

申 报	系列：教师系列
	教学科研并重型
	专业：畜牧学
	职称：副教授

业绩成果材料

（申报人的业绩成果材料包括论文、科研项目、获奖以及其他成果等）

单 位（二级单位） 动物科学学院

姓 名 张泽宾

材料核对人：

单位盖章：

核对时间：

华南农业大学制

目 录

一、科研项目

- 1 主持 A 类：关于多基因编辑的优质猪育种新材料创制项目的立项通知及有关佐证材料..... 5
- 2 主持 A 类：关于终端父本与二元母猪杂交组合试验测定项目的立项通知及有关佐证材料..... 37
- 3 主持 B 类：关于瘦肉型猪遗传缺陷病数据库构建及遗传机制研究项目的立项通知及有关佐证材料..... 53
- 4.主参 A 类：关于 2022 年云浮市-广东省新兴县温氏生猪种业创新园项目的立项通知及有关佐证材料..... 59

二、论文、著作等

- 1 检索证明..... 99
- 2 以第一作者发表本专业论文情况
 - 2.1 第一作者 T2 类，Zhang et al.,Molecular Biology and Evolution (2020,37(1):167-182)，中科院 1 区，封面文章，IF₂₀₂₀=16.24..... 101
 - 2.2 第一作者 T2 类，Zhang *et al.*, New Phytologist (2022, 236(6): 2344-2357)，中科院 1 区，IF₂₀₂₂=9.4..... 119
- 3 以通讯作者发表本专业论文情况
 - 3.1 通讯作者 A 类,Zhou *et al.*, BMC Genomics(2023,24 (1), 701)，中科院 2 区，IF₂₀₂₃=3.5..... 135

三、其他业绩

- 1 个人荣誉
 - 1.1 院级“青年教师教学能力比赛一等奖”证书..... 151

一、科研项目

课题编号：2023ZD0404703

密 级：公开

科技创新 2030—重大项目 课题任务书

课题名称	猪基因编辑干细胞育种技术研发与新种质 创制
所属项目名称：	优质猪新品种设计与培育
所属重大项目：	农业生物育种重大项目
重大项目实施管理机构：	农业农村部科技发展中心
课题牵头承担单位：	华南农业大学（公章）
课题负责人：	李紫聪
执行期限：	2023 年 9 月至 2025 年 12 月

中华人民共和国科学技术部制

2023 年 9 月

填 写 说 明

- 一、任务书甲方即项目牵头承担单位，乙方即课题承担单位。
- 二、任务书通过“国家科技计划管理信息系统公共服务平台”，按照系统提示在线填写。
- 三、任务书中的单位名称，请按规范全称填写，并与单位公章一致。
- 四、任务书要求提供乙方与所有参加单位的合作协议，需对原件进行扫描后在线提交。
- 五、任务书中文字须用宋体小四号字填写。
- 六、凡不填写内容的栏目，请用“无”表示。
- 七、乙方完成任务书的在线填写，提交甲方审核确认后，用 A4 纸在线打印、签章后上传电子扫描件。
- 八、如项目下仅设一个课题，课题任务书只需填报课题基本信息表与课题预算部分。
- 九、涉密课题请在“国家科技计划管理信息系统公共服务平台”下载任务书的电子版模板，按保密要求离线填写、报送。一式八份报项目牵头承担单位签章，其中课题承担单位一份，课题负责人一份，作为项目任务书附件六份。
- 十、《项目申报书》和《项目任务书》是本任务书填报的重要依据，任务书填报不得降低考核指标，不得自行对主要研究内容作大的调整。《项目申报书》和《项目任务书》和本任务书将共同作为课题过程管理、综合绩效评价（验收）和监督评估的重要依据。

课题基本信息表

课题名称		猪基因编辑干细胞育种技术研发与新种质创制										
课题编号		2023ZD0404703										
所属项目名称		优质猪新品种设计与培育										
项目编号		2023ZD04047										
所属重大项目		农业生物育种重大项目										
密级		<input checked="" type="checkbox"/> 公开 <input type="checkbox"/> 秘密 <input type="checkbox"/> 机密		单位总数		4						
课题成果技术就绪度		<div><input type="checkbox"/>1. 发现基本原理<input checked="" type="checkbox"/>2. 形成技术方案<input type="checkbox"/>3. 方案通过验证 <input type="checkbox"/>4. 形成单元并验证<input type="checkbox"/>5. 形成分系统并验证<input type="checkbox"/>6. 形成原型并验证 <input type="checkbox"/>7. 现实环境的应用验证<input type="checkbox"/>8. 用户验证认可<input type="checkbox"/>9. 得到推广应用</div>										
课题成果应用的主要国民经济行业		畜牧业										
课题的社会经济目标		一级目标：农林牧渔业发展 二级目标：畜牧业										
经费预算		课题总经费根据概算批复结果核定，课题年度经费按照项目任务书约定下达。配套经费不得低于项目任务书约定要求。										
课题周期节点		起始时间		2023 年 9 月		结束时间		2025 年 12 月				
		实施周期		共 28 个月		预计中期时间点		2024 年 12 月				
课题承担单位	单位名称		华南农业大学			单位法定代表人姓名		薛红卫				
	单位性质		高等学校			组织机构代码		124400004554165634				
	单位主管部门		广东省教育厅			隶属关系		地方				
	单位所属地区		广东省			地市（市、自治州、盟）		广州市天河区				
	通信地址		广东省广州市天河区五山路 483 号			邮政编码		510642				
	单位开户名称		华南农业大学									
	开户银行（全称）		中国工商银行股份有限公司广州五山支行			汇入地点		广东省 广州市				
	银行账号		3602002609000310520			银行机构代码		102581000546				
课题负责人	姓 名		李紫聪		性 别		<input checked="" type="checkbox"/> 男 <input type="checkbox"/> 女		出生日期		1979. 03. 06	
	证件类型		身份证		证件号码		440781197903068517					

	所在单位	华南农业大学			
	最高学位	<input checked="" type="checkbox"/> 博士 <input type="checkbox"/> 硕士 <input type="checkbox"/> 学士 <input type="checkbox"/> 其他			
	职 称	<input checked="" type="checkbox"/> 正高级 <input type="checkbox"/> 副高级 <input type="checkbox"/> 中级 <input type="checkbox"/> 初级 <input type="checkbox"/> 其他		职务	系主任
	电子邮箱	lizicong@sacu.edu.cn		移动电话	15013065924
课题 联系 人	姓 名	李紫聪	电子邮箱	lizicong@scau.edu.cn	
	固定电话	020-85284985	移动电话	15013065924	
	证件类型	身份证	证件号码	440781197903068517	
课题 财务 负责 人	姓 名	肖斐	电子邮箱	37115980@qq.com	
	固定电话	020-85288032	移动电话	13640883198	
	证件类型	身份证	证件号码	362421198305040024	
其他 参与 单位	序号	单位名称		单位性质	组织机构代码
	1	华中农业大学		高等学校	121000004200048172
	2	中粮肉食投资有限公司		国有企业	91110000717884253L
	3	湖北省农业科学院畜牧兽医研究所		事业型研究单位	12420000441438622D
课题 参加 人数	<u>11</u> 人。其中：		高级职称 <u>7</u> 人，中级职称 <u>4</u> 人，初级职称 <u>0</u> 人，其他 <u>0</u> 人；		
			博士学位 <u>10</u> 人，硕士学位 <u>1</u> 人，学士学位 <u>0</u> 人，其他 <u>0</u> 人。		
课题 简介 (限 500 字以内)	本课题拟围绕“猪基因编辑干细胞育种技术研发与新种质创制”，通过猪基因组多基因编辑技术和基于干细胞的克隆技术等，创制综合性能优秀且肉质性能突出的多基因编辑猪育种新材料，为高效培育兼具优质、高繁殖力、高产、高抗病力等特性的猪新品种提供育种新材料。				

填表说明：1. 组织机构代码指企事业单位国家标准代码，单位若已三证合一请填写单位统一社会信用代码，无组织机构代码的单位填写“000000000”；
2. 单位公章名称必须与单位名称一致；
3. 单位开户名称应与单位名称一致，如有开户名称不一致等特殊情况，必须提供证明文件。

一、课题目标及考核指标、评测方式/方法

（一）课题目标

本课题为农业生物育种重大项目“优质猪新品种设计与培育”的课题3“猪基因编辑干细胞育种技术研发与新种质创制”。课题拟通过猪基因组多基因编辑技术和干细胞克隆技术，创制综合性能优秀且肉质性状突出的多基因编辑猪育种新材料，为高效培育兼具优质、高繁殖力、高产、高抗病力等特性的猪新品种提供育种新材料。

（二）考核指标

1、课题目标及考核指标（2023-2025 年）

开展多基因编辑技术研发和优质猪育种新材料创制，建立能同时在猪基因组中高效编辑多个基因的技术1套，并创制多基因编辑的兼具优质、高繁力、高产、高抗病力等特性的猪育种新材料2种。

2、课题中期目标及考核指标（截至 2024 年底）

建立多基因编辑技术1套，筛选获得多基因编辑的用于制备兼具优质、高繁力、高产、高抗病力等特性的猪育种新材料的供体细胞系3个。

（三）评测方式/方法

申请多基因编辑技术和优质猪育种新材料相关的专利，以相关的专利申请受理通知书或技术总结报告或相关论文或行业专家论证报告为测评方式；

（四）预期标志性成果

获得多基因编辑的兼具优质、高繁力、高产、高抗病力等特性的猪育种新材料1种。

课题目标、成果与考核指标表

课题目标 ¹	成果名称	成果简述	成果类型	考核指标 ²					考核方式（方法）及评价手段 ⁴
				指标名称	立项时已有指标值/状态	立项时重点国别指标值/状态	中期指标值/状态 ³	完成时指标值/状态	
<p>本课题拟围绕“猪基因编辑干细胞育种技术研发与新种质创制”，通过猪基因组多基因编辑技术和基于干细胞的克隆技术等，创制综合性能优秀且肉质性状突出的多基因编辑猪育种新材料，为高效培育兼具优质、高繁殖力、高产、高抗病力等特性的猪新品种提供育种新材料。</p>	猪多基因编辑技术	建立在猪基因组中同时高效编辑多个基因的技术	<input type="checkbox"/> 新理论 <input type="checkbox"/> 新原理 <input type="checkbox"/> 新产品 <input checked="" type="checkbox"/> 新技术 <input type="checkbox"/> 新方法 <input type="checkbox"/> 关键部件 <input type="checkbox"/> 数据库 <input type="checkbox"/> 软件 <input type="checkbox"/> 平台 <input type="checkbox"/> 应用解决方案 <input type="checkbox"/> 实验装置/系统 <input type="checkbox"/> 临床指南/规范 <input type="checkbox"/> 生产工艺 <input type="checkbox"/> 标准 <input type="checkbox"/> 论文 <input checked="" type="checkbox"/> 发明专利 <input type="checkbox"/> 其他	建立同时高效编辑多个基因的技术1套	0	0	0	1	以相关专利、总结报告或行业专家论证报告为考核方式
	多基因编辑的兼具优质、高繁殖力、高产、高抗病力等特性的猪育种新材料	利用CRISPR/Cas9等系统在供体细胞中进行多基因编辑，通过体细胞克隆制备兼具优质、高繁殖力、高产、高抗病力等特性的猪育种新材料	<input type="checkbox"/> 新理论 <input type="checkbox"/> 新原理 <input type="checkbox"/> 新产品 <input type="checkbox"/> 新技术 <input type="checkbox"/> 新方法 <input type="checkbox"/> 关键部件 <input type="checkbox"/> 数据库 <input type="checkbox"/> 软件 <input type="checkbox"/> 平台 <input type="checkbox"/> 应用解决方案 <input type="checkbox"/> 实验装置/系统 <input type="checkbox"/> 临床指南/规范 <input type="checkbox"/> 生产工艺 <input type="checkbox"/> 标准 <input type="checkbox"/> 论文 <input type="checkbox"/> 发明专利 <input checked="" type="checkbox"/> 其他	创制多基因编辑的兼具优质、高繁殖力、高产、高抗病力等特性的猪育种新材料2种	0	0	1	3	以相关专利、总结报告或行业专家论证报告为考核方式
科技报告考核指标	序号		报告类型 ⁵	数量	提交时间			公开类别及时限 ⁶	

	1	2024 年度中期进展报告	1	2024 年 6 月	延期公开（5 年）
	2	项目中期进展报告	1	2024 年 11 月	延期公开（5 年）
	3	2025 年度技术进展报告	1	2025 年 12 月	延期公开（5 年）
	4	最终科技报告	1	2025 年 12 月	延期公开（5 年）
其他目标与考核指标完成情况					

备注：

1. **“课题目标”**，应从以下方面明确描述：（1）研发主要针对什么问题和需求；（2）将要解决哪些科学问题、突破哪些核心/共性/关键技术；（3）预期成果；（4）成果将以何种方式应用在哪些领域/行业/重大工程等，并拟在科技、经济、社会、环境或国防安全等方面发挥何种的作用和影响。（5）所列主要成果原则上不超过 5 项，如有其他重要成果放在“其他”成果中表述。
2. **“考核指标”**，指相应成果的数量指标、技术指标、质量指标、应用指标和产业化指标等，其中，数量指标可以为论文、专利、产品等的数量，论文代表作应注重质量，不以数量作为评价标准；技术指标可以为关键技术、产品的性能参数等；质量指标可以为产品的耐震动、高低温、无故障运行时间等；应用指标可以为成果应用的对象、范围和效果等；产业化指标可以为成果产业化的数量、经济效益等。同时，对各项考核指标需填写立项时已有的指标值/状态、课题完成时要到达的指标值/状态，以及立项时重点国别指标值/状态。同时，考核指标也应包括支撑和服务其他重大科研、经济、社会发展、生态环境、科学普及需求等方面的直接和间接效益。如对国家重大工程、社会民生发展等提供了关键技术支撑，成果转让并带动了环境改善、实现了销售收入等。若某项成果属于开创性的成果，立项时已有指标值/状态可填写“无”，若某项成果在立项时已有指标值/状态难以界定，则可填写“/”。
3. **“中期指标”**，各重大项目根据管理特点，确定是否填写，阶段目标明确的项目课题应填写中期指标。
4. **“考核方式方法”**，应提出符合相关研究成果与指标的具体考核技术方法、测算方法等。
5. **“科技报告类型”**，包括课题综合绩效评价（验收）前撰写的全面描述研究过程和技术内容的最终科技报告、课题年度或中期检查时撰写的描述本年度研究过程和进展的年度技术进展报告以及在课题实施过程中撰写的包含科研活动细节及基础数据的专题科技报告（如实验报告、试验报告、调研报告、技术考察报告、设计报告、测试报告等）。其中，每个课题在综合绩效评价（验收）前应撰写一份最终科技报告；研究期限超过 2 年（含 2 年）的项目，应根据管理要求，每年撰写一份年度技术进展报告；每个课题可根据研究内容、期限和经费强度，撰写数量不等的专题科技报告。科技报告应按国家标准规定的格式撰写。
6. **“公开类别及时限”**，公开课题科技报告分为公开或延期公开，内容需要发表论文、申请专利、出版专著或涉及技术诀窍的，可标注为“延期公开”。需要发表论文的，延期公开时限原则上在 2 年（含 2 年）以内；需要申请专利、出版专著的，延期公开时限原则上在 3 年（含 3 年）以内；涉及技术诀窍的，延期公开时限原则上在 5 年（含 5 年）以内。涉密课题科技报告按照有关规定管理。

二、课题研究内容、研究方法及技术路线

（一）课题的主要研究内容

拟解决的关键科学问题、关键技术问题，针对这些问题拟开展的主要研究内容，限 1000 字以内。

1. 拟解决的关键科学问题和关键技术问题：

- （1）能否在猪基因组同时进行多个基因编辑；
- （2）能否通过多基因同步编辑制备综合性能优秀且肉质性状突出的猪育种新材料；

2. 拟开展的主要研究内容：

- （1）多基因同步编辑载体构建和多基因同步编辑单克隆细胞筛选
- （2）多基因编辑优质猪育种新材料创制

（二）课题采取的研究方法

针对课题研究拟解决的问题，拟采用的方法、原理、机理、算法、模型等，限 1000 字以内。

（1）多基因同步编辑载体构建

针对不同调控基因的 sgRNA 通过 tRNA 串联，使用 U3/U3b 和 7SL 启动子启动串联 sgRNA，通过 golden gate assembly 的方法将多个 gRNA 组装串联进一个 Cas9 表达载体，联合 epi 等元件，构建成多基因同步编辑工具载体。将上述载体电转至猪肾成纤维细胞中，经嘌呤霉素药物筛选后提取细胞 DNA。针对基因编辑位点设计引物进行 PCR 检测。分析测序结果，计算不同位置 sgRNA 的编辑效率，评估多基因编辑同步性。

（2）多基因同步编辑单克隆细胞筛选

将验证成功的多基因同步编辑载体电转至猪胎儿成纤维细胞中，嘌呤霉素药物筛选后挑选单克隆细胞。利用 PCR、测序及 Western blot 验证单克隆细胞中多基因同步编辑结果并计算编辑效率，选择发生多基因同步编辑的纯合单克隆细胞作为供体细胞进行体细胞核移植实验。

（3）多基因编辑优质猪育种新材料创制

结合已报道的 CD163 等疾病抗性基因，挑选影响肉色、滴水损失和肌肉脂肪含量等肉质性状及瘦肉率、料肉比等高产性状相关大效应基因作为靶基因，根据每个靶基因 mRNA 序列设计 2 对 sgRNA，在细胞水平验证单基因编辑效率，选择效率最高的 sgRNA 进行多基因同步编辑；将筛选到的纯合突变单克隆细胞注入去核的猪卵母细胞中，电融

合并激活，置于三气培养箱中培养。对受体母猪进行同期发情处理后进行输卵管克隆胚胎移植，待受体母猪妊娠产仔后，采集新生仔猪的耳组织进行基因型鉴定和表达分析。

三、主要创新点

围绕基础前沿、共性关键技术或应用示范等层面，简述课题的主要创新点。具体内容应包括该项创新的基本形态及其前沿性、时效性等，并说明是否具备方法、理论和知识产权特征。每项创新点的描述限 500 字以内。

主要创新点：通过多基因编辑创制兼具优质、高产、高繁殖力、高抗病力等特性的猪育种新材料。

多基因同步编辑是一个技术难点。此外，多种不同性状之间可能存在拮抗或负相关。通过多基因同步编辑同时改良多个相应的性状，创制兼具优质、高产、高繁殖力、高抗病力等特性的猪育种新材料，是一种前沿创新。获得的猪育种新材料将为培育综合性能优秀且肉质性状突出的猪新品种提供重要基础。

四、预期经济社会效益

课题的科学、技术、产业预期指标及科学价值、社会、经济、生态效益。限 500 字以内。

本课题拟建立的猪基因组多基因编辑技术不但可为本课题创制多基因编辑的优质猪育种新材料提供技术支撑，也可以为同行提供技术借鉴或参考，用于多基因编辑制备其它育种新材料或其它用途的基因修饰猪；此外，本课题拟创制的综合性能优秀且肉质性状突出的猪育种新材料将为培育相应的猪新品种提供重要基础。如果能获得国家审批将该多基因修饰猪育种新材料用于新品种培育，将显著提高育种的效率和效果，产生良好的社会、经济和生态效益。

五、课题年度计划

（一）2023 年 9 月—2023 年 12 月

1.任务： 靶基因筛选。

2.考核指标： 筛选与优质、高产、高繁殖力、高抗病力等重要性状相关的靶基因 5 个以上。

3. 成果形式：技术总结报告。

(二) 2024 年 1 月—2024 年 6 月

1.任务：多基因同步编辑载体构建。

2.考核指标：构建多基因同步编辑载体 3 个以上。

3. 成果形式：技术总结报告或发明专利申请受理通知书或相关论文

(三) 2024 年 7 月—2024 年 12 月

1.任务：多基因同步编单克隆细胞筛选。

2.考核指标：筛选多基因同步编单克隆细胞系 3 个以上。

3.成果形式：技术总结报告或发明专利申请受理通知书或相关论文

(四) 2025 年 1 月—2025 年 6 月

1.任务：利用克隆法制备多基因编辑的优质猪育种新材料。

2.考核指标：制备多基因编辑的优质猪育种新材料 1 种以上。

3.成果形式：技术总结报告或发明专利申请受理通知书或相关论文

(五) 2025 年 7 月—2025 年 12 月

1.任务：多基因编辑的优质猪育种新材料检测分析。

2.考核指标：完成多基因编辑的优质猪育种新材料的检测分析。

3.成果形式：技术总结报告或发明专利申请受理通知书或相关论文或专家论证报告。

六、课题组织实施机制及保障措施

1、课题的内部组织管理方式、协调机制等，限 500 字以内。

- (1) 课题负责牵头组织子课题（任务）承担单位实施。
- (2) 建立课题成员微信群随时讨论和协调课题实施，定期举行课题工作推进会。
- (3) 项目经费支出严格按预算和有关的经费管理规定和办法进行支出。
- (4) 按任务书要求的考核指标按时完成。

2、课题实施的相关政策，已有的组织、技术基础，支撑保障条件，限 500 字以内。

- (1) 课题负责人起总负责和总协调作用，负责制订子课题的总体研究方案和经费预算，并负责监督子课题研究方案的实施和经费支出。
- (2) 子课题（任务）承担单位和成员按课题负责人分工实施课题研究内容和目标。
- (3) 课题承担单位的科研平台在课题内开放、共享，为课题实施服务。按照课题的研究

内容和目标配备研究人员，组成结构合理、创新能力强的科研团队，全力投入项目研究。

(4) 各单位聘请科研助理，会同财务人员负责日常项目管理、经费预算、报账等工作，保证研究人员精力不分散。

3、对实现项目总目标的支撑作用，及与项目内其他课题的协同机制，限 500 字以内。

(1) 本课题研发的多基因编辑育种技术为项目培育优质猪新品种提供技术支撑。

(2) 本课题创制的多基因编辑优质猪育种新材料为项目培育优质猪新品种提供基础。

(3) 本课题将与项目内课题 2“猪优质高产性状功能基因挖掘与基因组选择功能位点芯片研发”协调对接，将课题 2 挖掘的高产性状重要功能基因作为基因编辑的候选基因。

七、知识产权对策、成果管理及合作权益分配

(一) **知识产权对策：**课题主持单位与参与单位在申请课题之前各自所获得的知识产权及相应权益均归各自所有，不因共同申请本课题而改变。课题实施过程中涉及课题组内部合作与课题组外单位的合作研究，将由合作双方就责、权、利达成共识，并签订合作协议。由合作研究取得的知识产权按照合作协议约定的方式进行分配。

(二) **成果管理：**积极推动课题取得成果的推广应用。共同完成的科技成果的精神权利，如身份权、依法取得荣誉称号、奖章、奖励证书和奖金等荣誉权归完成方共有；由各方共同完成的技术秘密成果，各方均有独立使用的权利。未经其他各方同意，任何一方不得向第三方转让技术秘密。

(三) **合作权益分配：**双方对共有科技成果实施许可、转让专利技术、非专利技术而获得的经济收益由双方共享。收益共享方式应在行为实施前另行约定。

八、需要约定的其他内容

无。

九、课题参加人员基本情况表

填表说明： 1、专业技术职称：A、正高级 B、副高级 C、中级 D、初级 E、其他； 2、投入本课题的全时工作时间（人月）是指在课题实施期间该人总共为课题工作的满月度工作量；累计是指课题组所有人员投入人月之和； 3、课题固定研究人员需填写人员明细； 4、是否有工资性收入：Y、是 N、否； 5、人员分类代码： B、课题负责人 C、项目/课题骨干 D、其他研究人员； 6、工作单位：填写单位全称，其中高校要具体填写到所在院系。														
序号	姓名	性别	出生日期	证件类型	证件号码	专业技术职称	职务	最高学位	专业	投入本课题的全时工作时间（人月）	人员分类代码	在课题中分担的任务	是否有工资性收入	工作单位
1	李紫聪	男	19790306	身份证	440781197903068517	A	系主任	博士	动物遗传育种与繁殖	18	B	课题研发方案制订和组织实施	Y	华南农业大学动物科学学院
2	阮进学	男	19851205	身份证	421122198512050079	A	无	博士	生物化学与分子生物专业	10	C	多基因编辑载体构建	Y	华中农业大学动物科技学院
3	毕延震	男	19790206	身份证	370304197902061919	A	副所长	博士	生物化学与分子生物学	14	C	育种新材料检测分析	Y	湖北省农业科学院畜牧兽医研究所
4	周吉隆	男	19891109	身份证	32028319891109813	B	无	博士	动物繁殖学	7	D	多基因编辑供体细胞系筛选	Y	华中农业大学动科动医学院
5	王建华	男	19850601	身份证	421023198506011218	C	技术改良与应用经理	硕士	动物遗传育种与繁殖	10	C	育种新材料制备和检测分析	Y	中粮肉食投资有限公司
6	洪林君	男	19860503	身份证	362334198605037114	B	无	博士	动物遗传育种与繁殖	12	C	多基因编辑载体构建	Y	华南农业大学动物科学学院
7	顾 婷	男	19861128	身份证	420101198611287028	B	无	博士	动物遗传育种与繁殖	12	C	多基因编辑供体细胞系筛选	Y	华南农业大学动物科学学院
8	张泽宾	男	19870323	身份证	370724198703236933	B	无	博士	动物遗传育种与繁殖	12	C	育种新材料检测分析	Y	华南农业大学动物科学学院

9	招华兴	男	19940907	身份证	440823199409072714	C	无	博士	遗传育种与繁殖	12	C	育种新材料制备	Y	华南农业大学动物科学学院
10	周昌繁	男	19910318	身份证	430923199103183511	C	无	博士	动物遗传育种	10	C	育种新材料制备和检测分析	Y	湖北省农业科学院畜牧兽医研究所
11	陈 矾	女	19891130	身份证	422202198911303820	C	无	博士	动物繁殖学	10	C	多基因编辑载体构建和供体细胞系筛选	Y	湖北省农业科学院畜牧兽医研究所
固定研究人员合计										127	/	/	/	/
流动人员或临时聘用人员合计										11	/	/	/	/
累计										138	/	/	/	/

十、课题信息表

序号：1—1 课题编号：2023ZD0404703 课题名称：猪基因编辑干细胞育种技术研发与新种质创制

填表说明：单位类型为课题牵头单位或课题参与单位。					
序号	单位名称	纳税人识别号（统一社会信用代码）	单位类型	任务分工	研究任务负责人
	(1)	(2)	(3)	(4)	(5)
1	华南农业大学	124400004554165634	课题牵头单位	建立多基因编辑技术，创制多基因编辑的优质猪育种新材料 1 种	李紫聪
2	华中农业大学	121000004200048172	课题参与单位	建立多基因编辑技术，构建多基因编辑载体和筛选多基因编辑供体细胞系	阮进学
3	中粮肉食投资有限公司	91110000717884253L	课题参与单位	利用任务 2 筛选的多基因编辑供体细胞制备优质猪育种新材料 1 种	王建华
4	湖北省农业科学院畜牧兽医研究所	12420000441438622D	课题参与单位	建立多基因编辑技术，创制多基因编辑的优质猪育种新材料 1 种	毕延震

十一、相关附件

1. 乙方与参加单位有关协议（须加盖乙方与参加单位公章、法人签字签章；协议文件须扫描上传。如无参加单位，则不填）；
2. 申报指南规定的其他附件。

关于农业生物育种重大项目——“优质猪新品种设计与培育”合作协议

甲方：华南农业大学（课题牵头单位）

乙方：湖北省农业科学院畜牧兽医研究所（课题参加单位）

甲乙双方就共同参与课题“猪基因编辑干细胞育种技术研发与新种质创制”有关事宜，进行平等协商，达成如下合作协议，并由合作双方共同恪守：

一、合作内容：

- 1、研发多基因同步编辑技术，在猪基因组中实现多个基因同步精准编辑；
- 2、利用多基因编辑技术创制优质性状突出，兼具高繁殖力、高产、高抗病力等性状的猪育种新材料 1 种。

二、责任义务

- 1、甲方作为课题牵头单位，负责本课题的任务分工、研究进展监督、总结上报及整体协调。乙方作为课题参加单位，负责完成相关部分研究内容，保证按期完成考核指标，并根据项目需求提交相关进展汇报材料。
- 2、各单位均有义务配合完成配套系组装过程中涉及培育品系的杂交组合实验并开展数据采集、分析工作。配套系培育过程中的杂交组合试验主要通过但不限于引进种公猪精液的方式开展。
- 3、甲乙双方严格遵守国家科技项目申报、立项、实施和综合绩效评价的相关法律法规和政策要求，承诺对本单位提供的各种材料及其内容的真实性负责。
甲方负责相关材料的保密工作，并保证材料仅用于此项目的开展。
- 4、技术情报和资料的保密：各单位及其有关人员均应遵照《中华人民共和国保守国家秘密法》和《科学技术保密规定》的要求，承担保密责任，双方均负有为对方技术、生产数据等资料的保密责任，未经许可，不得向第三方透漏。
- 5、严格遵守国家科技计划相关专项经费管理办法，保证资金使用的规范性和合理性，专款专用。
- 6、项目执行过程中，甲、乙双方各自取得的研究成果和相关的知识产权归各单位自己所有。若为甲乙双方共同研究形成的成果（专利、论文等）由双方另

签协议约定知识产权共享。

三、合作期限

此协议有效期同项目执行期，协议不受场地、人员等因素变动的影响。

四、附则

- 1、本协议一式四份，每方各执两份，任何一方不得违背；
- 2、如因不可抗拒因素造成的损失，双方均不承担责任；
- 3、其他未尽事宜，由双方协商确定。

甲方法人代表（签章）：

甲方（公章）：

课题负责人（签字）：

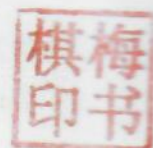
2023 年 12 月 12 日

乙方法人代表（签章）：

乙方（公章）：

课题参加单位负责人（签字）：

2023 年 12 月 11 日



Handwritten signature in black ink.

关于农业生物育种重大项目----“优质猪新品种设计与培育”合作协议

甲方：华南农业大学（课题牵头单位）

乙方：华中农业大学（课题参加单位）

甲乙双方就共同参与课题“猪基因编辑干细胞育种技术研发与新种质创制”有关事宜，进行平等协商，达成如下合作协议，并由合作双方共同恪守：

一、合作内容：

- 1、研发多基因同步编辑技术，在猪基因组中实现多个基因同步精准编辑；
- 2、筛选多基因编辑的猪克隆供体细胞系 3 个。

二、责任义务

- 1、甲方作为课题牵头单位，负责本课题的任务分工、研究进展监督、总结上报及整体协调。乙方作为课题参加单位，负责完成相关部分研究内容，保证按期完成考核指标，并根据项目需求提交相关进展汇报材料。
- 2、各单位均有义务配合完成配套系组装过程中涉及培育品系的杂交组合实验并开展数据采集、分析工作。配套系培育过程中的杂交组合试验主要通过但不限于引进种公猪精液的方式开展。
- 3、甲乙双方严格遵守国家科技项目申报、立项、实施和综合绩效评价的相关法律法规和政策要求，承诺对本单位提供的各种材料及其内容的真实性负责。
甲方负责相关材料的保密工作，并保证材料仅用于此项目的开展。
- 4、技术情报和资料的保密：各单位及其有关人员均应遵照《中华人民共和国保守国家秘密法》和《科学技术保密规定》的要求，承担保密责任，双方均负有为对方技术、生产数据等资料的保密责任，未经许可，不得向第三方透漏。
- 5、严格遵守国家科技计划相关专项经费管理办法，保证资金使用的规范性和合理性，专款专用。
- 6、项目执行过程中，甲、乙双方各自取得的研究成果和相关的知识产权归各单位自己所有。若为甲乙双方共同研究形成的成果（专利、论文等）由双方另签协议约定知识产权共享。

三、合作期限

此协议有效期同项目执行期，协议不受场地、人员等因素变动的影响。

四、附则

- 1、本协议一式四份，每方各执两份，任何一方不得违背；
- 2、如因不可抗拒因素造成的损失，双方均不承担责任；
- 3、其他未尽事宜，由双方协商确定。

甲方法人代表（签章）

甲方（公章）：

课题负责人（签字）：

2023 年 12 月 12 日

乙方法人代表（签章）：

乙方（公章）：

课题参加单位负责人（签字）：

2023 年 12 月 12 日

关于农业生物育种重大项目——“优质猪新品种设计与培育” 合作协议

甲方：华南农业大学（课题牵头单位）

乙方：中粮肉食投资有限公司（课题参加单位）

甲乙双方就共同参与课题“猪基因编辑干细胞育种技术研发与新种质创制”有关事宜，进行平等协商，达成如下合作协议，并由合作双方共同恪守：

一、合作内容：

利用体细胞克隆法创制优质性状突出，兼具高繁殖力、高产、高抗病力等性状的猪育种新材料 1 种。

二、责任义务

- 1、甲方作为课题牵头单位，负责本课题的任务分工、研究进展监督、总结上报及整体协调。乙方作为课题参加单位，负责完成相关部分研究内容，保证按期完成考核指标，并根据项目需求提交相关进展汇报材料。
- 2、各单位均有义务配合完成配套系组装过程中涉及培育品系的杂交组合实验并开展数据采集、分析工作。配套系培育过程中的杂交组合试验主要通过但不限于引进种公猪精液的方式开展。
- 3、甲乙双方严格遵守国家科技项目申报、立项、实施和综合绩效评价的相关法律法规和政策要求，承诺对本单位提供的各种材料及其内容的真实性负责。甲方负责相关材料的保密工作，并保证材料仅用于此项目的开展。
- 4、技术情报和资料的保密：各单位及其有关人员均应遵照《中华人民共和国保守国家秘密法》和《科学技术保密规定》的要求，承担保密责任，双方均负有为对方技术、生产数据等资料的保密责任，未经许可，不得向第三方透漏。
- 5、严格遵守国家科技计划相关专项经费管理办法，保证资金使用的规范性和合理性，专款专用。
- 6、项目执行过程中，甲、乙双方各自取得的研究成果和相关的知识产权归各单

位自己所有。若为甲乙双方共同研究形成的成果（专利、论文等）由双方另签协议约定知识产权共享。

三、合作期限


此协议有效期同项目执行期，协议不受场地、人员等因素变动的影响。

四、附则


- 1、本协议一式四份，每方各执两份，任何一方不得违背；
- 2、如因不可抗拒因素造成的损失，双方均不承担责任；
- 3、其他未尽事宜，由双方协商确定。

甲方法人代表（签章）：

甲方（公章）：

课题负责人（签字）：

2023 年12月12日

乙方法人代表（签章）：

乙方（公章）：

课题参加单位负责人（签字）：

2023 年12月12日

任务书签署

甲乙双方根据《国务院印发关于深化中央财政科技计划（专项、基金）管理改革方案的通知》（国发〔2014〕64号）、《国务院关于优化科研管理提升科研绩效若干措施的通知》（国发〔2018〕25号）、《国务院办公厅关于改革完善中央财政科研经费管理的若干意见》（国办发〔2021〕32号）、《科学技术活动违规行为处理暂行规定》（科学技术部令第19号）、《科技部 财政部关于印发<中央财政科技计划（专项、基金等）监督工作暂行规定>的通知》（国科发政〔2015〕471号）、《科技部 自然科学基金委关于进一步压实国家科技计划（专项、基金等）任务承担单位科研作风学风和科研诚信主体责任的通知》（国科发监〔2020〕203号）、《科技部、财政部、自然科学基金委关于进一步加强统筹国家科技计划项目立项管理工作的通知》（国科办资〔2022〕107号）等有关文件规定，以及有关法律、政策和管理要求，依据项目立项通知，签署本任务书。

同时，本单位和课题负责人郑重承诺：对本课题所有成果产出（包括但不限于新产品、新技术、标准、论文、专利等）的真实性、与项目（课题）的关联性等负责，将按要求落实科研作风学风和科研诚信主体责任；课题经费全部用于与本课题研究工作相关的支出，不截留、挪用、侵占，不用于与科学研究无关的支出；接受并积极配合相关部门的监督检查。如有违反，本单位和课题负责人以及相关成果产出者愿接受项目管理专业机构和相关部门做出的各项处理决定，包括但不限于终止课题执行、追回课题经费，取消一定期限国家科技计划项目（课题）申报资格，记入科研诚信严重失信行为数据库以及主要负责人接受相应党纪政纪处理等。

项目牵头承担单位（甲方）：

中粮肉食投资有限公司

法定代表人签字（签章）：

王明全

（公章）

2023 年 12 月 12 日

项目负责人签字（签章）：

马国刚

2023 年 12 月 12 日

课题牵头承担单位（乙方）：

华南农业大学

法定代表人签字（签章）：

薛红已



2023 年 12 月 12 日

课题负责人签字（签章）：

杨

2023 年 12 月 12 日

所属课题编号： 2023ZD0404703

密 级：公开

科技创新 2030—重大项目 子课题任务书

子课题名称	多基因编辑的优质猪育种新材料创制
重大项目实施管理机构：	农业农村部科技发展中心
课题牵头承担单位（甲方）：	华南农业大学（公章）
子课题承担单位（乙方）：	华南农业大学（公章）
子课题负责人：	张泽宾
执行期限：	2023 年 9 月至 2025 年 12 月

中华人民共和国科学技术部制

2023 年 9 月

填写说明

一、任务书甲方即课题牵头承担单位，乙方即子课题承担单位。

二、任务书中的单位名称，请按规范全称填写，并与单位公章一致。

三、任务书中文字须用宋体小四号字填写。

四、凡不填写内容的栏目，请用“无”表示。

五、《项目任务书》和《课题任务书》是本任务书填报的重要依据，任务书填报不得降低考核指标，不得自行对主要研究内容作大的调整。《项目任务书》和《课题任务书》和本任务书将共同作为课题过程管理、综合绩效评价（验收）和监督评估的重要依据。

一、子课题任务

利用体细胞克隆技术等创制多基因编辑的兼具优质、高繁力、高产、高抗病力等特性的猪育种新材料 1 种。

二、考核指标（具体化、量化）

1. 研究计划及考核指标（2023-2025 年）

（1）2023 年 9 月-2023 年 12 月

研究计划：研发体细胞克隆技术。

考核指标：初步建立猪体细胞克隆技术。

（2）2024 年 1 月-2024 年 12 月

研究计划：制备优质性状突出，兼具高繁力、高产、高抗病力等特性的猪育种新材料。

考核指标：制备优质性状突出，兼具高繁力、高产、高抗病力等特性的猪育种新材料 1 种。

（3）2025 年 1 月-2025 年 12 月

研究计划：检测和分析制备的优质猪育种新材料。

考核指标：完成优质猪育种新材料的检测和分析。

2. 课题中期考核指标（截至 2024 年底）

制备优质性状突出，兼具高繁力、高产、高抗病力等特性的猪育种新材料 1 种。

3. 预期标志性成果

制备优质性状突出，兼具高繁力、高产、高抗病力等特性的猪育种新材料 1 种。

三、知识产权保障措施及权益分配机制

本课题严格按照《中华人民共和国专利法》、《中华人民共和国著作权法》和《关于加强国家科技计划知识产权管理工作的规定》等国家有关法律法规和科技部、农业农村部有关知识产权保护的规定，对项目实施产生的知识产权进行管理。项目申请单位和参与单位之间签订项目实施协议，明确知识产权成果管理及合作权益分配的相关事项。

本项目研究过程中合作完成的成果归项目团队所有，无条件用于项目年度总结、中期检查、绩效评价等主管部门要求的管理环节汇报和材料报送。根据各方对成果的实际贡献大小来进行成果署名和权益大小分配。遵守严格保密制度，本项目涉及的相关申报材料、试验数据、研究成果、验收资料等全套项目档案管理要符合相关保密要求，实行论文发表、项目进展及成果公开宣传审核制度，未经审核的论文、成果不得擅自发表、宣传。

四、经费预算表

序号	预算科目名称	经费（万元）
1	一、中央财政专项资金	50.00
2	（一）直接费用	45.00
3	1. 设备费	0.00
4	（1）购置设备费	0.00
5	2.业务费	40.00
6	2.1 材料费	20.00
7	2.2 测试化验加工费	10.00
8	2.3 燃料动力费	0.00
9	2.4 出版/文献/信息传播/知识产权事务费	4.00
10	2.5 会议/差旅/国际合作交流费	6.00
11	3.劳务费	5.00
12	3.1 劳务费	5.00
13	3.2 专家咨询费	0.00
14	（二）间接费用	5.00
15	二、其他来源资金	0.00
16	三、 合计	50.00

五、双方责任

1、甲乙双方根据《国务院办公厅关于改革完善中央财政科研经费管理的若干意见》

（国办发〔2021〕32号）和项目管理各项规章制度等有关文件规定，及相关法律、政策和管理要求，签署本任务合同书。

根据本任务合同书，甲方从课题经费中分配 50.00 万元给乙方用于完成本子课题的研究任务。

2、课题负责人与课题参加人全力以赴开展研发，严格按照课题任务指标和本课题任务合同书要求，按时保质保量完成研发目标。不得以任何理由降低课题目标要求，拖延课题进展，推诿工作责任。

3、课题负责人与课题参加人定期向项目牵头单位报送课题进展，若经年度考核，课题牵头单位未按课题任务合同书要求完成考核指标，项目牵头单位有权建议调整经费或终止课题的实施单位、追回已拨付资金。对本课题所有成果产出（包括但不限于新产品、新技术、标准、专利等）的真实性、与课题的关联性等负责，落实科研作风学风和科研诚信主体责任。

4、课题经费实行专款专用、单独核算；全部用于与本课题研究工作相关的支出，不截留、挪用、侵占，不用于与科学研究无关的支出；接受并积极配合相关部门的监督检查。如有违反，课题牵头单位和课题负责人以及相关成果产出者愿接受项目主管单位和相关部门做出的各项处理决定，包括但不限于终止课题执行、追回课题（课题）经费，取消一定期限国家科技计划课题申报资格，记入科研诚信严重失信行为数据库以及主要负责人接受相应党纪政纪处理等

5、本合同一式叁份，报送项目牵头单位壹份，甲、乙双方各壹份，具有同等效用。

六、协议签约盖章



甲方:

课题承担单位(公章)

法人(签章)

薛红已

课题负责人(签字)

张泽兵

年 月 日



乙方:

子课题承担单位(公章)

法人(签章)

薛红已

子课题负责人(签字)

张泽兵

年 月 日

课题编号：NK2022110405

内部文件

子课题任务协议书 (2024-2026年度)

