

小麦 A 基因组测序与进化研究进展

史晓黎¹, 何伊琳^{1,2}, 凌宏清^{1,2}

1. 中国科学院遗传与发育生物学研究所, 植物细胞与染色体工程国家重点实验室, 北京 100101

2. 中国科学院大学生命科学学院, 北京 101408

摘要: 小麦是世界上广泛种植的主要粮食作物, 养活了全世界 35% 以上的人口。获取高质量的基因组图谱对于推动小麦基础理论和遗传育种研究至关重要。然而, 庞大而复杂的基因组一度使小麦基因组测序被认为是“不可能完成的任务”。随着高通量测序和组装技术的成熟, 近年来多个小麦基因组序列图谱陆续发布, 序列组装质量日臻完善。仅最近两年就公布了 5 个不同倍性的小麦参考基因组序列, 包括两个二倍体祖先种乌拉尔图小麦(*Triticum urartu*, AA)和粗山羊草(*Aegilops tauschii*, DD)、野生和栽培四倍体二粒小麦(*T. turgidum* ssp. *dicoccoides*, BBAA)和六倍体普通小麦(*T. aestivum*, BBAAADD)。其中, 作为多倍体小麦 A 亚基因组供体的乌拉尔图小麦基因组测序和分析是由中国科学院遗传与发育生物学研究所牵头完成。本文主要对小麦 A 基因组的结构解析和进化分析等方面的研究进展进行了综述, 以期对相关领域的科研人员提供参考信息, 促进小麦的基础和应用研究。

关键词: 小麦; 乌拉尔图小麦; 小麦 A 基因组; 基因组测序; 染色体演化

Progress on wheat A genome illustration and its evolutionary analysis

Xiaoli Shi¹, Yilin He^{1,2}, Hongqing Ling^{1,2}

1. State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

2. College of Life Sciences, University of Chinese Academy of Sciences, Beijing 101408, China

Abstract: Wheat is one of the main food crops and widely grown in the world. It feeds more than 35% of the world's population. Obtaining high-quality genome sequences of wheat is important for its basic and breeding researches. However, the large and complex genome of wheat once led to its genome sequencing as an "impossible task". Recently, with the development of high-throughput sequencing and assembly technology, many wheat genome sequences have been released, and their sequencing and assembly quality is being improved continuously. In the last two years, five wheat reference genomes with different ploidy levels have been published, including two diploid ancestors *Triticum urartu* (AA) and

收稿日期: 2019-08-10; 修回日期: 2019-09-10

基金项目: 国家自然科学基金项目(编号: 31871273)资助[Supported by the National Natural Science Foundation of China (No.31871273)]

作者简介: 史晓黎, 博士, 助理研究员, 研究方向: 生物信息和植物比较基因组学。E-mail: xlshi@genetics.ac.cn

通讯作者: 凌宏清, 研究员, 博士生导师, 研究方向: 小麦基因组学和植物营养分子生物学。E-mail: hqling@genetics.ac.cn

DOI: 10.16288/j.yczs.19-233

网络出版时间: 2019/9/16 14:55:15

URI: <http://kns.cnki.net/kcms/detail/11.1913.R.20190916.0912.001.html>

Aegilops tauschii (DD), wild and cultivated tetraploid wheat *T. turgidum* ssp. *dicoccoides* (BBAA) and hexaploid wheat *T. aestivum* (BBAADD). Among them, the sequencing and analysis of the *T. urartu* genome, a donor of polyploid wheat A subgenome, was led by the Institute of Genetics and Developmental Biology of the Chinese Academy of Sciences. In this review, we summarize the research progress on structure and evolution analyses of the *T. urartu* genome to provide some valuable information for promoting the basic and applied researches of wheat.

Keywords: wheat; *Triticum urartu*; wheat A genome; genome sequencing; chromosome evolution

小麦(*Triticum aestivum* L.)是世界上适应性最广、栽培面积最大的主要粮食作物,年种植面积达 2.2 亿公顷,年产量超过 7.4 亿吨,养活了世界上 35% 的人口,分别提供了人类所需热能和蛋白质的 19% 和 20% (<http://www.fao.org/>, <http://www.wheatgenome.org>)。我国是世界上小麦种植面积最大和总产量最高的国家,也是全球最大的小麦消耗国,常年种植面积在 2400 万公顷左右,年产量超过 1.2 亿吨(<http://www.xiaomai.cn/html/news/20150226/363917.html>)。随着世界人口的不断增加,预计到 2050 年将达到 96 亿,对小麦的需求量将在现有产能基础上增加 60%^[1],再加之全球气温升高和极端天气频发,使小麦生产面临着严峻挑战。为了保证我国乃至全世界的粮食安全,维持小麦生产的增产和稳产,迫切需要加强小麦的基础理论和育种应用研究,培育出更多环境友好型的高产优质小麦新品种。

