



中国农业科学院(深圳)农业基因组研究所  
Agricultural Genomics Institute at Shenzhen  
Chinese Academy of Agricultural Sciences

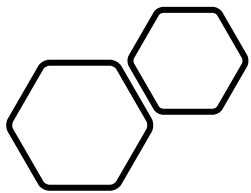
## 马铃薯图基因组的构建

## The construction of potato graph genome

汇报人：程林

小组成员：张文婧 王丹丹 程林

chenglin\_solab@163.com

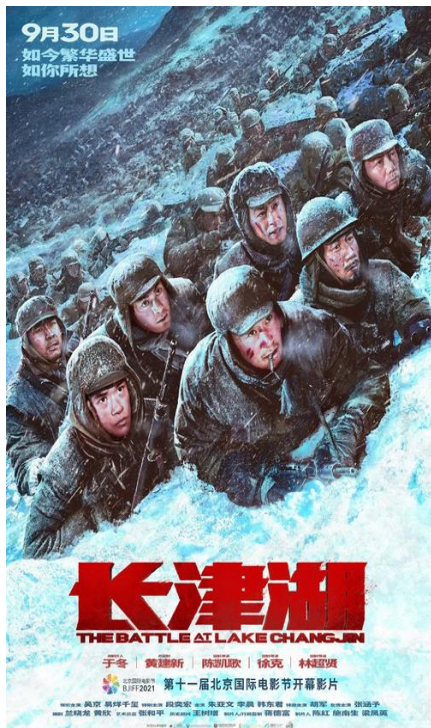


- **马铃薯(Potato)**
- **图基因组构建(Graph genome)**

# 一、马铃薯 特别？



# 马铃薯的重要性



一些国家把马铃薯列为战略储备粮  
马铃薯对种植环境要求不高  
马铃薯是世界第四大粮食作物。  
~120个国家  
~13亿人口



# 2013 年 世界十大主食排行榜

## Production [ edit ]

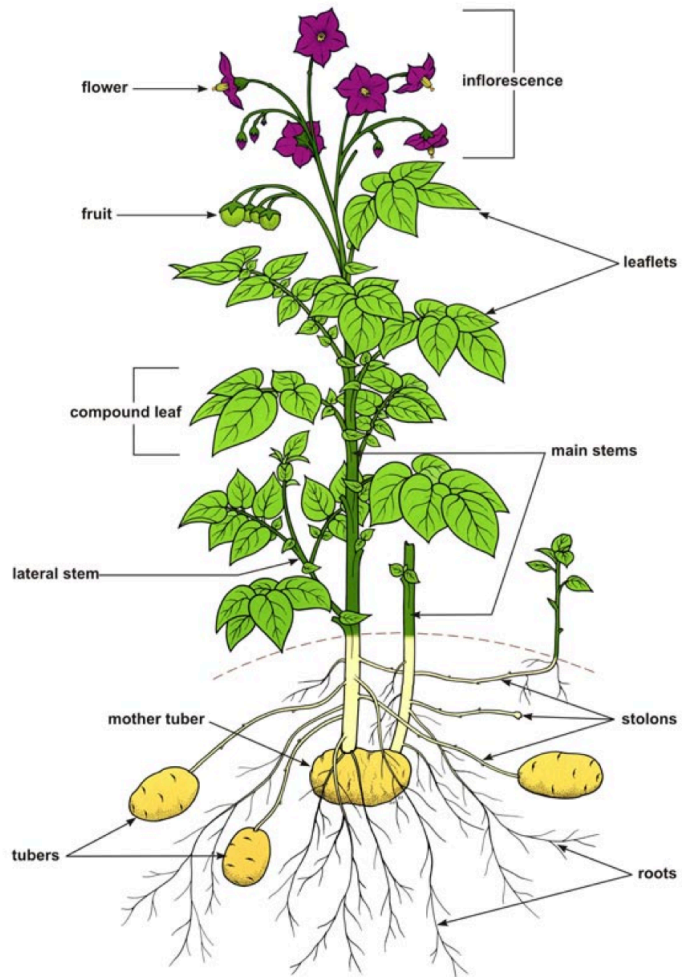
Most staple food is currently produced using modern, [conventional farming](#) practices. However, the production of staple food using [organic farming](#) methods is growing.

Ten staple foods of global importance (ranked by annual production)<sup>[12]</sup>

Rank <span>↕</span>	Crop <span>↕</span>	World production, 2012 <sup>[13]</sup>	Average world yield, 2010	World's most productive countries, 2012 <sup>[14]</sup>		World's largest producing countries, 2013 <sup>[16]</sup>	
		(metric tons) <span>↕</span>	(tons per hectare) <span>↕</span>	(tons per hectare) <span>↕</span>	Country <span>↕</span>	(metric tons) <span>↕</span>	Country <span>↕</span>
1	<a href="#">Maize (Corn)</a>	873 million	5.1	25.9	<a href="#">United States</a>	<b>354 million</b>	<a href="#">United States</a>
2	<a href="#">Rice</a>	738 million	4.3	9.5	<a href="#">Egypt</a>	204 million	<a href="#">China</a>
3	<a href="#">Wheat</a>	671 million	3.1	8.9	<a href="#">New Zealand</a>	122 million	<a href="#">China</a>
4	<a href="#">Potatoes</a>	365 million	<b>17.2</b>	45.4	<a href="#">Netherlands</a>	96 million	<a href="#">China</a>
5	<a href="#">Cassava</a>	269 million	12.5	34.8	<a href="#">Indonesia</a>	47 million	<a href="#">Nigeria</a>
6	<a href="#">Soybeans</a>	241 million	2.4	4.4	<a href="#">Egypt</a>	91 million	<a href="#">United States</a>
7	<a href="#">Sweet potatoes</a>	108 million	13.5	33.3	<a href="#">Senegal</a>	71 million	<a href="#">China</a>
8	<a href="#">Yams</a>	59.5 million	10.5	28.3	<a href="#">Colombia</a>	36 million	<a href="#">Nigeria</a>
9	<a href="#">Sorghum</a>	57.0 million	1.5	<b>86.7</b>	<a href="#">United States</a>	10 million	<a href="#">United States</a>
10	<a href="#">Plantain</a>	37.2 million	6.3	31.1	<a href="#">El Salvador</a>	9 million	<a href="#">Uganda</a>

玉米  
水稻  
小麦  
土豆  
木薯  
大豆  
甘薯  
山药  
高粱  
芭蕉

## The Potato Plant



International Potato Center (CIP)

名称：马铃薯

学名：Solanum tuberosum L.