课题名称：快长、节粮、优质配套系培育

课题承担单位（甲方）：佛山大学

子课题名称：组装终端父本与最优二元母本杂交组合试验及配合力测定

子课题承担单位（乙方）：华南农业大学

子课题负责人：袁晓龙

执行期限：2024年1月1日至2026年12月31日

二〇二四年一月

一、课题任务：

利用本项目培育的终端父本，分别与经杂交组合试验遴选出来的优秀二元母猪开展配套系杂交，生产商品猪。运用基因组精准选配技术指导各个组合配套系杂交试验的配种方案制订，对所有参与杂交组合试验的亲本个体进行基因芯片检测。测定不同杂交组合所生后代商品猪的生长速度、饲料利用率和背膘厚；随机挑选商品猪进行屠宰，测定胴体和肉质性状。比较不同组合试验商品猪的综合生产性能表现，整体确定最优配套系生产类型。

二、考核指标

1. 总体考核指标

获得第三方权威机构出具的杂交组合试验检测报告；开发针对快长、节粮和快长、优质杂交试验评估模型各 1 个，基因组选配效果分析方法各 1 个，发表论文 1~2 篇。

2. 阶段性目标

（1）2024 年度目标

开展终端父本与二元母猪杂交组合试验测定，并获得第三方机构出具的杂交组合试验检测报告 1 份。

（2）2025 年度目标

对参与杂交组合试验的亲本个体开展芯片检测，开发针对快长、节粮和快长、优质杂交试验评估模型各 1 个，发表文章 1 篇。

（3）2026 年度目标

育成快长、节粮和快长、优质的杂交配套系，并开发完成基因组选配效果分析方法 1 个，发表文章 1 篇

三、知识产权保障措施及权益分配机制

本课题严格按照《中华人民共和国专利法》、《中华人民共和国著作权法》和《关于加强国家科技计划知识产权管理工作的规定》等国家有关法律法规和科技部、农业农村部有关知识产权保护的规定，对项目实施产生的知识产权进行管理。项目申请单位和参与单位之间签订项目实施协议，明确知识产权成果管理及合作权益分配的相关事项。

本项目研究过程中合作完成的成果归项目团队所有，无条件用于项目年度总结、中期检查、绩效评价等主管部门要求的管理环节汇报和材料报送。根据各方对成果的实际贡献大小来进行成果署名和权益大小分配。遵守严格保密制度，本项目涉及的相关申报材料、试验数据、研究成果、验收资料等全套项目档案管理要符合相关保密要求，实行论文发表、项目进展及成果公开宣传审核制度，未经审核的论文、成果不得擅自发表、宣传。

四、全面落实保密制度

严格遵守国家保密法律法规和农业关键核心技术攻关保密管理规定，制定项目保密管理制度，从严限定攻关项目知悉范围。项目参与人员对所获得的项目文件及项目相关信息，不复制、不传播、不扩散至无关的单位和个人。不在任何公开场合发表关于项目的言论，不在任何媒体、媒介上对项目情况进行宣传。

五、经费预算表

序号	预算科目名称	经费（万元）
1	一、中央财政专项资金	80.00
2	（一）直接费用	68.00
3	1. 设备费	6.00
4	（1）购置设备费	6.00
5	2.业务费	39.90
6	2.1 材料费	7.90
7	2.2 测试化验加工费	20.00
8	2.3 燃料动力费	0.00
9	2.4 出版/文献/信息传播/知识产权事务费	5.00
10	2.5 会议/差旅/国际合作交流费	7.00
11	3.劳务费	22.10
12	3.1 劳务费	20.00
13	3.2 专家咨询费	2.10
14	（二）间接费用	12.00
15	二、其他来源资金	0.00
16	三、合计	80.00

六、双方责任

1、甲乙双方根据《国务院办公厅关于改革完善中央财政科研经费管理的若干意见》（国办发〔2021〕32 号）和项目管理各项规章制度等有关文件规定，及相关法律、政策 和管理要求，签署本任务合同书。

根据本任务合同书，甲方应向乙方拨付 2024 年度课题经费 80.00 万元。项目年度经费拨付到乙方指定的下列账户：

收款单位：华南农业大学

开户银行：中国工商银行广州五山支行

账 号：3602002609000310520

2、课题负责人与课题参加人全力以赴开展研发，严格按照课题任务指标和本课题任务合同书要求，按时保质保量完成研发目标。不得以任何理由降低课题目标要求，拖延课题进展，推诿工作责任。

3、课题负责人与课题参加人定期向项目牵头单位报送课题进展，若经年度考核，课题牵头单位未按课题任务合同书要求完成考核指标，项目牵头单位有权建议调整经费或终止课题的实施单位、追回已拨付资金。对本课题所有成果产出（包括但不限于新产品、新技术、标准、专利等）的真实性、与课题的关联性等负责，落实科研作风学风和科研诚信主体责任。

4、课题负责人与课题参加人严格遵守国家保密法律法规，从严限定课题知悉范围，并与核心涉密人员签订保密协议。

5、课题经费实行专款专用、单独核算；全部用于与本课题研究工作相关的支出，不截留、挪用、侵占，不用于与科学研究无关的支出；接受并积极配合相关部门的监督检查。如有违反，课题牵头单位和课题负责人以及相关成果产出者愿接受项目主管单位和相关部门做出的各项处理决定，包括但不限于终止课题执行、追回课题（课题）经费，取消一定期限国家科技计划课题申报资格，记入科研诚信严重失信行为数据库以及主要负责人接受相应党纪政纪处理等

6、本合同一式伍份，报送项目牵头单位壹份，甲、乙双方各贰份，具有同等效用。

七、协议签约盖章

甲方：

课题承担单位（公章）



课题负责人（签字）

李红

法人（签章）

马少明

2024年9月12日

乙方：

子课题承担单位（公章）



子课题负责人（签字）

李红

法人（签章）

薛红已

2024年9月12日

附、子课题参加人员基本情况表

填表说明:

1. 专业技术职称: A. 正高级 B. 副高级 C. 中级 D. 初级 E. 其他;
2. 投入本课题的全时工作时间(人月)是指在课题实施期间该人总共为课题工作的满月度工作量;累计是指课题组所有人员投入人月之和;
3. 人员分类代码: A. 课题负责人 B. 课题骨干(指子课题负责人) C. 其他人员(仅填固定工作人员);
4. 工作单位: 填写单位全称,其中高校要具体填写到所在院系。

序号	姓名	性别	出生日期	国籍	证件类型	证件号码	专业技术职称	职务	最高学位	专业	投入本课题的全时工作时间 (人月)	人员分类代码	工作单位	签名
1	袁晓龙	男	1989-04-15	中国	身份证	341621198904152331	B	无	博士	动物遗传育种与繁殖	3	B	华南农业大学	袁晓龙
2	张泽宾	男	1987-03-23	中国	身份证	370724198703236933	B	无	博士	动物遗传育种与繁殖	6	C	华南农业大学	张泽宾
累计											9	/	/	/

任务编号：NK202211040501

密级：公开

国家重点研发计划 课题任务合约书

任务名称：终端父本与二元母猪杂交组合试验测定

所属课题名称：组装终端父本与最优二元母本杂交组合试验及配
合力测定

所属课题编号：NK2022110405

所属项目名称：快长、节粮、优质配套系培育

所属项目编号：NK20221104

课题承担单位（甲方）：华南农业大学

课题负责人：袁晓龙

任务承担单位（乙方）：华南农业大学

任务负责人：张泽宾

执行期限：2024 年 1 月至 2024 年 12 月

二零二四年一月

填写说明

- 1、合约书为课题验收的依据，其各项内容应尽可能详细填写。
- 2、每个任务的目标要强调形成主推技术规程和规范、重要标准、决策支持方案、知识产权、种质资源收集数量和质量等；考核指标应具体、量化、可考核。
- 3、经费的使用应严格按有关经费管理办法执行。
- 4、本协议书要求用 A4 纸、正文小四号宋体、双面打印并装订。
- 5、协议书正式文本一式 6 份，甲方保留 2 份，乙方保留 2 份，项目牵头单位保留 2 份。

一、乙方主要研究内容		
<p>1、在本项目培育的终端父本，分别与经杂交组合试验遴选出来的优秀二元母猪开展配套系杂交，生产商品猪</p> <p>2、应用杂种优势预测技术筛选杂交配套组合，集成新品种高效生产配套技术，建立从核心育种群到优质商品猪生产的高效繁育体系，开展产业化。</p>		
二、乙方总体目标、年度任务及考核指标		
总体目标	开展终端父本与二元母猪杂交组合试验测定，并获得第三方机构出具的杂交组合试验检测报告 1 份。	
年度	年度任务	考核指标
2024	开展终端父本与二元母猪杂交组合试验测定，并围绕主要胴体指标和肉质指标进行测定，并获得第三方机构出具的杂交组合试验检测报告 1 份。其中主要胴体指标包含屠宰前活重、胴体重、屠宰率、眼肌面积；主要肉质指标包含肉色评分、大理石纹评分、滴水损失率、剪切力、肌肉脂肪含量。	(1)对至少 50 头终端父本与二元母猪杂交后代开展胴体和肉质测定。(2)12 月底前提交第三方机构出具的杂交组合试验检测报告 1 份。

三、乙方经费预算		
预算科目名称	金额（万元）	备注
一、中央财政专项资金	64.10	
（一）直接费用	64.10	
1、设备费	0.00	
其中：购置设备费	0.00	
2、业务费	60.00	包括材料费、测试化验加工费、燃料动力费、出版/文献/信息传递/知识产权事务费、会议/差旅/国际合作交流费
3、劳务费	4.10	包括劳务/专家咨询费
（二）间接费用	0.00	
二、其他来源资金	0.00	
三、合计	64.10	

注：1. 绩效支出在间接费用中无比例限制，要处理好合理分摊间接成本和对科研人员激励的关系，绩效支出安排与科研人员在任务工作中的实际贡献挂钩。

四、乙方参加人员基本情况表									
序号	姓 名	工作单位	身份证号	性别	专 业	职称/职务	责任分工	投入工作时 间（人月）	联系电话
1	张泽宾	华南农业大学	370724198703236933	男	动物遗传育种与繁殖	副高	任务方案制 定、组织协 调实施等	6	15600660868
								6	/
								0	/
								6	/
填表说明： 1、职称分类：正高级、副高级、中级、初级、其他； 2、投入工作时间（人月）是指在实施期间该人总共为任务工作的满月度工作量；累计是指任务所有人员投入人月之和； 3、任务固定研究人员需填写人员明细； 4、工作单位：填写单位全称，其中高校要具体填写到所在院系。									

五、需要约定的其他内容

（一）关于科研成果、知识产权和奖励的相关约定

在课题执行过程中涉及到的信息公开与分享、科研成果处理、知识产权申请与转让、奖励申报和收益分享等事宜按照以下约定执行，本协议未尽之处，应采取“一事一议”的方式签订补充协议：

1、课题执行期间，各方承诺尽最大可能互为提供资料数据，共享研究成果，但相关资料和数据仅限于各方的研究目的，任何方都不得将其他方未公开的材料和资料向其他方转移和泄露。

2、承担单位与参加单位在课题执行日之前各自所获得的知识产权及相应权益均归各自所有，不因共同合作本课题而改变。

3、在课题执行过程中，各方应对科技成果及时采取知识产权保护措施，并按照国家科技计划知识产权管理相关规定决定归属。独自完成的科技成果及获得的知识产权归各方独自所有，相关成果被授予的奖励归各方独自所有。各方共同完成的科技成果及其形成的知识产权归各方共有，共同享有知识产权使用权，相关成果获得的荣誉和奖励归完成各方共有。

4、共有知识产权所有权申请及转让需要各方共同同意，并另行起草签署书面约定明确归属和收益共享方式。无论是独有还是共有的知识产权转让，课题各参与方有以同等条件优先受让的权利。

5、各方承诺本项目产生的所有科学数据无条件、按期递交到科技部指定的平台，在专项约定的条件下对专项各承担单位，乃至今后面向所有的科技工作者和公众开放共享。项目内采集的种质资源，包括活体在内，全部按项目技术要求汇交到项目指定的种质资源库或保种场。

（二）课题资产管理和归属约定

本课题中央财政经费支持的设备购置与样机试制、资料材料以及数据等资产按国家科技计划管理办法相关要求执行；自筹经费提供方与使用方应另行签订相关协议约定自筹经费投入产生的资产归属问题。

1、中央财政经费投入研制或购置的设备、开发的样机、资料材料及数据等按国家科技计划有关法规决定所有权和归属，并至少在课题参加单位间提供共享。

2、配套自筹经费研制或购置的设备、开发的样机、资料材料及数据等，在不违反国家法律和相关规定的基礎上，按自筹经费提供方与使用方之间的协议处理。

（三）课题执行违约责任

1、根据任务书约定，课题参与方负有按时完成各自负责任务并达到相关考核指标的义务。参加单位研发进展滞后，课题承担单位有权督促相关责任方加快进度；出现进展严重滞后并影响课题甚至整个课题考核指标完成的情况，课题承担单位报项目承担单位、项目承担单位报专业机构后，有权缓拨、停拨、甚至追缴部分或全部课题经费。

2、参加单位为完成任务书规定研究任务的支出，超出预算的部分由其自行承担。经费承诺的自筹经费不能按时足额到位，课题承担单位有权督促，必要时向项目承担单位和专业机构进行报告，申请相应调整处罚措施。

3、违反本协议第四条关于科研成果和知识产权申请和权属等约定，违约方向知识产权所属人支付违约金赔偿相关损失；在项目（课题）承担单位或有关部门调节无法达成谅解的情况下，课题各方均有权通过法律途径追究违约方责任，但相关纠纷不作为影响课题研究进度的理由。

4、课题因难以克服的技术挑战或无法预见的客观条件变化而无法达到预期考核指标的情况，应及时通知项目（课题）承担单位，及时报告专业机构申请调整，责任和损失由各方协商共同决定承担方式。因责任方未及时通知项目（课题）承担单位造成的额外损失，由相关责任方自行承担。

5、因不可抗力不能履行任务书规定义务时，可以免除违约责任，但应及时通知课题承担单位和项目承担单位，并按相关流程及时报告专业机构。在出现不可抗力的情况下，各方均采取适当措施减轻损失。因责任方未及时采取应对措施或通知项目（课题）承担单位造成的额外损失，由相关责任方自行承担。

（四）课题过程管理及验收

课题承担单位应积极配合专业机构对课题执行的过程管理和验收，并采取召开会议、进展报告等方式协调和促进项目和课题的执行，督促和保证本课题的研究任务按时完成，并达到相应考核指标。

课题参加单位应严格按照国家科技计划管理的相关规定和办法执行课题预算，保证任务书规定的研究任务按时完成，并达到相应考核指标。

因一方或几方原因导致整个项目或课题验收不通过，相关参加单位负责承担责任。

（五）补充协议签署和争议解决办法

1、若课题执行过程中，任何重大调整（如任务考核指标调整、经费调整、参

加单位变化等)都应及时通知项目(课题)承担单位,报专业机构批准后签署补充协议。补充协议应对调整后的各方责任义务进行约定,与本协议具有同等效力。

2、在项目执行过程中,各参加单位发生争议应当友好协商解决。项目(课题)承担单位出面协调无法达成一致的,可请求专业机构或相关科技主管部门进行调解。

3、本协议自各方盖章(签字)之日起生效,有效期至项目验收合格之日。协议一式6份,甲方保留2份,乙方保留2份,上交项目承担单位2份。具有同等法律效力。

六、协议签约各方

甲 方(单位公章):

法定代表人(签章):

课题负责人(签字):

年 月 日

乙 方(单位公章):

法定代表人(签章):

任务负责人(签字):

年 月 日

特定高校学科建设专项（人才引进类）

项目（课题）任务书

项目名称：新型养殖模式下动物疫病“全健康”综合防控关键共性技术研究

课题名称：瘦肉型猪遗传缺陷病数据库构建及遗传机制研究

项目起止时间：2024年01月01日至2025年12月31日

管理单位（甲方）：华南农业大学

依托学院（乙方）：动物科学学院

课题负责人（丙方）：张泽宾 联系电话：15600660868

课题联系人：张泽宾 联系电话：15600660868

华南农业大学
二〇二二年制

一、研究计划

（一）主要研究内容及创新点

本项目拟在前期50K SNPs芯片的基础上，首先对杜洛克、长白、大白等纯种瘦肉型猪群体的部分代表性个体进行全基因组重测序，继而采用缺失基因型填充技术将基因组序列信息填充至整个实验群体，开展全基因组序列的GWAS分析，精细定位QTL区域；第二，通过进一步收集猪的阴囊疝、脐疝、血管瘤、皮肤不全等遗传缺陷表型个体的组织素材，对QTL精细定位区间内的 SNPs 进行基因分型并构建单倍型，针对与阴囊疝高度关联的单倍型，开展基于后裔同源的共享单倍型分析，充分利用重组事件实现QTL的再次精细定位。第三，将目的区域高深度重测序数据及全球猪种基因组重测序数据库对比，进行同源重组分析，锚定易感基因，搜寻和发现致病突变位点。通过生物信息学分析基因通路，以及基因位点之间的互作，初步解析猪阴囊疝发生的分子机理。本项目成果将为深入解析猪和人类阴囊疝发生的分子遗传机理提供重要依据。

创新点：

- 1、通过群体调查，从经济的角度清晰了解对温氏种猪影响最大的遗传缺陷。
- 2、从不同的角度分析遗传缺陷发生的概率，探索闭锁群体发生的遗传缺陷的合适概率范围，以便及时发现群体异常情况。
- 3、建立首个遗传缺陷与异常的特殊样本库，为深入研究提供了材料平台，为以分子育种新技术的研发奠定坚实基础。

（二）拟开展的研究在国际国内同领域所处的地位

猪的遗传缺陷是由于生殖细胞或受精卵中的遗传物质在结构或功能上发生了改变，从而使发育的个体所患的缺陷或异常。具有垂直传递和终生性特征"，其遗传机制不外乎染色体畸变和基因突变。在染色体畸变方面，有染色体数目变异引起的缺陷和染色体结构变异，如先天性卵巢发育不全属于染色体数目变异引起的缺陷。在基因突变方面，包括了单基因遗传缺陷与多基因遗传缺陷。猪的遗传缺陷与异常有多少种，目前无确切统计。在Nicholas（1998）的一篇综述中包括了161个性状和异常，其中有些属于正常的遗传性状，如毛色、血型等。

结合前人的研究和现场实践的发现，可以将猪的遗传缺陷和异常种类总结如下：毛皮系统方面的缺陷和异常有皮肤发育不全、植物性皮炎(Flatia, 1961)、稀毛症、内翻乳头（Clayton, 1981）、玫瑰糠疹、卷毛、羊毛状毛等；骨骼系统方面的主要有：矮小症（Anon, 1991）、骨软骨病、骨骼异常综合征（Pulawska因子病，Dabczewski, 1949）、脊椎异常、曲尾、前肢肥厚、后肢外展（八字脚）、无腿、多趾、无下颌并耳畸胎多趾（Malynicz, 1982）、并趾、三腿、关节屈曲、两后腿不对称综合征、肢蹄软弱综合征（蹄裂、蹄枕、蹄冠脓肿、偏蹄、关节肿胀、肢势不正、蹄质疏松而致跛行或不能站立）。头部方面

（四）科研组织管理、国内外合作设想

1. 项目由课题负责人课题执行负责制。课题负责人将根据课题任务，组织课题成员开展科研，包括制定具体研究计划，组织协调研究力量，合理使用研究经费，接受科技部、财政部等单位的检查；依据首席科学家及专家组的评估指导意见，调整研究策略，按时规范向项目主持单位提交年度研究报告及研究成果，项目结题时提交研究数据和资料。

2. 项目研究跟踪及成果管理制度。项目组定期开展年度学术交流，首席科学家及各课题负责人向项目专家组汇报研究进展及项目遇到的问题，听取专家组意见，并根据意见进行调整；项目研发成果及时汇总，所有需要署名资助情况的，都需要标注项目资助及项目编号。

本项目拟与温氏集团合作，温氏集团目前已经在EAS系统中有较为完备的遗传缺陷与系谱记录，可直接开展遗传缺陷的种类、分布、比例等调查分析。对于遗传缺陷的筛选与淘汰方案措施，现场已经开展多年，在闭锁选育的情况下，将发生频率控制在现有的水平，这一工作已经非常成熟。现在仅需根据遗传方式不同将各遗传缺陷进行归类，制定更有针对性的详细的筛选淘汰方案，对于有多年现场经验的员工，可以很快落实。另一方面，目前对于出生仔猪的样品采集已经规范化操作，针对特殊表型样品的采集工作仅需要做专业培训就可以很快在猪场实现，建立特殊样本库，为开发分子标记提供良好的材料平台。华南农业大学和温氏具有良好的合作传统，且前期已经利用80k基因芯片分型技术开展了关于大白的长趾甲的分子标记研发，获得了初步的结果。其具备全基因组关联分析、分子标记开发和挖掘的生物统计分析技术。这些工作基础可使得本项目顺利开展。

（四）个人能力提升、人才培养和团队建设

学术梯队与青年人才培养

团队注重学术梯队建设，通过引进高层次人才和培养青年科研人员，逐步形成多层次、多领域的科研队伍。每年组织定期的学术交流与培训，鼓励青年人才参与科研项目和发表学术论文。此外，通过设立博士后流动站，提供良好的研究条件和发展平台，促进青年人才的成长与发展。

创新团队文化

团队文化强调协作、创新与责任。团队鼓励开放的学术氛围，倡导自由讨论和思想碰撞，以激发创新灵感。同时，团队注重科学伦理和科研诚信，培养成员的责任感和使命感。定期举办团队建设活动和学术沙龙，增强团队凝聚力，提升整体科研能力和水平。通过这样的文化建设，团队致力于营造一个积极向上、共同进步的科研环境。

二、预期考核目标（参照人才引进合同指标填报）

1.教学任务（包括承担核心课程的讲授任务、必修课或选修课讲授任务等）：

根据《华南农业大学新进教师“双证”上讲台暂行实施办法》华南农办〔2020〕58号规定，原则上在报到入职1年内获得“双证”，持证上岗后需完成以下任务：

（1）承担本科生及研究生的教学任务，按照学院教学任务安排，讲授1门本科生课程以上，其中按照培养计划开设并计入学分的课程教学学时数不少于24。

（2）指导本科生创新创业实践或指导实践教学、毕业设计、毕业论文。

2. 科研任务（包括科研项目、经费、论文、论著、奖励、专利等）：

（1）至少主持获得1项国家自然科学基金及1项A类或B类科研项目，主持到位科研经费100万元以上。

（2）以第一作者或通讯作者在本领域顶级期刊（需由学校人才引进考核工作小组同行专家进行论证后确定）上发表具备原创性、前沿性、突破性创新内容的学术论文1篇以上，或T2类期刊上发表2篇以上论文，或A类期刊上至少发表6篇以上论文。

（3）授权发明专利2项，力争专利转化1项。

3. 学科建设任务：

（1）积极参与并推动学科建设。

（2）积极举荐学校第四层次青年才俊以上人才1名以上，并在人才引进过程中发挥有效的推动作用。

（3）在符合国家政策规定和申报条件的情况下，每年申报国家级、省级人才项目。

4. 学科专业人才培养任务（包括培养博士、硕士研究生，指导博士后研究人员、高级访问学者和青年教师等）：

按学校、学院规定培养或协助培养研究生。

5. 其他任务：

完成学院安排的其它工作任务。

三、经费预算

直接费用	经费额	用途说明
(1) 设备费	5	
(2) 材料费	10	
(3) 测试化验加工外协费	30	
(4) 燃料动力费		
(5) 差旅费/会议费/国际合作与交流费	5	
(6) 出版/文献/信息传播/知识产权事务费	5	
(7) 劳务费		
(8) 专家咨询费		
(9) 直接费用其他支出		
合计	55	
其他需说明的情况:		

四、签约各方

管理单位（甲方）： 华南农业大学（盖章）

科研部门负责人（签章）：

胡华双

2024年12月3日

依托学院： 动物科学学院（盖章）

学院负责人：

项目负责人： 肖琛（签字）

2024年12月04日

课题负责人（丙方）：

本人承诺由特定高校学科建设专项（Specific university discipline construction project）经费资助产出的相关科研成果，发表论文等成果将标注项目资助编号“2023B10564003”。

张泽宝（签字）

2024年12月4日

124
合同编号: 2022-XJS-21-001

2022年省级乡村振兴战略专项资金 种业振兴项目

合 同 书

项目名称: 2022年云浮市-广东省新兴县温氏生猪种业创新园

项目管理单位(甲方): 广东省农业农村厅

项目牵头承担单位(乙方): 广东温氏种猪科技有限公司

项目推荐(主管)单位(丙方): 云浮市农业农村局

项目负责人: 吴珍芳 联系电话: 13922715775

项目联系人: 蔡更元 联系电话: 13312828197

广东省农业农村厅制

第一条 为保障 2022 年省级乡村振兴战略专项种业振兴项目顺利实施，按时保质保量完成项目任务，根据《中华人民共和国民法典》、《广东省省级财政专项资金管理办法（试行）》、《广东省财政厅关于安排 2022 年省级乡村振兴战略专项资金种业振兴项目资金的通知》（粤财农[2022]184 号）、《广东省农业农村厅财政专项资金管理办法》等文件有关规定，经甲、乙、丙三方协商一致，签署本合同书。

第二条 甲方的权利义务：本合同履行过程中，甲方有权对乙方项目的实施情况和资金到位、使用情况进行监督、检查，提出改进要求。

第三条 乙方的权利义务：

1. 按财政资金管理规定，对甲方核拨的资金做到专款专用，单独列账，并随时配合甲方进行监督检查；

2. 认真填写本合同书附件 1《项目任务书》，《项目任务书》的内容应与乙方的《项目申报书》保持一致；

3. 严格按照本合同书及合同书附件 1《项目任务书》的要求及时完成项目建设内容，项目实施完成后，按照本合同附件 2 的要求提交验收报告；

4. 按照《广东省农业农村厅财政专项资金管理办法》规定，按季度向甲方、丙方报告项目实施情况、财政资金开支进度等内容；

5. 乙方需保留与所有参与单位的合作实施协议和相关财务凭证，并向甲方备案。

第四条 丙方的权利义务：

1. 为乙方项目实施提供必要的条件保障；

2. 负责对项目承担单位的实施条件、能力以及财务管理规范进行审查，对推荐项目的实施场地、申报资料等进行真实性审核，并监督项目实

施、资金预算执行情况；

3. 受甲方委托或协助甲方完成项目验收等工作，并及时向甲方报告情况。

第五条 本项目资金不得用于以下方向：1. 行政事业单位基本支出；2. 各项奖金、津贴和福利补助；3. 企业担保金和弥补企业亏损；4. 修缮楼堂馆所以及建造职工住宅；5. 弥补单位预算支出缺口和偿还债务；6. 购买交通工具及通讯设备；7. 形成地方政府债务的支出；8. 购买理财产品、发放借款及平衡预算等。

第六条 项目验收。对财政资金投资 500 万元（含）以上（科研项目财政资金在 200 万元（含）以上）的项目，及乙方直接向甲方申报的项目，由甲方负责组织验收；对财政资金投资 500 万元以下（科研项目财政资金 200 万元以下）的项目，甲方自主或委托推荐（主管）单位负责验收，由验收单位向甲方提交验收材料，甲方对验收材料进行审核确认。

第七条 在履行本合同的过程中，如出现相关政策法规重大改变等不可抗力情况，甲方有权对所核拨经费的数量和时间进行相应调整。因非不可抗力因素导致的项目未履行或未履行完毕，或因乙方责任造成项目不能继续开展的，甲方有权终止项目合同，收回尚未使用和使用不符合规定的财政经费。

第八条 在履行本合同的过程中，当事人一方发现可能导致项目整体或部分失败的情形时，应及时通知另一方，并采取适当措施减少损失，没有及时通知并采取适当措施，致使损失扩大的，应当就扩大的损失承担责任。

第九条 实施项目所获得的科技成果（项目成果）归属、成果转让和实施技术成果所产生的经济利益的分享，按照国家和广东省有关规定执行。项目研究成果应向省农业农村厅进行登记、备案，对外发布前应征求省农

业农村厅的意见。

第十条 本合同在履行过程中发生的任何争议，由甲乙丙三方友好协商解决。

项目管理单位(甲方) (盖章)：广东省农业农村厅



法定代表人(或授权代表) (签章)：

刘中国

签订日期：2023年2月23日

项目牵头承担单位(乙方) (盖章)：广东温氏种猪科技有限公司



法定代表人(或授权代表) (签章)：

项目负责人(签章)：吴永华



签订日期：2023年1月19日

乙方推荐(主管)单位(丙方) (盖章)：云浮市农业农村局



法定代表人(或授权代表) (签章)：

李永华

签订日期：2023年1月20日

项目任务书

填写说明

一、本项目任务书由乙方填写。

二、本项目任务书所列内容应实事求是填写，表达要明确、严谨。对填写不符合要求的，或填报内容出现虚报夸大、不切实际的，将退回项目承担单位修改。

三、项目任务书规定的项目考核指标、建设内容和绩效目标必须依据《项目申报书》填写，应遵循明确、量化、可考核的原则，其中技术指标应明确项目完成时达到的关键技术参数及预期可以形成的发明专利、标准、新技术、新产品、新装置、论文、专著等的数量。项目申报指南对项目技术、经济和成果等指标有明确要求的，应符合项目申报指南的要求，相关专项管理办法有特别规定的，应符合相关规定。

四、《项目申报书》及申报指南是本项目任务书填报的重要依据，项目任务书填报不得修改考核指标、绩效目标、资金预算等内容。《项目申报书》、申报指南和本项目任务书将共同作为项目过程管理、综合绩效评价（验收）和监督评估的重要依据。

五、省财政资金支出的预算计划应按照国家及省相关规定执行。

六、表格栏目不够可自行增加。

一、目的及意义

主要说明项目的建设目的、研究价值和意义。

2022年云浮市-广东省新兴县温氏生猪种业创新园旨在建成具有国际影响力的瘦肉型猪种业创新园区。通过该产业园建设，实现我区种业科技创新条件大幅改善、创新能力大幅提升，组建一支长期致力于瘦肉型猪育种的科技创新团队，培育一家品种优势突出、种业技术链条完整、科技含量高、创新实力强、种质资源最丰富、育种群体规模最大、竞争力最强的瘦肉型猪种业龙头企业，成为广东省瘦肉型猪种质资源保护和利用的引领区、现代育种技术和装备的集成区、育种联合攻关的样板区和种业成果转化和推广的示范区，进一步扩大广东省瘦肉型猪育种领先优势，助力广东省率先实现生猪种业振兴，打造民族种业“中国芯”。

二、项目建设内容

详细说明项目建设内容（项目需求或项目建设任务）。

（一）建设研究内容

1. 基因资源挖掘与利用

本部分研究旨在通过生物技术和生物信息学手段鉴别一批影响猪各项经济性状（生长、繁殖、抗病等）的重要功能基因及因果突变位点；并在细胞水平和活体水平进行功能验证，揭示功能基因或因果突变影响种猪特定经济性状的分子调控网络和遗传机理，为优质性状种猪的分子标记筛选、育种芯片优化等提供坚实的理论依据，为进行种猪性状改良和提升提供参考。

2. 瘦肉型猪种质创制

本部分研究拟采用生物育种策略进行瘦肉型猪种质创制。以动物转基因和基因编辑为代表的生物育种技术持续推动传统育种技术的升级，使得瘦肉型猪育种可以实现更为精准的遗传改良，取得更快的遗传进展。通过生物育种技术可以精确快速地获得具有特定优良性状的种猪个体和群体，尤其对传统选择性育种难以选育的性状，如种猪抗病力、肉质等的提升上具有明显优势。同时生物工程育种还可以引入外源优势遗传物质，赋予种猪新的优势遗传性状。本项目的研究目的是制备多种具有优良经济性状的种猪育种材料，并培育相关优势性状种猪育种群体。

本项目围绕肉质、生长、繁殖、毛色、抗病等经济性状进行生物育种开发，创制优势经济性状改良的种猪新材料。针对第一部分挖掘出的特定性状的有效基因位点和已知的基因位点进行基因修饰操作，根据基因型和

表型的关系，进行转基因或基因编辑敲除或替换等基因工程操作，结合体细胞克隆或胚胎注射技术，制备特定基因改良的种猪个体和群体。研究基因修饰种猪的特定性状，明确其具有预计改良的表型，同时对性状改良种猪的生长繁殖等生产性能进行分析，鉴定基因改造带来的其他的副作用。对优势生物育种个体进行群体组建和生物安全评估，研究基因改造的生物安全性，获得农业部生物安全证书。通过特定种猪多个优质家系进行相应的基因改造，培育具有优良经济性状的种猪品种并进行品种审定和转化应用。

3. 瘦肉型猪育种技术研发（基因组选择、基因组选配、智能测定、大数据遗传评估、分子血缘鉴定等）

开展瘦肉型猪基因组选择和精准选配关键共性技术研发，如基因分型技术、遗传评估算法、基因大数据分析技术等。针对不同性状的遗传特点和基因分型数据特点，优化基因组遗传评估模型，开发多种基因组遗传评估新方法，构建性状特异基因组遗传关系矩阵；提高瘦肉型猪基因组选择参考群体规模和质量，建立大规模的主流瘦肉型种猪包括杜洛克猪、长白猪、大白猪的基因组选择参考群体，相较传统育种方法提高低遗传力性状（繁殖力等）、难以度量性状（肉质等）的选种准确性，开发新的高附加值育种性状并应用全基因组选择进行选育。

4. 瘦肉型猪本土化选育和配套系筛选

我国猪肉市场庞大，标准化、规模化养殖逐渐成为主流，低成本和高品质将是未来养猪业企业提高产品竞争力的直接体现。因此，对瘦肉型猪进行本土化选育，培育高瘦肉率和饲料转换率、且高载肉能力、繁殖性能高的本土化品系及相应配套系，显得尤为重要。本项目围绕瘦肉型猪本土

化选育和配套系筛选，打造综合性能优、料肉比性能突出的杜洛克猪本土化品系；繁殖性能优秀、骨架高大的大白猪本土化品系；瘦肉率高、母乳性能优异的长白猪本土化品系，并形成配套效果好、市场经济价值大的商品猪。利用 4-5 年时间，根据各自特点培育各具特色的本土化专门品系，筛选出高竞争力具有自主知识产权的配套系。

5. 瘦肉型猪高效扩繁技术研发（体细胞克隆、低剂量深部输精、冷冻精液、性别控制等）

开展高效率种猪克隆技术，探索影响供体细胞、卵母细胞、克隆胚胎发育的分子机制，并通过小分子药物等调控手段，建立高效的种猪克隆新技术，能够显著提高克隆效率，降低克隆猪生产成本，使之能够大规模产业化应用。并针对瘦肉型克隆猪无效仔高、活仔死亡率高的特点展开研究，尤其是出现频率较高的一些畸形如大舌头，探讨其形成机制并在细胞或者胚胎阶段给予调控，在提高克隆猪存活率的同时，为常规生产解析畸形猪形成提供思路。建立种猪高效精液冷冻技术和低剂量深部输精技术，建立猪冻精技术、设备、试剂为一体的冻精制作体系及冻精实验室。研究高效精液冷冻稀释液配方和冷冻/解冻技术，实现对精液的高活力保存，提升冷冻精液的受胎率，使得冷冻精液的受胎率达到常温液态精液的水平。开发和比较手术法和非手术深部输精的效果，比较不同输精部位（输卵管，子宫体、子宫角）和输精剂量的受精效果，比较冷冻-解冻精子和性控分离精子的输精效果，开发有效的深部输精装置和设备，建立一套切实可行的猪低剂量深部输精技术。开展种猪性别控制技术研究，通过猪精子特异性抗体、小分子药物和基因工程性控技术的研发，建立猪精子体内/外有效性控分离的技术，达到对种猪性别控制的目标。

(二) 科研区平台建设

1. 瘦肉型猪种质资源保护平台建设

按照基因库与活体资源场相结合的原则，在广东温氏种猪科技有限公司组建现代化、智能化的瘦肉型猪种质资源基因库和活体资源保护场，实现冻精与克隆细胞系的长期冷冻保存。新建资源库库容可保存保存各品种、各类别遗传物质 20 万份以上，收集保存 4 个以上国内外引进瘦肉型品种和 2 个以上地方猪或地方猪培育品种的遗传物质，遗传物质的恢复率 50%以上；建成智能化种质资源保存管理数据库，建设具有全国重要影响力的瘦肉型猪种质资源保护利用大数据中心，加强育种相关数据采集、整合、挖掘与育种利用，搭建全省畜禽种质资源保护与开发利用共享交流平台。

2. 瘦肉型猪种质资源鉴评平台建设

持续收集国内外重要的瘦肉型种猪资源，进一步丰富瘦肉型猪良种繁育基地优质品种资源。开展资源评价和利用研究，为瘦肉型种猪资源选择提供技术依据和种质资源。根据瘦肉型种猪品种优势性状进行分类，发掘在特色性状、肉产量性状、肉品质性状、繁殖性状、抗病性状等具有明显优势的瘦肉型种猪品种资源，应用基因组重测序、转录组、ATAC-seq 等多组学技术，结合全基因组关联分析技术，分析瘦肉型种猪品种特异性遗传变异，针对不同的优势性状，挖掘与鉴定决定优势性状的重要基因和分子标记，为猪的精准育种奠定基础。最终，形成自主知识产权“基因源”，定向改良创制优质、抗病、高繁、饲料高效利用的新种质；建立高通量基因型-表型数据库，创建种质资源管理与共享平台。

3. 瘦肉型猪表型组智能测定平台建设

开发种猪个体身份电子化智能识别技术；建立体型体况智能测定技术

并应用示范；建立种猪 CT 智能测定技术，建立屠宰数据-CT 图像数据集；开发三维视频摄像，建立慧眼大数据自动分析平台和表单，实现表型数据的自动采集、实时传输、长期保存和智能分析；平台建成后将具备以下功能：（1）实现种猪育种测定数据实时自动上传；（2）实现种猪体尺数据非接触式智能测定；（3）实现产肉性状和肉质性状数据活体采集和评估；（4）实现种猪智能个体识别；（5）提高育种效率和准确性。

4. 瘦肉型猪基因组选择平台建设

建立高通量、全自动基因芯片检测平台，高效整合自动化基因组核酸提取系统、自动化移液工作站、核酸质控系统、基因芯片扫描仪以及高性能服务器等设备，实现年检测量可达 50 万份种猪样本的国内先进的瘦肉型猪基因组选择平台。设计和定制适用于瘦肉型种猪品系的育种专用功能基因芯片，有效提高瘦肉型猪选择和选配准确性。不断提升瘦肉型猪基因组选择参考群体规模和质量并加快推广应用，提高低遗传力性状（繁殖力等）、难以度量性状（肉质等）的育种效率，开发新的高附加值育种性状并应用。挖掘、鉴定与猪重要经济性状相关、具有重要育种价值的功能基因、分子标记、功能元件，重点揭示瘦肉型饲料效率性状、繁殖性状以及肉质性状的形成和调控机制，为优质瘦肉型猪选育奠定理论基础。

5. 瘦肉型猪大数据遗传评估平台建设

收集、保存瘦肉型猪生长、繁殖等重要经济性状表型和基因型测定数据，建设瘦肉型猪育种大数据库。建设种猪大数据基因组遗传评估系统，提升数据处理服务器功能，实现从表型到基因组原始数据的汇总、挖掘、分析、整合全流程自动化分析。持续优化“慧眼大数据”育种分析方法，开发“育种一张表”自动化表单，加强育种大数据的挖掘和应用。统一性

能测定和遗传评估方法，建立各育种场间遗传联系，实现跨场间联合遗传评估。不断优化种猪遗传评估模型，估计群体遗传参数，提高遗传评估的准确性。

6. 瘦肉型猪转基因与基因编辑平台建设

增添二氧化碳培养箱、PCR 仪、液氮罐、荧光倒置显微镜、离心机等实验设备，拓展和完善分子实验室、负压细胞房、正压细胞房、生物育种样本资源库功能，提高平台科研效率。进一步增强候选基因、抗病基因或病毒受体的挖掘能力；增强转基因技术研究和转基因猪育种材料创制能力；增强基因编辑技术研究和基因编辑猪育种材料创制能力；完善 DNA、RNA、蛋白质，病毒等分子检测平台；提高对各类生物育种新材料的生物安全评价的能力。建成集技术研发、种质资源创制、成果转化、人才培养、科普宣传为一体的综合性技术平台。

7. 瘦肉型猪体细胞克隆平台建设

增添万级洁净细胞室和胚胎室显微操作仪、倒置荧光显微镜、体视镜、融合仪、胚胎实时观测仪等设备，提高胚胎生产能力，该平台建成后，具备以下功能：（1）为高校联合培养生物工程领域研究生；（2）开展表观遗传修饰、胚胎发育调控机制等前沿研究；（3）进行瘦肉型种猪克隆，形成特级种猪群体，快速扩大特级种猪优秀基因的覆盖面；（4）为基因编辑、转基因技术提供克隆动物生产，加快基因编辑技术成果转化；（5）为养殖行业提供优秀种源的建系冻存、胚胎冷冻保存、特级种猪克隆等技术服务。最终形成体细胞克隆方面的一个产、学、研技术平台。

8. 瘦肉型猪肉质检测平台建设

升级肉质检测平台，购置肉质 pH 计、胴体直测仪、嫩度计、肉色仪、

Foodscan、HGP7、系水力仪等高端仪器设备，研究肉质测定技术，深挖肉质测定遗传价值，建成国内先进的瘦肉型猪肉质检测示范平台，构建肉质性状全基因组选择参考群体。与屠宰场相结合，建设与屠宰线相配套的大规模肉质分析能力。

(三) 基地建设

1. 瘦肉型猪科研基地建设

以新兴县新城镇为中心，依托广东温氏种猪科技有限公司总部的实验室和平台，构筑具有国际竞争力的种猪育种研发平台。建设内容包括改造和升级种质资源库、种质资源评鉴、全基因组选择、体细胞克隆、转基因、基因编辑实验平台，升级物联网和大数据分析平台和服务器，建设和升级育种科技团队。可满足各类育种新材料的生产性能测定、营养代谢试验、抗病性效果评价等设备的建设。

