等排名信号大面积不重现，则文献中"新发现的选择位点"这一类主张需要重新评估。**两种结果都构成有效结论**。
>
> 风险声明：若重现率接近 100%（差异不显著），必须给出自助法误差棒证明本设计有能力分辨假设中的分档差异，否则"高度稳健"的结论无效。

8 · 决策门槛（go / no-go）

- **第 4 周末**：确认 phase 3 GRCh38 重调用版可下载且与 30× 版坐标一致。若不可得，**立即降级**：其一，改用官方 chain 文件把 30× 版 liftOver 到 GRCh37，与原始 phase 3 比较，并把 liftOver 失败率作为一条限制条款报出；其二，把比较对象从"两个数据版本"改为"**同一 30× 数据的两种定相方式**"（如 SHAPEIT/Beagle 统计定相 vs 三口之家 pedigree 定相），此时研究问题变为"定相方法对选择扫描的影响"，主结论框架（分档重现率 + 不重现窗口特征刻画 + F_ST 对照）完全保留。
- **第 8 周末（硬门槛）**：验收标准第 1 条必须通过。若与已发表 iHS 的相关 < 0.85，先排查图谱版本与祖先等位（ancestral allele）标注（这两项是 iHS 最常见的错误源）；四周内仍不通过则**降级**：把校验基准从"数值相关"改为"经典位点的排名分位数复现"（更稳健，阈值定为三个经典位点均入前 1%），并在论文中说明数值不可比的原因。
- **第 14 周末**：确认磁盘与运行时间。若 5 条染色体的下载+运行超出可承受范围，**立即降级**：染色体数降到 3（chr2 + chr15 + chr22 对照），群体数降到 2（CEU + CHB），并在结论中相应收窄推广范围。主结论框架不变。
- **第 28 周末**：若 H1 与 H2 均未达阈值，按风险声明写"版本稳健"结论，并把 F_ST/PBS 对照与误差棒作为核心证据。
- **第 36 周**结果冻结，第 44 周英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) 30× 定相 VCF 的确切 FTP 路径与是否已按无关个体子集发布；(b) GRCh38 重组图谱的来源与是否覆盖全部目标染色体；(c) 祖先等位标注文件（Ensembl EPO 比对导出）在 GRCh38 上的可得性——iHS 的极性依赖此文件，缺失会使结果无意义。三项在第 6 周前查清。
- **选择前提**：适合已熟悉命令行、有耐心处理大文件与坐标系统的学生。本路线的最大困难是"两版数据的位点集不同"这一根本性不可比问题——**这个困难本身就是研究内容**，处理它（共有位点子集 + F_ST 对照）就是本课题方法学的核心。

- **预算裁剪顺序**：先砍 nSL（只留 iHS），再砍 H2 的定相一致性分析（保留重组率与 MAF 谱特征刻画），再砍群体数，最后砍染色体数。砍到只剩"chr2 + chr15、CEU + CHB、iHS 两版分档重现率 + F_ST 对照"时，主结论仍成立。

B06 · 公民科学物候数据中估计量选择造成的趋势偏倚：以 2015–2025 记录量爆发期为自然实验，比较首现日、分位数与 Weibull 校正估计量给出的昆虫出现期位移

`优先级：中高` · `分族：生态与物候公开数据` · `参赛子类：生态信息学 / 生物统计` · `资源需求：纯 CPU 笔记本，GBIF 下载 ≤ 3 GB，自助法为主要算力开销` · `技能取向：R（rgbif + phenesse）+ 统计推断`

1 · 研究问题

在 GBIF 上 2015 – 2025 年公民科学记录量增长约一个数量级的自然实验条件下，用"首次观测日（first observation date）""第 10/50 百分位日""Weibull 分布校正的起始日"三类物候估计量对同一批昆虫物种计算年际趋势，所得的趋势方向与幅度差异有多大？这些差异能否被各物种各年的记录数增长速率定量解释？

2 · 研究背景与空白

背景。 物候（phenology）位移是气候变化最直接的生物学指标之一。公民科学平台（iNaturalist、eBird 等）经 GBIF 汇聚了海量带日期与坐标的观测记录，成为物候研究的主力数据源。但这类数据没有标准化的采样设计：某物种某年的"首次被观测到的日期"同时取决于该物种何时真正出现，和当年有多少人在看——观测努力（observer effort）越大，首现日越早。这是一个经典的极值统计量偏倚问题。

已有工作到哪一步。 该偏倚在文献中被反复指出：加拿大 PlantWatch 研究（Scientific Reports, 2013）为控制观测者差异，只保留"至少 5 名观测者、至少 5 个地点"的首花日记录，报告 19 种植物 2001 – 2012 年首花日平均提前约 9 天；丹麦的对比研究显示定向公民科学项目（目标就是报告"首花"）记录到的花期显著早于标本馆与随机记录；Belitz 等（2020）提出并实现了基于 Weibull 分布的物候分位数估计量（R 包 `phenesse`），可给出对采样强度不敏感的起始日估计与置信区间。

空白在于：这些工作要么只做努力筛选、要么只提出更好的估计量，没有人把"估计量选择"当成自变量，在一个记录量剧烈变化的时间窗上量化它对趋势结论本身的影响。 换句话说，文献知道首现日有偏，但没有回答"这个偏差有多大、会不会把一个真实的物候提前趋势放大成两倍，或者把没有趋势的物种伪造出趋势"。2015 – 2025 恰好是 iNaturalist 记录量爆发的十年，构成一个天然放大器。这个空白适合学生课题：数据完全开放、算力可控、**结论正负都成立**。

3 · 可检验假设

- **H1**：以"首现日"估计的年际趋势斜率（天/年）系统性地比 Weibull 校正起始日估计的斜率更负（更"提前"），跨物种的中位差值 ⩾ 1.0 天/年，且配对自助法 95% 置信区间不跨 0。
- **H2**（机制假设）：物种层面的斜率差值与该物种 2015 – 2025 记录数的年均增长率呈正相关（Spearman ρ ⩾ 0.5，BH-FDR < 0.05）；若把每物种每年的记录数随机稀释到该物种十年内的最小年记录数，斜率差值应缩小 ⩾ 60%。若稀释后差值不缩小，说明偏倚来自记录的空间/物种构成变化而非单纯数量，需另作解释。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建管线复现两组已发表结果：(a) 在 Belitz 等（2020）`phenesse` 论文所用的模拟设定下重跑其 Weibull 估计量，要求覆盖率（nominal 95% CI 的实际覆盖率）与原文报告值偏差 ⩽ 3 个百分点；(b) 选取 1 篇已发表的公民科学物候趋势论文（须公开物种清单、区域与年份范围），在同源数据上重算其报告的趋势斜率，要求 ⩾ 70% 的物种斜率符号一致且中位斜率的绝对偏差 ⩽ 0.5 天/年。**任一组不达标，管线不可信，后续全部结论无效。**
2. **统计口径预先写死**：(a) 趋势用物种级线性回归（估计日 ~ 年）与跨物种混合效应模型（`lme4`，物种为随机截距与随机斜率）双口径都报；(b) 跨物种的多重检验一律报 BH-FDR，不报裸 p 值；(c) 所有估计量的不确定性用

≥ 500 次自助法给 95% 置信区间，`phenesse` 的内置区间与自助区间都报；(d) 时间序列性质必须尊重——**趋势检验不得做随机划分交叉验证**，若做预测评估须按年份前后划分并额外报"随机划分 vs 时间划分"的差值以量化高估幅度；(e) 空间混杂必须控制——每物种限定在一个固定的地理网格（如 100 km × 100 km 或固定行政区）内，并报告各年记录的空间质心漂移，漂移超过 50 km 的物种-年标记并做敏感性分析。

3. **样本量与检验力预先估计**：≥ 40 个物种，每物种每年 ≥ 30 条记录且 ≥ 8 个有效年份。开工前用模拟估计：在 n=40 物种、10 年序列下，能以 80% 检验力检出的最小斜率差值是多少；若大于 H1 的 1.0 天/年阈值，必须先扩大物种数或放宽年份要求。
4. **数据来源地位**：GBIF 记录中的物种鉴定（含 iNaturalist 的 research-grade 标记）**仅作为标签来源使用，不计入本项目的数据贡献**；必须报告各物种的鉴定可靠性代理指标（research-grade 比例）并做敏感性分析。
5. **可复现性**：GBIF 下载必须通过 `rgbif::occ_download` 生成**带 DOI 的可引用下载**，DOI 写入论文；全部脚本、物种清单、随机种子、区域定义开源；提供 10 物种的降规模复现包（≤ 1 小时可跑完）。

5 · 数据与工具

用途	来源 / 工具
观测记录	GBIF（https://www.gbif.org/ ），经 R 包 `rgbif` 的 `occ_download` 提交筛选式：`basisOfRecord = HUMAN_OBSERVATION`、`hasCoordinate = TRUE`、`hasGeospatialIssue = FALSE`、`year = 2015,2025`、限定目标分类群与地理范围。**限定单一区域 + 单一目（如鳞翅目或蜻蜓目）后，下载体积通常在数百 MB 至 3 GB（DwC-A 压缩包）**，`data.table::fread` 分块读取即可，16 GB 内存充裕。全球全类群下载（数十 GB 至 TB 级）属超出范围
为什么选昆虫成虫出现期而非植物花期	昆虫成虫的观测日期本身即物候信号，**无需依赖注释字段**；而植物花期需要 iNaturalist 的 "Plant Phenology" 注释，**该注释是否完整传递到 GBIF 的 DwC-A（`reproductiveCondition` / `dynamicProperties` 字段）需核实**。若核实为可用，则可把植物花期作为第二个类群加入（见 go/no-go）
物候估计量	R 包 `phenesse`（Belitz et al. 2020）：`weib_percentile()` 与 `weib_percentile_ci()`。单次估计（n≈100 观测、iterations=500）约数秒；40 物种 × 11 年 × 3 分位点 × 500 次自助 ≈ 数小时至一夜，**纯 CPU 可行**。另实现朴素首现日与经验分位数作为对照
趋势建模	R `lme4`（混合效应）、`boot`、`p.adjust(method="BH")`
气候协变量（可选）	ERA5-Land 或 WorldClim 月均温（公开下载），仅用于说明性分析，不作为主结论
观测努力代理	同区域同期**全部**该分类群记录数（不限物种）作为努力代理；亦可用 eBird/iNaturalist 的观测者数。**eBird Basic Dataset 需提交免费申请并等待审批**，虽通常获批但存在时间不确定性，因此**不设为主线依赖**，仅列为可选补充
对照基准	Belitz 等 2020 的模拟覆盖率结果、所选复现论文的趋势斜率。**仅用于校验与对比，不计入本项目的数据贡献**
伦理	全部为已公开的观测记录，不涉及人体数据、不涉及活体动物操作、不涉及可识别个人信息（GBIF 记录的观测者字段在分析中一律不使用），无需伦理审批

6 · 方法路径

1. 装环境（R + rgbif + phenesse + lme4），跑通 `phenesse` 内置示例，**先完成验收标准第 1 条的两组复现**，把覆盖率表与斜率对比表写入 `validation/`。
2. 按预先写死的规则确定研究区域、分类群与物种清单（规则：每年 ≥ 30 条记录、≥ 8 个有效年份、research-grade 比例 ≥ 0.8、空间质心年漂移 < 50 km），提交带 DOI 的 GBIF 下载。**规则先于数据探索确定，不得看到结果后再改**。

3. 核实注释字段可用性（决定是否加入植物花期第二类群）与各物种的记录量增长曲线，出具"物种 × 年 × 记录数"矩阵与空间质心漂移表。
4. 对每个物种每年计算三类估计量（首现日、第 10/50 经验百分位、Weibull 校正第 10/50 百分位）及其自助区间。
5. 分别用三类估计量拟合年际趋势（物种级 OLS + 跨物种混合效应），得到三套斜率，做配对比较并给自助区间，检验 H1。
6. 检验 H2：把斜率差值对记录数年均增长率做回归；再做稀释实验（每物种每年下采样到最小年记录数），重算三套斜率与差值。
7. 独立交叉校验：换一套**采样设计完全不同**的数据源复算至少 10 个物种——首选具备标准化协议的公开数据（如英国 UKBMS 蝴蝶监测的公开年度指数、NEON 的地面节肢动物取样数据），检验"标准化数据上三类估计量差异是否消失"。若消失，则确证差异源于公民科学的采样不均而非估计量本身的数学性质。

7 · 新颖性边界

> 本课题**不**提出新的物候估计量（Weibull 校正估计量由 Belitz 等 2020 提出并实现），**不**声称发现新的物候位移现象，**不**声称"首现日对采样强度敏感"为本项目发现——这一点自 Miller-Rushing 等以来是物候统计学的常识，PlantWatch 与丹麦对比研究都已明确报告。
>
> 已有工作具体完成了什么：一类工作用观测者/地点数下限**筛掉**有偏记录（PlantWatch，2001 – 2012，19 种植物，报出约 9 天提前）；另一类工作**提出更好的估计量**并用模拟验证其覆盖率（`phenesse`，2020）；还有工作对比了不同数据源（标本馆 vs 定向公民科学 vs 随机记录）的花期分布差异。**没有一项工作把"估计量选择"当作自变量，在记录量剧烈增长的真实时间窗上量化它对最终趋势结论的偏倚幅度**。
>
> 本项目的贡献（且是主结论）：利用 2015 – 2025 公民科学记录量增长约一个数量级这一**自然实验**，给出"估计量选择造成的趋势偏倚"的定量刻度（天/年），并检验该偏倚是否可由记录量增长率解释（H2）与是否在标准化数据上消失（第 7 步对照）。这是评价维度的转换——已有工作评估的是估计量的统计性质（覆盖率、偏差），本项目评估的是该选择对**最终气候生物学结论**的偏置。
>
> 为什么有价值：大量近年物候论文直接使用 GBIF/iNaturalist 的首现日。若偏倚达到 1 天/年量级，在十年窗口上就是 10 天——与文献报告的真实位移同量级，足以改变结论的解释。
>
> 风险声明：**"三类估计量给出的趋势无显著差异"同样是有效结论**，它会为大量已有物候文献提供一次重要的稳健性背书。但必须由第 3 条的检验力估计证明本设计有能力检出 1.0 天/年的差值，否则该结论无效。

8 · 决策门槛（go / no-go）

- **第 5 周末**：确认满足筛选规则的物种数。若合格物种 < 25，**立即降级**：其一，把区域扩大（从单省/单州扩到国家或大区）并把年记录数下限降到 20；其二，把时间窗前移到 2010 – 2025 以纳入更多年份。两条降级都保留"三类估计量并列 + 偏倚量化 + 增长率解释"的主结论框架。
- **第 8 周末（硬门槛）**：验收标准第 1 条必须通过。若 `phenesse` 覆盖率复现偏差 > 3 个百分点，先核对自助迭代次数与随机种子；若已发表趋势论文的斜率复现符号一致率 < 70%，须逐项排查区域定义与物种同义名（GBIF 的分类骨架会合并同义名，这是最常见的不一致来源）。六周内仍不通过则**降级**：把校验目标改为"在**模拟数据**上验证三类估计量的已知理论性质"（首现日随 n 增大而单调提前、Weibull 估计量渐近无偏），阈值改为模拟结果与理论预期一致。主结论框架不变。
- **第 12 周末**：核实植物花期注释字段是否可用。若可用则加入第二类群（增强结论普适性）；若不可用则**明确放弃植物分支**，只做昆虫，并在论文中写明原因——不得含糊带过。

- **第 24 周末**：完成第 7 步的标准化数据对照。若找不到与目标物种重叠的标准化监测数据，**立即降级**：改用"高观测努力年份的子集"作为伪标准化对照（在记录最密集的 3 年内做稀释，模拟低努力情形），并标注这是内部对照而非独立数据源。
- **第 36 周**结果冻结，第 44 周英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) GBIF 下载配额与 DwC-A 体积（先用 `occ_count` 预估）；(b) iNaturalist 物候注释在 GBIF 中的传递情况；(c) `phenesse` 在 $n < 20$ 时是否会失败或给出发散区间（须实测并写入物种筛选规则）。三项在第 8 周前查清。
- **选择前提**：适合对统计推断有兴趣、能接受"结论是一把尺子而不是一个发现"的学生。答辩时须能用英文清楚说明"为什么估计量的选择会改变生态学结论"，这是本课题的说服力核心。
- **预算裁剪顺序**：先砍气候协变量分析，再砍稀释实验（H2 后半），再砍第 50 百分位（只保留第 10 百分位与首现日的对比），最后把物种数从 40 降到 25。砍到只剩"25 物种 × 首现日 vs Weibull 第 10 百分位 × 混合效应趋势 + 增长率回归"时，主结论仍成立。

B07 · 肠道微生物组疾病分类器的临床有用性再评估：把跨队列评价维度从 AUROC 换成真实患病率下的 PR–AUC 与决策曲线净获益

优先级：中 · 分族：微生物组公开数据 · 参赛子类：微生物组信息学 / 生物统计 · 资源需求：纯 CPU 笔记本，物种丰度表 ≤ 2 GB，无需处理原始测序数据 · 技能取向：R/Python 机器学习 + 临床预测模型评价

1 · 研究问题

已发表的肠道微生物组疾病分类器在留一队列（leave-one-dataset-out, LODO）验证下报出的 AUROC 约 0.7 – 0.8，若把评价维度换成**真实人群患病率下的 PR-AUC 与决策曲线净获益（net benefit, decision curve analysis）**，这些分类器在筛查场景中还剩多少实际价值？在哪些疾病、哪个阈值概率区间内，模型的净获益仍高于"全部筛查"与"全部不筛查"两条参照线？

2 · 研究背景与空白

背景。 用肠道微生物组做疾病无创诊断是过去十年的热门方向：从粪便宏基因组或 16S 数据提取物种/功能丰度，训练随机森林或梯度提升机，报告 AUROC。但 AUROC 是一个与患病率无关的判别指标，而临床筛查的价值取决于在**真实患病率**下阳性预测值有多高、以及在医生实际使用的阈值处相对"不筛查"能净挽回多少真阳性。结直肠癌人群患病率约 0.05% – 0.5%，即便 AUROC 0.80 的模型在这种患病率下的阳性预测值也可能低于 5%。

已有工作到哪一步。 Duvallet 等（2017）建立了 MicrobiomeHD 跨疾病 16S 元分析；Wirbel 等（2019）做了结直肠癌宏基因组的多队列元分析并报 LODO AUROC；一项 2023 年的系统评估（Gut Microbes）用 curatedMetagenomicData 对 **20 种疾病**做跨队列验证，报告队列内 AUROC 约 0.77 而跨队列仅在肠道疾病上保持约 0.73，非肠道疾病显著更低，并报告了标志物一致性；另有 2022 年的 IBD 基准工作（15 个 16S 数据集、7,707 样本）系统比较了归一化、批次校正与模型的组合，并用 LODO 评估泛化。

空白在于：**这一整条文献线全部以 AUROC（及少量 AUPRC）为终点，没有任何一项做过真实患病率下的净获益评估。** 而这恰恰是"微生物组能否用于筛查"这一实际问题的唯一相关指标。决策曲线分析在临床预测模型领域早已是标准做法（TRIPOD 指南推荐），却几乎没有进入微生物组领域。这个空白适合学生课题：所有数据已统一处理并可通过 R 包一行代码获取、算力在小时级、方法学成熟且公开、**结论正负都成立**。

3 · 可检验假设

- H1：在文献报告的真实人群患病率下，跨队列（LODO）训练的分类器对**非肠道疾病**在全部临床相关阈值概率区间（0.5% – 20%）内的净获益均不高于"全部筛查"策略（差值 95% 自助法置信区间上界 < 0）；而对**肠道疾病**（结直肠癌、IBD）存在一个宽度 ⩾ 5 个百分点的阈值区间使净获益显著为正。
- H2（机制/量化假设）：AUROC 与净获益之间不存在单调对应——存在至少一对 (疾病 A, 疾病 B)，使 AUROC(A) > AUROC(B) 但在各自真实患病率下净获益(A) < 净获益(B)。若这样的反转对存在，即证明现有以 AUROC 排序疾病的做法会误导优先级判断。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建管线复现两项已发表结果：(a) 复现 Wirbel 等（2019）或上述 2023 年 Gut Microbes 工作中结直肠癌的 LODO AUROC，要求各队列的 AUROC 与原文报告值偏差 ⩽ 0.05（绝对值），且队列间排序的 Spearman 相关 ⩾ 0.7；(b) 复现该工作中"跨队列 AUROC 显著低于队列内 AUROC"的定性反差，方向与量级（差值 ⩾ 0.03）一致。**任一项不达标，管线不可信，后续净获益结论全部无效。**
2. **统计口径预先写死**：(a) 主指标为**在真实患病率下重标定后的 PR-AUC 与决策曲线净获益，明确不报 accuracy**——原因是在患病率 0.5% 的筛查场景中，全判阴性即得 99.5% accuracy，该指标完全无信息；(b) AUROC 仅作

为与文献对齐的辅助指标；(c) 由于研究队列中病例/对照比例被人为设为接近 1:1，必须做**患病率重标定**：用 Bayes 先验修正把模型输出概率映射到目标患病率，重标定公式与来源写入方法；(d) 校准必须评估——报 calibration slope/intercept 与 Brier 分数，并做 Platt scaling 或等距回归重校准后的对照；(e) 所有指标给 ⩾ 1,000 次自助法（**按队列整体重抽样，不按样本**，以尊重队列层级结构）95% 置信区间；(f) 多疾病比较报 BH-FDR。

3. **批次效应必须检查并写明校正方法**：全部使用 curatedMetagenomicData 的统一 MetaPhlAn 管线输出以消除生信管线差异；对剩余的测序平台/国家差异，用 PERMANOVA（`vegan::adonis2`）量化其解释的方差比例，并做一次 ConQuR 或 MMUPHin 批次校正的敏感性分析，报告校正前后 AUROC 与净获益的差值。
4. **规模下限**：⩾ 6 种疾病，每种疾病 ⩾ 3 个独立队列，每队列 ⩾ 50 例病例与 50 例对照；每种疾病必须能从一级文献（WHO/国家疾控/系统综述）查到可引用的人群患病率，患病率来源与年份逐条写入表格，并对患病率做 ±50% 的敏感性分析。
5. **公开注释与已发表模型的地位**：curatedMetagenomicData 的物种注释、已发表模型的超参数与特征列表**仅作为对照与输入使用，不计入本项目的数据贡献**。
6. **可复现性**：全部脚本、包版本、随机种子、每疾病的患病率来源表开源；提供 2 疾病的降规模复现包（⩽ 1 小时可跑完）。

5 · 数据与工具

用途	来源 / 工具
统一处理的宏基因组物种丰度	`curatedMetagenomicData`（Bioconductor R 包，https://waldronlab.io/curatedMetagenomicData/ ）。约 2 万份公开人体宏基因组样本，全部用同一 MetaPhlAn/HUMAnN 管线处理，附人工策展的元数据（年龄、BMI、国家、疾病状态、用药）。**按疾病切片下载，单疾病数据 < 200 MB**；全库物种丰度表约 1–2 GB，16 GB 内存可处理
16S 备用数据	MicrobiomeHD（Duvallet 等 2017，公开于 Zenodo）与 Qiita 公开研究。**仅在宏基因组队列不足时启用**（见 go/no-go）
受控访问核查	curatedMetagenomicData 与 MicrobiomeHD 收录的均为**已公开发表并开放下载**的数据，元数据已去标识化，**无需 DAC 审批**。HMP2/iHMP 的部分详细临床字段属受控访问，本项目**一律不使用**。不涉及人体实验与可识别个人信息，无需伦理审批
人群患病率	WHO、GBD（Global Burden of Disease，http://ghdx.healthdata.org/ ）、各国疾控年报、系统综述。每种疾病取一个主值 + 一个区间
机器学习	R `mlr3` / `caret` 或 Python `scikit-learn`（随机森林、弹性网 logistic、LightGBM）。样本量千级、特征数千级，单次 5 折交叉验证在 CPU 上 < 2 min；LODO 全流程（6 疾病 × 约 4 队列 × 3 模型）约 2–6 小时
决策曲线分析	R 包 `dcurves` 或 `rmda`（标准实现，秒级）；净获益公式须在论文中显式写出并自行实现一遍做交叉校验
校准与重标定	R `CalibrationCurves`、Platt scaling / 等距回归（`scikit-learn` 的 `CalibratedClassifierCV`）
批次校正	`MMUPHin`（Bioconductor）或 ConQuR
对照基准	Wirbel 等 2019、2023 Gut Microbes 20 疾病评估、2022 IBD 基准的已发表 AUROC 表。**仅用于校验与对比，不计入本项目的数据贡献**
明确排除的路径	**从原始 fastq 重跑 MetaPhlAn/HUMAnN 属超出范围**：单份粪便宏基因组原始数据 3–10 GB，2,000 份即 6–20 TB 下载；MetaPhlAn4 数据库索引约 20 GB 且比对阶段建议 ⩾ 32 GB 内存，单样本 4 核约 20–60 分钟。本项目**只消费已统一处理的丰度表**，此点须在论文方法部分明写

6 · 方法路径

1. 装环境（R + curatedMetagenomicData + mlr3 + dcurves），跑通包内示例，**先完成验收标准第 1 条的结直肠癌 LODO 复现**，把 AUROC 对比表写入 `validation/`。
2. 按预先写死的规则锁定 ⩾ 6 种疾病与其队列集合（规则：⩾ 3 队列、每队列 ⩾ 50/50、元数据完整度 ⩾ 80%），生成"疾病 – 队列 – 样本数 – 国家 – 平台"总表并冻结。
3. 从一级来源查取每种疾病的人群患病率，形成带出处与年份的患病率表；**此表必须先于任何净获益计算冻结**。
4. 跑 LODO 训练与预测，得到每个样本的预测概率；报 AUROC（对齐文献）、AUPRC、校准曲线与 Brier 分数。
5. 做患病率重标定与概率重校准，然后计算每种疾病在阈值概率 0.5% – 20% 网格上的净获益曲线，与"全筛"、"全不筛"两条参照线并列作图；自助法给区间。检验 H1。
6. 检验 H2：在疾病之间比较 AUROC 排序与净获益排序，用 Kendall τ 量化两种排序的一致性，并点名所有反转对。
7. 独立交叉校验：其一，用 MMUPHin 批次校正后的特征重跑全流程，验证结论不由批次伪迹驱动；其二，用一个**完全不同的模型族**（弹性网 logistic 而非树模型）重算主曲线；其三，对患病率做 ±50% 敏感性分析，检验结论是否稳健。

