

GWAS on agronomic traits (e.g. flowering time and leaf trichome) based on resequencing of 991 genomes of *Brassica napus*

Presentation on the Annual Congress of Chinese Crop Science Society, October 28-29, 2019, Hangzhou



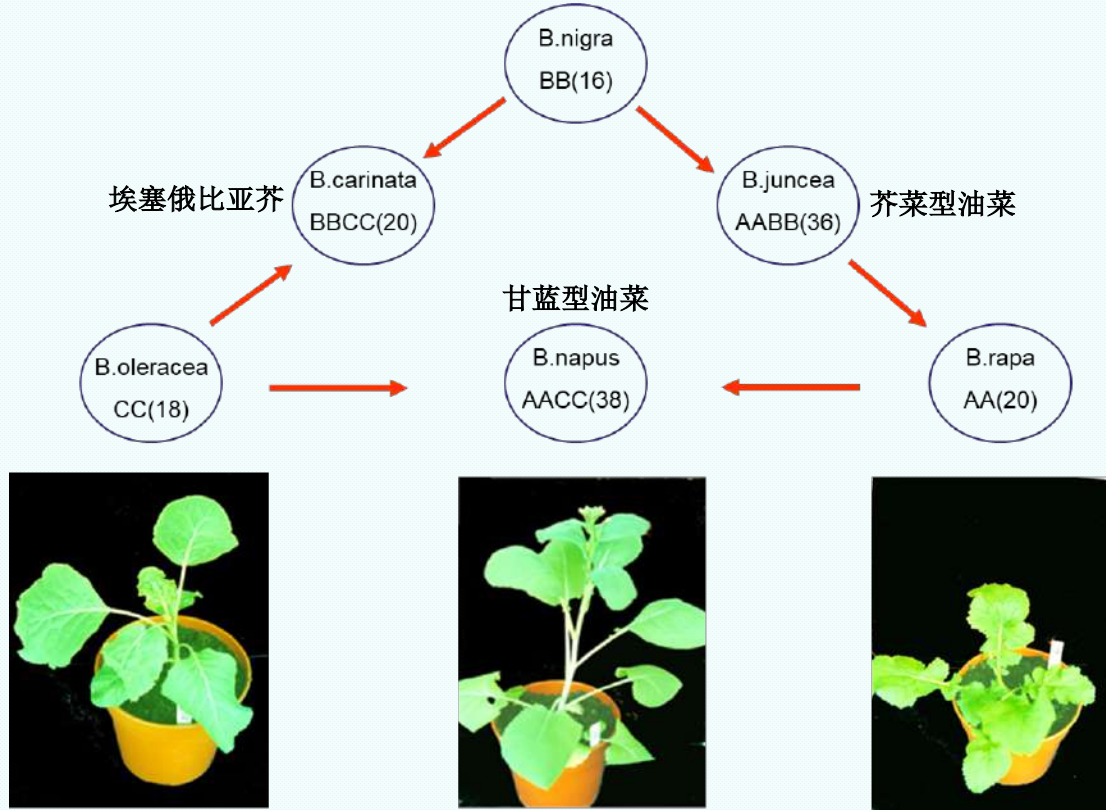
Prof. Dr. Lixi Jiang

Zhejiang University, Hangzhou China

jianglx@zju.edu.cn

The history of the species *Brassica napus* and its morphotypes

Nagaharu, 1935,
Jpn J Bot 7:389-452



Columella (70 AD)

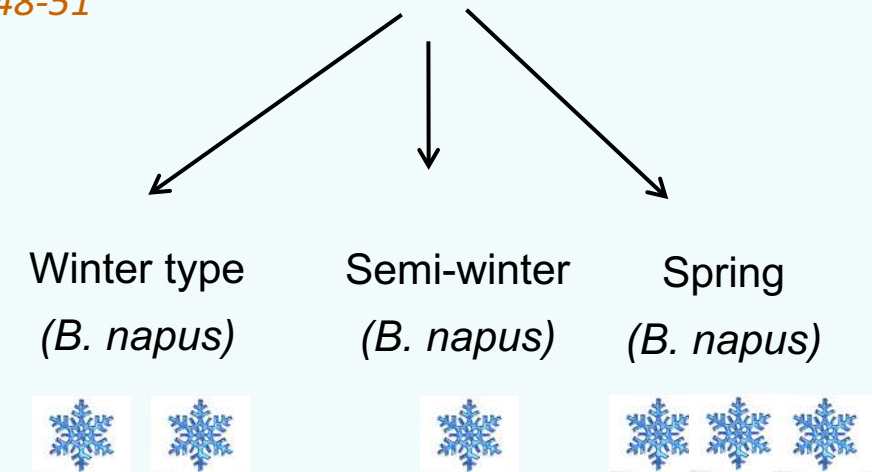


Fussell, 1955, *Nature* 9:48-51

~7500 Year ago: *Brassica napus*

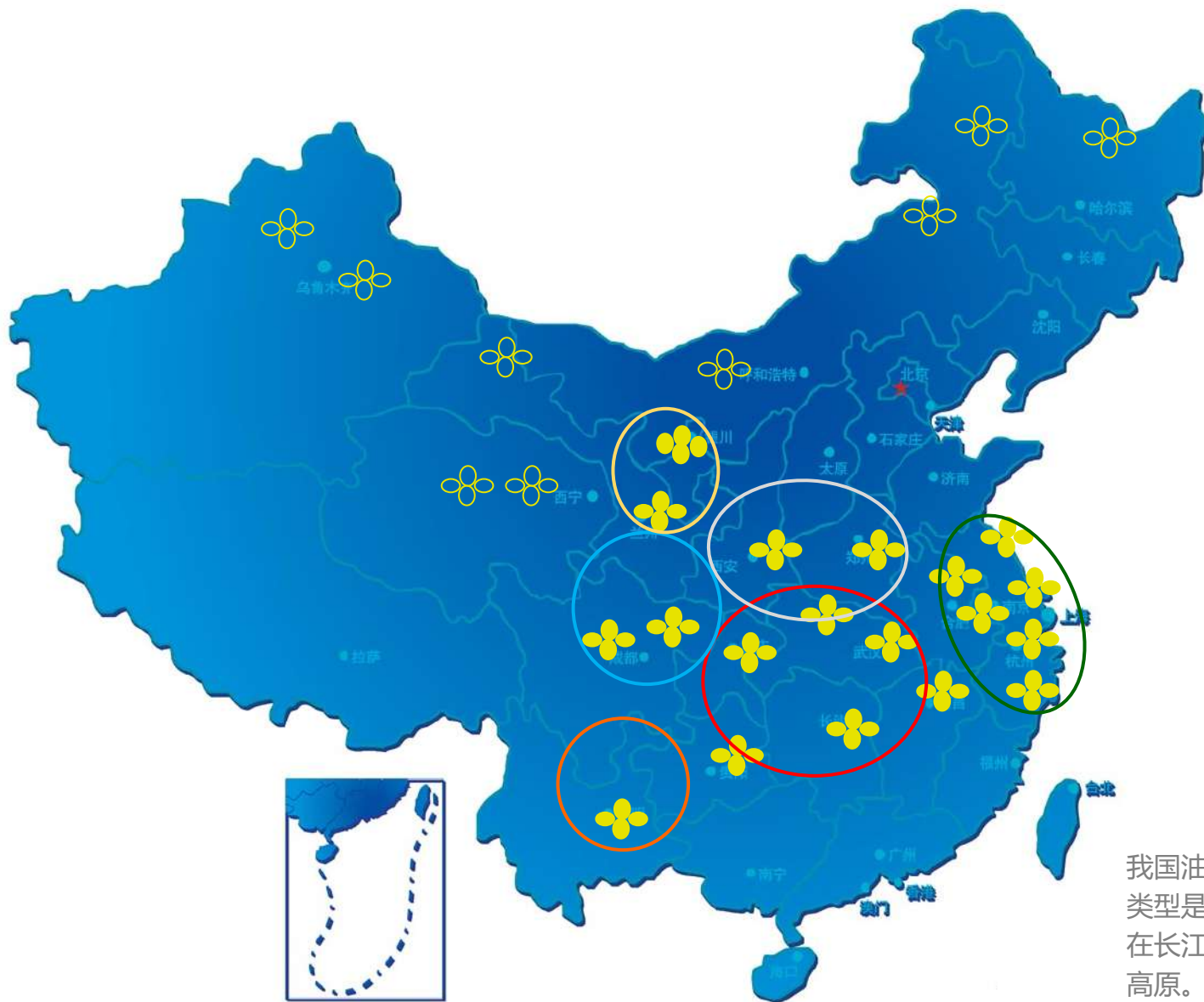
Middle age: *Rapa*, *Cole*

B. napus
8th B.C. - 6th A.D.



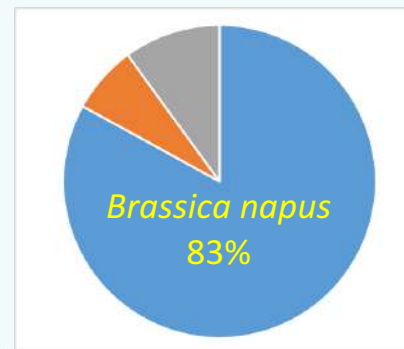
(1) 油菜是芸薹属多个物种的总称，但我们一般说的油菜，是世界上目前种植面积最大的甘蓝型油菜，它是由白菜和甘蓝天然杂交而成，是一个典型的异源四倍体物种，迄今大约只有7500年的历史。最早记录油菜的，英语古籍里面的cole，是由一位叫克留米拉的学者记录的。油菜最早分布在地中海沿岸与欧洲国家，在北美和澳洲、以及中国的历史都很短。(3) 甘蓝型油菜，在向世界各地扩散、引种的过程中，由于海拔、气候、光照等生态条件的不同，形成了三种生态类型。

The distribution of rapeseed production areas in China



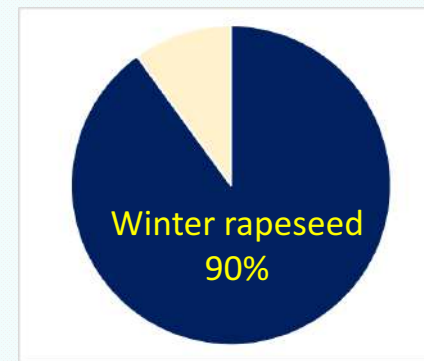
Brassica juncea
7%

Brassica rapa
10%



83% of rapeseed in China is *Brassica napu*, 10% is *Brassica rapa*, 7% is *Brassica juncea*.

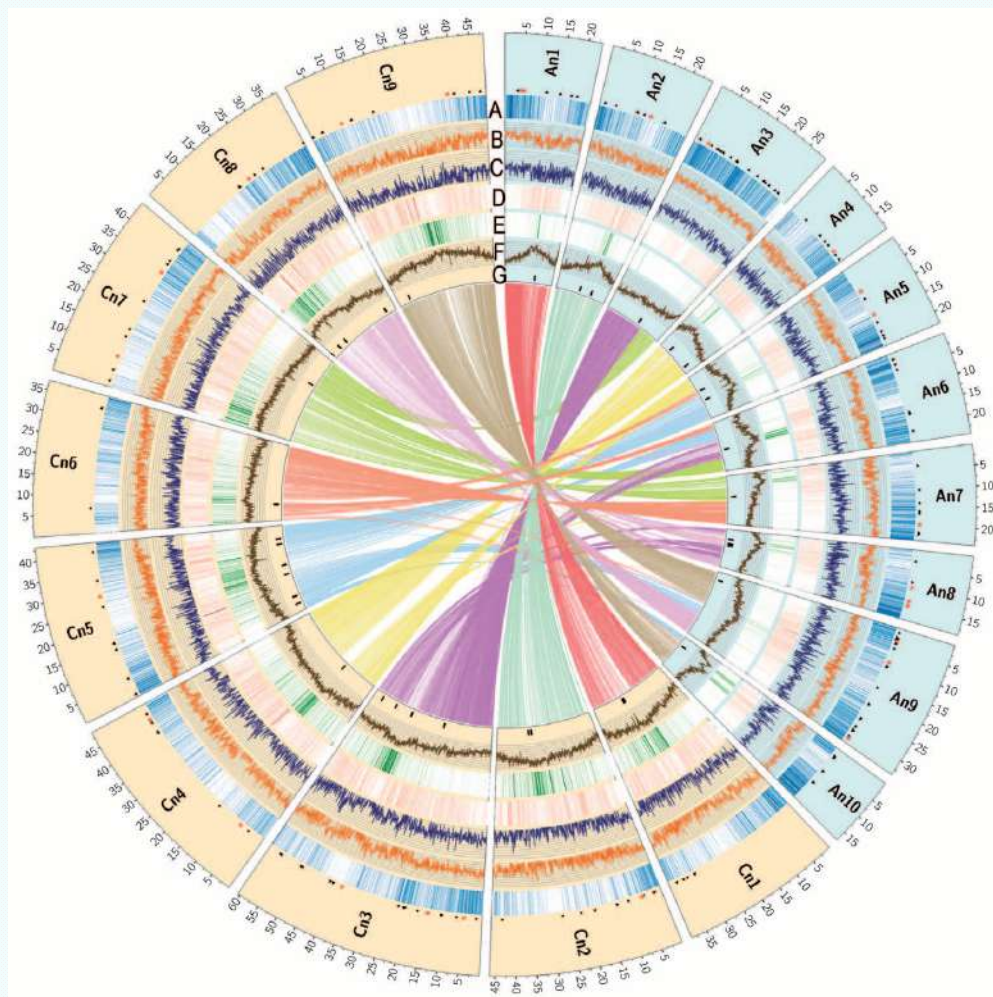
Spring type 10%



我国油菜生产上最重要的生态类型是半冬性油菜，主要分布在长江流域、黄淮、以及云贵高原。

The reference genomes of rapeseed

Chalhoub et al., 2004, *Science* 345:950-953



Darmor-bzh: Winter type

Chalhoub et al., 2014, Science 345:950-953

Tapidor: Winter type

Bayer et al., 2017, Plant Biotech J 345:950-953

Zhongshuang 11: Semi-winter

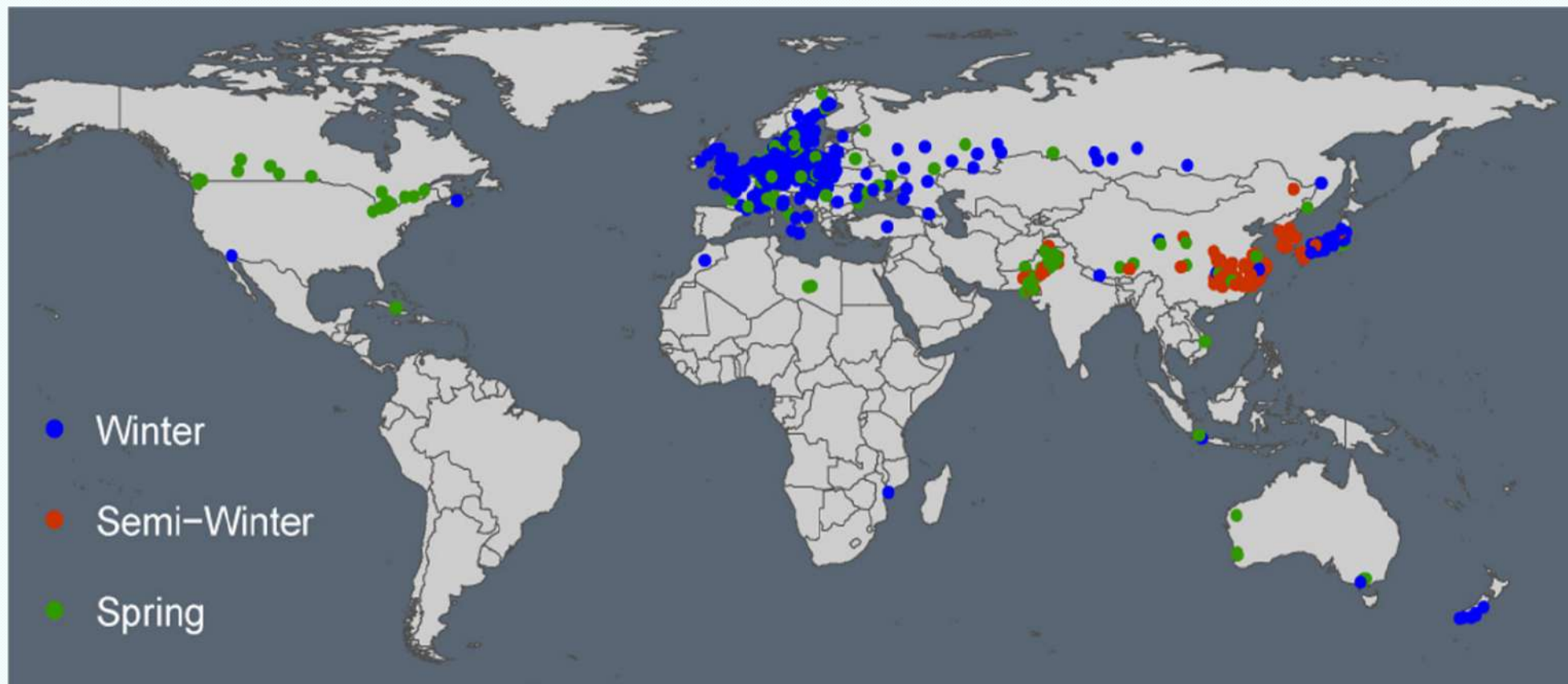
Sun et al., 2017, Plant J 92:452-468

Ningyou 7: Semi-winter

Zou et al., 2019, Plant Biotech J doi: 10.1111/pbi.13115

油菜第一个参考基因组发表于2014年，现在看来这个基因组有很大的完善空间，定位在19条染色体上面的序列只占有基因组总长度75%，850MB。2017年西澳大学组装的Tapidor参考基因组，也没有比Darmor好很多。中国组装了中双11号，和宁油7号，都是半冬性的品种，宁油7号是我们浙大樊龙江课题组组装的，这是一个传统的高芥酸/高硫代的品种。宁油7号有930MB左右的序列被mapping，拼装的质量比前面3个都好很多。