2. 瘦肉型猪育种基地建设

以新兴县水台原种场作为长白 W57、大白猪 W64 资源场、核心育种基地以及克隆猪、克隆猪创制基地，生物育种新材料创制基地。以温氏集团下属所有长白和大白种猪场为主体，开展联合遗传评估，场间通过精液进行遗传物质传递，建立规模化、协同化、智能化育种体系，选育长速快、饲料转化效率低、瘦肉率高、泌乳性能优秀的长白新品系，和长速快、繁殖力高、母性好、体型优秀的大白新品系，创立种猪新品系选育的典范。建设内容包括优秀种质资源的引进，场区内测定站猪舍改造，智能测定设备、料肉比测定系统、生物安全防控设备、疾病检测和净化设备等升级改造，用于护理克隆猪的仔猪护理平台的安装和维护。

3. 瘦肉型猪扩繁基地建设

以新兴县水台扩繁一场、水台扩繁二场作为长白 W57、大白 W64 扩繁群生产推广基地和配套系生产基地，场内配备用于自循环的模拟核心群，精液来自其他核心场公猪站，或利用克隆技术将最优秀种公猪的遗传基因进行全面覆盖。建设内容包括生物安全防护设施设备、猪繁殖技术装备、精液检测设备、疾病检测和净化设备、物联网设备的升级改造。

4. 瘦肉型猪成果展示基地建设

以新研究院展示厅和新兴县东成试验场作为品种选育和科研成果展示基地，通过升级改造现有设备，建设一个质量过硬、行业有名的核心示范基地，加快技术应用推广，全力助推行业向前发展。建设内容包括示范场和栏舍的改造升级，品种和技术展示设备，物联网监控设备的安装、改造和升级等。

（四）团队组建

依托国家生猪种业工程技术研究中心，以广东温氏种猪科技有限公司育种部及各专业育种技术实验室、科隆公司、水台原种场、水台扩繁场、东成试验场的研发人员和生产技术人员为核心，以华南农业大学生猪种业相关的师资力量为技术支撑。拟组建由国家级创新领军人才、省级领军人才、优秀青年后备人才和育种技术人员为主体的产学研结合的高水平人才梯队 300 余人，其中技术支撑团队里有国家“万人计划”科技创新领军人才、广东省生猪产业技术体系首席专家 1 人，广东省现代农业产业技术体系生猪创新团队岗位专家 1 名，正高职称 20 余人，博士学历的核心研究骨干 50 余人，博士和硕士研究生 50 余人，育种技术人员 150 余人。种业创新园人才团队年龄结构合理、梯度明显、专业互补、人岗匹配良好。人才团队研究领域涵盖遗传育种、生物信息、基因工程、繁殖技术、疾病防控、

种猪生产管理、智慧农业等，通过学科交叉融合，助力瘦肉型猪种业的创新与突破。

技术平台拟组建研究团队如下：①瘦肉型猪种质资源保护平台，博士 1 人，硕士 1 人，一般技术人员 2 名；②瘦肉型猪种质资源鉴评平台，副高级 1 名，博士后 1 名，博士研究生 3 名，研究生 3 名；③瘦肉型猪表型组智能测定平台，正高级 1 人，副高级 1 人，博士 1 人，硕士 2 人。④瘦肉型猪基因组选择平台，博士 2 人，硕士 3 人，一般技术人员 3 人；⑤瘦肉型猪大数据遗传评估平台，博士 1 人，信息技术人员 4 人；⑥瘦肉型猪转基因与基因编辑团队，博士 2 人、硕士 2 人，研究生 5 人；⑦瘦肉型猪体细胞克隆平台，博士 1 人，硕士 1 人，其他技术人员 12 人；⑧瘦肉型猪肉质检测平台建设有博士 1 人，硕士 1 人，实验员 3 人。⑨示范基地基层种业技术人员 150 余人。

备注：项目建设内容（项目需求或项目建设任务）按《项目申报书》内容填写。

三、项目绩效目标

主要说明项目实施后，预期达到的目标和产生的效果，相关表述应量化。

（一）种业创新园预期目标

到 2024 年底，基本建成具有国际影响力的瘦肉型猪种业创新园区，种业科技创新条件大幅改善、创新能力大幅提升，组建一支长期致力于瘦肉型猪育种的科技创新团队，培育一家品种优势突出、种业技术链条完整、科技含量高、创新实力强、种质资源最丰富、育种群体规模最大、竞争力最强的瘦肉型猪种业龙头企业，成为广东省瘦肉型猪种质资源保护和利用的引领区、现代育种技术和装备的集成区、育种联合攻关的样板区和种业成果转化和推广的示范区，打造民族种业“中国芯”，带动引领全国畜牧业加速转型升级。

（二）考核指标

1. 育种技术的先进性与引领性

主要猪育种核心关键技术保持与国内外前沿技术同步，其中在基于低深度重测序的基因组选择、节粮环保猪创制、成年猪体细胞克隆效率、基于三维重构的智能测定等方面建立领先优势。建立瘦肉型猪表型智能测定技术，智能化测定效率预期提高 3 倍以上；建立全基因组选择高精度液相芯片，预期育种成本降低 20%以上，育种效率提高 20%以上；提高低遗传力性状（繁殖力等）、难以度量性状（肉质等）的选种准确性 20%以上；建立瘦肉型种质资源鉴评技术，预期初步解析生猪的生长、繁殖、抗逆等重要经济性状的遗传调控网络 3 个以上，构建生猪多组学数据库 1 个以上，挖

掘筛选与重要经济性状相关的功能基因和有重要育种价值的分子标记50个以上。设计基因组选配技术和软件，提高内容丰富和运算速度、操作友好性，预期能支撑上千万种猪同时进行遗传评估。建立的遗传评估系统能够融合基因组信息，自动、高效、准确地分析种猪的遗传性能，为现场育种工作提供充分依据。建立成熟高效转基因和基因编辑技术，突破大片段基因定点技术，用其研发瘦肉型猪育种新材料达到国内首创。建成全国最大的瘦肉型猪克隆平台和全国最大克隆猪推广应用基地，窝均克隆猪总仔达6头以上。

2. 创制优异种质资源数

创制瘦肉型抗蓝耳病猪育种新材料10个以上，创制瘦肉型节粮环保猪育种新材料1个以上，创制瘦肉型优质猪育种新材料1个以上，创制黑毛瘦肉型猪育种新材料1个，创制大体型瘦肉型猪育种新材料1个。

3. 育成品种水平与数量

重点培育1个大体型、长速快、泌乳性能好的长白猪专门化品系，配套系中做母本父系（第一父本），核心群总产仔达到14.5头以上，公猪校正115kg天龄在165天以内。培育1个高繁、快长、节粮型的新配套组合，终端母本PSY达到27头以上，肉猪料肉比达2.55以内，达到国家畜禽新品种（配套系）审定条件。

4. 研发基地规模、良种繁育推广能力与水平

建成后创业园科研区总面积达3000平方米，育种基地规模达1400余亩，基地内各品系祖代纯种猪基础群8000头以上，每年可提供合格纯种母猪5万头以上，可推广应用父母代种母猪30万头以上，覆盖瘦肉型商品猪

500 万头以上。

5. 预期效益（经济效益、社会效益、生态效益）

经济效益：种业创新园建成后，其技术和平台可服务于广东温氏种猪科技有限公司在全国的 31 个育种基地，每年提供祖代纯种猪规模达 25 万头，推广商品代肉猪规模 3000 万头以上，预期产值达 500 亿元，相当于广东省全年 70% 以上瘦肉型猪消费量，对我省猪肉稳产保供具有举足轻重的作用。

社会效益：种业创新园建成后，其技术和平台可服务于温氏集团以外的其他种猪育种企业，为全行业提供基因检测、种猪克隆、健康检测、技术培训等社会化服务，还可向社会提供大量高质量的瘦肉型种猪，建立年产 5000 万头以上商品猪的种源保障体系，带动我国生猪育种技术进步，助力广东率先在生猪育种领域实现种业振兴。

生态效益：种业创新园培育瘦肉型新品种（配套系）生产的肉猪料肉比下降 0.2，达 120kg 上市体重，可节约 24 kg 饲料/头，若将其推广 3600 万头肉猪规模，预计每年可减少饲料消耗 8.6 亿公斤，节省饲料成本 30 亿元。预期每年可减少 4320 吨磷污染排放，每年减少 18000 吨氮污染排放。

四、项目进度安排

详细说明各阶段的工作内容和时间安排情况。

(一) 进度安排 (详细到季度)

时间	建设内容	建设目标
2023 年一 季度	1. 基因资源挖掘与利用; 2. 瘦肉型猪种质创制; 3. 瘦肉型猪育种技术研发; 瘦肉型猪本土化选育与配套系筛选; 6. 瘦肉型猪高效扩繁技术研究; 7. 完成资源库、资源鉴评、大数据遗传评估、表型组智能测定、肉质检测、生物育种平台, 猪体细胞克隆 7 大科研平台的归并、升级改造和统一管理; 8. 完成种源创制基地、育种示范基地, 成果转化展示基地基地的规划设计; 9. 制订生猪育种人才团队建设和培养计划;	1. 筛选繁殖、抗病性状的功能基因一批; 2. 进行基因修饰猪体细胞的筛选; 3. 构建基因组选择参考群体, 并开展分型方法研究和基因组选配系统开发; 4. 组建健康、性能突出的瘦肉型猪专门化父母系核心育种群, 制订选育目标和选育方案; 5. 进行体细胞克隆技术和冻精技术研究。6. 生猪育种人才团队得到加强, 储备后备人才; 7. 完成 7 大平台, 试验基地和示范基地的规划和设计。8. 完成育种团队和技术团队组团;
2023 年二 季度	1. 基因资源挖掘与利用; 2. 瘦肉型猪种质创制; 3. 瘦肉型猪育种技术研发; 4. 瘦肉型猪本土化选育与配套系筛选; 5. 瘦肉型猪高效扩繁技术研究。 6. 对 7 大平台开工改造, 完成 7 大试验平台仪器设备的招投标和购买; 7. 开展试验平台和示范基地, 展示基地工程装备升级改造。	1. 筛选繁殖、生长, 产肉、抗病性状的功能基因一批; 2. 进行基因修饰猪体细胞的筛选并进行克隆猪制备; 3. 组建基因组选择参考群体, 进行基因组选配系统开发; 4. 通过表型测定结合基因组选择等方法对专门化品系系核心群进行持续选育; 5. 进行体细胞克隆技术和冻精技术研究。7. 试验基地和示范基地建设准备工作。

2023 年三 季度	1. 基因资源挖掘与利用；2. 瘦肉型猪种质创制；3. 瘦肉型猪育种技术研发；4. 瘦肉型猪本土化选育与配套系筛选；5. 瘦肉型猪高效扩繁技术研究。6. 完成三维重构、电子标识等智能测定技术平台和场地搭建7. 完成新一代基于代深度全基因组测序的种猪全基因组选择技术研发，初步建设基因型数据库，8. 基本完成对7大平台开工改造。	进行特性性状基因功能研究与应用；2. 进行基因修饰猪体细胞的筛选并进行克隆猪制备；3. 完成基因组选择及遗传评估系统主体内容开发；4. 通过表型测定结合基因组选择等方法对专门化品系核心群进行持续选育；开展配套系筛选；5. 进行低剂量深部输精技术研究和性控技术研究。6. 收集智能测定基础数据收集和总结，建设表型数据库7.6. 完成7大科研平台设备交付；
2023 年四 季度	1. 基因资源挖掘与利用；2. 瘦肉型猪种质创制；3. 瘦肉型猪育种技术研发；4. 瘦肉型猪本土化选育与配套系筛选；5. 瘦肉型猪高效扩繁技术研究。9. 完成对各试验平台，育种基地，创制基地，推广基地，展示基地，建设工程验收。	1. 进行特性性状基因功能研究与应用；2. 进行基因编辑克隆猪制备与功能研究；3. 基因组选择进行研发和扩大应用；4. 通过表型测定结合基因组选择等方法对专门化品系核心群进行持续选育；进行配套系筛选；5. 进行低剂量深部输精技术研究和性控技术研究。6. 完成各试验平台工程验收，各基地工程建设，科研和育种条件得到大幅改善。
2024 年一 季度	1. 基因资源挖掘与利用；2. 瘦肉型猪种质创制；3. 瘦肉型猪育种技术研发；4. 瘦肉型猪本土化选育与配套系筛选；5. 瘦肉型猪高效扩繁技术研究。6. 完成全部仪器设备的调试运行；7. 完成对试验基地和示范基地投产和新育种技术应用；	1. 进行特性性状基因功能研究与应用；2. 进行基因编辑克隆猪制备与功能研究；3. 进行基因组选择和选配系统技术开发；4. 通过表型测定结合基因组选择等方法对专门化品系核心群进行持续选育；进行配套系筛选；5. 进行性控技术研究。6. 科研试验条件得到进一步改善；7. 生猪育种基层技术人才培养开始产出；8. 完成6大科研平台、试验基地和示范基地规章制度建立和完善

2024 年二 季度	1. 瘦肉型猪种质创制；2. 瘦肉型猪育种技术研发；3. 瘦肉型猪本土化选育与配套系筛选；4. 瘦肉型猪高效扩繁技术研究；5. 建成种猪表型组智能测定应用平台；	1. 进行基因编辑克隆猪制备与功能研究，进行群体组建和生物安全评价；2. 进行基因组选择技术研发和扩大应用；3. 通过表型测定结合基因组选择等方法对专门化品系核心群进行持续选育；进行配套系筛选；4. 进行性控技术研究。5. 7 大科研平台、试验基地和示范基地运营顺利；
2024 年三 季度	1. 瘦肉型猪种质创制；2. 瘦肉型猪育种技术研发；3. 瘦肉型猪本土化选育与配套系筛选；4. 瘦肉型猪高效扩繁技术研究。5. 开始与高效联合培养研究生，建立 7 大科研平台开放共享机制。	1. 进行基因编辑克隆猪制备与功能研究，进行群体组建和生物安全评价；2. 进行基因组选择育种技术研发和应用推广；3. 进行配套系筛选；4. 进行性控技术研究。5. 6 大平台开放共享，联合培养研究生 20 名以上，培养生产技术人员 100 人以上。
2024 年四 季度	1. 瘦肉型猪种质创制；2. 瘦肉型猪育种技术研发；3. 瘦肉型猪本土化选育与配套系筛选；4. 瘦肉型猪高效扩繁技术研究。5. 整合形成“表型-基因型”数据库。6. 瘦肉型猪（胚胎系）种业创新园平台开放共享；7. 形成管理制度和技术规范汇编	1. 进行基因编辑克隆猪制备与功能研究，进行群体组建和生物安全评价；2. 进行基因组选择育种技术研发和应用推广；3. 进行新品系和新配套系推广用用；4. 进行性控技术研究。5. 全面建成瘦肉型猪（胚胎系）种业创新园平台；6. 生猪育种整体技术水平达到国内领先，国际先进水平，部分技术达到国际领先。7. 达到创新园全部建设目标。

备注：项目绩效目标按《项目申报书》内容填写。

五、项目主要合作、参与单位(含牵头承担单位)

单位名称	单位性质	统一社会信用代码	通讯地址
广东温氏种猪科技有限公司	有限责任公司	91445321MA4UM4FJ3U	广东省云浮市新兴县新兴县新城镇东堤北路7温氏科技园
华南农业大学	事业单位	124400004554165634	广东省广州市天河区五山路

六、项目组主要成员(含项目负责人)

姓名	性别	身份证号	单位	职称/职务	电话
吴珍芳	男	420111197003145016	广东温氏种猪科技有限公司	博士/教授	13922715775
蔡更元	男	420203197004154315	华南农业大学	博士/研究员	13312828197
张献伟	男	410222198308193550	广东温氏种猪科技有限公司	博士/高级畜牧师	15914225968
董林松	男	371081198602094075	广东温氏种猪科	博士/无	15063189125

			技有限公司		
杨化强	男	413029198109190417	华南农业大学	博士/副 研究员	15521148340
叶健	男	342625199208110593	广东温氏种猪科 技有限公司	博士/无	18927157751
石俊松	男	411325191805060710	广东温氏种猪科 技有限公司	博士/高 级畜牧 师	13672525779
谈成	男	340881198808125233	广东温氏种猪科 技有限公司	博士/无	18810516382
莫健新	男	441224199307106359	广东温氏种猪科 技有限公司	硕士/无	18302005360
张健	女	342622198706284664	广东温氏种猪科 技有限公司	硕士/助 理研究 员	15013279492
王凯	男	340123199002062598	广东温氏种猪科 技有限公司	硕士/中 级畜牧 师	18819492878
向有为	男	430626198003218519	广东温氏种猪科 技有限公司	硕士/无	13143683913
刘珍云	男	430224197506275196	广东温氏种猪科	本科/无	13927147258

			技有限公司		
贺晓燕	女	432522198506054588	广东温氏种猪科 技有限公司	博士/无	13560378628
王声会	男	342601198104045314	广东温氏种猪科 技有限公司	硕士/无	1382685 8575
陈小强	男	513029197510193533	广东温氏种猪科 技有限公司	硕士/无	1382686 7801
彭妙莲	女	44128219820205044	广东温氏种猪科 技有限公司	硕士/无	15807665843
李紫聪	男	440781197903068517	华南农业大学	博士	15013065924
杨杰	男	430721198706211333	华南农业大学	博士	18602020087
洪林君	男	362334198605037114	华南农业大学	博士	13622230561
顾婷	女	420101198611287028	华南农业大学	博士	13662304975
郑恩琴	女	441322198108016041	华南农业大学	硕士	13826029036
丁荣荣	男	320682199204295430	华南农业大学	博士	19879303695
黄思秀	女	360424198208150042	华南农业大学	硕士	13826094680
周荣	女	420881198304213723	广东温氏种猪科 技有限公司	硕士/无	18927164621
吴丹	女	440811198907250923	广东温氏种猪科	硕士/中	15013019352

			技有限公司	级畜牧 师	
麦然标	男	445122198604185613	广东温氏种猪科 技有限公司	本科/无	15811771014
温淑贤	女	445321198910071627	广东温氏种猪科 技有限公司	本科/无	18316303891
曾海玉	女	432522197707122423	广东温氏种猪科 技有限公司	硕士/无	13826703848
纪红美	女	610323198502282644	广东温氏种猪科 技有限公司	本科/无	15360363767
罗绿花	女	440229198711212221	广东温氏种猪科 技有限公司	本科/无	13672615378
苏巧云	女	445321199506271920	广东温氏种猪科 技有限公司	本科/无	13437867576
宋长绪	男	412901196505173013	华南农业大学	博士	15914225968
徐铮	男	430626197912280032	华南农业大学	硕士	13602496031
王珊珊	女	421381199010136441	华南农业大学	博士	15907163745
曹露	女	500383199105240403	华南农业大学	博士	18310978172
全建平	男	42028119920612805X	华南农业大学	博士	15102030907
邢徐鹏	男	410621199002181119	华南农业大学	博士	18675940801
赵尧露	女	22052319920803014X	华南农业大学	博士	18643565285

七、资金使用预算

主要说明资金使用的范围或方向及资金使用进度安排。

(一) 项目资金使用范围和方向

资金来源		金额（万元）	备注			
一、申请省财政补助资金		1000	广东温氏种猪科技有限公司 800 万元，华南农业大学 200 万元。			
二、配套资金		0				
三、自筹资金		1000	广东温氏种猪科技有限公司 1000 万元			
合计		2000				
项目资金概算						
序号	费用名称	资金用途说明（概算）	资金来源（万元）			合计
			省级财政	配套资金	自筹资金	
一	农业设施	用于测定、防疫设施建设或改造	200.00	0.00	300.00	500.00
1	种猪核心群测定中心建造	生产性能测定设备	150.00	0.00	200.00	350.00
2	物资及车辆洗消中心扩建	消毒设备	50.00	0.00	100.00	150.00
二	科技研发	技术攻关和研发平台建设费用	634.80	0.00	0.00	634.80
1	试验材料费	相关研究试验试剂、耗材采购	193.99	0.00	0.00	193.99
2	出版印刷费	出版文献、信息传播、知识产权事务费	12.50			12.50
3	会议/差旅/国际合作交流费	会议/差旅/国际合作交流费	28.31			28.31

4	技术研发相关 测试化验费	测试化验费	50.00	0.00	0.00	50.00
5	育种软件智能 测定软件研发	软件开发	50.00	0.00	0.00	50.00
6	克隆猪研发和 示范平台	显微操作和手术设 备	100.00	0.00	0.00	100.00
7	资源鉴评平台	基因扩增设备、耗 材	50.00	0.00	0.00	50.00
8	肉质检测平台	肉质检测设备、耗 材	50.00	0.00	0.00	50.00
9	生物育种平台	细胞培养设备、耗 材	50.00	0.00	0.00	50.00
10	数据分析服务 器平台	服务器	50.00	0.00	0.00	50.00
三	示范推广	平台建设和育种示 范基地	110.00	0.00	700.00	810.00
1	示范区、样板区 展示平台建设	物联网设备、展示 牌、展示物料等	50.00	0.00	0.00	50.00
2	示范基地建设	智能化升级改造、 物联网、空气过滤 等相关设备设施、 升级、改造、修缮	60.00	0.00	700.00	760.00
四	劳务费	相关研究生、临时 雇用人员	55.20	0.00	0.00	55.20
五	其他业务相关 费用		0.00		0.00	0.00

(二) 省级财政资金测算依据

1. 设备费，合计 610.00 万元

(1) 电子饲喂站：预算 90.00 万元。计划采购 20 套，每套 4.5 万元，用于精准跟踪生长猪或母猪测定、泌乳或者怀孕期种猪耗料。

(2) B 超测定设备：预算 45.00 万元。计划购入 3 套，15 万元/套，可以准确测定种猪瘦肉率。

(3) 种猪测定相关小型设备：预算 15.00 万元。主要用于 A 超测定设备、笼称、便携式数据收集、处理和分析设备，为测定辅助设备。

(4) 车辆冲洗、烘干设备：预算 50.00 万元。计划购入冲洗机 5 台，1 万元/台；购入烘干设备 1 套，45 万元/套，为物资及车辆洗消中心核心设备。

(5) 计算服务器：预算 50.00 万元，主要用于数据分析服务器平台搭建。

(6) 肉质测定设备：预算 40.00 万元，计划购入肉色仪、系水力测定仪、肉质嫩度测定仪、Foodscan 等设备用于肉质测定平台搭建。

(7) 克隆猪培育设备：预算 110 万元，用于除土建外的培育栋舍栏架、饲喂器、仔猪补奶设备、仔猪保温箱、清粪系统、环控系统等相关设备购买。

(8) 分子检测相关设备：预算 100.00 万元。主要用于 DNA 自动提取设备、全自动热封膜机、PCR 仪、振荡混匀仪、三维冷冻研磨仪、垂直板电泳槽、移液器、电子分析天平、普通离心机、多功能脱色摇床（圆周）等，为相关分子实验常用设备，为建设资源鉴评平台和其他公共平台所必须。

(9) 数据分析服务器平台建设费用，50 万

应种猪遗传评估需求，预采购高性能服务器 4 台，单价 12.5 万/台，合计 50 万。

(10) 示范基地升级改造，60.00 万元。

主要用于智能化升级改造、物联网、空气过滤，消毒等相关设备设施、升级、改造、修缮等，共 60.00 万。

2. 业务费, 合计 334.80 万元

(1) 材料费: 预算 193.99 万元。

①基因芯片检测费: 预算 120.00 万元。用于购买定制种猪育种芯片及抽提试剂, 用于专门化品系基因组选择参考群体及备选群体的基因型分型。完成 2500 头基因组选择参考群体及备选群体构建, 共计 1.25 万头, 每头猪基因芯片费 80 元/头, 抽提试剂 16 元/头, 共计 120.00 万元;

②分子生物学研究各种试剂盒: 预算 73.99 万元。用于克隆猪制备, 购买分子生物学研究相关的各种试剂盒, 如 DNA 提取试剂盒、RNA 提取试剂盒、反转录试剂盒、荧光定量 PCR 试剂盒、Southern 杂交试剂盒、Northern 杂交试剂盒、蛋白抽提试剂盒、ELISA 检测试剂盒等。

(2) 测试化验加工费: 预算 50.00 万元。

用于基因测序、基因优化合成、宏基因组测序、肉成分分析, 饲料成分分析、粪成分分析、组织切片、血液检测等, 25 万/年, 2 年共 50 万元。

(3) 出版文献、信息传播、知识产权事务费: 预算 12.50 万元。

申请国家发明专利 5 件, 按 5000 元/件, 2.50 万元。

发表论文 4 篇, 按平均每篇文章 15000 元计算, 计 6 万元。

科技查新、制作各类视频、展板、技术资料手册等费用, 以 4 万元预计。

(4) 会议/差旅/国际合作交流费: 预算 28.31 万元。

①样品采集、实验等差旅费: 预算 12 万元。按研究人员共 400 人次往返于试验基地采样、实验跟踪、屠宰试验等计算, 平均每人次 300 元, 包括往返交通费、住宿费、伙食和工杂费补助等, 计 12 万元。

②交流、调研、考察等差旅费: 预算 5.51 万元。项目骨干课题间开展

交流、调研、考察等发生的差旅费用，按 5 个课题育种人员和科研单位研究人员共 29 人次参与，平均每人次 1900 元，包括往返交通费、会务费、住宿费、伙食和工杂费补助等，计 5.51 万元。

③参加各种会议差旅费：预算 7.60 万元。项目骨干参加国内与项目相关的学术讨论会，参加项目、课题启动会、年度会议费、总结交流会等差旅费，按 5 个课题育种人员和科研单位研究人员共 20 人次参与，平均每人次 3800 元，包括往返交通费、会务费、住宿费、伙食和工杂费补助等，计 7.60 万元。

④项目实施会议费：预算 3.20 万元。项目牵头单位拟组织启动会，约 40 人次参会，每人费用约 400 元/天（包含材料费、会场租金和工作餐等），会期 2 天，计 3.2 万元。

（5）软件开发，共 50 万元

①软件开发服务：预算 50.00 万元。育种管理系统完善和深度开发，包括，基因组与表型组分类管理、储存和涉密使用管理软件，数据开放管理平台软件，个体从采样到检测、分析、使用管理软件等，开发服务费预计 25 万元。

品系选育过程中电子耳牌应用场景等软件开发费用，完成 2 个应用场景的开发，费用预计 25 万元。

3. 劳务费，共 55.20 万

园区内科研平台预计联合培养研究生 12 人/年，2 年共培养 24 人次，科研劳务费 1500 元/月/人，合计 43.20 万元。

聘请猪场员工协助开展饲养、采样和试验辅助工作，支付加班劳务费 500 元/人/月，按项目累计 10 人次/年计算，预计 6 万/年，2 年共需劳务

费合计 12.00 万元。

4. 其他来源资金

本年度预算自筹资金 1000.00 万元，主要用于：

(1) 农业相关设施土建

预算 300 万元，用于种猪核心群测定中心建造、物资及车辆洗消中心扩建、克隆猪手术室及示范展示区土建工程。

(2) 育种基地建设费用

预算 700.00 万元。用于园区内育种基地栏舍修缮、饲料设备升级改造费，环保设施维护等费用，350 万/年，2 年共 700 万。

(三) 项目资金使用进度

广东省新兴县温氏生猪种业创新园分 2 年建设完成，其中 2022 年四季度完成预算总额 5%，预计支出金额 100 万；2023 年预计完成预算总额 70%，预算金额 1400 万，2014 年预计完成预算总额 25%，预算金额 500 万。

如经费未按期下拨，将 2022 年四季度预算累加到 2023 年预算。

八、保障措施

说明围绕完成项目任务、目标所要采取的具体措施。

1. 加强组织领导

为保障广东省新兴县温氏生猪种业创新园建设工作顺利进行，加强组织领导，实行园长责任制，并成立温氏种猪公司-华大农大省瘦肉猪型种业创新园工作领导小组，由牵头主体广东温氏种猪科技有限公司董事长吴珍芳任园长及领导小组组长，华南农业大学国家生猪种业工程技术中心副主任蔡更元任和广东温氏种猪科技有限公司副总裁向有为任副组长，广东温氏种猪科技有限公司、温氏食品集团股份有限公司研究院、广东温氏种猪科技有限公司下属在新兴县内的各个种猪场等种业创新园各实施主体及参与单位负责人刘珍云、张献伟、石俊松、叶健、董林松、谈成、范海峰、王凯为小组成员。领导小组负责统筹协调推进种业创新园建设管理工作，及时解决创新园建设工作中遇到的重大问题。

2. 人员支持和保障

在人才引进方面，给予引进人才一定的引进费用补贴和政策倾斜。充分利用现有科技人员，广东温氏种猪科技有限公司拥有育种方向博士12人，华南农业大学有大批相关的科技人员，为项目的顺利实施提供了可靠的保障。两单位给予所有项目参与人员特殊支持和激励政策，保证科技人员能够集中精力投入到项目建设和研究工作，并将本项目实施绩效纳入年度考核和岗位考核。出台科研激励政策，对做出优秀成绩的人员给予奖励补贴、

专家称号、经费支持等。

3. 配套资金支持和保障

参与种业创新园建设的广东温氏种猪科技有限公司承诺提供 1:1 以上的配套资金支持，负责提供配套资金的企业经济实力强，经营状况良好，资产质量好，资产负债率低，近几年的科技投入均在 1.2 亿元/年以上，完全有足够的能力支持和保障。因此，配套资金可以及时、足额到位。

4. 基地支持与保障

本次项目中，广东温氏种猪科技有限公司划拨 4 个种猪场参与基地建设，在育种核心群、扩繁群、成果展示等方面给予支持，公司总部克隆实验室和集团研究院所有育种相关实验室均参与平台建设。种猪场配套有测定站、料肉比测定系统、B 型超声波背膘测定仪等育种设备，科研平台配套有 SNP 多态分析系统、芯片分析仪、全自动生化分析仪、流式细胞仪、显微操作仪、计算机服务器等大型设备。育种和试验基地采用现代化的高效养猪工艺模式，生物安全条件好，养猪和育种设备设施条件完备，配备最优秀的干部员工，可以保证育种和试验基础正常运转，同时也保证项目顺利开展。

5. 技术和平台保障

广东温氏种猪科技有限公司在全基因组选择、体细胞克隆、转基因、基因编辑、大数据遗传评估、种质资源保存等多个种猪育种领域均已具备国内甚至世界领先的技术。公司建有国内唯一的国家生猪种业工程技术研究中心，还建有畜禽育种国家地方联合工程研究中心、岭南现代农业广东

省实验室、广东省农业动物基因组学与分子育种重点实验室、广东省畜禽种质资源库等国家级、省级科技创新平台。作为牵头实施主体，建有全国首个国家级畜禽种业产业园，承担建设了广东省畜禽种业功能性产业园。

6. 育种群体保障

广东温氏种猪科技有限公司在全国 8 个地区设有种猪分公司，另有 5 大区域公猪站，拥有杜洛克、长白、大白、皮特兰、温氏大黑猪五个品种、十二个专门化品系，共有育种核心群 1 万头，扩繁群 3.9 万头。品种种类齐全，不同选育方向的专门化品系众多，满足各类市场需求，可为本项目的实施提供充足的、特点各异的种质资源。

联合申报项目协议书

甲方：广东温氏种猪科技有限公司

乙方：华南农业大学

经合作各方友好协商，决定联合申报2022-2023年乡村振兴战略专项省级种业振兴行动项目种业创新园项目，项目名称：2022年云浮市-广东省新兴县温氏生猪种业创新园。并达成如下合作协议：

第一条：项目研究工作详细分工：

甲方：作为项目主持单位，负责项目统筹协调和组织实施，承担平台建设、基地建设、应用技术研发和集成、种质创制、品种培育等建设任务。

乙方：作为项目参加单位，为项目实施提供技术支撑，承担种质资源保存、遗传机制解析、功能基因挖掘、分子标记开发，以及全基因组选择、体细胞克隆、基因编辑等前沿技术迭代升级等建设任务。

第二条：经费分配：

1、如果本申报项目获批立项，按政府下达的资助经费，合作各方同意此经费分别按政府资助经费的甲方：80%，乙方：20%进行分配。

2、甲方在收到项目下达的资助经费后的一个月内将经费按上述约定支付到合作方指定帐户。

第三条：知识产权归属：

1. 项目实施过程中所产生的知识产权：①各方独立完成的所有权归各自所有，甲方具有知情权、使用权和优先转让权；双方共同完成的由双方共享，具体按照双方的贡献大小进行分配或双方另行商定。②项目成果的转让，须在双方同意的前提下进行，任何一方不得私自转让或许可实施。

2. 项目成果申报各级奖项、专利等，合作各方排名根据实际情况另行商定，人员排名原则上按贡献大小先后排名。

第四条：合作项目各方应严格遵守共同签订的合作协议书，除因不可抗拒的客观原因，不得中途撤消或中止合同。在合同期内，合作一方要求修改合同条款，须各方协商，确认后方能生效，必要时还需上报项目主管部门。

第五条：如合作方因各种原因无法履行合同条款时，由项目负责人报项目主管部门予以协商解决。

第六条：经批准中途退出合作的一方，应视具体情况将所余经费退回项目主持方，已用经费由项目负责人提出审查报告，报项目主管部门审批。

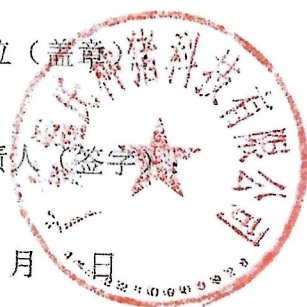
第七条：合作一方在工作进行中有问题不及时报告，影响项目整体的年度进展，将承担相关责任，项目主持人并上报主管部门。

第八条：本协议合作各方签字盖章之日起生效，项目完成之日起终止；若合作申请未获资助，本协议自动废止。

甲方单位（盖章）

项目负责人（签字）

年 月 日



乙方单位（盖章）

项目负责人（4）（签字）

合同专用章



蔡更元

科研空间

欢迎您，张泽宾老师



科研动态

我的办公

我的项目

我的经费

我的成果

我的学术交流

我的考核

我的主页

常用下载

项目立项

项目申报 (1)

所有项目

职称评定

主持的项目

参与的项目

博硕导评定

主持的项目

参与的项目

项目性质

纵向(1项)

项目分类

乡村振兴战略专项(1项)

参与形式

参与(1项)

总数:1项 (表中经费单位: 万元)

项目名称

查询

新增

删除

导出

全选	项目名称	项目分类	项目成员	起止日期	合同金额	项目状态	审核状态	操作
<input type="checkbox"/>	2022年云浮市-广东省新兴县温氏生猪种业创新园	乡村振兴战略专项	蔡更元,杨化强,李紫聪,杨杰,洪林君,顾婷,郑恩琴,丁荣荣,黄思秀,宋长绪,徐铮,刘琅青,张泽宾	2023-01-01--2024-12-31	200	进行	学校通过	

1/1 共1条 1

技术支持: 北京易普拉格科技有限责任公司[科研管理系统] 演示授权版本 信息反馈

二、论文、著作等

检索证明

根据委托人提供的论文材料, 委托人华南农业大学动物科学学院 张泽宾 4 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Recombining Your Way Out of Trouble: The Genetic Architecture of Hybrid Fitness under Environmental Stress	MOLECULAR BIOLOGY AND EVOLUTION 出版年: 2020 出版日期: JAN 卷期: 37 1 页码: 167-182 文献类型: Article	第一作者	T2 类	斯德哥尔摩大学	SCI	IF2-year=16.24 IF5-year=18.67 (2020)	生物学 1 区 Top 期刊: 是 (2020)
2	How broad is the selfing syndrome? Insights from convergent evolution of gene expression across species and tissues in the Capsella genus	NEW PHYTOLOGIST 出版年: 2022 出版日期: DEC 卷期: 236 6 页码: 2344-2357 文献类型: Article	第一作者	T2 类	乌普萨拉大学	SCI	IF2-year=9.4 IF5-year=10.5 (2022)	生物学 1 区 Top 期刊: 是 (2022)
3	Genome-wide association analysis unveils candidate genes and loci associated with aplasia cutis congenita in pigs	BMC GENOMICS 出版年: 2023 出版日期: NOV 21 卷期: 24 1 页码: - 文献号: 701 文献类型: Article	共同通讯作者 (倒数第一)	A 类	华南农业大学	SCI	IF2-year=3.5 IF5-year=4.1 (2023)	生物学 2 区 Top 期刊: 是 (2023)

4	Genome-Wide Association Study Reveals Genetic Mechanisms Underlying Intersex and Aproctia in Large White Pigs	ANIMALS 出版年: 2025 出版日期: APR 10 卷期: 15 8 页码: - 文献号: 1094 文献类型: Article	共同通讯作者 (倒数第一)	A类	华南农业大学	SCI	IF2-year=2.7 IF5-year=3.2 (2024)	农林科学 2区 Top 期刊: 否 (2025)
---	---	--	---------------	----	--------	-----	--	--------------------------------

说明: 论文等级和中科院大类分区按《华南农业大学学位论文评价方案(试行)》划分。

报告免责声明: 如未盖章, 报告无效



华南农业大学图书馆SCAU LIB20251934

MOLECULAR BIOLOGY AND EVOLUTION

academic.oup.com/mbe





Society for Molecular Biology and Evolution

Print ISSN 0737-4038
Online ISSN 1537-1719

- Adaptive radiation of food-borne flukes
- Functionality of long noncoding RNAs
- Advantages of a four-base alphabet
- Model selection with ModelTest-NG
- How homing pigeons find their way

第101页

Recombining Your Way Out of Trouble: The Genetic Architecture of Hybrid Fitness under Environmental Stress

Zebin Zhang,^{†,1} Devin P. Bendixsen ,^{†,1} Thijs Janzen,^{2,3} Arne W. Nolte,^{2,3} Duncan Greig,^{2,4} and Rike Stelkens ^{*,1,2}

¹Division of Population Genetics, Department of Zoology, Stockholm University, Stockholm, Sweden

²Max Planck Institute for Evolutionary Biology, Plön, Germany

³Institute of Biology and Environmental Sciences, University of Oldenburg, Oldenburg, Germany

⁴Centre for Life's Origins and Evolution (CLOE), Department of Genetics, Evolution, and Environment, University College London, London, United Kingdom

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: rike.stelkens@zoologi.su.se.

Associate editor: Aya Takahashi

Abstract

Hybridization between species can either promote or impede adaptation. But we know very little about the genetic basis of hybrid fitness, especially in nondomesticated organisms, and when populations are facing environmental stress. We made genetically variable F2 hybrid populations from two divergent *Saccharomyces* yeast species. We exposed populations to ten toxins and sequenced the most resilient hybrids on low coverage using ddRADseq to investigate four aspects of their genomes: 1) hybridity, 2) interspecific heterozygosity, 3) epistasis (positive or negative associations between nonhomologous chromosomes), and 4) ploidy. We used linear mixed-effect models and simulations to measure to which extent hybrid genome composition was contingent on the environment. Genomes grown in different environments varied in every aspect of hybridness measured, revealing strong genotype–environment interactions. We also found selection against heterozygosity or directional selection for one of the parental alleles, with larger fitness of genomes carrying more homozygous allelic combinations in an otherwise hybrid genomic background. In addition, individual chromosomes and chromosomal interactions showed significant species biases and pervasive aneuploidies. Against our expectations, we observed multiple beneficial, opposite-species chromosome associations, confirmed by epistasis- and selection-free computer simulations, which is surprising given the large divergence of parental genomes (~15%). Together, these results suggest that successful, stress-resilient hybrid genomes can be assembled from the best features of both parents without paying high costs of negative epistasis. This illustrates the importance of measuring genetic trait architecture in an environmental context when determining the evolutionary potential of genetically diverse hybrid populations.

Key words: *Saccharomyces*, hybridization, environmental stress, ddRADseq, heterozygosity, epistasis, genome evolution.

Introduction

Populations exposed to gene flow, introgression, or hybridization contain vast amounts of genetic variation, sometimes producing phenotypes with more extreme adaptations than found in the parent populations (Shahid et al. 2008; Stelkens and Seehausen 2009; Pritchard et al. 2013; Stelkens, Brockhurst, Hurst, Miller, et al. 2014; Holzman and Hulsey 2017). The average fitness of hybrid crosses is usually lower than that of nonhybrid crosses due to interspecific incompatibilities or other negative effects, for example, the break-up and loss of coadapted beneficial gene complexes for local adaptation (Coyne and Orr 2004). This applies especially to the hybrid offspring of genetically divergent lineages. However, some hybrids show high fitness (Rieseberg et al. 1999; Dittrich-Reed and Fitzpatrick 2013), which is often

exploited in agricultural breeding to improve the yield, taste, or other desirable traits of cultivars (Marullo et al. 2006; Kuczyńska et al. 2007; Shivaprasad et al. 2012; Koide et al. 2019). But fit and fertile hybrids are also relevant for adaptive evolution and the generation of biodiversity, especially when ecologically divergent hybrid phenotypes become reproductively isolated from the parents. Ultimately this process can lead to hybrid speciation (Anderson and Stebbins 1954; Lewontin and Birch 1966; Rieseberg et al. 2003; Seehausen 2004; Arnold 2006; Mallet 2007; Nolte and Tautz 2010; Abbott et al. 2013; Schumer et al. 2014).

The genetic mechanisms allowing some hybrid genomes to express high-fitness phenotypes, despite negative epistatic effects compromising their fitness, are poorly known, especially in nondomesticated organisms. We also know very little

about the impact of stressful and deteriorating environmental conditions on the evolutionary potential of hybrid populations, although this is becoming increasingly relevant in the face of global environmental change and invasive species management. So far, empirical studies putting hybrid fitness in an environmental context are rare and the results are mixed. Some found environmental stress to intensify the negative fitness effects of hybridization (Koevoets et al. 2012; Barreto and Burton 2013), others found hybrid fitness to increase with stress (Edmands and Deimler 2004; Willett 2012; Hwang et al. 2016). Again, others detected no effects of stress on hybrid fitness (Armbruster et al. 1999). Studies in interspecific yeast crosses suggest that F2 hybrids can outcompete their parent species in various stressful environments (Greig et al. 2002; Stelkens, Brockhurst, Hurst, and Greig 2014) and that hybridization can increase resistance against a range of toxins (Stelkens, Brockhurst, Hurst, Miller, et al. 2014).