世界三大主要粮食作物中,水稻和玉米基因组的测序分别于 2002 年和 2009 年相继完成^[2~4],并有力地推动了这两大作物的基础和分子育种研究。小麦基因组庞大(16 Gb)且含有大量重复序列,致使小麦基因组测序和组装变得极为困难。基因组测序的缓慢进展,严重制约了小麦的基础和应用研究。近年来,在科学家们的不懈努力下,随着高通量测序和生物信息分析技术的发展,小麦基因组测序研究取得了重大突破。自 2003 年普通小麦 3B 染色体物理图谱发表以来^[5],许多不同倍性的小麦基因组测序研究结果相继发表。仅在 2017 和 2018 年就先后公布了 A^[6]、D^[7]、AB^[8]和 ABD^[9] 4 个小麦基因组的参考序列,其中乌拉尔图小麦 A 基因组的测序与分析是由中国科学院遗传与发育生物学研究所植物细胞与染色体工程国家重点实验室主导完成。本文总结了本实验室多年来在乌拉尔图小麦 A 基因组测序

和进化研究领域的相关工作以及小麦 A 基因组的测序研究进展,以期为相关领域的科研人员提供参考信息,促进小麦的基础理论和分子育种研究。

1 小麦的起源及染色体组型

小麦是小麦属(*Triticum* L.)植物的统称,起源于中东的新月沃地(Fertile Crescent)。根据含有的染色体组型,可以将小麦属归纳为 6 个种,包括 2 个二倍体种、2 个四倍体种和 2 个六倍体种,有 A、B、D 和 G 4 种染色体组。二倍体小麦皆为 A 染色体组,如乌拉尔图小麦(*T. urartu* Thumanjan ex Gandilyan, AA, $2n=2x=14$)和一粒小麦(*T. monococcum* L., A^mA^m, $2n=2x=14$)。四倍体小麦中具有 AB 和 AG 两类染色体组型。具 AB 染色体组的四倍体小麦(*T. turgidum* L., BBAA, $2n=4x=28$)是由 A 基因组供体乌拉尔图小麦与 B 基因组供体拟斯卑尔脱山羊草(未完全确定)(*Aegilops speltoides*, SS, $2n=2x=14$)在 50~300 万年前经天然杂交和染色体加倍而形成,其中包括野生二粒小麦(*T. turgidum* ssp. *dicoccoides* (Körn. Ex Asch. & Graebn.) Thell.)、栽培二粒小麦(*T. turgidum* ssp. *dicoccum* (Schränk ex Schübl.) Thell.)、圆锥小麦(*T. turgidum* L.)和硬粒小麦(*T. durum* Desf.)等不同类型。而 AG 染色体组的四倍体小麦也是由含 A 基因组的乌拉尔图小麦与 1 种含有 G 基因组的拟斯卑尔脱山羊草(未知种)杂交和染色体加倍而来(*T. timopheevii* Zhuk, AAGG, $2n=4x=28$),包括野生阿拉拉丁小麦(*T. timopheevii* ssp. *araraticum*)和栽培的提莫菲维小麦(*T. timopheevii* ssp. *timopheevii*)两类。六倍体小麦在自然界中存在 ABD 和 A^mAG 两个染色体组类型。具 ABD 染色体组的普通小麦(*T. aestivum* L., BBAADD, $2n=6x=42$)是由栽培二粒小麦(*T. turgidum*

ssp. *Dicoccum*, BBAA, $2n=4x=28$)与具有 D 染色体组二倍体粗山羊草(*Ae. tauschii* (Coss.) Schmal, DD, $2n=2x=17$)在大约 9000 年前经天然杂交和染色体加倍而形成。而具 A^mAG 染色体组的茹科夫斯基小麦(*T. zhukovskyi* Menabde&Ericzjan, A^mA^mAAGG, $2n=6x=42$)是由四倍体提莫菲维小麦(AAGG)与一粒小麦(A^mA^m)杂交进化而来^[10,11]。

在小麦属的 6 个物种中,六倍体普通小麦的种植面积最广,占全世界小麦总生产的 95%,其次是四倍体硬粒小麦,约占 5%^[10]。其余的二倍体栽培一粒小麦、四倍体提莫菲维小麦和六倍体茹科夫斯基小麦只在局部地区或在植物园里有少量种植。

2 小麦 A 基因组测序与结构解析

如前所述,小麦属所有的 6 个物种中都含有 A 基因组,表明小麦 A 基因组是小麦进化的基础性基因组,在小麦进化过程中起着核心作用。含有 A 基因组的二倍体乌拉尔图小麦是小麦 A 基因组的原始供体种,主要分布在中东地区,如黎巴嫩、伊朗、伊拉克、土耳其和叙利亚等,其基因组大小是普通小麦基因组的 1/3,约为 5 Gb。因此,对二倍体乌拉尔图小麦基因组进行测序,将大大简化基因组测序研究的难点,同时可为普通小麦基因组分析提供参考,也将为多倍体小麦的进化与驯化研究提供重要信息。此外,乌拉尔图小麦基因组的测序还有助于鉴定和分离普通小麦 A 基因组上的重要农艺性状基因以及加快小麦的遗传改良。

鉴于小麦 A 基因组的重要性,中国科学院遗传与发育生物学研究所植物细胞与染色体工程国家重点实验室组织了小麦基因组测序攻关团队,以小麦 A 基因组原始供体种乌拉尔图小麦为研究对象,采取先草图后精细图的分步实施策略,开展乌拉尔图小麦基因组测序研究。