科：茄科

多倍体：栽培种  $4n=48$

单倍体基因组大小 DM6.1：743M

原产地：南美洲安第斯山区

（秘鲁南部地区 / 的喀喀湖）



- 1950 The Green Revolution



- BMW

- 2021



- BMW



- Russet Burbank



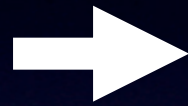
- Russet Burbank



# Hybrid potato-a new Green Revolution

## 同源四倍体

基因组杂合度高  
遗传分析复杂  
无法连续育种



## 二倍体

更快, 可以持续育种



## 无性繁殖

繁殖效率低(1:10-15)  
存储和运输效率低  
容易感染病虫害



## 种子繁殖

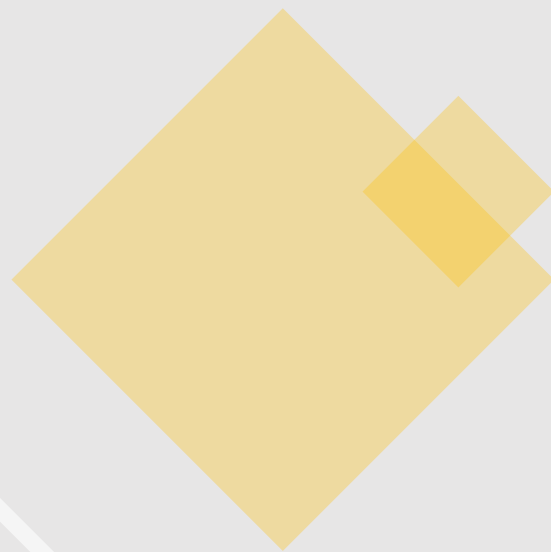
方便运输  
切断了薯块病虫害传播途径



1,000  
potato  
seeds

1,000  
sesame  
seeds

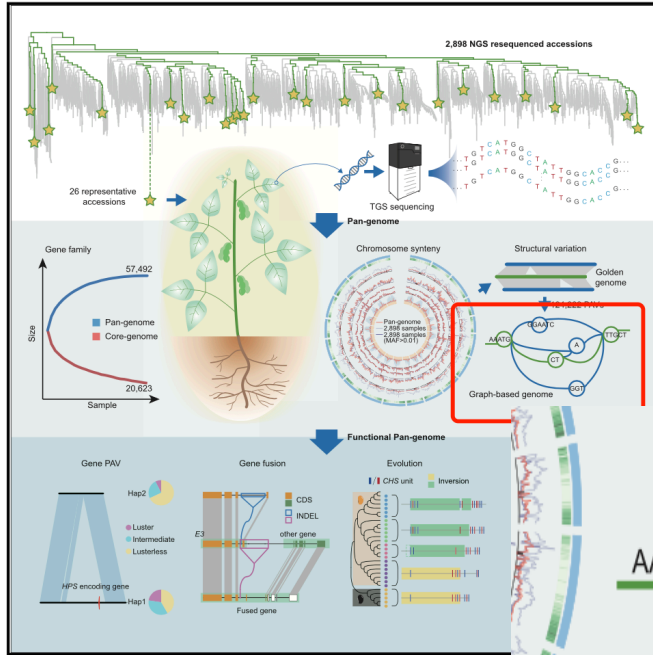
## 二、马铃薯图基因组



Cell

# Pan-Genome of Wild and Cultivated Soybeans

## Graphical Abstract



## Authors

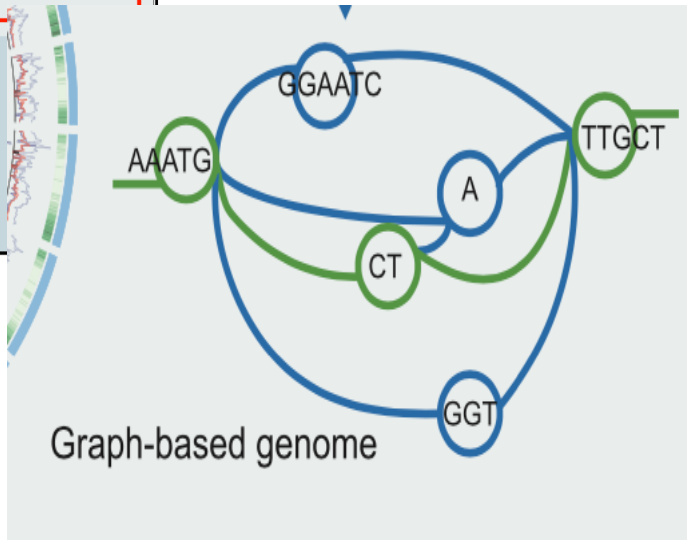
Yucheng Liu, Huilong Du, Pengcheng Li, ..., Bin Han, Chengzhi Liang, Zhixi Tian

## Correspondence

cliang@genetics.ac.cn (C.L.), zxtian@genetics.ac.cn (Z.T.)

## In Brief

A high-quality graph-based soybean pan-genome is constructed through *de novo* genome assemblies of 26 representative wild and cultivated soybean accessions,

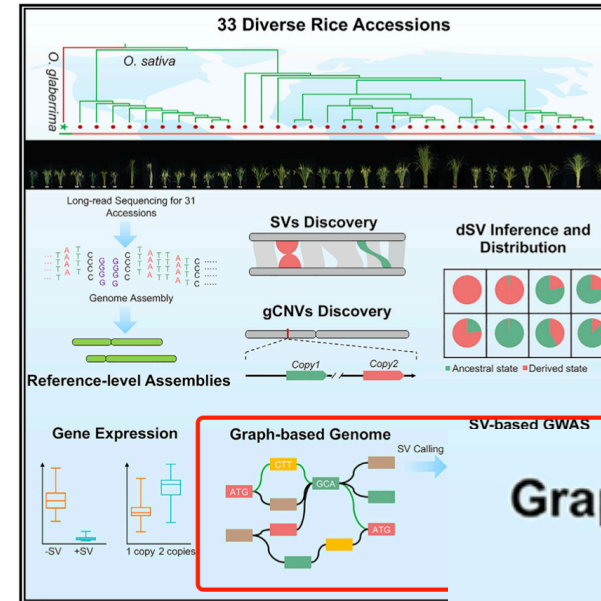


resource

Cell

# Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations

## Graphical abstract



## Authors

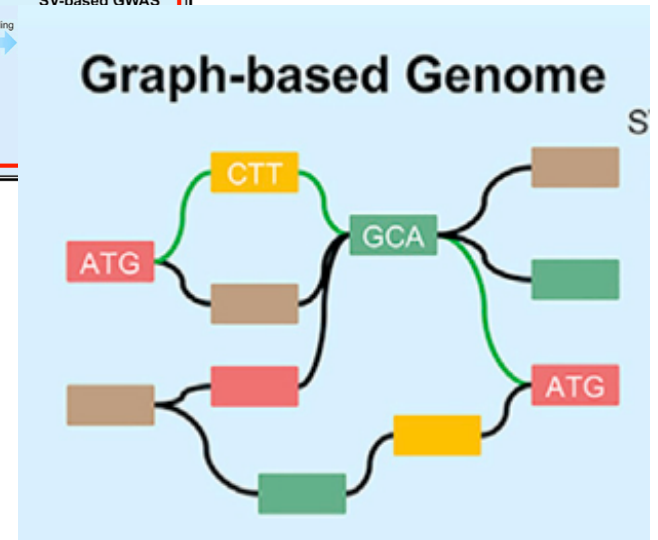
Peng Qin, Hongwei Lu, Huilong Du, ..., Xuewei Chen, Chengzhi Liang, Shigui Li

## Correspondence

qipeng@sicau.edu.cn (P.Q.), cliang@genetics.ac.cn (C.L.), lishigui@sicau.edu.cn (S.L.)