We made genetically highly variable F2 hybrids by crossing two divergent species of *Saccharomyces* budding yeast (*S. cerevisiae* and *S. paradoxus*). These species have well-sequenced, well-assembled genomes that differ at ~15% of nucleotides genome-wide (Liti et al. 2009). Due to their large divergence, these species produce only ~1% viable F2-hybrid offspring (Hunter et al. 1996; Boynton et al. 2018; Rogers et al. 2018). We exposed 240 populations of viable F2 hybrids to ten stressful environments containing high concentrations of different toxins (e.g., caffeine, ethanol, lithium acetate; [supplementary table S1, Supplementary Material online](#)). At the end of the growth period, we sequenced the genomes of 240 hybrid survivors at low coverage using double digest, restriction-site-associated DNA (ddRAD) markers and mapped hybrid genomes to both parental reference genomes to analyze their genome composition. **To measure the impact of the genetic and environmental factors shaping the hybrid genomes,** we measured four aspects of “hybridness”: 1) hybridity (proportion of hybrid genome mapping to one or other parent species), 2) interspecific heterozygosity (homologous chromosomes from opposite species), 3) epistasis (positive or negative associations between nonhomologous chromosomes from the same and opposite species), and aneuploidy (aberrant chromosome copy numbers compared with the euploid parents). By virtue of design, we were restricted to sampling only from the 1% viable subset of F2 hybrids here (i.e., we could only sequence survivors and not the inviable hybrids carrying lethal incompatibilities).

Hybridity is a continuous measure. It is the proportion of genomic material inherited from one over the other parental species (Gompert and Buerkle 2016), often referred to as hybrid index (Barton and Gale 1993; Buerkle 2005) or admixture proportion (Pritchard et al. 2000; Falush et al. 2003). Hybridity is useful for locating hybrid genomes on a single axis, ranging from zero to one with pure parental genomes at opposite ends. However, this simple measure can mask potentially important fitness effects of individual loci (or chromosomes) deviating from the average hybridity of the genome. For instance, dominance (the reciprocal masking of deleterious alleles at multiple loci; Bruce 1910) and overdominance (the intrinsic benefit of being heterozygous for at least one

locus) can produce hybrids with high fitness. This is known as “interspecific heterozygosity” or “interpopulation ancestry” (Barton 2000; Fitzpatrick 2012; Lindtke et al. 2012; Gompert and Buerkle 2013). As an example, although every chromosome in a diploid F1 hybrid is heterozygous, maximizing within-locus hybridity, the genome still carries a full haploid set of both parental chromosomes. At the same time, a diploid F2 or higher generation hybrid can, theoretically, be composed of fully homozygous chromosomes (i.e., homologous chromosomes are from the same species), which minimizes within-locus hybridity. But this F2 hybrid may contain half the chromosomal set from either parent, maximizing between-locus hybridity. Because these types of hybrids would be indistinguishable with a single hybrid index (which would be 0.5 in both examples), we also measured interspecific heterozygosity, that is, chromosomes in the F2 hybrid genome carrying interspecific, homologous combinations.

Epistasis, caused by interactions between alleles from at least two different loci that increase or decrease fitness more than the sum of the individual contributions of these loci, is common in nature and has been shown to be susceptible to changes in the environment (Kvitek and Sherlock 2011; de Vos et al. 2013). Jaffe et al. (2019) showed that adding more environmental conditions tripled the number of genetic interactions detected in fitness assays between double mutants of yeast (Jaffe et al. 2019). Filteau et al. (2015) found that the course of compensatory evolution rescuing yeast populations from the negative fitness effects of deleterious mutations was strongly contingent on the environment (different carbon sources) (Filteau et al. 2015). Lee et al. (2019), dissecting the genetic basis of a yeast colony phenotype, recently found that environmental stress affected the impact of epistasis, additivity, and genotype–environment interactions.

Epistasis, especially negative epistasis, is predicted to play a large role in hybrid fitness (Cubillos et al. 2011; Shapira et al. 2014). Negative epistasis is the core element of the Dobzhansky–Muller model of genetic incompatibilities (DMIs), which is often recruited to explain the evolution of reproductive isolation and outbreeding depression between biological species, with increasing negative impact the more divergent the parental genomes are (Dobzhansky 1936; Müller 1942; Lynch 1991). Here, we measured epistasis by testing for significant associations (presence and absence) between nonhomologous chromosomes from different species in F2 hybrid yeast genomes. We compared our data with epistasis- and selection-free simulations, assuming free segregation of chromosomes in a theoretical F2 hybrid population. Given the large nucleotide divergence between the parental genomes used here (~15% genome-wide), negative epistasis is expected to be prominent, and we expected to find more same species associations than opposite species associations in F2 hybrids. However, it is important to keep in mind that we were restricted to sampling only from the viable subset of F2 hybrids here (because we cannot sequence dead hybrids). As a result, we cannot make inferences about the impact of DMIs on the average fitness of this hybrid cross, and our results remain inconclusive as to the relative importance of DMIs compared with other genetic mechanisms of hybrid

breakdown, such as underdominance, or directional selection for one of the parental alleles in hybrid genomes.

Results

We sequenced a total of 240 F2 hybrid strains, of which 184 were mostly diploid and 53 were mostly haploid (three genomes were discarded due to low read quality). Thus, the majority of spores germinated and mated in the 96-well plates, forming diploid F2 hybrids homozygous for both drug resistance markers *cyh2r* and *can1r*. The haploid genotypes detected by sequencing were unmated F2 (i.e., F1 spores), and hemizygous for *cyh2r* and *can1r*. Because we saw no significant differences between F2 samples isolated from high concentration of a toxin and low concentration in any of the tests, we proceeded by pooling all samples for analysis.

Interspecific Heterozygosity

To measure interspecific heterozygosity (i.e., genomes having homologous chromosomes from both species) in our diploid samples, we considered chromosomes heterozygous if marker proportions fell between 0.25 and 0.75. Chromosomes with smaller or larger species content were considered homozygous for the dominant species. Assuming no selection and free segregation of chromosomes in our simulated chromosomes, interspecific heterozygosity was lower than expected in the 187 diploid F2 genomes across all toxins (mean heterozygosity = 0.44, median = 0.36; [fig. 1A](#)). **This low level of mean heterozygosity was found in <0.8% of simulations based on no interactions** ([supplementary fig. S1, Supplementary Material](#) online). On average, only 7 of the 16 chromosomes per genome carried a heterospecific combination. In total, there were almost 1.3 times more homozygous chromosomes ($n = 1,450$; 57%) than heterozygous chromosomes ($n = 1,126$; 43%). Environments differed substantially in the proportions of homo- and heterozygous chromosomes ([fig. 1B](#)). Although in some environments, F2 hybrids were mainly homozygous (e.g., zinc sulfate: 0.73), other environments promoted the growth of mainly heterozygote genotypes (e.g., NaCl: 0.56). This was confirmed using simulations. Five stress environments (zinc sulfate: 0.28, citric acid: 0.29, ethanol: 0.32, salicylic acid: 0.35, and caffeine: 0.40) produced mean interspecific heterozygosities below the range (0.41–0.59) expected based on random segregation simulations ([supplementary fig. S1, Supplementary Material](#) online). Seven environments produced mean homozygosities above the expected range: Three environments (zinc sulfate: 0.35, citric acid: 0.35, and ethanol: 0.36) produced F2 genomes more homozygous than the simulated range (0.16–0.34) of expected homozygosity for *S. cerevisiae*. Four environments (caffeine: 0.33, citric acid: 0.36, salicylic acid: 0.37, and zinc sulfate: 0.37) produced F2 genomes with higher homozygosity than the simulated range (0.15–0.33) of expected homozygosity for *S. paradoxus*.

Testing for variation in zygosity between chromosomes, across all environments, we found three chromosomes with a significantly larger proportion of *S. cerevisiae* homozygotes (chromosomes 4: 0.57, 5: 0.31, 14: 0.78), and three

chromosomes with a significantly larger proportion of *S. paradoxus* homozygotes (chromosomes 6: 0.45, 7: 1.0, 10: 0.36; [fig. 1C](#)). Note that chromosomes 5 and 7 carried resistance genes, which, as predicted, selected them to be homozygous for *S. cerevisiae* and *S. paradoxus*, respectively (chromosome 5 also carried some *S. paradoxus* content due to recombination at one end of the chromosome). Three chromosomes were more often heterozygous than expected by chance (chromosomes 2: 0.60, 3: 0.56, 12: 0.58).

Using a linear mixed-effect model to understand what causes variation in zygosity between environments, chromosomes, and genomes, we found that the interaction between *chromosome ID* and *environment* best predicted zygosity ([supplementary table S2, Supplementary Material](#) online). Thus, which chromosome was homozygous, and which chromosome was heterozygous, was largely determined by the stress environment from which it was sampled.

Chromosome Copy Number

To quantify the chromosome copy number within all F2 genomes and to detect potential aneuploidies, we considered a chromosome aneuploid if the mean copy number was >2.5 or <1.5 in diploids, and >1.5 in haploids. Using this cut-off, we found that 47 overall haploid genomes and 161 overall diploid genomes of the 237 F2 hybrids contained aneuploidy, affecting 15.3% of all chromosomes in haploids and 14.8% of chromosomes in diploids (i.e., on average 2.4 chromosomes per genome were aneuploid; [supplementary fig. S2A, Supplementary Material](#) online). Although no significant differences were detected between overall chromosome copy numbers of F2 hybrids and euploid F1 hybrids within environments (Wilcoxon test; [supplementary fig. S2B, Supplementary Material](#) online), using data from all environments, we found significant differences between chromosomes in average copy number, with nine chromosomes (1, 4, 5, 6, 7, 8, 12, 13, and 14) having significantly different ploidy from the chromosomes in euploid F1 genomes ([supplementary fig. S2C, Supplementary Material](#) online). For instance, chromosomes 2, 4, 13, and 14 showed high percentages of aneuploidy, whereas chromosomes 5 and 7 showed low percentages, probably due to the drug marker they contained ([supplementary fig. S2D, Supplementary Material](#) online). [Supplementary figure S3, Supplementary Material](#) online, provides an overview of the pervasive aneuploidy we found in the F2 hybrid genomes in ten stressful environments. [Supplementary figure S4, Supplementary Material](#) online, shows the chromosome copy number of one representative aneuploid F2 hybrid genome chosen from each environment for illustration.

Genome-Wide Hybridity

Average genome-wide hybridity (defined to be the percentage of markers inherited from *S. cerevisiae*) was 0.52 ± 0.05 ($\mu \pm \text{SD}$) across all chromosomes and all environments ([supplementary fig. S5A, Supplementary Material](#) online). This pattern was consistent with the analysis of published data ([Kao et al. 2010](#)) from an environment without added stress ([supplementary fig. S6A, Supplementary Material](#) online). Thus, genomic contributions were roughly balanced between

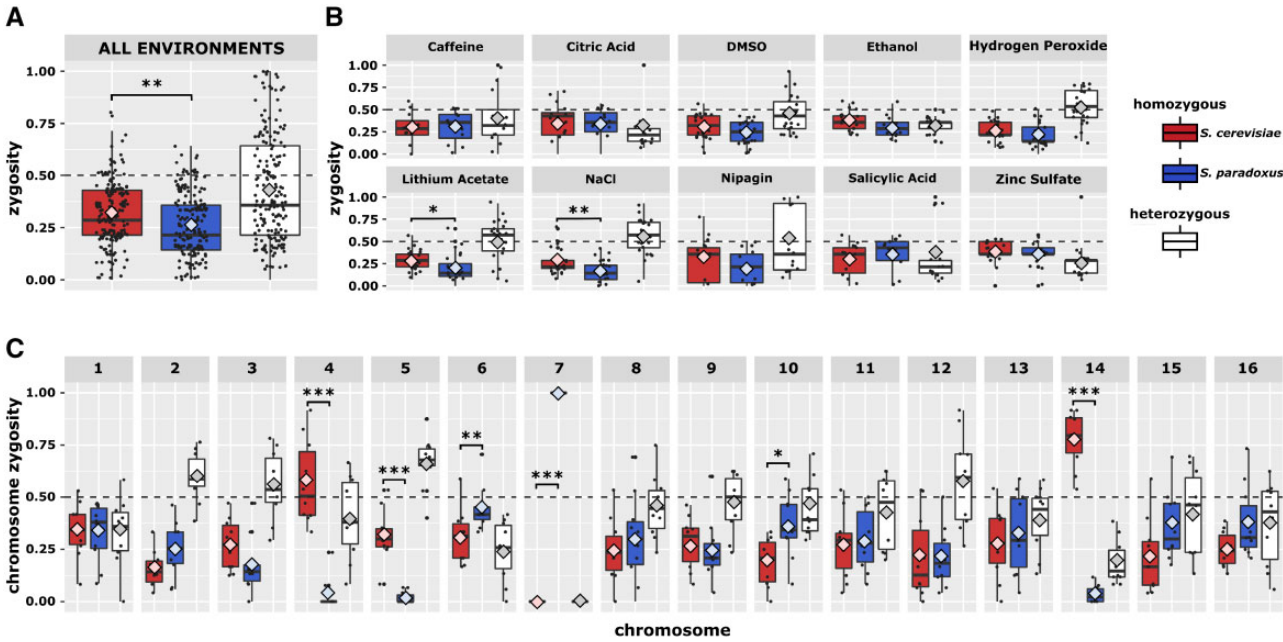


Fig. 1. Homo-/heterozygosity of diploid F2 hybrid genomes. (A) Mean zygosity of 187 diploid F2 hybrids across all chromosomes (except chromosomes 5 and 7) and across all environments. Dashed horizontal line (at 0.5) shows expected heterozygosity without selection and free segregation. Black lines in boxes are medians and large diamonds indicate means. Each black dot represents a genome. Boxes are interquartile ranges (IQR). Whiskers are $1.5 \times \text{IQR}$. (B) Mean F2 hybrid zygosity per environment. (C) Mean F2 hybrid zygosity per chromosome. Chromosomes 5 and 7 harbor drug markers selecting for *Saccharomyces cerevisiae* and *S. paradoxus* chromosomes, respectively. Each dot represents one of ten environments. Asterisks indicate significant differences in proportions of *S. cerevisiae* versus *S. paradoxus* homozygotes in Wilcoxon signed-rank tests ($P = *0.05$, $**0.01$, $***0.001$).

the two parental species with a slight overrepresentation of *S. cerevisiae*. The most biased genome toward *S. cerevisiae* was 0.68 and the most biased toward *S. paradoxus* was 0.35. Average hybridity per toxin ranged from 0.49 ± 0.07 in salicylic acid (*paradoxus*-biased) to 0.55 ± 0.07 (*cerevisiae*-biased) in Nipagin, but no significant differences were found between toxins (supplementary fig. S5B, Supplementary Material online). The distributions and means of genome-wide hybridity observed in each environment were all within the range (0.44–0.56) expected from simulations based on random chromosome segregation (supplementary fig. S7, Supplementary Material online). Chromosomes 5 and 7 were excluded from these analyses because they carried antibiotic resistance markers selecting for *S. cerevisiae* and *S. paradoxus* hemi- or homozygosity, respectively.

Chromosome Hybridity

A closer examination of hybridity (defined to be % markers inherited from *S. cerevisiae*) at the chromosomal level revealed species biases in inheritance patterns (fig. 2A). For most chromosomes the average chromosome hybridity across all environments was ~ 0.5 , suggesting equal inheritance from *S. cerevisiae* and *S. paradoxus*. As expected, chromosome 5 was inherited primarily from *S. cerevisiae* and chromosome 7 was inherited primarily from *S. paradoxus* because they carried species-specific drug markers. Interestingly, chromosomes 4 and 14 were also inherited predominantly from *S. cerevisiae* (0.77 and 0.88, respectively). Analysis of previously published data in an environment

without stress (Kao et al. 2010) also found this bias in chromosome 4 (supplementary fig. S6B, Supplementary Material online). Nine of the 16 chromosomes exhibited varying levels of hybridity depending on environment (ANOVA, $P < 0.05$), with six showing significant variation (fig. 2B, $P < 0.001$). For some chromosomes, the hybridity variation resulting from environment did not change the species bias (chromosomes 7 and 14). However, some chromosomes shifted from unbiased (~ 0.5 hybridity) to biased for *S. paradoxus* in some environments (chromosomes 10 and 15). Two chromosomes exhibited opposite species biases depending on the environment (chromosomes 12 and 13). In some environments these chromosomes were biased toward *S. cerevisiae* ($>0.65\%$), and in other environments they were biased toward *S. paradoxus* ($<0.35\%$).

Using a linear mixed-effect model, we found that interactions between *aneuploidy* and *environment*, *chromosome ID* and *environment*, and *aneuploidy*, *chromosome ID* and *ploidy* (haploid or diploid) best predicted the hybridity of the chromosome (supplementary table S3, Supplementary Material online). The optimal model did not change if data from chromosomes 5 and 7 were excluded.

By combining our chromosome hybridity data with published “relative fitness” data for the parental strains in each environment (Bernardes et al. 2017), we found a possible link between species-biased chromosomal inheritance and parental fitness. We derived “parental phenotypic divergence” from Bernardes et al. (supplementary fig. S8A, Supplementary Material online), defined as the difference between the competitive growth of the two parents relative to their

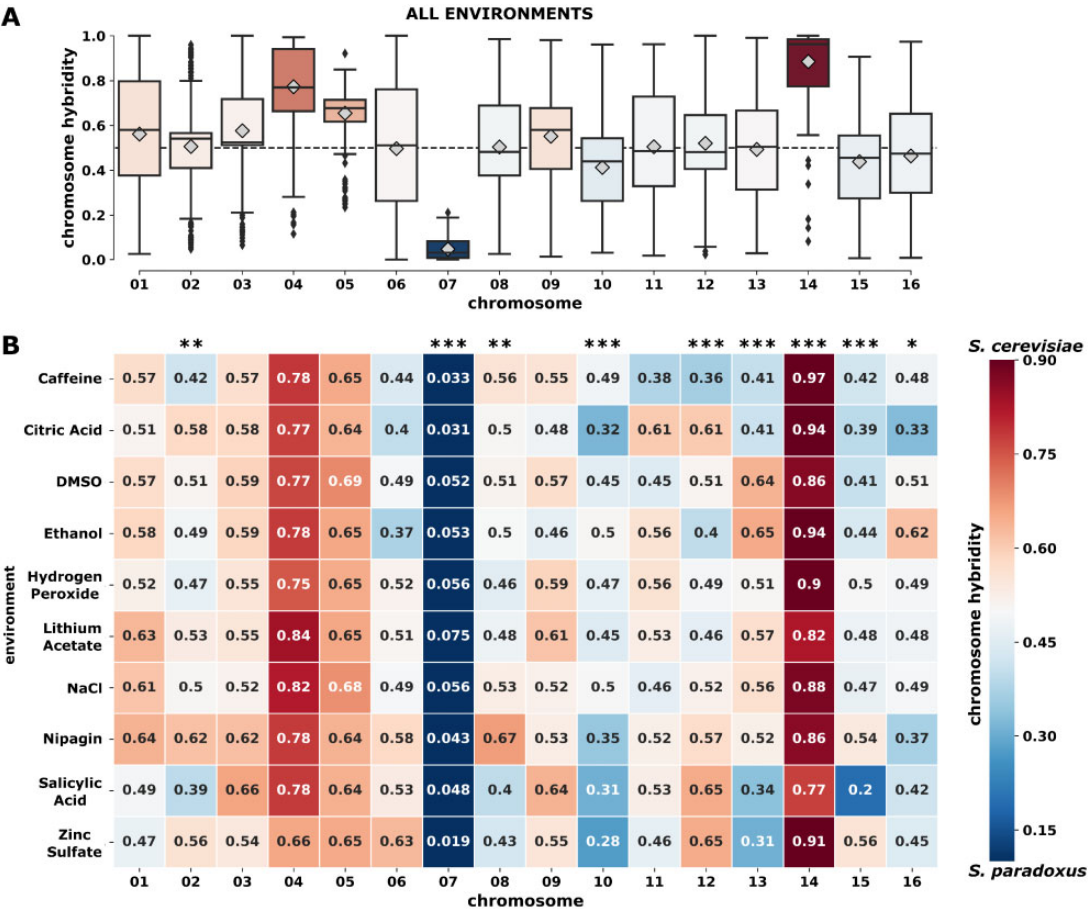


FIG. 2. Chromosome hybridity of 237 haploid and diploid F2 genomes. (A) Chromosome hybridity for each chromosome across all environments. Chromosome hybridity is measured as the percent chromosome mapping to *Saccharomyces cerevisiae*. Boxplots as in figure 1, but colored according to where the median falls. Dashed line (0.5) indicates equal amounts of the chromosome mapping to *S. cerevisiae*. Species biases on chromosomes 5 (~65% from *S. cerevisiae*) and 7 (~5% from *S. cerevisiae*) are by design and result from alternative recessive drug markers. (B) Chromosome hybridity for each chromosome within environments. Numbers in colored boxes indicate average chromosome hybridity across 24 F2 diploid genomes. Asterisks indicate significant differences in average chromosome hybridity across environments ($P = *0.05$, $**0.01$, $***0.001$).

interspecific F1 hybrid (*S. cerevisiae*–*S. paradoxus*). Negative parental phenotypic divergence suggests that the *S. cerevisiae* parent performed better in competition with the hybrid. Positive parental phenotypic divergence suggests that the *S. paradoxus* parent performed better during competition with the hybrid. Note that parental phenotypic divergence is a relative measure here and assumes transitivity of relative fitness between each of the two parents and the F1 hybrid. This analysis revealed a significant correlation ($P < 0.05$) between parental phenotypic divergence and chromosome hybridity in chromosomes 1 and 13 (supplementary fig. S8B, Supplementary Material online). This relationship suggests that in environments with a higher relative fitness for the *S. cerevisiae* parent, chromosome hybridity was biased toward *S. cerevisiae*. Inversely, in environments with higher relative fitness for the *S. paradoxus* parent, chromosome hybridity was biased toward *S. paradoxus*.

Chromosome Hybridity Interactions

Testing for chromosomal hybridity interactions within a given F2 genome revealed strong environment-dependent effects. A large difference in chromosome hybridity for each

chromosome combination, defined as delta chromosome hybridity (DCH), indicated that two chromosomes mapped primarily to opposing species within a given F2 genome. This could potentially result from negative epistatic interactions within species (i.e., chromosome incompatibility), from positive epistatic interactions between species or a combination of the two. Some chromosomes maintained high DCH suggesting that they were preferably from opposing species across all environments (fig. 3A, chr7 × chr4, chr7 × chr14, chr7 × chr5). Chromosome 7 and 5 were designed to come from opposing species and were therefore expected to have high DCH. However, the remaining high DCH interactions with chromosome 7 suggest potential chromosome incompatibilities or strong positive epistatic interactions between species. Chromosome 7 was designed to be inherited from *S. paradoxus* and potentially resulting from this, chromosome 4 and chromosome 14 were primarily inherited from *S. cerevisiae*. Analyzing the DCH in each stress environment independently showed that different environments exhibited varying levels and distributions of DCH (fig. 3B). Two environments (zinc sulfate: 0.37, salicylic acid: 0.39) exhibited higher mean DCH than the

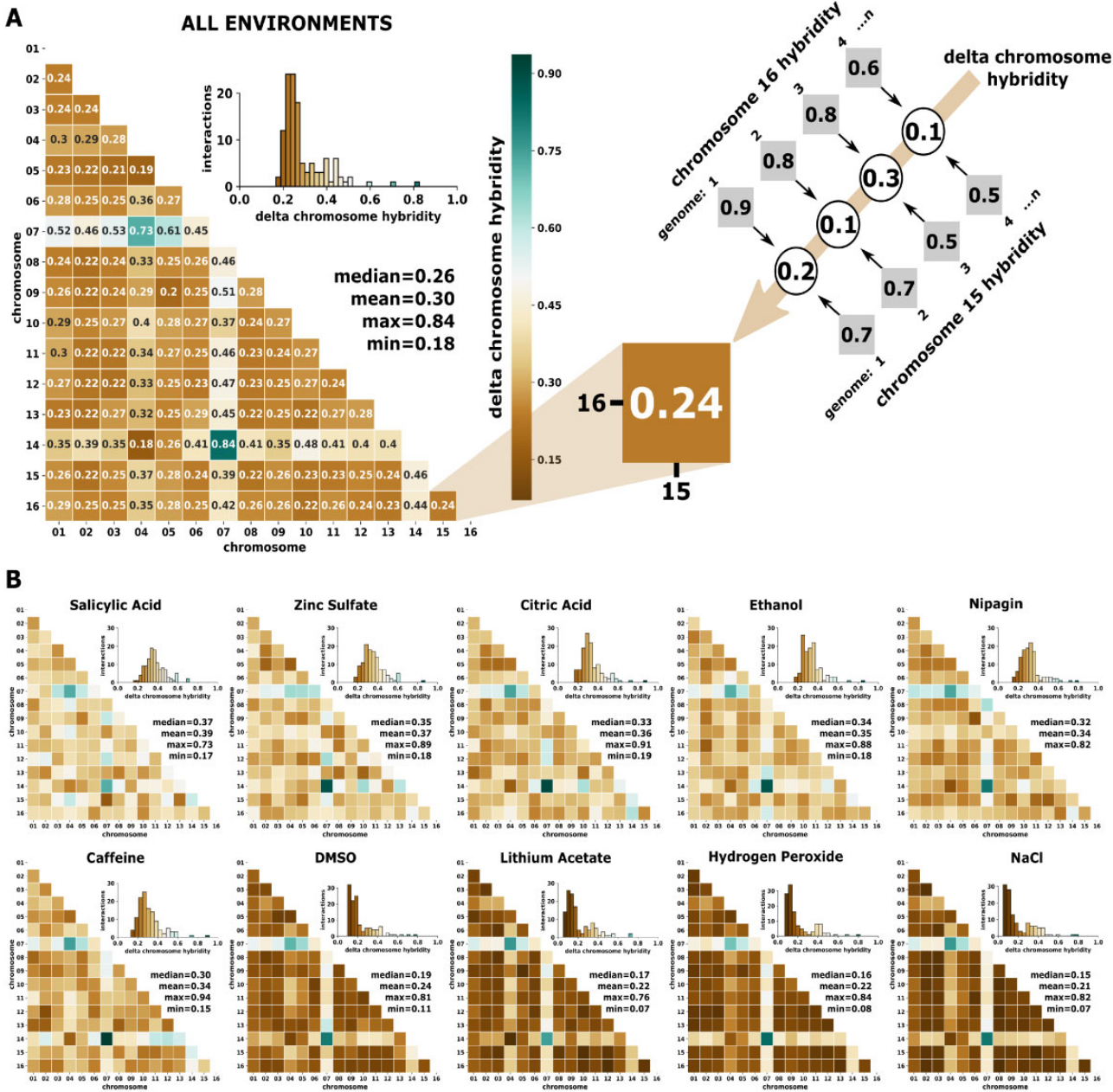


Fig. 3. Interactions of chromosome hybridities altered by environment. (A) Average change in chromosome hybridity (percent chromosome mapping to *Saccharomyces cerevisiae*) between chromosomes across all environments. Inset depicts the distribution of DCH for all chromosome interactions. DCH is determined for each chromosome interaction by taking the difference between the hybridity measurements for each chromosome within an F2 genome (n genomes = 237). These values are then averaged across all diploid F2 genomes, and medians and means are reported and colored accordingly. A large DCH (green) suggests that the chromosomes map primarily to opposing species. A small DCH (brown) suggests that the chromosomes have similar levels of hybridity. These chromosomes may come from primarily the same species or at least have similar hybridity proportions. (B) Average change in chromosome hybridity between chromosomes for each environment. Calculations are performed the same as panel (A) for each environment independently (n genomes ~ 24).

range (0.29–0.36) expected from simulations based on no chromosome interactions (fig. 4). Four environments (NaCl: 0.21, hydrogen peroxide: 0.22, lithium acetate: 0.22, DMSO: 0.24) exhibited lower mean DCH than the simulated expected range. An analysis of published data in an environment without added stress (Kao et al. 2010) found slightly higher mean DCH (0.42) than the simulated range (supplementary fig. S6C, Supplementary Material online).

Analyzing directional correlations between hybridities of chromosomes within a genome revealed similar environment-dependent interactions (fig. 5). Significant positive correlations suggest positive epistatic interactions within species, meaning that as chromosome hybridity in one chromosome shifted toward one species ($1 = S. cerevisiae$, $0 = S. paradoxus$), the chromosome hybridity in the linked chromosome also shifted toward the same species (fig. 5A). Inversely, negative correlations suggested negative epistatic

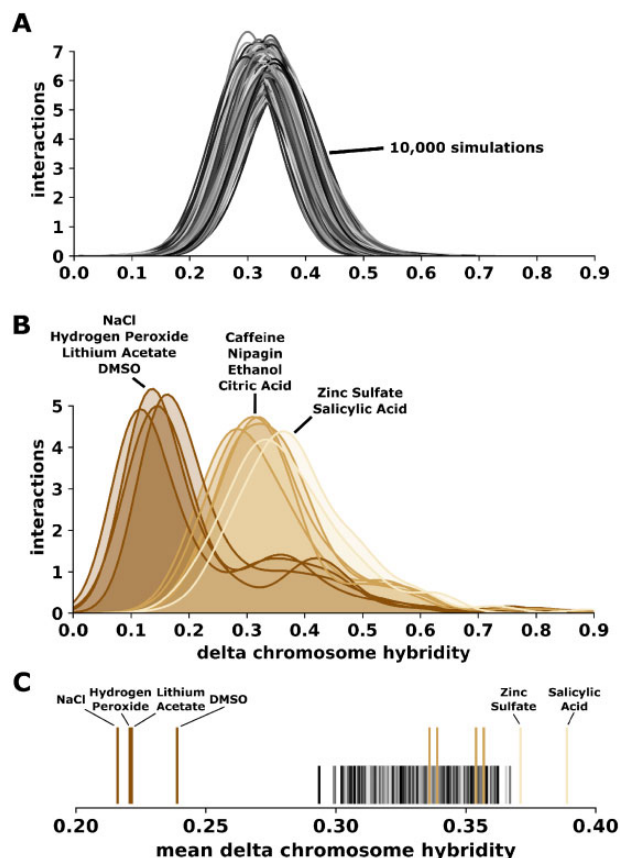


Fig. 4. DCH of simulated chromosomes. (A) The distribution of DCH in a simulated environment resulting from simulated chromosomes based on no chromosomal interactions. Each environment consists of 24 genomes, each consisting of 16 simulated chromosomes. Each simulated environment was repeated 10,000 times. (B) The distribution of DCH found in our experimental environments. Distributions are colored according to their mean DCH. (C) Mean DCH of our experimental environments as compared with the range of mean DCH found in simulations (gray-black lines).

interactions within species. A lack of correlation suggested no significant epistatic interaction. For all environments combined and individually, the distribution of correlations was approximately normal with a mean near zero. There was a range of significant correlations ($P < 0.01$) found across environments ranging in numbers from 0 in salicylic acid and citric acid to 11 in NaCl (fig. 5B). The distributions and means for each environment all fell within the range (-0.05 to 0.10) that was expected from simulations based on no chromosome interactions (supplementary fig. S9, Supplementary Material online). However, although the distributions and means were within the range simulated, three environments (lithium acetate: 7, caffeine: 8, and NaCl: 11) exhibited more significant correlations ($P < 0.01$) than the range expected (0–7, supplementary fig. S10, Supplementary Material online). The NaCl environment resulted in a total of 11 significant hybridity correlations (8 positive, 3 negative), whereas 10,000 simulations never encountered more than seven correlations by chance. Interestingly, the NaCl environment also contained the highest molarity range (0.234–0.4 M). However, this correlation between stress concentration and

interactions was not found in all environments. Analysis of published data in an environment without stress (Kao et al. 2010) found only a single significant correlation ($P < 0.01$, supplementary fig. S6D, Supplementary Material online).

Genotype \times Environment Interactions and Gene Ontology

Sampling all F2 hybrid genomes, a total of 315 10-kb bins fixed for either *S. cerevisiae* or *S. paradoxus* were detected (406 bins including chromosomes 5 and 7; supplementary fig. S11A, Supplementary Material online). We found large variation between environments ranging from 36 regions in salicylic acid to 372 regions in NaCl. Seventy-eight of these regions were environment-specific, that is, they were inherited from either *S. cerevisiae* or *S. paradoxus* across all 24 hybrid genomes per environment, but only found in one environment.

We identified 1,884 genes that were either located in or overlapped with the 315 fixed regions, ranging from 51 genes in salicylic acid to 1,533 genes in NaCl (supplementary fig. S11C, Supplementary Material online). The large differences between environments in the number of fixed alleles suggest that some stress conditions require more complex, quantitative adaptations with a polygenic basis, which is well-known to be the case for yeast adapting to salt for instance (Cubillos et al. 2011; Dhar et al. 2011). Gene annotation analysis showed that most of the 1,884 genes are involved in transport and pathways related to fungal cell-wall organization (supplementary fig. S11B, Supplementary Material online). Across environments, we found almost three times more genes fixed for the *S. cerevisiae* allele ($n = 1,373$) than fixed for the *S. paradoxus* allele ($n = 511$). Only the F2 hybrid genomes isolated from salicylic acid showed more genes fixed for *S. paradoxus* (46 of 51 genes in total).

Only one region was consistently inherited from *S. cerevisiae* across all environments (chromosome 4: 970,000–980,000 bp, harboring five genes), and one from *S. paradoxus* across all environments (chromosome 15: 10,000–20,000 bp, harboring seven genes). Among these are two general stress response genes, *HSP78* and *PAU20*. *HSP78* on chromosome 4 is associated with heat stress and mitochondrial genome maintenance (Leonhardt et al. 1993; von Janowsky et al. 2006). *PAU20* on chromosome 15 is associated with proteome stress and is upregulated during wine fermentation (Rossignol et al. 2003; Marks et al. 2008; Luo and van Vuuren 2009). It is surprising to find the *S. paradoxus* allele of *PAU20* fixed across all environments as we might expect the *S. cerevisiae* allele to provide higher fitness, especially in fermentation environments.

We did not identify fixations of alternative alleles at the same locus in different environments (e.g., a bin that was consistently inherited from *S. cerevisiae* across all 24 hybrid genomes in ethanol but from *S. paradoxus* in salicylic acid).

Discussion

Hybrid fitness varies among genotypes, generations, and environments (Nolte and Tautz 2010; Arnold et al. 2012; Stelkens,



Here, we describe the composition of 237 stress-resistant F2 hybrid genomes, made from two divergent *Saccharomyces*

yeast species. We measured variance in overall hybridity, interspecific heterozygosity, and epistasis found between and within hybrid genomes. We found that the composition of hybrid genomes was strongly contingent on environmental context. Genomes from different environments varied in every aspect of hybridness measured. **First, genomes from different environments differed in the proportion of interspecific heterozygosity.** Although some environments clearly selected for more homozygous genomes (e.g., zinc sulfate), others selected for more heterozygous genomes (e.g., NaCl; [fig. 1B](#)). **Second, individual chromosomes exhibited strong species biases depending on the type of stress they were exposed to ([fig. 2B](#)).** This is presumably because genes important for resistance to the specific toxin are located on these chromosomes, and alleles from parental species differ in how well they tolerate this toxin ([supplementary fig. S8, Supplementary Material](#) online). **However, we did not observe fixation of opposite species alleles in different environments ([supplementary fig. S11, Supplementary Material](#) online).** Third, we found evidence for environment-dependent, nonhomologous chromosomal associations ([fig. 3](#)). Positive same-species and to our surprise, also positive opposite-species chromosome associations occurred, depending in frequency and in type on the environment ([fig. 5](#)). Computer simulations in a selection- and epistasis-free space with independently segregating chromosomes ([fig. 4](#) and [supplementary figs. S7, S9, and S10, Supplementary Material](#) online), and a comparison to hybrid genomes grown in a stress-free environment from [Kao et al. \(2010\)](#) ([supplementary fig. S6, Supplementary Material](#) online), produced a smaller number of significant associations between chromosomes in hybrid genomes. This suggests that some heterospecific chromosomal combinations were indeed under positive epistatic selection. It is difficult to distinguish between environmental selection and the effects of (nonlethal) genetic incompatibility in our hybrids. However, the analyses of hybrids grown without stress ([supplementary fig. S6, Supplementary Material](#) online) suggest that at least a portion of the effect is driven by environmental selection.

Beneficial interactions between genetic materials from species with such large evolutionary divergence (~15%) are surprising. At this genetic distance, which is roughly three times larger than the distance between humans and chimpanzees, hybrid genomes are expected to contain Dobzhansky–Muller incompatibilities (or strong negative epistasis) and other negative fitness effects as a result of species divergence and reproductive isolation. In yeast, the missegregation of chromosomes due to antirecombination ([Greig et al. 2003](#)) is probably responsible for most of the >99% mortality observed in the F1 hybrid gametes of *S. cerevisiae* × *S. paradoxus* crosses ([Liti et al. 2006](#); [Hou et al. 2014](#)). It is important to keep in mind that we only sampled from the 1% viable subset of all F2 hybrids here, thus a priori excluding lethal hybrids or those with strongly compromised fitness. Still, we might expect selection to favor mainly same-species chromosomal combinations, which is not what we observed here (in the resolution possible given limited rounds of segregation and recombination in the F2 generation).

Interestingly, when screening for recessive incompatibilities in the same hybrid cross, a previous study found that replacing chromosomes 4, 13, and 14 in *S. cerevisiae* with homologous chromosomes from *S. paradoxus* did not yield any viable haploids ([Greig 2007](#)). In our experiment, focusing here on the NaCl environment because it revealed the most significant epistatic interactions, the same three chromosomes (4, 13, and 14) were involved in five of eight positive same-species interactions, suggesting that different-species combinations with any of these three chromosomes may indeed cause low fitness in hybrid genomes, and are selected against ([fig. 5B](#)). In addition, we found three positive opposite-species associations in NaCl involving five chromosomes to occur more often than expected by chance (6, 8, 9, 10, and 16). Of these, [Greig \(2007\)](#) also found four chromosomes (6, 8, 9, and 10) not to be problematic to transfer into the opposite species background (there was no suitable auxotrophic marker for chromosome 16, so this was a technical limitation, not an incompatibility). This is consistent with positive heterospecific epistasis, or at least an absence of incompatibilities on these chromosomes.

Individual chromosomes also differed in their level of interspecific heterozygosity independently of the environment. For instance, chromosomes 4, 6, and 14 were more beneficial to fitness as homozygotes, whereas chromosomes 2, 3, and 12 were more beneficial as heterozygotes ([fig. 1C](#)). Overall however, interspecific heterozygosity was unexpectedly low across all hybrid genomes ([supplementary fig. S1, Supplementary Material](#) online), suggesting that same-species allelic combinations (homozygosity) in an otherwise hybrid genomic background can provide high fitness. These results speak against strong effects of dominance or overdominance to hybrid fitness and are more consistent with directional selection for one of the parental alleles, or selection against heterozygotes (underdominance), which has been shown in *Saccharomyces* ([Laiba et al. 2016](#)), natural *Populus* ([Lindtke et al. 2012](#)), and *Helianthus* hybrid populations ([Lai et al. 2005](#)). Yet, dominance complementation of recessive alleles and overdominant interactions within loci are often reported from laboratory crosses of *Saccharomyces* where heterozygotes are fitter than both homozygotes ([Zörgö et al. 2012](#); [Plech et al. 2014](#); [Shapira et al. 2014](#); [Blein-Nicolas et al. 2015](#); [Laiba et al. 2016](#)). In some cases, this inconsistency with our results may be explained by the much smaller genomic divergence (usually <1%) between the parental lineages used in these studies (mostly all *S. cerevisiae* strains). Studies using the same genetically divergent *S. cerevisiae* × *S. paradoxus* cross as we suggested that heterosis in F1 hybrids was likely a result of dominance complementation of recessive deleterious alleles, but may also be due to additional overdominant or epistatic effects ([Zörgö et al. 2012](#); [Bernardes et al. 2017](#)). The excess of homozygosity in our experiment is also consistent with many if not most high fitness alleles being recessive and only becoming expressed in the homozygous state. Alternatively, selection on some aspects of fitness in the haploid phase (during sporulation and germination) may have caused strains with the same beneficial combinations of parental chromosomes to mate, resulting in homozygous diploids.

For instance, strains may be able to maximize their fitness by germinating early (Miller and Greig 2015; Stelkens et al. 2016), which could result in a process similar to assortative mating.

Aneuploidy, the gain or loss of chromosomes, is a common byproduct of missegregation in the F1 hybrid meiosis of interspecific yeast crosses (Hunter et al. 1996). We detected high levels of aneuploidy in our F2 hybrid genomes across all environments (208 of 237 genomes contained aneuploidy; supplementary figs. S2 and S3, Supplementary Material online) with significant variation between chromosomes. This may suggest that gene dosage effects due to copy number changes may have helped some hybrids in our experiment survive stress. Aneuploidy is known to be associated with stress and drug resistance in yeast and other fungi (Selmecki et al. 2009; Pavelka et al. 2010; Kwon-Chung and Chang 2012; Yang et al. 2019), and has been suggested to serve as a transient adaptation mechanism (Chang et al. 2013; Hose et al. 2015; Smukowski Heil et al. 2017). In our experiment, chromosome 2, 4, 13, and 14 showed high overall levels of aneuploidy, suggesting that aneuploidy of these chromosomes might be advantageous in stressful environments (supplementary fig. S2C and D, Supplementary Material online). Chromosome 4 and 13 aneuploidy has been shown to result from hydrogen peroxide (Linder et al. 2017) and galactose stress (Sirr et al. 2015), respectively. Inversely, the low aneuploidy in chromosomes 5 and 7 was likely linked to the selective drug markers located on these chromosomes, which we used to select for F2 hybrids. Interestingly, we did not find a negative correlation between chromosome size and rate of aneuploidy as previously suggested (Gilchrist and Stelkens 2019).

Hybridization mostly occurs at the margins of species ranges where conditions are more extreme than in the center of their distribution, that is, in more stressful environments. Hybridization also occurs more frequently in human- or otherwise perturbed habitats, where geographic and ecological species barriers are lost or weakened (King et al. 2015; Arnold 2016; Gompert and Buerkle 2016; McFarlane and Pemberton 2019). Thus, the circumstances leading to hybridization often coincide with times of increased environmental stress, which in turn creates ecological opportunity. The large number of significant chromosome-by-environment interactions we found in our hybrid populations illustrates the genetic variation typical for hybrid swarms. This can facilitate their functional diversification and, potentially, the colonization of novel and challenging environments as shown for instance in *Helianthus* sunflowers and African cichlid fish (Rieseberg et al. 2003; Seehausen 2004; Abbott et al. 2013). We acknowledge that the facultative asexual reproduction and the ability of yeast to self-fertilize (hybrids do not rely on backcrossing with the parents) can help restore fitness quickly after outbreeding depression in the F2 hybrid generation. Together, this can catalyze the propagation of hybrid genotypes even with small and temporary fitness advantages.

