

大麦泛基因组揭示了突变育种的潜在遗传痕迹



nature

大麦泛基因组揭示了突变育种的潜在遗传痕迹


The barley pan-genome reveals the hidden legacy of mutation breeding

- ◆ **发表时间:** 2020年11月25日
- ◆ **发表期刊:** Nature
- ◆ **影响因子:** 42.778
- ◆ **主要单位:** 德国莱布尼茨植物遗传与作物栽培研究所
- ◆ **关键词:** 大麦, 泛基因组

背景介绍

大麦 (*Hordeum vulgare*)

- **禾本科、大麦属**一年生草本植物。秆粗壮，光滑无毛，直立，高达100厘米。叶鞘松弛抱茎，多无毛或基部具柔毛；两侧有两披针形叶耳；叶舌膜质，叶片扁平。穗状花序，小穗稠密，小穗均无柄，颖线状披针形，外被短柔毛，内稃与外稃几等长。**颖果熟时粘着于稃内，不脱出。**
- **变种：青裸 (*Hordeum vulgare* var. *nudum*)**，**颖果成熟时易于脱出稃体。**
- 具有很强的环境适应性，**生长环境很广**，而且具有**春、冬生长习性**。
- 具有食用、饲用、酿造、药用等多种用途。
- 起源和驯化中心在**新月沃地**。

| Genus | Species | Subspecies | Authority | Chromosome Number (2n) | Ploidy Level (x) | Estimation Method  | DNA Amount 1C (pg) |
|---------|---------|------------|-----------|------------------------|------------------|--|--------------------|
| Hordeum | vulgare | cv. Sultan | L. | 14 | 2 | Fe | 5.50 |



背景介绍

| 物种 | 时间 | 期刊 | 主要单位 | 品系 | 组装策略 | 组装指标 |
|----|---------|-----------------------------|---------------------|-------------|------------------------------|--|
| 大麦 | 2012.11 | Nature | 国际大麦基因组测序联盟 (IBGSC) | Morex | Illumina 50X | Contig N50 266kb |
| | 2017.04 | Nature | 德国莱布尼茨植物遗传与作物栽培研究所 | Morex | Illumina/454,光学图谱,遗传图谱, Hi-C | Contig N50 79kb Scaffold N50 1.9 Mb |
| 青稞 | 2015.01 | PNAS | 西藏农牧科学院 | Goumang | Illumina 178X | Contig N50 18kb Scaffold N50 242kb |
| | 2017.11 | Plant Biotechnology Journal | 浙江大学 | Zangqing320 | Illumina 50X | Contig N50 5.9kb Scaffold N50 173.8kb |

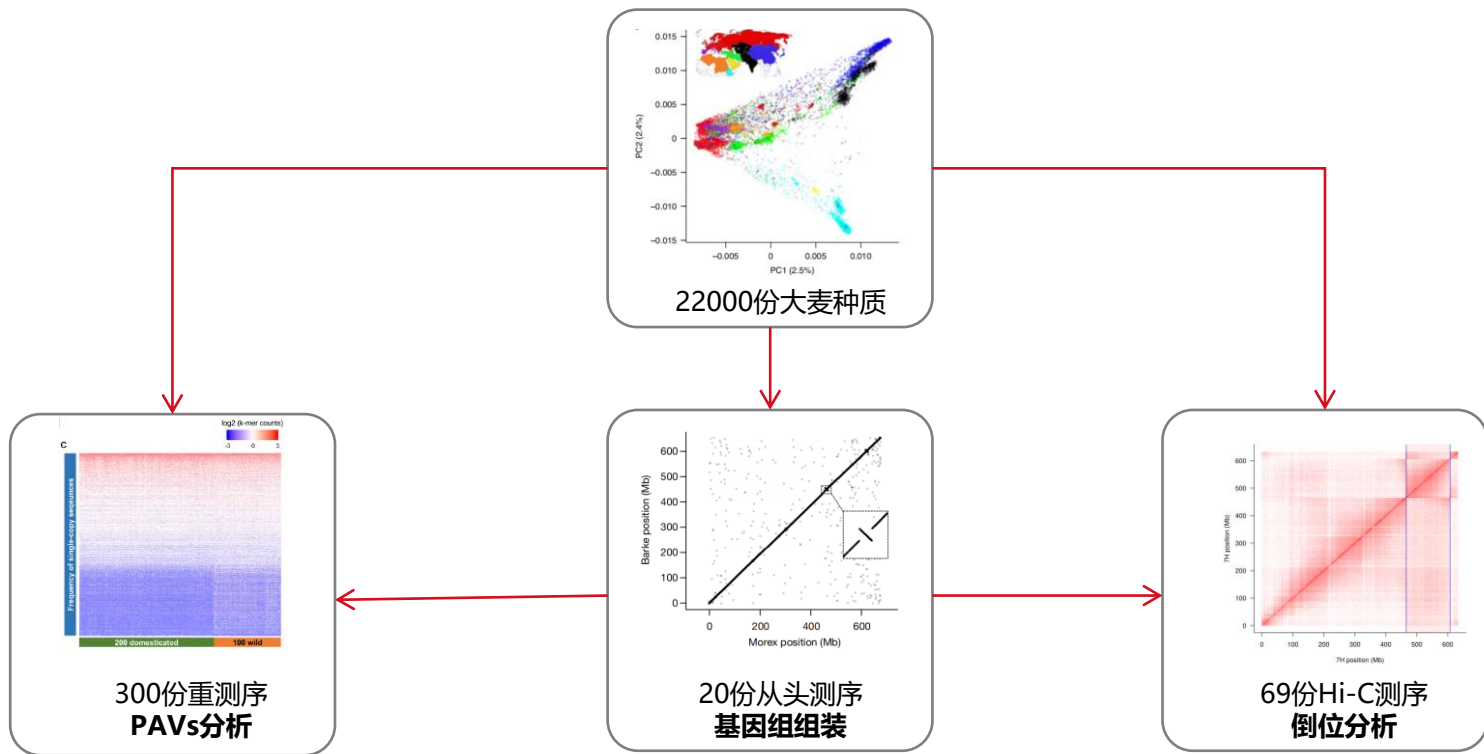
| 物种 | 时间 | 期刊 | 主要单位 | 材料 | 测序策略 | 研究内容 |
|-----------|---------|-----------------------|--------------------|---------------------|---------|---------------------|
| 大麦/ 青稞 | 2018.11 | Nature Genetics | 德国莱布尼茨植物遗传与作物栽培研究所 | 22626份大麦种质 | GBS | 种质库基因组学揭示全球大麦种质的多样性 |
| | 2018.10 | Nature Communications | 西藏农牧科学院 | 437份大麦 (大麦、青稞和野生大麦) | WGS/WES | 青稞在西藏的起源进化 |