## In brief

A high-quality rice pan-genome of genetically diverse rice accessions is constructed through *de novo* genome assemblies, demonstrating the impact of structural variation and gene copy number variations on environmental



Resource

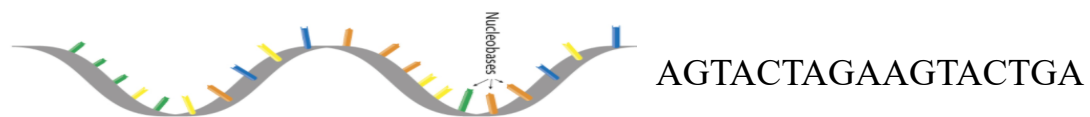




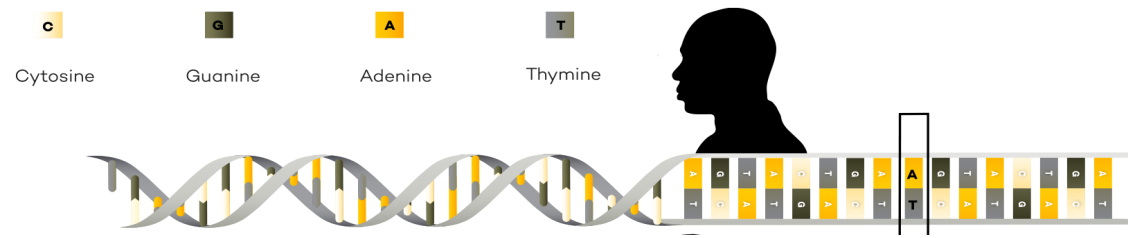
DNA

DNA是主要的遗传物质

Genome

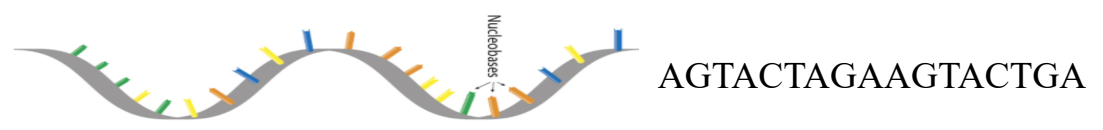


生物体的完整 DNA 集称为其基因组。



DNA

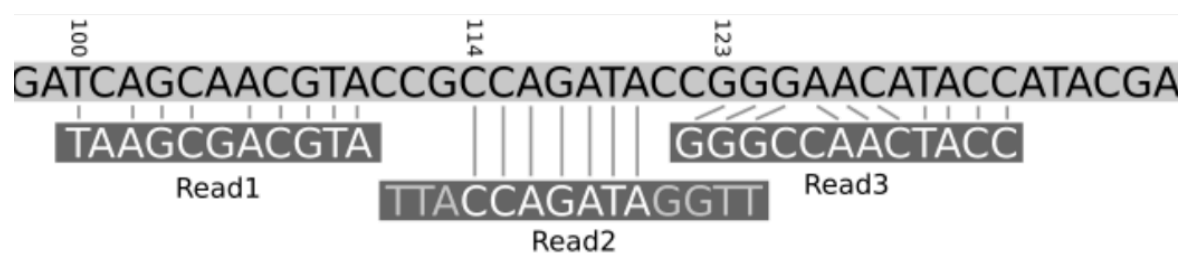
DNA是**主要的**遗传物质



Genome

生物体的完整 DNA 集称为其基因组。

Reference genome

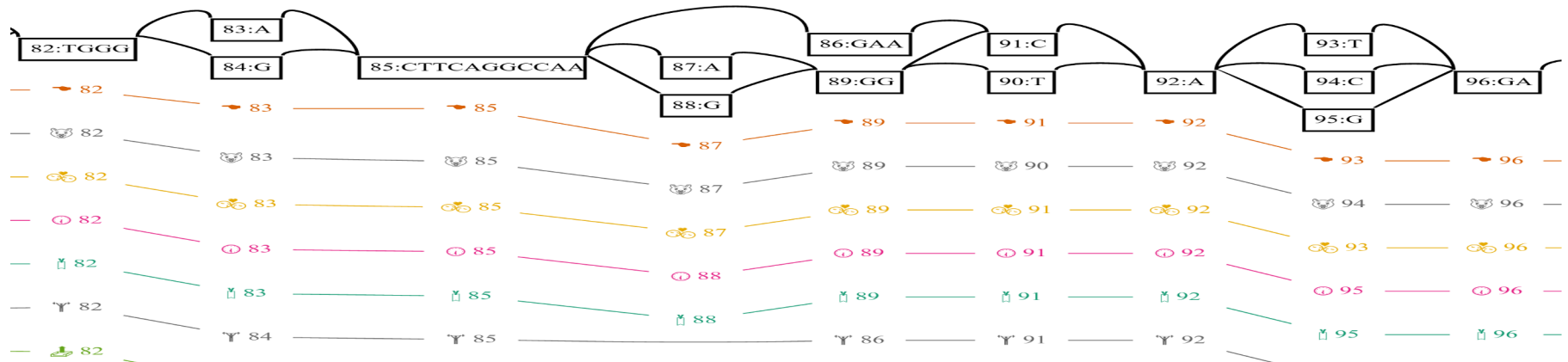
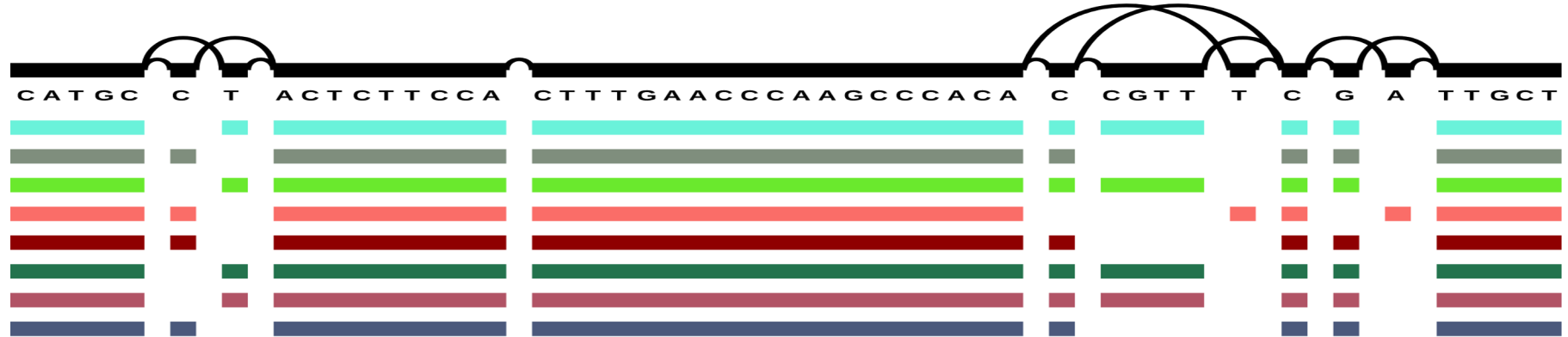
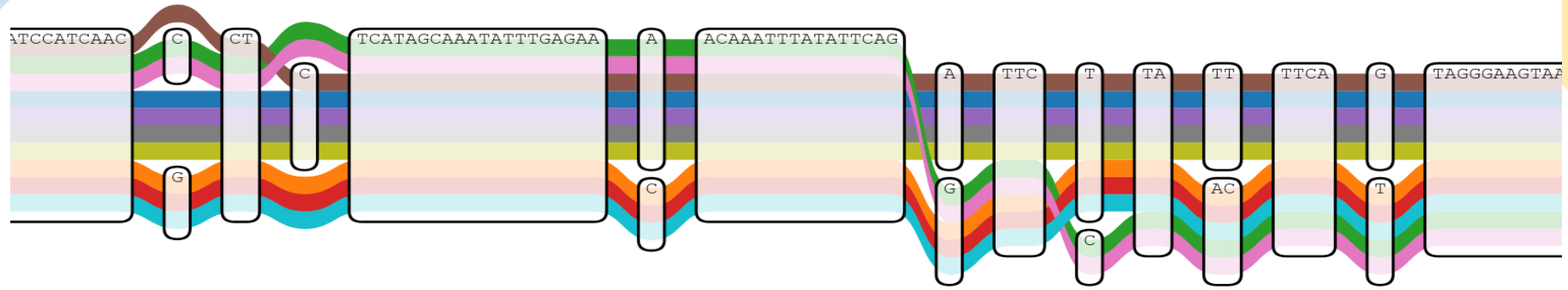