In conclusion, understanding the environmental and genetic mechanisms responsible for hybrid fitness is essential for any predictions regarding the role of hybridization in adaptive evolution (Gompert and Buerkle 2016). For instance, our data

suggests that each environment selects for different parental alleles and, also, for different epistatic interactions. Thus, a given hybrid could outcompete and drive a parent population to extinction in one locality, but may have inferior fitness in a different habitat, posing no threat to the parental lines. The results of our study show that in order to capture the risks and benefits of genetic exchange between lineages, it is important to measure multiple dimensions of hybridity and measure hybrid fitness in multiple environmental contexts.

Materials and Methods

Parent and F1 Hybrid Strains

We chose two genetically tractable laboratory strains as parents: *S. cerevisiae* haploid strain YDP907 (MAT α , *ura3::KanMX*, *can1r*), isogenic with strain background S288c, was crossed to *S. paradoxus* strain YDP728 (MAT α , *ura3::KanMX*, *cyh2r*), which is isogenic with strain background N17. This produced a diploid F1 hybrid (MAT α / α , *ura3::KanMX*, *cyh2r*, *can1r*), which was purified by streaking and then stored frozen at -80°C in 20% glycerol stock.

F2 Hybrids

To make F2 hybrids, we grew a 5-ml culture of the F1 hybrid in YEPD (1% yeast extract, 2% peptone, 2% dextrose) for 24 h at 30°C , and then transferred 200 μl of this to 50 ml of KAC sporulation medium (1% potassium acetate, 0.1% yeast extract, 0.05% glucose, 2% agar) which was shaken at room temperature for 5 days to induce meiosis and sporulation. Sporulation was confirmed under the microscope. Fifty microliters of the spore culture, in serial dilutions and using three replicates per dilution, was then plated onto arginine drop-out agar plates supplemented with the drugs canavanine (60 mg/l) and cycloheximide (3.33 mg/l). These drugs kill the fully heterozygous F1 hybrids because the resistance alleles *cyh2r* and *can1r* are recessive. But those meiotic spores (produced by the F1 hybrid) that contain both resistance alleles can, if viable, germinate and form colonies. This method ensured that only viable F2 hybrids were sampled and sequenced.

Environments and Stress

Fifty microliters of spore culture (containing 25.5 double drug-resistant viable spores on average confirmed by streaking out on double-drug medium) was used as the founding population for inoculation and growth on flat-bottomed, 96-well cell culture plates. Wells contained liquid minimal medium plus uracil, cycloheximide, and canavanine (0.67% yeast nitrogen base without amino acids, 2% glucose, 2% agar, 0.003% uracil, 60 mg/l canavanine, 3.33 mg/l cycloheximide), allowing for the germination, mating, and growth of the double drug-resistant F2 progeny. This was supplemented with a range of concentrations of ten toxins (one at a time): salicylic acid ($\text{C}_7\text{H}_6\text{O}_3$), caffeine ($\text{C}_8\text{H}_{10}\text{N}_4\text{O}_2$), ethanol ($\text{C}_2\text{H}_6\text{O}$), zinc sulfate (ZnSO_4), hydrogen peroxide (H_2O_2), methyl paraben ("Nipagin"; $\text{CH}_3(\text{C}_6\text{H}_4(\text{OH})\text{COO})$), sodium chloride (NaCl), lithium acetate (CH_3COOLi), dimethyl sulfoxide ("DMSO"; $\text{C}_2\text{H}_6\text{OS}$), and citric acid ($\text{C}_6\text{H}_8\text{O}_7$). These ten toxins are

arbitrary but do represent established stress environments for yeast, which we have used in previous assays (Stelkens, Brockhurst, Hurst, Miller, et al. 2014; Bernardes et al. 2017). Toxicity gradients (or “environmental clines”) were generated along the y-axis of the 96-well plates such that the bottom row contained the lowest concentration, and the top row contained the highest concentration of the toxin, lethal for all strains (exact concentrations in [supplementary table S1, Supplementary Material](#) online).

Plates were incubated at 30 °C for 4 days. Then, 1 µl from each well was transferred to the same position on a new 96-well culture plate containing identical concentrations. The optical density (OD) of each well was measured with a microplate reader (Infinite M200 Pro, Tecan) (time point t_0), and plates were incubated at 30 °C for another 3 days. After this, another OD measurement was taken (t_1) and plates were stored at 4 °C until further processing. We calculated, for each well, whether growth had occurred by subtracting OD_{t_0} from OD_{t_1} . Assuming that a doubling in OD approximately equals a doubling of cell numbers, cells in permissive environments at the bottom of the plate completed on average three cell cycles (2.78 ± 0.15 across toxins) from t_0 to t_1 whereas cells in the topmost row did not divide at all.

We sampled from all 12 wells of the bottom row (i.e., the lowest stress) and from the 12 wells of the highest stress that allowed growth from each column, from all ten plates. These 240 samples were streaked out for single colonies on YEPD plates and grown for 48 h at 30 °C. We then picked a single colony from each sample and froze it for sequencing.

DNA Extraction, Polymerase Chain Reaction, and Sequencing Protocols

DNA was extracted using the MasterPure Yeast DNA Purification Kit (Epicentre). A genotyping-by-sequencing protocol was modified from microsatellite library preparation and ddRAD sequencing approaches as follows (Nolte et al. 2005; Peterson et al. 2012). The library construction is based on an efficient combined restriction digest/adaptor ligation. The restriction enzymes *Csp6I* (which cleaves 5'-GT A C-3' sites) and *PstI* (which cleaves 5'-C T G C AG-3' sites) were used to digest genomic DNA to generate sticky ends. The reaction conditions permit that sticky end adaptors and T4 ligase are added to the reaction such that adaptors are ligated to the restriction sites. Importantly, the adaptors do not fully reconstitute the restriction sites. Thus, once an adaptor is ligated, this site will not be recut whereas an undesired ligation of two genomic DNA sticky ends will be recut until all DNA ends are saturated with an adaptor. For this purpose, we modified the ligation sites of the Illumina Truseq adapters such that they matched the sticky ends generated by the restriction enzymes. Further, we labeled the adapters to include 24 and 16 different molecular barcodes (MIDs), respectively, which could be combined in 384 different combinations for multiplexed sequencing (paired ends) of the libraries on an Illumina MiSeq sequencer.

Quality Filtering of Raw Reads and Mapping Protocol

We examined the quality of raw ddRADseq reads of each sample using FastQC (v0.11.8) (FastQC 2018). Illumina sequencing adapters, primer sequences, ambiguous, and low-quality nucleotides (PHRED quality score < 20) were removed from both paired-end reads according to the default parameters in the NGS QC Toolkit (v2.3.3) (Patel and Jain 2012). Three of 240 F2 genomes were abandoned due to low sequence yield. A total of 8.2-Gb high-quality reads were generated after quality control with mean depth of 12.84 per marker. Sequencing data are available at the European Nucleotide Archive (ENA accession number: PRJEB33368).

Parental strain sequences were downloaded from the *Saccharomyces* Genome Database website. The S288c reference genome used was R64-2-1_20150113. The N17 reference genome was obtained from <ftp://ftp.sanger.ac.uk/pub/users/dmc/yeast/latest/>, in the *para_assemblies/N_17* folder.

Given the high sequence divergence between *S. cerevisiae* and *S. paradoxus* (15%), correct assignment of reads to the parental species was efficient using NovoAlign (NovoAlign 2018) with default parameters. For this, both parental genomes were concatenated and used as a combined reference for mapping, and species affiliations and chromosomal positions of successfully mapped reads were written to the same output file. Mapped reads were considered correctly assigned to the reference sequence when the mapping quality was ≥ 20 , indicating a single and unique best match.

Ploidy Determination

To differentiate haploid and diploid F2 hybrid genomes (to be able to score the prevalence of homo- and heterozygosity), we mapped paired-end reads again, this time only to the *S. cerevisiae* S288c reference genome using NovoAlign (NovoAlign 2018) with the default parameters. We sorted mapped files according to their genomic coordinates using Picard Tools v2.18.23 (Picard 2019) and performed variant calling in FreeBayes (Garrison and Marth 2012) with at least five supporting reads required to consider a variant. Single nucleotide polymorphisms were selected and filtered using GATK (Li et al. 2008) according to stringent filtering criteria with the following settings: 1) quality > 30.0, 2) quality by depth > 5.0, 3) fisher strand < 60.0, 4) root mean square of mapping quality > 40.0, 5) MQRankSum > -12.5; and 6) ReadPosRankSum > -8.0. Additionally, if there were more than three SNPs clustered in a 10-bp window, we considered all three SNPs false positives and removed them. After filtering, we identified 14,975 SNPs, ranging between 10,725 and 12,345 among environments.

Haploidy or diploidy was called using the heterozygosity score of biallelic SNPs using a custom bash script. Theoretically, both the number of heterozygous SNPs and ratio of heterozygous SNPs over the total SNPs in haploids should be zero (except in case of sequencing errors and aneuploidies). However, the absolute number of heterozygous SNPs is affected by the overall sequence size of each sample, and the ratio is affected by the number of SNPs derived from *S. paradoxus* (because of the large nucleotide divergence between the parents, ~15% of the reads from *S. paradoxus* do

not map to the *S. cerevisiae* reference genome). Also, although unlikely, diploid F2 genomes can theoretically be fully homozygous with both copies of eight chromosomes inherited from each parental species, assuming that genetic material from both parental species is equally represented in the F2 generation. At zero heterozygosity, haploid and diploid genomes are indistinguishable by SNP calling and mapping to parental genomes. Thus, at the risk of omitting extremely homozygous diploids, we applied two criteria to consider a genome diploid: 1) The number of heterozygous SNPs in a diploid F2 genome should be larger than the average number of heterozygous SNPs found in both parental genomes, and 2) the ratio of heterozygous over hetero- plus homozygous SNPs in a diploid F2 genome should be larger than the average ratio of heterozygous over hetero- plus homozygous SNPs in parental genomes. As there are nearly zero heterozygous SNPs in the *S. cerevisiae* genome, this can be simplified to:

$$\text{ratio}(\text{SNP}_{\text{extreme}}^{\text{hetero}}) \approx \frac{n(\text{SNP}_{S.\text{par}}^{\text{hetero}})}{2} \cdot \frac{n(\text{SNP}_{S.\text{par}}^{\text{hetero}}) + n(\text{SNP}_{S.\text{par}}^{\text{homo}})}{n(\text{SNP}_{S.\text{par}}^{\text{hetero}})}$$

Using this analysis, we assigned 187 of the 237 F2 genomes to be diploid.

Chromosome Copy Number

To calculate chromosome copy number, the same mapping files used for ploidy determination of each genome were divided into 10-kb bins. Next, a series of filtering steps were performed to select an optimal set of bins to represent the chromosome copy number across the genome. To eliminate sequencing and fragmentation error, we first removed bins with a median read depth <2 in the nine euploid F1 hybrids and each F2 genome sequenced for this study. Then, the proportion of all reads aligning to that bin was calculated using a previously described method (Tan et al. 2013):

$$\text{Bin proportion} = \frac{\text{Read counts for bin}}{\text{Total read counts for strain}}.$$

To eliminate bin-specific coverage effects, for each bin in each query F2 strain, the bin proportion was normalized using the mean bin proportion of the same coordinate from the nine euploid F1 hybrids:

$$\text{Bin ratio} = \frac{\text{Query bin proportion}}{\text{Mean bin proportion of 9 F1 hybrids}}.$$

This ratio at the 10-kb bin level still showed sequencing bias as a result of the ddRAD sequencing approach, for example, chromosome-end bias, GC-content bias, fragment-length bias, etc. Considering that the effects of these biases are likely minimal at the chromosome level, we then calculated the chromosome ratio of each query strain and normalized the median chromosome ratio to 2 in diploids, or to 1 in haploids:

$$\text{Chromosome ratio}_{\text{Query strain normalized}} = \frac{\text{Chromosome ratio}}{\text{Median chromosome ratio for query strain}}.$$

After that, we transformed the 10-kb bin ratio using BoxCox (Box and Cox 1964) and normalized the data using the chromosome ratio. This was then used to represent the chromosome copy number:

$$\text{Chromosome copy number} = \text{Bin ratio}_{\text{BoxCox}} \times \text{Chromosome ratio}_{\text{Query strain normalized}}.$$

Chromosome Zygosity

To determine the zygosity of each chromosome in diploid F2 genomes, we mapped high-quality reads of each sample to the reference combining both parental genomes. We extracted the species affiliation of each read using a custom bash script. The chromosomal content per species per sample was extracted using custom perl code by calculating the proportion of markers mapping to only one species over markers mapping to both species. We considered chromosomes heterozygous if marker proportions fell between 0.25 and 0.75. Chromosomes with smaller or larger species content were assigned homozygous.

To understand what causes variation in zygosity between environments, chromosomes, and genomes, we modeled chromosome heterozygosity using a mixed-effect linear model. Chromosome zygosity was used as the response variable in the model and was predicted as a function of fixed predictors environment and chromosome ID and their interactions and sample ID as a random effect. We selected the most appropriate model by identifying the simplest model that maintained the lowest Akaike Information Criterion (AIC; Akaike 1974). AIC optimizes the relationship between the fit and complexity of a model by balancing the fit of the model with the number of parameters estimated (Harrison et al. 2018).

Genome-Level and Chromosome-Level Hybridity

To describe the genome-level and chromosome-level hybridity, we calculated the percentage of the entire genome or chromosome for each sample that mapped to either the *S. cerevisiae* or the *S. paradoxus* genome. We defined hybridity as the percent of markers per genome or chromosome mapping to *S. cerevisiae* (100% = *S. cerevisiae*, 0% = *S. paradoxus*). This was calculated across all environments, as well as for each environment individually. We then used mixed-effect linear models with genome and chromosome hybridity as response variables. We predicted hybridity to be a function of the fixed effects *aneuploidy*, *environment*, *chromosome ID*, *ploidy*, and their interactions and included sample ID as random effect. We selected the most appropriate model by starting with the full model and removing insignificant components, identifying the simplest model that maintained the lowest AIC.

To characterize the interactions between chromosomes within a F2 genome, we calculated the **difference in chromosome hybridity** for each chromosome combination, defined as DCH. Large DCH indicates that the two chromosomes map primarily to opposing species within a genome. Small DCH indicates that the chromosomes have similar levels of

hybridity, suggesting that either the chromosomes come from primarily the same species or at least they have similar hybridity proportions. DCH was calculated for all environments together, as well as each environment individually. To investigate the epistatic interactions between chromosomes, we determined the Pearson correlation (r) between chromosome hybridity for all pairwise chromosome combinations. A significant positive correlation between two chromosomes suggests a positive epistatic interaction within a species, meaning that as chromosome hybridity in one chromosome shifts toward one species ($1 = S. cerevisiae$, $0 = S. paradoxus$), the chromosome hybridity in the linked chromosome also shifts toward the same species. A significant negative correlation between two chromosomes suggests a negative epistatic interaction within a species. In this case, as one chromosome shifts toward one species, the linked chromosome shifts toward the other species. Custom Python scripts used for analyses of hybridity are available on GitLab (https://gitlab.com/devinbendixsen/yeast_hybrid_stress).

Simulated Chromosomes

In order to determine whether the observed zygosity, hybridity, and chromosomal interactions found in each environment deviated from neutral expectations, we developed simulations that determined the expected range for these measurements based solely on chance, **assuming no selection and free segregation of chromosomes in hybrid genomes**. Each simulation displayed the expected distribution of zygosity, hybridity, or chromosomal interactions that could be found in a selection-free environment with no epistasis between chromosomes. Each simulated environment was repeated 10,000 times. Each environment consisted of 24 genomes and each genome consisted of 16 chromosomes as found in our data. Each chromosome was randomly assigned a chromosome hybridity score between 0 and 1. To study the expected zygosity, chromosomes were labeled as either homozygous for *S. cerevisiae* or *S. paradoxus* or heterozygous based on the same rules used in the analysis of our data (see Chromosome Zygosity). The heterozygosity and hybridity for each genome were then calculated for each simulation. The means and distributions of these values were plotted and compared with the means and distributions we received from our data ([supplementary figs. S1 and S7, Supplementary Material](#) online). To study the expected chromosome interactions, we calculated the 120 interactions that occur within a simulated genome between chromosomes. DCH and Pearson correlation (r) were calculated in the same manner as our data was analyzed (see Genome-Level and Chromosome-Level Hybridity). The distributions of these interactions were then plotted and compared with the data from our study ([fig. 4 and supplementary fig. S9, Supplementary Material](#) online). For the Pearson correlation (r), we also determined the range of expected significant positive and negative correlations and compared them with our data ([supplementary fig. S10, Supplementary Material](#) online). Custom Python scripts used for simulations are available on GitLab (https://gitlab.com/devinbendixsen/yeast_hybrid_stress).

Genotype \times Environment Interactions and Gene Ontology

To explore for gene-by-environment interactions, that is, which gene from which species background is more frequently found in which environment, we divided genomes into 10-kb bins. Assuming each bin is a locus and the parental species affiliations are the two possible alleles, we scored bins as 1 if all markers within that bin mapped to the *S. cerevisiae* reference genome, and 0 if all markers mapped to the *S. paradoxus* reference genome across all 24 genomes within a given environment (excluding chromosome 5 and 7). Genes located in or overlapping with these fixed regions of each environment were further inspected using the gene annotation data base DAVID ([Huang et al. 2009](#)). Environment-independent fixations of alleles (bins fixed for the same species across all environments) and unique fixations (bin fixed only for one environment) were extracted using custom R code.

Analysis of Hybrid Genomes in Environment without Toxin

Data from *S. cerevisiae* \times *S. paradoxus* F2 hybrids grown in an environment without any toxin were obtained from [Kao et al. \(2010\)](#). These data were generated using a dual-species comparative genomic hybridization microarray analysis with species-specific DNA oligos for *S. cerevisiae* and *S. paradoxus*. Data for 36 hybrid genomes were analyzed by first determining the number of oligos that mapped significantly to each species (using $\text{LOG_RAT2N_MEAN} = -0.05$ as a cut-off). Genome and chromosome hybridity (defined as percentage of oligos mapping to *S. cerevisiae*) were calculated ([supplementary fig. S6A and B, Supplementary Material](#) online). DCH and Pearson correlation were analyzed as previously described for our own data ([supplementary fig. S6C and D, Supplementary Material](#) online). Custom Python scripts used for these analyses are available on GitLab (https://gitlab.com/devinbendixsen/yeast_hybrid_stress).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Elke Bustorf for the preparation of libraries and David Rogers for advice on mapping. We also thank Katy C. Kao for providing data for hybrids that were grown in an environment without toxin. This work was supported by the Max Planck Society (to D.G.), ERC Starting Grant (“EVOLMAPPING” to A.W.N.), Swedish Research Council (Grant Number 2017-04963 to R.S.), Knut and Alice Wallenberg Foundation (Grant Number 2017.0163 to R.S.) Carl Trygger Foundation (to Z.Z.), and Wenner-Gren Foundation (Grant Number UPD2018-0196 to D.P.B.).

Author Contributions

R.S. and D.G. received the project idea and designed the experiment. R.S. performed the experiment and collected the samples. A.W.N. planned and performed the library

preparation and sequencing. Z.Z. processed the ddRADseq data. Z.Z., D.P.B., and T.J. analyzed the data. R.S., Z.Z., and D.P.B. wrote the manuscript.

References

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman JW, Brelsford A, Buerkle CA, Buggs R. 2013. Hybridization and speciation. *J Evol Biol.* 26(2):229–246.
- Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 19(6):716–723.
- Anderson E, Stebbins GL. 1954. Hybridization as an evolutionary stimulus. *Evolution* 8(4):378–388.
- Armbruster P, Bradshaw WE, Steiner AL, Holzapfel CM. 1999. Evolutionary responses to environmental stress by the pitcher-plant mosquito, *Wyeomyia smithii*. *Heredity* 83(5):509–519.
- Arnold ML. 2006. Evolution through genetic exchange. Oxford/New York: Oxford University Press.
- Arnold ML. 2016. Anderson's and Stebbins' prophecy comes true: genetic exchange in fluctuating environments. *Syst Bot.* 41(1):4–16.
- Arnold ML, Ballerini ES, Brothers AN. 2012. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana lilies. *Heredity (Edinb).* 108(3):159–166.
- Barreto FS, Burton RS. 2013. Elevated oxidative damage is correlated with reduced fitness in interpopulation hybrids of a marine copepod. *Proc R Soc B.* 280(1767):20131521.
- Barton NH. 2000. Estimating multilocus linkage disequilibria. *Heredity* 84(3):373–389.
- Barton NH, Gale KS. 1993. Genetic analysis of hybrid zones. In: Harrison RG, editor. Hybrid zones and the evolutionary process. New York: Oxford University Press. p. 13–45.
- Bernardes JP, Stelkens RB, Greig D. 2017. Heterosis in hybrids within and between yeast species. *J Evol Biol.* 30(3):538–548.
- Blein-Nicolas M, Albertin W, da Silva T, Valot B, Balliau T, Masneuf-Pomarède I, Bely M, Marullo P, Sicard D, Dillmann C, et al. 2015. A systems approach to elucidate heterosis of protein abundances in yeast. *Mol Cell Proteomics.* 14(8):2056–2071.
- Box GE, Cox DR. 1964. An analysis of transformations. *J R Stat Soc Series B (Methodol).* 26:211–243.
- Boynton PJ, Janzen T, Greig D. 2018. Modeling the contributions of chromosome segregation errors and aneuploidy to *Saccharomyces* hybrid sterility. *Yeast* 35(1):85–98.
- Bruce AB. 1910. The Mendelian theory of heredity and the augmentation of vigor. *Science* 32(827):627.
- Buerkle CA. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Mol Ecol Notes.* 5(3):684–687.
- Buerkle CA, Rieseberg LH. 2008. The rate of genome stabilization in homoploid hybrid species. *Evolution* 62(2):266–275.
- Chang S-L, Lai H-Y, Tung S-Y, Leu J-Y. 2013. Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet.* 9(1):e1003232.
- Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.
- Cubillos FA, Billi E, Zörgo E, Parts L, Fargier P, Omholt S, Blomberg A, Warringer J, Louis EJ, Liti G. 2011. Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol Ecol.* 20(7):1401–1413.
- de Vos MGJ, Poelwijk FJ, Battich N, Ndika JDT, Tans SJ. 2013. Environmental dependence of genetic constraint. *PLoS Genet.* 9(6):e1003580.
- Dhar R, Sagesser R, Weikert C, Yuan J, Wagner A. 2011. Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. *J Evol Biol.* 24(5):1135–1153.
- Dittrich-Reed DR, Fitzpatrick BM. 2013. Transgressive hybrids as hopeful monsters. *Evol Biol.* 40(2):310–315.
- Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21(2):113–135.
- Edmands S, Deimler JK. 2004. Local adaptation, intrinsic coadaptation and the effects of environmental stress on interpopulation hybrids in the copepod *Tigriopus californicus*. *J Exp Mar Biol Ecol.* 303(2):183–196.
- Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587.
- FastQC. 2018. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, last accessed March 2019.
- Filteau M, Hamel V, Pouliot MC, Gagnon-Arsenault I, Dube AK, Landry CR. 2015. Evolutionary rescue by compensatory mutations is constrained by genomic and environmental backgrounds. *Mol Syst Biol.* 11(10):832.
- Fitzpatrick BM. 2012. Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evol Biol.* 12(1):131–114.
- Fitzpatrick BM, Shaffer HB. 2007. Hybrid vigor between native and introduced salamanders raises new challenges for conservation. *Proc Natl Acad Sci U S A.* 104(40):15793–15798.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907v2 [q-bio.GN], last accessed March 2019.
- Gilchrist C, Stelkens R. 2019. Aneuploidy in yeast: segregation error or adaptation mechanism? *Yeast.* <https://doi.org/10.1002/yea.3427>.
- Gompert Z, Buerkle CA. 2013. Analyses of genetic ancestry enable key insights for molecular ecology. *Mol Ecol.* 22(21):5278–5294.
- Gompert Z, Buerkle CA. 2016. What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evol Appl.* 9(7):909.
- Greig D. 2007. A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS Genet.* 3(2):e21–286.
- Greig D, Louis EJ, Borts RH, Travisano M. 2002. Hybrid speciation in experimental populations of yeast. *Science* 298(5599):1773–1775.
- Greig D, Travisano M, Louis EJ, Borts RH. 2003. A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J Evol Biol.* 16(3):429–437.
- Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, Robinson BS, Hodgson DJ, Inger R. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:e4794.
- Holzman R, Hulseley CD. 2017. Mechanical transgressive segregation and the rapid origin of trophic novelty. *Sci Rep.* 7:40306.
- Hose J, Yong CM, Sardi M, Wang Z, Newton MA, Gasch AP. 2015. Dosage compensation can buffer copy-number variation in wild yeast. *eLife* 4:e05462.
- Hou J, Friedrich A, de Montigny J, Schacherer J. 2014. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol.* 24(10):1153–1159.
- Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4(1):44–57.
- Hunter N, Chambers S, Louis E, Borts R. 1996. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J.* 15(7):1726–1733.
- Hwang AS, Pritchard VL, Edmands S. 2016. Recovery from hybrid breakdown in a marine invertebrate is faster, stronger and more repeatable under environmental stress. *J Evol Biol.* 29(9):1793–1803.
- Jaffe M, Dziulko A, Smith JD, St Onge RP, Levy SF, Sherlock G. 2019. Improved discovery of genetic interactions using CRISPRiSeq across multiple environments. *Genome Res.* 29(4):668.
- Kao KC, Schwartz K, Sherlock G. 2010. A genome-wide analysis reveals no nuclear Dobzhansky-Muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. *PLoS Genet.* 6(7):e1001038.
- King KC, Stelkens RB, Webster JP, Smith DF, Brockhurst MA. 2015. Hybridization in parasites: consequences for adaptive evolution, pathogenesis, and public health in a changing world. *PLoS Pathog.* 11(9):e1005098.

- Koevoets T, Van De Zande L, Beukeboom LW. 2012. Temperature stress increases hybrid incompatibilities in the parasitic wasp genus *Nasonia*. *J Evol Biol*. 25(2):304–316.
- Koide Y, Sakaguchi S, Uchiyama T, Ota Y, Tezuka A, Nagano AJ, Ishiguro S, Takamura I, Kishima Y. 2019. Genetic properties responsible for the transgressive segregation of days to heading in rice. *G3 (Bethesda)*. 9:1655–1662.
- Kuczyńska A, Surma M, Adamski T. 2007. Methods to predict transgressive segregation in barley and other self-pollinated crops. *J Appl Genet*. 48(4):321–328.
- Kvitek DJ, Sherlock G. 2011. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet*. 7(4):e1002056.
- Kwon-Chung KJ, Chang YC. 2012. Aneuploidy and drug resistance in pathogenic fungi. *PLoS Pathog*. 8(11):e1003022.
- Lai Z, Nakazato T, Salmaso M, Burke JM, Tang S, Knapp SJ, Rieseberg LH. 2005. Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* 171(1):291–303.
- Laiba E, Glikaitė I, Levy Y, Pasternak Z, Fridman E. 2016. Genome scan for nonadditive heterotic trait loci reveals mainly underdominant effects in *Saccharomyces cerevisiae*. *Genome* 59(4):231–242.
- Lee JT, Coradini ALV, Shen A, Ehrenreich IM. 2019. Layers of cryptic genetic variation underlie a yeast complex trait. *Genetics* 211(4):1469–1482.
- Leonhardt SA, Fearson K, Danese PN, Mason TL. 1993. HSP78 encodes a yeast mitochondrial heat shock protein in the Clp family of ATP-dependent proteases. *Mol Cell Biol*. 13(10):6304–6313.
- Lewontin RC, Birch LC. 1966. Hybridization as a source of variation for adaptation to new environments. *Evolution* 20(3):315–336.
- Li H, Ruan J, Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 18(11):1851–1858.
- Linder RA, Greco JP, Seidl F, Matsui T, Ehrenreich IM. 2017. The stress-inducible peroxidase TSA2 underlies a conditionally beneficial chromosomal duplication in *Saccharomyces cerevisiae*. *G3 (Bethesda)* 7(9):3177–3184.
- Lindtke D, Buerkle CA, Barabá T, Heinze B, Castiglione S, Bartha D, Lexer C. 2012. Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Mol Ecol*. 21(20):5042–5058.
- Liti G, Barton DBH, Louis EJ. 2006. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174(2):839–850.
- Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, et al. 2009. Population genomics of domestic and wild yeasts. *Nature* 458(7236):337–341.
- Luo Z, van Vuuren HJ. 2009. Functional analyses of PAU genes in *Saccharomyces cerevisiae*. *Microbiology (Reading, Engl)*. 155(Pt 12):4036–4049.
- Lynch M. 1991. The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution* 45(3):622–629.
- Mallet J. 2007. Hybrid speciation. *Nature* 446(7133):279–283.
- Marks VD, Ho Sui SJ, Erasmus D, van der Merwe GK, Brumm J, Wasserman WW, Bryan J, van Vuuren HJ. 2008. Dynamics of the yeast transcriptome during wine fermentation reveals a novel fermentation stress response. *FEMS Yeast Res*. 8(1):35–52.
- Marullo P, Bely M, Masneuf-Pomarede I, Pons M, Aigle M, Dubourdieu D. 2006. Breeding strategies for combining fermentative qualities and reducing off-flavor production in a wine yeast model. *FEMS Yeast Res*. 6(2):268–279.
- McFarlane SE, Pemberton JM. 2019. Detecting the true extent of introgression during anthropogenic hybridization. *Trends Ecol Evol (Amst)*. 34(4):315–326.
- Miller EL, Greig D. 2015. Spore germination determines yeast inbreeding according to fitness in the local environment. *Am Nat*. 185(2):291–301.
- Müller HJ. 1942. Isolating mechanisms, evolution and temperature. *Biol Symp*. 6:71–125.
- Nolte AW, Freyhof J, Stemshorn KC, Tautz D. 2005. An invasive lineage of sculpins, *Cottus* sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridisation between old phylogeographic groups. *Proc R Soc B*. 272(1579):2379–2387.
- Nolte AW, Tautz D. 2010. Understanding the onset of hybrid speciation. *Trends Genet*. 26(2):54–58.
- NovoAlign. 2018. Available from: <http://novocraft.com>, last accessed March 2019.
- Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7(2):e30619.
- Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R. 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468(7321):321–325.
- Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Mol Ecol*. 25(11):2337–2360.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7(5):e37135.
- Picard. 2019. Available from: <https://broadinstitute.github.io/picard/>, last accessed March 2019.
- Plech M, de Visser J, Korona R. 2014. Heterosis is prevalent among domesticated but not wild strains of *Saccharomyces cerevisiae*. *G3* 4:315–323.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
- Pritchard VL, Knutson VL, Lee M, Zieba J, Edmands S. 2013. Fitness and morphological outcomes of many generations of hybridization in the copepod *Tigriopus californicus*. *J Evol Biol*. 26(2):416–433.
- Rieseberg LH, Archer MA, Wayne RK. 1999. Transgressive segregation, adaptation and speciation. *Heredity* 83(4):363–372.
- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301(5637):1211–1216.
- Rogers DW, McConnell E, Ono J, Greig D. 2018. Spore-autonomous fluorescent protein expression identifies meiotic chromosome mis-segregation as the principal cause of hybrid sterility in yeast. *PLoS Biol*. 16(11):e2005066.
- Rosignol T, Dulau L, Julien A, Blondin B. 2003. Genome-wide monitoring of wine yeast gene expression during alcoholic fermentation. *Yeast* 20(16):1369–1385.
- Schumer M, Rosenthal GG, Andolfatto P. 2014. How common is homoploid hybrid speciation? *Evolution* 68(6):1553–1560.
- Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol Evol (Amst)*. 19(4):198–207.
- Selmecki AM, Dulmage K, Cowen LE, Anderson JB, Berman J. 2009. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet*. 5(10):e1000705.
- Shahid M, Han S, Yoell H, Xu J. 2008. Fitness distribution and transgressive segregation across 40 environments in a hybrid progeny population of the human-pathogenic yeast *Cryptococcus neoformans*. *Genome* 51(4):272–281.
- Shapira R, Levy T, Shaked S, Fridman E, David L. 2014. Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models. *Heredity (Edinb)* 113(4):316–326.
- Shivaprasad PV, Dunn RM, Santos B, Bassett A, Baulcombe DC. 2012. Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J*. 31(2):257–266.
- Sirr A, Cromie GA, Jeffery EW, Gilbert TL, Ludlow CL, Scott AC, Dudley AM. 2015. Allelic variation, aneuploidy, and nongenetic mechanisms suppress a monogenic trait in yeast. *Genetics* 199(1):247–262.
- Smukowski Heil CS, DeSevo CG, Pai DA, Tucker CM, Hoang ML, Dunham MJ. 2017. Selection on heterozygosity drives adaptation in intra- and interspecific hybrids. *Mol Biol Evol* 34:1596–1612.

- Stelkens RB, Brockhurst MA, Hurst GD, Greig D. 2014. Hybridization facilitates evolutionary rescue. *Evol Appl.* 7(10):1209–1217.
- Stelkens RB, Brockhurst MA, Hurst GDD, Miller EL, Greig D. 2014. The effect of hybrid transgression on environmental tolerance in experimental yeast crosses. *J Evol Biol.* 27(11):2507–2519.
- Stelkens RB, Miller EL, Greig D. 2016. Asynchronous spore germination in isogenic natural isolates of *Saccharomyces paradoxus*. *FEMS Yeast Res.* 16(3):fow012.
- Stelkens RB, Seehausen O. 2009. Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* 63(4):884–897.
- Tan Z, Hays M, Cromie GA, Jeffery EW, Scott AC, Ah Yong V, Sirt A, Skupin A, Dudley AM. 2013. Aneuploidy underlies a multicellular phenotypic switch. *Proc Natl Acad Sci U S A.* 110(30):12367–12372.
- von Janowsky B, Major T, Knapp K, Voos W. 2006. The disaggregation activity of the mitochondrial ClpB homolog Hsp78 maintains Hsp70 function during heat stress. *J Mol Biol.* 357(3):793–807.
- Willett CS. 2012. Hybrid breakdown weakens under thermal stress in population crosses of the copepod *Tigriopus californicus*. *J Hered.* 103(1):103–114.
- Yang F, Teoh F, Tan ASM, Cao Y, Pavelka N, Berman J. 2019. Aneuploidy enables cross-adaptation to unrelated drugs. *Mol Biol Evol.* 36(8):1768.
- Zörgö E, Gjuvsland A, Cubillos FA, Louis EJ, Liti G, Blomberg A, Omholt SW, Warringer J. 2012. Life history shapes trait heredity by accumulation of loss-of-function alleles in yeast. *Mol Biol Evol.* 29(7):1781–1789.

New Phytologist

December 2022

Vol. 236

No. 6

ISSN 0028-646X

eISSN 1469-8137

www.newphytologist.com

International Journal of Plant Science



How broad is the selfing syndrome? Insights from convergent evolution of gene expression across species and tissues in the *Capsella* genus

Zebin Zhang¹ , Dmytro Kryvokhyzha^{1,2} , Marion Orsucci^{1,3} , Sylvain Glémin^{1,4} , Pascal Milesi^{1,5*}  and Martin Lascoux^{1*} 

¹Program in Plant Ecology and Evolution, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden; ²Department of Clinical Sciences, Lund University Diabetes Centre, 214 28 Malmö, Sweden; ³Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala BioCenter, 750 07 Uppsala, Sweden; ⁴Université de Rennes, Centre National de la Recherche Scientifique (CNRS), ECOBIO (Ecosystèmes, Biodiversité, Evolution) – Unité Mixte de Recherche (UMR) 6553, F-35042 Rennes, France; ⁵Science For Life Laboratory (SciLifeLab), 752 37 Uppsala, Sweden

Summary

Authors for correspondence:
Martin Lascoux
Email: martin.lascoux@ebc.uu.se

Pascal Milesi
Email: pascal.milesi@scilifelab.uu.se

Received: 11 April 2022
Accepted: 1 September 2022

New Phytologist (2022)
doi: 10.1111/nph.18477

Key words: *Capsella*, convergent adaptation, gene expression, genomic and selfing syndromes, mating system change, relaxed selection.

- The shift from outcrossing to selfing is one of the main evolutionary transitions in plants. It is accompanied by profound effects on reproductive traits, the so-called selfing syndrome. Because the transition to selfing also implies deep genomic and ecological changes, one also expects to observe a genomic selfing syndrome.
- We took advantage of the three independent transitions from outcrossing to selfing in the *Capsella* genus to characterize the overall impact of mating system change on RNA expression, in flowers but also in leaves and roots. We quantified the extent of both selfing and genomic syndromes, and tested whether changes in expression corresponded to adaptation to selfing or to relaxed selection on traits that were constrained in outcrossers.
- Mating system change affected gene expression in all three tissues but more so in flowers than in roots and leaves. Gene expression in selfing species tended to converge in flowers but diverged in the two other tissues. Hence, convergent adaptation to selfing dominates in flowers, whereas genetic drift plays a more important role in leaves and roots.
- The effect of mating system transition is not limited to reproductive tissues and corresponds to both adaptation to selfing and relaxed selection on previously constrained traits.

Introduction

Mating system has a profound effect on organism evolution, as it strongly impacts both the level and structure of genomic variation and the ecology of a species (Charlesworth & Wright, 2001; Barrett, 2002; Campbell & Kessler, 2013; Cutter, 2019; Glémin *et al.*, 2019). Self-fertilizing species are characterized by higher homozygosity and lower effective recombination rate than outcrossing ones (Charlesworth & Wright, 2001). Their effective population size also is smaller, which leads to an overall reduction in selection efficiency although selection against recessive alleles will be stronger as a result of the higher level of homozygosity (Hamrick & Godt, 1996; Charlesworth & Wright, 2001; Charlesworth & Meagher, 2003). In the short term, selfing is beneficial because of its inherent transmission advantage, and reproductive assurance when mates and/or pollinators are scarce (Lande, 1988; Leimu & Fischer, 2008; Ellison *et al.*, 2011).

However, in the long term, the reduction in effective population size (N_e) associated to selfing leads to an accumulation of slightly deleterious mutations and may lower the adaptive capacity of the species. The shift in mating system also will be accompanied by major phenotypic changes (Wright *et al.*, 2013). In particular, self-fertilizing plant species no longer need to attract pollinators and are characterized by reduced flowers, and a low production of pollen, nectar and scent compared to their outcrossing relatives (Carr & Eubanks, 2002; Koslow & DeAngelis, 2006). Altogether, the most conspicuous changes associated with the transition from outcrossing to selfing have been called the *selfing syndrome* (Sicard & Lenhard, 2011; Rifkin *et al.*, 2019, 2021; Tsuchimatsu & Fujii, 2022), and there is evidence of convergent evolution across species, at least for floral traits (Woźniak *et al.*, 2020). Given the major impact of the shift from outcrossing to selfing on genetic diversity, population ecology and dynamics, one would expect the shift in mating system to also alter basic genomic processes such as gene expression (Thomas *et al.*, 2012; Cutter, 2019), and therefore to affect phenotypic traits beyond

*These authors contributed equally to this work.

those directly related to reproduction. In other words, in addition to a selfing syndrome related to reproductive traits (Woźniak *et al.*, 2020), a significant *genomic selfing syndrome* also is expected (Arunkumar *et al.*, 2016; Cutter, 2019; Wang *et al.*, 2021) and may include (1) reduced molecular diversity and increased linkage disequilibrium, (2) accumulation of deleterious mutations (Kryvokhyzha *et al.*, 2019a,b), (3) smaller genomes and reduced abundance or activity of transposable elements, and (4) greater structural chromosomal evolution (Cutter, 2019). The transition to selfing thus may also be expected to affect expression of genes globally. Here we shall investigate transcriptomic changes associated to selfing in three different tissues – flowers, roots and leaves. Patterns of gene expression are expected to be strongly altered in flowers (Woźniak *et al.*, 2020) but also in other tissues, although probably not as profoundly as in flower. Such a ‘broad’ transcriptomic selfing syndrome also is indirectly supported by recent developments in quantitative genetics and in our understanding of the evolution of gene regulation, which both indicate that mutations associated to changes in expression of a given gene are not limited to core genes (i.e. genes directly regulating the gene under study, nor genes known *a priori* to regulate those), but instead are found across the whole genome (Boyle *et al.*, 2017; Liu *et al.*, 2019; Hill *et al.*, 2021; Sinnott-Armstrong *et al.*, 2021). In other words, these results support the universal pleiotropy hypothesis that stems from Fisher’s geometric model of evolution (Fisher, 1930), or at least seem in line with the widespread pleiotropy predicted by the omnigenic model (Boyle *et al.*, 2017).

Generally, one would expect traits specifically related to reproduction to be more strongly affected by mating system shift than less related traits. Traits associated with the shift in mating system can evolve as a direct or indirect – in the case of correlated response – adaptation to the new conditions associated with self-fertilization, that can be viewed as a new phenotypic optimum, and/or because the selection pressure associated with outcrossing is relaxed (Cutter, 2019; Rifkin *et al.*, 2021; Tsuchimatsu & Fujii, 2022). Relaxed selection can take many forms (Lahti *et al.*, 2009). In the specific case of the shift in mating system, the effect of relaxed selection will be two-fold. First, because of the decrease in N_e associated to the shift in mating system, selection will be potentially less effective, especially for traits on which selection is intrinsically weak. Secondly, for some specific traits selection pressure will be alleviated independently of the decrease in N_e . For instance, nectar is no longer needed in selfers, and therefore selection on traits related to nectar production and display will be altered. Selection may initially be relaxed but positive selection for lower nectar production and display may follow in a second step because maintaining functions that are no longer required can represent substantial metabolic cost. Much less is known about the impact of change in mating system on RNA expression. If the shift in mating system is initially accompanied by relaxed selection, one would expect a more random pattern of expression than under purifying or positive selection. Most of the studies conducted so far have focused on changes in expression occurring in flowers (e.g. Woźniak *et al.*, 2020) or aimed at identifying genes underlying the change in mating system (Tsuchimatsu & Fujii, 2022),

and paid less attention to gene expression changes occurring in vegetative tissues (but see Frazee *et al.*, 2021). More generally, few studies have investigated the role of natural selection in mating system shift and fewer still have considered traits that were not *a priori* elements of the selfing syndrome.