2.1 乌拉尔图小麦基因组草图绘制

2009 年,小麦基因组测序攻关团队与华大基因研究院以及美国加州大学戴维斯分校的 Jan Dvorak 教授、罗明成博士等合作,利用当时刚出现的高通量二代测序技术,选取二倍体乌拉尔图小麦 G1812

系为材料,进行全基因组鸟枪法测序。利用 Hiseq2000 测序平台,对 57 种不同插入片段大小(200~20 000 bp)的测序文库进行测序,共获得 450 Gb 的有效序列,用华大基因研究院开发的 SOAPdenovo (v.1.05)基因组组装软件进行了拼接,通过 4 年的艰辛研究,于 2013 年率先在国际上完成和公布了乌拉尔图小麦基因组序列草图^[12]。该草图总长为 4.66 Gb, Scaffold N50 长度为 63.6 kb,不含 N 的完整序列总长为 3.92 Gb, Contig N50 长度为 3.42 kb,预测出 34 879 个蛋白编码基因。乌拉尔图小麦基因组序列草图的完成和公布,得到了国内外的广泛关注和新闻媒体的大量报道,研究成果入选了 2013 年度中国科学十大进展。

2.2 乌拉尔图小麦基因组精细图谱绘制

在完成乌拉尔图小麦基因组草图绘制的基础上,为了获得高质量的小麦 A 基因组的参考基因组序列,小麦基因组测序攻关团队与中国科学院遗传与发育生物学研究所基因组分析平台梁承志研究员合作,采用 BAC-by-BAC 测序策略,并与第三代测序技术 PacBio (Pacific Biosciences sequencing)大片段单分子实时(single-molecular real-time, SMRT) 测序和新的基因组物理图谱构建方法(BioNano genome map 和 10× Genomics linked reads)相结合,对乌拉尔图小麦 G1812 系基因组进行了测序和组装,并于 2018 年 5 月公布了乌拉尔图小麦基因组的高质量参考基因组序列^[6]。首先构建了 *EcoR*、*Hind* 和 *Mbo* 共 3 种不同酶切的 BAC 文库,对 451 584 个 BAC 克隆进行指纹分析,搭建出 20 702 个 BAC 重叠群,利用最小重叠法原则(minimal tiling path principle)筛选出 47 223 个 BAC 克隆,混合成 984 个 BAC 混池(48 BACs/池);提取 BAC DNA,构建了 300 bp 的插入片段测序文库,利用高通量的 Illumina HiSeq2500 平台进行测序,然后对单个 BAC 序列进行组装,并用纠错后的全基因组 SMRT 序列进行补洞;进一步根据 BAC 重叠群信息、BioNano 物理图谱和 10× Genomics linked reads 将获得的 BAC 序列进行拼接和组装,最终获得长度为 4.86 Gb 的组装基因组(Scaffold N50 为 3.67 Mb),其中无 N 序列总长为 4.79 Gb (Contig N50 为 344 kb),为乌拉尔图小麦基因组(4.94 Gb)的 97%。预测出 41 507 个蛋白编码基

因, 基因平均长度为 1453 bp, 编码蛋白的平均长度为 332 氨基酸。此外, 还鉴定出了 31 269 个 miRNAs、5810 个 lncRNAs、3620 个 tRNAs、80 个核糖体 rRNAs 和 2519 个 snRNAs。并利用高密度的单核苷酸多态性(single nucleotide polymorphism, SNP)遗传图谱将 4.67 Gb 序列铆钉到了染色体的相应位置, 绘制出小麦 A 基因组 7 条染色体的假分子(pseudomolecule), 其从第 1 至第 7 条染色体长度分别为 583 994 449、755 818 500、747 043 632、619 526 116、662 474 153、576 970 508 和 719 848 703 bp。基因组各组分(如基因、重复序列、SSR 等)在染色体上分布的偏好性分析显示, 在每条染色体上的两个末端区的基因密度和重组交换率显著高于染色体的中间区域, 而重复序列主要分布在染色体中间区域^[6]。

3 比较基因组学及染色体进化分析

利用获得的高质量乌拉尔图小麦基因组数据, 研究人员对基因组中重复序列、基因家族的组成特征、庞大基因组的形成原因、染色体演化、群体基因组学等方面进行深入分析。

3.1 乌拉尔图小麦基因组中的重复序列

基因组注释显示, 乌拉尔图基因组中 81.42% (3.9 Gb) 的组装序列为重复序列, 其中 RNA 反转座子重复序列占 71.83% (3.44 Gb), DNA 反转座子重复序列为 7.41% (355 Mb)。在反转座子重复序列中, 主要是长末端重复反转座子(long-terminal repeat retrotransposons, LTR-RTs), 由 Gypsy 和 Copia 两类 LTR 反转座子组成, 它们分别占全基因组的 42.71% 和 24.30%, 在乌拉尔图小麦基因组的进化过程中起到了重要作用。通过计算完整 Gypsy 和 Copia 反转座子 LTR 区的突变率估计其插入时间, 发现 Gypsy 类反转座子插入高峰大约在 1.5 百万年前, 而 Copia 类反转座子的插入高峰时间约在 0.8~0.9 百万年, 从而证明在乌拉尔图小麦基因组进化过程中先发生了 Gypsy 的爆发, 然后再出现 Copia 反转座子的大量插入。并且这两个最大组分的反转座子在染色体上的分布偏好性也存在明显差异, Gypsy 主要分布在染色体中间部位, 而 Copia 更多分布在染色体末端。