**正常**生物体的理想基因组

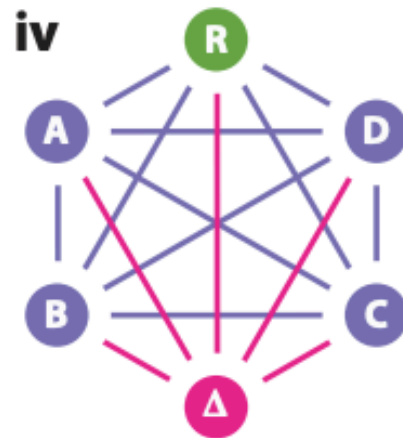
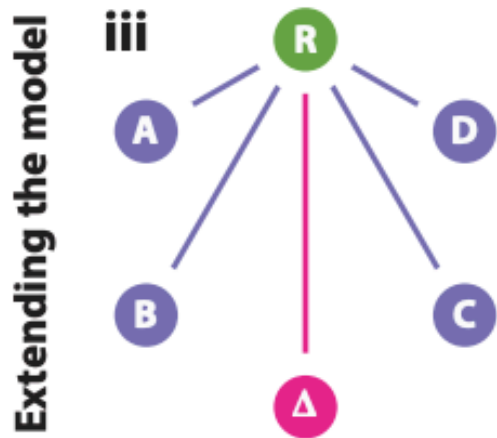
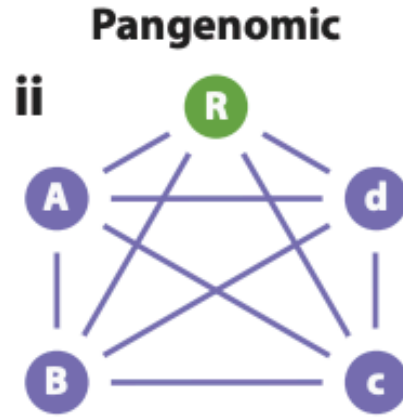
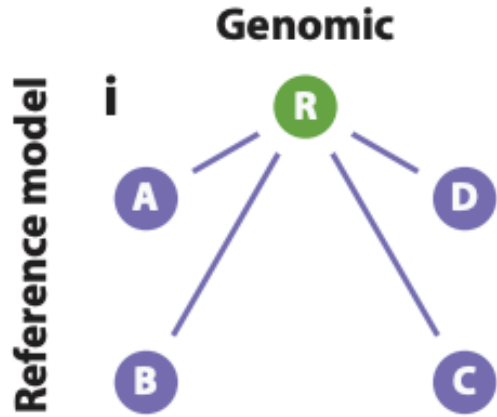