The origin of 991 rapeseed germplasm



Originating from:
39 countries/regions

Germany: 363

China: 131

Canada: 21

Pakistan: 43

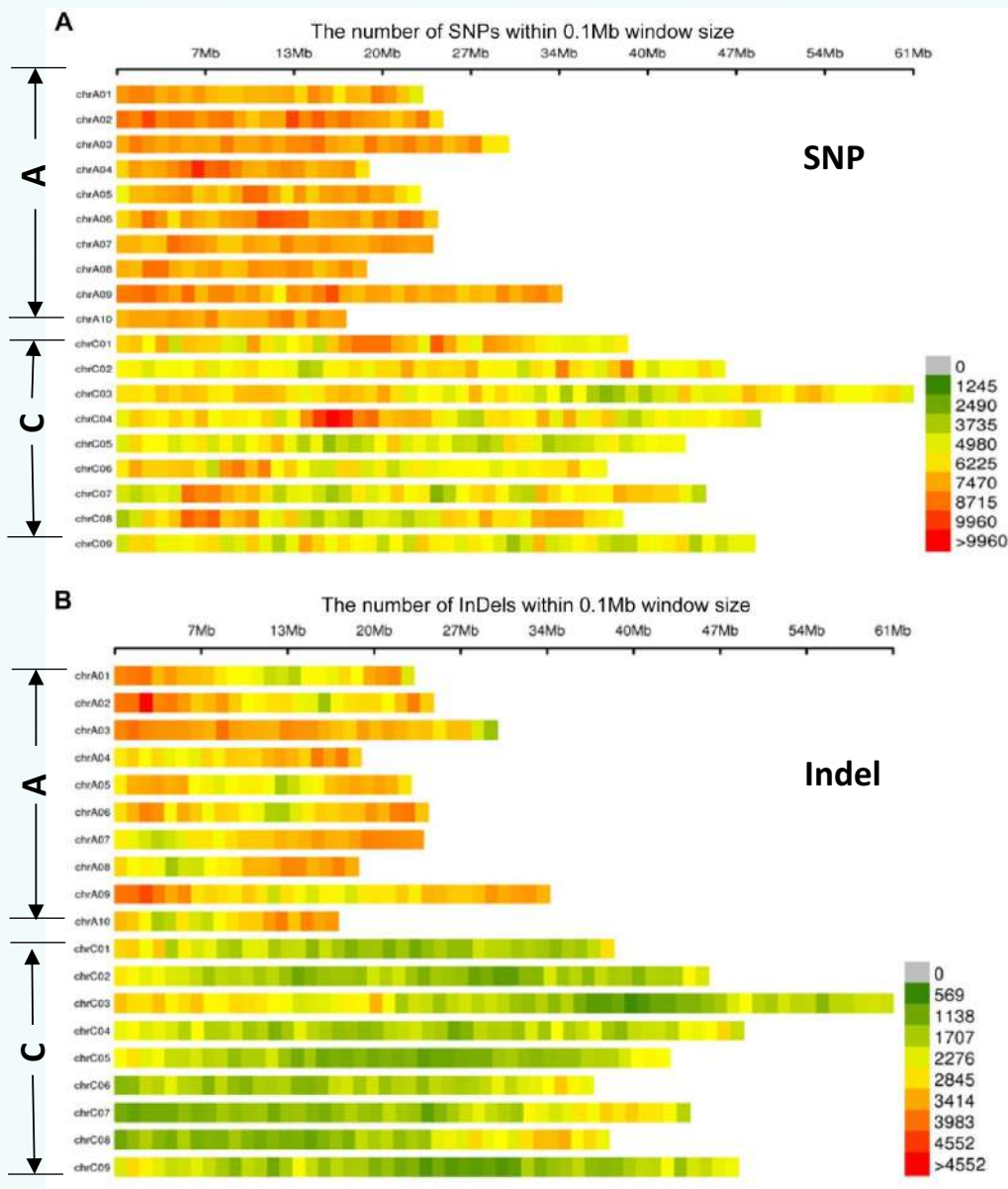
Winter: 658

Semi-winter: 145

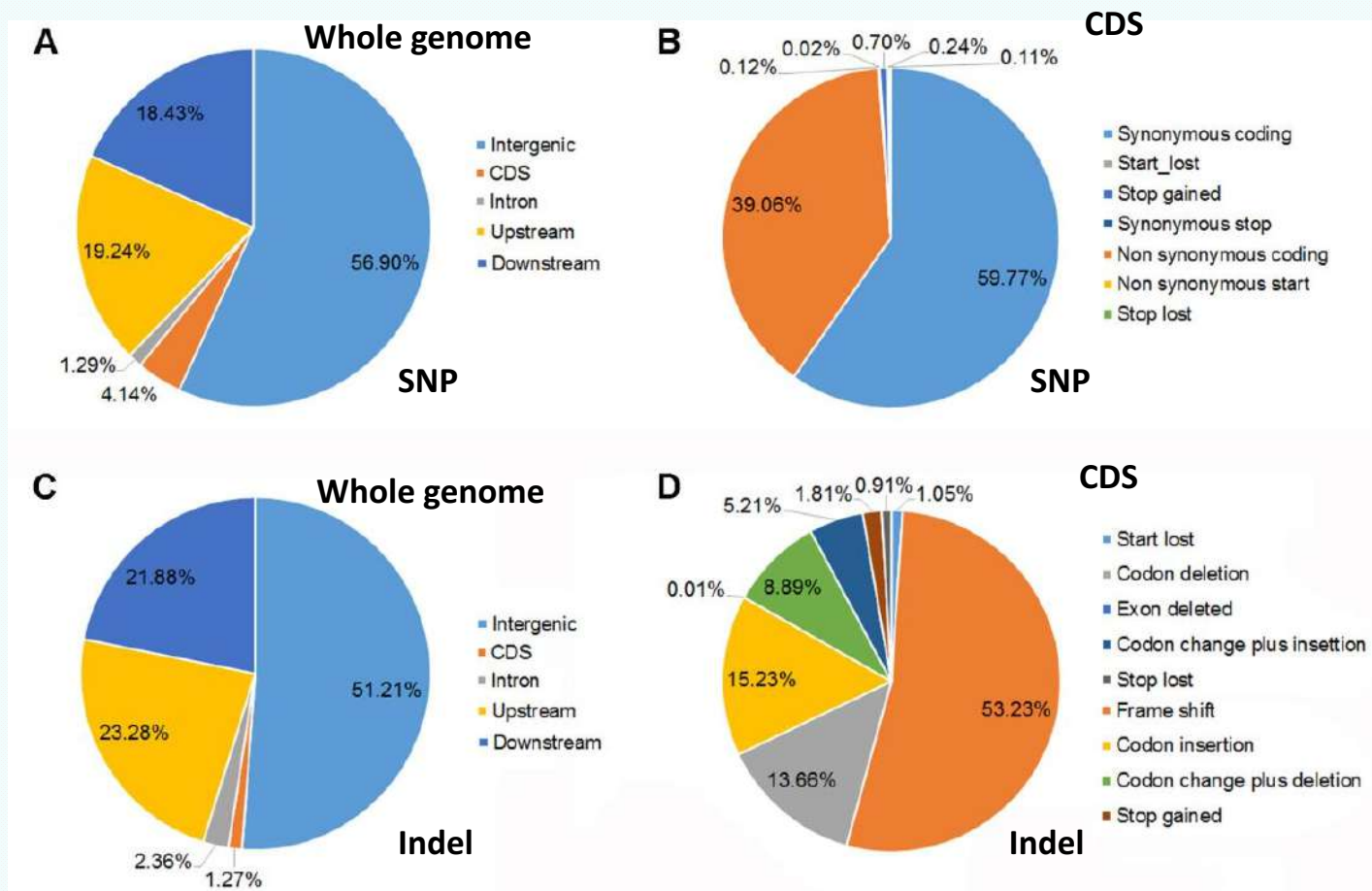
Spring: 188

2018年，我们以Darmor作为参考序列，对我们实验室收集到的991份品种资源进行了基因组重测序，这批材料对来自全球39个国家的，主要是欧洲、中国、加拿大、澳大利亚、印度和巴基斯坦等油菜主产国家。在这张世界地图上，我们对这991份资源的地理分布做了一个直观的呈现。这其中，冬油菜有658份，半冬油菜有145份，春油菜有188份。

Summary of genomic polymorphism and variants

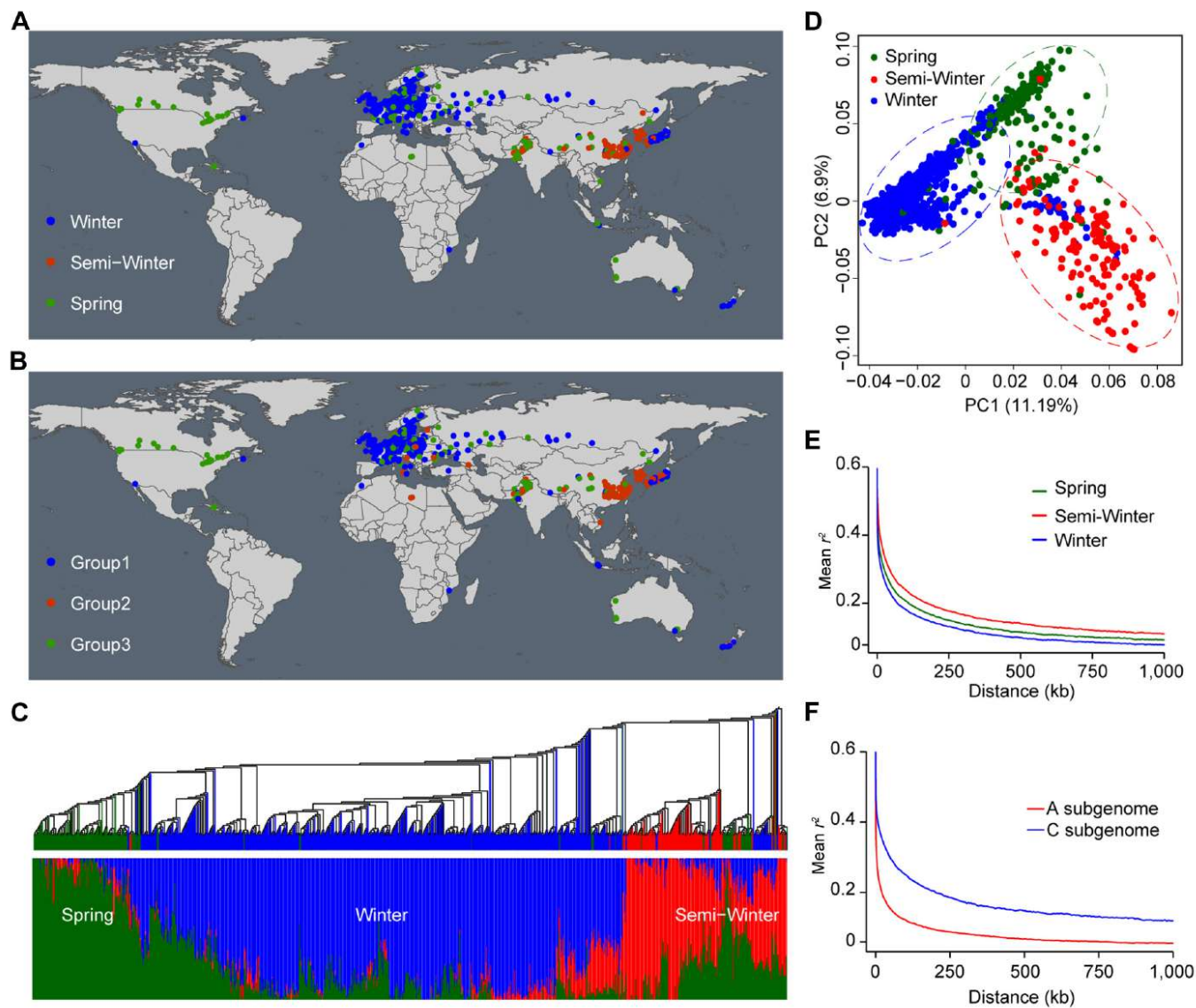


Total: SNPs 5.55 M; InDels 1.86 M



首先我们来看总体的SNP和Indel分布情况：(1) 总共有5.5M SNP，1.86M的Indel，(2) 多数 SNPs & indels 分布在基因间隔区；(3) 40% 落在编码区的SNPs造成氨基酸变化；(4) 二分之一以上的indel造成于阅读框架的位移；(4) A基因组的多态性高于C基因组：原因。

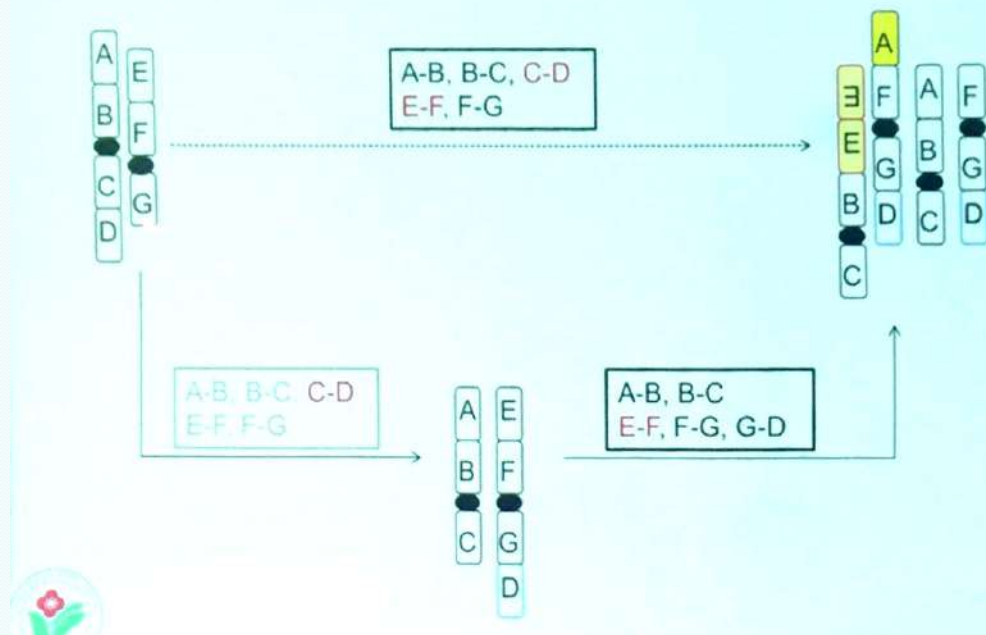
Population structure and genomic variations of the 991 accessions