Nils Stein

Leibniz Institute of Plant Genetics and Crop Plant Research · Department Genebank
+49 45.55 · Professor

研究内容概要



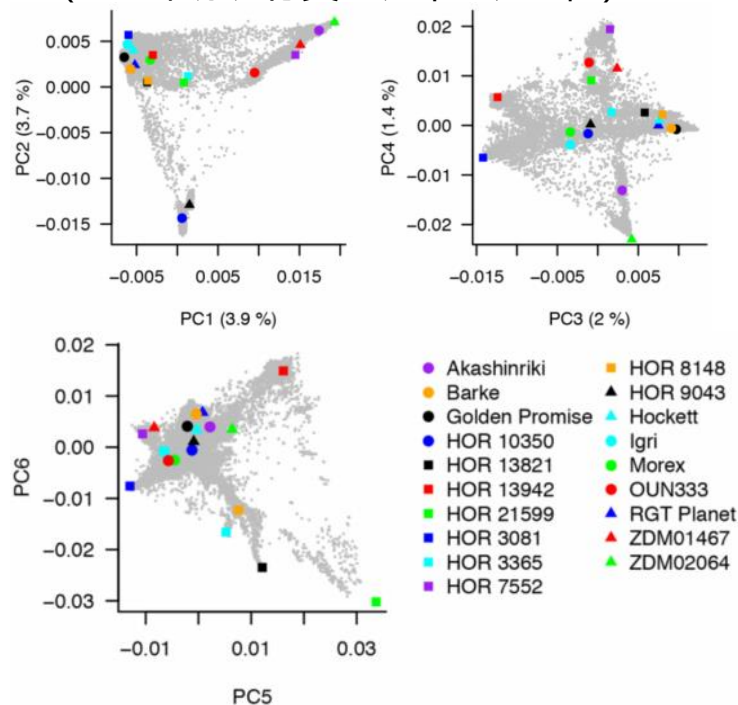
基于短读长的结构变异分析方法

全球代表性大麦染色体水平序列构建

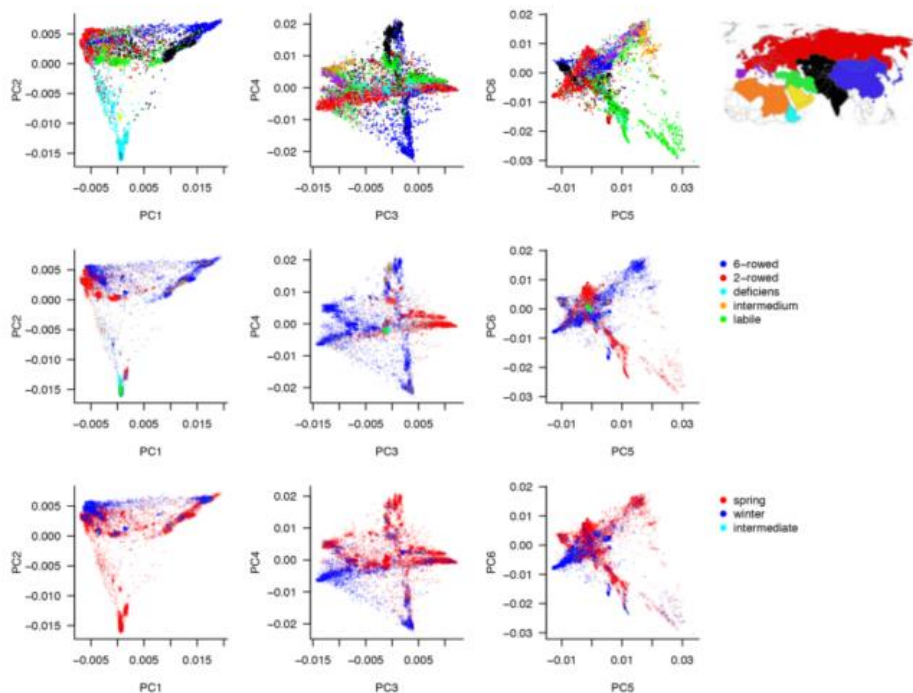
大麦品种中的大倒位多样性

一、20个大麦基因组组装

- **材料选择:** 通过22000份大麦种质的基因分型数据的PCA选择了20份大麦来代表其遗传多样性 (地理来源、行类型、年生长习性)



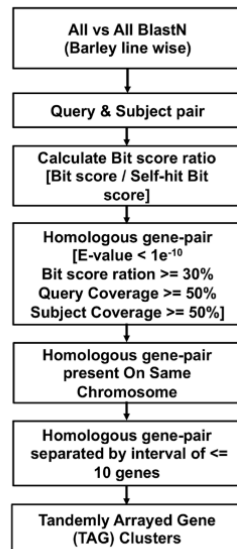
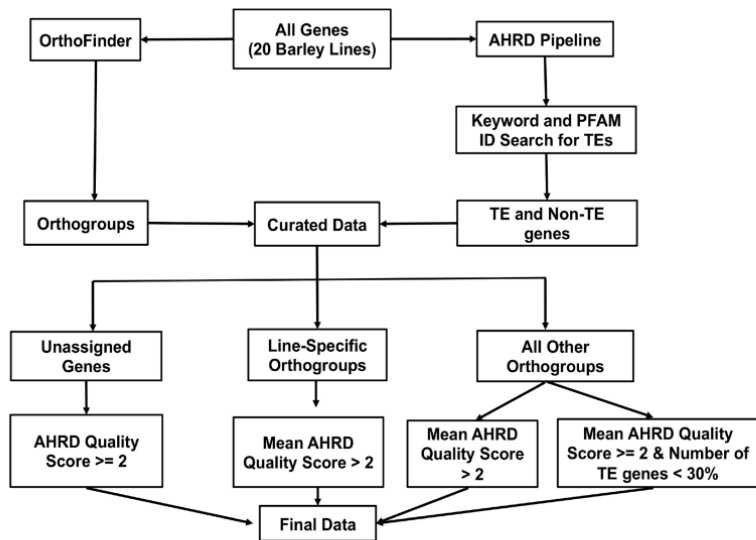
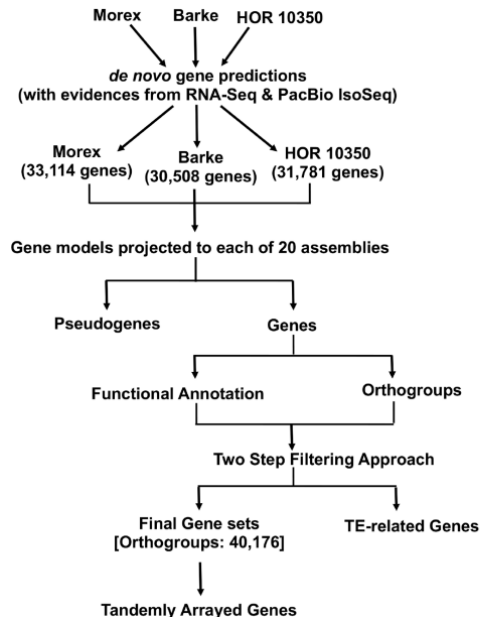
20份代表性大麦的选择