7 · 新颖性边界

本课题**不**提出新的分类模型、**不**提出新的批次校正方法、**不**声称发现新的微生物标志物，**不**声称"跨队列泛化性能下降"为本项目发现——这一点由 Duvallet 等（2017）、Wirbel 等（2019）与 2023 年 Gut Microbes 的 20 疾病评估明确报告。

已有工作具体完成了什么：Duvallet 等建立跨疾病 16S 元分析框架；Wirbel 等在结直肠癌上做多队列宏基因组元分析与 LODO；2023 年 Gut Microbes 工作在 curatedMetagenomicData 上系统评估 20 种疾病的队列内（约 0.77）与跨队列（肠道疾病约 0.73）AUROC 并比较 16S 与宏基因组、报告标志物一致性；2022 年 IBD 基准比较了处理流程与模型组合。**这四项工作的终点全部是判别力指标，均未做患病率重标定、未做校准评估、未做决策曲线净获益。**

本项目的贡献（且是主结论）：把评价维度从"判别力（AUROC）"换成"**真实患病率下的临床有用性（PR-AUC + 校准 + 净获益）**"，并检验一个可否证的命题——AUROC 排序与净获益排序会发生反转（H2）。交付物是一张"疾病 × 阈值概率 → 是否有净获益"的可查表。

为什么有价值：微生物组诊断的产业化宣传普遍引用 AUROC。给出同一批数据在真实患病率下的净获益，把"这个模型能不能用于筛查"从修辞变成可读数的判断。

风险声明：**"净获益普遍为正、AUROC 排序与净获益排序高度一致"同样是有效结论**，它会为该领域提供一次重要的正面背书。但必须给出按队列重抽样的自助法误差棒，证明本设计有能力分辨净获益曲线与参照线之间的差异。

8 · 决策门槛（go / no-go）

- **第 5 周末**：确认 curatedMetagenomicData 中满足"⩾ 3 队列、每队列 ⩾ 50/50"的疾病数量。若 < 4 种，**立即降级**：其一，把队列下限降到 30/30 并把疾病数补足；其二，启用 MicrobiomeHD 的 16S 队列作为补充，代价是需分数据类型报结果（16S 与宏基因组不合并）。两条降级都保留"重标定 + 校准 + 净获益 + 排序反转"的主结论框架。
- **第 9 周末（硬门槛）**：验收标准第 1 条必须通过。若结直肠癌 LODO AUROC 与文献偏差 > 0.05，先排查特征预处理（相对丰度 vs CLR 变换）与队列纳入范围；四周内仍不通过则**降级**：把校验基准改为"复现原文公开代码仓库的输出"（若有），阈值提到偏差 ⩽ 0.02；若原文无代码，则把校验改为"在同一队列上队列内 5 折交叉验证的 AUROC 落入原文报告区间"，并在论文中明写复现受限的原因。

- **第 14 周末**：确认能否为每种疾病从一级来源取到可引用的人群患病率。若某疾病取不到，**立即降级**：该疾病改用"患病率区间扫描"呈现（在 0.1% – 10% 全区间上画净获益热图），而非单点患病率。主结论框架不变，且这种呈现方式信息量更大。
- **第 26 周末**：若 H2 的反转对不存在（AUROC 与净获益排序 Kendall $\tau > 0.9$），按风险声明写"两种评价一致"结论，并把净获益可查表作为核心交付物。
- **第 36 周**结果冻结，第 44 周英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) `curatedMetagenomicData` 当前版本收录的疾病与队列清单（版本间会变动，须固定版本号）；(b) 目标复现论文是否公开了逐队列 AUROC 数值表；(c) 决策曲线在类别不平衡且经重标定后的正确实现方式（须手工推一遍公式，不能只调包）。三项在第 8 周前查清。
- **选择前提**：适合对临床统计有兴趣、能理解"判别力 ≠ 有用性"这一区分的学生。本课题的技术门槛不高，说服力全部来自评价框架的正确性，因此对概念清晰度要求高。
- **预算裁剪顺序**：先砍 MMUPHin 批次校正对照，再砍第三个模型族，再砍疾病数（从 6 降到 4）。砍到只剩"4 疾病 × LODO × 随机森林 × 重标定净获益曲线"时，主结论仍成立。

B08 · 已发表两样本孟德尔随机化因果宣称的系统性复核：用更新后的 GWAS 摘要统计重跑，检验重现率与工具变量强度、多效性检验完备性的关系

优先级：中 · 分族：群体遗传与流行病学 · 参赛子类：遗传流行病学 / 因果推断 · 资源需求：纯 CPU 笔记本，摘要统计按需下载 ≤ 20 GB，全部为聚合级公开数据 · 技能取向：R（TwoSampleMR）+ 因果推断方法论

1 · 研究问题

从近五年文献中系统抽取 ⩾ 60 条两样本孟德尔随机化（two-sample Mendelian randomization, MR）发表的"显著因果效应"宣称，用**发表之后才出现的更大样本 GWAS 摘要统计**在统一管线下重跑，有多大比例的效应方向一致且仍达显著？重现率能否由原研究的工具变量强度（F 统计量）、工具 SNP 数、以及原文是否完整报告了多效性（pleiotropy）敏感性分析来预测？

2 · 研究背景与空白

背景。 MR 用与暴露相关的遗传变异作为工具变量（instrumental variable）估计暴露对结局的因果效应，理论上可规避混杂与反向因果。因为全基因组关联研究（GWAS）摘要统计大量公开，两样本 MR 的执行成本极低——一台笔记本、一个 R 包、几分钟即可产出一篇论文。结果是过去五年 MR 论文数量爆炸式增长，其中大量是"暴露 X 对疾病 Y 有因果效应"的单一宣称，方法学质量参差。

已有工作到哪一步。 方法学社区已充分意识到问题：多项基准工作系统比较了 IVW、MR-Egger、加权中位数、MR-PRESSO、MR-APSS 等估计量在不同工具阈值（5×10^{-8} 至 5×10^{-5}）下的表现，并指出部分方法对阈值高度敏感、部分方法（如 MR-APSS）复现性更好；2026 年一项工作重新评估了两样本 MR 中的工具强度问题。STROBE-MR 报告规范也已发布。

空白在于：这些工作都在**方法层面**做模拟与基准比较，没有人对**已发表的具体因果宣称**做一次系统的、以更新数据为自变量的复核审计。 也就是说，我们知道 MR 方法可能不稳，但不知道文献中已经写进摘要的那些结论有多少经得起数据更新。这个空白适合学生课题：GWAS 摘要统计是聚合级公开数据、算力近乎为零、方法学完全公开、**结论正负都成立**（重现率高则为 MR 文献正名；重现率低则给出一个可引用的折扣系数与风险因子清单）。

3 · 可检验假设

- H1：在更新后的 GWAS 上重跑，已发表 MR 宣称中效应方向一致且 BH-FDR < 0.05 的比例 ⩽ 60%；且该重现率随原研究工具变量的平均 F 统计量分档单调上升，最高档与最低档之差 ⩾ 25 个百分点（自助法 95% 置信区间不跨 0）。
- H2（机制假设）：原文**未完整报告多效性敏感性分析**（未同时报 MR-Egger 截距、加权中位数与留一分析）的宣称，其重现率显著低于完整报告组，差值 ⩾ 20 个百分点；且不重现的宣称在重跑中更常出现 MR-Egger 截距显著（水平多效性证据）与 Cochran's Q 异质性显著。若 H2 被否证，说明报告完备性不是有效的质量筛选信号，则需另找预测子（如样本重叠比例、暴露的遗传力）。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：先做"同数据复现"——选 10 条原文明确给出所用 GWAS 数据集编号与工具 SNP 清单的宣称，在**原文所用的同一版本 GWAS** 上用自建管线重跑，要求 ⩾ 8 条的 IVW 效应估计与原文报告值的相对偏差 ⩽ 15%，且 95% 置信区间与原文区间重叠。**若 < 8 条达标，说明管线或数据对齐有误，全部后续审计结论无效。**
2. **统计口径预先写死**：(a) 主估计量为逆方差加权（IVW，随机效应），并强制同时报 MR-Egger、加权中位数、加权众数四种估计量，"重现"定义为 **IVW 方向一致 + BH-FDR < 0.05 + 至少两种稳健估计量方向一致**，定义在开

工前冻结；(b) 多重检验一律报 BH-FDR，**不报裸 p 值**；(c) 每条分析必须报平均 F 统计量（工具强度）与 Cochran's Q（异质性），F < 10 的分析标记为弱工具并单独统计；(d) 工具选择阈值固定为 5×10^{-8} + 局部 LD 剪枝（$r^2 < 0.001$，10 Mb 窗），并另做 5×10^{-6} 的敏感性分析，两套结果都报；(e) 重现率的分档比较用 1,000 次自助法（按宣称重抽样）给 95% 置信区间；(f) 样本量为 60 – 100 条宣称，须先做检验力估计：在此规模下能以 80% 检验力检出的最小分档差值是多少，写入方案；若大于 25 个百分点，必须扩大宣称数。

3. **样本重叠必须显式处理**：两样本 MR 的核心假设之一是暴露与结局 GWAS 来自不重叠人群。每条分析必须估计并报告样本重叠比例（依据各 GWAS 的队列构成），重叠 > 20% 的分析单独标记并做偏倚方向讨论；这一项在大量已发表 MR 中被忽略，本项目必须做。
4. **数据版本必须成对记录**：每条宣称记录"原文所用 GWAS（编号、样本量、发表年）"与"更新版 GWAS（编号、样本量、发表年）"两行，样本量增幅 < 30% 的对不纳入"更新"分析（因为不构成有意义的独立性提升）。
5. **发表文献与公开摘要统计的地位**：所有 GWAS 摘要统计与已发表 MR 结论**仅作为输入与对照，不计入本项目的数据贡献**。
6. **可复现性**：宣称抽取的检索式与纳入/排除流程图（PRISMA 式）、每条分析的 GWAS 编号、工具 SNP 清单、随机种子与全部脚本开源；提供 10 条宣称的降规模复现包（≤ 30 分钟可跑完）。

5 · 数据与工具

用途	来源 / 工具
GWAS 摘要统计（主入口）	GWAS Catalog 摘要统计 FTP（https://www.ebi.ac.uk/gwas/downloads/summary-statistics ），**完全开放，无需注册**；FinnGen 公开数据释放（https://finngen.gitbook.io/documentation/ ，开放下载）；UK Biobank Neale lab round 2（开放）
GWAS 摘要统计（便捷入口）	IEU OpenGWAS（https://gwas.mrcieu.ac.uk/ ）+ R 包 `ieugwasr`。**注意：OpenGWAS 已改为需要免费 API token 并设有速率限制，部分数据集访问受限，当前政策需核实**；因此方案以 GWAS Catalog / FinnGen 的直接下载为主线，OpenGWAS 仅作加速手段
受控访问核查	本课题**只使用聚合级（summary-level）统计量，不接触任何个体水平基因型或表型**，因此完全不涉及 dbGaP/UK Biobank 的受控访问审批。数据为等位基因频率与回归系数的汇总，不涉及可识别个人信息，无需伦理审批
MR 分析	R 包 `TwoSampleMR`（IVW / MR-Egger / 加权中位数 / 加权众数 / 留一分析 / 漏斗图）、`MRPRESSO`（水平多效性离群检测，10,000 次置换在 50 个 SNP 上约数十秒）、`MendelianRandomization` 作为第二实现做交叉校验
LD 剪枝参考面板	1000 Genomes EUR 子集（公开），配合本地 `plink 1.9` 做 clumping，避免依赖 OpenGWAS 的在线 clump 接口（该接口有配额）。参考面板约 1–2 GB
已发表 MR 宣称的抽取	PubMed 检索（英文检索式须写入论文），限定近五年、限定"two-sample Mendelian randomization"且摘要中给出明确因果方向宣称；按预先写死的纳入/排除规则由**两名队员独立筛选**，不一致处记录并裁决，报 Cohen's κ
算力口径	纯 CPU。单条 MR 分析（harmonize + 5 种估计量 + MR-PRESSO）约 10 s–2 min；100 条约 1–3 小时。主要成本是摘要统计下载（单个全基因组摘要统计 100 MB–2 GB，20–40 个即 10–20 GB 磁盘）与数据整理。**无算力风险，风险全部在文献抽取与数据对齐的工作量上**
对照基准	各原文报告的效应量与置信区间；MR 方法基准论文（如 2024 AJHG 的 MR 方法基准）给出的方法特性。**仅用于校验与对比，不计入本项目的数据贡献**

6 · 方法路径

1. 装环境（R + TwoSampleMR + plink），跑通包内示例（经典的 BMI → 冠心病算例），**先完成验收标准第 1 条的 10 条同数据复现**，把复现表写入 `validation/`。
2. 按预先写死的检索式与纳入/排除规则抽取候选宣称，两人独立筛选并报 κ；生成 PRISMA 式流程图与"宣称 – 暴露 – 结局 – 原文 GWAS – 效应量"总表并冻结。

3. 为每条宣称配对"更新版 GWAS"（同性状、样本量增幅 ⩾ 30%、发表晚于原文），配对规则写死；无法配对的宣称单独列出并统计比例。
4. **核实数据对齐**：效应等位基因、等位基因频率、链方向（palindromic SNP 处理规则）、基因组坐标版本必须逐条核对——这是 MR 中最常见且最隐蔽的错误源，须写成自动化检查脚本并留日志。
5. 在更新版数据上统一重跑，记录五种估计量、F 统计量、Cochran's Q、MR-Egger 截距、MR-PRESSO 结果、样本重叠估计；按冻结的"重现"定义判定。
6. 检验 H1 与 H2：按 F 统计量分档、按原文报告完备性分组比较重现率，自助法给区间，BH 校正。
7. 独立交叉校验：其一，用 `MendelianRandomization` 包（独立实现）重算全部 IVW 估计，验证结果不依赖单一软件；其二，做**反向 MR**（结局→暴露）检验反向因果；其三，用一组随机配对的"安慰剂"暴露 – 结局对（⩾ 100 对）建立假阳性零模型，报本管线在无真实因果时的显著率，作为整套判定的基线。

7 · 新颖性边界

> 本课题**不**提出新的 MR 估计量、**不**声称任何新的因果关系、**不**声称"MR 方法对工具阈值敏感"或"弱工具会导致偏倚"为本项目发现——这些是 MR 方法学的既有共识，已由多篇基准论文与 STROBE-MR 规范确立。
>
> 已有工作具体完成了什么：方法学基准工作（如 2024 年 AJHG 的 MR 方法基准、2026 年关于工具强度的重评估）在**模拟数据与少量示例**上比较了各估计量在不同工具阈值下的一类错误率与功效，指出 MR-APSS 等方法对阈值不敏感；STROBE-MR 给出了报告清单。**它们评估的是方法，不是已发表的具体宣称；也没有把"数据更新"当作自变量。**
>
> 本项目的贡献（且是主结论）：把评价对象从"方法"换成"**已发表宣称**"，把自变量换成"**GWAS 数据版本更新**"，交付一个可引用的重现率数字与一份**可事前使用的风险因子清单**（F 统计量分档、报告完备性、样本重叠）。这是研究手册所列"换样本做独立检验"这一最稳差异化轴在遗传流行病学中的应用。
>
> 为什么有价值：MR 结论已被大量综述与临床讨论引用。若重现率明显低于 60%，读者需要一把折扣尺与一份筛选规则；若重现率高，MR 这一低成本因果推断工具获得一次真正的外部验证。**两种结果都构成有效结论。**
>
> 风险声明：本课题的最大方法学困难是"更新版 GWAS 与原版 GWAS 存在样本重叠"——新版通常包含旧版样本，因此**不是严格独立的复现**。这个困难必须在论文中显式讨论并量化（报告每对的样本增量与重叠比例），把结论表述为"数据更新后的稳定性"而非"独立复现"。**这个困难本身就是研究内容。**

8 · 决策门槛（go / no-go）

- **第 6 周末（硬门槛）**：验收标准第 1 条必须通过。若 10 条同数据复现中 < 8 条达标，先排查等位基因对齐与 clumping 参数；四周内仍不通过则**降级**：把复现对象限制到"原文公开了完整工具 SNP 表（含 beta/se）"的宣称（这类文献比例较低但对齐无歧义），并把审计规模相应缩小到 40 条。主结论框架不变。
- **第 10 周末**：确认可抽取且可配对更新 GWAS 的宣称数量。若 < 40 条，**立即降级**：其一，把"更新"定义放宽为"任一未被原文使用的独立 GWAS（如 FinnGen 对应性状），不要求样本量更大"，此时问题从"数据更新稳定性"变为"**跨队列可移植性**"，主结论框架完全保留；其二，把宣称范围收窄到单一疾病领域（如心血管或精神疾病）以提高配对成功率，并相应收窄推广范围。
- **第 16 周末**：完成安慰剂零模型（第 7 步）。若本管线在随机配对上的假阳性率 > 10%（远高于名义 5%），**立即降级**：把判定标准从"FDR < 0.05"改为"经零模型经验校准后的阈值"，并把零模型校准本身作为一条独立方法学结论报出。
- **第 28 周末**：若 H1 与 H2 都未达阈值（重现率高且无分档差异），按风险声明写"MR 文献稳健"结论，并把安慰剂零模型与样本重叠量化作为核心证据支撑该结论的可信度。
- **第 36 周**结果冻结，第 44 周英文 PPT 初稿。

- **需提前核实而非边做边发现**：(a) OpenGWAS 的 API token 政策与配额（若不可用，主线必须完全走 GWAS Catalog/FinnGen 直接下载）；(b) 目标性状在 GWAS Catalog 中是否有可下载的**全基因组**摘要统计（很多研究只公开显著位点，无法做 clumping 与 MR）；(c) 各 GWAS 的队列构成文档是否足以估计样本重叠。三项在第 8 周前查清——**其中 (b) 是本课题最可能致命的前置风险**。
- **选择前提**：适合能读懂英文方法学文献、有耐心做系统性文献筛选的两人小队（第 2 步需要两人独立筛选以算 κ）。本路线技术门槛低但文献工作量大，不适合只想写代码的学生。
- **预算裁剪顺序**：先砍反向 MR，再砍 5×10^{-6} 阈值敏感性分析，再砍 MR-PRESSO（保留 MR-Egger 截距作为多效性证据），最后把宣称数从 60 降到 40。砍到只剩"40 条宣称 × IVW + 加权中位数 × F 分档重现率 + 安慰剂零模型"时，主结论仍成立。

B09 · 流感预测目标变量的选择如何改变模型排名：在 WHO FluNet 的 30 个无预测中心国家上，以绝对病例数 vs 阳性率为双目标的回顾性加权区间评分对照

`优先级：中` · `分族：流行病学建模` · `参赛子类：传染病建模 / 预测评估` · `资源需求：纯 CPU 笔记本，数据 < 200 MB，全流程数小时` · `技能取向：Python/R 时间序列 + 预测评分规则`

1 · 研究问题

在 WHO FluNet 覆盖但没有官方预测中心（forecasting hub）的 ⩾ 30 个国家中，对同一批候选预测模型做 1 – 4 周提前期的回顾性预测评估，当把预测目标从"每周实验室确诊流感病例绝对数"换成"**流感检测阳性率（percent positivity）**"时，模型按加权区间评分（weighted interval score, WIS）的排名会发生多大改变？排名改变的幅度能否由该国检测量（tested specimens）的年际波动幅度解释？

2 · 研究背景与空白

背景。 传染病预测已高度制度化：美国 CDC 的 FluSight、欧洲的 RESPICAST 等预测中心每周汇集多家团队的概率预测，用 WIS 与覆盖率统一评分。FluSight 的评估结果（Nature Communications, 2024）显示，2021 – 22 季 23 个模型中只有 6 个跑赢基线模型，2022 – 23 季 18 个中有 12 个；集合模型（ensemble）分别排第 2 与第 5。基线模型构造极简：中位数取上周观测值，区间由历史一周差分的正负值生成。

但这些评估的前提是有一个**质量稳定的目标变量**——美国用的是实验室确诊住院数，报告体系统一。全球大多数国家没有这个条件：WHO FluNet 汇总各国国家流感中心（NIC）上报的每周检测数与阳性数，而各国的检测量随疫情关注度、财政与政策剧烈波动。同一波真实疫情，在检测量翻倍的年份会表现为病例数翻倍。

空白在于：预测中心的评估体系从未在"报告伪迹显著"的环境中被检验过，也没有人系统比较过"绝对计数"与"阳性率"这两个目标变量会给出多不一致的模型排名。而全球绝大多数国家恰恰处在这种环境中，且 2020 – 2023 年的疫情扰动使检测量的波动达到历史极值。这个空白适合学生课题：数据完全开放且体积极小、算力可忽略、评分方法完全公开、**结论正负都成立**。

3 · 可检验假设

- **H1**：以国家为单位比较两个目标变量下的模型排名，Kendall τ 的中位数 $\leqslant 0.6$；且在 $\geqslant 25\%$ 的国家中，"最佳模型"在两个目标下不是同一个。
- **H2**（机制假设）：排名不一致度（1 − Kendall τ）与该国 2015 – 2025 年周检测量的变异系数呈正相关（Spearman $\rho \geqslant 0.5$，BH-FDR < 0.05）；且在检测量波动最小的四分位国家中，两目标排名基本一致（$\tau \geqslant 0.8$）。若 H2 被否证，说明不一致由疫情本身的动力学特征而非报告伪迹驱动，需另作解释。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建评分代码在 **FluSight 公开数据仓库**（模型预测与真实值均公开于 GitHub）上重算 2021 – 22 与 2022 – 23 两季的 WIS 与相对 WIS，要求：(a) 复现"2021 – 22 季 23 个模型中 6 个跑赢基线、2022 – 23 季 18 个中 12 个跑赢"这一已发表计数，误差 $\leqslant 1$ 个模型；(b) 复现集合模型的名次（2021 – 22 第 2、2022 – 23 第 5），名次偏差 $\leqslant 1$ 位；(c) 自建的 WIS 数值与官方评分文件在共有条目上的相对偏差 $\leqslant 1\%$。**三项任一不达标，评分代码不可信，后续全部结论无效。**
2. **统计口径预先写死**：(a) 主指标为 WIS 与相对 WIS（以基线模型归一化），并报 50%/95% 预测区间覆盖率；**不报点预测的 MAE 作为主指标**，原因是概率预测的质量必须用恰当评分规则（proper scoring rule）衡量，点误差会奖励过度自信的模型；(b) **时间序列必须按时间前后划分**：所有模型只用预测时点之前的数据训练，且必须模拟"实时可得性"——若某国数据存在回溯修订，须使用可得的最早版本或明确说明不可得并列为限制；额外做一

次"随机划分 vs 时间划分"对照，量化随机划分对性能的高估幅度；(c) 跨国比较报 BH-FDR；(d) 所有 WIS 汇总量给 ⩾ 1,000 次自助法（按"国家-季"块重抽样）95% 置信区间；(e) 排名一致性用 Kendall τ 并给自助区间。

3. **报告伪迹必须显式量化**：每个国家逐年报告检测量、阳性率与两者的相关；检测量周报告缺失率 > 20% 的国家-季剔除并列出。
4. **规模下限**：⩾ 30 个国家、⩾ 5 个流感季（须同时含疫情前 2016 – 2019 与疫情后 2023 – 2026 两段）、⩾ 5 个候选模型（须含 FluSight 式随机游走基线、历史同周中位数基线、ARIMA、带季节项的 GAM、一个简单机制模型如离散 SIR/SEIR 或 Serfling 回归）。每个国家-季的可评估周数 ⩾ 20。
5. **模型与数据的地位**：FluNet 的病例计数、FluSight 公开的他人模型预测**仅作为输入与校验对照，不计入本项目的数据贡献**。
6. **可复现性**：数据快照（含下载日期）、全部脚本、随机种子、逐国逐季评分表开源；提供 5 国的降规模复现包（⩽ 30 分钟可跑完）。

5 · 数据与工具

用途	来源 / 工具
全球流感监测周数据	WHO FluNet / FluMart（https://www.who.int/tools/flunet，经 GISRS 全球流感监测系统）。含各国每周处理标本数、A/B 型阳性数、分型结果，覆盖 2010 年至今、100+ 国家。**可导出 CSV，全球全时段体积仅数十 MB**，无算力压力。**当前导出接口与字段名需核实**（WHO 近年多次改版数据门户）
欧洲对照数据	ECDC ERVISS（European Respiratory Virus Surveillance Summary，开放 CSV），用于与 RESPICAST 覆盖区域做对照
校验用的预测与评分基准	FluSight 公开仓库（`cdcepi/FluSight-forecast-hub`，GitHub，含全部提交模型的概率预测、真实值与官方评分）。**仅用于校验与对比，不计入本项目的数据贡献**
评分实现	R 包 `scoringutils`（WIS、覆盖率、相对技巧的标准实现）；同时**自行实现一遍 WIS 公式**做交叉校验（WIS 是分位数损失的加权平均，公式须在论文中写出）
预测模型	R `forecast` / `fable`（ARIMA、ETS）、`mgcv`（季节 GAM）、简单 SEIR（自行实现，欧拉离散，每次拟合毫秒级）、Serfling 回归。概率预测通过残差自助法或分位数回归产生
算力口径	纯 CPU。30 国 × 5 季 × 25 周 × 4 提前期 × 5 模型 ≈ 75,000 次预测，单次拟合 10–500 ms，合计 **2–10 小时**，可后台跑。数据体积 < 200 MB。**无算力风险**
伦理	FluNet/ERVISS 为国家层面聚合监测数据，不涉及个体、不涉及可识别个人信息、不涉及人体或动物实验，无需伦理审批