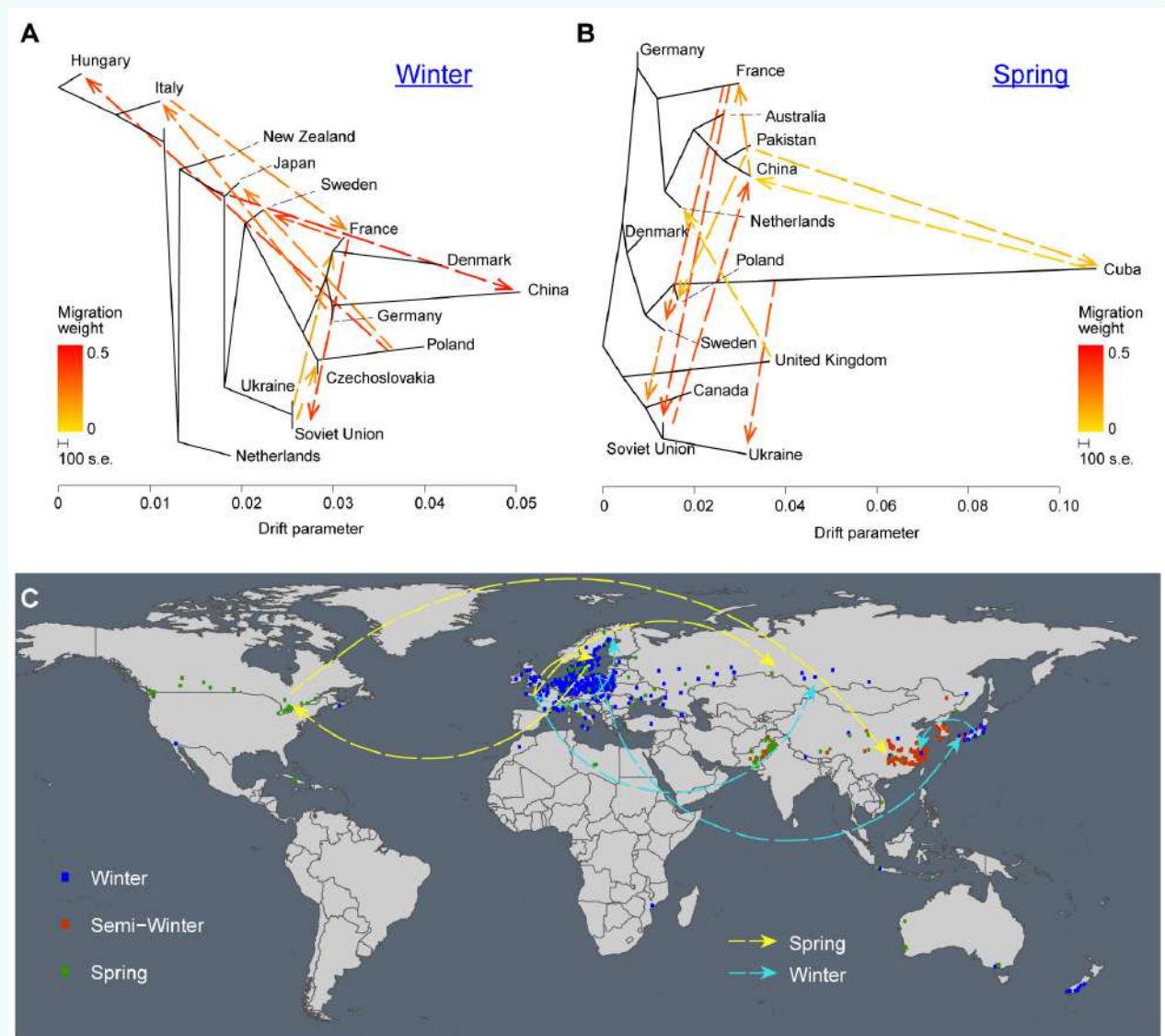
- 接下来，我们进行了群体遗传背景分析，如果K取3，991份材料的分布大致上正好和冬、半冬、春三个生态型符合；但是，也有大约有 (8.9%) 的材料例外。
- 我们从E和F两个图上，来看连锁不平衡衰减 (LD decay)。我们看到：冬油菜快于春油菜，春油菜快于半冬油菜；A亚基因组快于C亚基因组。也就是说，冬油菜基因组的重组率要高于半冬、春油菜的基因组；A基因组的重组率要高于C基因组。

Allelic drifting paths of populations among the major origins

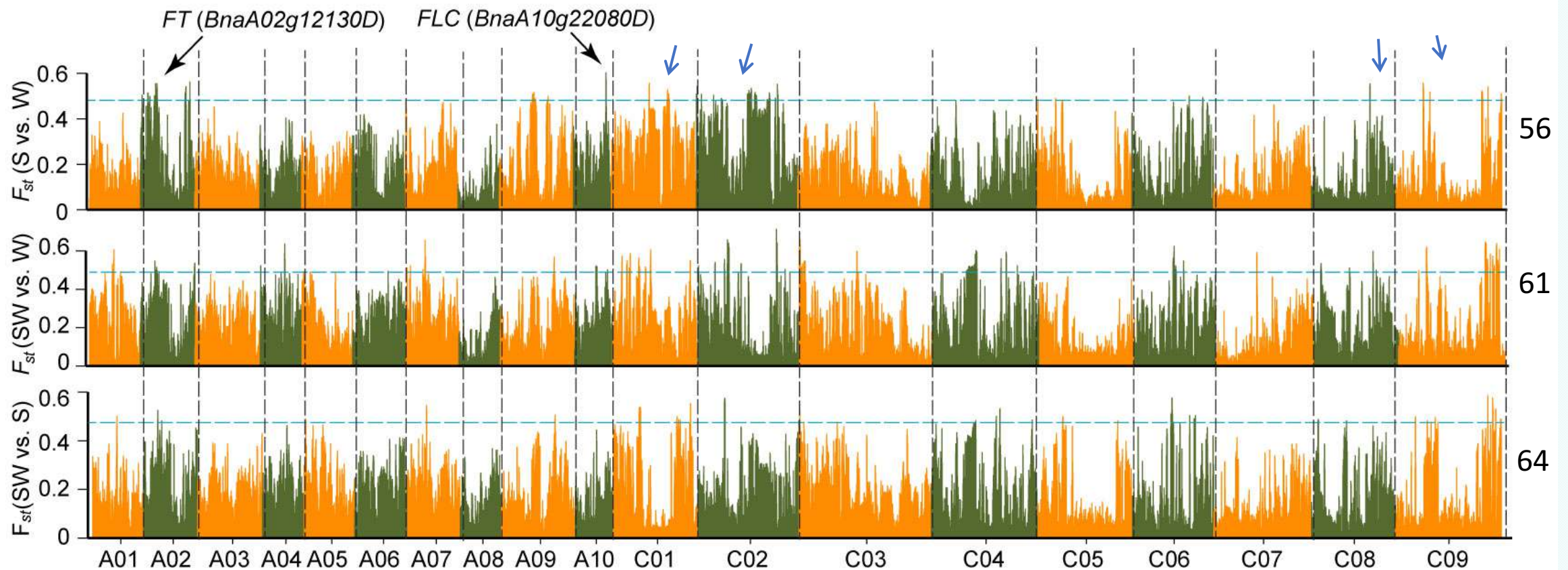
Older ancestor lost more block links



我们做的另一项工作是基于序列信息，进行了群体基因漂移分析（drifting path），这个分析大致可以告诉我们，油菜在世界上传播、扩散的轨迹。**主要计算依据是：**祖先（老品种）比现代品种，基因组区块之间的相邻关系数更少；我们可以以其中三条线路为例，来分析其真实性：罗马兵的征战路线、中国-日本、加拿大-中国。

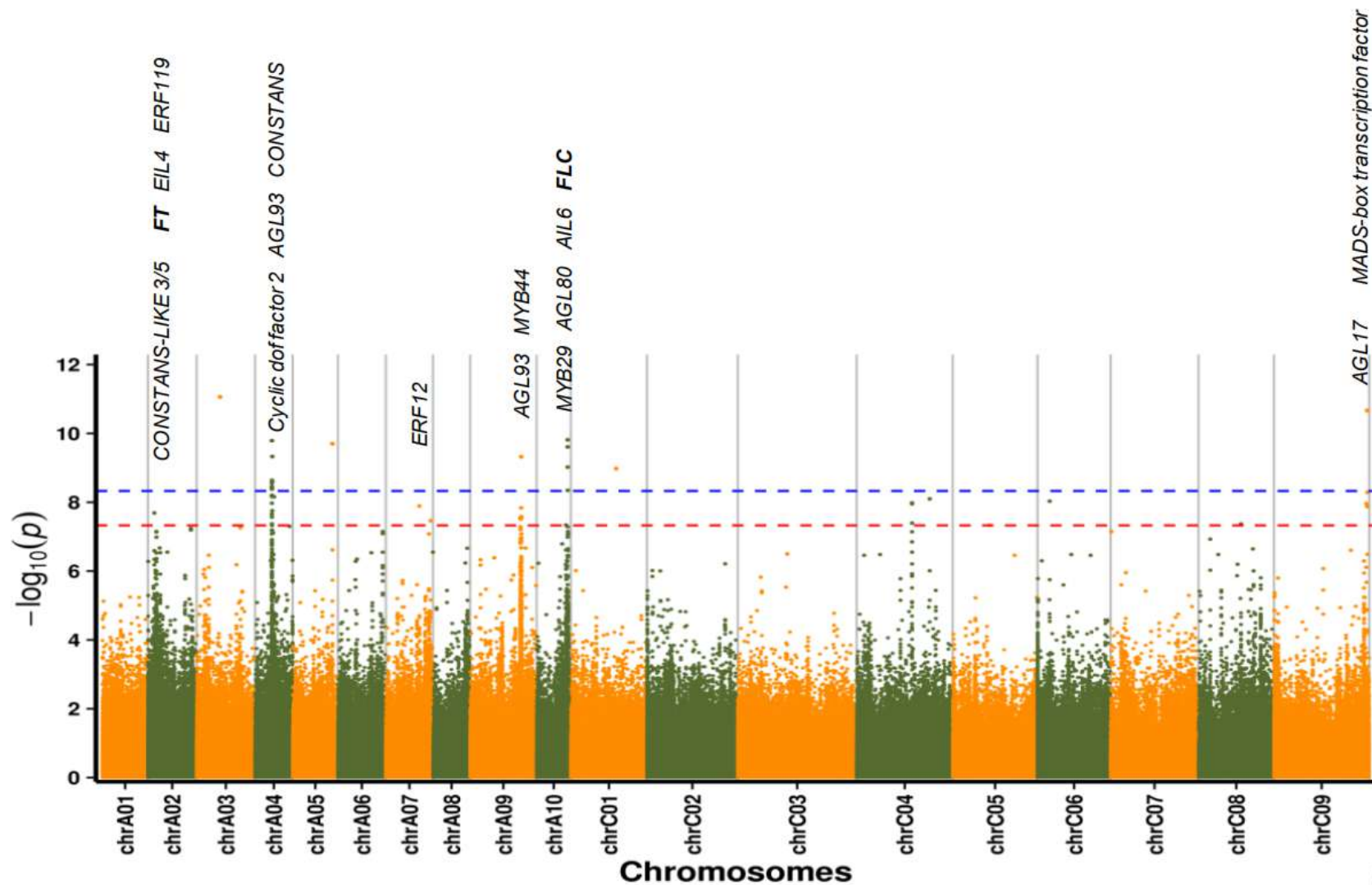


Selective signals produced by natural and artificial selection in *B. napus*



油菜不像棉花、水稻那样有明显的野生种，但是在育种和引种的过程中，自然选择加上人工选择，在油菜的基因组中留下各种选择的痕迹：genetic footprints (or selective sweeps / selective signals). 这张图的上面显示的是W and S, 中间是SW and W, 下面是SW and S, 分别存在的 56, 61 and 64 selective sweeps. 对这些选择性清楚分析峰下面的基因所注释的信息进行研究表明，很多的开花基因、乙烯途径基因、响应逆境抗性的基因，在不同的生态类型之间，选择痕迹非常明显。

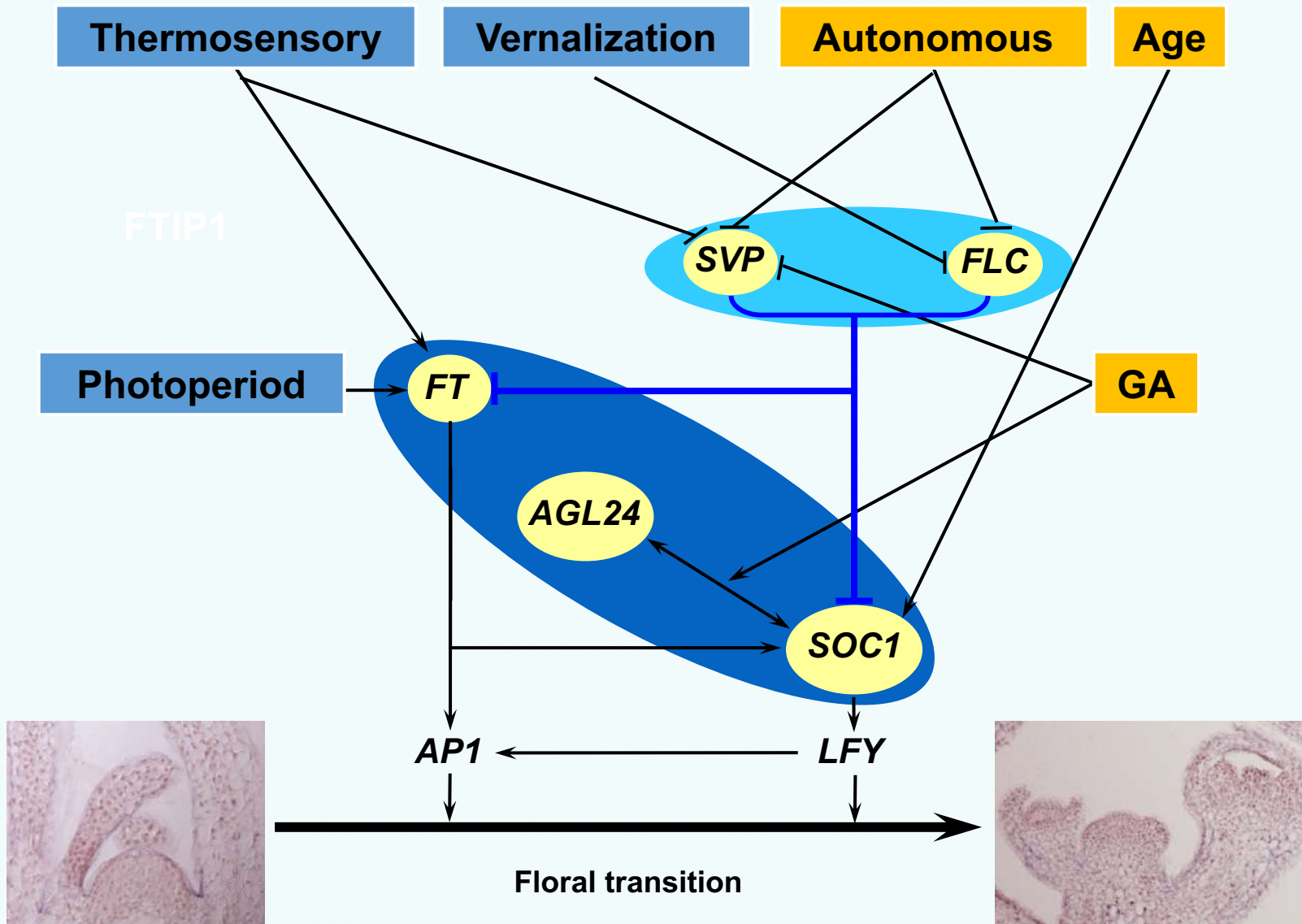
Selective signals produced by natural and artificial selection in *B. napus*



