



浙江大学  
ZHEJIANG UNIVERSITY

# 猪基因组选择技术及其应用

潘玉春

2020.11.10

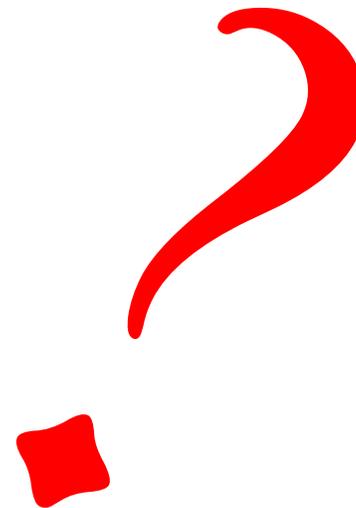
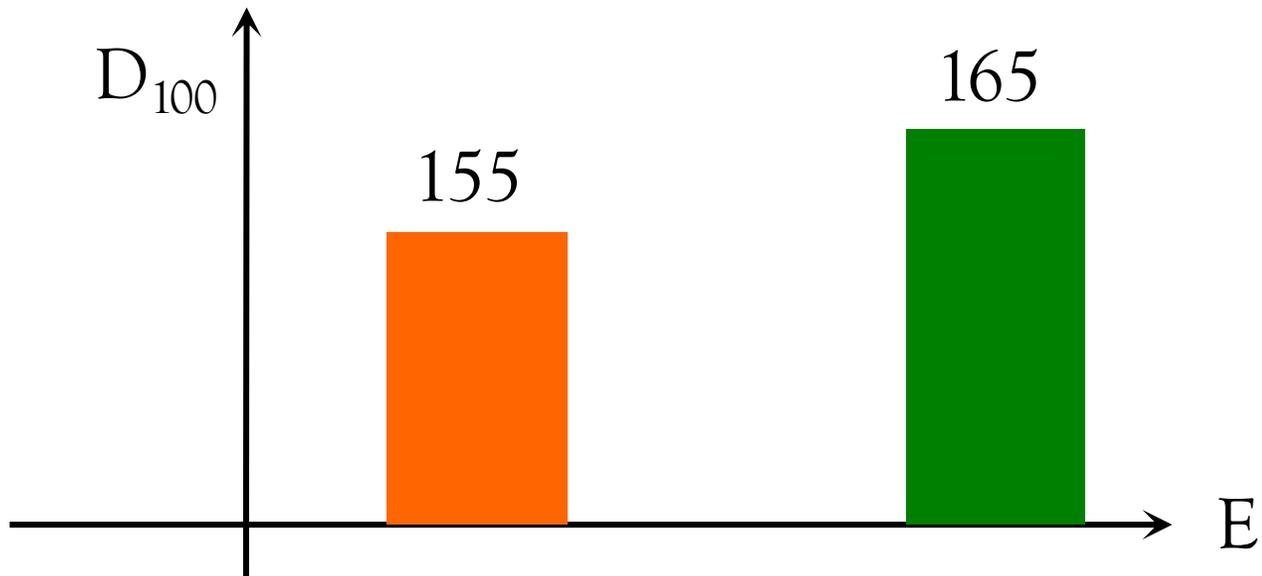


# □ 常规育种技术

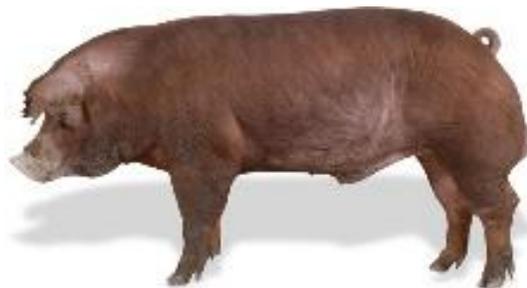
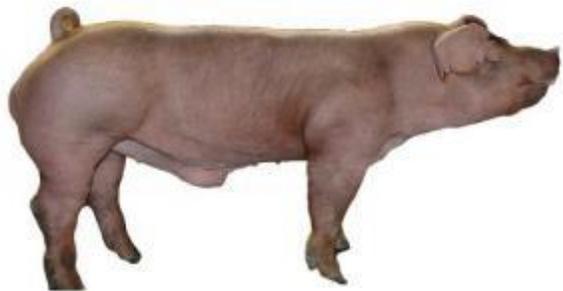