6 · 方法路径

1. 装环境（R + scoringutils + fable + mgcv），下载 FluSight 公开仓库，**先完成验收标准第 1 条的三项复现**，把复现表与已发表数值并列写入 `validation/`。
2. 下载 FluNet 全量数据，按预先写死的规则筛选国家-季（缺失率 < 20%、可评估周数 ⩾ 20、峰值阳性数 ⩾ 50），生成"国家 – 季 – 检测量 – 阳性数 – 阳性率"总表并冻结；同步计算各国检测量变异系数。
3. 构建实时预测框架：对每个国家-季、每个预测时点，只用该时点之前的数据拟合五个模型，输出 23 个分位数的概率预测（与 FluSight 格式一致），提前期 1 – 4 周。
4. 用**绝对确诊数**为目标做完整评分，得到每国的模型排名与相对 WIS。
5. 用**阳性率**为目标重跑第 3 – 4 步（模型族相同，仅目标变量与必要的变换不同），得到第二套排名。

6. 比较两套排名：逐国 Kendall τ、最佳模型是否切换、相对 WIS 差值；自助法给区间，检验 H1。再把不一致度对检测量变异系数做回归，检验 H2。
7. 独立交叉校验：其一，在 FluSight 覆盖的美国数据上做同样的双目标对照（美国报告体系稳定，预期不一致度应显著更低），作为"负对照"；其二，把疫情前（2016 – 2019）与疫情后（2023 – 2026）两段分开报，检验不一致度是否在疫情后放大；其三，用 `scoringutils` 与自实现 WIS 双份计算全部得分，验证评分代码无误。

7 · 新颖性边界

本课题**不**提出新的流行病学模型、**不**声称任何关于流感传播机制的新发现、**不**声称"WIS 是恰当评分规则"或"集合模型通常优于单模型"为本项目发现——这些是预测中心文献的既有结论。

已有工作具体完成了什么：FluSight 评估（Nature Communications, 2024）在**美国、单一目标变量（实验室确诊住院数）**上，用 WIS 系统评估了两季共 20 余个模型，报出跑赢基线的模型计数与集合模型名次；ECDC 的 RESPICAST 在欧洲做同类工作。**这些评估都在报告体系稳定的高收入国家、都只用一个目标变量、都没有检验目标变量选择对模型排名的影响**。

本项目的贡献（且是主结论）：把评价维度从"哪个模型更准"换成"**目标变量的选择会不会改变'哪个模型更准'这个答案**"，并在报告伪迹显著的 30 个国家上给出定量刻度与一条解释变量（检测量波动）。美国数据上的负对照使这个命题真正可被否证。

为什么有价值：随着预测中心模式向中低收入国家推广，目标变量的选择将成为一个必须在建制之初就定下的决策。若排名对目标变量高度敏感，则"照搬 FluSight 的目标定义"是有风险的；若不敏感，则推广障碍少一项。

风险声明：**"两个目标下模型排名高度一致"同样是有效结论**，但必须给出自助法误差棒证明本设计有能力分辨 $\tau = 0.6$ 与 $\tau = 0.9$ 的差异；同时必须报告本项目候选模型集合较小（5 个）这一限制——排名一致性在候选模型少时天然偏高，须在讨论中明确。

8 · 决策门槛（go / no-go）

- **第 5 周末**：确认 FluNet 的数据导出接口可用且字段满足需求（须同时含处理标本数与阳性数）。若阳性率无法计算（缺检测量字段），**立即降级**：其一，改用 ECDC ERVISS（明确提供检测量）+ 美国 CDC FluView（提供 percent positivity），国家数降到约 30 个欧洲国家 + 美国各州；其二，把第二个目标变量从"阳性率"换成"**对数变换后的病例数**"，此时研究问题变为"目标变换（尺度选择）对模型排名的影响"，主结论框架（双目标 + 排名一致性 + 解释变量）完全保留。
- **第 8 周末（硬门槛）**：验收标准第 1 条必须通过。若 WIS 数值与官方评分偏差 > 1%，逐项排查分位数集合、缺失预测的处理规则与截断规则；四周内仍不通过则**降级**：放弃"复现模型计数"这一项，改为"自实现 WIS 与 `scoringutils` 在随机构造的预测上完全一致（相对偏差 ⩽ 1e-6）"作为硬门槛，并在论文中说明未能复现官方计数的原因。
- **第 14 周末**：确认合格国家-季数量。若 < 20 个国家，**立即降级**：把可评估周数下限降到 15、峰值阳性数下限降到 30，并把国家数不足这一点写入结论的推广限制。
- **第 22 周末**：完成美国负对照。若美国数据上的不一致度与 FluNet 国家相当（即处处都不一致），说明现象与报告质量无关而是评分本身的性质，**立即转向**：把主结论改述为"WIS 排名对目标变量的普遍敏感性"，并把检测量回归降为次要分析。框架不变。
- **第 36 周**结果冻结，第 44 周英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) FluNet 数据是否存在回溯修订（若有，"实时可得性"模拟无法严格实现，须明确列为限制）；(b) 各国"季"的定义（南北半球流感季不同，须分半球处理）；(c) FluSight 公开仓库中官方评分文件的字段与本项目评分口径是否可直接对齐。三项在第 7 周前查清。

- **选择前提**：适合对时间序列与预测评估有兴趣、且能接受"结论是一条方法学警示"的学生。本路线是全部十条中算力与数据风险最低的之一，主要风险在数据接口稳定性。
- **预算裁剪顺序**：先砍机制模型（SEIR），再砍疫情前后分段分析，再砍提前期（只留 1 周与 4 周），最后把国家数从 30 降到 20。砍到只剩"20 国 × 3 季 × 4 模型 × 双目标 WIS 排名对照 + 美国负对照"时，主结论仍成立。

B10 · 纯 CPU 预算下的蛋白质变异效应零样本预测：语言模型规模、MSA 特征与结构特征的算力—性能帕累托前沿及其在 ProteinGym 分层上的稳定性

`优先级：低（风险最高，需明确接受可能跑不出显著结论）` · `分族：蛋白质序列与结构` · `参赛子类：计算蛋白质科学 / 机器学习` · `资源需求：纯 CPU 笔记本，模型权重 ≤ 3 GB，推理需数十小时后台运行` · `技能取向：Python + PyTorch 推理 + 基准评估`

1 · 研究问题

在"一台 16 GB 纯 CPU 笔记本"这一硬预算下，蛋白质语言模型（protein language model, PLM）的参数规模、基于多序列比对的位点无关基线、以及 AlphaFold 结构派生特征三者的组合，在 ProteinGym 深度突变扫描（deep mutational scanning, DMS）基准上构成怎样的算力—性能帕累托前沿？该前沿的形状在按 MSA 深度与实验类型分层后是否稳定，还是存在某一层上"小模型 + MSA 特征"反超大模型的区域？

2 · 研究背景与空白

背景。 零样本变异效应预测指：不使用任何该蛋白的实验数据，仅凭序列（或结构）模型给出的似然，预测每个氨基酸突变对蛋白功能的影响。它是变异致病性判读与蛋白工程的基础工具。ProteinGym 是该任务的标准基准，收录 200+ 个 DMS 实验、约百万级突变，并公开每个蛋白的预计算 MSA 与各方法的官方得分。

已有工作到哪一步。 ESM-2 系列（8M 至 15B 参数）的零样本性能已被系统报告，ESM2-3B 在 ProteinGym 上的平均 Spearman 约 0.443；基于结构的零样本方法（ESM-IF、ProteinMPNN 及其变体）与多模态方法（如 ProtSSN 等，Cell Research, 2024）也已有大量比较；2025 年有工作专门探讨"结构基零样本适应度预测"，指出 ESM-IF 因训练于 AlphaFold2 预测结构，在预测结构上反而优于实验结构。ProteinGym 官网本身即提供按 MSA 深度（低/中/高）与功能类别的分层排行榜。

空白在于：所有这些比较都以"性能"为唯一坐标，**没有任何一项把推理算力预算作为第二坐标**，也没有人在"纯 CPU、无 GPU"这一真实约束下刻画可行前沿。而这恰是绝大多数中学、地方院校与低资源实验室的实际处境：15B 模型不可能在笔记本上跑，3B 模型的权重在 fp32 下就要 12 GB。哪一档模型才是这个预算下的最优选择、能否用便宜的 MSA 特征把小模型补到大模型水平——这是一个有明确实用价值却无人回答的问题。**这条路线风险最高**，因为答案可能平淡（"越大越好，无反超区域"），因此它排在第 10 位，且必须与更稳的路线并行才可启动。

3 · 可检验假设

- H1：在 CPU 推理时间预算轴上，ESM-2 的性能—算力曲线在 150M 参数附近出现明显拐点——从 8M 到 150M 的平均 Spearman 增益 $\geqslant 0.08$，而从 150M 到 650M 的增益 $\leqslant 0.04$（推理时间却增加 $\geqslant 4$ 倍），两段增益之差的 95% 自助法置信区间（按 DMS 实验重抽样）不跨 0。
- H2（反超假设）：存在至少一个分层——预期为"MSA 深度高 + 蛋白长度短"的子集——在该层上"ESM-2 35M 的对数似然与 MSA 位点无关基线（PSSM）的秩和组合"的 Spearman 不低于 ESM-2 650M 单独使用，差值 95% 置信区间下界 $\geqslant -0.01$；而在"MSA 深度低"的层上不存在此反超。若 H2 在所有分层上都被否证，则结论为"低资源场景下没有捷径，模型规模不可替代"——这同样是一条对使用者有用的负面结论。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用自建推理与评分脚本在 ProteinGym 上复现**官方发布的 ESM-2 650M 零样本得分**：(a) 在 $\geqslant 30$ 个 DMS 实验的子集上，本项目计算的突变得分与 ProteinGym 官方发布的对应模型得分的 Spearman 相关 $\geqslant 0.95$；(b) 本项目算得的逐实验 Spearman 与官方排行榜数值的绝对偏差 $\leqslant 0.02$，均值偏差 $\leqslant 0.01$。**任一项**

不达标，推理或评分实现有误（最常见原因是打分策略选择——masked-marginals / wt-marginals / mutant-marginals 三者结果不同），后续全部前沿结论无效。

2. **统计口径预先写死**：(a) 主指标为**逐 DMS 实验的 Spearman 秩相关，跨实验取中位数与均值并都报**，不做突变级汇总（因为不同实验的突变数极不均衡，汇总会被少数大实验主导）；(b) 若做"有害/中性"二分类的补充分析，报 **PR-AUC 与 MCC，不报 accuracy**，原因是多数 DMS 中有害突变占比失衡，accuracy 会被多数类支配；(c) 所有模型间比较用按 **DMS 实验整体重抽样**的自助法（≥ 1,000 次）给 95% 置信区间，分层比较报 BH-FDR；(d) 算力必须以**同一台机器、同一线程数、同一 batch 设置**实测的墙钟时间与峰值内存报告，且必须给出"每单位 Spearman 增益的秒数"这一双成本口径；(e) 样本量为 DMS 实验数，须先做检验力估计：在 n 个实验下能以 80% 检验力检出的最小 Spearman 差值是多少，写入方案。
3. **算力预算必须显式声明并遵守**：全部结果必须在"16 GB 内存、纯 CPU"下产生；任何超出该预算的模型（见下表）一律标注为"未评估"而**不得**引用他人 GPU 结果冒充本项目结果。
4. **规模下限**：≥ 60 个 DMS 实验（须覆盖 ≥ 3 类实验终点：稳定性、结合、活性/适应度），每个实验 ≥ 500 个突变；≥ 4 档模型规模；≥ 3 类分层维度（MSA 深度、蛋白长度、实验终点）。
5. **公开基准与他人模型的地位**：ProteinGym 的 DMS 数据、预计算 MSA、官方得分文件，以及全部预训练模型权重，**均为他人成果，仅作为输入、校验与对照，不计入本项目的数据贡献**。
6. **可复现性**：全部推理脚本、打分策略实现、计时脚本、硬件规格、模型权重的 checksum、随机种子开源；提供 10 个 DMS 实验 + 两档模型的降规模复现包（≤ 3 小时可跑完）。

5 · 数据与工具

用途	来源 / 工具
DMS 基准与预计算 MSA	ProteinGym（https://proteingym.org/ ，数据托管于官方下载链接/Zenodo）。含 200+ DMS 实验、参考序列、突变表、**预计算 MSA** 与各方法官方得分。**MSA 打包文件体积较大（数十 GB 量级），需核实并按需只下载目标实验的 MSA**
蛋白质语言模型	Hugging Face `facebook/esm2_t6_8M_UR50D` 、`esm2_t12_35M` 、`esm2_t30_150M` 、`esm2_t33_650M` 。fp32 权重分别约 0.03 / 0.14 / 0.6 / 2.6 GB，**650M 及以下均可在 16 GB 内存下推理**
明确超出范围的模型	ESM-2 3B（fp32 权重约 12 GB，加上激活值在 16 GB 机器上不可行）、ESM-2 15B、ESM-3、以及任何需训练的方法（EVE、DeepSequence 需 GPU 训练数小时至数天）。**这些一律标注"未评估"，不得引用他人结果充数**
MSA 基线	自实现的位点无关模型（PSSM / 独立位点频率，含伪计数与序列加权），纯 numpy，秒级；以及 **GEMME**（Java，MSA 基，CPU 上单蛋白分钟级，在 ProteinGym 上性能接近顶尖方法）——这是本课题最重要的低算力对照
结构特征	AlphaFold DB 已发布模型的 pLDDT 与相对溶剂可及性（DSSP），以及残基接触数。**不运行任何结构预测推理**（理由与 B02 相同：AlphaFold2 完整流程需约 2.2 TB 数据库与 GPU；ColabFold 在 CPU 上单蛋白数小时，不可用于 60+ 蛋白的规模）
推理与评分	Python + PyTorch（CPU 后端，`torch.set_num_threads` 固定线程数）、`transformers` 、`scipy.stats.spearmanr`
算力口径（须实测校准，以下为规划估计）	打分策略决定成本：**wt-marginals 每个蛋白只需 1 次前向传播**；**masked-marginals 需要 L 次前向传播**（L 为序列长度），是 ESM 论文的标准做法但成本高 L 倍。以 L=300、8 线程为例，单次前向传播粗估 8M ≈ 0.05 s、35M ≈ 0.15 s、150M ≈ 0.5 s、650M ≈ 2 s；masked-marginals 下 650M 单蛋白约 600 s，60 个蛋白约 **10 小时**（长蛋白会显著更久）。四档模型全跑约 **13–20 小时**，可后台分夜运行。**这些数字必须在第 3 周用实测替换，规划估计不得写进论文**
对照基准	ProteinGym 官方排行榜与官方得分文件。**仅用于校验与对比，不计入本项目的数据贡献**
伦理	全部为公开序列、结构与已发表实验数据，不涉及人体实验、动物实验与可识别个人信息，无需伦理审批

6 · 方法路径

1. 装环境（PyTorch CPU + transformers），下载 ProteinGym 参考数据与官方得分，**先完成验收标准第 1 条的 650M 复现**；特别注意打分策略必须与官方一致，把三种策略的结果差异一并记录到 `validation/`（这本身是有用的方法学信息）。
2. **第一件事是实测算力**：在目标笔记本上测四档模型在不同序列长度下的单次前向时间与峰值内存，画出成本曲线，据此确定可评估的 DMS 实验子集（长度上限）。此步决定整个课题的规模，**不得跳过或用规划估计代替**。
3. 按预先写死的规则选定 ⩾ 60 个 DMS 实验（覆盖三类终点、长度 ⩽ 实测可行上限、突变数 ⩾ 500），并从 ProteinGym 按需下载对应 MSA。
4. 跑四档模型的零样本打分（固定 masked-marginals，若算力不足则固定 wt-marginals 并明确声明），逐实验算 Spearman，实测记录墙钟时间；画性能—算力帕累托前沿，检验 H1。
5. 实现 PSSM 基线与 GEMME 基线，同样记录算力；构造"小模型 + MSA 基线"的秩组合（简单的秩平均与按 MSA 深度加权两种组合规则，规则预先写死，不得事后调参）。
6. 按 MSA 深度、蛋白长度、实验终点三维分层，比较各组合的 Spearman，自助法给区间，BH 校正，检验 H2；明确报告是否存在反超区域及其边界。
7. 独立交叉校验：其一，加入结构派生特征（pLDDT、相对溶剂可及性）作为第三类廉价特征，检验其增益是否独立于 MSA 特征；其二，用一个 ProteinGym **未参与前沿构造**的留出实验子集（⩾ 15 个）做外部验证，报前沿结论在留出集上的成立情况——防止分层选择造成的事后挑选偏倚。

7 · 新颖性边界

> 本课题**不**训练任何模型、**不**提出新的打分方法、**不**声称任何模型性能超过已发表的最优方法，**不**声称"更大的 PLM 通常更好"或"结构特征有帮助"为本项目发现——这些已由 ESM-2 论文、ProteinGym 排行榜、2025 年结构基零样本适应度工作与 Cell Research 2024 的多模态工作确立。
>
> 已有工作具体完成了什么：ProteinGym 发布了 200+ DMS 的统一基准、预计算 MSA、官方得分与**按 MSA 深度与功能类别的分层排行榜**；ESM-2 的规模—性能关系（ESM2-3B 平均 Spearman 约 0.443）已公布；结构基方法与多模态方法的比较也已系统开展。**这些工作全部在 GPU 环境下进行，性能是唯一坐标轴；没有任何一项报告 CPU 推理成本，也没有任何一项刻画"给定算力预算下的最优选择"**。
>
> 本项目的贡献（且是主结论）：把评价维度从"性能"换成"**性能—算力二维帕累托前沿**"，在一个明确声明的硬件预算下给出可执行的选型建议，并检验一个可否证的命题（H2 的反超区域）。这属于研究手册所列"评价维度的转换"与"方法层面的可复现性贡献"，在计算领域是被承认的贡献类型；但**定位必须讲清楚，否则会被误读为重复基准测试**——答辩时这是最可能被追问的一点，须提前准备英文表述。
>
> 为什么有价值：变异效应预测的使用者中，有大量人手上只有 CPU。现有排行榜按性能排序，对他们没有直接指导意义；一条带算力坐标的前沿把"我应该用哪个模型"变成可读数的答案。
>
> 风险声明：**本课题最可能的负面结果是"前沿平凡"——即性能随算力单调增加、无拐点、无反超区域**。这仍是有效结论（它明确否定了"小模型 + 廉价特征可替代大模型"这一在低资源社区中流行的期望），但价值低于发现反超区域。**必须由第 2 条的检验力估计证明本设计有能力检出 0.04 的 Spearman 差值**，否则连负面结论也不成立。这正是本课题排在第 10 位的原因。

8 · 决策门槛（go / no-go）

- **第 3 周末（第一道硬门槛）**：完成算力实测。若 650M 模型在 masked-marginals 下的实测速度使 60 个实验需要 > 60 小时，**立即降级**：其一，打分策略改为 wt-marginals（成本降 L 倍，但性能会下降，须在论文中明确标注并

用 10 个实验量化两种策略的性能差）；其二，把最大模型档从 650M 降到 150M，前沿的横轴范围相应收窄。两条降级都保留"性能—算力前沿 + 分层稳定性 + 反超检验"的主结论框架。

- **第 7 周末（第二道硬门槛）**：验收标准第 1 条必须通过。若与官方得分的 Spearman < 0.95，逐项排查打分策略、突变编号偏移（1-based vs 0-based）、以及 ESM 的特殊 token 处理；四周内仍不通过则**本课题终止**，学生转入 B02（同为结构/序列分析族、共用 AlphaFold DB 与 Python 技能栈，前期投入不浪费）。**这是全部十条路线中唯一设有"终止并转线"门槛的课题，启动前必须让学生知情。**
- **第 12 周末**：确认 ProteinGym MSA 的下载可行性。若 MSA 无法按需下载（只提供整包且体积超出磁盘），**立即降级**：改用自行从 UniRef50 构建的浅 MSA（MMseqs2，单蛋白分钟级，但需 UniRef50 数据库约 30 – 60 GB 磁盘），或彻底放弃 MSA 分支，把 H2 改述为"小模型 + **结构特征**能否反超大模型"。框架保留。
- **第 24 周末**：若 H1 的拐点不存在且 H2 无反超区域，按风险声明写"前沿平凡"结论，并把实测算力表与留出集外部验证作为核心交付物。
- **第 36 周**结果冻结，第 44 周英文 PPT 初稿。
- **需提前核实而非边做边发现**：(a) ProteinGym 当前版本的 DMS 数量、MSA 打包方式与官方得分文件格式；(b) ESM-2 各档权重在 Hugging Face 上的可下载性与 fp32/fp16 CPU 推理的数值一致性（fp16 在 CPU 上可能不加速甚至更慢，须实测）；(c) GEMME 的安装依赖（Java + MSA 格式要求）。三项在第 5 周前查清。
- **选择前提：仅在学生已具备 Python/PyTorch 基础、已完成至少一条更稳路线的前期训练、且明确接受"可能跑不出显著结论"这一风险时才启动。** 本课题的交付物质量高度依赖实测算力表的严谨性与外部验证的诚实执行；若学生倾向于"调到好看为止"，这条路线会变成事后挑选偏倚的重灾区，反而在答辩中失分。
- **预算裁剪顺序**：先砍结构特征分支（第 7 步前半），再砍 GEMME（保留 PSSM 基线），再砍分层维度（只保留 MSA 深度），最后把 DMS 实验数从 60 降到 35、模型档从 4 降到 3。砍到只剩"35 个 DMS × 三档模型 × PSSM 基线 × 实测算力前沿 + 留出集验证"时，主结论仍成立。

B11 · 城市化梯度上路旁土壤浸提液的植物毒性剂量–反应：以生菜种子萌发与根伸长 EC50 判定毒性排序

`推荐优先级 P0（最稳）` · `实验生态族` · `预算档 ≤ 500 元` · `周期档 短轮次（7 天/轮，全年可跑 30+ 轮）` · `技能取向 定量实验 + 剂量–反应拟合`

1 · 研究问题

沿「城市主干道路肩 → 次干道绿化带 → 城市公园 → 城郊林地」的城市化梯度采集表层土壤，其水浸提液（aqueous extract）对生菜（Lactuca sativa）种子根伸长的半数抑制浓度 EC50，是否随城市化强度单调下降（即毒性单调上升）？若不单调，偏离点能否由可现场测量的协变量（电导率 EC、pH、车流量）解释？

2 · 研究背景与空白

背景。种子萌发 / 根伸长生物测定（seed germination and root elongation bioassay）是 OECD Test No. 208 与 ISO 11269 系列采纳的陆生植物毒性标准方法。其原理是：种子萌发阶段尚无成熟根系屏障与解毒代谢积累，根尖分生组织对水溶态离子和有机物极为敏感，因此根伸长抑制率是水溶态生物有效性（bioavailability）的综合读数，而不是某一化学品的浓度读数。这个性质对中学课题极关键——学校没有 ICP-MS 与 GC-MS，无法测「有什么」，但可以严格定量地测「有多毒」。

已有工作到哪一步。Lactuca sativa 生物测定已被系统用于药物废水（12 种常见药物的胚根 EC50 落在 170 – 5656 mg/L、胚轴 188 – 4558 mg/L）、处理后污水以及土壤可溶性元素的毒性评价。已确立的定量结论有两条：根伸长终点显著比萌发率灵敏（存在浓度高达 3000 mg/L 仍不影响萌发率的案例）；不同生菜品种的剂量 – 反应曲线并不一致，EC50 需逐样本重新测定。

空白在：已发表工作绝大多数是「某一类污染物 → EC50」，而把 EC50 本身当作**空间梯度上的可比排序指标**、并检验它能否被廉价现场协变量预测的城市尺度研究极少，中国城市的标准化重复几乎没有。这个空白适合一年期学生课题：方法完全公开、单轮 7 天、耗材极廉、结论正负都成立。

3 · 可检验假设

- **H1**：路肩土壤浸提液的根伸长 EC50（以浸提液体积分数表示）显著低于城郊林地样点，差异 ⩾ 2 倍，且两者自助法 95% 置信区间不重叠。
- **H2（机制备择）**：lg(EC50) 与浸提液 lg(电导率) 呈负线性关系（|Pearson r| ⩾ 0.6）。若成立，说明毒性主要由可溶盐（融雪剂、道路扬尘）驱动而非有机污染，则冬季采样应显著比夏季更毒。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用参考毒物 $ZnSO_4 \cdot 7H_2O$ 做完整剂量 – 反应，自测根伸长 EC50 与 OECD 208 / 已发表参考区间偏差 ⩽ 30%；同时去离子水对照萌发率 ⩾ 90%、对照组根长变异系数 CV ⩽ 25%。任一项不达标即判该轮无效、全部数据作废并重跑。**此项不过关，后续全部结论无效。**
2. 采样规模写死：⩾ 4 类样点 × 每类 3 个地理独立位置（**生物学重复 n = 3，且必须在不同日期采集**）× 5 个稀释度 × 3 个平行皿（**技术重复 n = 3**）× 每皿 20 粒种子。生物学重复与技术重复在数据表中分列，统计时不得混算。
3. 剂量 – 反应用四参数 log-logistic 模型（R 的 drc 或 Python 的 scipy.optimize）拟合，EC50 必须报**自助法 2000 次重抽样的 95% 置信区间**，并给出残差图与拟合失当（lack-of-fit）检验。

4. 统计口径：萌发数是计数数据，用二项 / 准二项 GLM，**禁止直接对萌发百分比做 t 检验**；根长为连续量，先在皿内取均值消除伪重复，再做单因素 ANOVA + Tukey 事后检验，并检查方差齐性（Levene）与正态性（Shapiro – Wilk）。一律**报效应量（Cohen's d 或偏 η^2）+ 95% CI，不得只报 p 值**。
5. 根长测量用固定支架拍照 + ImageJ/Fiji，同批图像由两名成员盲法各测一次，组内相关系数 ICC ⩾ 0.90；不达标则改为单人盲法测量并在论文中说明。
6. 全部原始图像、测量表与拟合脚本上传公开仓库，第三方一键重跑所得 EC50 差异 ⩽ 1%。