接着，我们用重测序的991份材料，来进行开花性状的全基因组关联分析（GWAS）；

我们发现在A2，A4，A9，A10，C10这些染色体上，有一些明显的SNP的峰。一些高于显著性门槛的SNP，它们要么是直接发生在开花基因、乙烯途径基因的上面，要么和开花基因或乙烯途径基因紧密地连锁在一起，这些基因包括 CO, FT, ERFs, AGL-like genes, MYB29, and FLC.，一共有二十多个。

The genetic model of floral transition

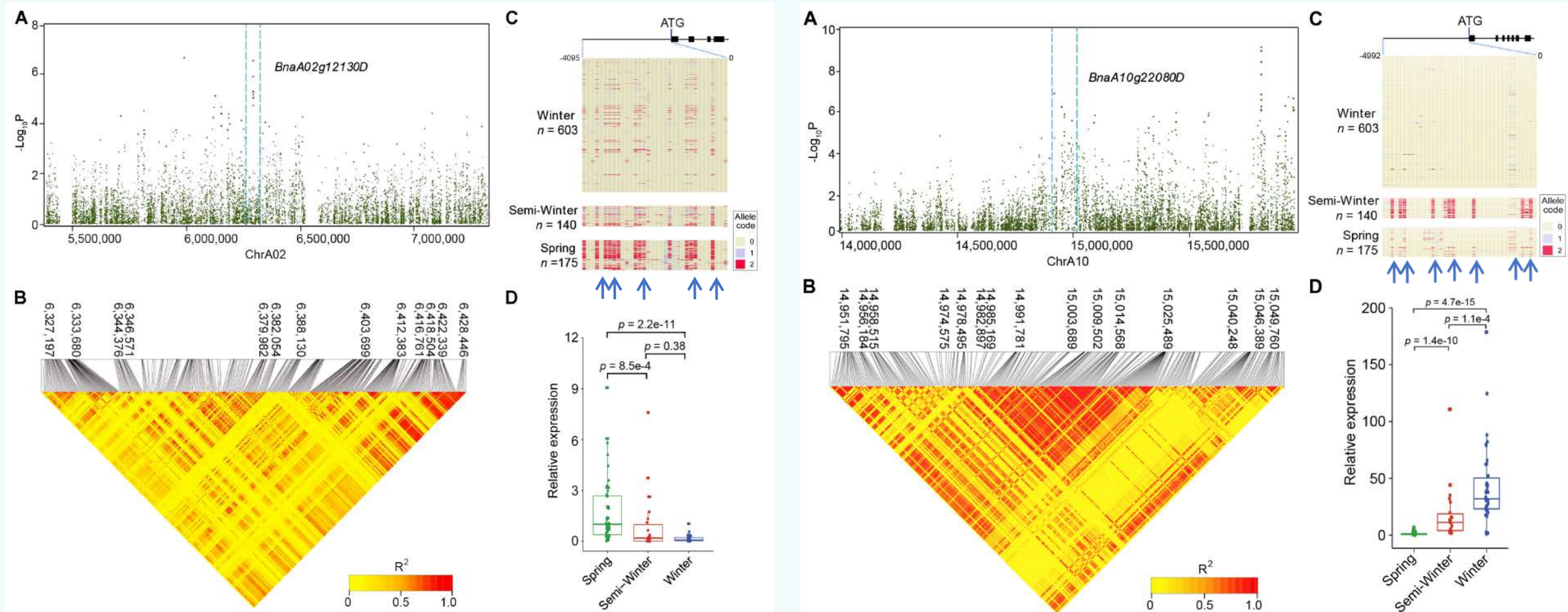


我们知道植物必须选择在合适的条件下开花，才能让它们自己在特定的生态条件下生存下来，开花期的早或晚，对于一个作物的成熟期、产量都是至关重要的。

我们在拟南芥中，已经知道分别有温度感应、春化感应、光周期、自发感应与年龄（age）感应等途径控制开花，但是殊途同归，这些不同的途径最终都要激活成花素基因 **FT** 来诱导开花，而 **FLC** 基因在 **FT** 上有，它是抑制 **FT** 基因的表达的。**FT** 和 **FLC** 是最重要的2个开花基因。

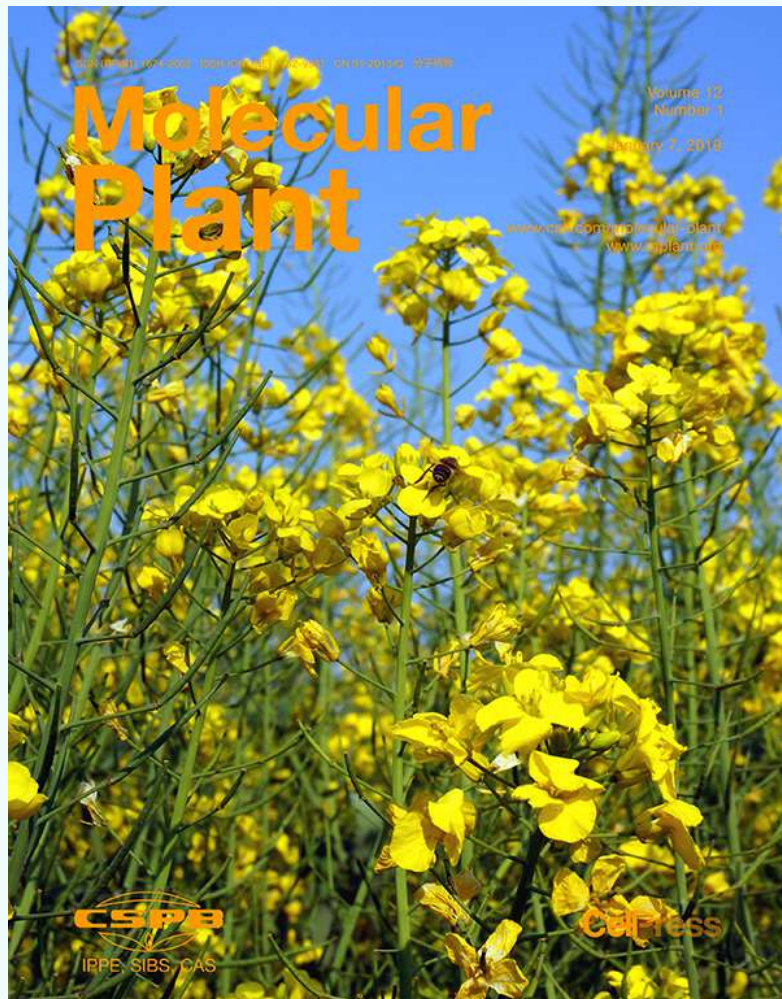
Li et al., 2008, Dev Cell, 15: 110–120; Liu et al., 2009, Development, 136: 3379–3391

GWAS on flowering time divergence



我们在A2染色体上，发现一系列的SNP和FT和FLC紧密连锁。有趣的是，这两个基因，它们的编码区域，在三种生态类型中都是高度保守的，也就是说，SNP都发生在这2个基因的调控区域，特别是5'端的调控区域：（1）大约95%的冬油菜，它们在这些箭头所指的位置，和参考基因组具有相同的SNP；（2）大约85%的春油菜，它们在这些箭头所指的位置，和参考基因组具有不同的SNP；（3）而大约有22%的半冬性油菜，它们在箭头所注释的位置，和春油菜有着相同的SNP变化；（4）再看这2个基因表达情况，FT和FLC的表达趋势在这三种生态类型中完全相反，FT在春油菜表达最高，冬油菜最低；而FLC的情况相反。

Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions



CellPress
PARTNER JOURNAL

Molecular Plant
Research Article

Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence

Dezhi Wu^{1,8}, Zhe Liang^{2,8}, Tao Yan¹, Ying Xu¹, Lijie Xuan¹, Juan Tang³, Gang Zhou³, Ulrike Lohwasser⁴, Shuijin Hua⁵, Haoyi Wang¹, Xiaoyang Chen⁷, Qian Wang¹, Le Zhu¹, Antony Maodzeka¹, Nazim Hussain¹, Zhilan Li¹, Xuming Li³, Imran Haider Shamsi¹, Ghulam Jilani⁶, Linde Wu³, Hongkun Zheng³, Guoping Zhang¹, Boulos Chalhoub¹, Lisha Shen^{2,*}, Hao Yu^{2,*} and Lixi Jiang^{1,*}

¹Institute of Crop Science, Zhejiang University, Hangzhou 310058, China

²Temasek Life Sciences Laboratory and Department of Biological Science, National University of Singapore, Singapore 117543, Singapore

³Biomarker Technologies Corporation, Beijing 101300, China

⁴Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research, 06466 Stadt Seeland, Germany

⁵Institute of Crop and Nuclear Agricultural Sciences, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

⁶Office of Research, Innovation & Commercialization, PMAS-Arid Agricultural University Rawalpindi, 46300 Rawalpindi, Pakistan

⁷Institute of Crop Science, Jinhua Academy of Agricultural Sciences, Jinhua 321017, China

⁸These authors contributed equally to this article.

*Correspondence: Lisha Shen (lisha@tll.org.sg), Hao Yu (dbsyuhao@nus.edu.sg), Lixi Jiang (jianglx@zju.edu.cn)

<https://doi.org/10.1016/j.molp.2018.11.007>

Molecular Plant
Spotlight

CellPress
PARTNER JOURNAL

Illuminating Crop Adaptation Using Population Genomics

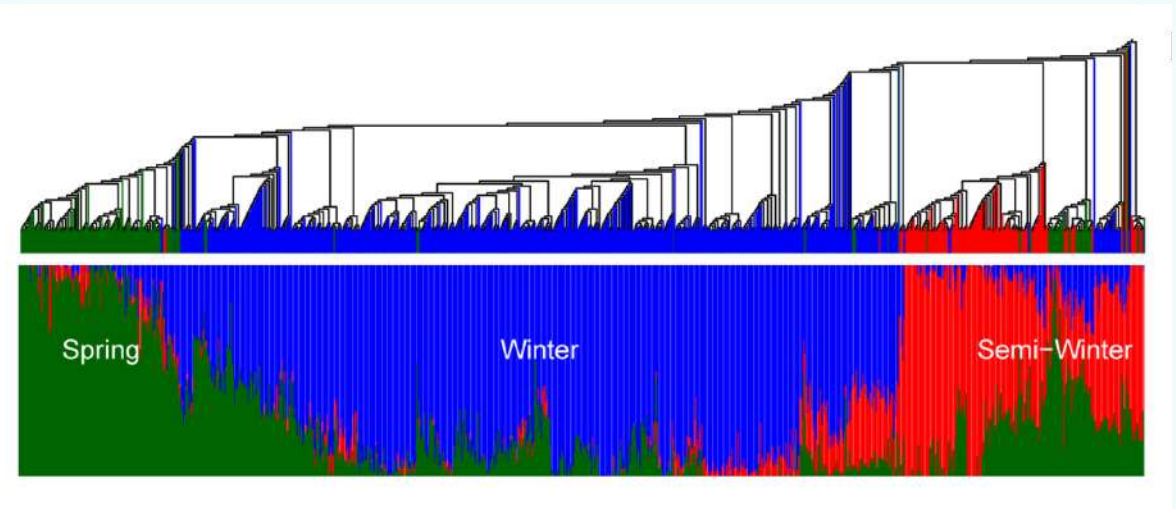
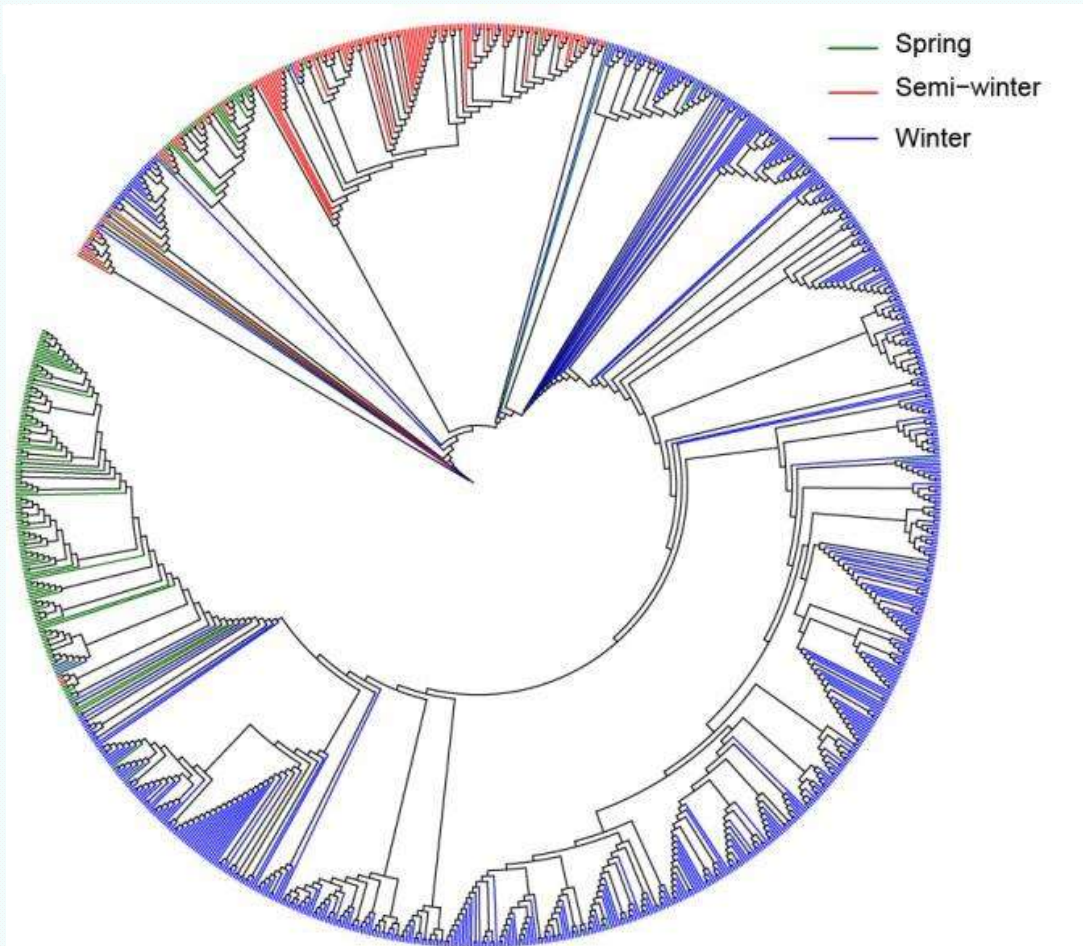
Rod J. Snowdon and Sarah Schiessl*

Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

*Correspondence: Rod J. Snowdon (rod.snowdon@agrar.uni-giessen.de)
<https://doi.org/10.1016/j.molp.2018.12.014>