The *Capsella* genus comprises one outcrossing species, *C. grandiflora*, and three self-fertilizing ones, *C. orientalis*, *C. rubella* and *C. bursa-pastoris*. *Capsella grandiflora*, *C. orientalis* and *C. rubella* are diploids, whereas *C. bursa-pastoris* is an allotetraploid. *Capsella orientalis* and *C. grandiflora* derived from a common ancestor some 1 million years ago, and *C. rubella* and *C. grandiflora* diverged c. 50 000 yr ago (Douglas *et al.*, 2015). *Capsella bursa-pastoris* originated from a hybrid between *C. grandiflora* and *C. orientalis* c. 100 000 yr ago (Fig. 1a; Douglas *et al.*, 2015). Hence, all three transitions to selfing occurred independently (Hurka *et al.*, 2012; Douglas *et al.*, 2015; Bachmann *et al.*, 2019), and in *Capsella*, as in other genera, the change from outcrossing to selfing is accompanied by a characteristic set of changes to the morphology and function of flowers, with evidence of convergent evolution of floral traits in *C. rubella* and *C. orientalis* (Woźniak *et al.*, 2020). In Kryvokhyzha *et al.* (2019a) we studied the expression changes in the allo-tetraploid *C. bursa-pastoris* and showed that the expression of each subgenome was biased towards the expression of their corresponding parental species, thereby demonstrating a conserved phylogenetic signal in expression across tissues. In addition to the phylogenetic signal, RNA expression also showed tissue-specific patterns. In flowers, expression was globally biased towards the selfing parent, *C. orientalis*, whereas in leaves and roots the expression was globally biased towards the outcrossing parent, *C. grandiflora*. We proposed that selection associated to the change in mating system drove the convergence in expression in flowers between the two selfing species (*C. bursa-pastoris* and *C. orientalis*), whereas a lower accumulation rate of deleterious mutation on the subgenome inherited from the outcrossing species, *C. grandiflora*, explained the bias in expression towards the outcrossing parental species in leaves and roots. In summary, the impact of selfing on both flower morphology (Woźniak *et al.*, 2020) and genetic load (Kryvokhyzha *et al.*, 2019a) have already been thoroughly described in *Capsella*, and in both cases there was evidence of a strong selfing syndrome. In the present study, we took advantage of the three independent transitions from outcrossing to selfing that occurred in the *Capsella* genus to further study the overall impact of mating system change on RNA expression, not only in flowers but also in leaves and roots. In particular, we quantified the extent of the genomic syndrome and tested whether changes in gene expression correspond to adaptation to selfing or to relaxed selection on genes that were involved in the control of traits constrained in outcrossers (Tsuchimatsu & Fujii, 2022). We focused our analysis on the three diploid species and only used the tetraploid self-fertilizing *C. bursa-pastoris* to confirm the results obtained with the diploid species. Finally, we carried out a GO analysis of differentially expressed genes between selfing and outcrossing species in order to facilitate comparison between our study and other studies on convergent evolution in selfing species and assess the breadth of the genomic selfing syndrome.

Materials and Methods

Samples and sequencing

In the present study, we re-used the RNA-seq data that were generated in Kryvokhyzha *et al.* (2019a) for three of the four species of the *Capsella* genus: *C. grandiflora* (Fauché & Chaub.) Boiss. (diploid, outcrosser), *C. orientalis* Klokov (diploid, selfer) and *C. bursa-pastoris* (L.) Medik. (allotetraploid, selfer). We added to this dataset, four accessions from *C. rubella* Reut. (diploid, selfer). These additional *C. rubella* accessions were part of the same experiment as accessions from the three other *Capsella* species. They were thus pre-processed simultaneously with those and

according to the same protocol. Briefly, seeds were surface-sterilized and sown into agar plates as described in Kryvokhyzha *et al.* (2019a,b). Following sterilization and germination, seedlings were transplanted into pots (10 × 10 × 10 cm) filled with soil 7 d after germination and cultivated in growth chamber (22°C, 16 h : 8 h, light : dark photoperiod, light intensity 150 μmol m⁻² s⁻¹). One week after the onset of flowering, flower buds, leaves and roots were collected. We collected closed green buds (without visible white petals) one week after the onset of flowering. We waited for a week after the onset of flowering to allow more buds to form and thus to have enough tissue for extraction. Tissues were immediately frozen in liquid nitrogen and stored at -80°C before RNA extraction. RNA was extracted

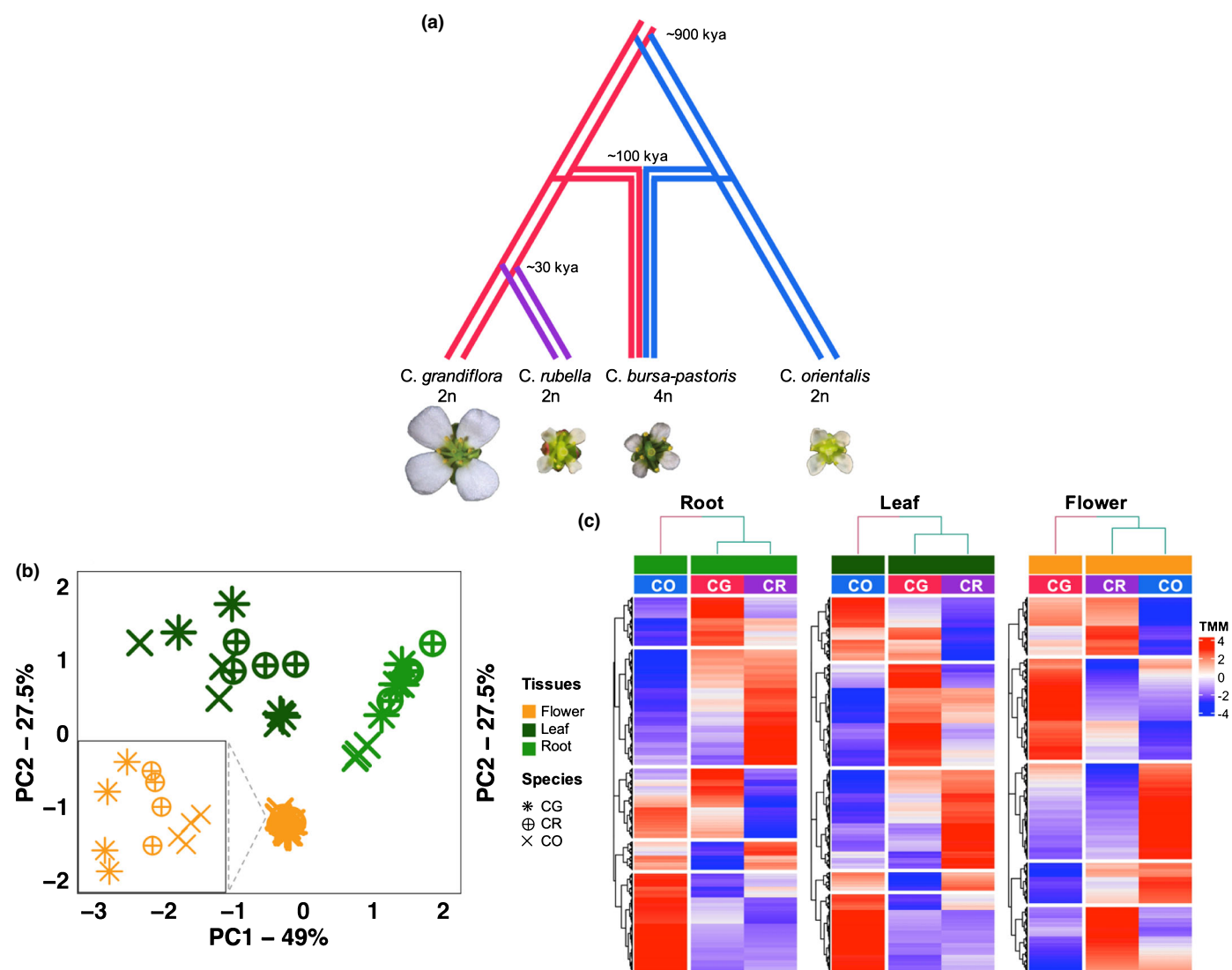


Fig. 1 Overall gene expression clusters by tissue, and by mating system in flower and phylogeny in root and leaf. (a) Evolutionary history of four species in *Capsella* genus. Species are labeled with specific colors: pink, *C. grandiflora* (CG); purple, *C. rubella* (CR); sky-blue, *C. orientalis* (CO). (b) Principal component analysis. Different tissues are represented by different colors and species by different shapes: orange, flower; dark green, leaf; light green, root. The zoomed-in plot for flowers is shown at the bottom left. (c) Heatmaps based on TMM normalized reads counts for 17 307 transcripts of *CR* (selfer), *CG* (outcrosser) and *CO* (selfer) for the three tissues: 'Root', left panels, 'Leaf', middle panels and 'Flower', right panels. For each heatmap, each row represents a transcript and each column represents the mean expression value of a species. The higher the expression value the darker the red, and conversely the lower the blue the lower the expression. The upper dendrogram shows the hierarchical relationships among species, and the row dendrogram shows the hierarchical relationships of gene expression for a specific tissue. Species are labeled with specific colors: pink, CG; purple, CR; sky-blue, CO.

from three different tissues: flowers, leaves and roots. Number of samples, accession names and sampling locations, and total library size per sample and per tissue, as well as the references of the sequences, are provided in Table S1. Total RNA of all four species was extracted and sequenced at the same time, following the same protocol (Kryvokhyzha *et al.*, 2019a), described in Methods S1. For all downstream analyses we filtered transcripts and considered a total of $n = 17\,307$ genes (see below).

Analysis of gene expression in diploid species.

Only samples from the three diploid species (*C. grandiflora*, *C. orientalis* and *C. rubella*) were used in the initial analyses to detect genes potentially involved in the transition from outcrossing to selfing. *Capsella bursa-pastoris* was not included at that stage in order to limit confounding effects of polyploidy and hybridization, but was used as a control later on (see below).

RNA expression variation pattern

Initially, in order to assess the overall gene expression pattern (e.g. across tissues, across species and between species with different mating systems) principal component analyses (PCAs) were performed on normalized reads count (TMM normalization, EDGER package (v.3.32.1; Robinson *et al.*, 2010) using the *prcomp* function, in R (R Core Team, 2013)). For any given gene, missing data were imputed as being the average expression value of each species and tissue. Then, for each tissue taken separately, more subtle variation in expression was analyzed using heatmaps (*Heatmap* function in R/COMPLEXHEATMAP (v.2.7.10; Gu *et al.*, 2016), with distance method 'euclidean' and clustering method 'ward.D').

Differential gene expression analysis

Differential gene expression analyses were conducted using TMM normalization in R/EDGER (Robinson *et al.*, 2010). We used strict criteria when trimming the transcripts. First, in order to be retained a transcript should have a nonzero expression in at least one sample per species, and this should be satisfied in all three tissues. Second, each transcript should have a nonzero expression in at least five samples in each tissue ($n = 17\,307$). In each comparison, genes differentially expressed between groups and with an FDR-adjusted P -value of 0.05 were considered as differentially expressed and hereafter called 'DE genes'.

Quantification of global effect of mating system transition on gene expression

We developed two indices to compare the relative distance in gene expression between different pairs of species and account for the effect of mating system change and phylogenetic inertia (Fig. 1a). Some pairs are phylogenetically distant but either share the same mating system (*C. rubella*, CR , and *C. orientalis*, CO , both selfers), or have different ones (*C. grandiflora*, CG , and *C. orientalis*, CO , respectively outcrosser and selfer), whereas

others are phylogenetically close but differ in mating system (*C. rubella* and *C. grandiflora*). Both indices are centered on 0 and constrained between -1 and 1 .

For each transcript i , the first distance index, D_{CR} was computed as the difference between the absolute difference in normalized expression between the two selfers (E_{iCO} and E_{iCR}) and the absolute differences in normalized expression between CG and CR (E_{iCG} and E_{iCR}).

$$D_{i,CR} = \frac{|E_{iCO} - E_{iCR}| - |E_{iCG} - E_{iCR}|}{\max(|E_{iCO} - E_{iCR}|, |E_{iCG} - E_{iCR}|)} \quad \text{Eqn 1}$$

$D_{i,CR} < 0$ indicates that the expression of a gene, i , in *C. rubella* is closer to the expression of the same gene in *C. orientalis* than in *C. grandiflora*. $D_{i,CR} > 0$ indicates that the expression of a given gene in *C. rubella* is closer to that in *C. grandiflora* than in *C. orientalis*. Finally, $D_{i,CR} \approx 0$ indicates that the expression in *C. rubella* is as distant from *C. orientalis* as it is from *C. grandiflora*.

Likewise, D_{CO} for transcript i is defined as:

$$D_{i,CO} = \frac{|E_{iCR} - E_{iCO}| - |E_{iCG} - E_{iCO}|}{\max(|E_{iCR} - E_{iCO}|, |E_{iCG} - E_{iCO}|)} \quad \text{Eqn 2}$$

$D_{i,CO} < 0$ indicates that CO expression is closer to that of CR than to that of CG , $D_{i,CO} > 0$ is the opposite (CO closer to CG than to CR), $D_{i,CO} \approx 0$ indicates that CO expression is equidistant from CG and CR .

In order to test whether the observed distributions of D_{CR} or D_{CO} could have been observed in the absence of selection we used simulations. The evolution of gene expression can be simulated as an Ornstein–Uhlenbeck (OU) process running on the phylogenetic tree (Felsenstein, 1988; Hansen, 1997; Nourmohammad *et al.*, 2017). An OU process can be described by the following stochastic differential equation, with parameters α , θ and σ :

$$dX = \alpha(\theta - X) + \sigma dB \quad \text{Eqn 3}$$

where θ is the optimum trait value, α is the strength of the force moving the trait towards the optimum (\sim selection) and dB is a Wiener process (aka Brownian motion) with variance σ (\sim drift). We used the function *rcOU* of R/SDE. For simplicity, we assume that the ancestral value is $X_0 = 0$ and that the optimum value also is $\theta = 0$. As we are interested in simulating the null hypothesis, we consider constant parameters along the phylogeny. When α tends towards 0, the process is equivalent to a Brownian motion, so expression is evolving under a pure random genetic drift model. If $\alpha > 0$, the process corresponds to uniform stabilizing selection along the phylogeny. In practice the parameter α cannot be set exactly to 0, so we used a very low value to simulate a Brownian motion. The evolution of 20 000 genes were simultaneously simulated. Details are given in Notes S1.

In order to further test for selection and, more specifically, assess whether the selfing syndrome involved adaptation and/or relaxed selection, we calculated the $\pi_N : \pi_S$ ratio for each transcript, where π_N is the nonsynonymous nucleotide diversity and π_S is the synonymous nucleotide diversity (Chen *et al.*, 2017).

π_N/π_S measures the efficiency of purifying selection: efficient purifying selection will lower π_N/π_S whereas weak selection will lead to high values. For each transcript in each tissue, the per-site π_S and π_N diversity in focal species were calculated using the 'dNdSpiNpiS' function, which is part of the software POPHYL, a specific pipeline designed to handle transcriptome-based NGS data (Cahais *et al.*, 2012). Details are provided in Methods S1. To minimize the variance between genes resulting from SNPs number differences between genes, π_N and π_S values first were weighted by the respective number of complete sites:

$$w\pi_{Ni} = ncs_i \times \pi_{Ni} \text{ and } w\pi_{Si} = ncs_i \times \pi_{Si}$$

where ncs_i is the number of complete sites used for π_N and π_S calculation in transcript i . Negative values of the D_{CO} index indicate convergence of expression in the self-fertilizing species, whereas positive values indicate less constraints. If the convergence is driven primarily by natural selection, then we would expect the $\pi_N : \pi_S$ ratio to be smaller for negative D_{CO} than for genes with positive values of the index and even more so in selfing species. The dataset thus was ordered according to increasing D_{CO} values and $w\pi_{Ni}$ and $w\pi_{Si}$ were summed over sliding windows of 100 transcripts. Average $\pi_N : \pi_S$ ratios then were computed for bins of $n = 10$ transcripts:

$$\mu\left(\frac{\pi_N}{\pi_S}\right) = \frac{1}{n} \times \sum_{i=1}^n w\pi_{Ni} / \sum_{i=1}^n w\pi_{Si}$$

Classification of differentially expressed genes and mating system transition gene identification

In order to identify differentially expressed genes associated to mating system shift (hereafter DEself genes), we first compared the whole gene expression dataset between pairs of species (*CR* vs *CG*, *CO* vs *CG*, and *CR* vs *CO*) for each tissue separately. For each transcript, in each comparison, we use the same method and threshold than for the identification of DE genes described above. We then compared DE gene lists across pairwise comparisons to define the set of mating system differentially expressed genes. A gene is DEself when it satisfied the two following criteria: (1) expression significantly differed between the outcrosser and both selfers but not between selfers (i.e. DE between *C. grandiflora* and *C. orientalis* and between *C. grandiflora* and *C. rubella*, but not between *C. orientalis* and *C. rubella*) and (2) for a given gene, this pattern is specific to flower tissue. Following Woźniak *et al.* (2020) co-differentially expressed genes (coDEGs) were defined as genes that are DE between each of the selfing species and the outcrosser, and that show changes in the same direction relative to the outcrosser.

In flower only, we further investigated genes putatively involved in mating system transition using the gene co-expression network analysis implemented in the 'Weighted correlation network analysis' R/WGCNA package (v.1.70-3, optimal soft threshold of 'power = 10'; Langfelder & Horvath, 2008). The cutreeDynamic function (minClusterSize = 30 &

deepSplit = 2) build in WGCNA (minClusterSize = 30 & deepSplit = 2) was used for tree trimming of the gene hierarchical clustering dendrograms to define co-expression modules. Modules with dissimilarity of module eigengenes (ME) lower than 0.25 were merged. The strength of association between co-expression modules and mating system was quantified using Pearson's correlation coefficient.

Genes expressed in flowers that are DEself and that belong to co-expression network associated with mating system were called mating system transition related genes, hereafter MST genes.

Gene ontology enrichment analysis

Gene ontology (GO) enrichment tests were performed using R/TOPGO (v.2.42; Alexa & Rahnenfuhrer, 2020). The GO term database for *C. rubella* was downloaded from PLANTREGMAP (<http://plantregmap.gao-lab.org>) and was used as reference for enrichment tests. We used a custom background list of genes that included only genes with a nonzero expression value in at least one sample per species and tissue ($n = 17\,307$). Significance of enrichment was tested using Fisher's exact-tests (FDR < 0.05) either for molecular functions (MF) or biological processes (BP) related GO terms. Finally, we used REVIGO (<http://revigo.irb.hr>; Supek *et al.*, 2011) with the *Arabidopsis thaliana* database to remove GO terms redundancy (medium, 0.7) and to cluster the remaining terms in a 2D space derived by applying multidimensional scaling to a matrix of the GO terms semantic similarities.

Ascertaining the global effect of mating system shift using the tetraploid selfer *C. bursa-pastoris*

Capsella bursa-pastoris (*Cbp*) originated from the hybridization between *C. orientalis* and *C. grandiflora* but retained both parental genomes, one inherited from the outcrosser (*Cbp_{CG}*) and one from the selfer (*Cbp_{CO}*). As an additional test of our hypothesis about the global effect of mating system transition on gene expression, we computed a similarity index (S_i) exactly as in Kryvokhyzha *et al.* (2019a) except that we replaced *C. grandiflora* with *C. rubella*. The S_i index quantifies the similarity between the expression level of each subgenome and the expression level in the 'parental' species. For each transcript i and each subgenome $j \in \{Cbp_{CG}, Cbp_{CO}\}$, S was computed as the subgenome relative expression deviation from the mean expression level in the 'parental' species (μ_i). In the present study, $\mu_i = (E_{iCO} + E_{iCG})/2$ as in Kryvokhyzha *et al.*'s (2019a) initial analysis, or $\mu_i = (E_{iCO} + E_{iCR})/2$ when *CG* was substituted by *CR*.

$$S_{ij} = \frac{E_{ij} - \mu_i}{\mu_i} \quad \text{Eqn 4}$$

where E_{ij} is the expression of a given transcript i , from a given genetic background j , (*Cbp_{CG}* or *Cbp_{CO}* subgenomes of *C. bursa-pastoris*). This index is centered and oriented to one parental species so that $S_{ij} > 0$ (respectively, $S_{ij} < 0$) indicates that the expression of transcript i is more similar (respectively, different) to the expression of parental species j .

Results

Global pattern of gene expression changes in diploids across three tissues

We used a PCA to explore the global pattern of gene expression variation ($n = 17\,307$) among the three diploid *Capsella* species (*CR*, *CG* and *CO*) and the three tissues (flowers, leaves and roots). Most of the variance in expression was between tissues (PC1 and PC2 together explained 76.5%; Fig. 1b). Consequently, tissues were analyzed separately in later analyses. Species clustered according to phylogenetic relationships in roots and leaves (i.e. *CR* and *CG* vs *CO*) and according to mating system in flowers (Figs 1c, S1). This difference in clustering was a consequence of both up- and downregulation in the selfing species (*CR* and *CO*) compared with the outcrossing species (block 2 and 4 in Fig. 1c). However, the shift to selfing was more often accompanied by downregulation (downregulated genes were at least twice as many as upregulated ones).

Disentangling phylogenetic signal from mating system transition effect on gene expression using distance indices

The first index, D_{CR} , measures the relative distance in expression between *CR* and *CO*, on the one hand, and *CR* and *CG*, on the other. As expected, the overall distribution of D_{CR} is biased towards 1 (*CR* closer to *CG* than to *CO*), indicating that phylogenetic relationships explain most of the variation in gene expression in all three tissues (median $D_{CR_flower} = 0.19$, $D_{CR_root} = 0.37$ and $D_{CR_leaf} = 0.37$). However, the distribution of D_{CR} in flowers differed from that in leaves and roots, and was characterized by an excess of negative values (Fig. 2a). For many genes, the expression between the two selfers (*CR* and *CO*) is thus closer than what would be expected given their phylogenetic relationships, suggesting the presence of convergent evolution in gene expression. The pattern was similar but more pronounced when D_{CR} index was calculated only for genes that are differentially expressed (DE, FDR < 0.05) between *CG* and *CO* (median $D_{CR_flower} = 0.42$, $D_{CR_leaf} = 0.69$ and $D_{CR_root} = 0.71$, Fig. 2c). The negative fraction of the distribution of the D_{CR} index in all three tissues implies that there are still a substantial number of genes with similar expression levels in *CR* and *CO* when only DE genes between *CG* and *CO* are considered (38.67%, 21.68% and 18.94%, respectively for flower, leaf and root).

The second index, D_{CO} , measures the relative distance in expression between *CR* and *CO*, on the one hand, and *CG* and *CO*, on the other, and allows controlling for the effect of phylogenetic distance on gene expression. The distribution of D_{CO} is slightly positively biased (*CO* closer to *CG* than to *CR*) in leaves and roots (median $D_{CO_leaf} = 0.05$ and $D_{CO_root} = 0.15$) and slightly negatively biased (*CO* closer to *CR* than to *CG*) in flowers (median $D_{CO_flower} = -0.09$). This indicates that in flowers constraints on gene expression evolution probably associated to mating system transition makes *CO* and *CR* more similar than expected, supporting the pattern of convergent evolution detected with D_{CR} . Interestingly, even though negative D_{CO}

values were less frequent in leaves and roots than in flowers, they are still far from negligible, indicating that a substantial number of genes in leaves and roots also were influenced by mating system transition through the genomic selfing syndrome. When we only considered DE genes between *CG* and *CR*, the distribution of D_{CO} was similar but the expression of *CO* is even closer to *CG* than to *CR* in leaves and roots (median $D_{CO_leaf} = 0.33$ and $D_{CO_root} = 0.43$) whereas, as expected, the expression of the two selfing species are even closer in flowers (median $D_{CO_flower} = -0.51$) (Fig. 2d). Note that considering only DE genes led to a larger increase in bias in flower (0.42) than in root and leaves (0.28 each).

In order to test whether the observed distributions of D_{CR} or D_{CO} could have been observed under a pure drift model, we simulated a OU process for 20 000 independent genes under a pure drift model. No combination of parameters in the OU simulations led to the observed distribution of D_{CR} or D_{CO} and departure from the expected distribution was the strongest in flowers (Notes S1). To test further whether the shift in expression associated with mating system transition was a consequence of adaptation and/or relaxed selection we compared the distribution of $\pi_N : \pi_S$ ratios between $D_{CO} \leq -0.5$ and $D_{CO} \geq 0.5$ (Fig. 3). As for the distribution of D_{CO} index described above, in flowers, $\pi_N : \pi_S$ ratios were higher for $D_{CO} \geq 0.5$ than for $D_{CO} \leq -0.5$, in particular for the two selfers. The same trend was observed in leaves although the differences in selfers were weaker and no specific trend was observed in roots. All of these analyses support a predominant role of natural selection in explaining convergence in expression in selfers and in particular in flowers.

Classification of differentially expressed genes

We then used pairwise comparisons between all three diploid species, to identify genes whose expression was most affected by mating system transition. In flower, there were more DE genes between *CG* and *CO* than in any other pairwise comparison, whereas in leaf and root there were more DE genes between *CO* and *CR* and, in particular, more than between *CO* and *CG* which are at the same phylogenetic distance (Fig. 4a–c). Also, the proportion of DE genes shared by the comparisons involving mating system transition, *CR/CG* and *CO/CG*, was four-fold larger than in leaves and roots. Therefore, gene expression in selfers tends to converge in flower but also tends to diverge in the two other tissues.

DEself genes are genes that are differentially expressed between the outcrossing species (*CG*) and each of the two selfers (*CO* and *CR*), but are not differentially expressed between the two selfers (Fig. 4a–c, colored sector). In agreement with results obtained with D_{CO} and D_{CR} , the number of DEself genes (786) was higher in flowers than in leaves (86) and roots (215). Overlap in DEself genes between the three tissues was limited, indicating tissue specific expression for these genes (Fig. 4d). Only 7% of the DE genes were classified as DEself in at least two tissues (1% in the three tissues). Finally, coDEGs are genes that are DE between each of the selfing species and the outcrosser, and that show changes in the same direction relative to the outcrosser (Fig. 4e).

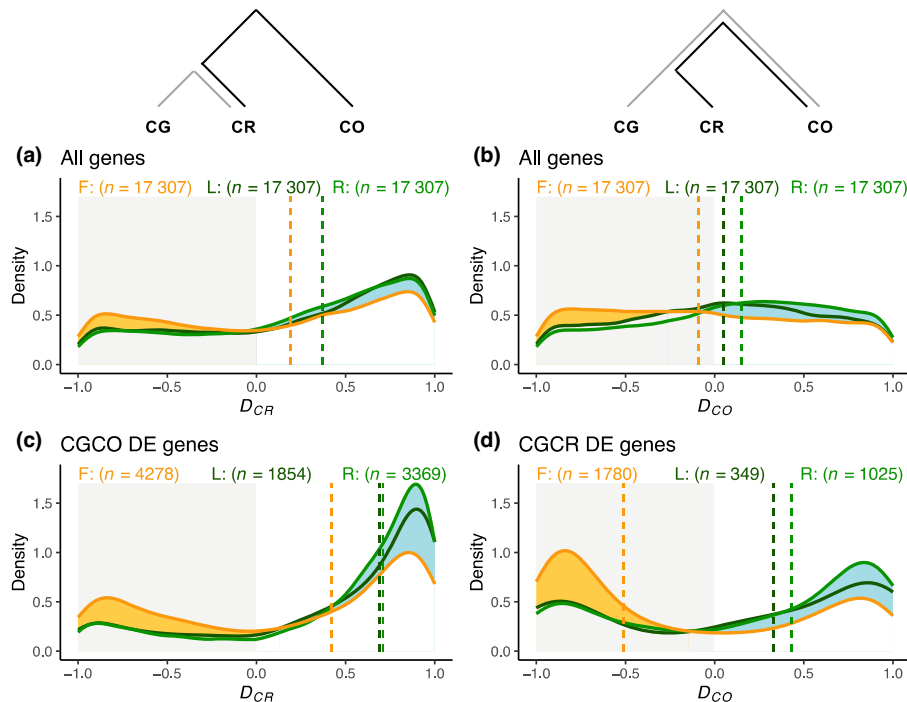


Fig. 2 Global effect of mating system transition and phylogeny on gene expression. (a, c) The D_{CR} index compares gene expression difference between *Capsella rubella* (CR) and *C. orientalis* (CO) (black line in the upper schematic plot), on the one hand, and the difference between *C. rubella* (CR) and *C. grandiflora* (CG) (gray line in the upper schematic plot), on the other (see Eqn 2 in the text). $D_{CR} < 0$ indicates that CR expression is closer to that of CO than to that of CG, $D_{CR} > 0$ is the opposite (CR closer to CG than to CO), $D_{CR} \approx 0$ indicates that CR expression is equidistant from CG and CO. Dotted lines are the median of the distribution of the D_{CR} index for the three tissues. (a) all genes; (c) differentially expressed (DE) genes between CG and CO. Both D_{CO} and D_{CR} value ranges from -1 to 1 . (b, d) The D_{CO} index compares gene expression difference between *C. orientalis* (CO) and *C. rubella* (CR) (CO–CR) (black line in the upper schematic plot), on the one hand, and the difference between *C. orientalis* (CO) and *C. grandiflora* (CG) (gray line in the upper schematic plot), on the other (see Eqn 1 in the text). $D_{CO} < 0$ indicates that CO expression is closer to that of CR than to that of CG, $D_{CO} > 0$ is the opposite (CO closer to CG than to CR), $D_{CO} \approx 0$ indicates that CO expression is equidistant from CG and CR. Flower, yellow; leaf, dark green; root, light green. Dotted lines are the median of the distribution of the D_{CO} index for the three tissues. (b) All genes; (d) DE genes between CG and CR.

The DE genes were much more common in flower than in leaves and roots (1061, 125 and 404, respectively) and the fraction of coDEG (i.e. both upregulated, coDEG⁺ or both downregulated, coDEG[−]) was significantly different between the three tissues (flower, 0.89, leaves, 0.8, and roots 0.68, binomial tests all $P < 0.05$). In flower coDEG[−] genes (64%) were significantly more common than coDEG⁺ (37%) (binomial test, $\chi^2 = 142$, df = 1, $P < 0.001$), whereas they were in similar proportion in leaves and roots (binomial test, $P = 0.32$ and $P = 0.55$, respectively; Fig. 4e). Hence, the pattern of coDEGs further supports higher convergence in expression in flower for genes associated to mating system transition and shows that it is driven mainly by downregulation of these genes.

Definition of mating system transition related genes and functional analysis

In order to narrow down the list of 786 DE genes potentially involved in mating system transition in flower, we conducted a weighted gene correlation network analysis (WGCNA v.1.70-3; Fig. S2). Three gene clusters ('modules' in WGCNA) showed a particularly strong association with mating system in flower (light green, 233 genes, Pearson's $r = 0.92$, $P = 3e^{-12}$; tan module, 467 genes, $r = 0.71$, $P = 3e^{-5}$; and turquoise module, 2070 genes,

$r = -0.84$, $P = 2e^{-8}$; Fig. S2). A total of 482 genes were detected using both approaches (Fig. S3; Table S2) and were defined, hereafter, as mating system transition genes (MST genes). Expression of these MST genes clustered according to mating system, the clustering being mainly due to genes that were downregulated in both selfing species (Fig. S3). We used a GO terms analysis to classify the MST genes according to BP and MF. For BP, there were 44 nonredundant GO terms ($P < 0.05$) and many were functionally related to pollen, fruit and anther development processes (Fig. S4; Table S3). At the MF level, MST genes clustered into 24 GO terms (Fig. S5; Table S4). One pathway seems particularly interesting as it relates to the process of L-ascorbic acid binding (GO:0031418).

Using the tetraploid self-fertilizing *C. bursa-pastoris* to further test the effect of mating system shift across tissues

In Kryvokhyzha *et al.* (2019a) we defined a similarity index, S , to quantify the shift in expression level of each subgenome in *Cbp* relative to the mean in expression level in the 'parental' species. In each tissue the expression of each subgenome was closest to its own 'parental' species – thus, there was a phylogenetic signal (S_{CO} and S_{CG} in Fig. 5a) – but the global expression was biased toward different parental species in the different tissue (ΔS in Fig. 5a).

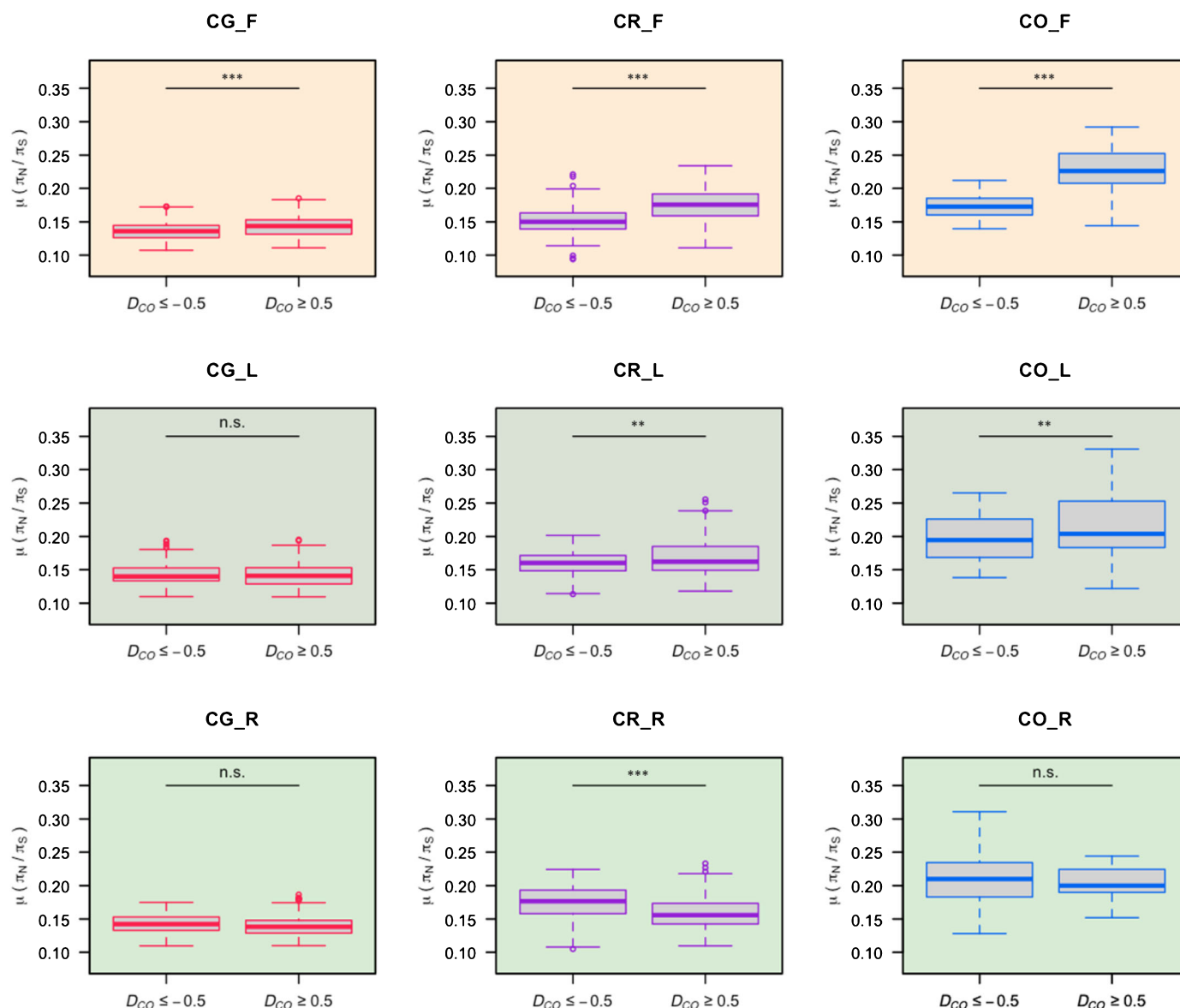


Fig. 3 $\pi_N : \pi_S$ ratios for different categories of D_{CO} index. The distribution of $\pi_N : \pi_S$ ratios is compared between $D_{CO} \leq -0.5$ and $D_{CO} \geq 0.5$ for the different species (*Capsella grandiflora*, *CG*, left column; *C. rubella*, *CR*, middle column; *C. orientalis*, *CO*, right column) and for the different tissues (flowers, top row; leaves, middle row; roots, bottom row). Within each box plot, the bold horizontal line represents the median value; box region means values within interquartile range (IQR) from first quartile (Q1) to third quartile (Q3); up and down whiskers indicate 1.5 IQR above the Q3 ($Q3 + 1.5$ IQR) and 1.5 IQR below the Q1 ($Q1 - 1.5$ IQR), separately; circles representative outliers. Wilcoxon test: ns, $P > 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

In flowers, the pattern of expression of the subgenome inherited from the selfer *CO* (*Cbp_{CO}*) tended to dominate that of the subgenome inherited from the outcrosser *CG* (*Cbp_{CG}*), leading to a bias towards the expression of the self-fertilizing parent *CO* ($\Delta S > 0$). In leaves, and even more so in roots, we observed the opposite pattern where the subgenome *Cbp_{CO}* tended to be dominated by the subgenome *Cbp_{CG}* ($\Delta S < 0$), the bias in expression thus being towards the outcrossing parent. Given that *Cbp* is self-fertilizing, this pattern was interpreted as follows: in tissues more closely associated to the mating system transition the *CO*-like expression pattern sets the standard, whereas in tissues less associated to mating system but more closely related to vigor and to the overall divergence between the two parental species, the *CG* pattern

dominates. If this is indeed the case, replacing *CG* by *CR* as a 'parental' species should remove the mating system effect without affecting the general vigor and phylogeny effects. We would thus expect to see a drastic change in *S* in flowers, from a bias toward *CO* to no bias if mating system of the parental species were the only factor explaining the relative expression between the two subgenomes. For leaves and roots much less change in *S* index is expected, although one would expect a slightly higher bias toward *CR* due to genes associated to selfing in these tissues. Replacing *CG* by *CR* led to a negative ΔS and a shift towards *CR* in all three tissues (Fig. 5b). However, the bias of ΔS away from 0 was smallest in flowers ($\Delta S = -0.084$) indicating that flowers were less influenced by the phylogenetic signal than leaves ($\Delta S = -0.128$)

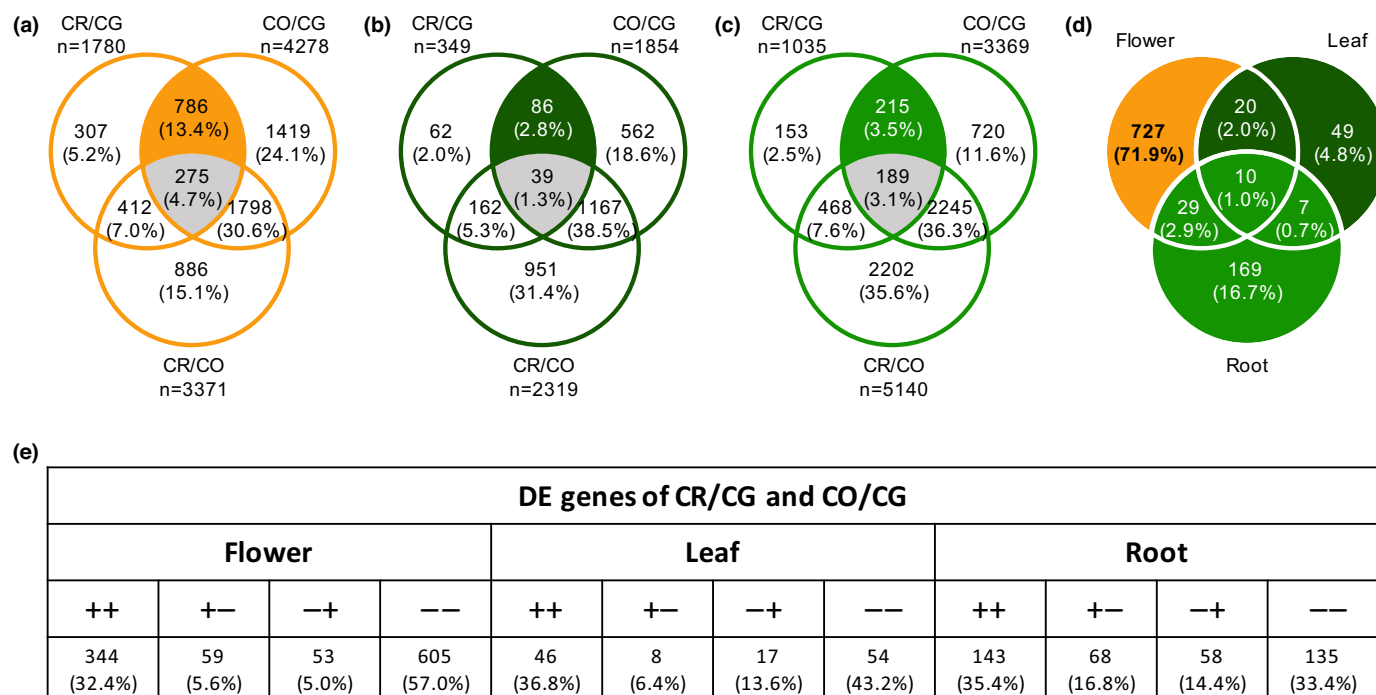


Fig. 4 Differentially expressed (DE) genes analysis and mating system transition (MST) related candidate genes identification. (a) Venn diagrams showing the number of detected DE (FDR < 0.05) genes in flower from groups of species pairwise comparison (i.e. between CR and CG, CO and CG, CR and CO), as well as overlaps between groups in number and in percentage. The orange highlighted overlap represents co-differentially-expressed genes (coDEGs) of CR and CO in flower. The coDEGs of CR and CO were defined as being differentially expressed between CG and CR and between CG and CO, but not between CR and CO. The gray highlighted overlap represents the coDEGs of CG and CO in flower. (b) In leaf (dark green), the same role as (a). (c) In root (light green), the same role as (a). (d) Identification of MST-related candidate genes. Each coDEGs of CR and CO should be flower-specific (orange highlighted) to be an MST-related candidate. (e) Summary of DE genes of CR/CG and CO/CG across tissues. '++' and '--' means genes either upregulated in both CR/CG and CO/CG, or downregulated in both comparisons. '+-' and '-+' means genes upregulated in one comparison but downregulated in another. CG, *Capsella grandiflora*; CO, *C. orientalis*; CR, *C. rubella*.

and roots ($\Delta S = -0.197$) or, conversely, more strongly affected by the mating system than other tissues.