5 · 数据与工具

用途	来源 / 工具
指示种子	生菜（Lactuca sativa）单一批号，一次买齐全年量（⩾ 5000 粒，约 30 元），4 °C 干燥避光保存，记录批号
参考毒物	$ZnSO_4 \cdot 7H_2O$ 分析纯 100 g（约 25 元）；**仅用于方法学校验与对比，不计入本项目的数据贡献**
浸提与培养	9 cm 培养皿 200 套 + 定性滤纸 + 移液器（约 200 元）；恒温培养箱 25 °C 暗培养（学校现有）
现场协变量	便携 EC/TDS 笔 + pH 笔（约 150 元）；土壤含水量用烘干称重法（烘箱现有）；车流量人工 5 分钟计数 × 3 次
降水 / 气温协变量	中国气象数据网公开日值数据（data.cma.cn），用于剔除采样前 7 日强降水样本
图像与统计	ImageJ/Fiji（免费）；Python（scipy、statsmodels）或 R（drc、lme4），纯 CPU，笔记本秒级
方法标准	OECD Test No. 208、ISO 11269-1/2 公开摘要；**仅作规程依据，不计入贡献**

预算合计约 350 – 500 元。

6 · 方法路径

1. 写死标准操作规程（SOP）：浸提比 1:10（w/v）、振荡 2 h、静置澄清 16 h、滤纸润湿体积、种子摆放模板，做成带照片的文档，全程不得中途更改。
2. 完成方法学校验：连续两轮参考毒物 EC50 相对偏差 ⩽ 20%，且对照萌发率达标，方可开始正式采样。
3. 布点采样：4 类 × 3 个独立位置，避开私人地界与保护区；记录 GPS、距路缘距离、车流量、采样日期、近 7 日降水。
4. 每批浸提液当场测 EC、pH，并留 50 mL 冷藏备份，供后续复测与季节对照。
5. 跑剂量 – 反应：5 个稀释度（100 / 50 / 25 / 12.5 / 6.25 %）+ 去离子水对照，25 ° C 暗培养 120 h，第 120 h 一次性拍照。**萌发判据写死为「胚根突破种皮 ⩾ 2 mm」**；判读用编号盲法，双人独立判读并计算 Cohen's κ（要求 ⩾ 0.80）。
6. 拟合 EC50 与置信区间，做梯度排序检验与 lg(EC50) – lg(EC) 回归；第 30 周左右补一轮冬季采样检验 H2。
7. 独立交叉校验：在最毒与最不毒的两个样点上换指示种（萝卜或小麦）重跑一次完整剂量 – 反应，检验毒性排序是否物种依赖。

7 · 新颖性边界

本课题**不**声称提出新的生物测定方法，**不**声称鉴定出具体污染物，**不**声称建立因果链——浸提液毒性与污染源的关联只能写成相关性并明确标注为推测。

已有工作完成了什么：OECD 208 与 ISO 11269 已固定了标准化协议；针对药物、处理后污水、土壤可溶元素的 EC50 已系统发表；「根伸长比萌发率灵敏」与「品种依赖」两条定量结论已确立。历届丘奖生物获奖论文中，2023 年北京王府学校《北京废弃矿山现状与修复策略》、2024 年同校《蚯蚓对土壤微塑料降解的影响》都涉及城市土壤，但前者以现状调查与修复策略为终点，后者以降解率为终点，**均未把带置信区间的标准化 EC50 作为跨样点可比的定量毒性指标**——这是本课题的差异化轴。

本项目的贡献（写成主结论，不是附加内容）：在一条明确定义的城市化梯度上，用同一批种子、同一 SOP、带自助置信区间的 EC50，给出可比且可复现的毒性排序，并检验它能否被一支 20 元的电导率笔预测。若强相关，则交付一条低成本城市土壤毒性快筛判据；若不相关，则证明毒性由非盐组分主导，本身即为有信息量的结论。

「差异不显著」同样是有效结论，但必须用参考毒物校验与置信区间宽度证明本体系有能力分辨 2 倍的 EC50 差异。

8 · 决策门槛（go / no-go）

- **第 4 周末**：方法学校验必须通过。若参考毒物 EC50 连续三轮偏差 > 30%，**立即降级**：其一，终点从根伸长改为萌发率（变异更小、灵敏度低），并把 H1 阈值由 2 倍放宽到 3 倍；其二，换萝卜或小麦作指示种重建参考曲线。两条路径都保留「梯度上的 EC50 排序」主结论框架。
- **第 8 周末**：确认可合法采样的样点 ⩾ 4 类 × 3 个。若城郊林地受保护法规限制，降级为**同一公园内的「距路缘距离梯度」**（0.5 / 5 / 30 / 100 m）——梯度定义变了，分析框架与主结论不变。
- **第 20 周末**：完成第一轮完整数据。若各样点 EC50 的 95% CI 全部重叠，不放弃，转 H2 主线：重心移到协变量回归与季节效应，并把「该城市化梯度不足以产生可检测的毒性差异」作为主结论报告，附检验效能（power）计算说明本设计能分辨的最小差异。
- **第 30 周（2027 年 1 月）**：完成冬季采样。若当地不使用融雪剂，季节对照改为「雨季 / 旱季」或「车流高峰 / 长假期」。
- **需提前核实（第 2 周内）**：种子批号萌发率与整齐度；培养箱温度波动是否 ⩽ ±1 °C；采样点是否受城市绿地管理条例限制（须取得公园管理方书面同意并留存）。
- **预算裁剪顺序**：先砍 pH 笔（改用 pH 试纸，精度降但主结论不受影响）→ 再砍第 7 步第二指示种交叉校验（主结论仍成立，但物种依赖性无法排除，须在讨论中明写）。**不可砍**参考毒物与培养皿数量。

安全与伦理要点

- 全程无微生物培养，无 BSL 风险。$ZnSO_4$ 为刺激性化学品，配制须戴手套与护目镜；含锌废液按学校无机废液流程收集，**不得倒入下水道**。
- 野外采样仅取表层 0 – 10 cm 土壤，每点 ⩽ 500 g，不破坏植被、不涉及保护动植物；进入城市公园须事先取得管理方书面同意。
- 采样若在居民区附近，不拍摄可识别人脸，车流量计数不记录车牌。
- 萌发后的种子材料高温灭活后按一般垃圾处理；生菜为栽培种且全程在培养皿内，无外来种扩散风险。

B12 · 酿酒酵母在乙醇 × 高渗双重胁迫下的产气与生长动力学：以失重法 CO_2 速率与比生长速率判定两种胁迫是否可加

`推荐优先级 P0` · `模式生物实验族` · `预算档 ≤ 600 元` · `周期档 极短轮次（48–72 h/轮，可跑 40+ 轮）` · `技能取向 定量动力学 + 析因设计`

1 · 研究问题

当乙醇胁迫与高渗胁迫（osmotic stress）同时施加于酿酒酵母（Saccharomyces cerevisiae）时，比生长速率 μ 与单位初糖 CO_2 产率的下降幅度，是否等于两种胁迫单独作用时下降幅度的乘积（即 Bliss 独立模型）？若存在显著交互（协同或拮抗），交互的符号是否随碳源（葡萄糖 vs 蔗糖）而改变？

2 · 研究背景与空白

背景。酵母发酵的化学计量是刚性的：1 mol 六碳糖（180 g）→ 2 mol 乙醇（92 g）+ 2 mol CO_2（88 g）。因此在敞开但只允许气体逸出的发酵瓶中，**质量损失速率 dm/dt 就是 CO_2 产生速率**，这使得一台百元级电子天平可以充当发酵动力学在线监测仪，精度和分辨率都足以刻画滞后期、指数期与稳定期。与此并行，用分光光度计测 OD_{600} 得到生长曲线，两者构成同一过程的独立读数——这正是学生课题少见的「双通道自洽性检验」机会。

已有工作到哪一步。乙醇对稳定期细胞发酵动力学的抑制、高糖引起的渗透胁迫（关闭甘油通道 Fps1p、胞内甘油积累以恢复膨压）都已有明确机理描述；施加 0.4 M NaCl 后 CO_2 产生速率立即下降、约 2 h 后完全恢复的动力学也已报道。更关键的一条是：**酵母的渗透胁迫响应是碳源依赖的**（Sci Rep 2017），不同碳源下同一渗透压产生不同的响应强度。

空白在：绝大多数文献分别研究单一胁迫，或在工业「极高浓度发酵」（very high gravity）条件下把两种胁迫捆绑在一起而无法分离；把乙醇 × 渗透压做成**完整析因设计并明确检验交互项符号**、且再叠加碳源这一第三因子的公开数据很少。这个空白适合中学课题：单轮 48 – 72 h，一年可完成 40 轮以上，样本量足以支撑交互项检验（这恰恰是最需要样本量的统计目标），且交互显著与不显著都是有效结论。

3 · 可检验假设

- H1：在葡萄糖为碳源时，乙醇与高渗胁迫对 μ 的联合效应显著偏离 Bliss 独立预测，交互项在双因素 ANOVA 中 $p < 0.05$ 且偏 $\eta^2 \geqslant 0.06$（中等效应量），方向为**协同**（联合抑制强于乘积预测）。
- **H2（碳源依赖备择）**：将碳源换为蔗糖后，交互项的符号或量级发生显著改变（两碳源下交互效应量的 95% CI 不重叠），说明渗透胁迫响应的碳源依赖性会传导到与乙醇胁迫的交互上。

4 · 量化验收标准

1. **方法学校验（硬门槛，两条都要过）**：(a) 失重法标定——在同型发酵瓶中用已知质量 $NaHCO_3$ + 过量酸释放理论量 CO_2，实测失重 / 理论值 ⩾ 95%，重复 5 次 CV ⩽ 3%，并测定纯水在同条件下的蒸发失重本底，本底须 ⩽ 信号的 5% 或被系统扣除；(b) 复现已知抑制曲线——在 0/2/4/6/8/10 % (v/v) 乙醇梯度下，μ 随乙醇浓度单调下降，且无胁迫对照的 μ 落在本实验室连续三轮建立的基线区间内（相对偏差 ⩽ 15%）。**任一条不过关，后续全部结论无效。**
2. 设计写死为 3（乙醇 0/4/8 %）× 3（初糖 2/20/35 % w/v）× 2（葡萄糖/蔗糖）全因子；**生物学重复 = 独立活化的种子液、不同日期接种，n ⩾ 3；技术重复 = 同批次内平行瓶，n = 3**。两者在数据表分列，统计时以生物学重复为分析单位。
3. μ 的提取方式预先写死：对 OD_{600} 取自然对数后，在指数期用滑动窗口线性回归取最大斜率，窗口长度与 $R^2 \geqslant 0.98$ 的判据固定不变；CO_2 通道同样提取最大 dm/dt。两通道给出的胁迫响应排序必须一致，不一致须在

论文中作为方法学发现讨论，不得挑一个报。

4. 统计口径：三因素 ANOVA 检验主效应与全部交互项，事前用 Levene 与残差 QQ 图检查前提，方差不齐则做 log 变换或改用 Welch-ANOVA；**报偏 η^2 + 95% CI，不得只报 p 值**；Bliss 独立性偏差（interaction index）另行给出自助法置信区间。多重比较用 Tukey HSD。
5. 每个处理的 OD 上限须落在分光光度计线性范围内（OD $\leqslant$ 0.8，超出必须稀释后测），并在报告中给出稀释校正的验证数据。
6. 全部原始称重记录（含时间戳）、OD 表、拟合脚本开源，第三方可重跑复现全部 μ 值（差异 $\leqslant$ 2%）。

5 · 数据与工具

用途	来源 / 工具
菌株	商品化活性干酵母（*S. cerevisiae*，单一批号，约 20 元）；BSL-1，食品级。若能获得实验室 BY4741 更佳，但**不作为前提**
培养基	葡萄糖、蔗糖、酵母浸粉、蛋白胨（YPD 组分，约 250 元）；蒸馏水由学校纯水机提供
产气监测	电子天平 0.01 g 量程 $\geqslant$ 500 g（约 150 元，若学校已有可省）；250 mL 三角瓶 + 水封发酵栓（约 100 元）
生长监测	学校现有分光光度计（600 nm）；比色皿若干
温控	恒温培养箱或恒温水浴 30 °C（学校现有）；温度记录仪或最高最低温度计
灭菌	高压灭菌锅（学校现有）；**若无，降级方案**：培养基煮沸 15 min × 3 次的间歇灭菌（tyndallization），须在方法中说明并用空白瓶做污染对照
分析	Python（numpy/scipy/statsmodels）或 R，纯 CPU
文献基线	乙醇抑制与渗透胁迫动力学的公开论文；**仅用于校验与对比，不计入本项目贡献**

预算合计约 350 – 600 元（天平已有则显著下降）。

6 · 方法路径

1. 建体系：配制 YPD、灭菌、活化种子液，跑通一条完整发酵曲线（称重每 2 h 一次、OD 每 4 h 一次，共 72 h），确认曲线三期清晰可辨。
2. 完成两条方法学校验（$NaHCO_3$ 定量释放标定 + 乙醇梯度单调性），并建立本实验室的无胁迫 μ 基线（连续三轮）。
3. 核实能力边界：确认分光光度计在高糖高浊度下的线性范围与必要稀释倍数；确认发酵栓在 35% 糖高粘度介质中不堵塞；确认水分蒸发本底随温度和瓶口面积的变化（这一步必须做在正式实验之前，不能边做边发现）。
4. 跑 3 × 3 × 2 全因子，每格 3 生物学 × 3 技术重复，随机化接种顺序与培养箱内瓶位（避免位置效应），瓶身编号盲法记录。
5. 提取 μ 与 $\max dCO_2/dt$，做三因素 ANOVA 与 Bliss 偏差分析。
6. 加做一条机制性旁证：在最强交互的处理组测发酵结束后残糖（用班氏试剂半定量或市售葡萄糖试纸）与终点 pH，判断抑制是「停止发酵」还是「发酵变慢」。
7. 独立交叉校验：用第二种独立方法（排水集气法量 CO_2 体积）在 4 个代表性处理组上复测速率，与失重法结果的相关系数须 $\geqslant$ 0.95。

7 · 新颖性边界

本课题**不**声称发现新的酵母胁迫响应机制，**不**声称改良菌株，**不**做任何遗传操作。甘油通道 Fps1p、HOG 通路等机制均为已发表知识，只能作为解释框架引用，不得写成本项目的发现。

已有工作完成了什么：乙醇对稳定期细胞发酵动力学的抑制、高糖 / NaCl 渗透胁迫下 CO_2 速率的即时下降与 2 h 恢复、渗透胁迫响应的碳源依赖性，均已在同行评议文献中定量报道。历届丘奖生物获奖论文中，2022 年 Milton Academy 的《丙酮酸脱氢酶突变酵母的乙醇耐受性与产量改良》与本课题最近，但它**做的是菌株改造后的耐受性提升**（依赖分子操作与测序），本课题不做任何改造，**做的是野生型上两种胁迫的交互结构**——研究对象与判据完全不同。2025 年复旦附属复兴中学的 PHB 生物合成课题属发酵工艺方向，终点是产物产率而非胁迫交互。

本项目的贡献（主结论）：给出一套只用天平与分光光度计即可复现的双通道发酵动力学测定规程，并在此规程下交付乙醇 × 渗透压交互项的符号、效应量与置信区间，以及该交互是否随碳源改变的直接检验。

价值：交互项是文献里被单因子设计系统性回避的量，而它恰恰决定了「两种胁迫叠加时能否用单因子结果外推」。**若交互项不显著，结论同样成立且有用**——它意味着 Bliss 独立模型在此参数区间可用；但必须用效应量置信区间宽度证明本设计有能力检出中等大小的交互。

8 · 决策门槛（go / no-go）

- **第 3 周末**：失重法标定回收率 $\geqslant 95\%$ 且 CV $\leqslant 3\%$。若达不到（多因瓶口蒸发或发酵栓漏气），**立即降级**：其一，改用排水集气法作为主通道（精度略降但绝对量可靠）；其二，把主终点从 CO_2 速率整体改为 OD_{600} 比生长速率，CO_2 降为旁证。两条路径都保留「交互项检验」这一主结论框架。
- **第 6 周末**：确认无胁迫 μ 基线的轮间 CV $\leqslant 15\%$。若 $> 15\%$，先排查温控与接种量标准化（须用 OD 标定接种密度而非按体积），排查后仍不达标则把设计从 $3\times3\times2$ 缩减为 3×3（单一碳源），把节省的重复数投入提高每格 n 到 5——**放弃 H2，保住 H1**。
- **第 12 周末**：确认高糖组（35%）在 72 h 内能进入可辨识的指数期。若严重迟滞导致无法提取 μ，把高渗档从 35% 下调至 30%，并在论文中说明参数区间的选择依据。
- **第 30 周末**：完成全部析因数据。若交互项 $p > 0.05$ 且效应量 CI 窄（上限 < 0.06），直接以「在该参数区间内两种胁迫近似 Bliss 独立」为主结论成文，并把重心转向双通道方法学一致性这一次级贡献。
- **需提前核实（第 2 周内）**：学校是否有可用高压灭菌锅（无则启用间歇灭菌降级方案，并加做空白污染对照）；分光光度计的线性上限；恒温箱在满载时的温度均匀性（不同层温差 $> 1.5\,^{\circ}$C 则必须随机化瓶位并把层作为区组因子）。
- **预算裁剪顺序**：先砍蔗糖臂（等于放弃 H2，H1 完整保留）→ 再砍第 6 步残糖测定（机制解释力下降，主结论不变）。**不可砍**生物学重复数与 $NaHCO_3$ 标定。

安全与伦理要点

- 仅使用食品级酿酒酵母，属 BSL-1，无致病性；不涉及任何脊椎动物或人体材料。
- 全程不得密封发酵容器（必须用水封栓或棉塞留出气路），防止 CO_2 累积致容器爆裂；高糖高醇培养基须在通风处操作。
- 乙醇为易燃品，配制与储存远离明火与电热设备，单次领用量 $\leqslant$ 500 mL 并登记。
- 发酵产物含乙醇，**严禁品尝或饮用**，实验结束后所有液体经高压灭菌（或煮沸 15 min）后排放。
- 高压灭菌锅须在教师在场时操作；若采用间歇灭菌降级方案，须以空白瓶做污染对照并在结果中报告污染率。

B13 · 城市绿地踩踏压实梯度上的土壤微生物呼吸：以碱吸收滴定法 CO_2 通量与可培养菌落数判定压实的功能阈值

`推荐优先级 P0` · `土壤微生物族` · `预算档 ≤ 500 元` · `周期档 中轮次（每轮 10 天，可跑 20+ 轮 + 四季重复）` · `技能取向 野外采样 + 滴定定量 + 阈值回归`

1 · 研究问题

在同一城市公园内部由人为踩踏形成的土壤压实梯度（步道中心 → 步道边缘 → 草坪 → 灌丛下）上，土壤基础呼吸速率（basal respiration，以 mg CO_2 · kg^{-1} 干土 · d^{-1} 计）随容重（bulk density）的下降是线性的，还是存在一个可定位的折点（breakpoint）阈值？若存在折点，其位置是否与文献中「限制根系生长的容重阈值」一致？

2 · 研究背景与空白

背景。土壤呼吸是微生物分解有机质的直接功能读数，可用极廉的碱吸收滴定法（alkali absorption / titration）测定：密闭容器内置一小杯已知浓度 NaOH 吸收土壤释放的 CO_2，反应后用 $BaCl_2$ 沉淀碳酸根，再用标准 HCl 滴定剩余碱量，差值即 CO_2 量。这套方法在 1950 年代即已标准化，需要的全部仪器就是天平、滴定管和密闭罐——中学实验室完全覆盖，而它测的是货真价实的生态系统功能指标。压实之所以重要，是因为它同时降低总孔隙度与通气孔隙比例，理论上应在某个容重处使氧扩散成为限制因子，从而让呼吸速率出现非线性下降。

已有工作到哪一步。米兰城市绿地的系统调查给出了具体数字：城市土壤容重均值 1.03 ± 0.13 g · cm^{-3}，整体低于限制根系生长的阈值，但容重与穿透阻力在类别间显著变化——城郊林地最低、中心公园与城市绿岛最高，反映植被、管理与游憩压力差异；同时中心公园 pH 偏中性至微碱、速效磷升高，显示更强的人为影响。另有研究指出城市土壤 pH 升高与微生物功能下降相关，高容重可能解释新建城市土壤中微生物丰度与活性的降低。

空白在：已发表工作大多在**类别之间**（林地 vs 公园 vs 绿岛）比较均值，很少在**单一场地内部**沿连续压实梯度密集布点、并明确检验呼吸 – 容重关系是线性还是分段线性（即是否存在功能阈值）。这个空白特别适合一年期学生课题：它把「城市土壤更差」这个类别命题改写成一个可拟合、可定位的参数问题，同时把气候、母质、植被类型这些最大的混杂因子通过「同一公园内取样」自动控制掉。

3 · 可检验假设

- **H1**：呼吸速率与容重的关系用分段线性（piecewise linear）模型拟合显著优于单一线性模型（AIC 差 $\geqslant 4$，或分段模型的 F 检验 $p < 0.05$），且折点落在 1.3 – 1.6 g · cm^{-3} 区间。
- **H2（机制备择）**：把充气孔隙度（air-filled porosity，由容重、含水量与假定颗粒密度 2.65 g · cm^{-3} 计算）作为自变量时，非线性消失、关系变为单一线性——若成立，说明压实对呼吸的抑制是通过通气受限而非机械阻力介导的。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：用已知质量的 Na_2CO_3 / $NaHCO_3$ 加过量酸在密闭罐中定量释放 CO_2，全流程走一遍碱吸收滴定，**回收率须落在 95 – 105%，重复 6 次 CV ⩽ 5%**；同时空白罐（无土无碳源）的滴定读数漂移须 ⩽ 满量程信号的 5%。此外用同一土样做一次分割重复（split-sample），两份读数相对偏差 ⩽ 10%。**此项不过关，后续全部呼吸数据无效**。
2. 布点写死：4 个压实带 × 每带 ⩾ 6 个独立采样点（**生物学重复 n ⩾ 6**），每个采样点的土样在实验室分为 2 个培养罐（**技术重复 n = 2**）。容重用环刀法原状取样，每点 3 次取均值。总样本量 ⩾ 24 点 × 2 罐。
3. 呼吸测定条件全部标准化后再比较：全部土样过 2 mm 筛、调至同一含水量（田间持水量的 60%）、25 ° C 预培养 7 d 消除扰动脉冲，再测 24 h（或 48 h）累积 CO_2。**未做预培养的数据不得进入主分析**。

4. 统计口径：呼吸速率对容重做线性 vs 分段线性模型比较，折点位置报**自助法 2000 次重抽样的 95% 置信区间**（若 CI 跨越整个容重范围，则判定折点不可分辨，H1 否证）；协变量（含水量、有机质、pH）用多元回归或偏相关控制。**报回归斜率与效应量 + 95% CI，不得只报 p 值。**
5. 平板计数为计数数据，用泊松或负二项 GLM 建模（过离散检验决定用哪个），**禁止对菌落数直接做 t 检验**；每样点 3 个稀释度 × 3 个平板，仅采用落在 30 – 300 CFU 区间的平板计数。
6. 季节重复：秋、冬、春三季各完整跑一轮，检验折点位置是否季节稳定（三季折点 CI 是否重叠）。
7. 全部滴定原始记录、容重计算表、拟合脚本开源，第三方可重跑复现折点位置（差异 $\leqslant 0.05\ g \cdot cm^{-3}$）。

5 · 数据与工具

用途	来源 / 工具
CO_2 吸收与滴定	NaOH、HCl 标准液、$BaCl_2$、酚酞（约 120 元）；学校现有滴定管、移液管、容量瓶
培养容器	广口密封玻璃罐 60 个 + 小烧杯（约 150 元）；恒温培养箱 25 °C（学校现有）
容重取样	环刀（100 cm^3）3 只 + 铝盒（约 120 元）；烘箱 105 °C（学校现有）；分析天平 0.001 g（学校现有）
微生物计数	营养琼脂或牛肉膏蛋白胨培养基 + 平皿（约 100 元）；高压灭菌锅；**若无超净工作台，降级方案**：酒精灯火焰圈内操作 + 每批设 3 个空白平板做污染对照，污染率 > 10% 则该批作废
现场协变量	pH 笔、EC 笔（与 B11 共用）；有机质用灼烧失重法（muffle furnace 若无，则用 550 °C 马弗炉替代方案：**须核实**学校是否具备，无则改用重铬酸钾外加热法或直接放弃该协变量）
分析	Python（scipy、statsmodels、piecewise-regression 包）或 R（segmented），纯 CPU
文献基准	米兰城市绿地容重与理化性质数据（Soil Use Manage 2026）；**仅用于对比与情境化，不计入本项目数据贡献**

预算合计约 400 – 500 元。

6 · 方法路径

1. 建立并跑通碱吸收滴定流程，完成 Na_2CO_3 定量释放的回收率校验与空白漂移测定（第 1 条验收标准）。
2. 选定一个面积足够、内部压实差异明显、且管理方同意采样的公园；沿步道横断面用穿透阻力（可用简易土壤硬度计或自制落锤贯入器，**须先做与容重的相关性验证，$r \geqslant 0.7$ 才可用作快速布点依据**）初筛，确定 4 个压实带的位置。
3. 采样：每点同步取环刀原状样（容重）、扰动样（呼吸与计数）、现场测含水量与 pH；记录 GPS、距步道距离、地表植被盖度（用手机拍照 + ImageJ 阈值分割定量，判据须写死）。
4. 实验室处理：过筛、调水分、25 ° C 预培养 7 d，再测 24 – 48 h 累积 CO_2 ；同批做稀释涂布平板计数（10^{-3} ~ 10^{-5} ），28 ° C 培养 48 h 计数。
5. 拟合线性与分段线性模型，做 AIC 比较与折点自助置信区间；再以充气孔隙度为自变量重跑一遍检验 H2。
6. 季节重复（秋 / 冬 / 春各一轮），检验折点稳定性。
7. 独立交叉校验：在最松与最紧两个带各取 3 个样点，改用**基质诱导呼吸**（substrate-induced respiration，加葡萄糖后测短期 CO_2 峰）测一次微生物活性生物量，与基础呼吸的排序须一致。