表型(P) = 遗传(G)+环境(E)+互作(G×E)

$$P = A + D + I + E + G \times E$$



# 测定站测定



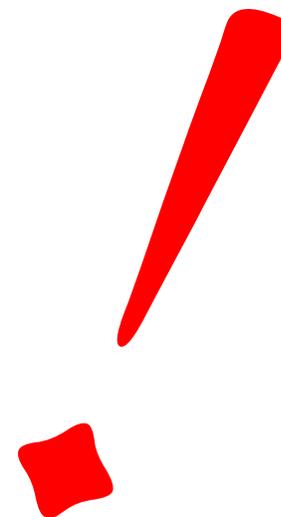
## ➤优点

- 能严格控制统一规范的测定条件；
- 利于比较不同环境的猪；
- 具有准确、公正、**权威**的特点。

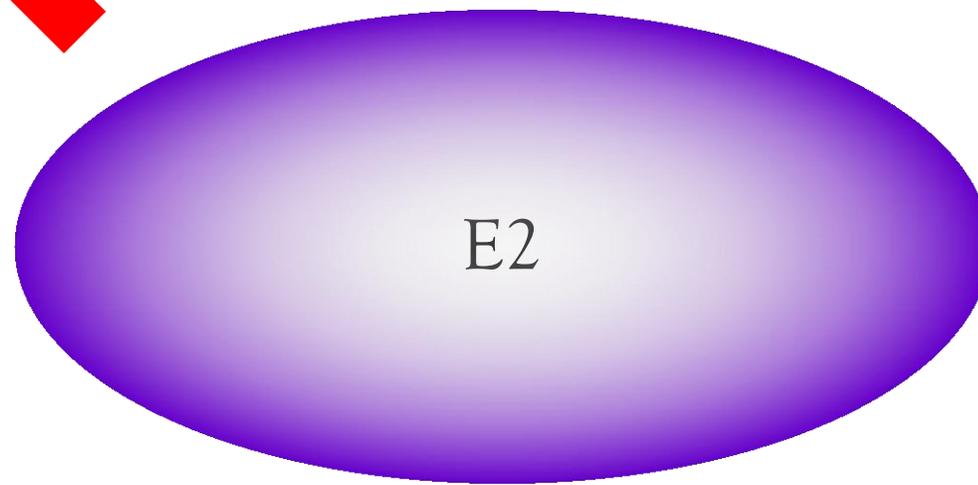
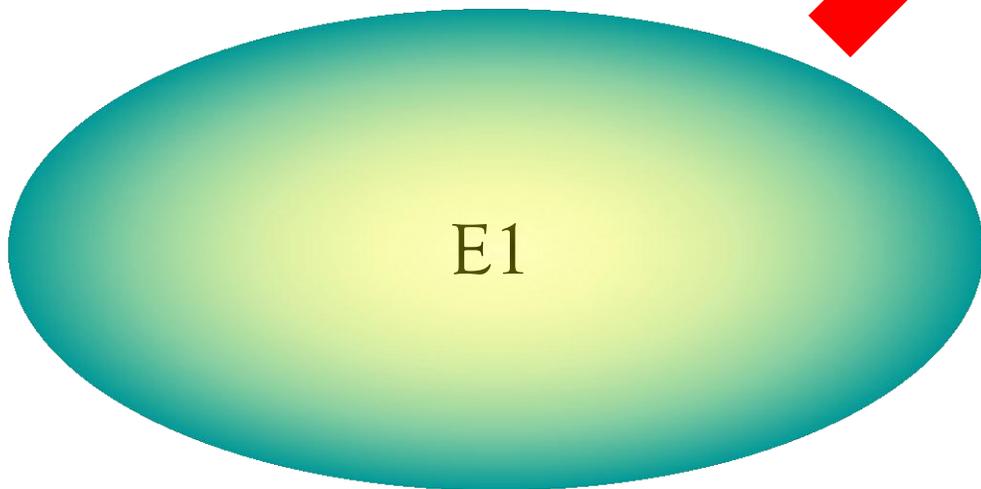
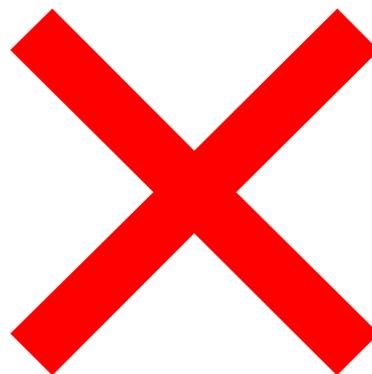
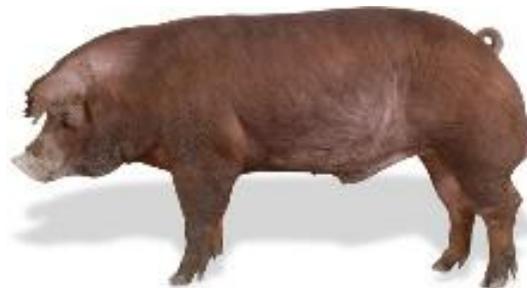
国外已不常用或者改变用途  
(测定难以测定的性状)