Discussion

In the present study, we analyzed changes in gene expression in flowers, leaves and roots between self-fertilizing *Capsella* species and their sole outcrossing relative, *C. grandiflora*. Our results show that gene expression was more affected in flower tissues than in leaves and roots and that there was an excess of convergent genes between selfing species (CO and CR) in flowers compared to roots and leaves. However, leaves and roots also were affected by mating system shift. In flowers, many changes were convergent across selfing species and adaptive, whereas changes in roots and leaves were more random and relaxed selection probably explained some of them.

This is not the first study of genome-wide expression in the *Capsella* genus although it is the first to look at the nature of the selective change associated to the mating system transition and include all four species and three tissues. Slotte *et al.* (2013) identified DE genes between *C. rubella* and *C. grandiflora* in flower tissues, and compared those to DE genes between *A. lyrata* and *A. thaliana*, noting some overlap suggesting some convergent evolution in flower. Combining transcriptomic and gene mapping in *C. rubella* and *C. orientalis*, Woźniak *et al.* (2020)

showed that petal size had a similar genetic basis in the two species. We compared our list of MST genes with the list of convergent genes between *C. rubella* and *A. thaliana* established by Slotte *et al.* (2013), on the one hand, and with the list of coDEGs between *C. rubella* and *C. orientalis* from Woźniak *et al.* (2020), on the other. Because Slotte *et al.* (2013) list of MST genes was unavailable, we first used their gene expression data to establish two MST gene lists using two different approaches (by permutation and using R/LIMMA). Of our list of 482 MST genes, 218 overlapped with at least one of the two previous studies (Fig. S6).

These 218 overlapped genes belonged to pathways functionally related to pollen development, and more precisely pollen exine formation (GO:0010584) (Tables S5, S6). In general, our results correspond to the observation in Slotte *et al.* (2013) that many DE genes between *C. grandiflora* and *C. rubella* are functionally related with pollen number differences. However, Woźniak *et al.* (2020) did not report such pollen-related pathways, even though they observed a dramatic decrease (around four-fold) of the number of pollen grain in both selfing species compared to *C. grandiflora*. Another interesting discrepancy is that both Slotte *et al.* (2013) and Woźniak *et al.* (2020) detected many genes associated to petal development but we did not. One potential reason is difference in sample and sampling time across these three studies. In Slotte *et al.* (2013) total RNA was harvested

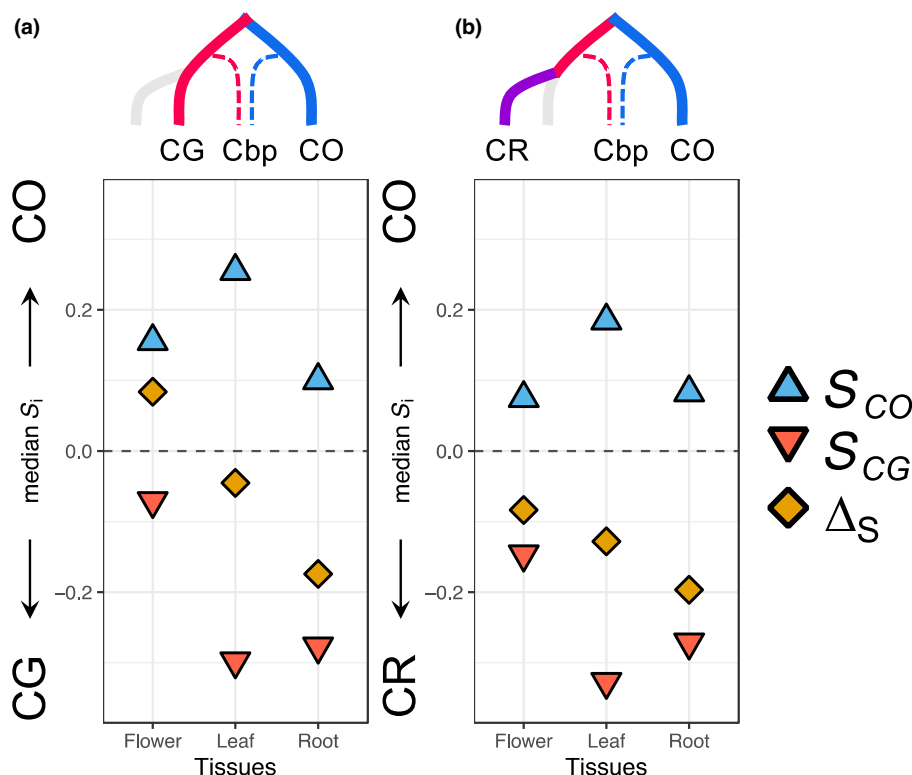


Fig. 5 Similarity indices of subgenomes of *Capsella bursa-pastoris* (Cbp) to different 'parental' species. In a given pair of diploid species, for each tissue and each subgenome, the median of similarity indices for each subgenome, either inherited from CO (S_{CO}) or from CG (S_{CG}), are presented as well as the difference between the two indices (ΔS) (see the **Materials and Methods** section for definitions). ΔS measures the dominance of one parental genetic background. Gray dotted line ($S = 0$) corresponds to an absence of bias. Phylogenetic trees on the top of each panel represent the phylogenetic relationships of Cbp with the diploid species used for the comparison. CG, *Capsella grandiflora*; CO, *C. orientalis*; CR, *C. rubella*.

from mixed flower buds from five *C. grandiflora* genotypes and six *C. rubella* genotypes, and the expression patterns in *Capsella* compared with expression patterns of stage-12 floral buds of *A. thaliana* and *A. lyrata*. In Woźniak *et al.* (2020) total RNA was extracted from dividing flower buds (so before stage 10 according to Bowman *et al.*'s (1991) development classification), and expanding/maturing flower buds. In the present study, buds were collected one week after the onset of flowering. Another possible explanation is that we used all four species, including both diploid, and three tissues as well as two different methods for detecting genes related with mating system transition, whose criteria are stricter than those used by Slotte *et al.* (2013) or Woźniak *et al.* (2020).

The classification of the MST genes based on GO showed that many terms are functionally related to pollen, fruit and anther development processes. Pollen is critical for plant reproductive success and is predicted to be under strong selective force in outcrossing species as a result of the pervasive male–male gamete competition, resulting in the outcrossing species being prone to generating a larger number of pollen grains (Harvey & May, 1989). By contrast, selfing species generally show a lower number of pollen grains, because fertilization is easier to complete in a selfing species than in its outcrossing relatives (Willis, 1999; Cruden, 2000; Shimizu & Tsuchimatsu, 2015; Liao *et al.*, 2022). Because the selfing species does not need to attract pollinators, some rewarding products for pollinators, such as scents, nectars and feeding anthers, that are costly for the plant to produce, will no longer be needed and thus will be under relaxed selection (Smith, 2016; Wessinger & Hileman, 2016; Liao *et al.*, 2022). Selfing species also exhibit an early mature time of

pollen development (de Jong *et al.*, 1993) and a lower amount of fruit set (Bellusci *et al.*, 2009; Jacquemyn & Brys, 2015). The fact that MST genes clustered into categories functional related to pollen sperm cell differentiation (GO:0048235), fruit development (GO:0010154) and anther wall tapetum development (GO:0048658) indicate that MST genes have a broad effect on reproductive functions at different stages of the life cycle.

When genes are classified according to MF, the most interesting seems to be the process of L-ascorbic acid binding. L-ascorbic acid appears highly pleiotropic and plays a role in seed germination, floral induction and senescence (Barth *et al.*, 2006). The antagonistic role of ascorbic acid in regulating seed germination has been well-documented (Koornneef *et al.*, 2002; Finch-Savage *et al.*, 2006). In many plant species, underexpression of ascorbic acid leads to enhanced germination potential and sometimes produce viviparous seeds (McCarty, 1995; Salaita *et al.*, 2005), whereas overexpression of ascorbic acid results in enhanced dormancy or delayed germination (Thompson *et al.*, 2000; Qin & Zeevaert, 2002; Okamoto *et al.*, 2010). In general, there is evidence that mating system influences phenology and, in particular, seed phenology and germination (e.g. Carta *et al.*, 2015; Baskin & Baskin, 2017). Hence, as for GO, the classification according to molecular function points at changes in developmental timing between selfers and outcrossers.

Understandably, studies of the selfing syndrome have largely focused on floral and reproductive traits and genes associated to traits such as the reduction in flower size have been identified as part of the selfing syndrome (e.g. Woźniak *et al.*, 2020). These results generally were obtained through quantitative trait loci (QTL) studies focusing on targeted traits and therefore

intrinsically ignoring the possibility of an impact of mating system shift beyond floral and reproductive traits. However, there are many reasons to expect changes to extend beyond those traits (Tsuchimatsu & Fujii, 2022). Selfing species differ from outcrossing ones in many factors that directly impact the rate of evolution: the effective population size is divided by two and the effective recombination rate is reduced. Thus, we can expect the genomic selfing syndrome to affect all genes and not only those directly related to selfing. We showed that the expression of a large fraction of genes deviated from what would be expected considering the phylogenetic relationships between the three species in all three tissues. This pattern is not explained by some ubiquitous genes given the small overlap that we found between the three tissues.

One central question is whether selfing syndrome trait evolution arises from relaxed selection (random genetic drift), from adaptive re-allocation of resources or from a mixture of both (Cutter, 2019; Rifkin *et al.*, 2021; Tsuchimatsu & Fujii, 2022). As pointed out by Cutter (2019), deciphering the relative contribution of adaptive changes vs relaxed selection for any trait individually remains challenging and few studies have attempted it. In morning glory (*Ipomoea lacunosa*), Rifkin *et al.* (2021) used Q_{ST} – F_{ST} comparisons to identify selection associated to the divergence between selfers and outcrossers on traits ranging from corolla size to early growth. Divergence between selfers and outcrossers was found to be under selection for corolla size and nectar traits, but not for early growth or pollen traits. The authors concluded that some aspects of the selfing syndrome were under selection but evolution in others was either due to drift or to correlated selection. The design of our study also allowed insights on the relative importance of these two evolutionary processes. First, the availability of three tissues, one closely related to reproduction and two others more distantly related to it means that we shall be able to capture both the selfing syndrome *sensu stricto* via flowers and the genomic selfing syndrome via comparison of expression across the three tissues. Second, the availability of three independent mating system transitions implies that we can use the phylogeny to derive expectations and identify genes that are more likely to be under directional selection than under relaxed selection.

Combining both allowed us to show that the observed pattern in gene expression cannot be obtained under a strictly neutral model, but instead is the result of a mixture of convergent adaptive changes and relaxed selection. Support for convergent evolution between the selfing species was particularly pronounced in flowers (both D_{CO} and D_{CR} indices <0), but also was observed in the two other tissues. Convergent evolution can be the result of either adaptation to a same new optimum in gene expression in selfing species or a change in the same direction in expression following relaxed selection. The later could create an enrichment for negative D index values (thereby mimicking adaptive convergence) if a consistent bias toward either up- or downregulation existed for genes evolving through drift, once the selection has been relaxed. Though we cannot completely rule out such a hypothesis, further observations support the role of positive selection in driving the pattern of convergent evolution in flower.

First, the enrichment for positive values in the D_{CO} index in leaves and roots indicates that gene expression evolved at a relatively higher rate on the branches leading to CR and CG . This could, for instance, occur if selection on CR is relaxed after mating system shift. Differentially expressed gene analyses also revealed that in leaves and roots many more genes showed a difference in expression between the two selfing species, CO and CR , than in any other comparisons, and in particular between CG and CO that are at the same phylogenetic distance. If changes in expression were to be biased towards either up- or downregulation after the selection is relaxed, such a pattern would not be observed. By contrast, the expression in the two self-fertilizing species tended to drift apart showing that drift following relaxed selection can lead to both up- or downregulation without preference. This, in turn, further supports the role of selection in shaping the convergence in expression in flower. Second, the presence of convergent evolution is strongly supported by the fraction of coDEGs among DE genes in the three tissues (i.e. genes whose expression in CO and CR compared to CG is in the same direction): *c.* 90% in flowers, *c.* 80% in leaves and *c.* 70% in roots (Fig. 3). In leaves and roots, the coDEGs are evenly distributed between coDEG[−] (co-downregulated genes) and coDEG⁺ (co-upregulated genes). However, there is significantly more coDEG[−] than coDEG⁺ in flower (64%) further supporting the role of convergent adaptation in flower. Third, we used the tetraploid selfing species *C. bursa-pastoris*, to test and confirm our predictions about the respective role of natural selection and genetic drift in the evolution of RNA expression in the various tissues. Concretely, one would not expect different patterns of S_i index between the three tissues, as was observed, unless the relative importance of selection and drift in shaping RNA expression changed between them. Finally, the global pattern of variation of $\pi_N : \pi_S$ ratio with the D_{CO} index further supports the respective roles of selection and drift in shaping pattern of expression following mating system change. The ratio π_N/π_S measures selection efficacy, and is expected to be small when purifying selection is effective and to increase as drift becomes more potent (Chen *et al.*, 2017). In our case, this suggests a more important role for drift in shaping RNA expression in leaves and roots in selfing species, and for selection for the genes showing a convergent pattern of RNA expression between selfing species, mainly in flowers.

Acknowledgements

The study was supported by grant 2019-00806 from the Swedish Research Council to ML. ML also acknowledges support from the Erik Philip-Sörensen Foundation. The computation and data handling were provided by the Swedish National Infrastructure for Computing (SNIC) at Uppmax, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

Author contributions

ML, PM and SG designed the study; ZZ and PM analyzed the data; MO and DK performed experiments; and ZZ, PM, SG and ML drafted the manuscript. All authors read and approved the

final version of the manuscript. PM and ML contributed equally to this work.

ORCID

Sylvain Glémin  <https://orcid.org/0000-0001-7260-4573>
Dmytro Kryvokhyzha  <https://orcid.org/0000-0001-6498-1977>
Martin Lascoux  <https://orcid.org/0000-0003-1699-9042>
Pascal Milesi  <https://orcid.org/0000-0001-8580-4291>
Marion Orsucci  <https://orcid.org/0000-0001-8516-1361>
Zebin Zhang  <https://orcid.org/0000-0002-9828-4245>

Data availability

All data are available and were deposited. The sequence references are given in Table S1. SRR8944173–SRR8944194 are from Kryvokhyzha *et al.* (2019a) and SRR18361966–SRR18361976 are newly sequenced.

References

- Alexa A, Rahnenfuhrer J. 2020. *TOPGO: enrichment analysis for gene ontology 2.40.0*. R package v.2(0) [WWW document] URL <http://www.bioconductor.org/packages/release/bioc/html/topGO.html>.
- Arunkumar R, Maddison TI, Barrett SCH, Wright SI. 2016. Recent mating-system evolution in *Eichhornia* is accompanied by cis-regulatory divergence. *New Phytologist* 211: 697–707.
- Bachmann JA, Tedder A, Laenen B, Fracassetti M, Désamoris A, Lafon-Placette C, Steige KA, Callot C, Marande W, Neuffer B *et al.* 2019. Genetic basis and timing of a major mating system shift in *Capsella*. *New Phytologist* 224: 505–517.
- Barrett SCH. 2002. The evolution of plant sexual diversity. *Nature Reviews. Genetics* 3: 274–284.
- Barth C, Tullio MD, Conklin PL. 2006. The role of ascorbic acid in the control of flowering time and the onset of senescence. *Journal of Experimental Botany* 57: 1657–1665.
- Baskin JM, Baskin CC. 2017. Seed germination in cleistogamous species: theoretical considerations and a literature survey of experimental results. *Seed Science Research* 27: 84–98.
- Bellusci F, Pellegrino G, Musacchio A. 2009. Different levels of inbreeding depression between outcrossing and selfing *Serapias* species. *Biologia Plantarum* 53: 175–178.
- Bowman JL, Drews GN, Meyerowitz EM. 1991. Expression of the *Arabidopsis* floral homeotic gene AGAMOUS is restricted to specific cell types late in flower development. *Plant Cell* 3: 749–758.
- Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169: 1177–1186.
- Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, Weinert L, Chiari Y, Belkhir K, Ranwez V, Galtier N. 2012. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources* 12: 834–845.
- Campbell SA, Kessler A. 2013. Plant mating system transitions drive the macroevolution of defense strategies. *Proceedings of the National Academy of Sciences, USA* 110: 3973–3978.
- Carr DE, Eubanks MD. 2002. Inbreeding alters resistance to insect herbivory and host plant quality in *Mimulus guttatus* (Scrophulariaceae). *Evolution* 56: 22–30.
- Carta A, Bedini G, Giannotti A, Savio L, Peruzzi L. 2015. Mating system modulates degree of seed dormancy in *Hypericum elodes* L. (Hypericaceae). *Seed Science Research* 25: 299–305.
- Charlesworth D, Meagher TR. 2003. Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358: 1051–1070.
- Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics & Development* 11: 685–690.
- Chen J, Glémin S, Lascoux M. 2017. Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution* 34: 1417–1428.
- Cruden RW. 2000. Pollen grains: why so many? *Plant Systematics and Evolution* 222: 143–165.
- Cutter AD. 2019. Reproductive transitions in plants and animals: selfing syndrome, sexual selection and speciation. *New Phytologist* 224: 1080–1094.
- Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R, Ågren JA, Hazzouri KM, Wang W *et al.* 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proceedings of the National Academy of Sciences, USA* 112: 2806–2811.
- Ellison A, Cable J, Consuegra S. 2011. Best of both worlds? Association between outcrossing and parasite loads in a selfing fish. *Evolution* 65: 3021–3026.
- Felsenstein J. 1988. phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* 19: 445–471.
- Finch-Savage WE, Leubner-Metzger G. 2006. Seed dormancy and the control of germination. *New Phytologist* 171: 501–523.
- Fisher R. 1930. *The genetical theory of natural selection*. Oxford, UK: The Clarendon Press.
- Fraze LJ, Rifkin J, Maheepala DC, Grant A-G, Wright S, Kalisz S, Litt A, Spigler R. 2021. New genomic resources and comparative analyses reveal differences in floral gene expression in selfing and outcrossing *Collinsia* sister species. *G3: Genes, Genomes, Genetics* 11: jkab177.
- Glémin S, François CM, Galtier N. 2019. Genome evolution in outcrossing vs. selfing vs. asexual species. In: Anisimova M, ed. *Evolutionary genomics*. New York, NY: Humana, 1910: 331–369.
- Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32: 2847–2849.
- Hamrick JL, Godt MJW. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351: 1291–1298.
- Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51: 1341–1351.
- Harvey PH, May RM. 1989. Out for the sperm count. *Nature* 337: 508–509.
- Hill MS, Zande PV, Wittkopp PJ. 2021. Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews. Genetics* 22: 203–215.
- Hurka H, Friesen N, German DA, Franzke A, Neuffer B. 2012. “Missing link” species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae). *Molecular Ecology* 21: 1223–1238.
- Jacquemyn H, Brys R. 2015. Pollen limitation and the contribution of autonomous selfing to fruit and seed set in a rewarding orchid. *American Journal of Botany* 102: 67–72.
- de Jong TJ, Waser NM, Klinkhamer PG. 1993. Geitonogamy: the neglected side of selfing. *Trends in Ecology & Evolution* 18: 321–325.
- Koornneef M, Bentsink L, Hilhorst H. 2002. Seed dormancy and germination. *Current Opinion in Plant Biology* 5: 33–36.
- Koslow JM, DeAngelis DL. 2006. Host mating system and the prevalence of disease in a plant population. *Proceedings of the Royal Society B: Biological Sciences* 273: 1825–1831.
- Kryvokhyzha D, Milesi P, Duan T, Orsucci M, Wright SI, Glémin S, Lascoux M. 2019a. Towards the new normal: transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (*Capsella bursa-pastoris*). *PLoS Genetics* 15: e1008131.
- Kryvokhyzha D, Salcedo A, Eriksson MC, Duan T, Tawari N, Chen J, Guerrina M, Kreiner JM, Kent TV, Lagercrantz U *et al.* 2019b. Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid *Capsella bursa-pastoris* (Brassicaceae). *PLoS Genetics* 15: e1007949.
- Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K, Foster SA. 2009. Relaxed selection in the wild. *Trends in Ecology & Evolution* 24: 487–496.
- Lande R. 1988. Genetics and demography in biological conservation. *Science* 241: 1455–1460.

- Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9: 559.
- Leimu R, Fischer MA. 2008. Meta-analysis of local adaptation in plants. *PLoS ONE* 3: 1–8.
- Liao IT, Rifkin JL, Cao G, Rausher MD. 2022. Modularity and selection of nectar traits in the evolution of the selfing syndrome in *Ipomoea lacunosa* (Convolvulaceae). *New Phytologist* 233: 1505–1519.
- Liu X, Li YI, Pritchard JK. 2019. Trans effects on gene expression can drive omnigenic inheritance. *Cell* 177: 1022–1034.
- McCarty DR. 1995. Genetic control and integration of maturation and germination pathways in seed development. *Annual Review of Plant Biology* 46: 71–93.
- Nourmohammad A, Rambeau J, Held T, Kovacova V, Berg J, Lässig M. 2017. Adaptive evolution of gene expression in *Drosophila*. *Cell Reports* 20: 1385–1395.
- Okamoto M, Tatematsu K, Matsui A, Morosawa T, Ishida J, Tanaka M, Endo TA, Mochizuki Y, Toyoda T, Kamiya Y *et al.* 2010. Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of *Arabidopsis* using tiling arrays. *The Plant Journal* 62: 39–51.
- Qin X, Zeevaert JA. 2002. Overexpression of a 9-cis-epoxycarotenoid dioxygenase gene in *Nicotiana plumbaginifolia* increases abscisic acid and phaseic acid levels and enhances drought tolerance. *Plant Physiology* 128: 544–551.
- R Core Team. 2013. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rifkin JL, Cao G, Rausher MD. 2021. Genetic architecture of divergence: the selfing syndrome in *Ipomoea lacunosa*. *American Journal of Botany* 108: 2038–2054.
- Rifkin JL, Liao IT, Castillo AS, Rausher MD. 2019. Multiple aspects of the selfing syndrome of the morning glory *Ipomoea lacunosa* evolved in response to selection: a Q_{st} – F_{st} comparison. *Ecology and Evolution* 9: 7712–7725.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
- Salaita L, Kar RK, Majee M, Downie AB. 2005. Identification and characterization of mutants capable of rapid seed germination at 10 °C from activation-tagged lines of *Arabidopsis thaliana*. *Journal of Experimental Botany* 56: 2059–2069.
- Shimizu KK, Tsuchimatsu T. 2015. Evolution of selfing: recurrent patterns in molecular adaptation. *Annual Review of Ecology, Evolution, and Systematics* 46: 593–622.
- Sicard A, Lenhard M. 2011. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany* 107: 1433–1443.
- Sinnott-Armstrong N, Naqvi S, Rivas M, Pritchard JK. 2021. GWAS of three molecular traits highlights core genes and pathways alongside a highly polygenic background. *eLife* 10: e58615.
- Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Newman LK *et al.* 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics* 45: 831–835.
- Smith SD. 2016. Pleiotropy and the evolution of floral integration. *New Phytologist* 209: 80–85.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* 6: e21800.
- Thomas CG, Li R, Smith HE, Woodruff GC, Oliver B, Haag ES. 2012. Simplification and desexualization of gene expression in self-fertile nematodes. *Current Biology* 22: 2167–2172.
- Thompson AJ, Jackson AC, Symonds RC, Mulholland BJ, Dadswell AR, Blake PS, Burbidge A, Taylor IB. 2000. Ectopic expression of a tomato 9-cis-epoxycarotenoid dioxygenase gene causes over-production of abscisic acid. *The Plant Journal* 23: 363–374.
- Tsuchimatsu T, Fujii S. 2022. The selfing syndrome and beyond: diverse evolutionary consequences of mating system transitions in plants. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 377: 20200510.
- Wang X-J, Barrett SCH, Zhong L, Wu Z-K, Li D-Z, Wang H, Zhou W. 2021. The genomic selfing syndrome accompanies the evolutionary breakdown of heterostyly. *Molecular Biology and Evolution* 38: 168–180.
- Wessinger CA, Hileman LC. 2016. Accessibility, constraint, and repetition in adaptive floral evolution. *Developmental Biology* 419: 175–183.
- Willis JH. 1999. The contribution of male-sterility mutations to inbreeding depression in *Mimulus guttatus*. *Heredity* 83: 337–346.
- Woźniak NJ, Kappel C, Marona C, Altschmied L, Neuffer B, Sicard A. 2020. A similar genetic architecture underlies the convergent evolution of the selfing syndrome in *Capsella*. *Plant Cell* 32: 935–949.
- Wright SI, Kalisz S, Slotte T. 2013. Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B: Biological Sciences* 280: 20130133.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Heatmap of 17 307 transcripts based on individual TMM normalized reads count for three *Capsella* species (CG, CR, CO) across tissues.

Fig. S2 Weighted gene correlation network analysis (WGCNA) revealed three modules highly related with mating system transition traits in *Capsella*.

Fig. S3 Putative selfing syndrome related genes identification and most of them were downregulated in selfers.

Fig. S4 Nonredundancy GO terms of selfing syndrome related genes in biological process category.

Fig. S5 Nonredundancy GO terms of selfing syndrome related genes in molecular function category.

Fig. S6 MST related genes overlapping with MST genes detected by Slotte *et al.* (2013) and CR-CO co-convergency genes from Woźniak *et al.* (2020).

Methods S1 Supplementary methods.

Notes S1 Simulation of Ornstein–Uhlenbeck processes for *Capsella* RNA seq evolution.

Table S1 Accessions used in the present study together with RNASEQ library size for the different tissues together with their SRA reference numbers.

Table S2 Reputed *Capsella* MST-related gene list, and overlaps with Slotte *et al.* (2013) and Woźniak *et al.* (2020).

Table S3 Nonredundancy GO terms of 482 putative MST genes (biological process).

Table S4 Nonredundancy GO terms of 482 putative MST genes (molecular function).

Table S5 Nonredundancy GO terms of 218 overlapping genes (biological process).

Table S6 Nonredundancy GO terms of 218 overlapping genes (molecular function).

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

BMC Genomics

RESEARCH

Open Access



Genome-wide association analysis unveils candidate genes and loci associated with aplasia cutis congenita in pigs

Fuchen Zhou^{1†}, Shenghui Wang^{2†}, Haojun Qin¹, Haiyu Zeng², Jian Ye², Jie Yang¹, Gengyuan Cai^{1,2}, Zhenfang Wu^{1,2*} and Zebin Zhang^{1*}

Abstract

Background Aplasia cutis congenita (ACC) is a rare genetic disorder characterized by the localized or widespread absence of skin in humans and animals. Individuals with ACC may experience developmental abnormalities in the skeletal and muscular systems, as well as potential complications. Localized and isolated cases of ACC can be treated through surgical and medical interventions, while extensive cases of ACC may result in neonatal mortality. The presence of ACC in pigs has implications for animal welfare. It contributes to an elevated mortality rate among piglets at birth, leading to substantial economic losses in the pig farming industry. In order to elucidate candidate genetic loci associated with ACC, we performed a Genome-Wide Association Study analysis on 216 Duroc pigs. The primary goal of this study was to identify candidate genes that associated with ACC.

Results This study identified nine significant SNPs associated with ACC. Further analysis revealed the presence of two quantitative trait loci, 483 kb (5:18,196,971–18,680,098) on SSC 5 and 159 kb (13:20,713,440–20,729,443 bp) on SSC13. By annotating candidate genes within a 1 Mb region surrounding the significant SNPs, a total of 11 candidate genes were identified on SSC5 and SSC13, including *KRT71*, *KRT1*, *KRT4*, *ITGB7*, *CSAD*, *RARG*, *SP7*, *PFKL*, *TRPM2*, *SUMO3*, and *TSPEAR*.

Conclusions The results of this study further elucidate the potential mechanisms underlying and genetic architecture of ACC and identify reliable candidate genes. These results lay the foundation for treating and understanding ACC in humans.

Keywords Aplasia cutis congenita, Congenital disorder, Genome-wide association study, Pig, Candidate gene

Background

Aplasia cutis congenita (ACC) is a rare congenital disorder characterized by the localized or widespread absence of skin in humans and animals [1, 2]. The lesions can range from a few millimeters to more than 10 cm in diameter. Some defects can have a membranous covering that can be filled with fluid, giving it a bullous appearance [3]. In 1826, Campbell initially documented cases of skin aplasia on the scalp of infants [4], with the potential for involvement in other body regions and leading to neonatal mortality [5]. In human research, studies on ACC

[†]Fuchen Zhou and Shenghui Wang contributed equally to this work.

*Correspondence:

Zhenfang Wu

wzfemail@163.com

Zebin Zhang

zbzhang@scau.edu.cn

¹ College of Animal Science and National Engineering Research Center for Breeding Swine Industry, South China Agricultural University, Guangdong 510642, P.R. China

² Guangdong Wens Breeding Swine Technology Co., Ltd, Guangdong 527400, P.R. China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

have primarily been confined to case analyses of individual patients, with a lack of in-depth investigation into the disease [6–8]. This may be attributed to the rarity of the condition, making it difficult to gather a large number of samples. Coi et al. [3] conducted a study on ACC cases in Europe from 1998 to 2017 and found a prevalence rate of 5.2 cases per 100,000 live births among newborns. The presence of ACC in pigs has implications for animal welfare. It contributes to an elevated mortality rate among piglets at birth, leading to substantial economic losses in the pig farming industry. In 2006, Odile et al. [9] reported that ACC commonly occurs in the tail, back, thighs, and abdomen of piglets. The affected areas exhibit red lesions with a depressed appearance, lacking skin and hair coverage in the surrounding regions. ACC may cooccur with developmental anomalies such as, muscular and skeletal abnormalities [10], while its precise etiology and pathogenesis remain elusive. Genetic factors have been proposed as possible contributors, and case studies have identified instances with both autosomal dominant and recessive inheritance patterns [9, 11]. Nonetheless, the genetic underpinnings of ACC have yet to be fully elucidated, highlighting the need for further research in this area.

Since the successful completion of the Human Genome Project in 2003, remarkable advancements in genome sequencing technology have led to a substantial reduction in sequencing costs [12]. These breakthroughs have paved the way for a comprehensive understanding of the phenotypic variations and disease occurrences resulting from genomic variations [12]. In 2005, Klein et al. pioneered the application of Genome-Wide Association Study (GWAS) in investigating human macular degeneration [13], marking the beginning of GWAS as a powerful tool for unraveling candidate genes associated with both human diseases [14, 15] and economically significant traits [16, 17] in agriculture. In our previous research, we utilized GWAS to identify potential genes related to several crucial economic traits in pigs, including lean meat percentage [18], teat number [19], and intramuscular fat content [20].

Due to the low probability of ACC occurrence, it took us 5 years (2017–2021) to collect a total of 116 ACC-affected piglets (58 from American Duroc population and 48 from Canadian Duroc population) and 100 healthy piglets (50 from American Duroc population and 50 Canadian Duroc population). Subsequently, we conducted a genome-wide association analysis on a total of 216 Duroc piglets to elucidate candidate genetic loci associated with ACC. The primary goal of this study was to identify candidate genes that contribute to the development of ACC, thus providing valuable insights into the genetic architecture underlying skin aplasia. By

leveraging the pig as a model organism, our research aimed to unravel the fundamental mechanisms governing skin development disorders, ultimately offering valuable implications for treating of related human diseases.

Result

Phenotype and SNP genotyping

ACC can occur in various parts of piglets' bodies (Additional file 1: Fig S1). Figure 1 illustrates a case of ACC on the left forehoof of a piglet, characterized by a well-defined border, a bright red base, a relatively small lesion area, and no apparent discharge. After DNA extraction from tail tissue samples of the 216 piglets, we performed genotyping using the GenoBaits Porcine 50 K single nucleotide polymorphism (SNP) Chip. The quality of genotyping of the 216 Duroc pigs was examined using PLINK v1.9. The distribution and visualization of the SNP dataset across chromosomes are summarized in Fig. 2a and Additional file 2: Table S1. These SNPs were roughly proportionally distributed on all 18 chromosomes of pigs, with the longest chromosome having the most significant number of SNPs. As shown in Fig. 2a, we observed a higher density of SNP distribution towards the ends of the chromosomes. This pattern may be attributed to the limitations of the sequencing technology, which is consistent with the observed distribution and provides further evidence for the reliability of our sequencing results. Additionally, *Sus scrofa* chromosome (SSC) 6 exhibited the most uniform distribution of SNPs.

Population structure and LD decay

Considering the involvement of two Duroc populations in this study, we were concerned about the potential presence of population stratification. Therefore, we conducted a Principal component analysis (PCA). As presented in Fig. 2b, the PCA plot showed that the American



Fig. 1 Illustration of a piglet with ACC on the left forehoof

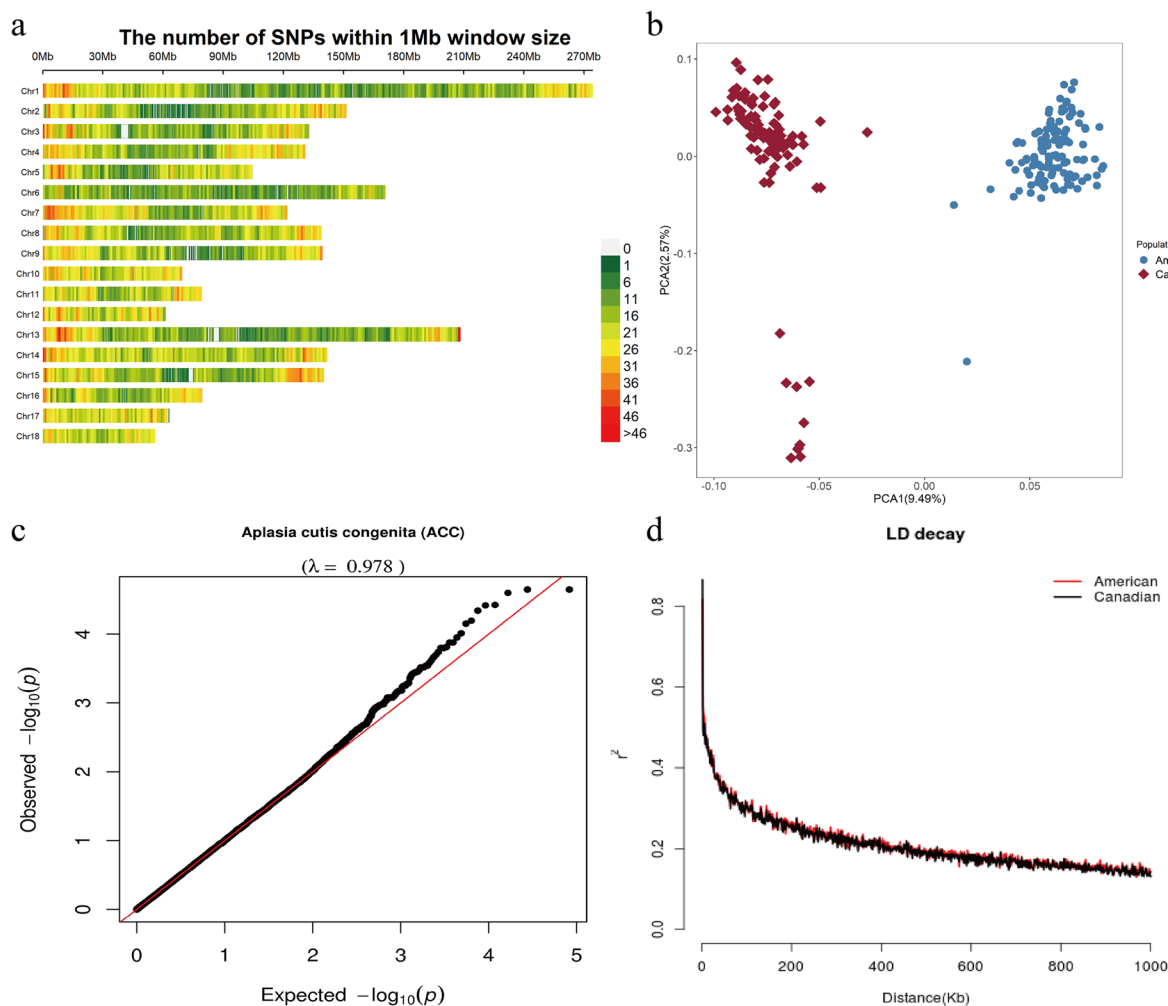


Fig. 2 Distribution of SNPs across Chromosomes after quality control (a), population structure for two Duroc populations (b), Quantile–quantile (Q–Q) plots of genome-wide association studies for ACC in the two populations (c), and LD decay across the whole genome of the association population (d). For a, the Y-axis values are autosomes and X-axis is chromosome position in Mb. For b, the PCA1 present first principal component, PCA2 is second principal component. The first two PCs derived from the genomic kinship matrix were extracted to assess the population structure. The plot c shows the observed versus expected negative log10 *P*-value. In plot 4, the red and black colors represent LD decay in the American and Canadian Duroc populations, respectively

and Canadian Duroc pigs did not coincide, indicating that these two populations have different genetic backgrounds. In addition to incorporating the genetic relationship matrix as a covariate in the Mixed linear model (MLM) model, we included the first five principal components from PCA as covariates to adjust for population structure. After adjust, Q-Q plots with genomic inflation factors were generated to evaluate the influence of population structure on single-locus GWAS (Fig. 2c). Notably, no significant inflation of test statistics was observed in the admixed population (American and Canadian Duroc population), indicating that the results were not biased by population stratification. The average Linkage disequilibrium (LD) decay distances of the American and

Canadian Duroc pig populations ($r^2=0.2$) were approximately 500 kb (Fig. 2d), suggesting a comparable LD level between the two populations. Consequently, we treated the American and Canadian Duroc populations as a unified group for further analysis.

Single-locus GWAS results

After analyzing the quality-controlled SNPs and phenotype data using the MLM model in GEMMA software [21], significant SNPs detected through 50 K chips GWAS are presented in Table 1 and Fig. 3a. A total of 2 SNPs surpassed the suggestive significance threshold ($P=2.42E-5$) and were identified as associated with ACC. Two SNPs (13_207176753 ($P=2.25E-5$))

Table 1 Significant and strong LD SNPs associated with ACC in Duroc pigs identified by 50 K chips GWAS

SSC ^a	SNP ID	Position (bp) ^b	MAF	P-value	Nearest gene	Distance (bp) ^c
5	5_18403003	18,403,003	0.051	9.72E-5	<i>CSAD</i>	within
5	5_18617273	18,617,273	0.051	6.37E-5	<i>SP1</i>	within
13	13_207176753	207,176,753	0.051	2.25E-5	<i>PFKL</i>	within
13	13_207265708	207,265,708	0.051	2.25E-5	<i>TRPM2</i>	726
13	13_207251317	207,251,317	0.049	2.52E-5	<i>TRPM2</i>	within

^a *Sus scrofa* chromosome
^b SNP position in Ensembl
^c The SNP located upstream/downstream of the nearest gene

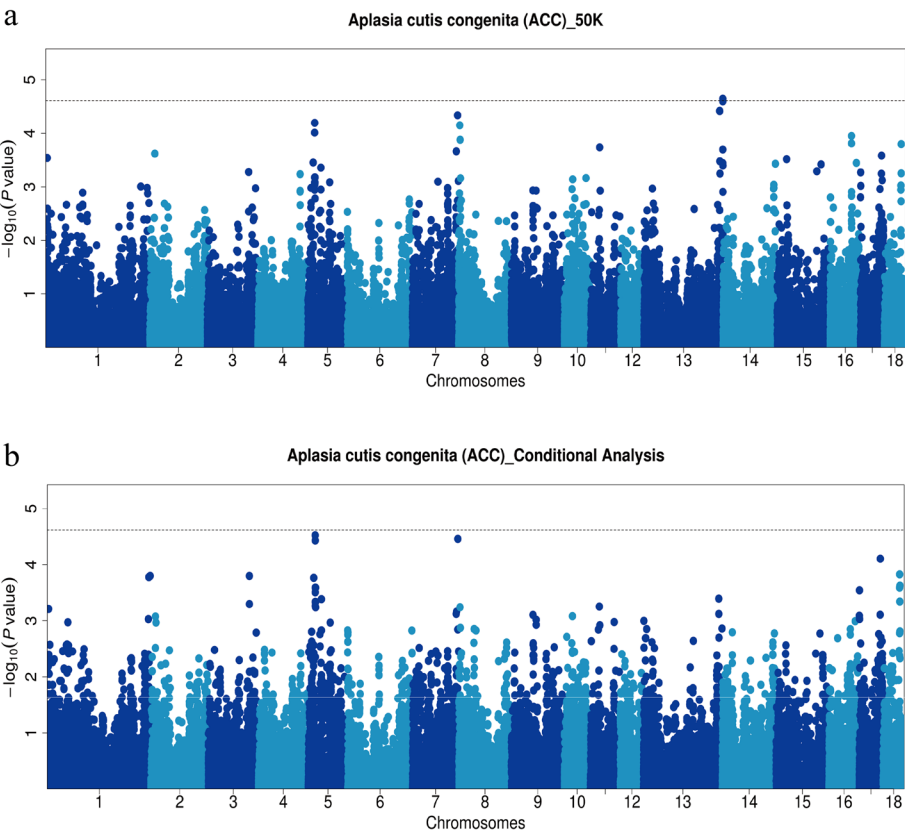


Fig. 3 Manhattan plots of GWAS (a) and conditional analysis (b) for ACC in 216 pigs using 50 K chip data. The x-axis represents the chromosomes, and the y-axis represents the $-\log_{10}(P\text{-value})$. The dashed lines indicate the suggestive thresholds for ACC ($P = 2.42 \times 10^{-5}$)

and13_207265708 ($P = 2.25 \times 10^{-5}$) on SSC13 were found to be the lead SNPs. They are in complete LD and located within a 159 kb haplotype block (13:20,713,440–207294431 bp), which suggests that mutations near the potential quantitative trait locus (QTL) may have essential effect on ACC (Fig. 4a). Furthermore, GWAS conducted by imputed data revealed nine and six significant SNPs on SSC 13 and 10, respectively. Among these, the top SNP on SSC 13 was identified as 13_206783027_C ($P = 2.49 \times 10^{-7}$). For more

comprehensive information, please consult Fig. S2 and Table S2. To further investigate, we conducted a conditional analysis by fitting the lead SNP (13_207176753) as a covariate into the MLM using GEMMA software (Fig. 3b). The signal peak on SSC13, where the lead SNP is located, disappeared, and the P -value of the associated SNPs all dropped below the threshold line (Fig. 5a and b). This indicates a strong LD among these SNPs, indicating a significant influence of the lead SNP on ACC.