# 图基因组——代表性的不同基因组的集合

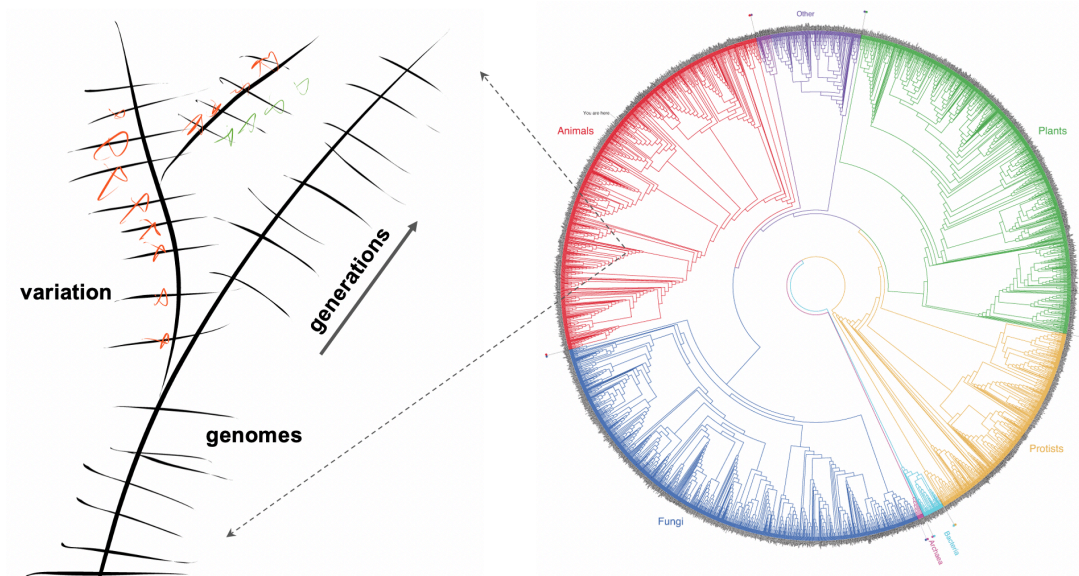


# Why graph genome ?



线性参考基因组存在的问题：

1. 线性参考基因组获取信息有限
2. 更新困难
3. 存在参考偏差 reference bias



Reference bias ? 参考偏差

Deletion

非洲人参考基因组

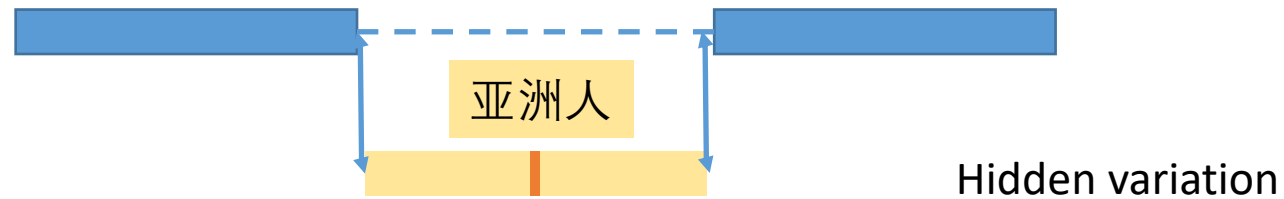


亚洲人



亚洲人 Reads

Graph genome



- Traditional linear referencing only represents one version of each locus,
- Reads are mapped as they similar enough to the liner reference genome



图基因组可以解决的问题：统一reference

物种	拉丁名	ID	参考基因组大小	基因数量
----	-----	----	---------	------



D1186

In addition, please cite the relevant publication(s) (listed on each genome's info page and in the Data Policy) to acquire from Phytozome.

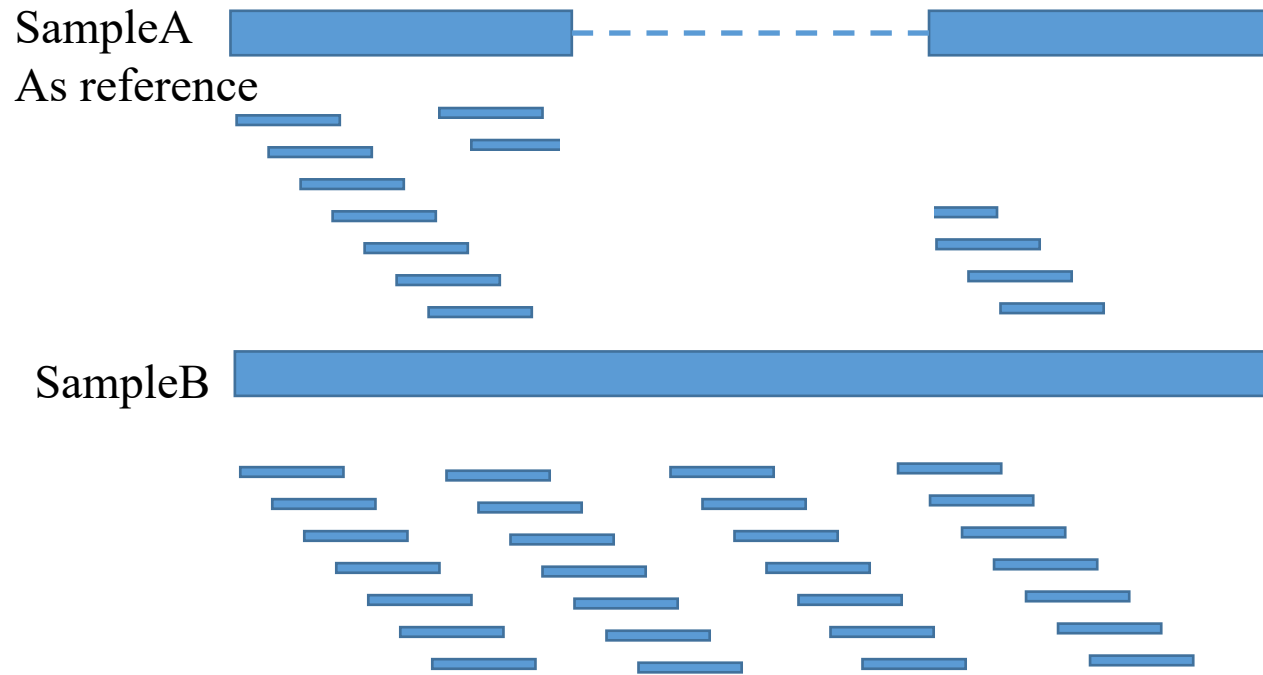
1.  0 genes

- Oryza sativa v7.0 – rice
- Oryza sativa Kitaake v3.1 – rice

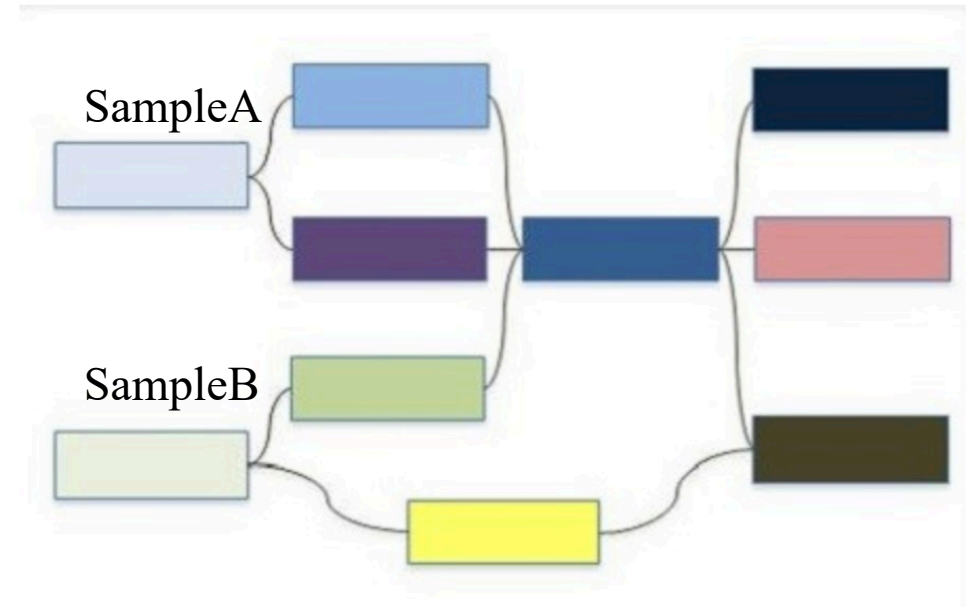
烟草	<i>N. tabacum</i>	Nitab4.5	~4Gb	69,500
----	-------------------	----------	------	--------