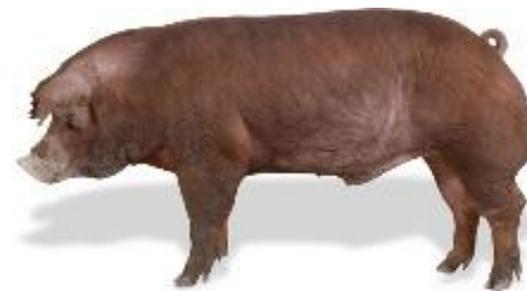
## ➤缺点

- 测定费用昂贵、测定规模有限，容易对猪产生应激；
- 增大场与场间疫病交叉传染的可能性；
- 需要考虑测定前不同环境的影响等。



# 场内测定





遗传联系

E1

E2

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{X}\mathbf{b} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; V \begin{bmatrix} \mathbf{V} & \mathbf{Z}\mathbf{A}\sigma_a^2 & \mathbf{I}\sigma_e^2 \\ & \mathbf{A}\sigma_a^2 & 0 \\ \text{sym.} & & \mathbf{I}\sigma_e^2 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + k\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Best Linear Unbiased Prediction (BLUP)

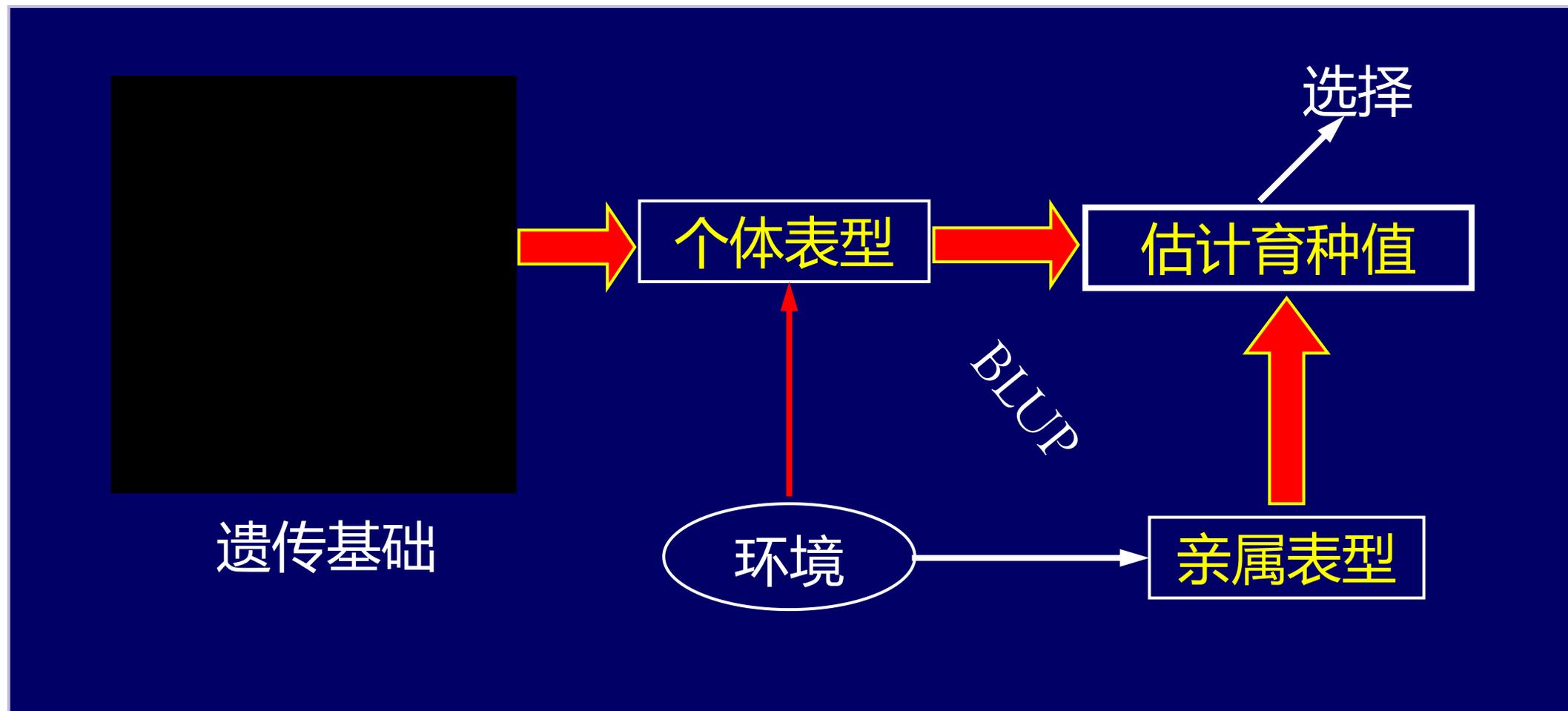
## 【案例】两个猪场通过公猪建立遗传联系，联合进行遗传评估

| 母猪 | 父亲 | 母亲 | 猪场 | 年份 | 场-年 | 产仔 | EBV   |
|----|----|----|----|----|-----|----|-------|
| 3  | 1  | —  | 1  | 1  | 1   | 10 | 1.11  |
| 4  | 2  | —  | 1  | 1  | 1   | 6  | -0.82 |
| 5  | 1  | 4  | 1  | 2  | 2   | 7  | -0.37 |
| 6  | 2  | 3  | 1  | 2  | 2   | 10 | 0.91  |
| 7  | 1  | 6  | 1  | 2  | 2   | 5  | -0.46 |
| 8  | 2  | —  | 2  | 1  | 3   | 8  | -0.63 |
| 9  | 2  | —  | 2  | 2  | 4   | 13 | 0.35  |
| 10 | 1  | 8  | 2  | 1  | 3   | 12 | 0.43  |
| 11 | 1  | 9  | 2  | 2  | 4   | 10 | -0.38 |

遗传力为0.1

|      |       |             |
|------|-------|-------------|
| hy11 | 7.86  | } 场年<br>效应  |
| hy12 | 7.31  |             |
| hy21 | 10.10 |             |
| hy22 | 11.52 |             |
| 1    | 0.02  | } 公猪<br>EBV |
| 2    | -0.15 |             |