3.2 基因家族比较分析

利用组装的乌拉尔图小麦基因组序列预测的基因家族与已报道的禾本科植物二穗短柄草(*Brachypodium distachyon*)、水稻(*Oryza sativa*)、高粱(*Sorghum bicolor*)和玉米(*Zea mays*)的基因家族进行了比较分析, 鉴定出 1567 个乌拉尔图小麦基因组特异基因家族, 含 4610 个基因, 功能注释显示多数基因参与逆境胁迫反应^[6]。另外, 还发现含 NB-ARC 功能域的抗病基因在乌拉尔图小麦基因组中发生了大量扩增^[12], 并推断这些基因的扩张可能赋予小麦抵御恶劣生存环境的广泛适应性。对转录因子家族的比较分析显示, B3 转录因子家族中的 REM (reproductive meristem) 亚家族在普通小麦、乌拉尔图小麦和粗山羊草的基因组中均呈现显著扩增。REM 基因主要参与春化及花发育, 编码 B3 REM 转录因子的基因在麦类作物的扩增很可能和小麦的春化反应有关^[6]。

3.3 乌拉尔图小麦基因组扩张机制及结构变异

乌拉尔图小麦基因组大小约为 5000 Mb, 是玉米基因组(2300 Mb^[4])的 2.3 倍、高粱基因组(730 Mb^[13])的 6.8 倍、水稻基因组(420 Mb^[2])的 12 倍、二穗短柄草基因组(272 Mb^[14])的 18.4 倍, 这表明乌拉尔图小麦基因组在进化过程中发生了大量扩增。基因组测序为揭示基因组扩增的机制提供了机遇。

通过对二穗短柄草与乌拉尔图小麦的共线性分析, 发现乌拉尔图小麦基因组中约 21% 的基因在基因密度上与对应的二穗短柄草共线性区段的基因密度相当, 但 79% 的乌拉尔图小麦基因在进化过程中被大量的 Gypsy 和 Copia 反转座子插入隔离, 使其基因间的平均距离与二穗短柄草基因组相比增大了近 20 倍^[12], 这是首次从全基因组水平上解释了小麦基因组庞大的原因。

在乌拉尔图基因组内识别了 5 个成对的共线性旁系同源基因区域, 它们是 7000 万年前的一次全基因组复制事件的产物。对该复制事件在水稻基因组内遗留的成对共线性旁系同源基因在乌拉尔图小麦中的保守性分析显示, 高达 85% 的水稻同源基因对在乌拉尔图小麦中已检测不到。其中, 47% 的基因对

在乌拉尔图小麦中发生了单拷贝缺失, 38% 的水稻基因对的两个拷贝在对应的乌拉尔图小麦的同源共线区域内均缺失^[6]。这些结果表明在与水稻分离后, 乌拉尔图小麦的基因组发生过大规模的重组或变异。

比较乌拉尔图小麦基因组序列和已报道的中国春小麦基因组序列, 鉴定出了 3 个大的染色体结构变异: (1) 乌拉尔图小麦 4 和 5 号染色体间的长臂末端发生相互易位, 该易位发生在 A、B 和 D 基因组分开后, 四倍体小麦形成前; (2) 中国春小麦基因组中 7B 向 4A 的单向染色体片段易位; (3) 中国春 4A 染色体中的一个含着丝粒的大片段倒位。后面两个结构变异时间发生在四倍化过程中或四倍化后^[6]。此外, 对乌拉尔图小麦 7 号染色体和中国春 7A 染色体的 DNA 序列比较发现, 两个染色体的 DNA 序列相似性为 90%, 但各具各自独有的转座子插入^[6]。

3.4 乌拉尔图小麦染色体进化分析

染色体是基因的载体, 不同物种具有不同的染色体数目。从进化角度来看, 禾本科的所有物种都是从一个共同祖先逐渐演化而来。在已测序的禾本科物种中, 水稻 $1n = 12$ 、高粱和玉米 $1n = 10$ 、二穗短柄草 $1n = 5$, 而乌拉尔图小麦 $1n = 7$ 。那么乌拉尔图小麦基因组的 7 条染色体是如何从禾本科植物共同祖先的染色体组演化而成? 通过将乌拉尔图小麦基因组中的基因与水稻、高粱和二穗短柄草基因组内旁系同源基因的共线性分布的分析, 结合 Salse 等^[15]的染色体进化假设, 研究人员推演出乌拉尔图小麦 7 条染色体的进化模型。便于理解本文对该模型进行了进一步总结归纳, 绘制了乌拉尔图小麦 7 条染色体进化示意图 (图 1)。如图所示, 禾本科植物鼻祖的染色体基数为 7, 在距今大约 7000 万年前经历了一次全基因组复制^[16], 变为 14, 其中 4 条染色体发生两两融合, 形成了禾本科植物的 12 条祖先染色体。这 12 条禾本科祖先染色体在水稻中得以很好保存, 而这 12 条祖先染色体再经断裂、融合和染色体片段重排等, 演化出了乌拉尔图小麦的 7 条染色体。乌拉尔图小麦的 3 号和 6 号染色体分别来自禾本科祖先的 1 号和 2 号 (也即水稻 1 号和 2 号) 染色体。而乌拉尔图小麦的 1 号、2 号、4 号和 7 号染色体分别由祖先的两条染色体融合而成。其中, 祖先