20份大麦的多样性代表性

一、20个大麦基因组组装

- 组装策略：**将Illumina的双末端和配对短序列组装成scaffolds，使用10X Genomics和Hi-C将scaffolds排列成染色体假分子
- 注释方法：**利用RNA-seq和Iso-Seq对Morex、Barke和HOR 10350进行从头基因注释，注释的基因模型被合并并比对到其他17个基因组



注释流程

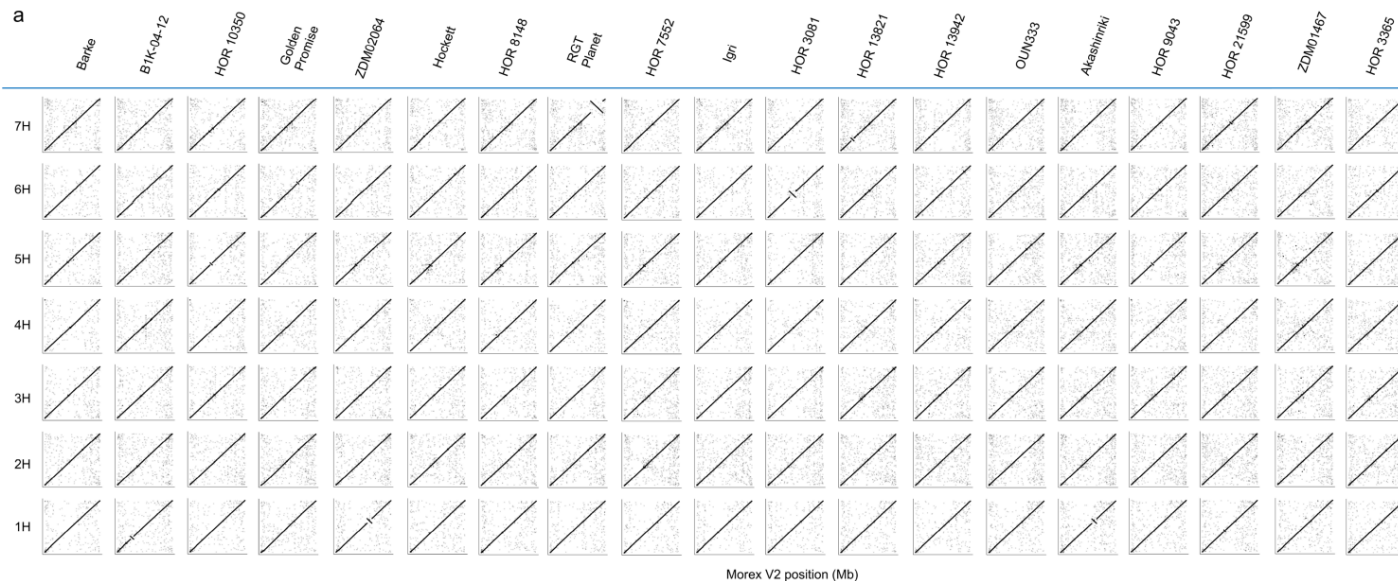
一、20个大麦基因组组装

- 结果：** Scaffolds均达Mb级别，注释到35,859-40,044个基因，直系同源基因聚类产生了40,176个直系同源组

| Accession | Status | Row-type | Growth habit | Country of origin | No. of super-scaffolds [†] | Super-scaffolds N50 (Mb) | Size (Gb) [†] | No. of projected gene models [§] | % of transposons | single-copy sequence (Mb) [¶] | Single-copy unique sequence (Mb) |
|----------------|----------|-------------|--------------|-------------------|-------------------------------------|--------------------------|------------------------|---|------------------|--|----------------------------------|
| Akashinriki | cultivar | 6-rowed | winter | Japan | 345 | 34.3 | 4.4 | 36,948 | 80.2 | 475.8 | 1.9 |
| B1K-04-12 | wild | 2-rowed | spring | Israel | 347 | 32.5 | 4.2 | 36,366 | 80.4 | 478.9 | 13.2 |
| Barke | cultivar | 2-rowed | spring | Germany | 284 | 34.8 | 4.1 | 38,302 | 80.3 | 478.3 | 1.5 |
| Golden Promise | cultivar | 2-rowed | spring | Europe | 1595 | 18.4 | 3.8 | 35,859 | 80.1 | 467.9 | 1.4 |
| Hockett | cultivar | 2-rowed | spring | USA | 482 | 18.4 | 4.0 | 38,725 | 79.9 | 450.0 | 1.6 |
| HOR 10350 | landrace | 6-rowed | spring | Ethiopia | 273 | 30.5 | 4.1 | 38,074 | 80.3 | 477.7 | 2.4 |
| HOR 13821 | landrace | 2-rowed | spring | Turkey | 338 | 33.3 | 4.3 | 37,199 | 80.3 | 477.1 | 2.1 |
| HOR 13942 | landrace | 6-rowed | spring | Southern Europe | 492 | 20.2 | 4.2 | 37,252 | 80.3 | 476.9 | 3.2 |
| HOR 21599 | landrace | 2-rowed | winter | Syria | 318 | 39.4 | 4.3 | 36,834 | 80.3 | 478.1 | 4.7 |
| HOR 3081 | cultivar | 6-rowed | winter | Poland | 552 | 18.9 | 4.2 | 37,502 | 80.2 | 479.4 | 2.2 |
| HOR 3365 | landrace | 6-rowed | winter | Russia | 377 | 31.1 | 4.5 | 40,044 | 79.4 | 476.1 | 2.7 |
| HOR 7552 | landrace | 6-rowed | spring | Pakistan | 558 | 17.7 | 4.2 | 37,047 | 80.2 | 477.9 | 3.7 |
| HOR 8148 | landrace | 2-rowed | spring | Turkey | 519 | 18.8 | 4.2 | 37,474 | 80.3 | 478.5 | 2.0 |
| HOR 9043 | landrace | 6-rowed | spring | Ethiopia | 282 | 42.7 | 4.2 | 37,459 | 80.3 | 478.2 | 2.2 |
| Igri | cultivar | 2-rowed | winter | Germany | 646 | 12.5 | 4.2 | 37,590 | 80.2 | 475.9 | 1.6 |