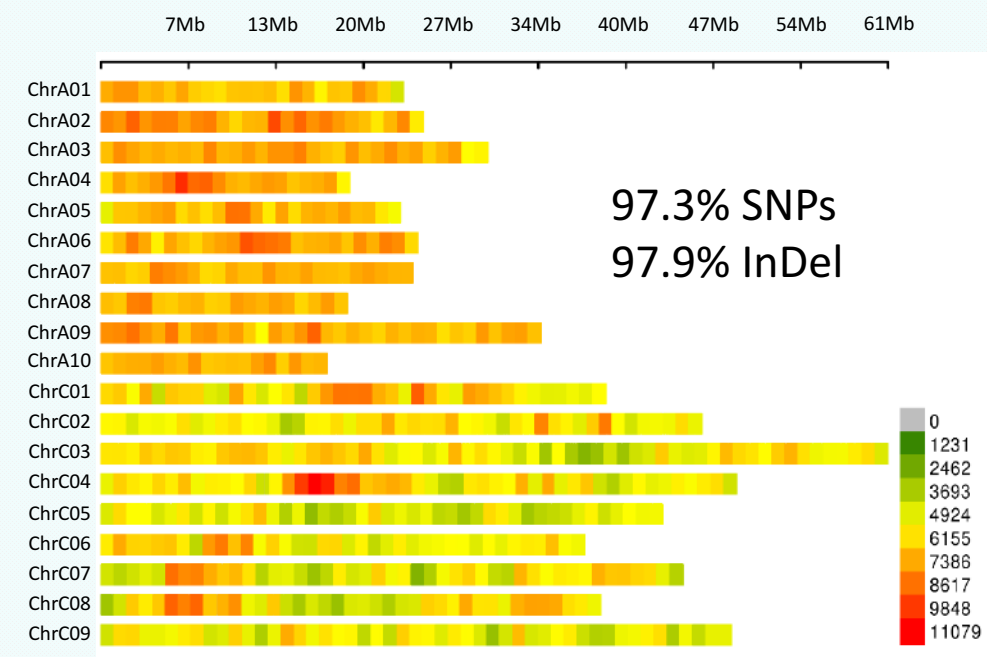
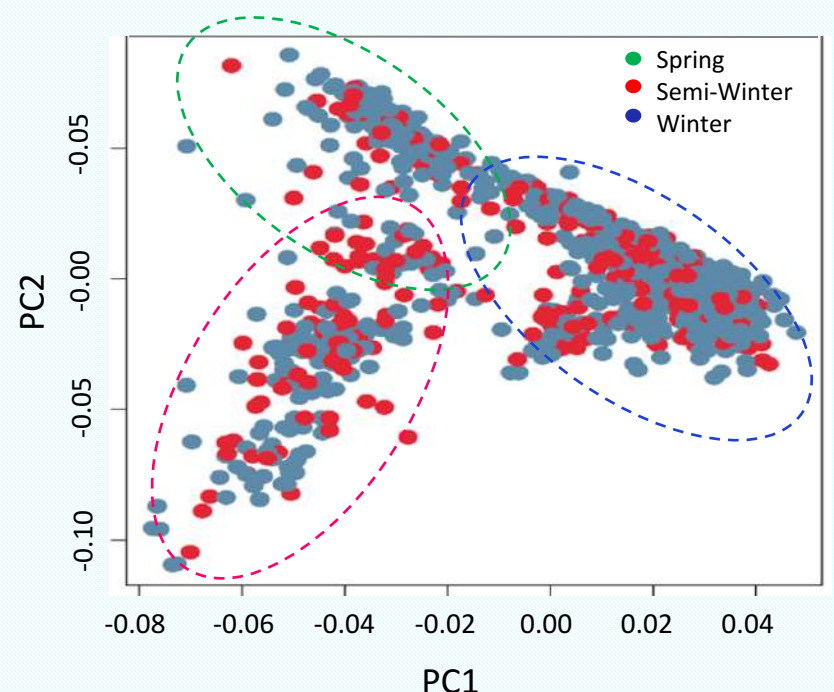
上面这项开花期的GWAS分析，作为封面论文，发表2019年第一期的在MP上，不到10个月已有很多引用，被Web of Science 预测高被引论文。

Establishment of core accessions for GWAS



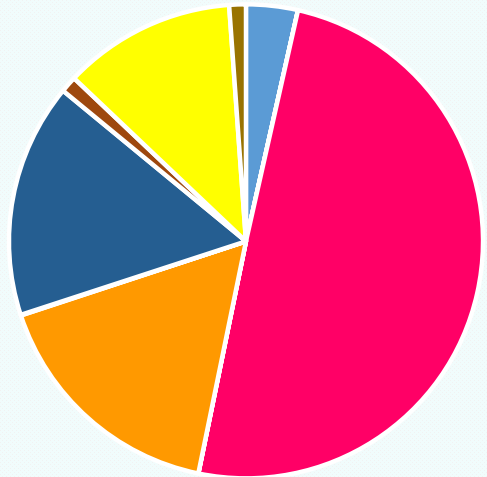
但是，用991份材料来进行GWAS分析，如果我们设计3-4个重复、外加多个环境的话，操作起来并不容易。所以，我们进一步选择了290份核心的种质资源，作为可操作的GWAS分析平台。

Genetic polymorphism of the core accessions for GWAS

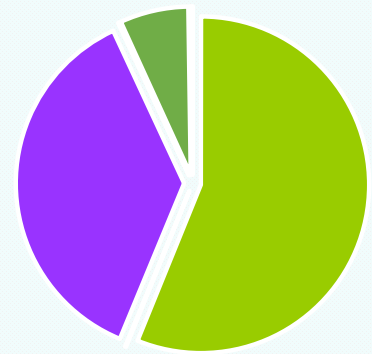


虽然290份和991份相比规模小很多，但是这个290份的核心油菜种质群体，保留了大群体97.3%的SNPs和97.9%的 indels 遗传多态性。用这个核心群体来计算SNP和Indel在油菜基因组不同区间的分布频率和利用大群题来计算，得到的结果类似。

- CDS (3.52%)
- Splice site donor (0.00%)
- Intergenic (49.69%)
- Splice site acceptor (0.00%)
- Upstream (16.71%)
- Splice site region (0.04%%)
- Downstream (15.99%)
- Intron (1.12%)
- Others (11.78%)
- Intragenic (1.13%)



The CDS



- Non synonymous start (0.02%)
- Synonymous coding (59.66%)
- Synonymous stop (0.11%)
- Non synonymous coding (39.15%)
- Start lost (0.12%)
- Stop gained (0.70%)
- Stop lost (0.24%)

因此，可以说虽然核心资源库群体小，但是遗传多态性却非常丰富。

油菜科学研究合作协议

甲方：浙江大学 浙江省作物种质资源重点实验室

乙方：江苏省农业科学院经济作物研究所

为了更好地利用甲方油菜种质资源和基因组序列信息，推动油菜基础生物学研究的发展与油菜育种进程，甲、乙双方签订以下合作协议：

1. 甲方向乙方提供基于千份种质基因组重测序基础上的（230-280 份）世界油菜核心种质种子、重测序原始序列数据，以及经提炼的 SNP 数据（以下简称“甲方资源”）；

2. 乙方可以利用甲方资源开展基础生物学研究（诸如全基因组关联分析、分子标记辅助育种研究、作物育种或生理研究等）或育种应用；乙方不得直接将甲方资源用于盈利目的的繁殖销售；不得将种子提供给第三方；

3. 乙方利用甲方资源所产生的科研成果，若以论文形式发表，乙方为第一产权单位（第一作者与主要通讯作者单位），甲方为合作单位（共同第一、或第二作者，以及非主要通讯作者单位）；乙方若将有关资源用于育种，在品种审定与其它文件中需明确亲本来源甲方；

4. （未来）甲方以牵头单位申报教育部或浙江省成果，乙方（视贡献）以合作单位参与浙大成果申报。

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

蒋经

甲方代表签字：

乙方：江苏省农业科学院经济作物研究所

(盖章)



乙方代表签字：

胡其长

2019年9月9日

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：甘肃省干旱生境作物学重点实验室

(盖章)

乙方代表签字：

王新明

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：

(盖章)

乙方代表签字：

王新明

2019年9月9日

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：

(盖章)

乙方代表签字：

丁江

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：江苏大学 生命科学研究院

(盖章)

乙方代表签字：

朱立明

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：山西省农业科学院棉花研究所

(盖章)

乙方代表签字：

杜春芳

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：

(盖章)

乙方代表签字：

徐瑞

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：

(盖章)

乙方代表签字：

张锦芳

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：西北农林科技大学 生命科学院

(盖章)

乙方代表签字：

谢长根

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：西南大学 重庆市油菜工程中心

(盖章)

乙方代表签字：

王坤

2019年9月16日

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：西南大学 资源环境学院

(盖章)

乙方代表签字：

李楠楠

2019年9月16日

5. 本合同双方签字后生效，一式两份，甲乙各持一份。

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：江苏省农业科学院经济作物研究所

(盖章)

乙方代表签字：

李纪辉

2019年9月16日

甲方：浙江大学 浙江省作物种质资源重点实验室

(盖章)

甲方代表签字：

乙方：甘肃省农业科学院作物研究所

(盖章)

乙方代表签字：

董云

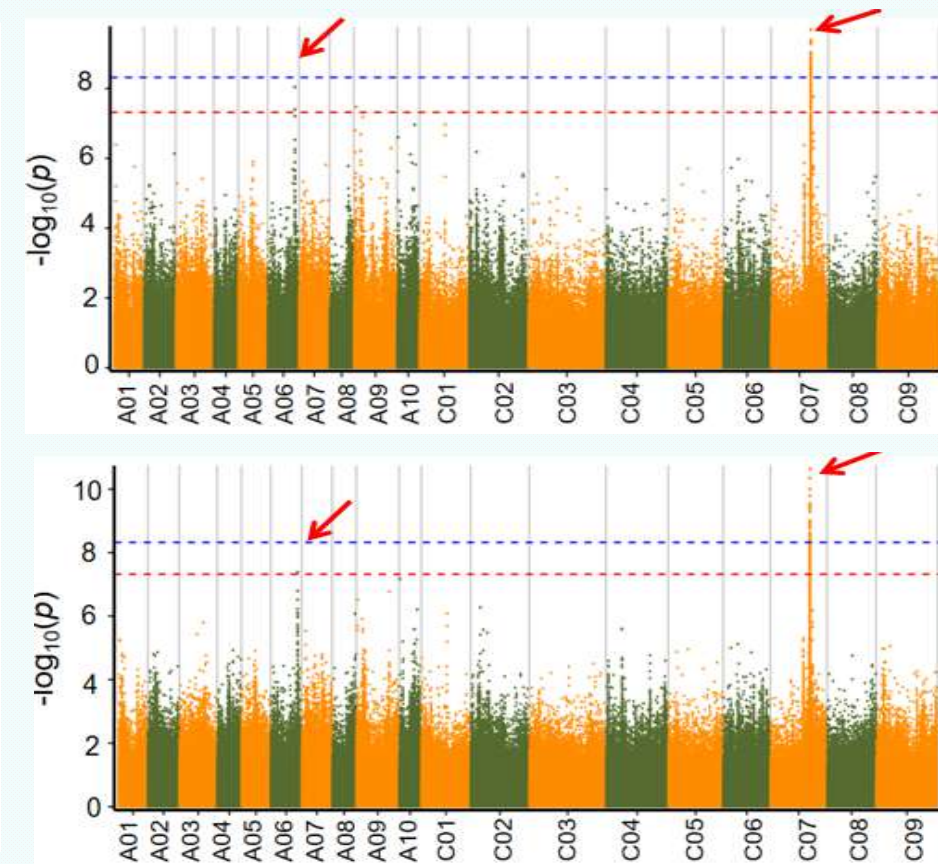
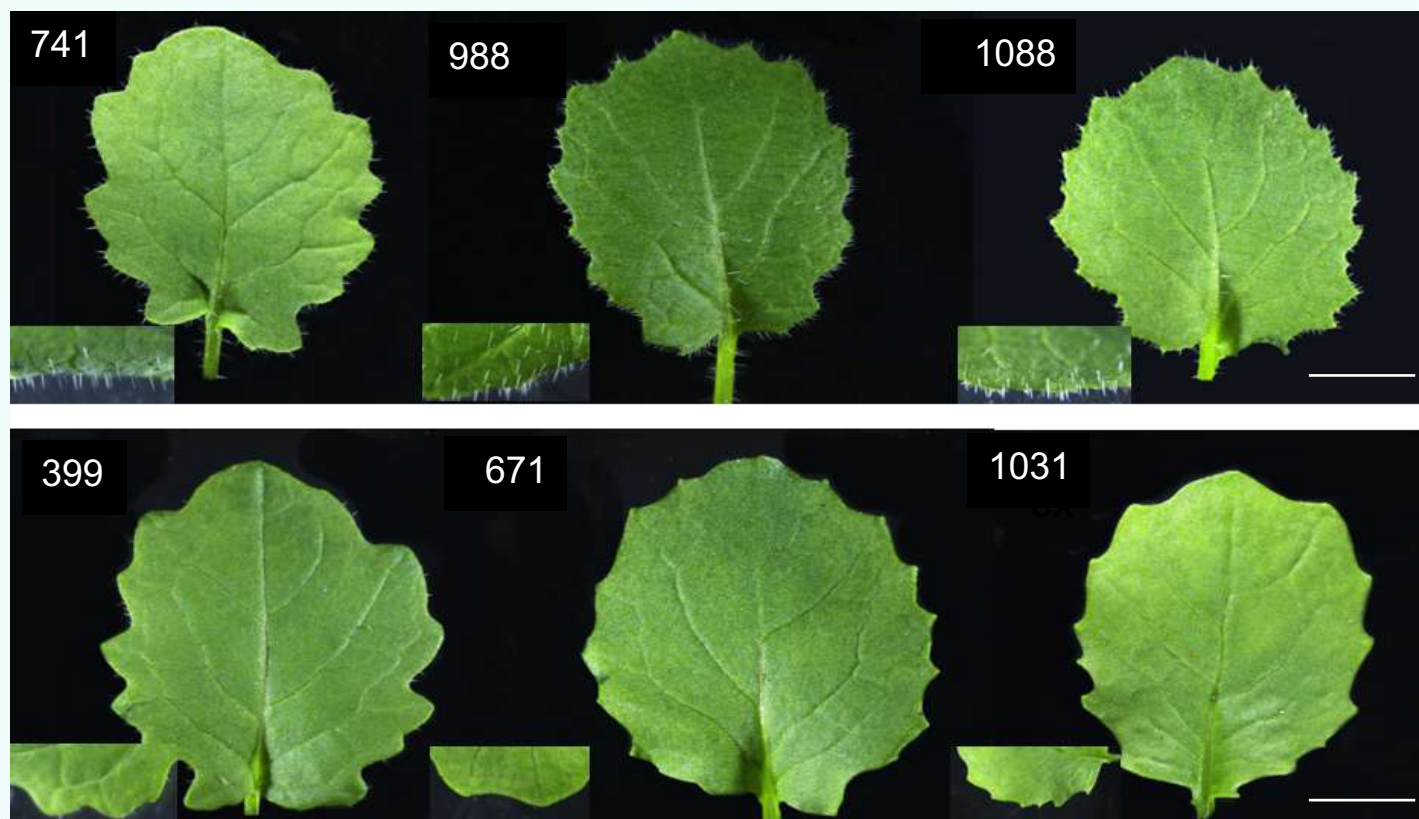
2019年9月16日