7 · 新颖性边界

本课题**不**声称提出新的呼吸测定方法（碱吸收滴定是 70 年前的标准法），**不**做任何微生物鉴定或群落组成分析（无 PCR、无测序），**不**声称压实与呼吸之间的因果关系——观测设计只能给出关联，因果表述须明确标注为推测。可培养菌落数只代表可培养部分（通常 < 1% 的总群落），这一局限必须在论文中显式声明，不得把它当作「微生物丰度」。

已有工作完成了什么：米兰城市绿地研究给出了容重（1.03 ± 0.13 g · cm^{-3}）、穿透阻力、pH、速效磷在城郊林地 / 中心公园 / 城市绿岛间的差异排序；城市土壤 pH 升高与微生物功能下降、高容重与微生物活性降低的关联也已报道。历届丘奖生物获奖论文中，2021 年北京德威《小麦-大豆轮作农业生态系统土壤细菌与古菌群落的垂直与时间变异》与本课题同属土壤微生物，但它依赖高通量测序做群落组成，本课题**完全不依赖测序，做的是功能速率与物理参数之间的响应形状**；2024 年北京王府学校的蚯蚓-微塑料课题终点是降解率，与本课题的阈值定位问题不同。

本项目的贡献（主结论）：在单一场地内控制掉气候、母质与植被类型后，给出土壤呼吸对压实响应的**函数形状**（线性 vs 分段）与折点位置的置信区间，并通过「换自变量为充气孔隙度」的对照，检验非线性是否由通气机制介导。

价值：类别均值比较无法回答「压到多紧才出问题」，而这正是城市绿地管理最需要的量。**若折点不可分辨（CI 过宽），同样是有效结论**——它意味着在该场地的容重范围内呼吸响应近似线性，此时须报告本设计能分辨的最小折点效应，证明结论不是样本量不足的假象。

8 · 决策门槛（go / no-go）

- **第 4 周末**：滴定回收率必须落在 95 – 105%。若连续三轮不达标，**立即降级**：其一，改用市售 CO_2 检测管或土壤呼吸室 + 廉价 NDIR CO_2 传感器模块（约 150 元，须自行标定，成为新的方法学校验对象）；其二，把主终点改为**基质诱导呼吸的相对值**（组间比值受系统偏差影响小）。两条路径都保留「响应形状与折点」主结论框架。
- **第 8 周末**：确认所选公园内容重的实测跨度 ⩾ 0.4 g · cm^{-3}。若跨度不足（例如全部落在 1.0 – 1.2），单一场地无法定位折点，**降级为多场地设计**：增补 2 – 3 个管理强度不同的绿地（施工便道、球场边缘、封闭养护区），代价是引入场地混杂因子，须在分析中加入场地随机效应——分析框架与主结论保留。
- **第 14 周末**：完成秋季完整一轮（24 点 × 2 罐）。若单点呼吸测定的技术重复相对偏差 > 15%，先排查密封性与滴定终点判读（改用 pH 计辅助判定终点），排查后仍不达标则把每点技术重复提高到 3 并相应减少采样点数至 5 点/带。
- **第 26 周末（2026 年 12 月末）**：完成冬季轮。冬季土温低、呼吸信号弱是已知困难——**把它写成研究内容而非障碍**：季节间呼吸绝对值差异本身就是数据，只需在低信号季延长培养至 72 h 保证滴定差值 > 空白漂移的 5 倍。
- **需提前核实（第 2 周内）**：公园管理方是否书面同意环刀取样（会留小坑，须承诺回填）；学校是否有 105 ° C 烘箱与 0.001 g 天平；是否有高压灭菌锅与可用的洁净操作条件（无超净台时的降级方案与污染对照必须在正式实验前验证通过）。
- **预算裁剪顺序**：先砍平板计数臂（呼吸 – 容重主结论完整保留，仅失去「功能变化是否伴随可培养菌数变化」的旁证）→ 再砍有机质测定（须在讨论中承认无法排除有机质混杂）。**不可砍**环刀（容重是核心自变量）与滴定标定试剂。

安全与伦理要点

- 微生物培养仅做土壤异养菌总数的稀释涂布计数，属环境样本常规操作；**不做任何致病菌的分离、富集或鉴定**，不使用选择性致病菌培养基，不使用抗生素平板。所有平板 **不得开盖观察**，计数隔盖进行，用后一律高压灭菌（或 84 消毒液浸泡 24 h）后按医疗/实验废弃物流程处理。

- 无超净工作台时采用酒精灯火焰圈操作，须由教师在场监督；酒精灯与酒精棉远离易燃物。
- NaOH、HCl、$BaCl_2$ 均有腐蚀性/毒性（$BaCl_2$ 为有毒品），须戴手套护目镜，含钡废液**单独收集**交学校统一处置，严禁排入下水道。
- 野外采样须取得公园或绿地管理方书面同意，取样坑当场回填复原；不在保护区、古树名木保护范围内取样；不采集任何动植物活体。
- 全程不涉及脊椎动物。

B14 · 果蝇负趋地性爬升速度的年龄衰退曲线：以高糖饮食与温度双因子判定运动衰老速率是否被饮食加速

`推荐优先级 P1` · `模式生物实验族` · `预算档 ≤ 700 元` · `周期档 世代 10–14 天，全年可完成 6–8 个独立队列（cohort）` · `技能取向 活体维持 + 视频行为定量 + 生存/衰退曲线拟合`

1 · 研究问题

在高糖饮食（high-sugar diet）下饲养的黑腹果蝇（Drosophila melanogaster），其负趋地性（negative geotaxis）爬升速度随年龄的衰退**速率**（即衰退曲线的斜率），是否显著大于标准饮食组？该饮食效应在 25 ° C 与 29 ° C 两个饲养温度下是否量级一致，还是被高温放大？

2 · 研究背景与空白

背景。负趋地性是果蝇受惊后向上攀爬的先天行为，被广泛用作运动功能的定量读数。关键的一条已发表结论是：负趋地性的衰老**主要源于爬升速度（climbing speed）随年龄下降**，而非「爬没爬上去」这个二值指标。这直接决定了实验设计——传统的「N 秒内越线百分比」把连续量压成二值，损失大量信息且天花板效应严重；用手机高帧率录像 + 逐帧位置追踪提取速度曲线，可以在同样的果蝇数量下得到高得多的统计效力。这恰好是低成本自动化能带来真实提升的场景：一部手机（120 fps 慢动作）+ ImageJ/Fiji 的 TrackMate 插件或 DeepLabCut CPU 版即可。

已有工作到哪一步。高糖饮食对果蝇运动的影响文献结论并不一致：一方面有研究报告适度提高糖浓度可提升繁殖力、寿命与运动表现；另一方面 2025 年的工作在负趋地性与探索行为两种范式中都观察到**性别特异、剂量依赖**的运动缺陷（尤其在三氯蔗糖暴露下），并有工作报告高糖饮食降低多巴胺水平从而损害运动与探索活动。方法学上，「连续追踪受惊果蝇」已被明确提出作为负趋地性爬升测定的替代方案。

空白在：文献大多报告**某个年龄点上的组间差异**，而把饮食效应写成「衰退曲线斜率的差异」——即饮食是改变了运动能力的**起点**还是改变了**衰老速率**——的定量分离很少，且几乎没有把温度作为第二因子来检验该效应是否被代谢速率放大。这个空白适合中学课题：果蝇世代仅 10 – 14 天，一年可跑 6 – 8 个独立队列，样本量足以支撑斜率比较；且「饮食只降低起点、不改变斜率」与「饮食加速衰退」两种结果都是干净有力的结论。

3 · 可检验假设

- **H1**：高糖组的爬升速度 – 年龄回归斜率（mm · s^{-1} · d^{-1}）的绝对值显著大于标准组，两组斜率差的 95% 置信区间不含 0，且相对差异 ⩾ 25%。
- **H2（温度交互备择）**：29 ° C 下的饮食效应量（斜率差）显著大于 25 ° C（交互项 $p < 0.05$ 且偏 $\eta^2 \geqslant 0.06$）。若成立，支持「饮食效应由代谢速率放大」的解释；若不成立而两温度下饮食效应一致，则支持效应独立于总体代谢速率。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：在标准饮食、25 ° C 条件下重建一条年龄 – 爬升速度衰退曲线，其**衰退趋势方向与显著性必须与已发表的年龄依赖性下降一致**，且第 5 日龄的平均爬升速度与文献报道的野生型基线区间偏差 ⩽ 25%（文献基线须在开题时锁定并写入方案，不得事后挑选）。同时做追踪精度校验：用已知长度标尺与匀速移动的物体（如恒速下落小球或秒表配合手动标记）验证像素 – 毫米换算与速度提取误差 ⩽ 5%。**此项不过关，后续全部数据无效**。
2. 队列规模写死：每处理组（2 饮食 × 2 温度 = 4 组）每个队列 ⩾ 60 只成虫，分装为 ⩾ 6 个测试管（每管 10 只）；**生物学重复 = 独立建立的队列（不同批次亲本、不同日期收集处女蝇），n ⩾ 3；技术重复 = 同一队列同**

一日的重复轻敲试验，n = 3（每次间隔 ⩾ 60 s 以消除疲劳）。性别分开饲养与分析（文献已报性别特异效应，混合分析无效）。

3. 测定时间点写死：第 5、10、15、20、25、30 日龄各测一次，全部在每日同一时段（±1 h）进行以控制昼夜节律；测定前 30 min 将管移入测定室适应室温。
4. 统计口径：以「测试管」为随机效应、日龄为连续自变量的线性混合模型（linear mixed model）估计各组斜率，检验 饮食 × 日龄 与 饮食 × 温度 × 日龄 交互项；**报斜率估计值 + 95% CI 与偏 η^2，不得只报 p 值**。存活率作为协变量与竞争解释（若高糖组死亡更快，则存活者选择偏倚必须用 Kaplan – Meier 生存曲线并行报告并在讨论中处理）。多组比较用 Tukey HSD 并检查残差正态性与方差齐性。
5. 视频判读的可复现性：爬升速度的提取判据写死（起跑线、计时起点为轻敲结束帧、只取前 4 s 的位移）；随机抽取 20% 的视频由第二名成员独立重跑追踪，两次结果的 ICC ⩾ 0.90；若采用人工判读辅助（如剔除跌落个体），须做双人一致性检验，Cohen's κ ⩾ 0.80。
6. 全部视频、追踪配置文件与统计脚本开源，第三方可重跑复现全部斜率（差异 ⩽ 5%）。

5 · 数据与工具

用途	来源 / 工具
虫源	野生型（Canton-S 或 Oregon-R）果蝇，来源须**第 1 周内落实**：高校实验室赠予、生物教学器材商（部分供应野生型教学品系）或校内自捕野生种群（后者须承认遗传背景不纯并作为局限说明）。**这是本题最大的外部依赖，列入 go/no-go 风险项**
培养基	玉米粉–琼脂–酵母–蔗糖标准培养基组分 + 丙酸/尼泊金酯防霉（约 200 元）；高糖组用提高蔗糖至标准 4–5 倍配制
饲养	果蝇培养管 200 支 + 海绵塞（约 150 元）；两台恒温培养箱或一台带隔层控温（25 ˚C / 29 ˚C，**须核实学校是否有两个独立温区**，无则改为顺序进行两个温度并加入批次因子）
麻醉与转移	低温麻醉（冰盘）优先，避免乙醚/CO_2；若必须用 CO_2，须核实气源与安全操作
行为录像	手机 120 fps 慢动作 + 固定支架 + 均匀 LED 背光 + 标尺（约 150 元）
追踪与分析	ImageJ/Fiji + TrackMate（免费，CPU）；或 DeepLabCut CPU 版（**需核实**：CPU 训练一个小模型的时长可达数小时至一天，须在第 8 周前实测，超出承受则固定用 TrackMate 阈值追踪）
统计	R（lme4、emmeans、survival）或 Python（statsmodels、lifelines），纯 CPU
文献基线	已发表的年龄依赖爬升速度衰退数据与高糖饮食运动效应论文；**仅用于校验与对比，不计入本项目数据贡献**

预算合计约 500 – 700 元。

6 · 方法路径

1. 落实虫源并建立稳定传代（第 1 – 4 周）：连续 2 代确认培养基无霉变、羽化数稳定、无螨虫污染。
2. 搭建录像与追踪管线，完成标尺与匀速物体的速度提取误差校验，并复现标准饮食 25 ˚ C 的年龄衰退曲线（第 1 条硬门槛）。
3. 核实工具能力边界：实测 TrackMate 在多只果蝇同管场景下的丢轨率与身份互换率；若 > 10%，改为**单只果蝇单管测定**（数据质量升、通量降，须相应把每组管数提高）或降级为「N 秒越线高度」的半定量指标（此时 H1 的

阈值须从速度斜率改写为高度斜率）。

4. 建立 4 组处理队列：同批亲本产卵、幼虫密度标准化（每管定量投卵，避免密度效应混杂）、羽化后 0 – 24 h 内按性别分装。
5. 按写死的时间点测定，同步记录每日死亡数用于生存分析；每次测定后随机化管序并盲法编号。
6. 拟合线性混合模型，估计各组斜率与交互项；并行做 Kaplan – Meier 生存分析排除选择偏倚。
7. 独立交叉校验：在第 15 日龄用第二种独立行为范式（水平开放场地的自发活动距离，同样用视频追踪）复测 4 个组，与爬升速度给出的组间排序须一致；不一致则作为「运动衰退的范式特异性」写入讨论，本身即为发现。

7 · 新颖性边界

本课题**不**做任何遗传操作、不做分子测量（无 qPCR、无测序、无多巴胺定量），因此**不得声称**揭示了机制；多巴胺、胰岛素信号等解释只能作为引用的文献假说出现，并明确标注为推测。**不**声称提出新的行为范式——连续追踪替代传统越线计数已被明确提出。

已有工作完成了什么：负趋地性衰老主要由爬升速度的年龄依赖下降驱动，这一定量结论已发表；高糖饮食的运动效应文献结论不一致（既有提升繁殖力/寿命/运动表现的报告，也有 2025 年报告的性别特异、剂量依赖运动缺陷，以及高糖降低多巴胺进而损害运动的报告）；连续追踪法作为替代范式已有专文。历届丘奖生物获奖论文中，2021 年北京十一学校《亚致死农药选择压力下豆蚜的翅发育与性别分化》同属昆虫实验生态，但对象、处理与终点完全不同；2024 年成都外国语学校的涡虫再生课题与 2025 年多篇神经行为课题均依赖分子手段，本课题**刻意设计为零分子依赖**。

本项目的贡献（主结论）：把「高糖饮食是否影响果蝇运动」这个结论互相矛盾的问题，改写为可分离的两个参数——**截距（能力起点）与斜率（衰老速率）**——并在两个温度下同时估计，给出带置信区间的斜率差与温度交互项。文献分歧的一个可能来源正是不同研究测了不同年龄点，本课题的设计直接针对这一点。

价值：若饮食只降低截距而不改变斜率，说明它损害的是发育期建立的运动能力而非衰老过程本身，这与「高糖加速衰老」的流行叙述相悖，是一个有分量的负结论。**两种结果都成立**，但必须用斜率差的置信区间宽度证明本设计有能力检出 25% 的斜率差异（须在开题时做效能计算并写入方案）。

8 · 决策门槛（go / no-go）

- **第 4 周末（最关键的一道门）**：必须已获得可稳定传代的果蝇品系并完成 2 代传代。若未落实，**立即降级**：其一，改用同样易得且世代更短的模式材料——面包虫/黄粉虫（Tenebrio molitor）幼虫的负趋性或爬行速度（世代长，但幼虫可直接购买且量大），或家蚕/赤拟谷盗；其二，改用**自捕野生果蝇种群**（用水果诱捕），承认遗传背景不纯，并把研究问题相应改为「同一野生种群内饮食处理的效应」——分析框架（截距 vs 斜率分离、双温度）与主结论完全保留。
- **第 8 周末**：追踪管线的丢轨率与速度提取误差必须达标。若 DeepLabCut CPU 训练超过 24 h 或精度不足，直接固定用 TrackMate；若 TrackMate 多虫丢轨 > 10%，切单虫单管方案。**不允许把这个问题拖到数据采集期再解决**。
- **第 12 周末**：完成第一个完整队列（4 组 × 6 个时间点）。若 30 日龄前某组存活率 < 30%，缩短观测窗至第 5 – 20 日龄并相应调整效能计算；若高糖组死亡显著更快，把生存分析从协变量升级为并列的第二个主结果。
- **第 20 周末**：确认两个温度可同时或顺序完成。若学校只有一个温区且顺序执行导致批次与温度完全混杂，**放弃 H2**，把节省的资源投入把 H1 的生物学重复提高到 n = 5——主结论（斜率差）完整保留。
- **第 36 周末**：完成 ⩾ 3 个独立队列。若斜率差的 95% CI 包含 0 且宽度 < 25%（即有能力检出却未检出），以「高糖饮食不改变运动衰老速率」为主结论成文，并把截距差异作为次级结果报告。

- **需提前核实（第 2 周内）**：虫源渠道（最高优先级）；培养箱温区数量；手机慢动作模式的实际帧率与是否有自动曝光漂移；学校是否允许饲养活体昆虫及逃逸管理措施。
- **预算裁剪顺序**：先砍第 7 步开放场地交叉校验（主结论保留，范式特异性无法排除，须在讨论中说明）→ 再砍 29 °C 温度臂（等于放弃 H2）。**不可砍**队列数与每组虫数。

安全与伦理要点

- 果蝇为无脊椎动物，**不涉及脊椎动物有创实验**；行为测定为非侵入式（轻敲惊扰属标准范式，无损伤），全程不做手术、注射或致痛处理。
- 麻醉优先使用低温冰盘，避免乙醚（易燃、麻醉过量致死率高）；若使用 CO_2 ，须在教师监督下操作并控制暴露时间。
- **逃逸管理**：所有操作在带纱窗的封闭房间内进行；培养管一律用海绵塞封口；实验结束的果蝇一律冷冻致死（−20 °C ⩾ 24 h）后作为一般生物废弃物处理，**严禁活体释放到环境中**。若使用野捕种群，同样须全部灭活，不得跨地区转移虫源。
- 培养基含丙酸 / 尼泊金酯，配制时戴手套并在通风处操作；霉变培养管立即高温灭活弃置，不得敞开放置。
- 若课题涉及自捕野生果蝇，诱捕装置须置于校内或经许可的场地，不在自然保护区内布设，不误捕保护物种（诱捕器口径与诱饵仅吸引果蝇类）。

B15 · 康普茶菌膜细菌纤维素的产率标度律：以「单位气液界面面积产率」判定产量受界面控制还是受体积控制

`推荐优先级 P1` · `发酵与应用微生物族` · `预算档 ≤ 600 元` · `周期档 中轮次（14 天/轮，可跑 15–20 轮）` · `技能取向 发酵操作 + 几何设计 + 标度关系拟合`

1 · 研究问题

康普茶共生菌群（kombucha SCOBY）在静置培养中形成的细菌纤维素（bacterial cellulose, BC）菌膜，其干重产率随培养容器几何形状的变化，是与气液界面面积成正比（界面控制），还是与培养液体积成正比（体积控制）？在固定界面面积后，碳源浓度与氮源浓度对**单位面积产率**的影响是否仍与文献报道的体积产率结论一致？

2 · 研究背景与空白

背景。BC 由醋酸菌属（Komagataeibacter 等）在静置培养的**气液界面**分泌形成，因为该菌为专性好氧菌，只有界面附近氧供应充足。这给出了一个干净、可否证的物理预测：在营养不限的前提下，产率应正比于界面面积 A 而非体积 V。文献中几乎所有产率都以 g/L（每升培养液）报告，这个单位内建了「体积控制」的假设，导致不同容器几何下的数据无法互相比较——而中学实验室恰恰可以用最廉价的方式直接检验这个假设：用不同直径的容器装同样体积、或同样直径装不同深度，做一个二维几何扫描。

已有工作到哪一步。响应面法优化已被系统应用：中心复合设计考察起始茶液体积（18 – 54 mL）、葡萄糖量（18 – 36 g）与温度（25 – 35 °C），联合优化得到 70 g/L SCOBY 与 2.94 g/L BC（66.27 mL 起始茶液、22.37 g 葡萄糖、38.4 °C）；优化后 BC 抗拉强度 4 MPa、断裂伸长率 3.8%。另有工作用糖蜜（75 g/L）替代碳源、茶末（10 g/L）替代氮源，在 pH 5、30 °C 下得到 2.56% 的最大 SCOBY 产率；也有工作报告绿茶 + 棕榈糖组合给出 0.194 mm 厚度与最高抗拉强度。

空白在：已发表优化研究全部在**固定容器几何**下报告体积产率（g/L），没有把几何本身作为自变量做系统扫描，因而无法判断这些优化结论是否只是「换了个容器就失效」的表观参数。这个空白适合中学课题：几何扫描的成本几乎为零（不同直径的玻璃罐十几元一个），单轮 14 天，一年可跑 15 – 20 轮；而「产率是界面控制」与「产率是体积控制」两种结论都干净、都有用。

3 · 可检验假设

- **H1**：BC 干重产率 m 与界面面积 A 和体积 V 的关系用幂律 $m \propto A^{\alpha} \cdot V^{\beta}$ 拟合时，α 的 95% 自助置信区间包含 1 且不含 0，而 β 的置信区间包含 0——即产率由界面面积单独决定。
- **H2（营养耦合备择）**：当把液层深度增至 A/V 极低（深液层）时，α 显著偏离 1，说明在深液层下营养/代谢产物扩散成为共同限制因子；此时单位面积产率随液深单调下降，且下降的拐点可定位。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：(a) 干重测定标定——BC 菌膜的纯化流程（0.1 M NaOH 90 °C 处理 30 min 去菌体、去离子水洗至中性、60 °C 恒重干燥）须给出**恒重判据**（连续两次称量差 ≤ 0.5%），并用同一批菌膜做 6 份分割重复，干重 CV ≤ 8%；(b) 复现文献产率区间——在文献报道的标准条件（约 30 °C、蔗糖 ~20 – 75 g/L、红茶）下，自测体积产率须落在已发表区间（约 1 – 3 g/L · 14 d）内，偏差 ≤ 50%（BC 产率的文献间离散度本就很大，故阈值放宽，但必须事前锁定所对标的具体文献值）。**任一条不过关，后续全部产率数据无效**。
2. 几何扫描写死：≥ 5 种「面积 × 体积」组合（同体积不同直径 3 档 + 同直径不同深度 3 档，共享中心点），**生物学重复 = 独立接种批次（不同日期、独立分割的母菌膜），n ≥ 3；技术重复 = 同批次内同规格平行罐，n = 3**。接种量须按**界面面积**与**体积**两种口径各做一次标准化对照，以免接种量本身混入几何效应。

3. 统计口径：对 lg(m) 对 lg(A)、lg(V) 做多元线性回归得到 α、β，报**自助法 2000 次重抽样的 95% 置信区间**；同时给出普通最小二乘之外的稳健回归（Huber 或 Theil – Sen）交叉验证，两者指数差异 > 0.2 须在论文中讨论。多组比较用 ANOVA + Tukey 并检查前提；**报效应量 + CI，不得只报 p 值**。
4. 力学表征（可选但推荐）：用自制悬挂砝码装置测干膜的抗拉强度，须报告样条尺寸、加载速率与至少 5 条样条；结果与文献值（约 4 MPa 量级）作数量级对照即可，**不得声称精确测量**。
5. 每轮设 ⩾ 2 个未接种空白罐做污染对照；出现霉菌污染（肉眼可见绒毛状菌落）的罐当轮作废并记录污染率，污染率 > 20% 须排查流程后重跑。
6. 全部称重记录、几何参数表、拟合脚本开源，第三方可重跑复现 α、β（差异 ⩽ 0.05）。

5 · 数据与工具

用途	来源 / 工具
菌种	市售康普茶 SCOBY 母菌膜（约 60 元）或未杀菌的商品康普茶自行培养起始；单一来源、全年不换
培养基	红茶/绿茶、白砂糖、食醋（调初始 pH 至 4.5）（约 100 元）
容器几何扫描	不同直径的直筒玻璃罐 5 规格 × 各 9 只（约 200 元）；直径与液深用游标卡尺/直尺测量，界面面积按 πr^2 计算并**实测校核**（罐口非正圆时以照片 + ImageJ 面积分割为准）
纯化与干燥	NaOH、pH 试纸（约 50 元）；60 °C 烘箱（学校现有）；分析天平 0.001 g（学校现有）
温控	恒温培养箱或恒温房 28–30 °C（学校现有）；温度记录仪
力学测试	自制悬挂加载装置（滑轮 + 砝码 + 夹具，约 80 元）+ 手机录像判断断裂点
分析	Python（numpy、scipy、statsmodels）或 R，纯 CPU
文献基准	BC 响应面优化论文（Cellulose 2025）、糖蜜/茶末替代碳氮源论文（Int J Polym Sci 2024）；**仅用于校验与对比，不计入本项目数据贡献**

预算合计约 400 – 600 元。

6 · 方法路径

1. 建立稳定的 SCOBY 传代与标准培养流程，连续 2 轮确认菌膜形成时间、厚度与污染率可控。
2. 完成两条方法学校验（干重恒重判据 + 文献产率区间复现），并锁定所对标的文献数值。
3. 核实边界：确认罐口面积测量误差 ⩽ 3%；确认蒸发导致的液面下降在 14 d 内是否显著改变界面面积（若 > 5%，必须加盖透气膜并把蒸发量作为协变量记录）——**这一步必须在正式扫描前完成**。
4. 跑几何扫描（5 组合 × 3 生物学 × 3 技术），随机化罐位与接种顺序，编号盲法称重。
5. 拟合 $\lg m \sim \lg A + \lg V$，得到 α、β 与置信区间；再做深液层加密扫描检验 H2 的拐点。
6. 在固定为「最优几何」的条件下，做一个小型碳源 × 氮源二因子实验（各 3 水平），把结果同时以 g/L 和 g/m^2 两种口径报告，直接展示结论是否随口径改变——这是本课题最有说服力的一张图。
7. 独立交叉校验：换一个独立的 SCOBY 来源（另一品牌/另一家庭培养株）在 3 个代表性几何上重跑，检验 α 是否菌源依赖。

7 · 新颖性边界

本课题**不**声称提出新的 BC 生产工艺，**不**做菌群组成鉴定（无测序），**不**声称开发新材料或新产品，**不**做任何食品安全性或功效声称。抗拉强度测定为自制装置的数量级对照，**不得**与标准拉力机数据直接并列比较。