| Morex | cultivar | 6-rowed | spring | USA | 273 | 40.1 | 4.2 | 38,352 | 80.2 | 479.7 | 2.1 |
| OUN333 | landrace | intermedium | intermediate | Nepal | 337 | 35.1 | 4.4 | 37,042 | 80.3 | 477.2 | 3.7 |
| RGT Planet | cultivar | 2-rowed | spring | Australia | 320 | 37.2 | 4.2 | 37,822 | 80.2 | 479.6 | 1.5 |
| ZDM01467 | landrace | 6-rowed | spring | China | 1397 | 5.0 | 4.4 | 37,116 | 80.3 | 461.3 | 2.5 |
| ZDM02064 | landrace | 6-rowed | spring | China | 720 | 10.9 | 4.0 | 37,292 | 80.2 | 475.5 | 2.2 |

一、20个大麦基因组组装

- 结果：20个基因组具有相似的参考基因模型，基因组彼此高度共线



b

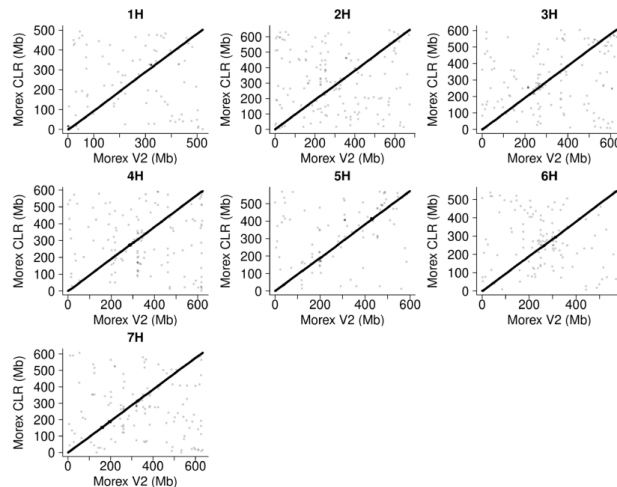
| Genotype | Morex | Barke | B1K-04-12 | HOR 10350 | Golden Promise | Chiba Damia | Hockett | HOR 8148 | RGT Planet | HOR 7552 | Igri | HOR 3081 | HOR 13821 | HOR 13942 | OUN333 | Akashinriki | HOR 9043 | HOR 21599 | Du-Li Huang | HOR 3365 |
|-----------------------------|-------|-------|-----------|-----------|----------------|-------------|---------|----------|------------|----------|------|----------|-----------|-----------|--------|-------------|----------|-----------|-------------|----------|
| % of complete gene coverage | 99.8 | 94.8 | 93.1 | 94.7 | 94.2 | 94.1 | 95.8 | 95.0 | 95.3 | 94.5 | 95.5 | 95.8 | 94.9 | 94.9 | 94.3 | 94.2 | 94.8 | 94.2 | 94.3 | 94.9 |
| % of cDNA alignment | 91.0 | 91.0 | 88.0 | 89.0 | 91.0 | 89.0 | 92.0 | 91.0 | 91.0 | 89.0 | 91.0 | 91.0 | 91.0 | 90.0 | 89.0 | 89.0 | 89.0 | 89.0 | 89.0 | 90.0 |

20个基因组的共线性和基因完整性

一、20个大麦基因组组装

- 结果：** Morex品种的短读长与PacBio长读长组装具有高共线性，并且具有检测PAV的相似能力

| | Morex V2 | Morex CLR |
|-------------------------|---------------|---------------|
| Pseudomolecule size# | 4,257,712,555 | 4,072,877,080 |
| contig N50 | 32.7 kb | 10.2 Mb |
| contig N90 | 1.4 kb | 1.9 Mb |
| Bionano label sites (%) | 90.5 | 93.1 |
| % cDNA alignment | 91.0 | 89.5 |



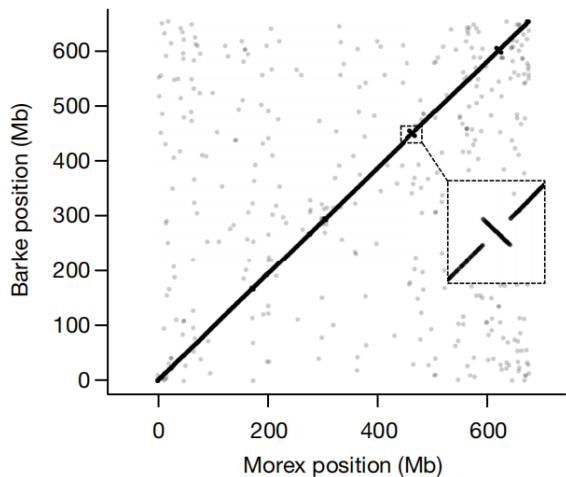
| Reference | Query | No. PAVs (> 1 kb) | Presence# | Absence# | Single-copy overlapping presence | Single-copy overlapping absence |
|-----------|-------|-------------------|-----------------|-----------------|----------------------------------|---------------------------------|
| Morex V2 | Barke | 14,636 (66.8 Mb) | 7,013 (38.5 Mb) | 7,623 (28.3 Mb) | 3,378 (5.4 Mb) | 4,686 (5.5 Mb) |
| Morex CLR | Barke | 15,532 (87.4 Mb) | 8,226 (44.2 Mb) | 7,306 (43.2 Mb) | 3,598 (5.6 Mb) | 4,232 (5.4 Mb) |