Genetic polymorphism of the core accessions for GWAS



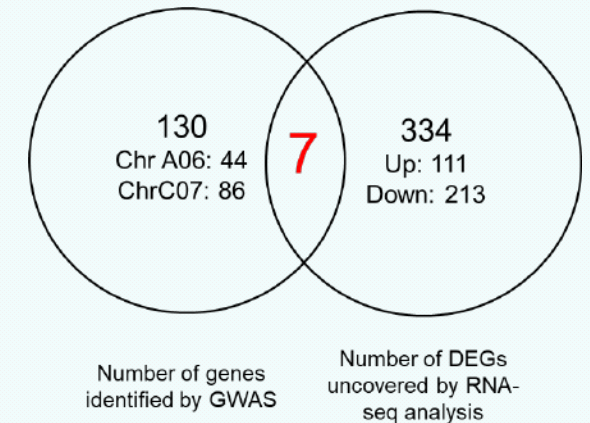
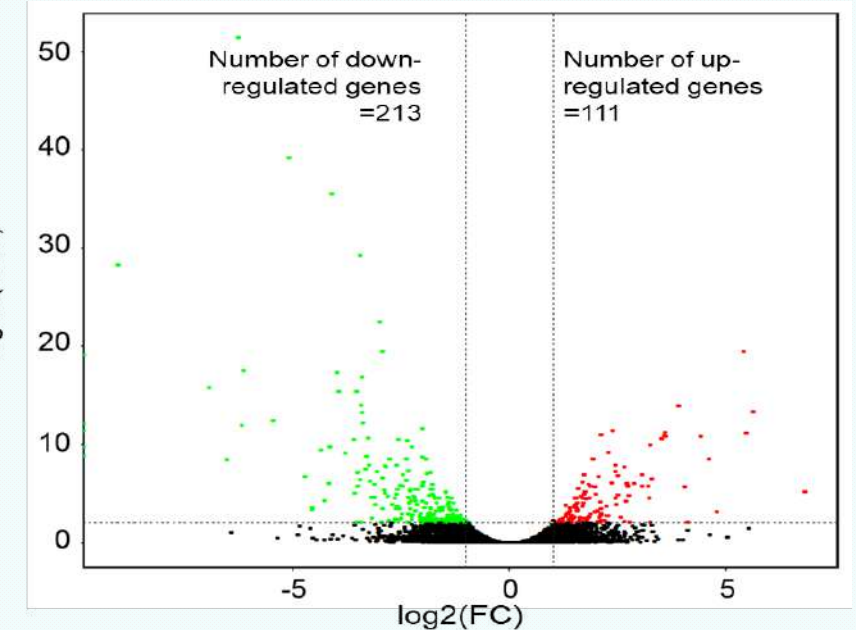
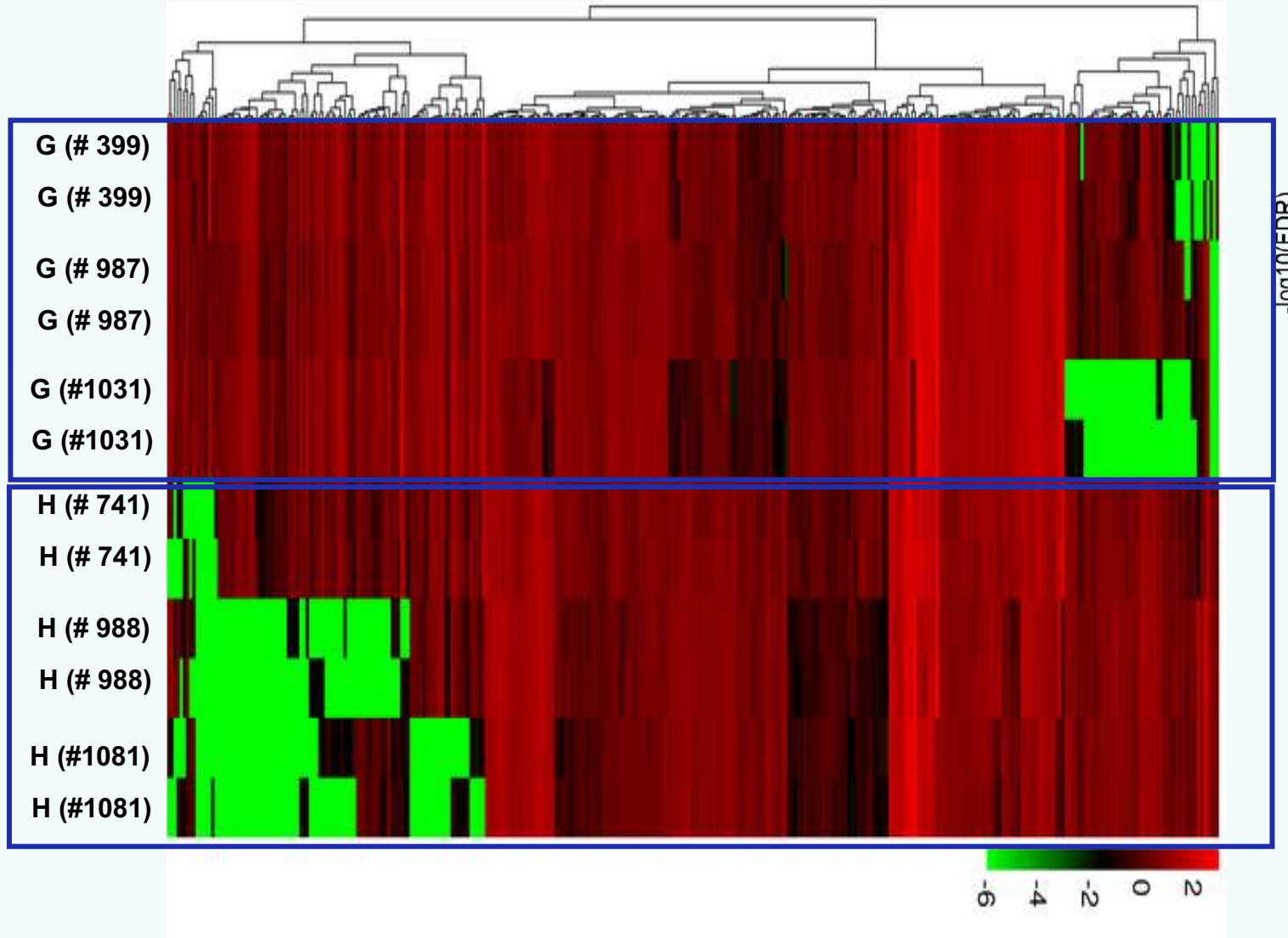
接下来，我们用叶毛性状的GWAS分析，来介绍这个核心群体的有效性。叶毛性状（1）叶毛有利于油菜的嫩叶抵御早期害虫的侵害；（2）在寒冷地区，叶毛有利于防止叶片的表皮结霜、霜冻；（3）在多风、干燥的气候环境中，叶毛可以防止水分过分蒸发，起到抗旱保水作用；（4）浓密的叶毛还能反射太阳光，在高温的气候环境起到保护嫩叶组织的作用；

GWAS on trichome initiation on leaves of young seedlings



我们在浙大紫金港和长兴农场2个地点，对油菜核心种质的叶毛有无情况进行了表型观察。紫金港和长兴农场2个点的GWAS结果基本上一致。GWAS分析显示，（1）在A6和C7两条染色体上，都出现SNP的峰谱；（2）100-KB的区间内，A6染色体上的SNP和包括MYB基因在内的44基因紧密连锁（3）C7染色体上SNP和包括糖转运蛋白和生长素转运在内的86个基因紧密连锁。

Cluster analysis of the gene expression patterns of H- and G-leaf types



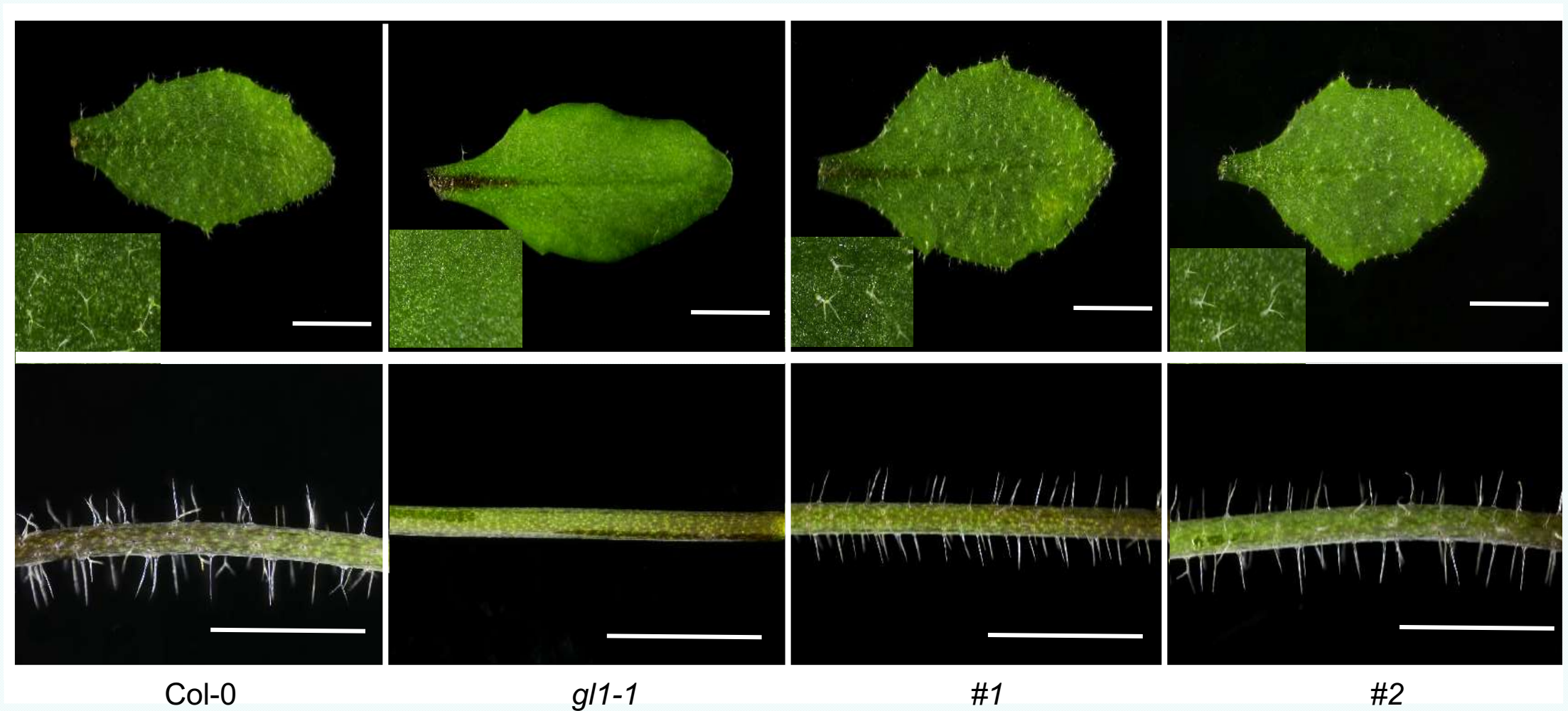
为了缩小GWAS分析初步得到的叶毛调控基因，我们进一步采用 RNA-seq 分析来比对有叶毛的油菜和没有叶毛的油菜基因表达的差异。左边的聚类分析图，右上部分的火山图显示，没有叶毛的叶片和有叶毛的叶片基因表达的状况有很大的不同。交叉分析进一步把调控叶毛性状的候选基因缩小到7个基因的范围。

The candidate genes narrowed down by GWAS and RNA-seq and the annotation

Gene ID	Regulation	Pfam_annotation	nt_annotation
BnaC07g24860D	down	Sugar efflux transporter	<i>Brassica rapa</i> sugar transporter SWEET4
BnaC07g24950D	down	EamA-like transporter family	<i>Brassica rapa</i> WAT1-related protein
BnaC07g24960D	down	EamA-like transporter family	<i>Brassica rapa</i> WAT1-related protein
BnaC07g24970D	down	--	<i>Brassica rapa</i> hsp70-binding protein 1
BnaC07g25000D	down	Glycosyltransferase like family 2	<i>Brassica rapa</i> xyloglucan glycosyltransferase 4
BnaC07g49070D	up	--	<i>Brassica rapa</i> uncharacterized LOC103875260
BnaA06g31780D	up	Myb-like DNA-binding domain	Cotton fibre differentiation protein GL1

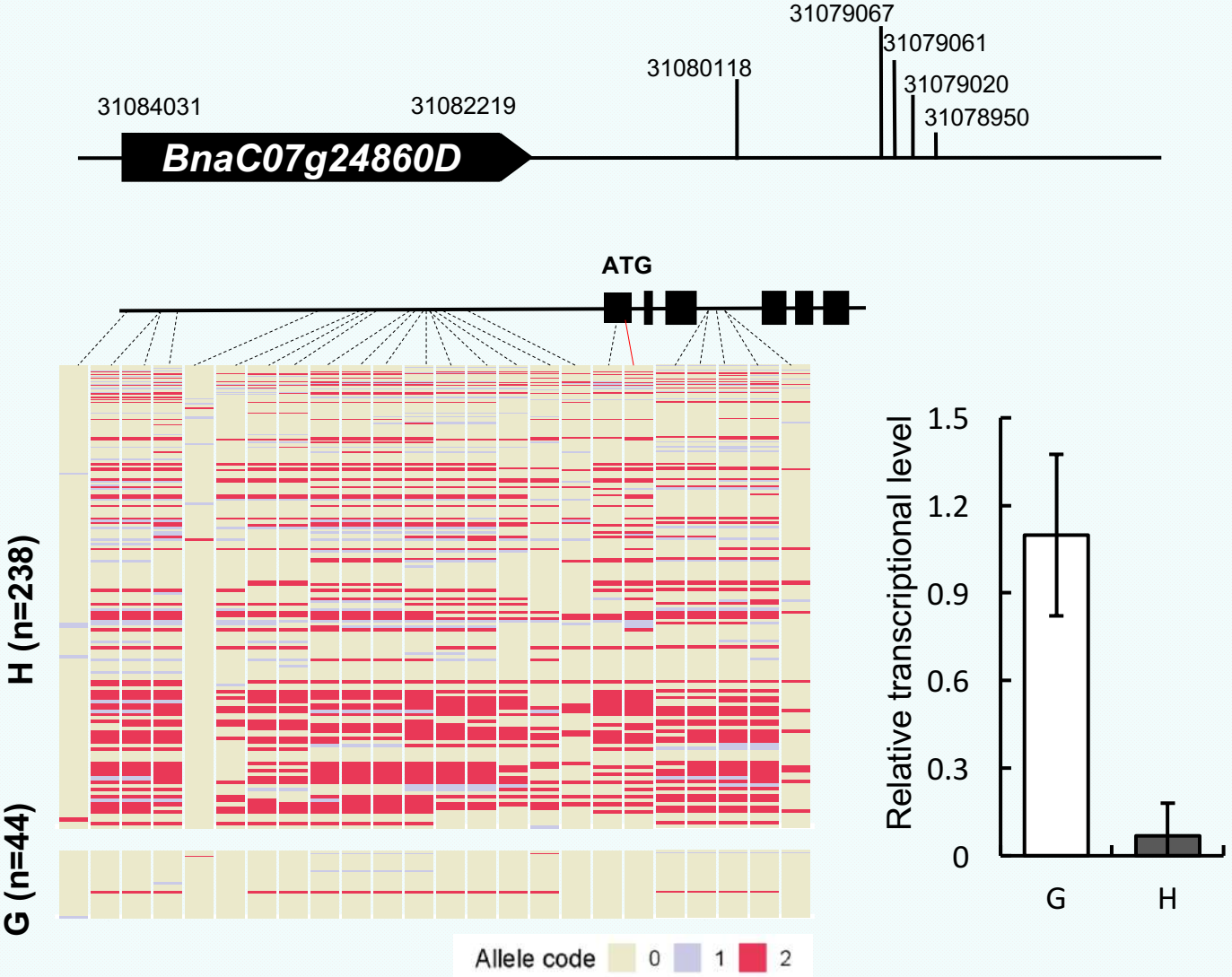
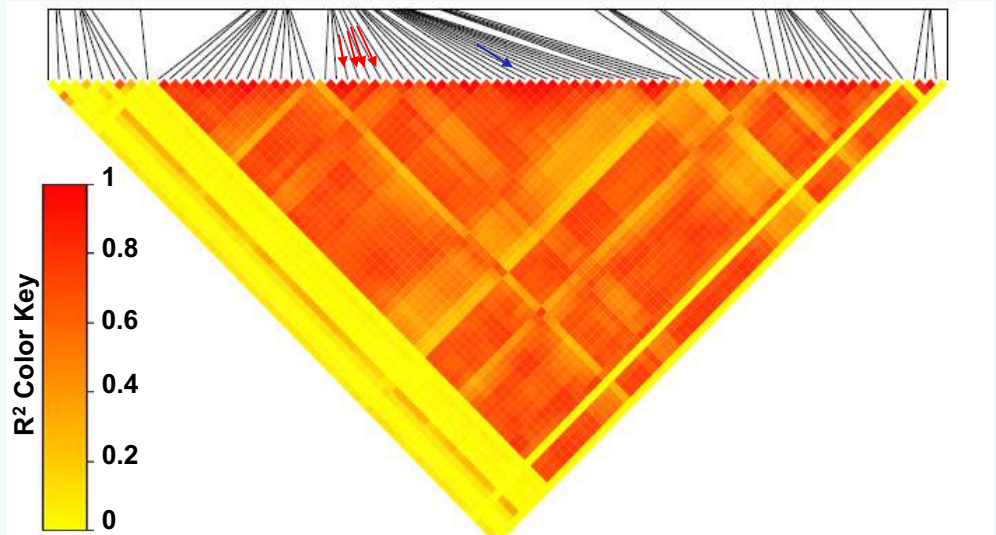
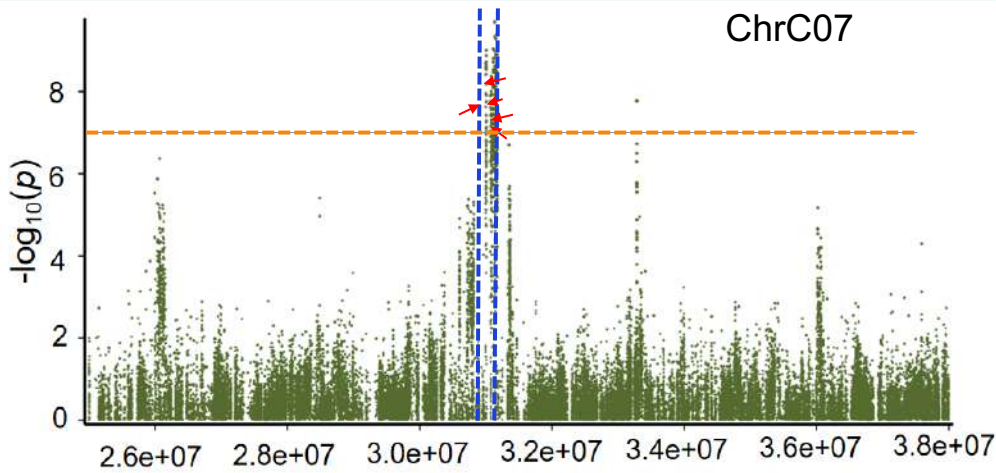
这7个基因中，有3个基因它们在有毛和无毛材料之间不存在SNP的差异，而另外存在SNP差异的4个基因，其中一个是MYB家族的GL1，这个基因在棉花中和棉纤维的发育有关；有一个基因编码糖的运输蛋白，有两个基因编码生长素运输蛋白。