图基因组可以解决的问题 3) 使用二代数据精确检测结构变异

**Deletion 500bp (SV)**



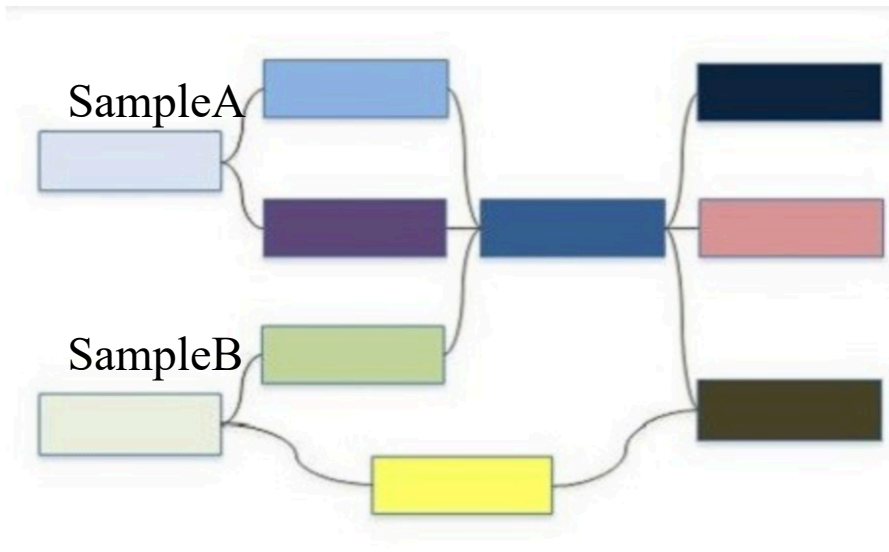
**Graph genomes**



Conclusion : Illumina short reads also can genotype high quality structural variation

# 构建图基因组的作用：

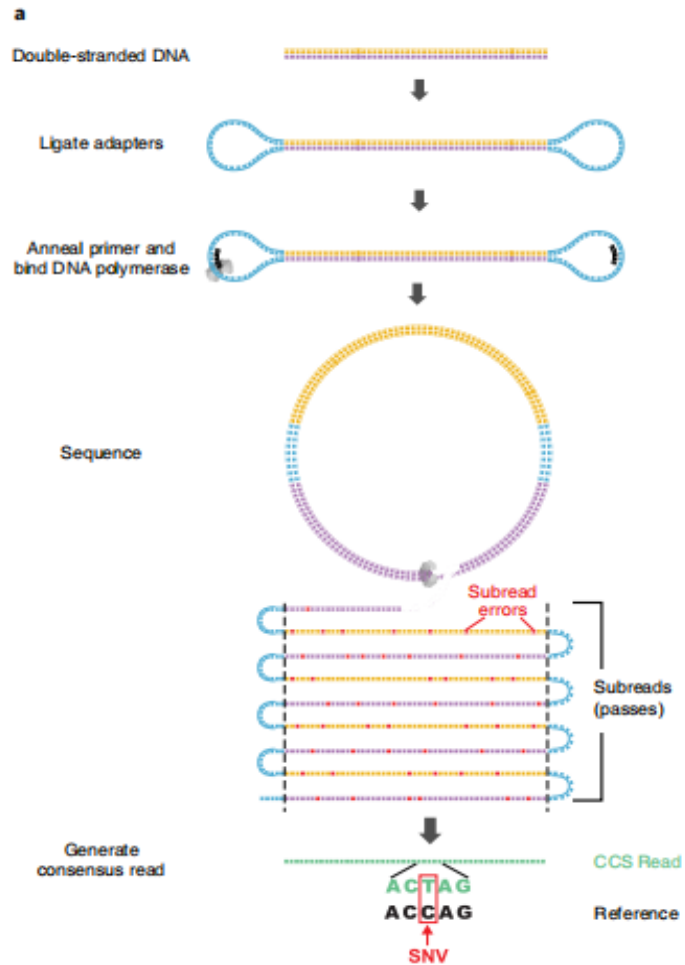
## Graph genomes



- ◆ 提供一个代表物种/群体的参考基因组
- ◆ 图基因组可以减少 **reference bias** 的错误，分析更准确，找到更全的变异信息
- ◆ 图基因组 可以用**二代测序数据来找结构变异**，这对育种来说是一块很大且未被发掘的资源。



# 图基因组构建----选样,测序 CCS + HIC



Article | [Published: 12 August 2019](#)

## Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome

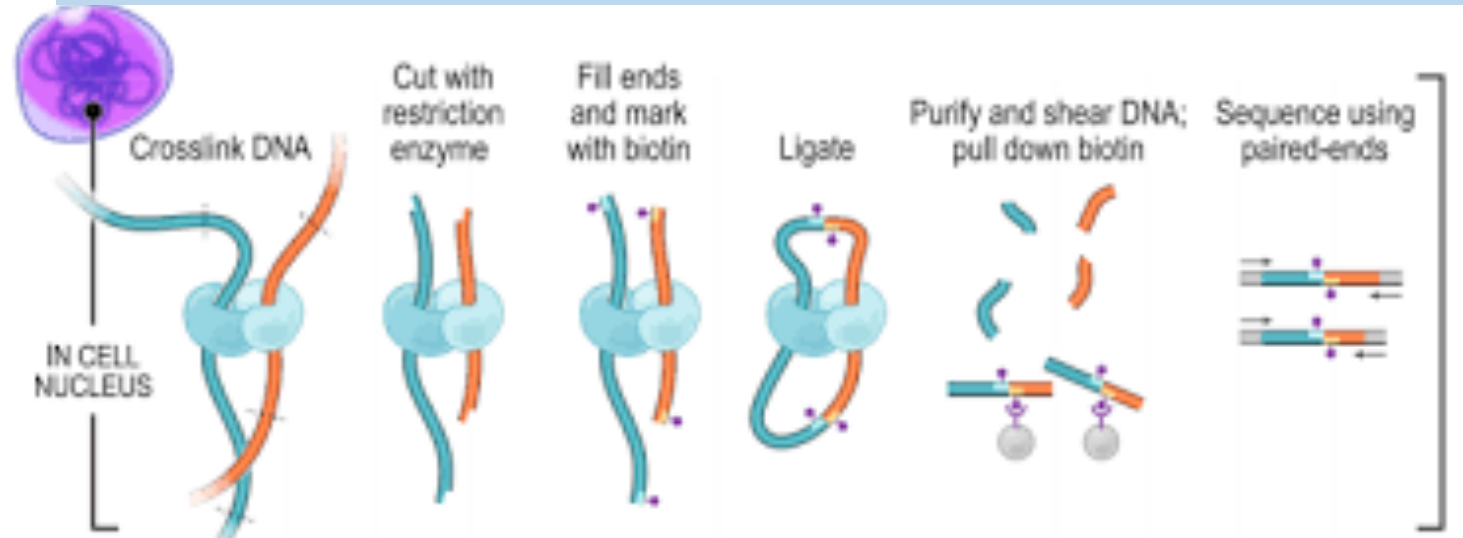
[Aaron M. Wenger](#), [Paul Peluso](#), ... [Michael W. Hunkapiller](#)

[+ Show authors](#)

[Nature Biotechnology](#) **37**, 1155–1162 (2019) | [Cite this article](#)

23k Accesses | 260 Citations | 157 Altmetric | [Metrics](#)

### HIC



HIFI (CCS)