已有工作完成了什么：响应面优化（起始茶液体积、葡萄糖、温度）已给出优化点与 2.94 g/L BC 产率、4 MPa 抗拉强度、3.8% 伸长率；糖蜜 + 茶末的碳氮源替代已给出 2.56% SCOBY 产率的优化条件；绿茶 + 棕榈糖给出的厚度与强度数据也已发表。**历届丘奖生物获奖论文中有两篇直接相关**：2021 年 Carmel Pak U Secondary School《Kombuchas from tannin-rich fruit skins as bio-disposables》与 2023 年同校《Antimicrobial Edible Bio-disposables of Kombucha of Fruit Skins with Chitosan Coating》。这两篇的定位是**用果皮基质做康普茶菌膜、制成可降解/可食用餐具并加壳聚糖赋予抗菌性**——即「换基质 + 做产品 + 加功能」的应用路线。本课题**刻意不走应用路线**，做的是「产率对容器几何的标度关系」这一方法学/物理定标问题，判断依据是幂指数 α、β 的置信区间，与上述两篇的对象、终点、判据均不重叠。这一点必须在论文引言中主动点明，不能等评委问。

本项目的贡献（主结论）：给出 BC 产率对界面面积与体积的幂律指数及其置信区间，并据此指出以 g/L 报告 BC 产率在跨几何比较时的系统性偏误；进而用同一批数据展示碳氮源优化结论在 g/L 与 g/m^2 两种口径下是否一致。

价值：这是一条**可复现性贡献**——物理机制（界面好氧）早已知，本课题提供的是量化的标度指数与一条可操作的报告口径建议。这类贡献是被承认的，但定位必须讲清楚，否则会被误读为重复已有的优化工作。**若 β 显著不为 0（体积也有贡献），同样是有效结论**，它意味着营养限制在常用条件下已经在起作用，需要用置信区间宽度证明本设计能分辨 0.2 量级的指数差异。

8 · 决策门槛（go / no-go）

- **第 5 周末**：SCOBY 传代稳定且污染率 ⩽ 20%。若污染反复（多为霉菌），**立即降级**：其一，提高初始酸度（醋添加量）并改用一次性透气封口膜；其二，把研究体系整体换为**醋酸菌纯培养的 BC 生产**（若能获得菌株）或换为「静置乳酸发酵的酸凝乳膜厚度」这一同样受界面/几何影响的体系。降级后「产率 – 几何标度」的主结论框架不变。
- **第 8 周末**：干重 CV 必须 ⩽ 8%。若纯化流程导致菌膜破碎丢失，改用**湿重 + 独立测定含水率换算干重**的双通道法，并把两法一致性作为方法学结果报告。
- **第 16 周末**：完成首轮几何扫描。若 α 与 β 的置信区间都很宽（宽度 > 0.6），说明信噪比不足，**降级**：把几何跨度加大（直径比从 2 倍拉到 4 倍）、生物学重复提到 n = 5，并放弃第 6 步的碳氮源二因子实验以腾出资源。主结论保留。
- **第 28 周末**：完成碳氮源 × 双口径对照实验。若两种口径给出相同排序，把这一点作为「优化结论在本参数区间对几何稳健」的结论报告——**这是负结论但同样有价值**。
- **需提前核实（第 2 周内）**：SCOBY 来源的稳定性与是否可长期传代；恒温箱在 28 – 30 ° C 的均匀性；学校是否允许在实验室长期放置发酵液（气味与虫害管理）。
- **预算裁剪顺序**：先砍力学表征（第 4 条验收标准，主结论完全不受影响）→ 再砍第 7 步第二菌源交叉校验（菌源依赖性无法排除，须在讨论中明写）。**不可砍**几何规格数量与生物学重复数。

安全与伦理要点

- 康普茶 SCOBY 为食品级共生菌群，属 BSL-1；**不做任何致病菌培养、不做抗菌活性的致病菌挑战实验**。若要做抑菌旁证，指示菌只能用 BSL-1 的 *E. coli* K-12 或枯草芽孢杆菌，且须教师批准。
- **发酵产物一律不得饮用或品尝**，无论气味、外观如何；实验完成的发酵液煮沸 15 min 或高压灭菌后排放。
- 出现霉菌污染（绒毛状、彩色菌落）的罐**不得开盖嗅闻**，直接整罐煮沸灭活后弃置——部分霉菌可产生毒素。

- NaOH 纯化步骤在 90 °C 热碱条件下进行，须戴耐热手套与护目镜，在通风处操作，教师在场；碱废液中和后排放。
- 长期发酵罐须密闭防果蝇滋生，放置于通风且远离食品的独立区域。

B16 · 盆栽植物渐进干旱的 RGB 图像早期预警：以绿度指数拐点相对目视萎蔫与相对含水量阈值的提前天数判定预警能力

`推荐优先级 P1` · `植物生理与逆境族` · `预算档 ≤ 800 元` · `周期档 长周期（每轮 6–9 周，全年 4–5 轮）` · `技能取向 图像分析 + 生理测定 + 时序拐点检测`

1 · 研究问题

用固定光照条件下的手机 RGB 图像计算的绿度指数（如超绿指数 ExG、绿色色度坐标 GCC）在渐进干旱过程中出现统计显著拐点的时刻，比目视萎蔫首次出现早多少天？该提前量在三个形态差异明显的物种间是否一致，还是随叶片形态（肉质 / 薄叶 / 禾本科）系统变化？

2 · 研究背景与空白

背景。水分亏缺会引起叶片衰老与色素变化，使红光波段反射相对于绿蓝波段上升，因此普通 RGB 相机的色度坐标就能承载水分状态信息——这一点已在综述中明确。真正的价值不在于「能否检测」，而在于**能否比人眼更早**：目视萎蔫是一个已经发生了显著生理损伤的滞后指标，而灌溉决策需要提前量。本课题把这个提前量本身定义为交付物，是一个天然量化、天然可否证的目标。所需设备只有手机、固定支架、恒定人工光源和一台天平（用称重法精确控制土壤含水量），全部在中学实验室范围内。

已有工作到哪一步。RGB 指数与反射率指数的对应关系已被系统研究：小麦与豌豆在土壤干旱与盐渍下，部分 RGB 指数（如 ExG）的变化**在干旱作用早期即已启动**，且 RGB 指数可用于估计 NDVI、PRI 与 Fv/Fm。另一项工作用 ImageJ 测量日本落叶松幼苗的顶端针叶角度，发现该形态指标**第 2 天**即对干旱产生显著响应，而叶绿素荧光与针叶温度要到**第 6 天**才响应——即已有工作明确给出过「提前 4 天」量级的对照。城市灌木的 RGB 成像与灌溉管理研究也已给出水分胁迫阈值。

空白在：已发表的提前量结论多为单物种、单指数，且对照的是**仪器指标**（叶绿素荧光、冠层温度）而非**目视萎蔫**这一实际管理判据；跨形态类群比较提前量是否稳定的系统研究很少。这个空白适合中学课题：不需要任何高端仪器，靠称重法可以把干旱进程控制得极精确；而「提前量在物种间一致」与「提前量随形态变化」两种结果都构成有效结论。

3 · 可检验假设

- **H1**：ExG 时序的拐点显著早于目视萎蔫首次判定日，提前量 $\geqslant 2$ 天，且三个物种的提前量均值的 95% 置信区间下限 > 0。
- **H2（形态依赖备择）**：三物种的提前量存在显著差异（单因素 ANOVA $p < 0.05$，偏 $\eta^2 \geqslant 0.14$），且薄叶物种的提前量最短——若成立，说明 RGB 早期预警的可用性依赖叶片形态，不能跨类群外推。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：(a) 成像系统标定——每次拍摄同框放入标准灰卡与彩色标准色块，做白平衡/曝光归一化；对同一株植物在同一天连续拍 10 次，归一化后 ExG 的 CV $\leqslant$ 2%；连续 3 天在无处理对照株上的 ExG 漂移 $\leqslant$ 处理组预期变化幅度的 10%。(b) 生理对照复现——用称重法建立土壤相对含水量与叶片相对含水量（relative water content, RWC）的关系，并确认充分供水对照的 RWC $\geqslant$ 85%（文献公认的非胁迫区间），干旱终点 RWC $\leqslant$ 60%。**任一条不过关，全部图像数据无效。**
2. 设计写死：3 物种 × 2 处理（充分供水 / 渐进断水）× **生物学重复 = 独立植株 n $\geqslant$ 8**（每株为一个独立单位，同盆多株不得视为重复）；每株每日固定时刻（±30 min）拍 **3 张技术重复照片**。全部植株随机化摆位并每 3 天轮换位置（消除光照梯度）。

3. 拐点检测方法预先写死：对每株的 ExG 时序做分段线性回归（segmented regression）或 CUSUM 变点检测，拐点日期报**自助法 95% 置信区间**；同一时序须用两种方法各算一次，两法给出的拐点相差 > 2 天则该株数据标记并单独讨论，不得只报较有利的一种。
4. 目视萎蔫判据必须写死并盲法判读：给出明确判据（例如「顶端 3 片完全展开叶中 ⩾ 2 片的叶缘或叶尖出现可见下垂/卷曲」），照片随机编号后由两名判读者独立判定，**Cohen's $\kappa \geqslant 0.80$**；不达标须细化判据并重新培训后重判。
5. 统计口径：提前量（天）为连续量，用单因素/双因素 ANOVA + Tukey 检验物种差异，检查方差齐性与正态性；**报效应量（偏 η^2、Cohen's d）+ 95% CI，不得只报 p 值**。RWC 与 ExG 的关系用回归并报自助置信区间。
6. 破坏性测定（RWC、干重）在每轮的固定日期从**独立的牺牲株**取样，不得从被追踪株上反复摘叶（会自身构成胁迫混杂）。牺牲株每物种每处理每时间点 ⩾ 3 株。
7. 全部原始照片（含灰卡）、ExG 计算脚本、拐点检测脚本、判读记录开源，第三方可重跑复现全部拐点日期（差异 ⩽ 1 天）。

5 · 数据与工具

用途	来源 / 工具
植物材料	3 个形态类群各一种，例如绿萝/多肉（肉质）、罗勒或菜豆（薄叶双子叶）、小麦或黑麦草（禾本科）；统一播种/扦插，苗龄一致（约 200 元）
栽培	统一规格花盆 60 个 + 统一基质（泥炭:蛭石 = 2:1，一次配齐全年用量，约 250 元）；托盘、标签
水分控制	电子天平 0.1 g、量程 ⩾ 3 kg（约 200 元，若学校已有可省）；**称重法**每日记录盆重换算土壤含水量
成像	手机 + 固定支架 + 黑色背景板 + 两侧 LED 面光源（约 150 元）+ 18% 标准灰卡与色卡（约 60 元）；房间遮光，全程仅用人工光源
图像分析	ImageJ/Fiji（阈值分割提取植株像素）或 Python + OpenCV，计算 ExG = 2G−R−B、GCC = G/(R+G+B)，纯 CPU
生理对照	RWC 用鲜重/饱和重/干重三次称量法（需 0.001 g 天平与 60 °C 烘箱，学校现有）；叶绿素含量可选用 SPAD 仪（**若学校无，不作为前提**，改用叶片浸提丙酮 + 分光光度计 665/649 nm 测定，属常规配置）
环境记录	温湿度记录仪（约 80 元），全程记录以排除环境漂移
统计	Python（statsmodels、ruptures）或 R（segmented、changepoint），纯 CPU
文献基准	小麦/豌豆 RGB 指数与反射率指数对应研究、落叶松针叶角度早期检测研究、城市灌木 RGB 灌溉阈值研究；**仅用于校验与对比，不计入本项目数据贡献**

预算合计约 600 – 800 元。

6 · 方法路径

1. 建立成像间：固定相机高度与角度、固定光源、遮蔽自然光；完成灰卡归一化流程与重复拍摄 CV 校验（第 1 条硬门槛）。
2. 育苗与均一化：统一播种，选取株高与叶片数落在中位数 ±10% 的个体入组，其余淘汰；记录初始盆重（饱和后沥干 2 h 的田间持水量重）。
3. 完成称重法—RWC 标定曲线，并确认对照株 RWC 稳定 ⩾ 85%（第 1 条硬门槛第二部分）。

4. 启动渐进干旱：处理组完全断水，对照组每日补水至田间持水量的 80%；每日固定时刻称重、拍照、盲法记录目视萎蔫；同步记录温湿度。
5. 每 3 天从牺牲株取样测 RWC 与叶绿素含量，构建「图像指数—生理状态」对应关系。
6. 计算 ExG/GCC 时序，用两种变点方法检测拐点，与目视萎蔫日比较求提前量；做物种间比较检验 H2。
7. 独立交叉校验：在第二轮实验中引入**复水处理**（在拐点检出后立即复水），检验 ExG 是否同步回升——若图像指数能同时捕捉损伤与恢复，其作为预警指标的可信度显著提升；若不能回升而植株恢复，说明指数捕捉的是不可逆衰老而非可逆水分状态，这本身是重要发现。

7 · 新颖性边界

> 本课题**不**声称提出新的植被指数（ExG、GCC 均为标准指数），**不**做任何高光谱或叶绿素荧光测量，**不**声称建立通用的灌溉决策模型（三物种、单一环境、盆栽尺度，不得外推到田间）。图像指数与生理状态的关联为相关性，因果表述须标注为推测。
>
> 已有工作完成了什么：RGB 色度坐标对水分状态的敏感性、ExG 在干旱早期即启动变化、RGB 指数可估计 NDVI/PRI/Fv/Fm，均已在小麦与豌豆上发表；落叶松幼苗的针叶角度在第 2 天响应而叶绿素荧光与针叶温度在第 6 天响应，已给出明确的「提前 4 天」量级对照；城市灌木的 RGB 成像水分胁迫阈值也已发表。历届丘奖生物获奖论文中，2023 年上海中学的《爬山虎攀爬形态发生机制》与 2025 年上海中学国际部的《赤霉素延长牵牛花花期》同属植物生理表型方向，但前者研究形态发生机制、后者研究激素对花期的调控，**均未涉及无损图像指标相对于目视判据的提前量**；2025 年清华附中的《植物茎内超声信号机制与实时监测系统》与本课题的「无损早期胁迫检测」目标最接近，但其信号通道是声学、依赖自制超声硬件，本课题的通道是光学、仅用手机——**通道与判据完全不同，且本课题的判断依据是相对目视萎蔫的提前天数，而非信号本身的存在性**。
>
> 本项目的贡献（主结论）：把「RGB 能否检测干旱」这一已被回答的问题，改写为「相对于实际管理判据（目视萎蔫）能提前几天，且这个提前量是否跨叶片形态稳定」，并给出带置信区间的提前量与跨物种比较。
>
> 价值：提前量是唯一能决定该方法有没有实用意义的量，而它恰恰是文献里最少被直接测量的量。**若提前量的置信区间包含 0（即 RGB 并不比人眼早），同样是有效结论**——但必须用重复拍摄 CV 与样本量证明本设计有能力分辨 2 天的差异。

8 · 决策门槛（go / no-go）

- **第 4 周末**：成像重复 CV 必须 $\leqslant$ 2%。若达不到（多因自动曝光/自动白平衡漂移），**立即降级**：其一，锁定手机专业模式的 ISO、快门与白平衡（多数机型支持，须第 2 周内核实）；其二，改用固定参数的 USB 摄像头 + 电脑采集（约 100 元）。主结论框架不变。
- **第 6 周末**：完成称重法—RWC 标定并确认对照株健康。若对照株出现病虫害或徒长，先解决栽培条件（光照强度、通风）再启动正式轮次；宁可推迟一轮，不可在不健康材料上采数据。
- **第 12 周末**：完成第一个完整干旱轮次（3 物种 $\times$ 2 处理）。若目视萎蔫的双人一致性 $\kappa < 0.80$，**立即细化判据**（改为按叶片计数的连续评分而非二值判定），并在论文中把判据的可操作性作为方法学讨论点。
- **第 20 周末**：若某物种（典型如肉质植物）在 6 周内不出现目视萎蔫，**降级**：把该物种的对照终点从「目视萎蔫」改为「RWC 首次跌破 70%」，并在结果中把两种对照口径的提前量并列报告——这不是妥协，而是把「目视判据对肉质植物失效」写成结论。
- **第 32 周末**：完成 $\geqslant$ 3 个独立轮次。若提前量的 95% CI 包含 0 但宽度 $\leqslant$ 2 天，以「RGB 指数在本条件下不早于目视萎蔫」为主结论成文。
- **需提前核实（第 2 周内）**：手机是否支持锁定 ISO/快门/白平衡；实验室是否有可遮光且温度稳定的固定空间供全年占用（这是本课题的隐性硬需求）；三物种在校内是否能同步育苗成功。

- **预算裁剪顺序**：先砍叶绿素含量测定（生理解释力下降，提前量主结论不变）→ 再砍第三个物种（放弃 H2，H1 完整保留）。**不可砍**灰卡/色卡（归一化是全部数据的基础）与天平（称重法是干旱进程控制的唯一手段）。

安全与伦理要点

- 全程为植物栽培与非破坏性成像，无微生物培养、无脊椎动物、无有害化学品的主要暴露。
- 叶绿素浸提使用丙酮（易燃、有毒），须在通风橱或通风良好处操作、远离明火，戴手套护目镜；废液单独收集交学校统一处置，**不得倒入下水道**。
- 使用的植物材料须为常见栽培种，**不得使用国家重点保护植物或有入侵性风险的物种**；实验结束后的植株按园艺废弃物处理，不得随意弃置于野外，种子不得散播。
- LED 面光源长时间使用须注意散热与用电安全，实验室无人时断电。
- 若使用带刺或有汁液刺激性的多肉植物，操作时戴手套并在实验记录中注明品种。

B17 · 线虫化学趋性指数的浓度翻转点：以饥饿时长为自变量判定「吸引—排斥」阈值是否随营养状态系统移动

`推荐优先级 P2` · `模式生物实验族` · `预算档 ≤ 900 元` · `周期档 短轮次（3–5 天/轮），但材料落实期长` · `技能取向 显微操作 + 行为计数 + 剂量–反应`

1 · 研究问题

秀丽隐杆线虫（Caenorhabditis elegans）对挥发性引诱剂（如异戊醇 isoamyl alcohol）的趋化指数（chemotaxis index, CI）随刺激浓度升高会由正转负（吸引变排斥），这个**翻转浓度**是否随虫体饥饿时长（0 / 3 / 6 / 12 h）系统移动？若移动，方向是使线虫在更高浓度下仍保持吸引（阈值上移），还是相反？

2 · 研究背景与空白

背景。趋化指数是行为神经生物学中最标准化的定量读数之一：CI = (到达测试点的虫数 − 到达对照点的虫数) / 总虫数，取值 −1 到 +1。它的价值在于**天然无量纲、天然可跨实验室比较**，且有公认的基线值可供校验——这正是中学课题最需要的方法学锚点。C. elegans 的 AWC 感觉神经元检测异戊醇、丁酮等挥发物，AWA 检测双乙酰与吡嗪；挥发性气味的检测灵敏度可低至纳摩尔量级。整套实验只需要平板、体视显微镜（或高倍放大的手机微距）、移液器和计时器。

已有工作到哪一步。经典范式源自 Bargmann 等人的工作：平板分为象限，2 个含测试化学品、2 个含溶剂对照，让线虫自由爬行至少 60 min 后计数计算 CI；标准做法是同时设置引诱剂对照（异戊醇）与排斥剂对照（1-辛醇）以保证可重复性。2025 年发表的 Current Protocols 专文《Enhancing Reproducibility in Chemotaxis Assays for C. elegans》系统梳理了影响可重复性的因素。此外已有免麻醉的「池塘测定法」用于检测极低浓度响应。

空白在：「高浓度引诱剂转为排斥」是教科书级的已知现象，但**翻转浓度本身作为一个可测参数、并检验它对内部状态（饥饿）的依赖性**的系统剂量扫描很少见于公开数据；多数研究在单一浓度下比较处理组均值。这个空白适合学生课题：单轮 3 – 5 天、完全依赖行为计数、无需任何分子手段，而「翻转点随饥饿移动」与「翻转点不随饥饿移动」两种结论都干净可发表。**但本题的可行性瓶颈不在方法而在材料获取，因此排在 P2。**

3 · 可检验假设

- **H1**：异戊醇的 CI – 浓度曲线存在可定位的翻转浓度（CI 过零点），且饥饿 12 h 组的翻转浓度显著高于饱食组（对数浓度差 ⩾ 0.5 个数量级，两者自助法 95% 置信区间不重叠）。
- **H2（特异性备择）**：该饥饿依赖的移动在 AWC 介导的异戊醇上出现，但在 AWA 介导的双乙酰上不出现（或方向相反）——若成立，说明状态依赖的调制作用于特定感觉通路而非全局运动能力。

4 · 量化验收标准

1. **方法学校验（硬门槛，两条都要过）**：(a) 复现公认基线——在标准条件（饱食、1:1000 异戊醇、60 min）下自测 CI 须为正且落在已发表的野生型 N2 基线区间内（须在开题时锁定所引文献的具体数值区间，偏差 ⩽ 0.15 个 CI 单位）；同一平板范式下 1-辛醇（排斥剂对照）的 CI 须显著为负。(b) 运动能力对照——每个饥饿处理组必须做一次**溶剂对照平板**（两侧均为溶剂），CI 的 95% CI 须包含 0，且爬出起始圈的虫比例 ⩾ 80%；若饥饿组运动能力本身下降，CI 的解释即被混杂，须在结果中显式处理。**任一条不过关，后续全部结论无效。**
2. 规模写死：4 个饥饿时长 × ⩾ 6 个浓度（10 倍稀释系列）；**生物学重复 = 独立同步化的虫群批次（不同日期、独立漂白同步化），n ⩾ 3；技术重复 = 同批次内的重复平板，n = 3**；每板投放 50 – 200 条虫，实际投放数逐板记录。

3. 虫群同步化与年龄写死：全部使用同一发育期（青年成虫 young adult），同步化方法（碱－次氯酸钠处理或定时产卵）固定不变并在方法中详述；不同批次的发育期差异 > 2 h 的数据须标记。
4. 统计口径：CI 由计数构成，用**二项/准二项 GLM** 或对计数直接建模，**禁止把 CI 当作正态连续量直接做 t 检验**；翻转浓度由 CI－lg(浓度) 的回归过零点估计，报**自助法 2000 次重抽样的 95% 置信区间**；组间比较报效应量 + 95% CI，多组用 ANOVA/GLM + 事后检验并检查前提。
5. 计数的可复现性：终点照片编号盲法，由两名成员独立计数，象限归属判据写死（虫体质心所在象限），双人计数的 ICC $\geqslant$ 0.95；对判读存疑个体（跨线）的处理规则事先写死。
6. 每轮设 $\geqslant$ 2 块无虫空白板检查污染；出现真菌或杂菌蔓延的板当轮作废并记录作废率。
7. 全部平板照片、计数表、拟合脚本开源，第三方可重跑复现翻转浓度（对数浓度差异 $\leqslant$ 0.1）。

5 · 数据与工具

用途	来源 / 工具
虫源（**最大外部依赖**）	野生型 N2 品系。渠道：高校线虫实验室赠予（最常见且免费）、国内生物试剂公司、或 CGC（Caenorhabditis Genetics Center，国际订购，周期与手续须核实）。**须在第 1 周内落实并写入 go/no-go**
食源细菌	E. coli OP50（随虫源一并索取）；**若无法获得 OP50，降级方案**：用 BSL-1 的 E. coli K-12 或市售酵母菌苔，须在方法中说明并做一次两种食源下 CI 基线的对照
培养基	NGM 平板组分（琼脂、蛋白胨、NaCl、胆固醇、$CaCl_2$、$MgSO_4$、磷酸缓冲液）（约 350 元）；LB 培养基（约 80 元）
化学刺激	异戊醇、1-辛醇（排斥剂对照）、双乙酰（可选）、无水乙醇（溶剂）（约 200 元）；叠氮化钠**不使用**（见安全要点，改用免麻醉的固定时长计数法）
观察与计数	学校现有体视显微镜；或手机微距镜头 + 背光平板 + 固定支架（约 150 元）拍照后 ImageJ 计数
温控	20 °C 恒温培养箱（**须核实**学校培养箱能否稳定在 20 °C；多数中学培养箱设定下限为 25–30 °C，若不能，须配廉价半导体恒温箱或利用空调房，列为核实项）
灭菌与无菌操作	高压灭菌锅；**无超净台时**在酒精灯火焰圈内倒板，每批设 3 块空白板做污染对照
分析	Python（statsmodels、scipy）或 R，纯 CPU
文献基线	Bargmann 经典范式与 Current Protocols 2025 可重复性专文；**仅用于校验与对比，不计入本项目数据贡献**

预算合计约 700－900 元（虫源与 OP50 若为赠予则显著下降）。

6 · 方法路径

1. **第 1－4 周专攻材料**：落实 N2 与 OP50，建立传代与同步化流程，确认 20 ° C 温控可行；连续 2 轮无污染传代方可继续。
2. 完成两条方法学校验（标准条件 CI 基线复现 + 排斥剂对照 + 运动能力对照）。
3. 核实边界：确认平板上气味梯度在 60 min 内的稳定性（用不同点样体积做一次预实验，确认 CI 不随点样体积在合理范围内剧烈变化）；确认不使用麻醉剂时的计数窗口（须固定为点样后 60 min 的瞬时计数，且明确记录「累计到达」与「瞬时停留」的判据差异）。
4. 跑饱食组的完整浓度扫描（6 个浓度 × 3 生物学 × 3 技术），定位翻转浓度并给出置信区间。
5. 加入饥饿处理（转移至无菌无食平板 0/3/6/12 h），每个饥饿时长重跑完整浓度扫描，同步做运动能力对照。
6. 拟合各组翻转浓度，做组间比较；若资源允许，在双乙酰上重跑关键浓度点检验 H2。