The *gl1* plants harboring the *35S:BnaA06.GL1.a* cassette had trichomes on leaves and stems as WT plants



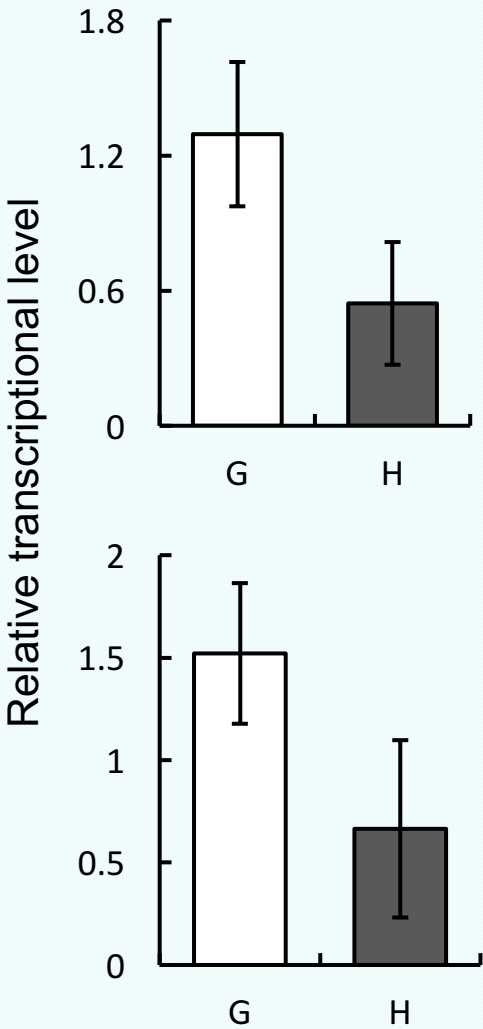
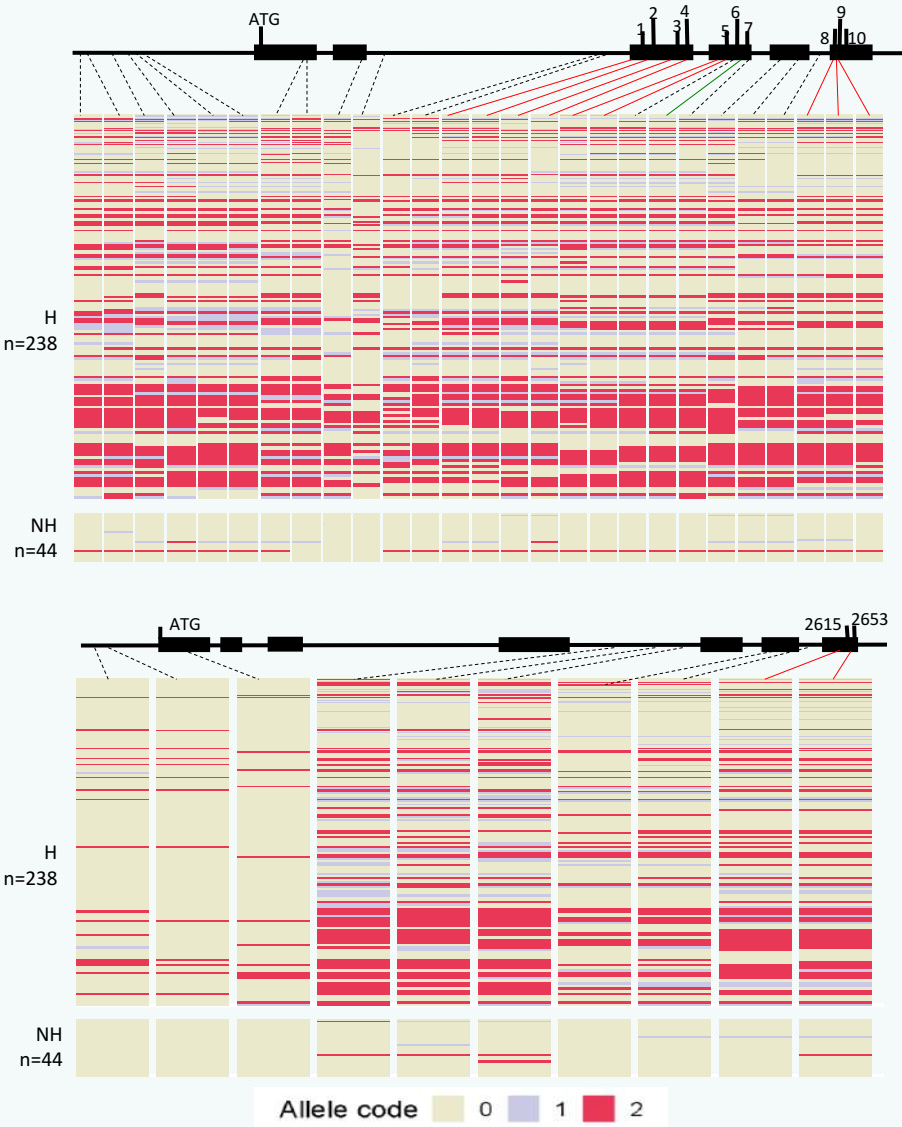
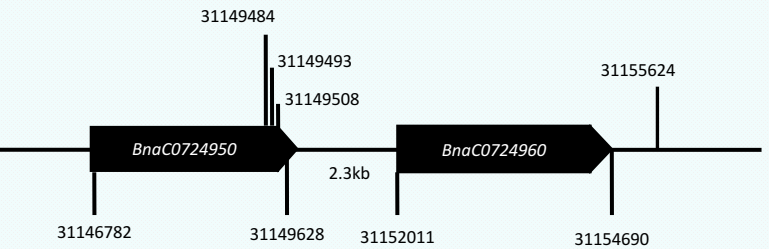
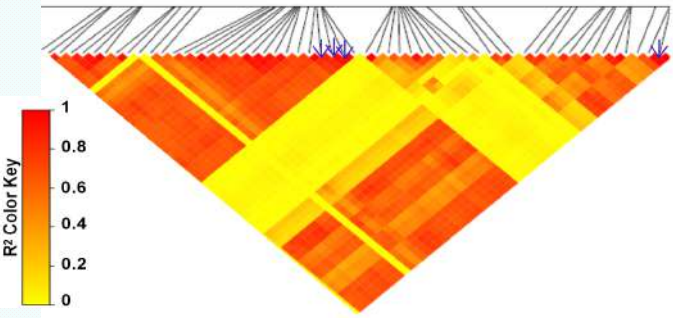
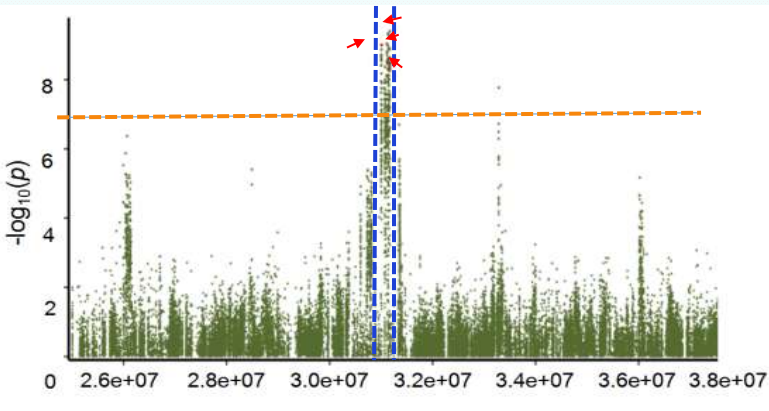
我们用A6染色体上的GL1同源基因来（超表达）互补拟南芥无毛突变体，转基因植株有毛性状得以很好的恢复，说明油菜A6染色体上GL1同源基因确实具有调控叶毛生成的功能。

Negative regulation of *BnaC07.SWEET4.a* on leaf-hair initiation



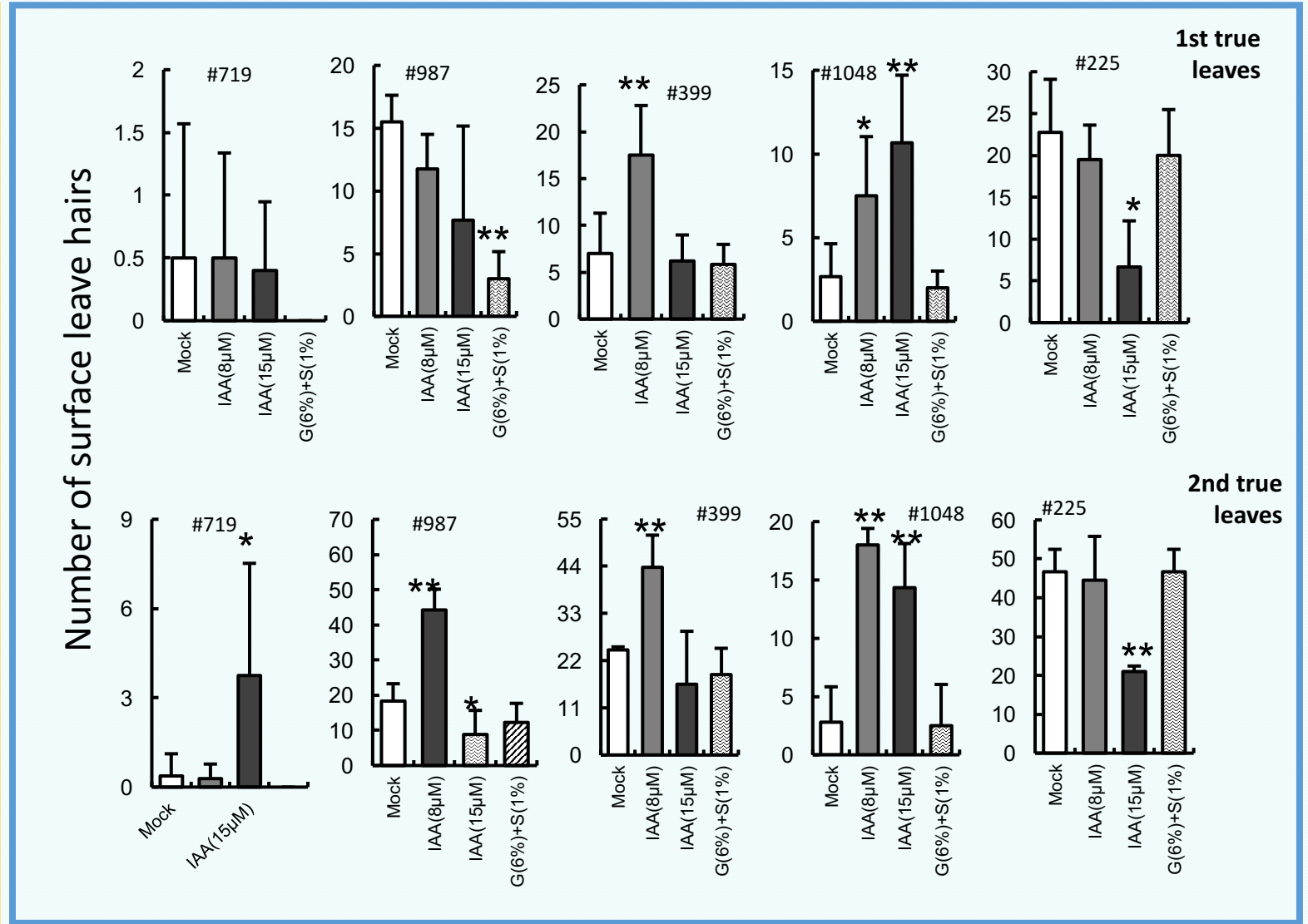
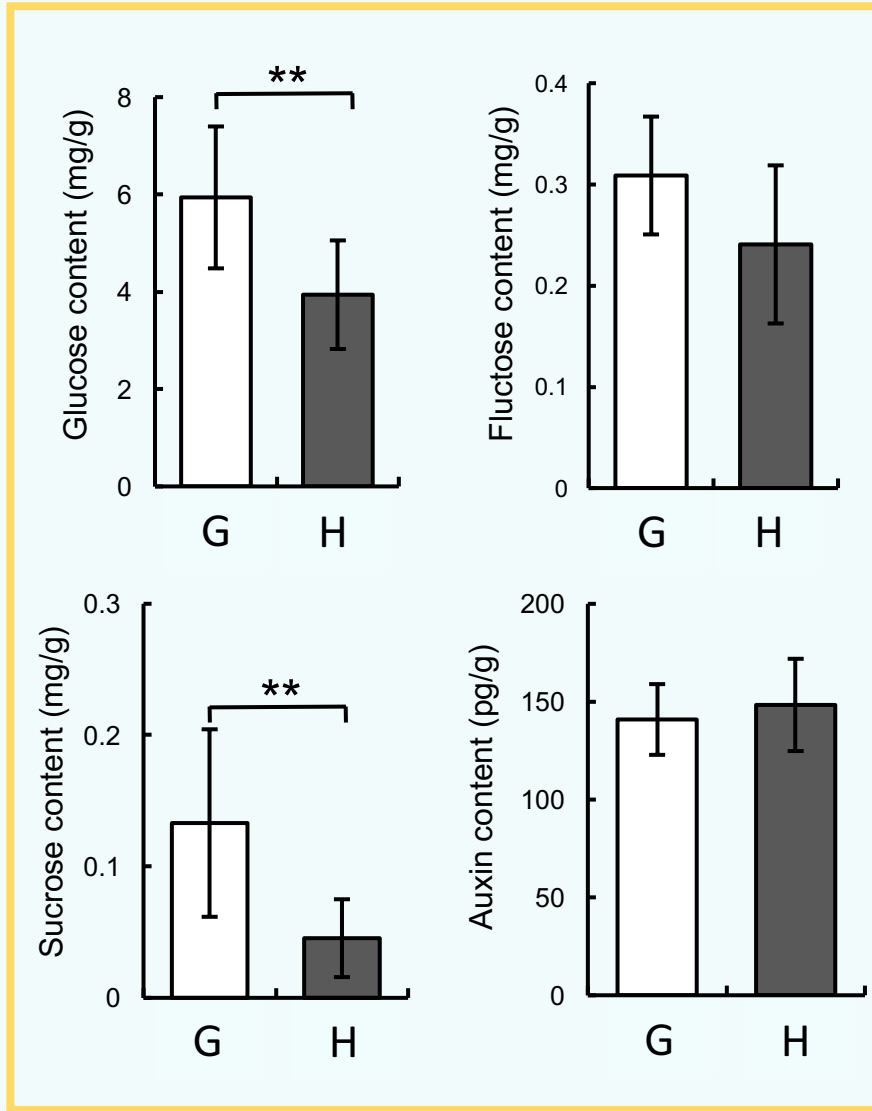
我们用这张图来呈现C07号染色体上与SWEET4紧密连锁的5个SNP的位置，我们用右下角这张分型图来展示SWEET4基因调控区域和编码区域所有SNP的分布，以及这两个基因在有毛和无毛材料中表达水平的差异。其中红的实心线条表示引起氨基酸差别的SNP，虚线表示调控区域的SNP差异。