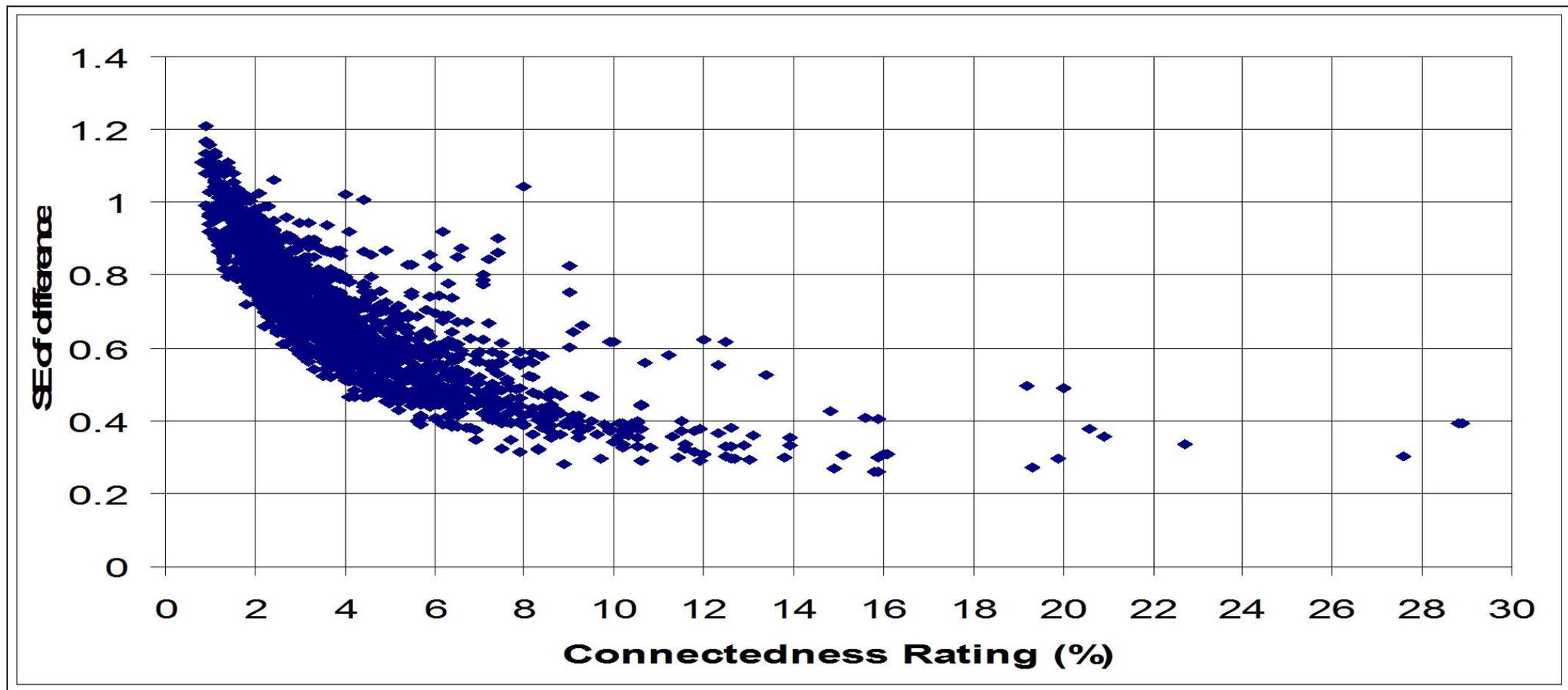
# 常规育种流程



- **优点：**可降低测定费用、扩大测定规模；避免仔猪应激和疾病的广泛传播

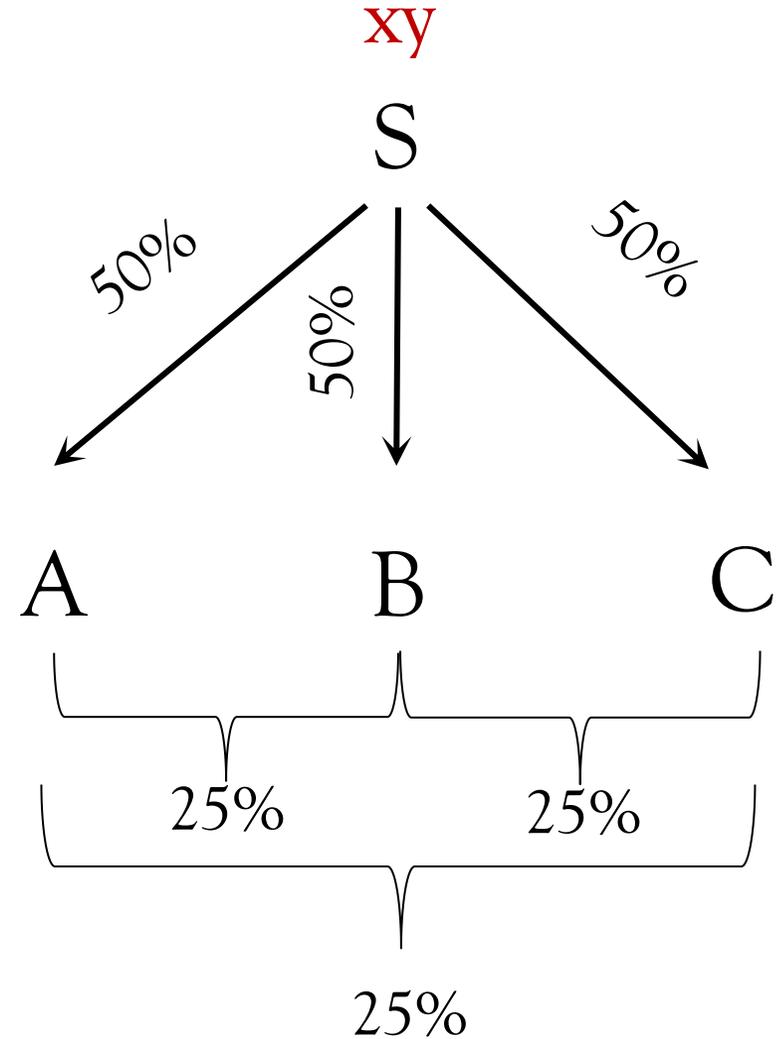
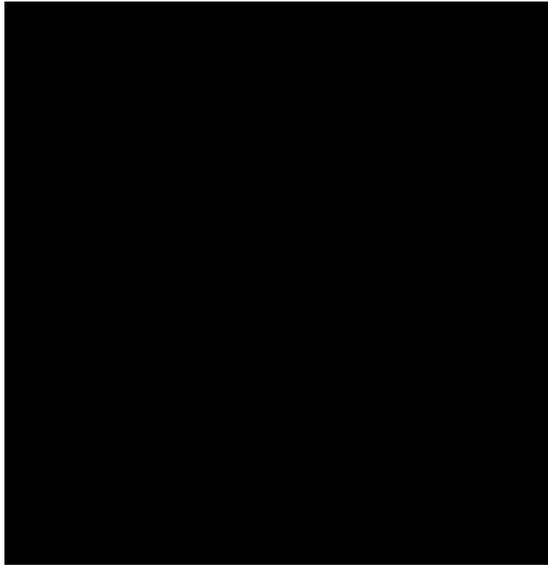
## ➤ 缺点

- 场间必须存在一定关联



关联率与场间比较误差间的关系（引自张勤）

– 将个体遗传基础视为黑箱，个体间的遗传关系只是概率估计而非实际状态



- 依赖大规模且持续性的表型测定

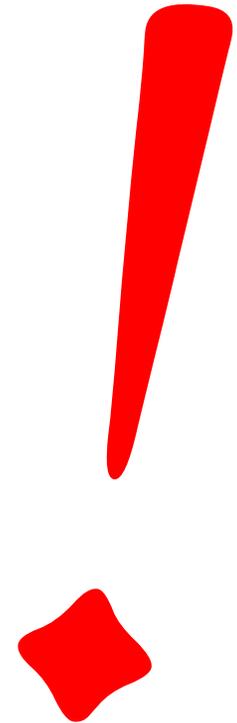
- 体型外貌；繁殖性状；
- 生长发育（含饲料转化效率）；
- 屠宰与肉质；健康性状、……