的 10 号 (水稻 10 号) 染色体插入祖先的 5 号 (水稻的 5 号) 染色体的着丝粒区, 形成乌拉尔图小麦的 1 号染色体; 祖先的 7 号 (水稻 7 号) 染色体插入到祖先的 4 号 (水稻 4 号) 染色体的着丝粒区, 形成乌拉尔图小麦的 2 号染色体; 祖先的 11 号 (水稻 11 号) 染色体插入祖先的 3 号 (水稻 3 号) 染色体的着丝粒区, 形成乌拉尔图小麦的 4 号染色体; 祖先的 8 号 (水稻 8 号) 染色体插入到祖先的 6 号 (水稻 6 号) 染色体的着丝粒区, 形成了乌拉尔图小麦的 7 号染色体^[6]。这种由着丝粒区插入完成的染色体融合在二穗短柄草中也有报道, 并且在二穗短柄草中还观察到更为复杂的多层嵌套插入^[14]。区别于上述染色体融合方式, 乌拉尔图小麦的 5 号染色体由祖先的 12 号 (水稻 12 号) 和 9 号 (水稻 9 号) 染色体头尾相接而成, 此外其 3' 端粒区还融合了位于祖先 3 号 (水稻 3 号) 染色体的 3' 近端粒区和 5' 端粒区的两个染色体小片段^[6]。水稻、高粱、玉米、乌拉尔图小麦和二穗短柄草的进化树显示^[14], 二穗短柄草与乌拉尔图小麦有更近的亲缘关系, 但这两个物种的染色体演化模式完全不同, 表明乌拉尔图小麦和二穗短柄草的染色体组进化是独立发生的。

3.5 乌拉尔图小麦群体基因组学分析

如前所述, 乌拉尔图小麦主要分布在新月沃地。利用从亚美尼亚、伊朗、伊拉克、叙利亚、土耳其和黎巴嫩等 6 个国家不同生态环境收集到的 147 份乌拉尔图小麦系 (accession) 为材料, 进行了群体基因组学分析; 通过叶片转录组测序, 鉴定出 144 806 个高质量 SNP, 根据获得的 SNP 进行系统发育分析, 可将 147 份乌拉尔图小麦聚类成 3 个大类群^[6]。类群 1 (第 1 组) 包含来自多个国家的 30 个品种; 类群 2 (第 2 组) 包含 64 个品种, 其中 88% 来自黎巴嫩; 类群 3 (第 3 组) 包含 53 个品种, 其中 92% 来自土耳其。这 3 个类群材料的采集地点在海拔高度上存在明显差异, 类群 2 大多数材料的采集海拔高度超过 1000 米, 而绝大多数类群 1 和类群 3 材料的采集海拔高度低于 1000 米。通过对小麦白粉菌 *Blumeria graminis* f. sp. *tritici* (Bgt, race E09) 接种鉴定发现, 乌拉尔图小麦的这 3 个类群在抗白粉病菌侵染方面存在明显差异。来自黎巴嫩高海拔的类群 2 (第 2 组) 中, 92.2% 的系表现出抗性, 而来自于低海拔的类群 1 (第 1 组)

