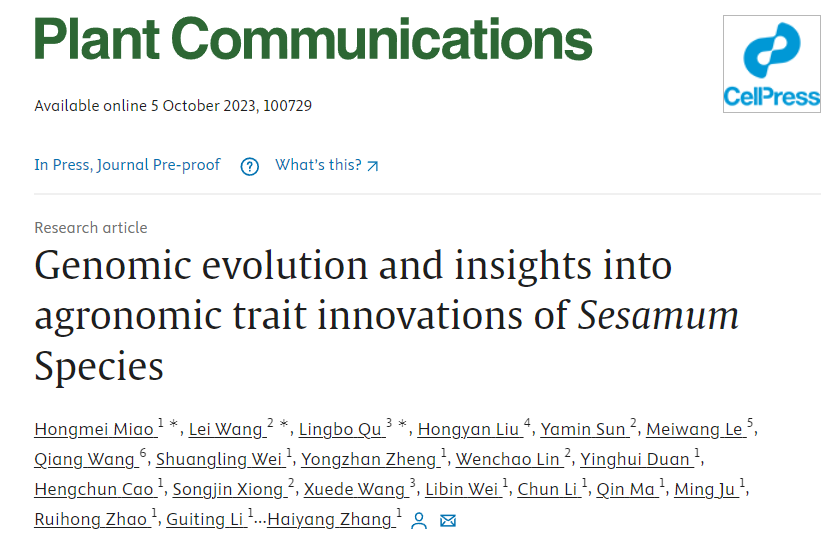
**我院张海洋团队在芝麻和野生近缘种基因组研究上获重大突破，揭示了胡麻属进化规律和芝麻含油量高但抗病性差的机理**

2023年10月2日，河南省农业科学院张海洋研究员带领的芝麻基因组研究小组（The Sesame Genome Working Group）在《*Plant Communications*》杂志在线发表了题为“Genomic evolution and insights into agronomic trait innovations of *Sesamum* species”的研究论文。该论文报道了胡麻属芝麻栽培种和6个野生种染色体级别的基因组精细图，首次揭示了胡麻属物种的进化规律及其在双子叶植物进化中的重要地位。研究发现，在胡麻属不同种群进化中，基因组结构变异对芝麻栽培种株型、含油量及枯萎病抗性水平等重要农艺性状产生了重大影响。发掘出了芝麻高含油量基因*SiNAC1*（提高6.2个百分点)、抗枯萎病基因*ScaDIR40*（免疫），首次揭示了芝麻栽培种较野生种含油量高而抗病性差的机理。



芝麻（*Sesamum indicum* L.）属胡麻科胡麻属，起源于非洲，是世界上最为古老的油料作物之一。芝麻主要种植在热带和亚热带地区，具有喜温耐旱、耐瘠薄等优点，籽粒富含不饱和脂肪酸和抗氧化物质木酚素，被誉为“油料皇后”。我国是世界芝麻四大主产国之一，也是最大的消费国。为推动芝麻分子育种研究进程，2010年河南省农业科学院、南开大学、河南工业大学、江西省农业科学院、安徽省农业科学院等6家芝麻科研团队组建了芝麻基因组研究小组，启动了芝麻基因组计划，并于2013年首次公布了芝麻栽培种*De novo*拼接基因组图。

胡麻属包含36个种（Species），涉及3个染色体组型（n = 13、16和 32），芝麻是胡麻属中的唯一栽培种。作为芝麻基因组计划的主体内容，本研究构建了胡麻属芝麻栽培种（*S. indicum*，n = 13）和6个野生种（*S. alatum*，n=13； *S. latifollum*， n=16；*S. angolense*，n=16；*S. calycinum*，n=16；*S. angustifolium*，n=16；*S. radiatum*，n=32）染色体水平的的基因组精细图，涵盖了属内全部3个染色体组型。栽培种基因组拼接大小为312.95 Mb，6个野生种基因组拼接大小为300.7 Mb（*S. angustifolium*）- 668.4 Mb（*S. radiatum*）。BUSCO结果显示，7个基因组的拼接完整度为91.2-98.5%。



图1 芝麻基因组圈图

基因组直系同源基因Ks分布结果显示，这7个基因组均经历了全基因组复制（WGD）和基因组三倍体化（WGT-γ）事件。此外，*S. radiatum* 还发生有第二次全基因组复制（WGD-2）事件，表明该种为四倍体（**图2A**）。7个种的染色体组原位杂交（GISH）及着丝粒重复序列荧光原位杂交（FISH）的结果进一步证明，*S. radiatum* 基因组包含了两套不同的着丝粒重复序列（Centromere repeats），为异源四倍体，被命名为AABB（**图2B**）。其中，AA*Sra*亚基因组与n = 16的二倍体野生种 *S. latifolium* 同为AA类，BB*Sra*亚基因组与n = 16的*S. angolense*、*S. calycinum*、*S. angustifolium*等3个二倍体野生种为BB类。二倍体栽培种 *S. indicum* （n = 13）和*S. alatum* （n = 13）具有各自特异的着丝粒重复序列，分别被命名为CC和DD类基因组（**图2C**）。