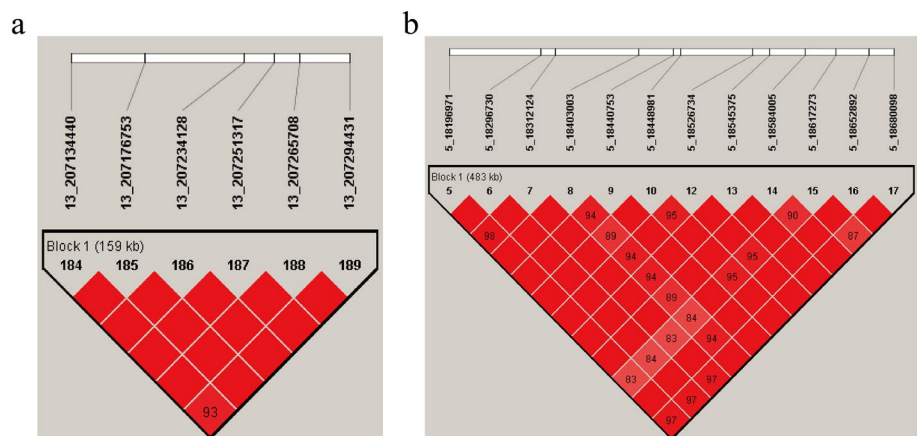


Fig. 4 LD blocks in the significant region on SSC13 (a) and SSC5 (b). LD blocks are marked with triangles. Values in boxes are LD (r^2) between SNP pairs and the boxes are colored according to the standard Haploview color scheme

Surprisingly, the signal peak on SSC5 became even more pronounced, implying the presence of a potential QTL associated with ACC (Fig. 3b). Furthermore, haplotype analysis revealed strong linkage disequilibrium (LD) between SNPs 5_18403003 ($P=9.72E-5$) and 5_18617273 ($P=6.37E-5$), which are located within a 483 kb haplotype block (5:18,196,971–18,680,098) (Fig. 4b). This implies that mutations in the vicinity of this region may also have substantive implications for ACC. Finally, to assess the impact of population stratification on our analysis results, we conducted GWAS using logistic mixed model with the GMMAT R package. The results were similar to those of the MLM model and showed strong signal on SSC 13. This further validates the scientific rigor of our use of the MLM model for binary trait GWAS (Additional file 5: Fig S3).

Candidate genes search and functional annotation

Among the identified five SNPs on 50 K chips GWAS, four SNPs were found to be located within five genes: cysteine sulfinic acid decarboxylase (*CSAD*), Sp1 transcription factor (*SP1*), phosphofructokinase, liver type (*PFKL*), transient receptor potential cation channel subfamily M member 2 (*TRPM2*), respectively (Table 1). Based on the LD decay distance observed in the population used in this study, we observed that LD (r^2) dropped below 0.2 after a distance of approximately 500 kb (Fig. 2d). This indicates that the causal mutations and genes may be located within a region spanning 500 kb upstream and downstream of the identified GWAS signals. To further understand the GWAS results, a total of 58 protein-coding genes within a 1 Mb region centered on the five SNPs (50 K chips GWAS) and 15 SNPs (imputed GWAS) were annotated. Additionally,

we identified seven genes from the imputed GWAS that overlap with those discovered in the 50 K chip data, further underscoring the accuracy of the 50 K chip GWAS results (Additional file 6: Table S3). Our pathway enrichment analysis revealed several significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) terms related to ACC, including cellular cytoskeleton organization and biological development, such as skeletal muscle tissue regeneration, prostate gland epithelium morphogenesis, establishment of skin barrier, face development (Fig. 6a and b, Additional file 7: Table S4). After conducting non-redundant GO analysis on all GO terms that exceeded the threshold using the REVIGO website, a total of 87 GO terms were clustered. (Fig. 7a and b). One pathway stood out as particularly intriguing, as it is related to the process of skin barrier (GO:0061436). Subsequently, we employed the GeneCards, Mouse Genome Informatics databases, and conducted an extensive literature review to explore the functional roles of the identified genes. As a result, we identified a total of 11 candidate genes with potential relevance to ACC. These genes, namely *KRT71*, *KRT1*, *KRT4*, *ITGB7*, *CSAD*, *RARG*, *SP7*, *PFKL*, *TRPM2*, *SUMO3*, and *TSPEAR*, exhibit promising associations with ACC based on their known functions and previous research findings. The genes *KRT71*, *KRT1*, *KRT4*, *RARG*, *PFKL*, *TSPEAR*, and *SUMO3* are involved in hair follicle formation and skin development. On the other hand, the genes *ITGB7*, *CSAD*, and *TRPM2* are associated with immune system diseases that accompany skin deficiencies. Surprisingly, both the *PFKL* and *TRPM2* genes were identified by both the 50 K chips and Imputed GWAS, further substantiating their reliability as candidate genes. Additionally, the *SP7* gene is related to skeletal developmental abnormalities in ACC.

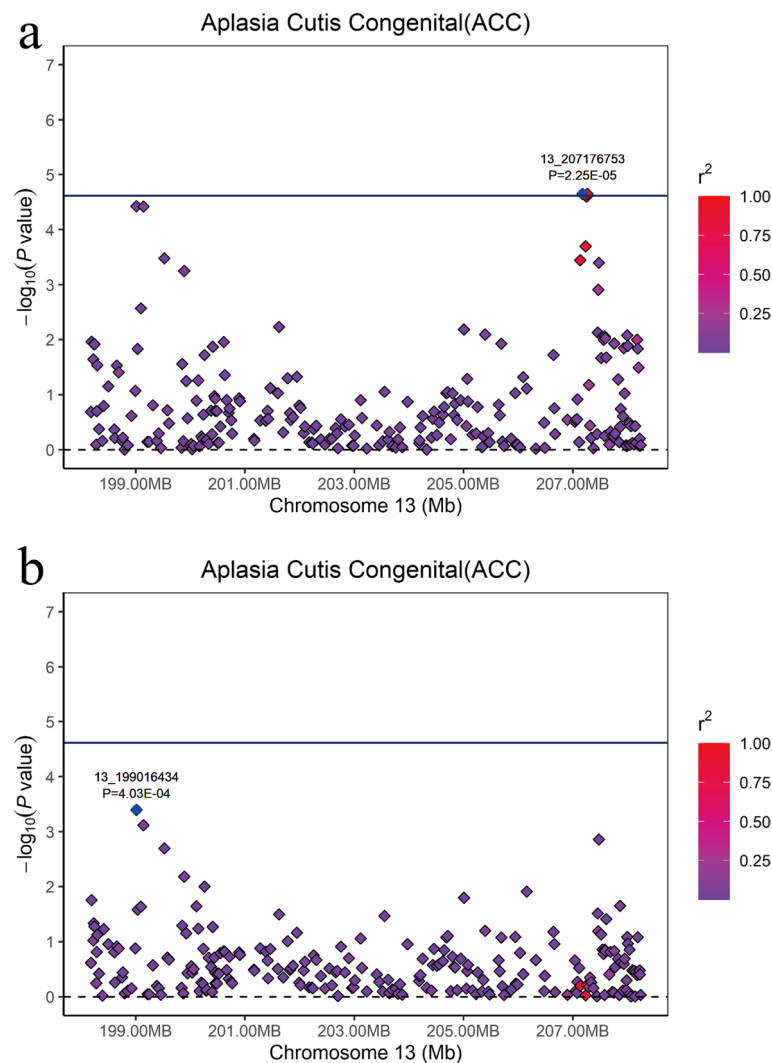


Fig. 5 Regional association plot of the lead SNP (13_207176753) associated with ACC at SSC13. For **a** and **b** plots, the $-\log_{10}(P\text{-value})$ of SNP (Y-axis) are presented according to their chromosomal position (X-axis) on SSC13. The lead SNP of GWAS is denoted by large blue square. Other SNPs are represented by colored rhombi according to the target SNP with which they were in strongest LD. The plots **a** and **b** show the association results for ACC before and after conditional analysis on lead SNP (13_207176753)

Discussion

The skin, the largest organ in the human body, plays a vital role in immune surveillance, wound healing, and protection against environmental challenges [22]. Skin-related diseases contribute significantly to healthcare expenditures. Extensive congenital skin aplasia, in particular, can result in neonatal mortality, while congenital skin aplasia in pig farming is associated with postpartum piglet mortality. Genetic factors have been shown to impact on many skin diseases significantly [23]. Therefore, uncovering candidate genes related to skin development and diseases can greatly facilitate diagnosing and treating of skin-related disorders.

Candidate genes function for ACC on SSC5

Through gene function annotation, we identified seven candidate genes associated with ACC on SSC5, namely *KRT71*, *KRT1*, *KRT4*, *ITGB7*, *CSAD*, *RARG*, and *SP7*. The *KRT71* gene encodes a protein expressed in the inner root sheath of hair follicles and is implicated in disorders such as Hypotrichosis 13 and Familial Woolly Hair Syndrome (<https://www.ncbi.nlm.nih.gov/gene/112802#summary>). *KRT71* exhibits differential expression in various human skin regions, including the buttocks, arms, abdomen, and legs (<https://www.bgee.org/gene/ENSG00000139648>). Studies have highlighted the pivotal role of *KRT71* in hair formation, as it is involved in abnormal hair follicle

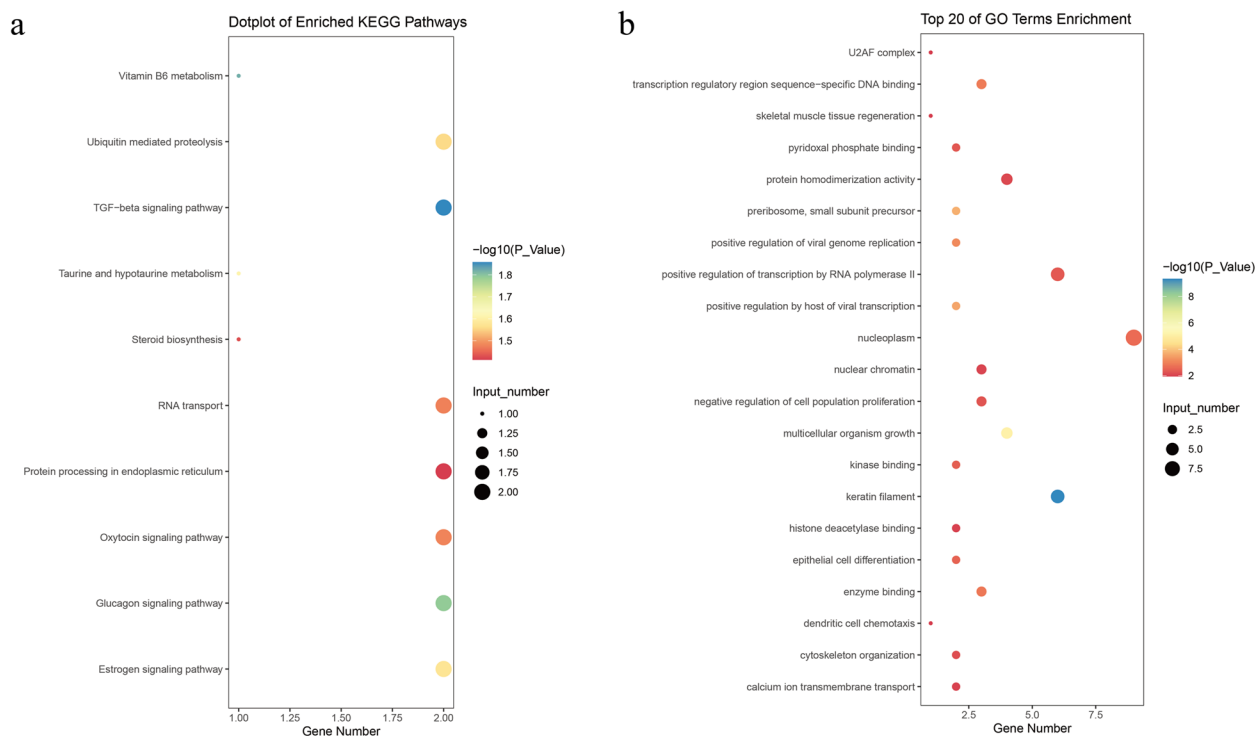


Fig. 6 Significant KEGG pathways and GO terms associated with ACC ($P < 0.05$). The plot **a** represents the KEGG PATHWAY of the biological process for protein-coding genes within a 1 Mb region centered on the significant SNPs. The plot **b** shows the top 20 terms of the GO enrichment

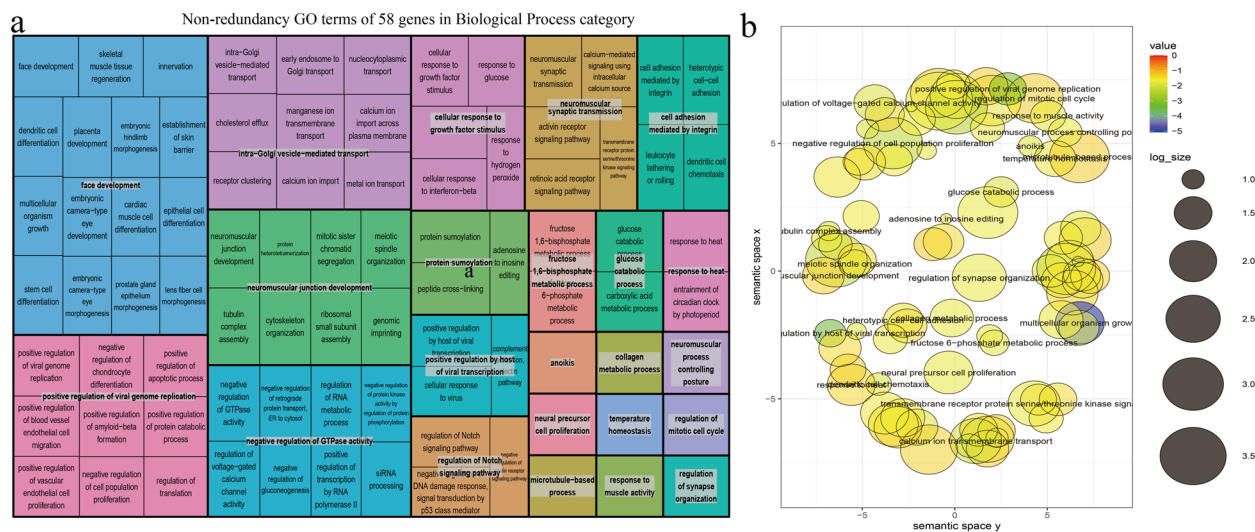


Fig. 7 Non-redundancy GO terms of ACC-related genes in Biological Process category. The plot **a** shows the "TreeMap" view of REVIGO. Each rectangle represents a representative cluster. These representatives are combined into "superclusters," representing loosely related terms and visualized using different colors. The size of the rectangles is adjusted to reflect the P -value and frequency of the GO term in the Mus musculus GOA database. The plot **b** displays the scatter plot of representative clusters. The log size indicates the frequency of the GO term in the Mus musculus GOA database, with larger sizes indicating more common terms. The numerical value represents the $-\log_{10}(P\text{-value})$, with colors ranging from red to blue indicating increasing significance

morphology, focal hair loss, Male-pattern baldness, and abnormal hair cortex morphology [24–27]. The protein encoded by the *KRT1* gene is a member of the keratin gene family. Type II cytokeratins consist of acidic or neutral proteins that form heterotypic pairs of keratin chains and are co-expressed during the differentiation of simple and stratified epithelial tissues (<https://www.ncbi.nlm.nih.gov/gene/3848>). Notably, *KRT1* has been implicated in various conditions, including abnormal skin morphology, impaired skin barrier function, abnormal epidermal layer morphology, and neonatal lethality with incomplete penetrance [28, 29]. Like *KRT1*, another candidate gene, *KRT4*, encodes a protein that belongs to the keratin gene family. It has been implicated in a range of skin and hair abnormalities, including epidermal hyperplasia, abnormal morphology of the suprabasal layer in the epidermis, impaired skin barrier function, skin inflammation, abnormal morphology of the dermal layer, and diluted coat color [30–32]. All three candidate genes mentioned above are members of the keratin gene family and play crucial roles in hair and skin development. They are likely to be closely associated with the occurrence of ACC, although the specific mechanisms of their actions have not yet been elucidated.

Research findings suggest that ACC may be accompanied by developmental deficits or musculoskeletal deficiencies [10]. The candidate gene *RARG* is not only associated with skin abnormalities, such as abnormal epidermal layer morphology, abnormal epidermis stratum corneum morphology, and impaired skin barrier function [33], but also implicated in skeletal developmental anomalies, including abnormal basioccipital bone morphology, rib fusion, cervical vertebral transformation, abnormal ventral tubercle of atlas morphology, and abnormal cervical axis morphology [34–38]. The candidate gene *SP7* also correlates with skeletal developmental abnormalities [39–41]. Due to the high postpartum mortality rate in affected piglets, *CSAD* gene on SSC5, the peak SNP (5_18403003) located within, is associated with neonatal lethality and incomplete penetrance [42]. Furthermore, another candidate gene, *ITGB7*, is implicated in abnormal immune system function in animals [43, 44]. Therefore, we hypothesize that the postpartum mortality in piglets may be attributed to the actions of these two genes.

Candidate genes function for ACC on SSC13

The *PFKL* gene on SSC13, the lead SNP (13_207176753) located within, encodes the liver (L) subunit of an enzyme that catalyzes the conversion of D-fructose 6-phosphate to D-fructose 1,6-bisphosphate, which is a crucial step in glucose metabolism (glycolysis). Previous studies have linked the *PFKL* gene to hair follicle degeneration and

abnormal hair cycle catagen phase [45]. Another gene, *TRPM2*, located near another lead SNP (13_207265708), is closely associated with the innate immune system and may contribute to increased susceptibility to bacterial infections, potentially leading to mortality [46–48].

During embryonic development, the ectoderm undergoes differentiation to form the skin, hair, and skeletal system. Research has demonstrated that the *TSPEAR* gene participates in ectodermal development and plays a crucial role in tooth and hair follicle morphogenesis by regulating the Notch signaling pathway [49, 50]. It is also a candidate gene affecting wool production traits in Merino sheep [51]. The *SUMO3* gene also influences the ectoderm development in embryos, resulting in embryonic growth arrest and delayed somite formation [52]. In summary, these four strong candidate genes on SSC13 may play a crucial role in the development of ACC, ultimately leading to the manifestation of the condition in affected individuals. However, the precise mechanisms underlying their involvement require further investigation.

Conclusion

The present study aimed to investigate the candidate genes and loci associated with ACC. GWAS analysis was conducted on a herd of 226 Duroc pigs, identifying nine significantly associated SNPs. Further investigation revealed the presence of a 483 kb QTL (5:18,196,971–18,680,098) on SSC 5 and a 159 kb QTL (13:20,713,440–20,729,443 bp) on SSC13. By annotating candidate genes within a 1 Mb region surrounding the significant SNPs, a total of 11 candidate genes were identified on SSC5 and 13, including *KRT71*, *KRT1*, *KRT4*, *ITGB7*, *CSAD*, *RARG*, *SP7*, *PFKL*, *TRPM2*, *SUMO3*, and *TSPEAR*. These findings shed light on the genetic underpinnings of ACC in pigs, contribute to the understanding and treatment of ACC and provide a foundation for future studies elucidating the underlying molecular mechanisms.

Material and methods

Ethics statement

The animals and experimental methods used in this study follow the guidelines of the Ministry of Agriculture of China and the Use Committee of South China Agricultural University (SCAU). The ethics committee of SCAU (Guangzhou, China) approved all animal experiments. The experimental animals were not anesthetized or euthanized in this study. We confirmed that all methods are reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>) for the reporting of animal experiments.

Animals and phenotype

Between 2017 and 2021, a total of 108 recently farrowed (58 ACC-affected and 50 healthy) American Duroc piglets and 98 recently farrowed (48 ACC-affected and 50 healthy) Canadian Duroc piglets were collected from breeding farms in Guangdong Wen's Foodstuffs Co., Ltd. (Guangdong, China). The farrowing sows were raised under the same environmental conditions, and all affected piglets were identified on-site personnel.

Genotype data acquisition and quality control

The genomic DNA necessary for this study was obtained by applying the standard phenol/chloroform method to isolate and extract DNA from the tail tissues of 216 piglets. Subsequently, rigorous quality control measures were implemented, including assessment of the light absorption ratios (A260/280 and A260/230), gel electrophoresis analysis, and quantification of DNA concentration at 50 ng/μL. GenoBaits Porcine 50 K Panel was used for genotyping. After genotyping, we enhanced the genotype data to the whole-genome sequence level using an imputation strategy. We employed the Swine Imputation (SWIM) Server tool [53] with default parameter settings to perform genotype imputation, bridging the target and reference genotype data. The reference haplotype panels were constructed from whole-genome sequencing data collected from 2259 pigs, representing 44 breeds. The genotype imputation accuracy consistently demonstrated a high average concordance rate exceeding 97%, a non-reference concordance rate of 91%, and an r^2 value of 0.89. This ensured the reliability and robustness of our imputed data. The genotype quality control of the 216 pigs was conducted by PLINK v1.9 software [54]. Individuals with a call rate of less than 90%, SNP with a call rate of less than 90% and a minor allele frequency of less than 1% were also deleted. SNP that failed the Hardy–Weinberg equilibrium test ($P < 10^{-6}$) and was unmapped or located on the sex chromosome was also removed. After quality control, the final 41,362 (50 K) and 14,942,886 (SWIM imputed) SNPs are retained for subsequent analysis, respectively.

Population structure and LD estimation

PCA and LD analysis were performed using the SNPs that passed quality control standards to investigate the population structure of the two Duroc pig populations. PCA was performed with GCTA software v1.92.4beta [55]. The average LD decay distance across the whole

genome of the American and Canadian Duroc pigs was calculated using PopLDdecay software [56].

Single-locus GWAS analysis

In the present study, the GEMMA software v0.98.1 [21] was used to implement MLM for single-locus GWAS of ACC. GEMMA calculated the genomic relatedness matrix (GRM) between individuals within each population to account for the population structure. The first five principal components calculated by GCTA tool [55] are embedded into the MLM as covariables to eliminate the mixed influence of population structure. The MLM is as follows:

$$y = W\alpha + X\beta + u + \varepsilon$$

where y represents a vector of the phenotype, with affected piglets assigned as 1 and healthy individuals assigned as 0; W is the incidence matrix of covariates, including fixed effects of the sex, and the top five principal components from PCA analysis; α represents the vector of corresponding coefficients including the intercept; X is the vector of all marker genotypes; β specifies the corresponding effect size of the marker; u is the vector of random effects, with $u \sim \text{MVN}_n(0, \lambda\tau^{-1}K)$; ε is the vector of random residuals, with $\varepsilon \sim \text{MVN}_n(0, \tau^{-1}I_n)$; λ signifies the ratio between two variance components; τ^{-1} is the variance of the residual errors; K is GRM; I is an $n \times n$ identity matrix and n refers to the number of pigs. In the 50 K chips GWAS, the Bonferroni method was used to determine the genome-wide significant threshold ($0.05/N$), where N is the number of SNPs. Since that is a stringent criterion, we set a more lenient threshold to detecting the suggestive SNPs ($1/N$) [20, 57]. Furthermore, to address the potential type I errors introduced by population stratification, we employed the GMMAT R package to perform logistic mixed model GWAS specific to binary traits [58]. Based on human GWAS results, we set the genome-wide significance threshold and suggestive significance threshold for GWAS based on imputed data at $5.00\text{E-}8$ and $1.00\text{E-}6$, respectively [59, 60].

Haplotype block analysis and conditional analysis

The PLINK v1.9 [54] and Haploview v4.2 [61] were used for chromosomal regions with multiple significantly clustered around the top SNP to calculate the LD pattern of the region. GWAS often identifies a significant set of SNPs associated with target traits in putative regions, possibly due to high LD between them. Conditional analysis was conducted by fitting the genotypes of peak SNP as covariates to the univariate MLM of GEMMA to evaluate the independence of the signal peak in the putative region.

Identification of candidate genes and functional enrichment analysis

Based on the LD decay distances of the American and Canadian Duroc populations, the candidate genomic regions were determined to 500 kb on either side of the significant SNPs. All SNPs refer to the latest version of the *Sus scrofa* 11.1 genome (http://ensembl.org/Sus_scrofa/Info/Index). Functional gene annotation (v105) was downloaded in GIFF3 format from the Ensembl website (http://ftp.ensembl.org/pub/release-105/gff3/sus_scrofa/). The R package BioMart [62] efficiently retrieved functional genes. KEGG [63] and GO analyses were conducted using KOBAS 3.0 [64] to investigate the functions of all candidate genes. Enriched terms with a significance threshold of P value < 0.05 were selected to explore further the genes invoked in pathway and biological processes [19]. Subsequently, we employed REVIGO (<http://revigo.irb.hr>) in conjunction with the Mus musculus database to eliminate GO term redundancy (medium threshold, 0.7) and cluster the remaining terms in a 2D space [65, 66]. This space was derived by applying multidimensional scaling to a matrix of GO term semantic similarities. Mouse Genome Informatics website (<https://www.informatics.jax.org/>), GeneCards (<http://www.genecards.org/>) and Ensembl (www.ensembl.org/biomart/martview) were used to query gene functions.

Abbreviations

ACC	Aplasia cutis congenita
SSC	<i>Sus scrofa</i> Chromosome
GWAS	Genome-wide association study
SNP	Single nucleotide polymorphism
QTL	Quantitative trait locus
LD	Linkage disequilibrium
MLM	Mixed linear model
PCA	Principal component analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09803-6>.

Additional file 1: Fig S1. Photos of ACC-afflicted piglets.

Additional file 2: Table S1. Distributions of SNPs after QC and the average SNPs on each chromosome for ACC trait.

Additional file 3: Fig S2. Manhattan plots of Imputed GWAS for ACC in 216 pigs. Description: The x-axis represents the chromosomes, and the y-axis represents the $-\log_{10}(P\text{-value})$. The dashed lines indicate the thresholds for ACC ($P = 1.00E-6$).

Additional file 4: Table S2. Significant SNPs associated with ACC in Duroc pigs identified by imputed GWAS. Description: This file provides information about significant SNPs discovered in the GWAS with imputed data.

Additional file 5: Fig S3. Manhattan plots of 50K GWAS using logistic mixed model for ACC in 216 pigs. Description: The x-axis represents the chromosomes, and the y-axis represents the $-\log_{10}(P\text{-value})$. The dashed lines indicate the suggestive thresholds for ACC ($P = 2.42E-5$).

Additional file 6: Table S3. Genes annotation within 1 Mb surrounding nine significant SNPs identified in GWAS. Description: This file provides the positional information of protein-coding genes within a 1 Mb region surrounding significant SNPs.

Additional file 7: Table S4. Pathway Enrichment Analysis Reveals Significantly Enriched KEGG and GO Pathways Related to ACC. Description: This file provides the enrichment of protein-coding genes in KEGG and GO terms ($P < 0.05$).

Acknowledgements

The authors would like to thank all staff at the pig core breeding farms of Wens Foodstuff Group Co., Ltd. (Guangdong, China) for the help of sample collection.

Authors' contributions

JianY and ZW conceived and designed the experiment. SW, HQ, HZ and JeY collected the samples and recorded the phenotypes. FZ, HQ and HZ extracted the DNA for genotyping. FZ, SW and ZZ analyzed the data. FZ and ZZ wrote the manuscript. ZW and GC contributed to the materials. All authors reviewed and approved the manuscript.

Funding

This research was supported by the Guangdong Provincial Key R&D Program(2022B0202090002). The funders had no role in study design, data collection and analysis, publication decisions, or manuscript preparation.

Availability of data and materials

The SNP genotyping data presented in this study are deposited in the figshare (https://figshare.com/articles/dataset/Duroc_ACC_50K_vcf_gz/24346768).

Declarations

Ethics approval and consent to participate

The animals and experimental methods used in this study follow the guidelines of the Ministry of Agriculture of China and the Use Committee of South China Agricultural University (SCAU). The ethics committee of SCAU (Guangzhou, China) approved all animal experiments. The informed consent was obtained from Wens Foodstuff Group Co., Ltd. (Guangdong, China) for data collection. There was no use of human participants, data, or tissues. We confirmed that all methods are reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>) for the reporting of animal experiments.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 August 2023 Accepted: 11 November 2023

Published online: 21 November 2023

References

- Humphrey SR, Hu X, Adamson K, Schaus A, Jensen JN, Drolet B. A practical approach to the evaluation and treatment of an infant with aplasia cutis congenita. *J Perinatol*. 2018;38(2):110–7.
- Colon-Fontanez F, Fallon Friedlander S, Newbury R, Eichenfield LF. Bullous aplasia cutis congenita. *J Am Acad Dermatol*. 2003;48(5 Suppl):S95–98.
- Coi A, Barisic I, Garne E, Pierini A, Addor MC, Aizpurua Atxega A, Ballardini E, Braz P, Broughan JM, Caverro-Carbonell C, et al. Epidemiology of aplasia cutis congenita: A population-based study in Europe. *J Eur Acad Dermatol Venereol*. 2023;37(3):581–9.
- Campbell W. Case of congenital ulcer on the cranium of a fetus, terminating in fatal hemorrhage, on the 18th day after birth. *Edin J Med Sci*. 1826;2:82–3.

5. Taifour Suliman M, Quazi A. Aplasia cutis congenita of the trunk in a Saudi newborn. *Br J Plast Surg*. 2004;57(6):582–4.
6. Perry BM, Maughan CB, Crosby MS, Hadenfeld SD. Aplasia cutis congenita type V: a case report and review of the literature. *Int J Dermatol*. 2017;56(6):e118–21.
7. Ugowe OJ, Balogun SA, Adejuyigbe EA. Aplasia Cutis Congenita: A Case Report. *West Afr J Med*. 2021;38(4):391–4.
8. Yang MY, Ha DL, Kim HS, Ko HC, Kim BS, Kim MB. Aplasia cutis congenita in Korea: Single center experience and literature review. *Pediatr Int*. 2020;62(7):804–9.
9. Benoit-Biancamano MO, Drolet R, D'Allaire S. Aplasia cutis congenita (epitheliogenesis imperfecta) in swine: observations from a large breeding herd. *J Vet Diagn Invest*. 2006;18(6):573–9.
10. Verhelle NA, Heymans O, Deleuze JP, Fabre G, Vranckx JJ, Van den hof B. Abdominal aplasia cutis congenita: case report and review of the literature. *J Pediatr Surg*. 2004;39(2):237–9.
11. Itin P, Pletscher M. Familial aplasia cutis congenita of the scalp without other defects in 6 members of three successive generations. *Dermatologica*. 1988;177(2):123–5.
12. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333–51.
13. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308(5720):385–9.
14. Zhang XJ, Huang W, Yang S, Sun LD, Zhang FY, Zhu QX, Zhang FR, Zhang C, Du WH, Pu XM, et al. Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. *Nat Genet*. 2009;41(2):205–10.
15. Padmanabhan S, Dominiczak AF. Genomics of hypertension: the road to precision medicine. *Nat Rev Cardiol*. 2021;18(4):235–50.
16. Cho IC, Park HB, Ahn JS, Han SH, Lee JB, Lim HT, Yoo CK, Jung EJ, Kim DH, Sun WS, et al. A functional regulatory variant of MYH3 influences muscle fiber-type composition and intramuscular fat content in pigs. *PLoS Genet*. 2019;15(10):e1008279.
17. Ma J, Yang J, Zhou L, Ren J, Liu X, Zhang H, Yang B, Zhang Z, Ma H, Xie X, et al. A splice mutation in the PHKG1 gene causes high glycogen content and low meat quality in pig skeletal muscle. *PLoS Genet*. 2014;10(10):e1004710.
18. Zhou S, Ding R, Meng F, Wang X, Zhuang Z, Quan J, Geng Q, Wu J, Zheng E, Wu Z, et al. A meta-analysis of genome-wide association studies for average daily gain and lean meat percentage in two Duroc pig populations. *BMC Genomics*. 2021;22(1):12.
19. Zhuang Z, Ding R, Peng L, Wu J, Ye Y, Zhou S, Wang X, Quan J, Zheng E, Cai G, et al. Genome-wide association analyses identify known and novel loci for teat number in Duroc pigs using single-locus and multi-locus models. *BMC Genomics*. 2020;21(1):344.
20. Ding R, Yang M, Quan J, Li S, Zhuang Z, Zhou S, Zheng E, Hong L, Li Z, Cai G, et al. Single-Locus and Multi-Locus Genome-Wide Association Studies for Intramuscular Fat in Duroc Pigs. *Front Genet*. 2019;10:619.
21. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012;44(7):821–4.
22. Lee HJ, Kim M. Skin Barrier Function and the Microbiome. *Int J Mol Sci*. 2022;23(21):13071.
23. Dyer JA. Practice gaps. Propranolol to treat hemangiomas of infancy: safety and side effect recognition. *JAMA Dermatol*. 2013;149(4):485–6.
24. Runkel F, Klasten M, Koch K, Böhnert V, Büssow H, Fuchs H, Franz T, Hrabé de Angelis M. Morphologic and molecular characterization of two novel Krt71 (Krt2-6g) mutations: Krt71rco12 and Krt71rco13. *Mamm Genome*. 2006;17(12):1172–82.
25. Poirier C, Yoshiki A, Fujiwara K, Guénet JL, Kusakabe M. Hague (Hag). A new mouse hair mutation with an unstable semidominant allele. *Genetics*. 2002;162(2):831–40.
26. Peters T, Sedlmeier R, Büssow H, Runkel F, Lüers GH, Korthaus D, Fuchs H, Hrabé de Angelis M, Stumm G, Russ AP, et al. Alopecia in a novel mouse model RCO3 is caused by mK6irs1 deficiency. *J Invest Dermatol*. 2003;121(4):674–80.
27. Yap CX, Sidorenko J, Wu Y, Kemper KE, Yang J, Wray NR, Robinson MR, Visscher PM. Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. *Nat Commun*. 2018;9(1):5407.
28. Roth W, Kumar V, Beer HD, Richter M, Wohlenberg C, Reuter U, Thiering S, Startschek-Jox A, Hofmann A, Kreusch F, et al. Keratin 1 maintains skin integrity and participates in an inflammatory network in skin through interleukin-18. *J Cell Sci*. 2012;125(Pt 22):5269–79.
29. Wallace L, Roberts-Thompson L, Reichelt J. Deletion of K1/K10 does not impair epidermal stratification but affects desmosomal structure and nuclear integrity. *J Cell Sci*. 2012;125(Pt 7):1750–8.
30. Fischer H, Langbein L, Reichelt J, Praetzel-Wunder S, Buchberger M, Ghanadan M, Tschachler E, Eckhart L. Loss of keratin K2 expression causes aberrant aggregation of K10, hyperkeratosis, and inflammation. *J Invest Dermatol*. 2014;134(10):2579–88.
31. Ness SL, Edelman W, Jenkins TD, Liedtke W, Rustgi AK, Kucherlapati R. Mouse keratin 4 is necessary for internal epithelial integrity. *J Biol Chem*. 1998;273(37):23904–11.
32. McGowan KA, Fuchs H, Hrabé de Angelis M, Barsh GS. Identification of a Keratin 4 mutation in a chemically induced mouse mutant that models white sponge nevus. *J Invest Dermatol*. 2007;127(1):60–4.
33. Calléja C, Messaddeq N, Chapellier B, Yang H, Krezel W, Li M, Metzger D, Mascréz B, Ohta K, Kagechika H, et al. Genetic and pharmacological evidence that a retinoic acid cannot be the RXR-activating ligand in mouse epidermis keratinocytes. *Genes Dev*. 2006;20(11):1525–38.
34. Folberg A, Kovács EN, Huang H, Houle M, Lohnes D, Featherstone MS. Hoxd4 and Rarg interact synergistically in the specification of the cervical vertebrae. *Mech Dev*. 1999;89(1–2):65–74.
35. Ghyselink NB, Dupé V, Dierich A, Messaddeq N, Garnier JM, Rochette-Egly C, Chambon P, Mark M. Role of the retinoic acid receptor beta (RARbeta) during mouse development. *Int J Dev Biol*. 1997;41(3):425–47.
36. Lohnes D, Mark M, Mendelsohn C, Dollé P, Dierich A, Gorczy P, Gansmüller A, Chambon P. Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. *Development*. 1994;120(10):2723–48.
37. Lohnes D, Kastner P, Dierich A, Mark M, LeMeur M, Chambon P. Function of retinoic acid receptor gamma in the mouse. *Cell*. 1993;73(4):643–58.
38. Houle M, Sylvestre JR, Lohnes D. Retinoic acid regulates a subset of Cdx1 function in vivo. *Development*. 2003;130(26):6555–67.
39. Akiyama H, Kim JE, Nakashima K, Balmes G, Iwai N, Deng JM, Zhang Z, Martin JF, Behringer RR, Nakamura T, et al. Osteo-chondroprogenitor cells are derived from Sox9 expressing precursors. *Proc Natl Acad Sci U S A*. 2005;102(41):14665–70.
40. Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrughe B. The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell*. 2002;108(1):17–29.
41. Nishimura R, Wakabayashi M, Hata K, Matsubara T, Honma S, Wakisaka S, Kiyonari H, Shioi G, Yamaguchi A, Tsumaki N, et al. Osterix regulates calcification and degradation of chondrogenic matrices through matrix metalloproteinase 13 (MMP13) expression in association with transcription factor Runx2 during endochondral ossification. *J Biol Chem*. 2012;287(40):33179–90.
42. Park E, Park SY, Dobkin C, Schuller-Levis G. Development of a novel cysteine sulfinic Acid decarboxylase knockout mouse: dietary taurine reduces neonatal mortality. *J Amino Acids*. 2014;2014:346809.
43. Park EJ, Mora JR, Carman CV, Chen J, Sasaki Y, Cheng G, von Andrian UH, Shimaoka M. Aberrant activation of integrin alpha4beta7 suppresses lymphocyte migration to the gut. *J Clin Invest*. 2007;117(9):2526–38.
44. Wagner N, Löhler J, Kunkel EJ, Ley K, Leung E, Krissansen G, Rajewsky K, Müller W. Critical role for beta7 integrins in formation of the gut-associated lymphoid tissue. *Nature*. 1996;382(6589):366–70.
45. DiTommaso T, Jones LK, Cottle DL, Gerdin AK, Vancollie VE, Watt FM, Ramirez-Solis R, Bradley A, Steel KP, Sundberg JP, et al. Identification of genes important for cutaneous function revealed by a large scale reverse genetic screen in the mouse. *PLoS Genet*. 2014;10(10):e1004705.
46. Knowles H, Heizer JW, Li Y, Chapman K, Ogden CA, Andreassen K, Shapland E, Kucera G, Mogan J, Humann J, et al. Transient Receptor Potential Melastatin 2 (TRPM2) ion channel is required for innate immunity against *Listeria monocytogenes*. *Proc Natl Acad Sci U S A*. 2011;108(28):11578–83.
47. Zou J, Ainscough JF, Yang W, Sedo A, Yu SP, Mei ZZ, Sivaprasadarao A, Beech DJ, Jiang LH. A differential role of macrophage TRPM2 channels in Ca²⁺ signaling and cell death in early responses to H₂O₂. *Am J Physiol Cell Physiol*. 2013;305(1):C61–69.

48. Yamamoto S, Shimizu S, Kiyonaka S, Takahashi N, Wajima T, Hara Y, Negoro T, Hiroi T, Kiuchi Y, Okada T, et al. TRPM2-mediated Ca^{2+} influx induces chemokine production in monocytes that aggravates inflammatory neutrophil infiltration. *Nat Med*. 2008;14(7):738–47.
49. Peled A, Sarig O, Samuelov L, Bertolini M, Ziv L, Weissglas-Volkov D, Eskin-Schwartz M, Adase CA, Malchin N, Bochner R, et al. Mutations in TSPEAR, Encoding a Regulator of Notch Signaling, Affect Tooth and Hair Follicle Morphogenesis. *PLoS Genet*. 2016;12(10):e1006369.
50. Shieh JT, Penon-Portmann M, Wong KHY, Levy-Sakin M, Verghese M, Slavotinek A, Gallagher RC, Mendelsohn BA, Tenney J, Belefors D, et al. Application of full-genome analysis to diagnose rare monogenic disorders. *NPJ Genom Med*. 2021;6(1):77.
51. Wang Z, Zhang H, Yang H, Wang S, Rong E, Pei W, Li H, Wang N. Genome-wide association study for wool production traits in a Chinese Merino sheep population. *PLoS ONE*. 2014;9(9):e107101.
52. Wang L, Wansleben C, Zhao S, Miao P, Paschen W, Yang W. SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. *EMBO Rep*. 2014;15(8):878–85.
53. Ding R, Savegnago R, Liu J, Long N, Tan C, Cai G, Zhuang Z, Wu J, Yang M, Qiu Y, et al. The SWINE Imputation (SWIM) haplotype reference panel enables nucleotide resolution genetic mapping in pigs. *Commun Biol*. 2023;6(1):577.
54. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
55. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
56. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. 2019;35(10):1786–8.
57. Yang Q, Cui J, Chazaro I, Cupples LA, Demissie S. Power and type I error rate of false discovery rate approaches in genome-wide association studies. *BMC Genet*. 2005;6 Suppl 1(Suppl 1):S134.
58. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, Szpiro AA, Chen W, Brehm JM, Celedón JC, et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet*. 2016;98(4):653–66.
59. Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, O'Brien SJ. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*. 2010;11:724.
60. Huang Y, Cai L, Duan Y, Zeng Q, He M, Wu Z, Zou X, Zhou M, Zhang Z, Xiao S, et al. Whole-genome sequence-based association analyses on an eight-breed crossed heterogeneous stock of pigs reveal the genetic basis of skeletal muscle fiber characteristics. *Meat Sci*. 2022;194:108974.
61. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263–5.
62. Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A. BioMart—biological queries made easy. *BMC Genomics*. 2009;10:22.
63. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. 2023;51(D1):D587–d592.
64. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res*. 2021;49(W1):W317–w325.
65. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. 2011;6(7):e21800.
66. Zhang Z, Kryvokhyzha D, Orsucci M, Glémin S, Milesi P, Lascoux M. How broad is the selfing syndrome? Insights from convergent evolution of gene expression across species and tissues in the *Capsella* genus. *New Phytol*. 2022;236(6):2344–57.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



三、其他业绩

聘

张译宾 老师:

书

在动物科学学院 2023 年青年教师教学能力比赛中
荣获“一等奖”

特发此证，以资鼓励



2023 年 12 月 22 日