二、泛基因组作为遗传学和育种的工具

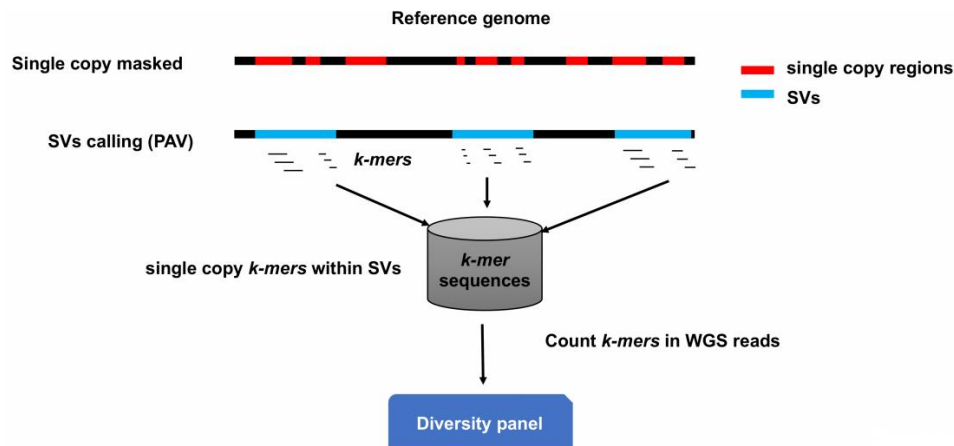
两种方法检测结构变异:

1) **基因组比对**: 将19个组装分别与Morex对比

2) **单拷贝序列聚类**: 从每个组装中提取单拷贝区域, 生成非冗余序列 (单拷贝泛基因组), 利用多样性群体 (200驯化+100野生) 的短读长序列进行k-mers ($k = 31$) 计数, 估计与结构变异重叠的单拷贝簇的丰度



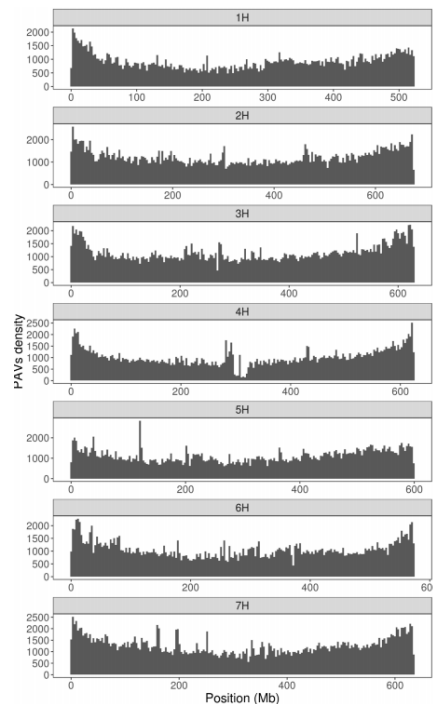
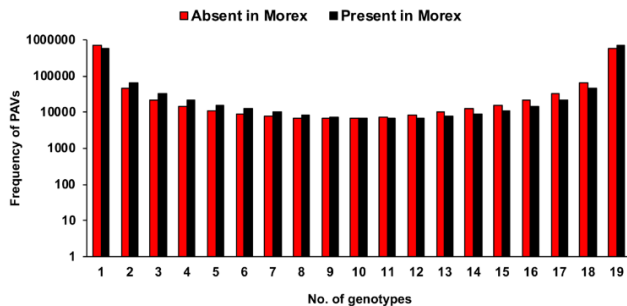
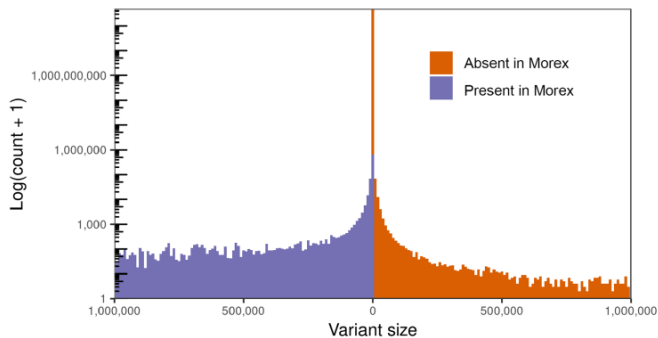
Barke与Morex的H7染色体比对



利用短读长序列k-mers的结构变异分析方法

二、泛基因组作为遗传学和育种的工具

- 结果：**基因组比对鉴定到1,586,262个PAVs，大小从50到999,568 bp，呈现低频变异的富集，在远端密度较高

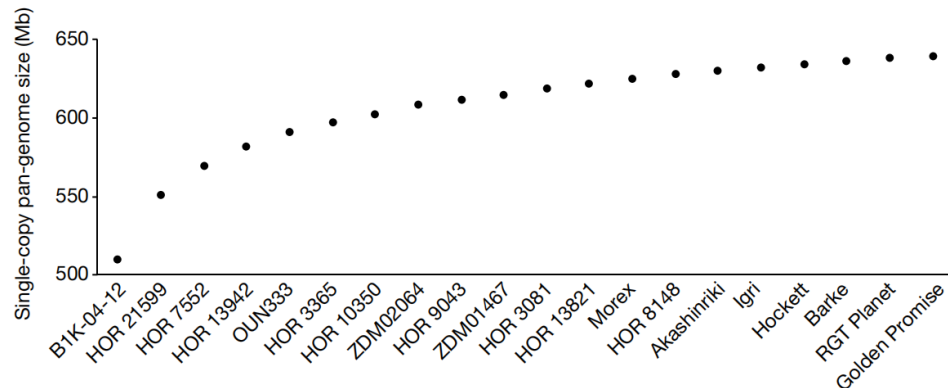


基因组比对检测SV的长度、频率和染色体分布

二、泛基因组作为遗传学和育种的工具

- 结果：**基因组中单拷贝序列的平均累积大小为478 Mb（基因组的9.5%），非冗余单拷贝序列的大小为638.6 Mb（1,472,508个簇，N50为1,087 bp），其中402.5 Mb在所有20个基因组之间共享