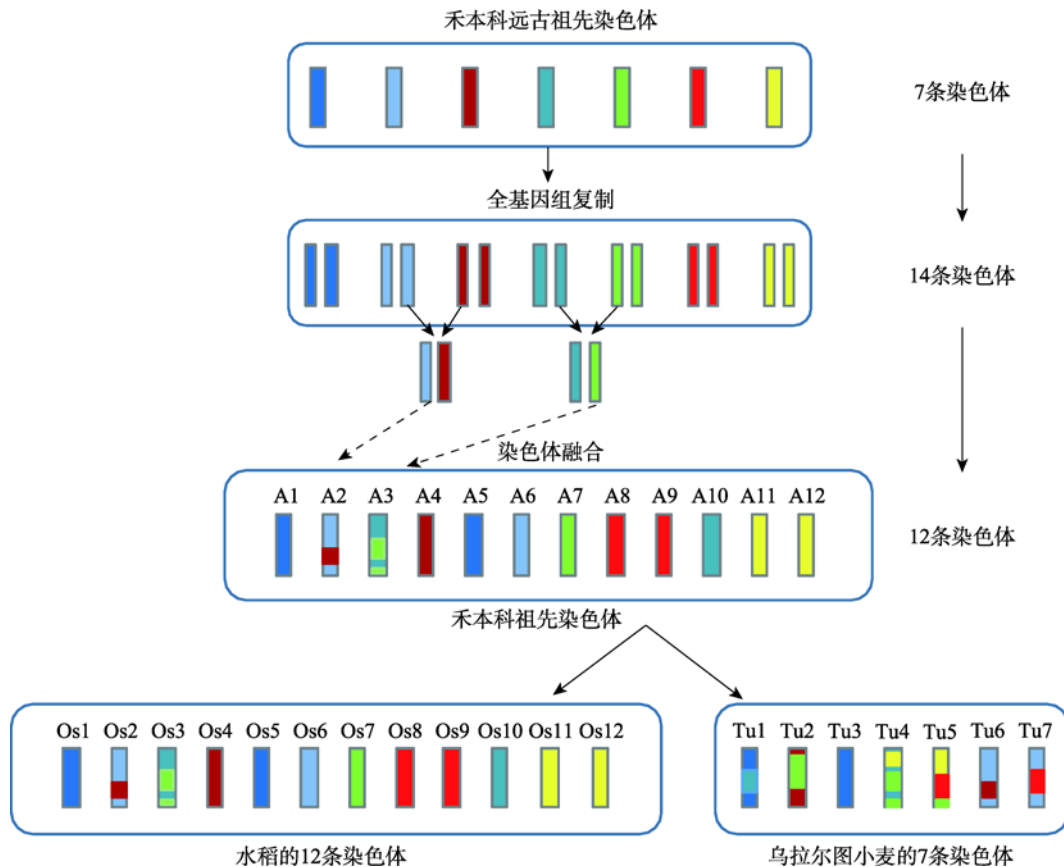


图 1 乌拉尔图小麦 7 条染色体的进化模型

Fig. 1 Evolutional model of the 7*T. urartu* chromosomes

和 3 (第 组)的大部分系(分别为 96.7%和 90.6%)为易感型^[6]。这些研究结果暗示不同海拔高度生态环境与乌拉尔图小麦适应性进化密切相关。

3.6 乌拉尔图小麦基因组的测序促进了小麦的基础研究

乌拉尔图小麦是普通小麦 A 亚基因组的供体, 具有较六倍体小麦更为简单的基因组结构, 对于研究小麦 A 基因组基因型和表型的相互关联能够提供更加直接和确定性的证据支持。受多倍化和长期驯化的影响, 普通小麦 A 亚基因组与之相比已经有了相当的变化, 为研究来自不同亚基因组同源基因在多倍化和驯化过程中发生的变化提供依据。乌拉尔图小麦是小麦抗白粉病基因的重要来源。利用乌拉尔图小麦基因组序列设计的分子标记, 结合遗传作图和 RNA 测序, Zou 等^[17]在很短时间从具有白粉病抗性的乌拉尔图小麦 PI428309 品系中定位并克隆

了白粉病抗性基因, 这也是第一个从乌拉尔图小麦中克隆的抗病基因。此外, 研究人员利用水稻等作物的已知基因序列与乌拉尔图小麦序列进行比对, 鉴定和克隆了多个小麦重要农艺性状基因, 如控制小麦粒长的 *GASR7* 基因。Dong 等^[18]利用水稻 *GASR7* 基因序列与乌拉尔图小麦基因组序列比对, 获得了水稻 *GASR7* 在小麦 A 基因组上的同源基因, 利用乌拉尔图小麦基因组序列设计引物, 扩增出了六倍体小麦基因组中 A 亚基因组的 *TaGASR7*, 并通过 *TaGASR7-A1* 基因的单倍型分析, 结合表型数据证明了该基因在普通小麦中参与对籽粒性状的调控。这些结果表明, 乌拉尔图小麦基因组结构解析推动了小麦的基础和应用研究。

4 小麦 A 基因组研究的未解之谜

综上所述, 小麦的 A 基因组在多倍体小麦进化

过程中起到了核心作用。它的起源很明确,来自于二倍体乌拉尔图小麦。普通小麦的 A 亚基因组经历了从供体乌拉尔图小麦→野生四倍体小麦→栽培四倍体小麦→栽培六倍体小麦的进化和驯化。在这一系列进化和驯化过程中, A 基因组如何改变其自身的结构、基因组成、基因表达等来适应多倍化环境和接纳 B、D 基因组这一基础科学问题目前仍是未解之谜。二倍体供体乌拉尔图小麦^[6]、野生二粒小麦^[8]、栽培二粒小麦^[19]和六倍体普通小麦^[9]基因组测序的完成和高质量参考基因组序列的公布,为研究这一基础科学问题提供了前所未有的新机遇。通过深入比较分析将揭示多倍体物种进化和人工驯化过程的分子机制及基础理论,同时也有助于小麦基因组的功能分析和遗传改良。

参考文献(References):

- [1] Rosegrant MW, Cline SA. Global food security: challenges and policies. *Science*, 2003, 302(5652): 1917–1919. [\[DOI\]](#)
- [2] Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Japonica*). *Science*, 2002, 296(5565): 92–100. [\[DOI\]](#)
- [3] Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 2002, 296(5565): 79–92. [\[DOI\]](#)
- [4] Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddelloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 2009, 326(5956): 1112–1115. [\[DOI\]](#)
- [5] Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeier W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Bergès H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C. A physical map of the 1-gigabase bread wheat chromosome 3B. *Science*, 2008, 322(5898): 101–104. [\[DOI\]](#)
- [6] Ling HQ, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, Yu Y, Du H, Qi M, Li Y, Lu H, Yu H, Cui Y, Wang N, Chen C, Wu H, Zhao Y, Zhang J, Li Y,