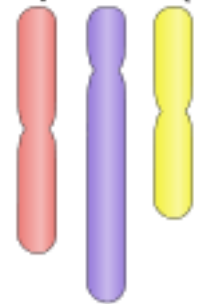
Long reads 15K

High accuracy 99.9%

- long reads( average 15 k)
- highly accurate (99.9%)
- Haplotype resolved genome

图基因组构建----组装(Haplotype-resolved)

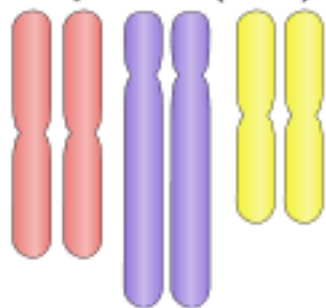
Haploid (N)



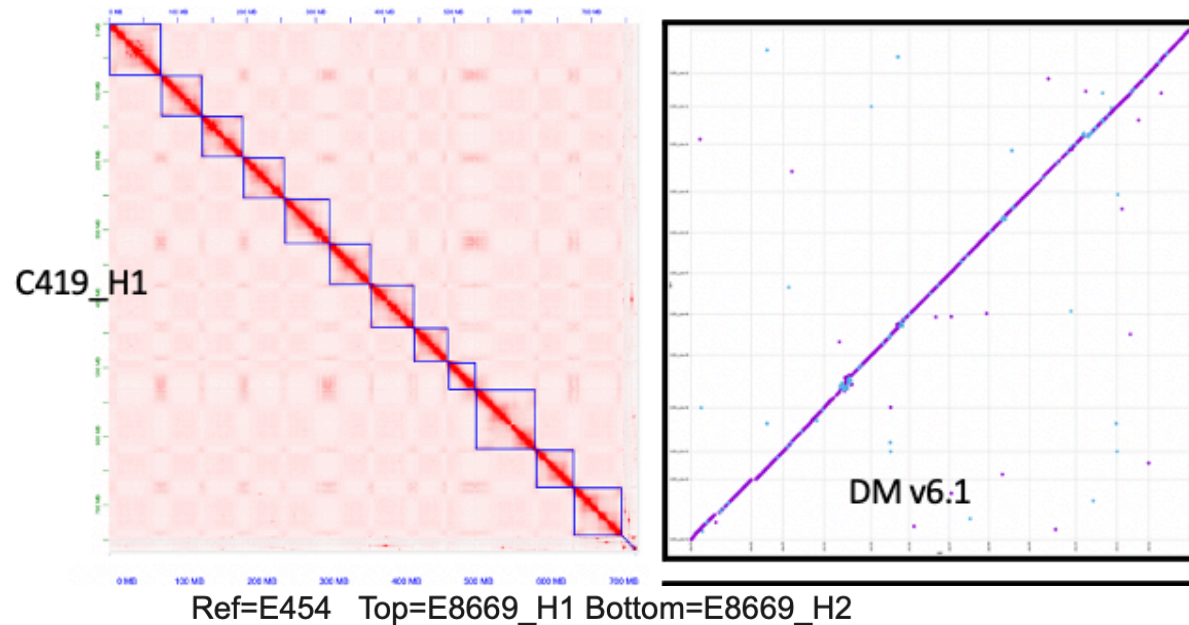
Reference genome

Potato 12

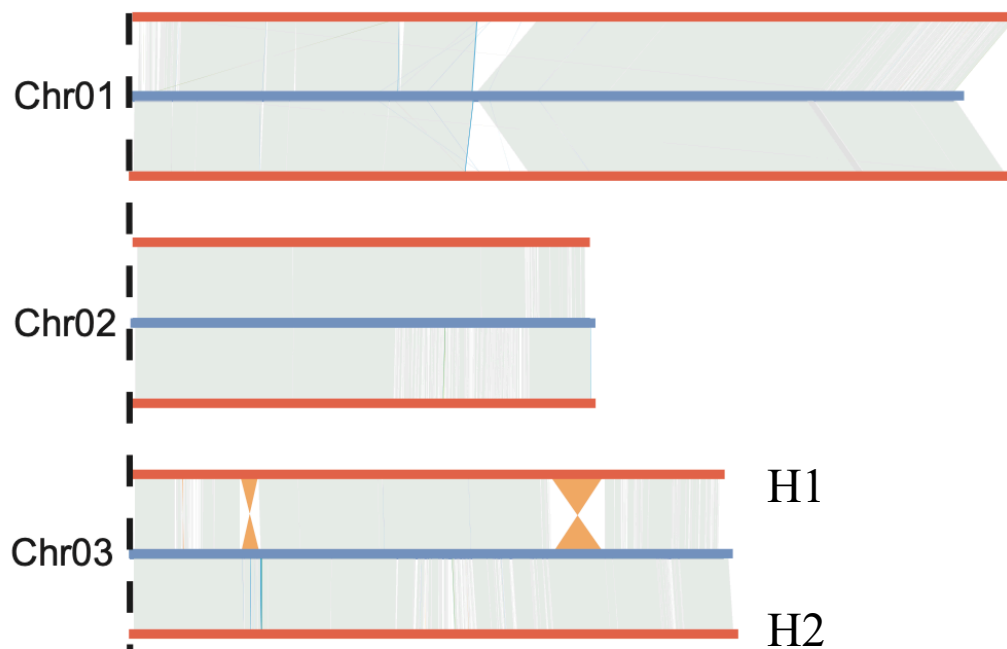
Diploid (2N)