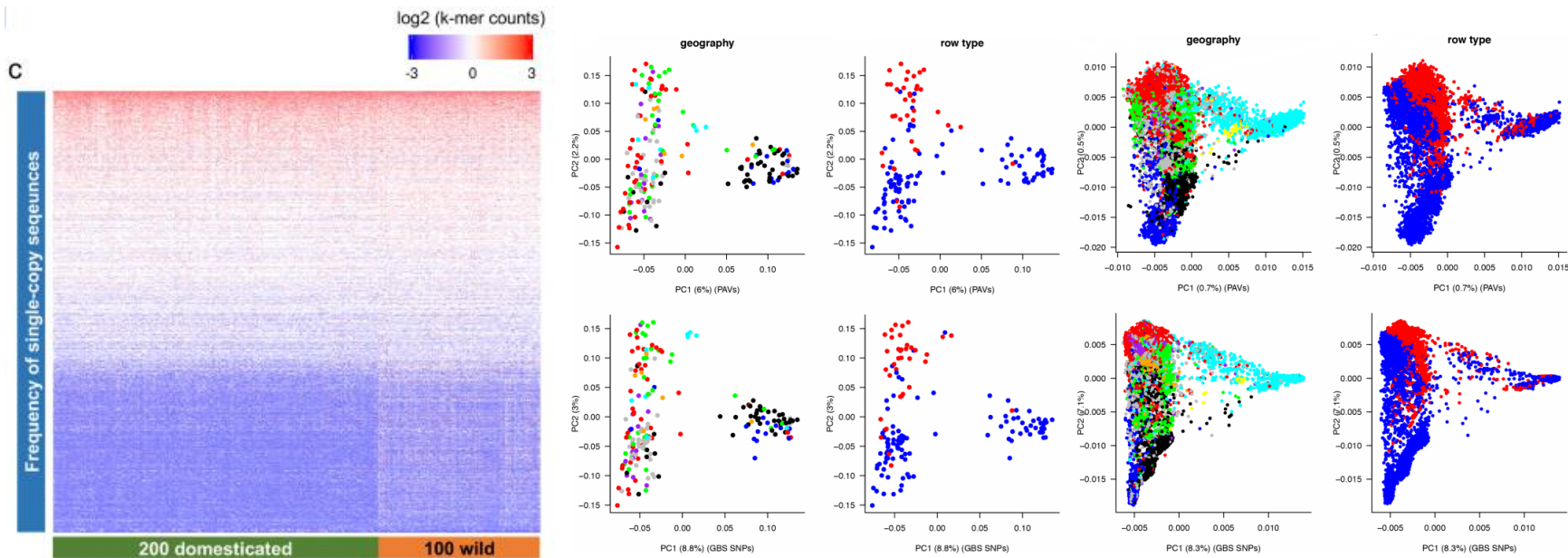
| | Size (Mb) | Number of sequences | N50 |
|--|------------|---------------------|----------|
| single-copy sequence per genotype (mean, SD) | 478.5, 4.8 | 905,318 | 1,177 bp |
| all clusters | 638.6 | 1,472,508 | 1,087 bp |
| PAV clusters | 236 | 1,013,765 | 269 kb |
| singletons | 59.1 | 395,235 | 132 kb |