7. 独立交叉校验：在 2 个代表性浓度上改用**免麻醉的池塘/象限变体范式**复测 CI，与主范式结果的排序须一致；不一致则作为范式依赖性写入讨论。

7 · 新颖性边界

本课题**不**做任何遗传操作、成像或分子测量（无荧光、无 qPCR、无测序），因此**不得声称**鉴定了神经通路或分子机制；AWC/AWA 的功能归属为引用的已发表知识，不得写成本项目发现。**不**声称提出新的行为范式。

已有工作完成了什么：Bargmann 等确立的象限平板 CI 范式、异戊醇作为引诱剂对照与 1-辛醇作为排斥剂对照的标准配置、AWC/AWA 的气味特异性、挥发物纳摩尔级检测灵敏度，均为已发表内容；2025 年 Current Protocols 的可重复性专文系统整理了影响 CI 的因素；免麻醉的池塘测定法已发表。历届丘奖生物获奖论文中，2025 年北京海淀稻香湖学校《Tubby-like protein 2 对莱茵衣藻纤毛驱动运动行为的多尺度研究》与本课题同属「微小生物运动行为定量」，但其自变量是特定蛋白（依赖分子遗传学资源），本课题的自变量是**内部营养状态**且零分子依赖；2020 年山东师大附中的蚂蚁视觉巢伴识别属宏观昆虫行为，范式与判据不同。

本项目的贡献（主结论）：把教科书级的「高浓度引诱剂转排斥」现象参数化为一个可测量、带置信区间的**翻转浓度**，并系统检验它对饥饿时长的依赖性；同时以运动能力对照排除「饥饿只是让虫爬得慢」这一最显然的替代解释。

价值：单浓度比较无法区分「响应幅度改变」与「响应阈值移动」，而这两者对应完全不同的调制机制。**若翻转浓度不随饥饿移动，同样是有效结论**——它把状态依赖调制定位在增益而非阈值上；但必须用置信区间宽度证明本设计能分辨 0.5 个数量级的阈值差异。

8 · 决策门槛（go / no-go）

- **第 4 周末（生死门）**：必须已获得可稳定传代的 C. elegans 与可用食源细菌。若未落实，**立即降级、不得拖延**：其一，改用**本地土壤分离的自由生活线虫**（Baermann 漏斗法分离 Rhabditidae，成本近零），承认物种未鉴定并把研究问题改写为「某一自由生活线虫类群的 CI 翻转点及其饥饿依赖」——分析框架、统计口径与主结论完全保留，代价是失去与文献基线的直接可比性（此时第 1 条硬门槛的 (a) 改为**自建基线的轮间一致性**：连续三轮标准条件 CI 的极差 ⩽ 0.15）；其二，把体系整体换为**草履虫（Paramecium）或四膜虫的化学趋性**，同样保留 CI 框架。
- **第 6 周末**：确认 20 ° C 稳定温控可行。若不可行，改在恒定 25 ° C 下工作并重新建立全部基线（须在论文中说明温度偏离标准条件，且不再与 20 ° C 文献值直接比较）。
- **第 12 周末**：完成饱食组完整浓度扫描并定位到翻转点。若在可用浓度范围内 CI 始终为正（未见翻转），**降级**：把主终点从「翻转浓度」改为「CI 达到半数最大值的浓度（EC50 型阈值）」，饥饿依赖性的检验框架完全不变。
- **第 24 周末**：完成全部 4 个饥饿时长。若饥饿组运动能力对照不达标（爬出比例 < 80%），把最长饥饿时长从 12 h 缩短至 8 h，并把运动能力作为协变量纳入模型。
- **第 34 周末**：完成 ⩾ 3 个独立批次。若翻转浓度的组间差异 CI 包含 0 且宽度 ⩽ 0.5 个数量级，以「饥饿不改变阈值」为主结论成文，并把 CI 幅度的变化作为次级结果。
- **需提前核实（第 1 – 2 周）**：虫源渠道与到货周期（最高优先级，直接决定本题能否启动）；OP50 可得性；20 ° C 温控；无超净台条件下 NGM 平板的污染率是否可控在 10% 以下。
- **预算裁剪顺序**：先砍双乙酰臂（放弃 H2，H1 完整保留）→ 再砍体视显微镜依赖（改手机微距 + ImageJ 计数，须先验证计数一致性）。**不可砍**排斥剂对照与运动能力对照（它们是全部结论可信度的支柱）。

安全与伦理要点

- 线虫为无脊椎动物，非侵入式行为观测，**不涉及脊椎动物实验**；实验中不做任何有创操作。

- 食源细菌 E. coli OP50 / K-12 均为 BSL-1 非致病实验室菌株；**严禁培养任何致病菌**，不使用选择性致病菌培养基与抗生素筛选平板。全部菌液与平板用后高压灭菌（或 84 消毒液浸泡 24 h）后弃置。
- **明确不使用叠氮化钠（NaN_3）**——经典 CI 范式常用它在终点麻醉线虫，但叠氮化钠为剧毒品且与酸反应生成叠氮酸气体，不适合中学环境。本方案改用固定时长瞬时计数的免麻醉变体，这一取舍须在方法学中说明，并作为与文献基线比较时的已知偏差来源。
- 异戊醇、1-辛醇、双乙酰均有刺激性气味与易燃性，须在通风处操作、远离明火，单次领用量最小化。
- 若启用降级方案自土壤分离线虫，采样须遵守当地规定、不在保护区内取样；分离出的线虫全部在实验结束后高温灭活，**不得释放**；土壤样本不跨地区转移。
- 无菌操作在酒精灯火焰圈内进行时须有教师在场；平板不得开盖长时间观察。

B18 · 城市噪声与鸟类鸣声最低频率的因果分离：以同一样点内「高噪声时段 vs 低噪声时段」的时间对照判定频率上移是否独立于栖息地结构

推荐优先级 P2 · 城市生态与行为观测族 · 预算档 ≤ 700 元 · 周期档 季节受限（繁殖期鸣唱高峰，须抢 3–6 月窗口） · 技能取向 野外录音 + 声学定量 + 混合模型

1 · 研究问题

在**同一录音点位、同一物种**上比较交通高峰时段与低噪声时段（清晨/节假日）的鸣声，最低频率（minimum frequency）是否随环境噪声即时上移？该时间尺度上的上移幅度，与传统「城市 vs 郊区」空间比较得到的幅度是否一致——若空间比较的幅度显著更大，差额可归因于栖息地结构等空间混杂因子而非噪声本身吗？

2 · 研究背景与空白

背景。鸟类在噪声环境中提高鸣声频率是行为生态学中被引用最多的现象之一，但其解释一直有争议：是**适应性的频谱调整**（主动避开低频交通噪声的掩蔽），还是**唱得更响的物理副产物**（Lombard 效应导致振幅升高、而振幅与频率在发声机制上耦合）。区分这两者需要控制振幅或控制空间混杂——而空间比较（城区 vs 郊区）天然混入了栖息地结构、植被密度、种群密度等一整套差异。

已有工作到哪一步。城市梯度上的频率上移已被反复证实并具体到物种：麻雀（Passer montanus）与长尾缝叶莺（Orthotomus sutorius）在城区的鸣唱频率均显著高于近郊；红耳鹎（Pycnonotus jocosus）在城区仅将鸣唱的**第一个音节**上移；欧乌鶇（Turdus merula）的城市梯度变异已有专文。机制侧也有明确进展：有工作论证高频鸣唱可能只是唱得更响的生理后果（Lombard 效应），但**振幅升高并不能解释暗眼灯草鸦最低频率的升高**。更重要的一条是，2021 年 Behavioral Ecology 的工作指出城市鸟类通信同时受**非生物噪声与生物噪声**约束，且在同时存在虫鸣时，麻雀与缝叶莺**不再**提高最低频率——说明这个响应是情境依赖的。栖息地结构与噪声共同塑造优势频率也已有专文。

空白在：绝大多数研究是**跨样点的空间比较**，把噪声与栖息地结构、种群密度绑在一起；在**同一点位内利用时段差做时间对照**、从而把空间混杂完全消掉的设计非常少，而这正是中学生用一部录音笔就能做的设计。这个空白适合一年期课题，但**季节窗口是本题最大的死穴**：鸣唱高峰集中在繁殖期，错过就要等一年。

3 · 可检验假设

- **H1**：同一点位同一物种，高噪声时段的鸣声最低频率显著高于低噪声时段，效应量对应每 10 dB(A) 噪声升高带来 $\geqslant$ 100 Hz 的上移，且混合模型中噪声固定效应的 95% CI 不含 0。
- **H2（混杂量化备择）**：跨样点空间比较得到的「每 10 dB 上移量」显著大于同点位时间对照得到的量（两者 95% CI 不重叠）；若成立，差额即为空间混杂因子（栖息地结构等）对表观噪声效应的贡献，可给出定量估计。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：(a) 声压计与频率标定——用手机/录音笔录制**已知频率的标准音**（用另一设备播放 1、2、4 kHz 纯音），频谱峰值频率误差 $\leqslant$ 2%；用已知声压级参照（廉价校准声级计或播放定标信号）确认噪声测量的相对可比性，重复测量 CV $\leqslant$ 5%。(b) 最低频率测量的方法学陷阱必须显式处理——最低频率对信噪比与频谱图参数（窗长、动态范围阈值）高度敏感，须证明：在人工加噪的**同一段录音**上，所采用的测量流程给出的最低频率变化 $\leqslant$ 100 Hz；若变化超过阈值，必须改用对信噪比稳健的测量方式（如 −20 dB 相对阈值法或峰值频率），并在论文中把这一点作为方法学结论报告。**此项不过关，后续全部频率数据无效——这是本课题最容易被评委击穿的地方**。

2. 采样规模写死：$\geqslant 2$ 个物种 $\times \geqslant 6$ 个固定点位 $\times$ 每点位 $\geqslant 2$ 个噪声时段 $\times$ 每时段 $\geqslant 5$ 个可用鸣唱片段；**生物学重复 = 点位（近似个体/领域），$n \geqslant 6$；技术重复 = 同一录音会话内的多个鸣唱片段**。同一录音会话内的多个片段不得当作独立样本，必须以点位为随机效应。
3. 每次录音同步记录：等效连续 A 计权声级 L_Aeq（1 min）、时间、气温、风速（> 3 级风不录）、与最近道路距离、点位周边 25 m 半径内的植被盖度（照片 + ImageJ 定量，判据写死）。
4. 统计口径：以「点位」为随机截距的线性混合模型，固定效应为噪声级、时段、物种及其交互；**报固定效应系数 + 95% CI 与边际/条件 R^2，不得只报 p 值**；同一模型再加入植被盖度作为协变量，比较噪声系数的变化幅度以量化混杂。空间比较（H2）用同一模型框架在跨点位数据上重跑，两个系数用自助法比较。
5. 鸣唱片段的选取判据写死并盲法执行：信噪比阈值、片段完整性、不重复计入同一次连续鸣唱；由两名成员独立筛选片段，一致性 Cohen's $\kappa \geqslant 0.80$；物种鉴定须双人独立确认，存疑片段剔除并记录剔除率。
6. 全部原始音频、标注表、声学测量脚本（Praat 脚本或 Python/librosa）开源，第三方可重跑复现全部最低频率值（差异 $\leqslant$ 50 Hz）。

5 · 数据与工具

用途	来源 / 工具
录音	便携数字录音笔（线性 PCM，44.1 kHz/16 bit 以上）+ 防风罩（约 350 元）；或高质量手机录音 App（须核实是否有自动增益控制 AGC——**AGC 必须可关闭**，否则振幅信息全部作废）
噪声测量	廉价数字声级计（A 计权，约 150 元）；**须核实其与标准声级计的偏差**，本课题只用相对值，绝对精度不作强要求但须在论文中说明
声学分析	Praat（免费）或 Python + librosa/scipy（免费），纯 CPU；频谱图参数（窗长、重叠、动态范围）全程固定并写入方法
物种识别	现场目视 + 图鉴 + 鸣声比对；可用开源自动识别工具（如 BirdNET）作**辅助**，但**每一条纳入分析的片段必须经人工确认**，自动识别结果不得直接入库
环境协变量	手持风速计（约 80 元）、温湿度计；植被盖度用手机照片 + ImageJ
道路/土地利用	公开地图服务的道路等级与距离量算；土地利用可用公开遥感产品（如 ESA WorldCover）**需核实**下载入口与分辨率
统计	R（lme4、lmerTest、performance）或 Python（statsmodels），纯 CPU
文献基准	麻雀/缝叶莺城市梯度频率数据、红耳鹎首音节上移、暗眼灯草鹀振幅—频率解耦、生物噪声调制效应等；**仅用于对比与定位，不计入本项目数据贡献**

预算合计约 500 – 700 元。

6 · 方法路径

1. 建立录音与分析规程：固定录音增益（关闭 AGC）、固定录音距离范围、固定频谱图参数；完成标准音频率标定与噪声测量重复性校验。
2. 完成最低频率的信噪比稳健性检验（第 1 条硬门槛 b）：对同一批录音人工加入不同水平的白噪声/交通噪声，量化测量流程的偏移，据此**选定最终采用的频率测量口径**并锁定不再更改。
3. 踏查选点：在城市内选 $\geqslant 6$ 个点位，覆盖噪声梯度，且每个点位内部都存在明确的时段噪声差（例如临主干道公园入口）；记录点位环境参数。

4. 采样：在繁殖期鸣唱高峰（当地通常 3 – 6 月，须先核实目标物种物候）按「同点位、同物种、不同噪声时段」配对录音，每点位每时段 ⩾ 3 次会话分布于不同日期。
5. 声学测量与筛选：盲法编号、双人筛选片段、双人确认物种，提取最低频率、峰值频率、频宽、持续时间。
6. 拟合混合模型，估计时间尺度的噪声效应；再用同一批点位做跨样点空间比较，得到空间尺度的噪声效应，两者对比检验 H2。
7. 独立交叉校验：在 2 个点位做一次**播放实验的替代方案——被动噪声事件对照**（利用突发交通噪声事件前后 30 s 的自然对照），检验个体在秒 – 分钟尺度上是否即时调整；若无法实施，改为在同一点位比较工作日与节假日的同一时段。

7 · 新颖性边界

> 本课题**不**声称发现噪声导致鸣声频率上移（这一现象已被反复证实），**不**做任何声音回放实验（playback）以免干扰繁殖期鸟类，**不**捕捉、标记或接触任何鸟类，因此**无法**追踪个体身份——「同一点位」只是个体的近似，这一局限必须在论文中显式声明并作为不确定性来源讨论。**不**声称区分了适应性调整与 Lombard 效应的最终机制。
>
> 已有工作完成了什么：城市梯度上的频率上移已在麻雀、长尾缝叶莺、红耳鹎（仅第一音节）、欧乌鶇等物种上定量报道；「高频可能只是唱得更响的生理后果」的 Lombard 解释与「振幅无法解释暗眼灯草鹀最低频率升高」的反证均已发表；2021 年 Behavioral Ecology 的工作证明生物噪声（虫鸣）在场时麻雀与缝叶莺不再提高最低频率；栖息地结构与噪声共同塑造优势频率已有专文。历届丘奖生物获奖论文中，2024 年 The Harker School 的《跨注意力多模态视听融合的蜜蜂健康评估系统》同样用声学做动物状态判断，但其目标是分类系统的性能，本课题的目标是**效应量的因果分离**，判据与产出物完全不同。
>
> 本项目的贡献（主结论）：用同一点位内的时间对照给出噪声效应的估计，并与传统空间对照的估计做定量比较，从而给出「空间比较高估了多少」的一个具体数字与置信区间。这是**评价维度的转换**——已有工作评估的是「有没有效应」，本项目评估的是「常用设计对该效应的偏置有多大」。
>
> 价值：如果时间对照的效应显著小于空间对照，说明大量文献中的效应量含有栖息地混杂；如果两者一致，则空间设计得到一次独立的方法学支持。**两种结果都是有效结论**，但必须用置信区间宽度证明本设计有能力分辨两者的差异。

8 · 决策门槛（go / no-go）

- **第 6 周末（2026 年 10 月中，季节性前置门）**：必须完成录音设备标定与最低频率稳健性检验，并用秋季非繁殖期的零星鸣声/鸣叫（call）跑通一次完整流程。**这一步必须在冬季之前完成**，因为繁殖期窗口（次年 3 – 6 月）一旦开始就没有调试时间。若第 6 周未跑通，**立即降级**：把研究对象从「鸣唱（song）」改为全年可得的「鸣叫（call）」或改为**城市昆虫（蟋蟀/蝉）的鸣声**——后者季节窗口不同、密度高、易采样，分析框架与主结论完全保留。
- **第 10 周末**：确认目标物种在本地的繁殖期物候与鸣唱高峰月份（查阅地方鸟类志与本地观鸟记录）。若目标物种在本地不常见，改选本地最常见的 2 个鸣唱物种，并在第 12 周前完成踏查定点。
- **第 12 周末**：确认 ⩾ 6 个点位同时满足「可安全到达 + 有明确时段噪声差 + 目标物种稳定出现」。若不足，**降级**：把设计从「时间对照 + 空间对照」缩减为纯时间对照（放弃 H2），把资源集中于把每点位的会话数提高到 ⩾ 6。
- **第 28 – 40 周（2027 年 3 – 6 月，唯一的数据采集窗口）**：这是不可延期的硬窗口，须在第 26 周前完成全部准备。窗口内每两周复盘一次可用片段数，若到第 34 周可用片段 < 目标量的 50%，立即增加点位或延长每次会话时长，**不得等到窗口结束才发现数据不够**。
- **第 44 周末**：完成分析。若时间对照的噪声系数 CI 包含 0，以「在秒 – 小时尺度上未检出频率调整」为主结论成文，并把它与空间对照结果的对比作为核心论证——**这恰恰支持了「空间比较的效应含混杂」这一命题**，是强

结论而非失败。

- **需提前核实（第 2 周内）**：录音设备能否关闭 AGC（决定整个课题成立与否）；本地是否有观鸟/野生动物记录规范或保护区限制；学生在清晨时段野外活动的安全与监护安排。
- **预算裁剪顺序**：先砍风速计（改用手机气象数据 + 现场定性判断，须写明）→ 再砍第二个物种（统计效力下降，主结论仍成立）。**不可砍**可关闭 AGC 的录音设备与声级计。

安全与伦理要点

- 全程为**非侵入式被动观测**：不捕捉、不标记、不接触、不投喂、不使用回放（playback）——回放会干扰繁殖期鸟类的领域行为，本方案明确排除。
- 录音距离保持在不引起警戒或飞离的范围（观察到警戒姿态或飞离即停止并后退），不接近巢区，不为拍摄或录音而搜寻鸟巢。
- 严格遵守《中华人民共和国野生动物保护法》与地方规定；不在自然保护区、禁入区域内作业；若目标物种属保护物种，仅做远距离被动录音，不做任何可能构成"干扰"的行为，必要时事先向当地林业/园林主管部门报备。
- 清晨野外作业须两人以上同行、告知教师与家长、避开无照明的偏僻路段与水边；不进入私人领地。
- 录音不可避免会采集到路人语音——**须在方法中写明处理规则**：不针对人声录音，含可识别对话的片段在分析后立即删除、不公开发布，公开的音频数据集须做人声段静音处理。
- 不记录车牌、不拍摄可识别人脸。

B19 · 城市内部热岛梯度上同种行道树的春季物候提前量：以固定重复摄影绿度时序判定夜间人工光在控制温度后是否有额外效应

`推荐优先级 P3（季节性单次机会，风险高）` · `城市生态与物候族` · `预算档 ≤ 900 元` · `周期档 极强季节约束（秋季 2026 预演 + 春季 2027 唯一正式窗口）` · `技能取向 野外定点监测 + 图像时序 + 多元回归`

1 · 研究问题

在同一城市内部、同一树种（且尽量同一无性系来源的行道树）上，春季展叶日（leaf-out date）随城市热岛温度梯度的提前量为多少天/°C？在把日均温、夜间最低温与积温（growing degree days, GDD）都作为协变量控制之后，样点的夜间人工光照度（artificial light at night, ALAN）是否仍对展叶日有可检出的额外效应？

2 · 研究背景与空白

背景。春季物候是气候响应最灵敏、也最容易被中学生用定点重复摄影量化的生态指标。方法上有一个关键选择：人工目视判定展叶日主观性大，而**固定视角重复摄影 + 绿度色度坐标（GCC）时序 + 变点/S 形拟合**可以给出连续、可复现、带置信区间的展叶日，这正是 PhenoCam 类网络采用的做法，而它需要的全部硬件是一台带定时拍摄的相机和一个稳固支架。城市是天然的温度梯度实验室，能在几公里内制造出相当于数百公里纬度差的温差。

已有工作到哪一步。结论并不像直觉那样简单：城市化在**寒冷地区**提前展叶与开花，但在温暖的温带与亚热带地区反而**推迟**；人口密度对物候的提前作用在暖区消失甚至反转。美国本土的研究显示城市增温提前了春季物候，但同时**降低了物候对温度的敏感性**。城市内部的物候变异与土地覆被的关系也已有专文。最关键的一条是：**即使在统计上扣除了热岛效应之后，城市化程度更高处的展叶与开花时间仍有显著变化**，说明存在热岛之外的其他城市环境因子在起作用。GLOBE 项目的学生监测数据显示同一城市内展叶时间差异可达 1 – 23 天。

空白在：文献明确指出「热岛之外还有别的因子」，但**把某一个具体的候选因子（夜间人工光）在控制温度后单独检验**、且在中国城市用同种同源个体做的公开研究很少。这个空白的问题方向清晰、结论正负都成立。**但本题的致命约束是季节：春季展叶只发生一次，错过即等一年**，因此排在 P3。

3 · 可检验假设

- H1：展叶日与样点春季（2 – 4 月）平均气温呈显著负相关，提前量落在 2 – 6 天/°C 区间，回归斜率的 95% 置信区间不含 0。
- **H2（额外因子备择）**：在模型中加入 lg(ALAN 照度) 后，其偏回归系数的 95% CI 不含 0，且模型的调整 R^2 提升 ⩾ 0.05——即夜间人工光在控制温度后仍有可检出的额外提前（或推迟）效应。

4 · 量化验收标准

1. **方法学校验（硬门槛）**：(a) **秋季预演必须先跑通**——用 2026 年 9 – 11 月的秋季叶变色/落叶物候完整走一遍「定点摄影 → GCC 提取 → S 形/变点拟合 → 日期估计」全流程，并把图像法给出的日期与同步的人工目视记录比较，两者偏差 ⩽ 3 天、且双人目视判读的 Cohen's κ ⩾ 0.80。(b) 图像归一化——每张照片含标准灰卡或固定参照物，逐日 GCC 在无物候变化期（如深冬）的漂移 ⩽ 全年变幅的 5%；阴天/雨天/雾霾照片的剔除判据事先写死（基于参照物亮度阈值）。**秋季预演不过关，春季正式窗口不得启动**——这是本课题唯一的纠错机会。
2. 样本规模写死：同一树种 ⩾ 8 个样点 × 每样点 ⩾ 5 株个体（**生物学重复 = 个体树，n ⩾ 40**）；每株每日 ⩾ 1 张照片（**技术重复 = 同日多张，n ⩾ 2**）。样点须覆盖 ⩾ 2.5 °C 的实测温差，否则无法估计斜率。
3. 温度必须**实地测量而非依赖气象站外推**：每样点挂 1 个防辐射罩内的温度记录仪，全程逐时记录；同时下载最近气象站数据做交叉验证（相关 r ⩾ 0.9 方可作为备份）。GDD 的基温与起算日事先写死并不得事后调整。

4. 统计口径：展叶日为响应变量，用以样点为随机截距的线性混合模型；**报回归斜率 + 95% CI 与边际/条件 R^2，不得只报 p 值**；温度与 ALAN 共线性须用方差膨胀因子（VIF）检查，VIF > 5 须在论文中显式讨论并给出岭回归或分层建模的稳健性检验。展叶日估计的不确定度（拟合置信区间）须作为权重或误差传递进入回归。
5. 混杂控制写死：树木胸径、树冠郁闭度、朝向、距建筑距离、是否位于绿化带内，全部记录并作为候选协变量；同源性尽力保证（优先选同一批栽植的行道树，须向绿化管理部门核实**栽植年份与苗木来源，若无法核实必须在论文中明确标注为不确定项**）。
6. 全部原始照片、温度记录、GCC 脚本与建模脚本开源，第三方可重跑复现展叶日（差异 ≤ 1 天）。

5 · 数据与工具

用途	来源 / 工具
定点摄影	8 台廉价定时相机（户外延时相机或旧手机 + 定时 App + 充电宝，约 400 元）或人工每日拍摄（**零成本但依赖人力，须评估学生能否坚持每日到场**）；固定支架与参照标板（约 100 元）
温度记录	8 个纽扣式温度记录仪或 DS18B20 + 存储模块（约 250 元）+ 自制百叶箱防辐射罩（约 50 元）
夜间光照度	廉价照度计（lux meter，约 120 元），每样点在固定夜间时刻多次测量取中位数；辅以公开的 VIIRS 夜间灯光年度合成产品（NOAA/Colorado School of Mines 公开下载，约 500 m 分辨率，**须核实**分辨率是否足以分辨样点间差异——很可能不足，故以实测照度为主、遥感为辅）
气象背景	中国气象数据网（data.cma.cn）日值数据，用于交叉验证与背景说明
物候人工记录	参照中国物候观测网/BBCH 编码的展叶期定义，判据写死并双人判读
图像分析	ImageJ/Fiji 或 Python + OpenCV，提取感兴趣区（ROI）内 GCC = G/(R+G+B)，纯 CPU
曲线拟合	Python（scipy、ruptures）或 R（phenopix、segmented），纯 CPU
文献基准	城市化对物候影响随区域温度而异的研究、城市增温提前物候但降低温度敏感性的研究、城市内部物候变异与土地覆被研究；**仅用于对比与情境化，不计入本项目数据贡献**