Potato 12 + 12

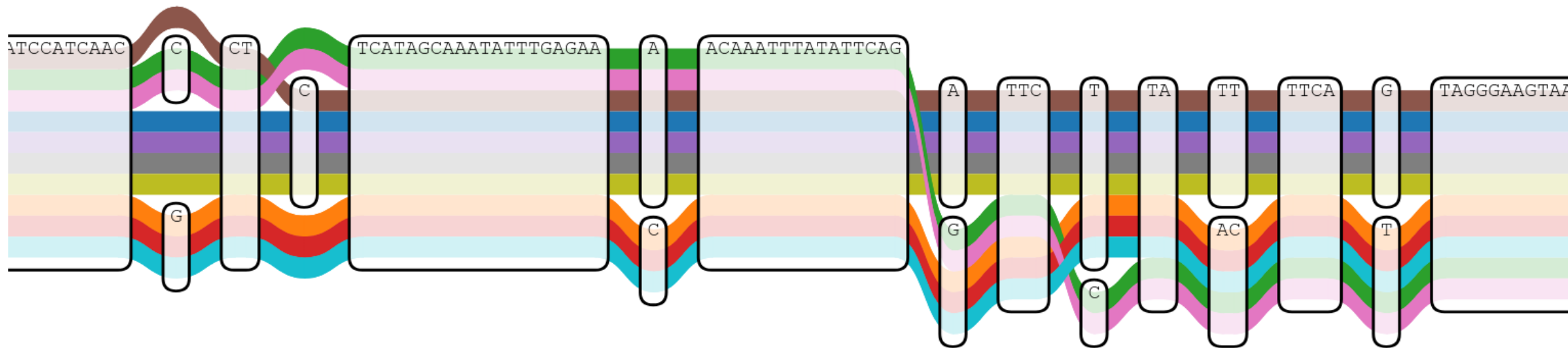


Reference Syntenic Translocation  
Query Inversion Duplication



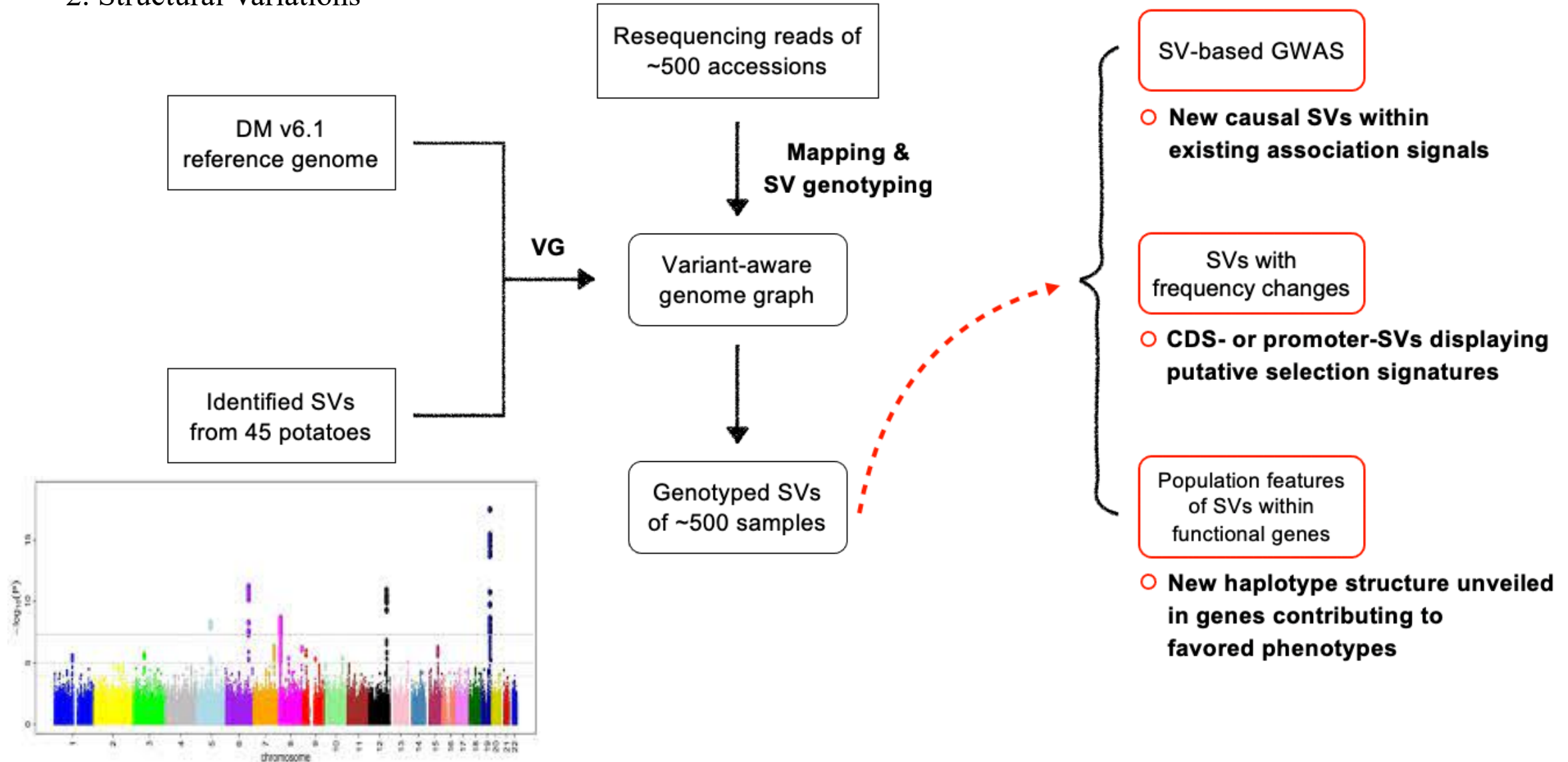
# 图基因组构建----基因组整合

Haplotype 1 **CAAATAAG**GCTTGGAAATTTACCCGCTCCTGCCCGCGTCTGGAGTTC**ACCCGCTCCTGCCCGCG**TATTATATT**C**CAACTCTCTG  
Haplotype 2 **CAAATAAG****C**GCTTGGAAATTTACCCGCTCCTGCCCGCGTCTGGAGTTC-----TATTATATT**C**CAACTCTCTG  
Haplotype 3 **CAAATAAG**GCTTGGAAATTTACCCGCTCCTGCCCGCGTCTGGAGTTC-----TATTATATT**A**CAACTCTCTG



# 图基因组的应用----GWAS

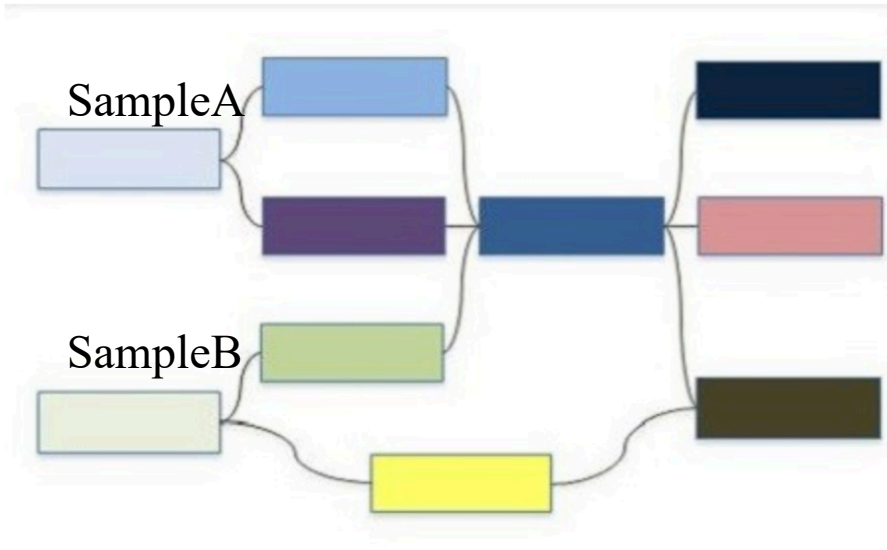
1. 更多更准确的SNP, InDels
2. Structural Variations





# 构建图基因组的作用：

## Graph genomes



- ◆ 提供一个代表物种/群体的**参考基因组**
- ◆ 图基因组可以减少 **reference bias** 的错误，分析更准确，找到更全的变异信息
- ◆ 图基因组 可以用**二代测序数据来找结构变异**，这对育种来说是一块很大且未被发掘的资源。



# Thanks

---

Any questions?



程林

chenglin\_solab@163.com