- Zhou W, Zhang B, Hu W, van Eijk MJT, Tang J, Witsenboer HMA, Zhao S, Li Z, Zhang A, Wang D, Liang C. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature*, 2018, 557(7705): 424–428. [DOI]
- [7] Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, McGuire PE, Liu S, Long H, Ramasamy RK, Rodriguez JC, Van SL, Yuan L, Wang Z, Xia Z, Xiao L, Anderson OD, Ouyang S, Liang Y, Zimin AV, Pertea G, Qi P, Bennetzen JL, Dai X, Dawson MW, Müller HG, Kugler K, Rivarola-Duarte L, Spannagl M, Mayer KFX, Lu FH, Bevan MW, Leroy P, Li P, You FM, Sun Q, Liu Z, Lyons E, Wicker T, Salzberg SL, Devos KM, Dvořák J. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature*, 2017, 551(7681): 498–502. [DOI]
- [8] Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K, Jordan KW, Golan G, Deek J, Ben-Zvi B, Ben-Zvi G, Himmelbach A, MacLachlan RP, Sharpe AG, Fritz A, Ben-David R, Budak H, Fahima T, Korol A, Faris JD, Hernandez A, Mikel MA, Levy AA, Steffenson B, Maccaferri M, Tuberosa R, Cattivelli L, Faccioli P, Ceriotti A, Kashkush K, Pourkheirandish M, Komatsuda T, Eilam T, Sela H, Sharon A, Ohad N, Chamovitz DA, Mayer KFX, Stein N, Ronen G, Peleg Z, Pozniak CJ, Akhunov ED, Distelfeld A. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, 2017, 357(6346): 93–97. [DOI]
- [9] International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Stein N, Choulet F, Distelfeld A, Eversole K, Poland J, Rogers J, Ronen G, Sharpe AG, Pozniak C, Ronen G, Stein N, Barad O, Baruch K, Choulet F, Keeble-Gagnère G, Mascher M, Sharpe AG, Ben-Zvi G, Josselin AA, Stein N, Mascher M, Himmelbach A, Choulet F, Keeble-Gagnère G, Mascher M, Rogers J, Balfourier F, Gutierrez-Gonzalez J, Hayden M, Josselin AA, Koh C, Muehlbauer G, Pasam RK, Paux E, Pozniak CJ, Rigault P, Sharpe AG, Tibbits J, Tiwari V, Choulet F, Keeble-Gagnère G, Mascher M, Josselin AA, Rogers J, Spannagl M, Choulet F, Lang D, Gundlach H, Haberer G, Keeble-Gagnère G, Mayer KFX, Ormanbekova D, Paux E, Prade V, Šimková H, Wicker T, Choulet F, Spannagl M, Swarbreck D, Rimbart H, Felder M, Guilhot N, Gundlach H, Haberer G, Kaithakottil G, Keilwagen J, Lang D, Leroy P, Lux T, Mayer KFX, Twardziok S, Venturini L, Appels R, Rimbart H, Choulet F, Juhász A, Keeble-Gagnère G, Choulet F, Spannagl M, Lang D, Abrouk M, Haberer G, Keeble-Gagnère G, Mayer KFX, Wicker T, Choulet F, Wicker T, Gundlach H, Lang D, Spannagl M, Lang D, Spannagl M, Appels R, Fischer I, Uauy C, Borrill P, Ramirez-Gonzalez RH, Appels R, Arnaud D, Chalabi S, Chalhoub B, Choulet F, Cory A, Datla R, Davey MW, Hayden M, Jacobs J, Lang D, Robinson SJ, Spannagl M, Steuernagel B, Tibbits J, Tiwari V, van Ex F, Wulff BBH, Pozniak CJ, Robinson SJ, Sharpe AG, Cory A, Benhamed M, Paux E, Bendahmane A, Concia L, Latrasse D, Rogers J, Jacobs J, Alaux M, Appels R, Bartoš J, Bellec A, Berges H, Doležel J, Feuillet C, Frenkel Z, Gill B, Korol A, Letellier T, Olsen OA, Šimková H, Singh K, Valárik M, van der Vossen E, Vautrin S, Weining S, Korol A, Frenkel Z, Fahima T, Glikson V, Raats D, Rogers J, Tiwari V, Gill B, Paux E, Poland J, Doležel J, Číhalíková J, Šimková H, Toegelová H, Vrána J, Sourdille P, Darrier B, Appels R, Spannagl M, Lang D, Fischer I, Ormanbekova D, Prade V, Barabaschi D, Cattivelli L, Hernandez P, Galvez S, Budak H, Steuernagel B, Jones JDG, Witek K, Wulff BBH, Yu G, Small I, Melonek J, Zhou R, Juhász A, Belova T, Appels R, Olsen OA, Kanyuka K, King R, Nilsen K, Walkowiak S, Pozniak CJ, Cuthbert R, Datla R, Knox R, Wiebe K, Xiang D, Rohde A, Golds T, Doležel J, Čížková J, Tibbits J, Budak H, Akpinar BA, Biyiklioglu S, Muehlbauer G, Poland J, Gao L, Gutierrez-Gonzalez J, N'Daiye A, Doležel J, Šimková H, Číhalíková J, Kubaláková M, Šafář J, Vrána J, Berges H, Bellec A, Vautrin S, Alaux M, Alfama F, Adam-Blondon AF, Flores R, Guerche C, Letellier T, Loaec M, Quesneville H, Pozniak CJ, Sharpe AG, Walkowiak S, Budak H, Condie J, Ens J, Koh C, MacLachlan R, Tan Y, Wicker T, Choulet F, Paux E, Alberti A, Aury JM, Balfourier F, Barbe V, Couloux A, Cruaud C, Labadie K, Mangenot S, Wincker P, Gill B, Kaur G, Luo M, Sehgal S, Singh K, Chhuneja P, Gupta OP, Jindal S, Kaur P, Malik P, Sharma P, Yadav B, Singh NK, Khurana J, Chaudhary C, Khurana P, Kumar V, Mahato A, Mathur S, Sevanthi A, Sharma N, Tomar RS, Rogers J, Jacobs J, Alaux M, Bellec A, Berges H, Doležel J, Feuillet C, Frenkel Z, Gill B, Korol A, van der Vossen E, Vautrin S, Gill B, Kaur G, Luo M, Sehgal S, Bartoš J, Holuřová K, Plíhal O, Clark MD, Heavens D, Kettleborough G, Wright J, Valárik M, Abrouk M, Balcárková B, Holuřová K, Hu Y, Luo M, Salina E, Ravin N, Skryabin K, Beletsky A, Kadnikov V, Mardanov A, Nesterov M, Rakitin A,