预算合计约 700 – 900 元。

6 · 方法路径

1. **第 1 – 3 周**：选定树种与 ⩾ 8 个样点，向绿化管理部门核实栽植年份与苗木来源，取得挂设备的许可；安装温度记录仪与相机支架。
2. **第 4 – 12 周（2026 秋）**：跑秋季物候预演，完成第 1 条硬门槛的两项校验；同期检验设备越冬可靠性（电池、防水、被移动/损坏的概率）。
3. **第 13 – 20 周（冬季）**：完成 GCC 漂移校验、建立分析脚本、锁定全部判据与建模方案（预注册式地写死，避免春季数据出来后调参数）；同步做 ALAN 的夜间实测（冬季夜长，测量更方便）。
4. **第 21 – 32 周（2027 年 2 – 4 月，唯一正式窗口）**：高频采集春季展叶数据；每周检查设备与数据完整性，任一样点连续缺失 > 3 天立即补救。
5. 提取 GCC 时序、拟合展叶日与置信区间，并与人工目视记录交叉验证。
6. 拟合混合模型：先只含温度，再加入 ALAN，比较系数与模型改善；做共线性诊断与稳健性检验。

7. 独立交叉校验：用同一批样点的**第二个树种**（若可得）或同一树种的开花期（若有花）重跑全流程，检验温度斜率与 ALAN 效应是否一致。

7 · 新颖性边界

> 本课题**不**声称发现城市热岛提前物候（这一现象已被大量研究证实），**不**声称建立因果关系——ALAN 与展叶日的关联是观测性的、无法排除未测量的混杂因子（如土壤水分、人为修剪、地下管线余热），因果表述必须标注为推测。**不**做任何生理或分子测量，因此不得声称揭示光敏色素等机制。
>
> 已有工作完成了什么：城市化在寒冷地区提前、在温暖地区推迟物候的区域依赖性已发表；城市增温提前春季物候同时降低物候的温度敏感性已发表；城市内部物候变异与土地覆被的关系已有专文；**「即使扣除热岛效应后城市化仍有显著额外效应」这一关键事实已被明确报道**，本课题正是接着这句话往下做；GLOBE 学生监测已显示同城内 1 – 23 天的展叶差异。历届丘奖生物获奖论文中，2023 年上海中学的爬山虎攀爬形态发生、2025 年上海中学国际部的赤霉素延长牵牛花花期，均为受控条件下的植物发育研究，**没有野外物候时序监测**；2023 年北京王府学校的废弃矿山课题为城市生态调查但终点是修复策略。本课题的野外定点时序监测 + 控制温度后的单因子检验，在丘奖生物获奖谱系中尚无同类。
>
> 本项目的贡献（主结论）：给出一个中国城市内部的展叶提前量（天/°C）及其置信区间，作为对已发表区域依赖性结论的**独立样本检验**；并在控制温度后对夜间人工光这一具体候选因子给出带置信区间的效应估计。
>
> 价值：「换样本做独立检验」是学生课题最稳的差异化轴——若提前量与暖区文献一致（甚至出现推迟），则区域依赖性获得一次独立支持；若不一致，则指出该规律的适用边界。**ALAN 效应不显著同样是有效结论**，但必须报告本设计能分辨的最小效应，并承认样点数 8 对多元回归而言偏少这一局限。

8 · 决策门槛（go / no-go）

- **第 12 周末（2026 年 11 月底，第一道生死门）**：秋季预演必须通过（图像法与目视法偏差 $\leqslant$ 3 天、$\kappa \geqslant 0.80$）。若不通过，**立即降级、绝不拖到春天**：其一，放弃图像法、改为纯人工目视物候观测（每 2 天一次，判据写死、双人判读），代价是日期精度降至 ±2 天、需相应把样点温差要求从 2.5 °C 提高到 4 °C；其二，把研究对象整体改为**秋季物候（叶变色/落叶）**，此时 2026 秋已有一轮数据、2027 秋在提交截止（9 月 15 日）之后不可用——因此这条路要求在 2026 年秋一次性采足数据，风险须提前告知学生。
- **第 16 周末（2026 年 12 月）**：确认设备越冬存活率 $\geqslant$ 80%（被盗/损坏/断电）。若 < 80%，改为人工每日拍摄（须确认学生每日可达性）或把样点收缩到校园及学生上下学沿线可覆盖的范围。
- **第 20 周末（2027 年 1 月末）**：全部分析脚本与判据必须锁定并公开存档（预注册），此后不得因数据不好看而修改。这一条是防止事后调参的制度保证，也是可以写进论文的方法学亮点。
- **第 21 – 32 周（唯一窗口）**：每周复核数据完整性。若到第 26 周某样点数据缺失 > 20%，立即启用备用样点（须在第 12 周前预备 3 个备用点位）。
- **第 34 周末**：完成建模。若温度斜率 CI 包含 0（即本城市尺度上检不出热岛物候效应），**这是有价值的结论**（与暖区推迟/敏感性下降的文献一致），须以效能分析证明本设计能分辨 2 天/°C 的斜率，并把重心转向「同城内展叶日的总变异有多大、由什么解释」。
- **选择前提**：仅在学生（团队）能承诺**跨越整个冬春、每周到场**、且学校支持在校外挂设备时才启动本路线。**这条路线最大的风险不是科学风险而是执行连续性风险**。
- **预算裁剪顺序**：先砍照度计（改用手机照度传感器 + 一次性交叉标定，精度降但主结论框架不变）→ 再砍定时相机（改人工拍摄）。**不可砍**温度记录仪（自变量的实测是本课题不可替代的基础）。

安全与伦理要点

- 全程为野外非破坏性观测：不采摘枝叶（若需验证展叶期，每株每季取样不超过 1 个枝条且须经绿化管理方同意）、不攀爬树木、不损伤树皮、不在树体上钉挂固定件（支架须独立设置或用可拆卸绑带）。
- 在公共绿地/道路旁挂设温度记录仪与相机，**必须事先取得绿化或城管主管部门的书面同意**，设备贴标注明用途、所属学校与联系人。
- 相机 ROI 必须朝向树冠、避开人行道与住宅窗户；若不可避免拍到行人，须在方法中写明处理规则：不做人物分析、含可识别人脸的帧在分析后立即删除，公开数据集须做模糊处理。
- 夜间测照度须两人以上同行、在有照明的公共区域进行、告知教师与家长，不进入偏僻区域。
- 不涉及任何动物、微生物或化学品；无 BSL 风险。
- 树种选择须为常见栽培行道树，不涉及古树名木与保护植物；不在自然保护区内布点。

B20 · 蚂蚁双分叉桥上的集体路径选择：以 Deneubourg 非线性选择函数的拟合参数判定模型在本地蚁种上是否成立

`推荐优先级 P4（风险最高）` · `动物行为定量观测族` · `预算档 ≤ 900 元` · `周期档 依赖蚁群活动季（须解决越冬期问题）` · `技能取向 行为实验设计 + 视频计数 + 模型拟合与仿真`

1 · 研究问题

在经典双分叉桥（binary bridge）范式下，本地常见蚁种的集体路径选择结果，能否被 Deneubourg 非线性选择函数 $P_A = (k + A)^n / [(k + A)^n + (k + B)^n]$ 定量描述？拟合得到的非线性指数 n 与阈值参数 k 及其置信区间，是否与文献报道值（$n \approx 2$）一致，且在不同巢间保持恒定？

2 · 研究背景与空白

背景。双分叉桥是自组织研究中最干净的实验范式之一：蚁巢与食源之间架一座菱形桥，给出两条路径，蚂蚁只依据信息素浓度做局部决策，群体却能收敛到单一路径（等长时对称破缺，不等长时选短路）。Deneubourg 等（1990）用一个非线性选择函数刻画了这一过程，而后续工作把它推进到个体层面：蚂蚁对信息素的响应遵循**韦伯定律（Weber's Law）**——两侧信息素量之差除以其和决定转向角的大小；在个体的韦伯型响应上叠加方向噪声，就能推导出经典 Deneubourg 模型中的 S 形集体响应；基于实验参数的模型检验也已确认该模型能解释最短路径选择。这个课题的迷人之处在于它同时包含**实验**与**模型拟合/仿真**两条腿，而仿真部分纯 CPU、代码量极小。

已有工作到哪一步：Deneubourg 模型的原始双分叉桥结果、阿根廷蚁（Linepithema humile）的个体轨迹形成规则（PLoS Comput Biol 2012）、基于实测参数的最短路径选择模型验证、拥挤条件下的集体觅食决策，均已发表。

空白在：模型的参数几乎全部标定自少数几个欧美常见蚁种（尤以阿根廷蚁与 Lasius 属为主），**在其他蚁种、尤其是不同觅食策略（集群觅食 vs 个体觅食）的类群上，n 与 k 是否守恒，公开的定量检验很少**；而「换样本做独立检验」正是学生课题最稳的差异化轴。这个空白技术门槛低（只需数过桥的蚂蚁）、结论正负都成立。**但本题被排在最后是因为三重执行风险：蚁群获取与维持、越冬期活动停滞、视频计数的工作量。**

3 · 可检验假设

- **H1**：在等长双桥上，群体最终选择比例呈显著的双峰分布（≥ 70% 的重复试验中一侧占比 > 80%），即出现对称破缺；随机选择的零模型可被拒绝（二项检验 $p < 0.01$）。
- **H2（参数守恒备择）**：由不等长桥系列（长度比 1:1、1:1.2、1:1.4、1:2）拟合得到的非线性指数 n 的 95% 自助置信区间包含文献值 2；若不包含，则 n 与该蚁种的觅食策略（集群 vs 个体觅食）相关，且不同巢间 n 的差异小于种间差异。

4 · 量化验收标准

1. **方法学校验（硬门槛，两条都要过）**：(a) **对称情形复现**——在等长双桥上完成 ≥ 15 次独立试验（每次用新桥、新巢或充分清洗），对称破缺发生率须与文献一致（≥ 70%），且左右两侧被选中的次数无系统偏倚（二项检验 $p > 0.05$，若有偏倚说明装置本身不对称，必须找出并消除）；(b) **仿真校验**——自写的 Deneubourg 模型仿真在文献给定参数下必须复现已发表的选择概率曲线，偏差 ≤ 5%。**任一条不过关，后续全部结论无效。**
2. 试验规模写死：4 个长度比 × 每比 ≥ 12 次独立试验；**生物学重复 = 独立蚁巢（n ≥ 3 巢）**，每巢在每个长度比下 ≥ 4 次试验（**技术重复**）。同一巢的连续试验间隔 ≥ 24 h 且更换桥面材料（避免残留信息素），残留检验须做：用洗净桥做一次「无食源空跑」确认无偏好。

3. 计数判据写死：在桥的两臂各设一条虚拟计数线，记录每分钟通过的蚂蚁数（方向分开计），试验时长固定（如 30 min），"选择比例"定义为最后 10 min 内两臂通过量之比。视频用固定机位俯拍，计数由 ImageJ 的 kymograph/虚拟线计数或人工逐帧完成；**随机抽 20% 的片段由第二名成员独立复计，ICC ⩾ 0.95**。
4. 统计口径：通过量为计数数据，用**泊松或负二项 GLM**（先做过离散检验决定），以巢为随机效应；**禁止对通过量直接做 t 检验**。选择比例的分布用二项混合模型或 beta-二项模型建模。n 与 k 由非线性最小二乘拟合，报**自助法 2000 次重抽样的 95% 置信区间**；拟合须给出残差诊断，并与替代模型（线性响应模型）做 AIC 比较。
5. 仿真与实验的对照必须双向：用实验拟合出的参数跑仿真，检验仿真产出的对称破缺率与选择曲线是否落在实验的置信区间内。
6. 环境标准化：温度、湿度、光照、试验时段固定并记录；饥饿处理（试验前禁食时长）固定。这些若不控制，n 的估计将不可比。
7. 全部视频、计数表、拟合与仿真代码开源，第三方可重跑复现 n（差异 ⩽ 0.2）。

5 · 数据与工具

用途	来源 / 工具
蚁群	本地常见非保护蚁种（须先做物种鉴定至属级，请教高校或博物馆专家；**不得采集保护物种或外来入侵种并转移**）。采集与人工巢维持：石膏巢或试管巢 + 觅食盒（约 250 元）
桥装置	亚克力/卡纸自制菱形桥若干套（可拆洗、可换臂长），约 150 元；桥面材料须一致且可更换（推荐可替换的描图纸条，用完即弃，彻底消除信息素残留）
食源与饲养	糖水、蜂蜜、昆虫饲料（约 100 元）
录像	手机/网络摄像头俯拍 + 支架 + 均匀 LED 光源（约 200 元）
计数与分析	ImageJ/Fiji（虚拟线 kymograph 计数）；**DeepLabCut CPU 版仅作为可选增强，须核实 CPU 训练时长，若 > 24 h 直接放弃**；本课题的主计数手段不依赖深度学习
环境控制	温湿度记录仪（约 80 元）；室内恒温（空调房或简易保温箱）
仿真与拟合	Python（numpy、scipy），Deneubourg 模型仿真代码量约 100 行，纯 CPU 秒级
文献基准	Deneubourg 等的原始双分叉桥结果、阿根廷蚁个体轨迹规则（PLoS Comput Biol 2012）、基于实测参数的最短路径模型验证；**仅用于校验与对比，不计入本项目数据贡献**

预算合计约 700 – 900 元。

6 · 方法路径

1. **第 1 – 6 周**：采集并建立 ⩾ 3 个可维持的蚁巢，确认工蚁数量足够（建议每巢 ⩾ 200 只工蚁）且能稳定觅食；同期完成物种鉴定（至少到属，请专家确认）。
2. 先写仿真：在实验之前完成 Deneubourg 模型的代码实现与文献参数复现（第 1 条硬门槛 b）。**这一步在等蚁群稳定的同时做，不占额外时间，且能让学生在开始实验前就明确要拟合什么参数**。
3. 建装置并完成对称性校验：等长桥 ⩾ 15 次试验，检验对称破缺率与左右无偏倚（第 1 条硬门槛 a）；同步确认信息素残留的清除方案有效。
4. 核实能力边界：实测 ImageJ 虚拟线计数在高流量下的漏计率（用人工逐帧计数的片段作金标准），若漏计率 > 10% 则降低食源诱引强度或缩短计数窗口。**必须在正式实验前完成**。

5. 跑不等长桥系列（4 个长度比 × 3 巢 × ⩾ 4 次），提取选择比例时序。
6. 拟合 n 与 k 及其置信区间，与文献值比较；做巢间比较检验参数守恒性；与线性响应模型做 AIC 对照。
7. 独立交叉校验：用实验拟合参数跑仿真，检验仿真是否复现实验的对称破缺率与收敛时间分布；若不能，差异本身即为发现（说明模型缺失某个机制，如个体的 U 形折返或信息素挥发速率差异），须写入讨论。

7 · 新颖性边界

本课题不声称提出新的自组织模型，不声称发现信息素通讯（这是几十年前的已知），不做任何化学分析或信息素成分鉴定，不做遗传或解剖。个体韦伯定律响应、S 形集体响应的推导均为已发表结论，只能作为引用的理论框架。

已有工作完成了什么：Deneubourg 等用非线性选择函数刻画了双分叉桥结果；阿根廷蚁的个体轨迹形成规则（对信息素的韦伯型响应 + 方向噪声 → S 形集体响应）已在 PLoS Comput Biol（2012）给出；基于实测参数的最短路径选择模型验证已完成；拥挤条件下的集体决策也已发表。**历届丘奖生物 2020 年金奖即为蚂蚁课题**——山东师范大学附属中学《Ants' nestmate recognition ability based on visual cue perception》，研究的是**巢伴识别中的视觉线索感知**，即个体识别问题；本课题研究的是**集体路径决策的定量模型参数**，对象层级（个体识别 vs 群体决策）、实验范式（识别测试 vs 双分叉桥）、判断依据（识别正确率 vs 拟合参数 n/k 的置信区间）三者全部不同。这一区分必须在引言中主动写明——评委很可能记得那篇金奖论文。

本项目的贡献（主结论）：把一个标定于少数欧美蚁种的定量模型，在本地蚁种上做一次**参数层面的独立检验**，给出 n 与 k 的估计与置信区间，并检验其巢间一致性；同时通过「实验拟合参数 → 仿真 → 回验」的闭环，检验模型的预测能力而非仅拟合能力。

价值：若 n 与文献值一致，说明该非线性响应是跨物种守恒的自组织特征，模型的普适性获得一次独立支持；若显著偏离，则该参数依赖于物种的觅食策略而非机制本身——**两种结果都是有效结论**。必须用置信区间宽度证明本设计能分辨 0.5 量级的 n 差异；若 CI 过宽，须诚实报告「本样本量不足以区分」，这比给出一个无误差棒的数字诚实得多。

8 · 决策门槛（go / no-go）

- **第 6 周末（第一道生死门）**：必须已建立 ⩾ 3 个能稳定觅食的蚁巢，且完成物种鉴定至属级。若未达成，**立即降级**：其一，改为**野外原位双分叉桥**——在自然蚁道上架设人工桥，不采集不饲养，代价是无法控制巢的身份与环境条件（须把巢作为随机效应并大幅增加重复数）；其二，把体系整体改为**其他可全年室内维持的社会性/群集性无脊椎动物**（如黄粉虫幼虫的集群趋暗选择、或涡虫的双分叉趋性），主结论框架（非线性选择函数的参数拟合 + 仿真回验）完全保留。
- **第 10 周末**：仿真代码必须已复现文献曲线（偏差 ⩽ 5%），且等长桥对称破缺率与无偏倚校验通过。若装置存在系统性左右偏倚且无法消除，改为**每次试验随机交换左右臂并把臂侧作为区组因子**——不解决问题但可量化并扣除。
- **第 16 周末（越冬期，本课题特有的时间陷阱）**：温带蚁种冬季活动显著下降甚至滞育。**必须在第 10 周前就确认加温饲养能否维持觅食活性**（试养观察）。若不能，**立即把全部正式实验压缩到两个活动窗口**：2026 年 9 – 11 月与 2027 年 3 – 6 月，并把冬季（第 14 – 26 周）整段用于仿真、代码、视频计数与预分析——**这不是浪费，而是把不可用的时间投入到唯一不受季节影响的工作上**。若第 10 周判断加温可行，则冬季可继续实验，风险解除。
- **第 24 周末**：完成至少 2 个长度比的完整数据。若 ImageJ 计数的工作量导致进度落后 > 30%，缩减为 3 个长度比（1:1、1:1.4、1:2），并把节省的资源用于提高每比的重复数——参数拟合对长度比的**跨度**比对点数更敏感，因此这个缩减代价可控。
- **第 38 周末**：完成全部拟合与仿真回验。若 n 的 95% CI 宽度 > 1.5（无法区分线性与非线性），以「本样本量下模型参数不可分辨」为诚实结论，并把重心转向**已通过的对称破缺率与收敛时间分布**这一组无需拟合即可报告

的定量结果。

- **选择前提**：仅在学生已具备 Python 基础（能独立实现并调试一个随机仿真）、且团队明确接受「可能拿不到可用的 n 估计」这一风险时才启动本路线。**这是全部十条中失败概率最高的一条，也是成功时最漂亮的一条。**
- **预算裁剪顺序**：先砍 DeepLabCut 相关的一切尝试（本课题不依赖它）→ 再砍第三个蚁巢（放弃巢间守恒性检验，H2 削弱但 H1 与 n 的点估计保留）。**不可砍**可更换的桥面材料（信息素残留是最大的系统误差来源）与温湿度记录。

安全与伦理要点

- 蚂蚁为无脊椎动物，**不涉及脊椎动物实验**；全部行为观测为非侵入式，不做手术、标记（除非使用无毒水性颜料点标且经教师批准）、不施加疼痛刺激、不做致死处理。
- 采集须遵守当地野生动物与自然保护区规定：**只在校园、城市绿地等非保护区域采集常见种**，采集量控制在维持一个小型实验巢的最低限度；**不得采集保护物种**；不得采集或转移外来入侵蚁种（如红火蚁 *Solenopsis invicta*）——发现疑似红火蚁应立即停止并按规定上报，**严禁带回学校**。
- **逃逸管理**：饲养装置须设防逃屏障（滑石粉/凡士林环 + 双层容器），置于独立房间；实验结束后蚁群**原地放归至采集点**（同一地点、同一栖息环境），不得跨区域释放、不得弃置于非原采集地。若某巢无法放归（如已衰亡），按生物废弃物处理。
- 饲养期间须保证食物与水源不间断，避免因照料中断导致群体死亡——这是本课题的动物福利底线，须在方案中安排假期值班表。
- 亚克力切割、热熔胶等装置制作环节须在教师指导下进行，注意刀具与高温安全。
- 全程不使用杀虫剂、麻醉剂或任何化学制剂；桥面清洁只用清水与乙醇擦拭并充分晾干。

附录 A · 倒排时间表（52 周，通用骨架）

> 终点是 **2027 年 9 月 15 日提交截止**（按 2026 赛季日程外推，须以当年官方公告为准）。第 0 周为 2026 年 9 月第一周。每个阶段都有可测的硬门槛，不是任务描述。

周次	阶段	硬门槛（可测，不达标即触发降级）
0–2	选题定稿	从本方案选定 1 条主路线 + 1 条备选路线；确认指导老师；核对当年官方参赛指南的学科研究范围
2–6	环境与方法学校验	**跑通该路线验收标准第 1 条的方法学校验**。这是全程最重要的门槛：不过关不得进入下一阶段
6–10	数据/器材到位	数据完成下载并通过完整性检查；或器材到齐且装置重复性达标（实验类 RSD ≤ 阈值）
10–14	第一次 go/no-go	按各路线写明的阈值判定。不达标立即切换到该路线的具名降级路径，**不要拖到第 20 周**
14–28	主体研究	完成主实验/主计算的全部参数点；中期产出可展示的核心图表 3 张
28–34	第二次 go/no-go	主结论方向已明确（含"差异不显著"这一有效结论）；误差棒已能证明有分辨该差异的能力
34–40	交叉校验	用第二种独立方法重算核心量，两者一致或差异有可解释来源
40–46	中文初稿	全文完成；图表定稿；代码与数据整理为可一键重跑的仓库
46–50	英文稿与查重	**完成英文研究报告与查重报告**（总决赛强制英文，且提交材料含查重报告）
50–52	提交与答辩准备	系统提交（9 月 15 日截止，逾期不可修改）；英文 PPT 与答辩问答演练

两个容易被忽略的时间陷阱：

其一，英文化不是最后一周的事。总决赛全程英文（研究报告、PPT、答辩），从第 0 周起就要用英文记录关键术语与方法描述，否则第 46 周会成为瓶颈。

其二，11 月 3 – 10 日的公示期会把研究报告全网公开。这意味着新颖性问题会暴露在公众视野，查重与文献边界必须在提交前就处理干净。

附录 B · 资源清单（全赛道通用底座）

用途	来源 / 工具	说明
文献检索	Google Scholar、arXiv、PubMed、Semantic Scholar、Connected Papers	**必须用英文检索**；中文检索会系统性漏掉近三年工作
全文获取	arXiv、bioRxiv、PMC、作者主页、通过学校图书馆或指导老师获取	不要使用非法下载渠道，公示期会暴露
代码与数据托管	GitHub（公开仓库）+ Zenodo（存档并获取 DOI）	可复现性是评审加分项，也是自证独立完成的证据
计算环境	Python 3.11+（conda/uv 管理）、Jupyter	环境须导出为 `environment.yml` 或 `requirements.txt`
统计与作图	numpy / scipy / statsmodels / matplotlib	图表须含误差棒；配色对色盲友好
实验记录	纸质或电子实验记录本，逐日、带日期、不得回填	评委可能追问原始记录；实验类还需按要求提交实验视频
写作	LaTeX（Overleaf）或 Word；参考文献用 Zotero 管理	英文稿建议请母语者或老师润色，但**内容必须学生本人撰写**
版本管理	Git，从第 0 周开始提交	连续的提交历史是"学生本人完成"最有力的旁证

附录 C · 红线清单

> 以下每一条，触碰即可能导致取消参赛/获奖资格，或在评审中被直接判负。

资格红线 1. 指导老师不得来自公司、培训机构或任何谋利机构。组委会有理由相信学生受培训机构指导完成研究报告，即保留取消资格的权利。 2. 不得跨校组队；同一学科每人只能报名一次；同一篇论文不得报两个学科。 3. 参赛信息须据实填写；报告提交截止后不得更改团队与指导老师信息。 4. 曾参加或拟参加其他竞赛、已投稿或发表的，**必须在报名时如实申报**。

学术诚信红线 5. 不得将他人（含指导老师、高校课题组）已完成的工作写成本人贡献。研究报告须明确区分"他人已发表成果""公开数据库/事件表"与"本项目贡献"。 6. 公开数据库、已发表事件表、预训练模型只能作为对照与输入，**不得计入学生的数据或方法贡献**。 7. 不得给出未标注为推测的物理/因果外推。讨论部分可以推测，但必须明示。 8. 提交材料含查重报告；公示期报告全网公开，抄袭与过度重合会被公开发现。 9. AI 使用政策以当年官方参赛规则为准（公开公告中未见明文，**须核实**）。无论政策如何，研究报告的科学内容必须由学生本人完成；若使用 AI 辅助，按当年规则要求致谢。

方法学红线（不违规，但会被评委问倒） 10. 没有方法学校验的结果不可信——所有路线的验收标准第 1 条都是这个门槛。 11. 不平衡分类报 accuracy、时间序列随机划分、幂律只做双对数最小二乘、模拟不做收敛性检查：这四项是评审最常见的追问点，方案里已逐条写死正确口径。 12. "差异不显著"是有效结论，但必须有误差棒证明有能力分辨该差异；没有误差棒的"无差异"等于没有结论。

附录 D · 本方案的使用方式

1. **先读赛道导语**，判断这个赛道的评委口味是否匹配学生的能力结构。跨赛道选题（例如数学好的学生去做经济金融建模的理论建模族）常常比在本赛道硬拼更有胜算。
2. **按可行性顺序从上往下读**，不要从最精彩的那条开始。编号越小越稳。
3. **选定一条主路线后，必须再选一条备选路线**，且两条最好落在不同的族——这样第一次 go/no-go 触发时有真正的退路。
4. **第 2 – 6 周的方法学校验是全案的生死线。** 这一步跑不通，说明学生的技能栈与该路线不匹配，此时换题的成本远低于第 20 周换题。
5. **本方案的赛制信息取自官方公告（yau-awards.com，2020 – 2026 全部公示文章），但规则每年会改。** 报名开放后必须重新核对当年的《参赛规则》与六个学科各自的《参赛指南》，尤其是研究范围、评审标准与 AI 使用政策。