单拷贝同源基因进化分析显示，胡麻属分化发生在第三纪时期的48.5–19.7百万年前（MYA）（**图2C**）。其中n = 13的野生种*S. alatum*（DD）最早先分化出来，随后是栽培种（n = 13, CC），其次是n = 16的野生种*S. latifolium*（AA）。*S. latifolium*分化时间在34.2 MYA，其他四个野生种 (n = 16) 的分化时间为 27.2–19.7 MYA。推测*S. radiatum*异源四倍体化的时间应该在~5 MYA。在该时期有多种植物发生多倍体化事件。研究认为胡麻属物种的进化路线为：n = x= 13→ n= x= 16→ n = 2x = 32；物种向着染色体数目增多的方向进化。胡麻属不同种的长时期进化特点，也反映出该属物种在双子叶植物进化中的相对古老地位。



**图2. 胡麻属基因组进化特征**

泛基因组分析结果显示，胡麻属7个基因组共有25,401个基因家族，包含202,685个putative基因（**图**3）。核心基因家族（Core families）共有9,164（36.1%）个；保守线性基因家族比例较大豆属（*Glycine*，50.1%）和棉属（*Gossypium*，61.8%）的低，进一步表明胡麻属物种的发生应处于重要的物种扩展时期。



**图3 胡麻属7个种基因组共线性比较与泛基因组分析**

同时，研究小组系统分析了栽培种及野生种的基因组结构变异（Structural variations，SV）特征，发现栽培种基因组中发生有多个SV变异，并对芝麻关键农艺性状产生重大影响。例如，4166bp的Copia\_LTR\_反转座子序列插入和24.9 Kb的大片段基因组序列缺失，分别引发了PEBP家族中的*SiPT1*和*SiDT1*基因发生SV变异，进而导致芝麻株型和花序类型性状发生改变（**图4**）。



图4**胡麻属株型和花序类型调控基因**

芝麻枯萎病是芝麻两大真菌病害之一。与已测序的*S. calycinum*、*S. latifolium*、*S. angolense*、*S. radiatum* 等高抗病野生种相比，芝麻栽培种对强致病力枯萎病菌*F. oxysporum* f. sp. *sesami* （FOS）菌株则表现为高感水平。为揭示栽培种与野生种在枯萎病抗性方面的差异，研究小组开展了*S. calycinum*（高抗枯萎病）×*S. angustifolium*（高感枯萎病）种间杂交群体的GWAS关联分析和SV分析（**图5**），成功克隆出芝麻抗枯萎病基因*ScaDIR40*。该基因编码Dirigent（DIR）蛋白，调控木酚素和木质素合成过程中的有机松柏醇类耦合反应，并能在次生细胞壁发生过程中提高植物的抗病抗逆性。在高感病野生种*S. angustifolium*中，等位基因*Saudir32*的5′-UTR -210bp处，插入了一段1411bp 的Copia\_LTR\_反转座子（**图 5D**）。在高感病栽培种中，同源基因*Sindir37* 的-1bp到-1400bp的5′-UTR 序列以及部分CDS序列则发生了丢失（**图5E**）。上述SVs的独立发生均导致了*DIR*基因功能发生改变，并最终导致植株枯萎病抗性丧失。研究为芝麻远缘杂交、基因编辑等抗病育种研究提供了重要理论和技术支持。



图5**胡麻属枯萎病抗性调控基因*DIR***

为揭示芝麻含油量高的分子机理，研究小组开展了油脂合成代谢相关基因家族（FA）分析（**图6**）和560份芝麻核心种质3年3点8个环境下的GWAS分析（**图7**）。发掘出了芝麻高含油量基因*SiNAC1*，首次证实*SiNAC1*参与调控植物油脂合成过程（**图7**）；该基因的T333G碱基突变对芝麻含油量表型变异的最高解释率为43.8%，籽粒含油量可提高6.2个百分点。研究发现*NAC1*基因在野生种中存在结构变异；*SiNAC1*能够共表达调控*FAD6*、*NPC1*、*CYP96A10*、*FAB1B*、*FAR2*、*DPL1*、P*DAT*等不饱和脂肪酸生物合成与代谢调控相关基因（**图 7**）。同时，研究小组还对另一个含油量相关基因*SiPPO*进行了家族分析和过表达转基因功能验证。研究为芝麻及其他油料作物品种改良提供了重要组学信息。



**图6 芝麻脂肪酸合成路径及人工选择区域的重要优异基因群分布**



**图7 芝麻含油量性状相关基因功能分析**

本研究得到了国家特色油料产业技术研发中心和河南省特色油料作物基因组学重点实验室的支持。感谢中国科学院生物物理研究所Joy Fleming博士对本研究给予的重要建议及论文修改贡献！