Negative regulation of *BnaC07.WAT1.a* and *BnaC07.WAT1.b* on leaf trichome initiation



这张图呈现C07号染色体上与WAT1和WAT2紧密连锁的4个SNP的位置，同样我们用右SNP分型图来展示2个WAT1基因调控区域和编码区域所有SNP的分布，在WAT1上有比较多的引起编码氨基酸变化的SNP。

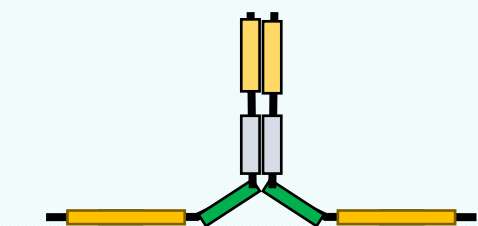
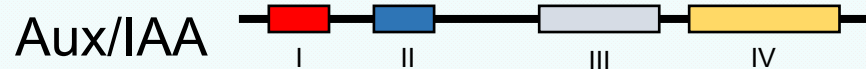
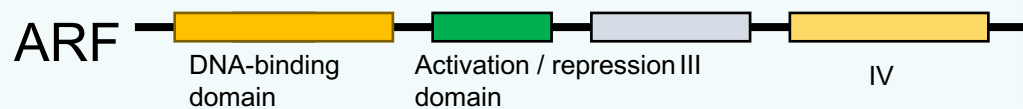
The effect of sugar and auxin on trichome initiation



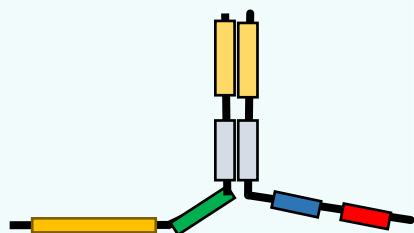
由于 SWEET4 是糖运输蛋白, 而WAT1是保持细胞之间生长素平衡的运输蛋白, 所以我们怀疑糖与生长素信号在叶片叶毛形成的过程中扮演了重要的角色。左边的图显示, 有毛叶片中糖的含量 (无论是葡萄糖、果糖、蔗糖) 都要比无毛叶片要高, 而生长素含量却没有明显的差异。右图显示, 分别用外源的糖和生长素来处理顶端分生组织, 如果浓度合适的话, 会引起叶毛数量的增多, 当然, 这种增加的效果在第1片叶与第2片真叶之间, 在不同基因型之间是有所差别的。

IAAs act as negative regulators on the auxin signaling pathway

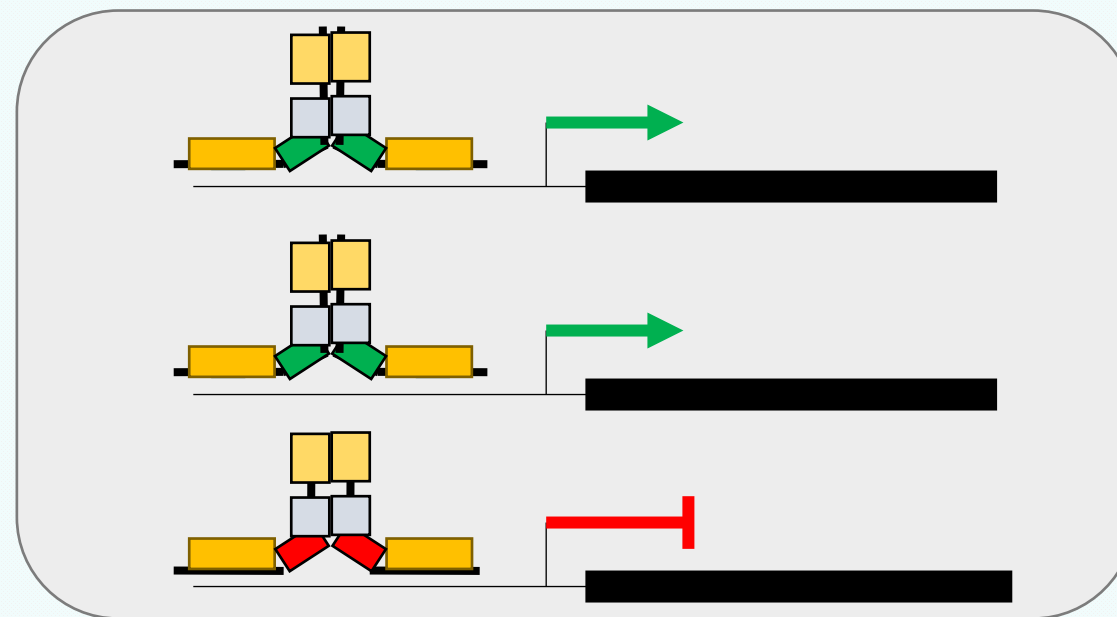
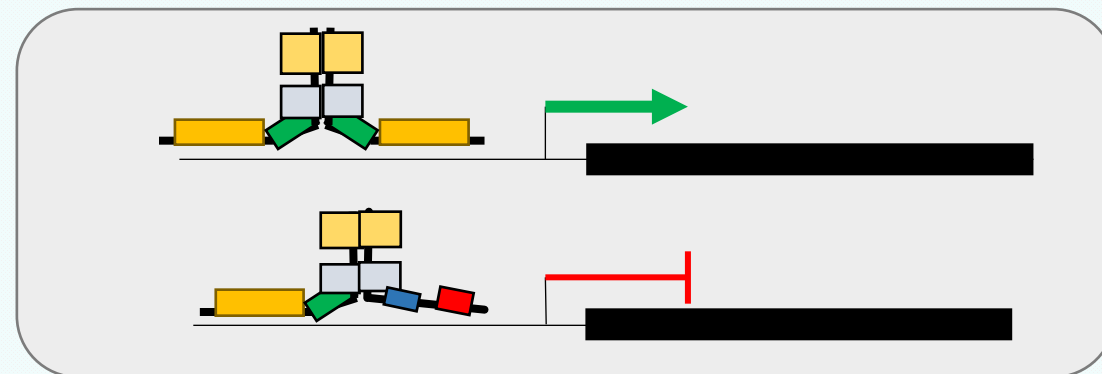
Guilfoyle et al., 2007. Curr. Opin. Plant Biol.



ARF homodimer

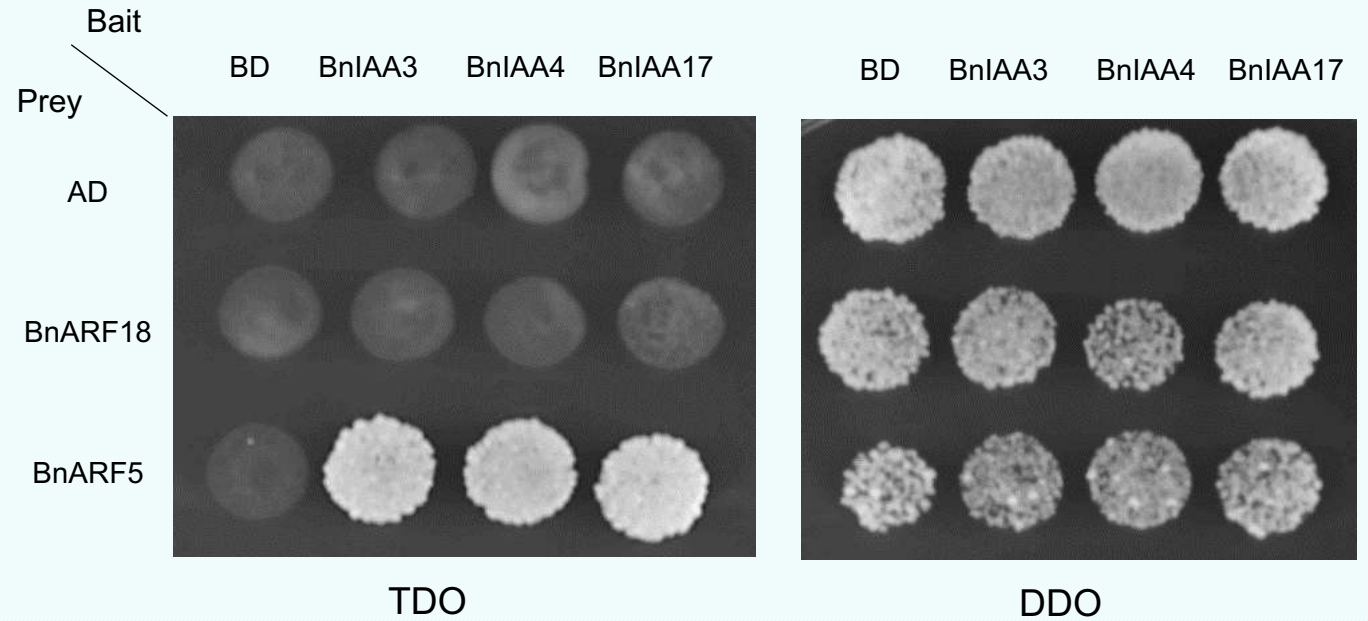
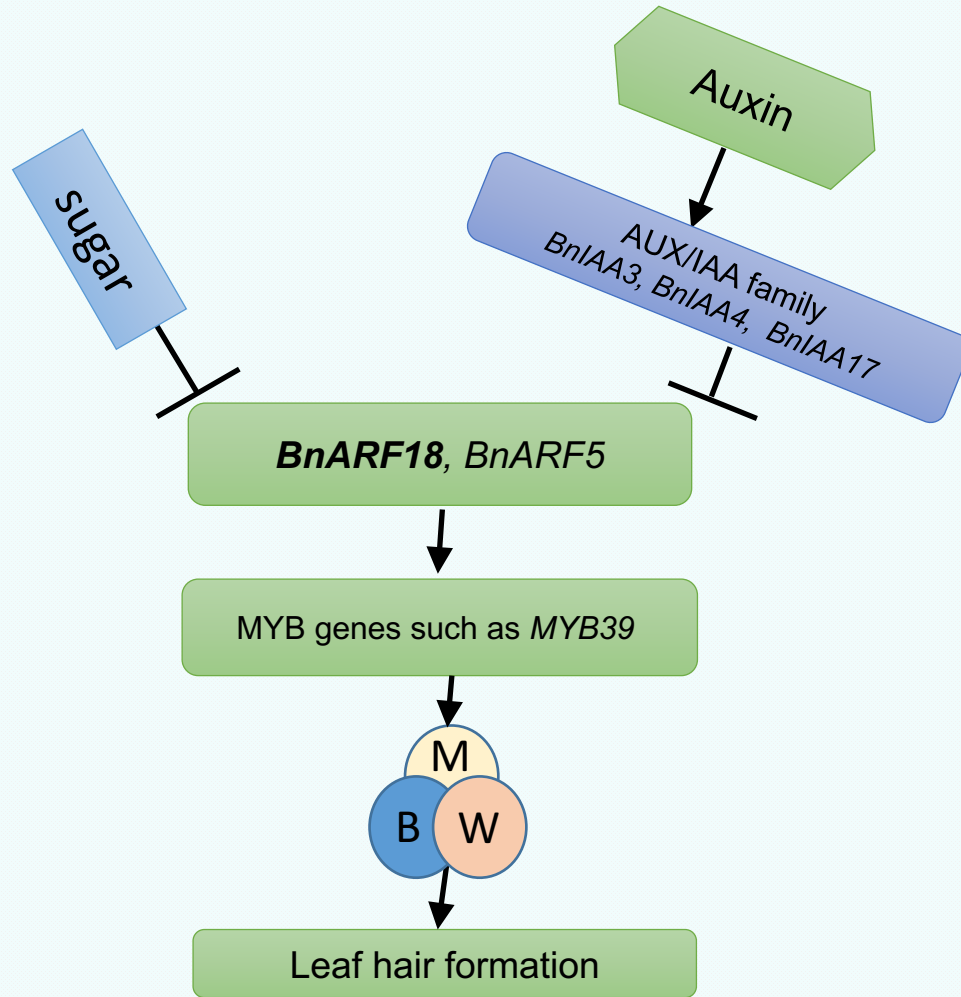


ARF Aux/IAA heterodimer



IAAs实际上与ARF形成异质二聚体，外喷生长素起到分解IAA蛋白释放生长素相应因子，从而激活一系列下游的反应。

A sketch summarizing the probable regulatory factors involving trichome initiation in SAMs of young rapeseed seedlings



我们用这个示意图来汇总一下我们的发现：（1）生长素信号促进摧毁生长素传导途径上的负向调控因子IAA家族基因，而这些基因不利于下游ARF基因的表达；（2）ARF促进MYB家族基因的表达，而MYB基因是BMW三聚复合体的成员，BMW三聚复合体诱导叶毛形成；（3）糖信号途径抑制生长素响应蛋白的表达。。

小 结

1. 我们通过将近1000份油菜种质资源的基因组重测序，解释了世上油菜的遗传多态性，为油菜育种提供了丰富的分子标记。
2. 我们鉴定到了一系列在自然和人工选择过程中留下的选择印迹，以及开花时间、不同生态类型相互区别的分子基础。
3. 基于序列信息，我们进行了群体漂移（ allelic drifting path ）分析，大致明确了世界上油菜栽培扩散的轨迹。
4. 我们建立了核心资源库，该核心库群体小，但是遗传变异的丰富程度高，SNP数量比已经发表的高1-2个数量级。
5. 我们通过核心资源库的GWAS分析，发现 BnaA.GL1.a, BnaC.SWEET4.a, BnaC.WAT1.a and BnaC.WAT1.b等基因和油菜的叶毛性状存在非常强的关联，显示糖信号和生长素信号途径对油菜叶毛的形成具有影响。

Acknowledgement