单拷贝序列泛基因组大小

二、泛基因组作为遗传学和育种的工具

- 结果：**基于k-mers的丰度估计（泛基因组标记）表明，单拷贝序列的位点在不同种质中的拷贝数变异很大，泛基因组标记和与SNP生成相似的群体结构

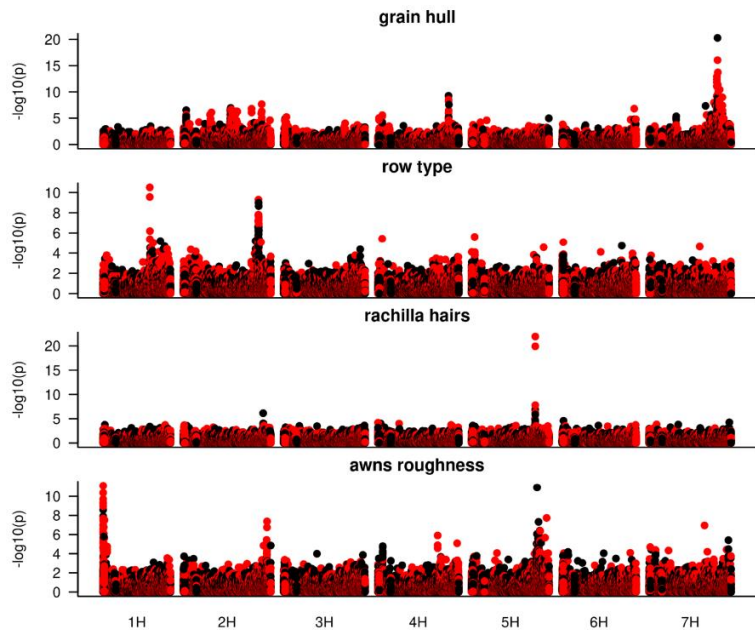


300份重测序样本中的结构变异k-mers丰度估计

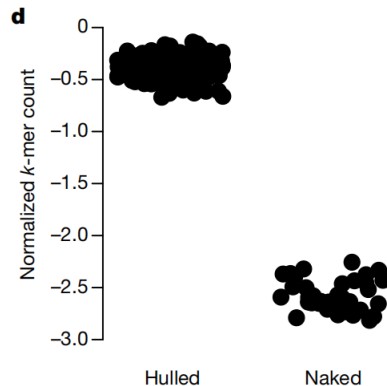
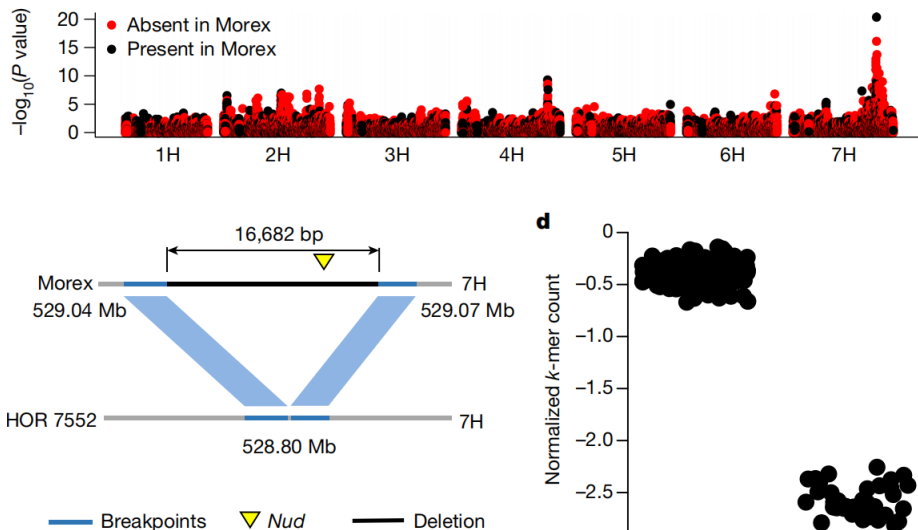
基于泛基因组标记的300份重测序（左）和20000份GBS大麦（右）的PCA

二、泛基因组作为遗传学和育种的工具

- 结果：**基于泛基因组标记的GWAS显示与先前研究一致的峰值。与籽粒依赖性关联的泛基因组标记包含 *NUDUM* (*NUD*) 基因，36个裸粒品种均含有缺失



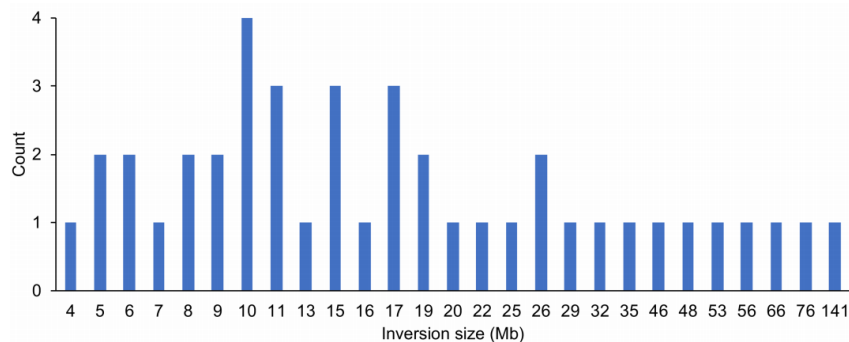
基于泛基因组标记的GWAS



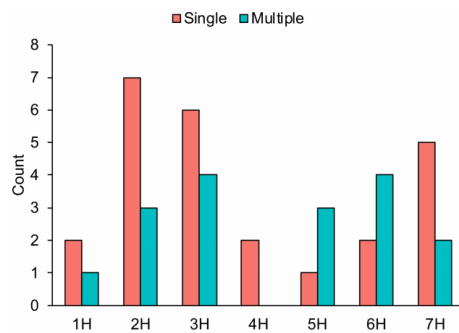
基于泛基因组标记的籽粒依赖性GWAS揭示裸粒大麦中的缺失

三、多态性倒位图谱

- **材料方法：**67驯化和2野生种，Hi-C测序
- **结果：**鉴定到总共42个倒位事件，大小在4到141 Mb之间，大多数为低频



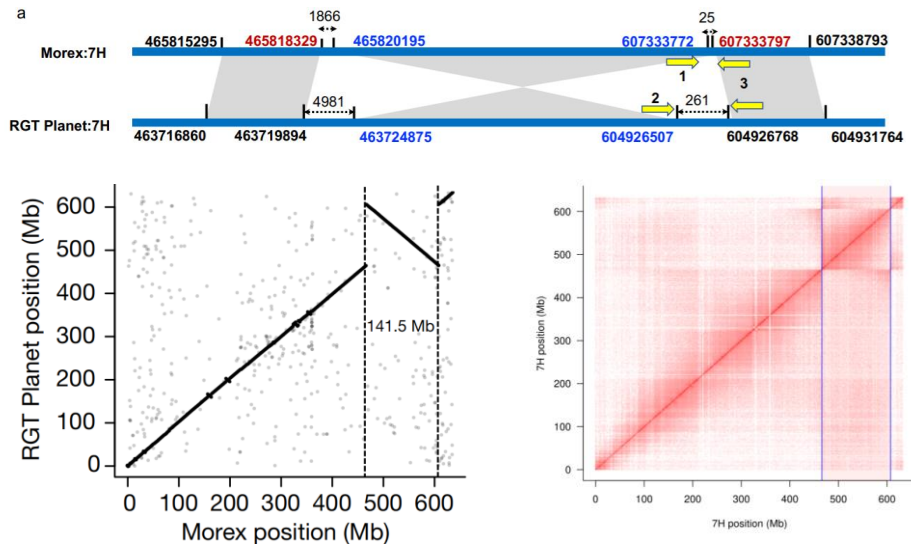
69份大麦中大倒位的大小分布



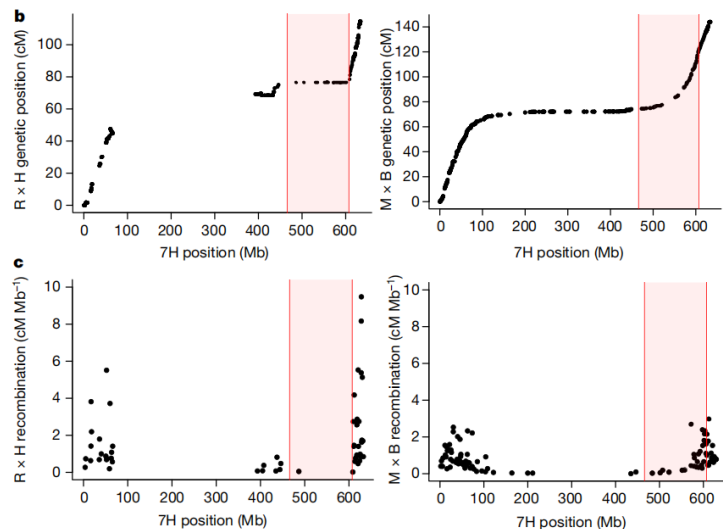
69份大麦中大倒位的频率分布

三、多态性倒位图谱

- 结果：** RGT Planet中检测到的7H染色体倒位是最大事件（141 Mb）；相对于Morex×Barke, RGT Planet×Hindmarsh群体倒位侧翼区域的重组频率增加，表明存在抑制大量重组缺失时补偿机制