- Sergeeva E, Handa H, Kanamori H, Katagiri S, Kobayashi F, Nasuda S, Tanaka T, Wu J, Appels R, Hayden M, Keeble-Gagnère G, Rigault P, Tibbits J, Olsen OA, Belova T, Cattonaro F, Jiumeng M, Kugler K, Mayer KFX, Pfeifer M, Sandve S, Xun X, Zhan B, Šimková H, Abrouk M, Batley J, Bayer PE, Edwards D, Hayashi S, Toegelová H, Tulpová Z, Visendi P, Weining S, Cui L, Du X, Feng K, Nie X, Tong W, Wang L, Borrill P, Gundlach H, Galvez S, Kaithakottil G, Lang D, Lux T, Mascher M, Ormanbekova D, Prade V, Ramirez-Gonzalez RH, Spannagl M, Stein N, Uauy C, Venturini L, Stein N, Appels R, Eversole K, Rogers J, Borrill P, Cattivelli L, Choulet F, Hernandez P, Kanyuka K, Lang D, Mascher M, Nilsen K, Paux E, Pozniak CJ, Ramirez-Gonzalez RH, Šimková H, Small I, Spannagl M, Swarbreck D, Uauy C. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 2018, 361(6403): eaar7191. [DOI]
- [10] Fuller DQ, Lucas L. Wheats: Origins and Development. In: Encyclopedia of Global Archaeology. New York, Springer, 2014. [DOI]
- [11] Peng JH, Sun DF, Nevo E. Domestication evolution, genetics and genomics in wheat. *Mol Breeding*, 2011, 28(3): 281–301. [DOI]
- [12] Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo MC, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*, 2013, 496(7443): 87–90. [DOI]
- [13] Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang HB, Wang XY, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang LF, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, 2009, 457(7229): 551–556. [DOI]
- [14] International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, 2010, 463(7282): 763–768. [DOI]
- [15] Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell*, 2008, 20(1): 11–24. [DOI]
- [16] Wang X, Shi X, Hao B, Ge S, Luo J. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol*, 2005, 165(3): 937–946. [DOI]
- [17] Zou S, Wang H, Li Y, Kong Z, Tang D. The NB-LRR gene Pm60 confers powdery mildew resistance in wheat. *New Phytol*, 2018, 218(1): 298–309. [DOI]
- [18] Dong LL, Wang FM, Liu T, Dong ZY, Li A, Jing RL, Mao L, Li YW, Liu X, Zhang KP, Wang DW. Natural variation of TaGASR7-A1 affects grain length in common wheat under multiple cultivation conditions. *Mol Breeding*, 2014, 34(3): 937–947. [DOI]
- [19] Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG, Himmelbach A, Mascher M, Bagnaresi P, Faccioli P, Cozzi P, Lauria M, Lazzari B, Stella A, Manconi A, Gnocchi M, Moscatelli M, Avni R, Deek J, Biyikliglu S, Frascaroli E, Corneti S, Salvi S, Sonnante G, Desiderio F, Marè C, Crosatti C, Mica E, Özkan H, Kilian B, De Vita P, Marone D, Joukhadar R, Mazzucotelli E, Nigro D, Gadaleta A, Chao S, Faris JD, Melo ATO, Pumphrey M, Pecchioni N, Milanesi L, Wiebe K, Ens J, MacLachlan RP, Clarke JM, Sharpe AG, Koh CS, Liang KYH, Taylor GJ, Knox R, Budak H, Mastrangelo AM, Xu SS, Stein N, Hale I, Distelfeld A, Hayden MJ, Tuberosa R, Walkowiak S, Mayer KFX, Ceriotti A, Pozniak CJ, Cattivelli L. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet*, 2019, 51(5): 885–895. [DOI]

(责任编辑: 夏先春)