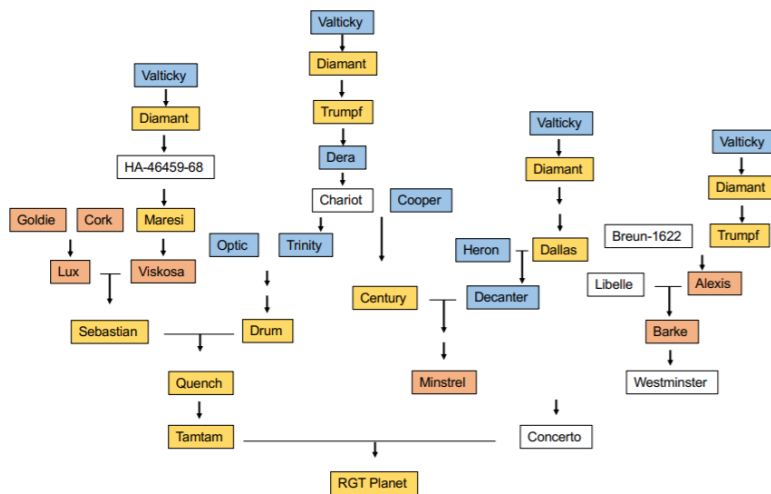
RGT Planet 7H染色体上的大倒位事件



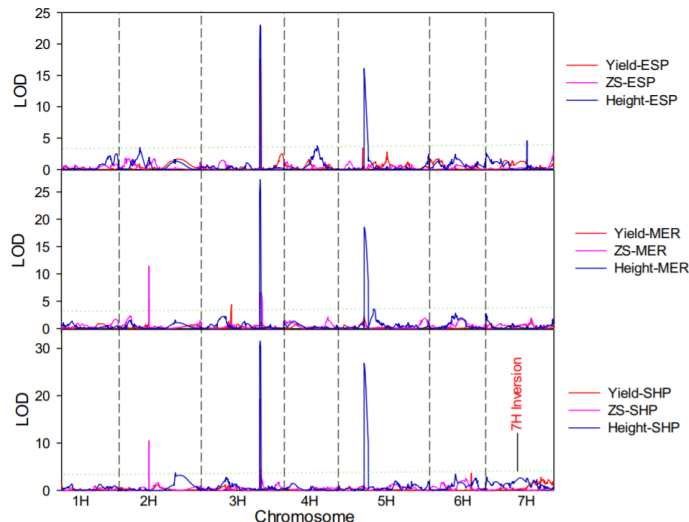
Morex×Barke和RGT Planet×Hindmarsh群体7H染色体上的倒位事件附近的重组率比较

三、多态性倒位图谱

- 方法：**对该倒位设计诊断性PCR检测方法，对全部种质进行分型
- 结果：**RGT Planet谱系的许多系中均被发现该倒位，最早出现在Diamant（由Valticky辐射诱变而来）中，Valticky样品均未进行倒位，而在Diamant样品中分离；RGT Planet×Hindmarsh群体中产量相关性状的定位未在7H染色体上显示信号，与倒位的选择性中立一致，表明诱变育种导致这一育种家未知的大倒位



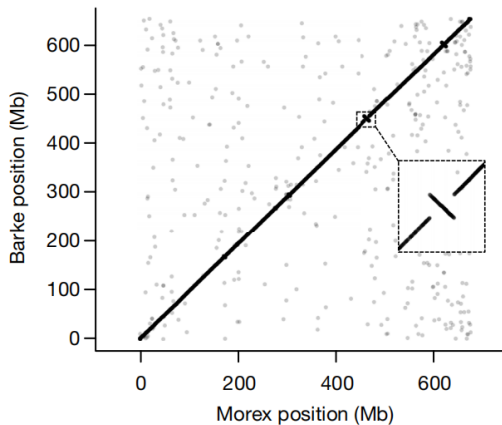
RGT Planet谱系中7H倒位的PCR基因分型
(黄/蓝，携带/不携带RGT Planet等位，红，为扩增出)



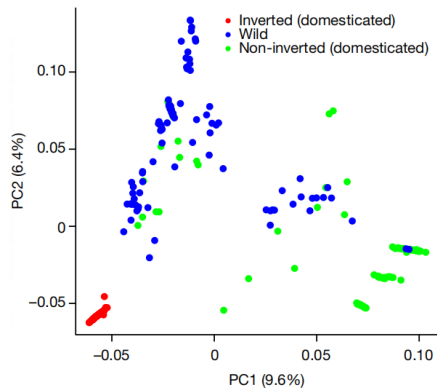
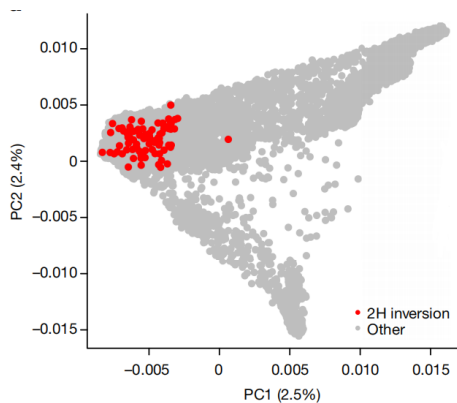
RGT Planet×Hindmarsh群体中产量相关性状的定位

三、多态性倒位图谱

- 结果：**2H上跨度一个10 Mb的倒位出现在69个Hi-C样本中的26个中，对200个驯化的大麦品种和100个野生大麦品种进行的区域PCA和单倍型分析表明，倒位单倍型是单一来源



Barke 2H染色体上的大倒位事件



具有2H染色体上大倒位的样本分布

总结

亮点:

- 首个大麦泛基因组
- 基于大规模群体重测序的材料选择代表性个体
- 基于短读长数据的单拷贝序列k-mer计数PAV检测方法

不足:

- 组装策略不先进, 未使用三代测序
- 泛基因组的深入研究不足
- 生物学故事不够, 没有与驯化改良和重要性状机制解析结合起来

Thank You!



官方网站



官方微信

www.berrygenomics.com