许多性状  
测定困难

- 统计分析平台必须完善

- 有的性状育种效率很低

- 低遗传力性状（如繁殖性状）
- 限性性状（繁殖性状）
- 难以度量的性状（饲料转化效率、胴体与肉质、健康性状）
- 无法早期度量的性状（如使用寿命）

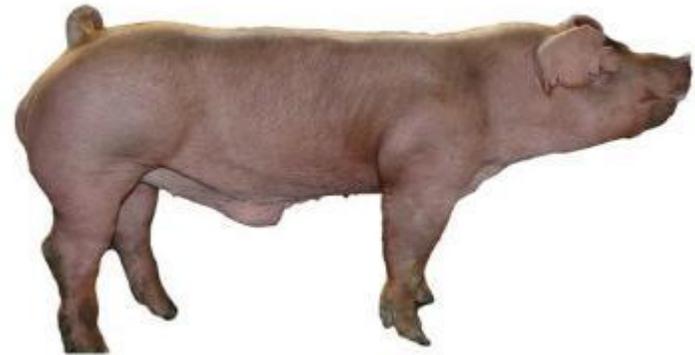


# 全基因组选择

战前  
演习档案  
派兵布阵  
战场环境  
参考群体  
Reference  
Population

预测

评估每一头猪对下一代种用价值如何，仿佛战前评估一种作战方案胜算如何！



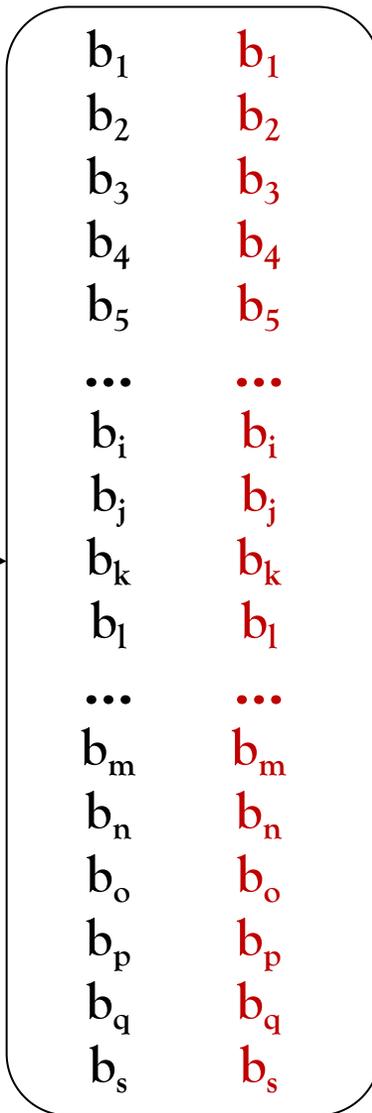
遗传物质

A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
...  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O

约28  
亿个  
碱基  
位点

每个  
位点  
5种  
可能

精子 卵子



个体

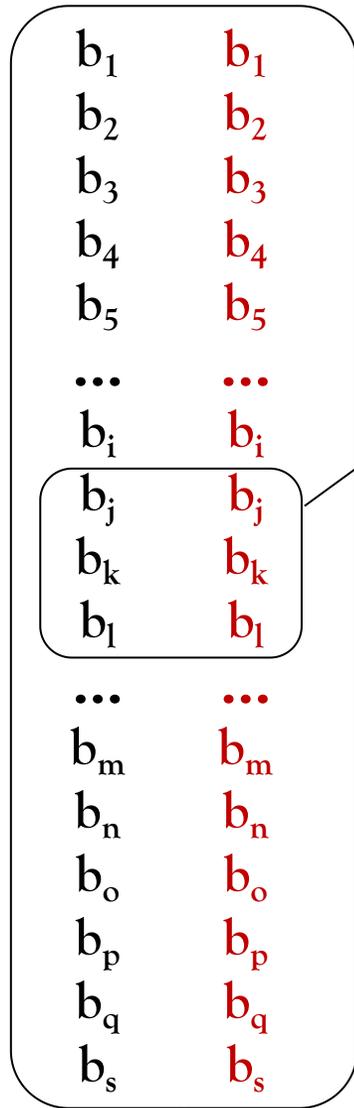
遗传物质

A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
...  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
...  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O  
A T G C O

约28  
亿个  
碱基  
位点

每个  
位点  
5种  
可能

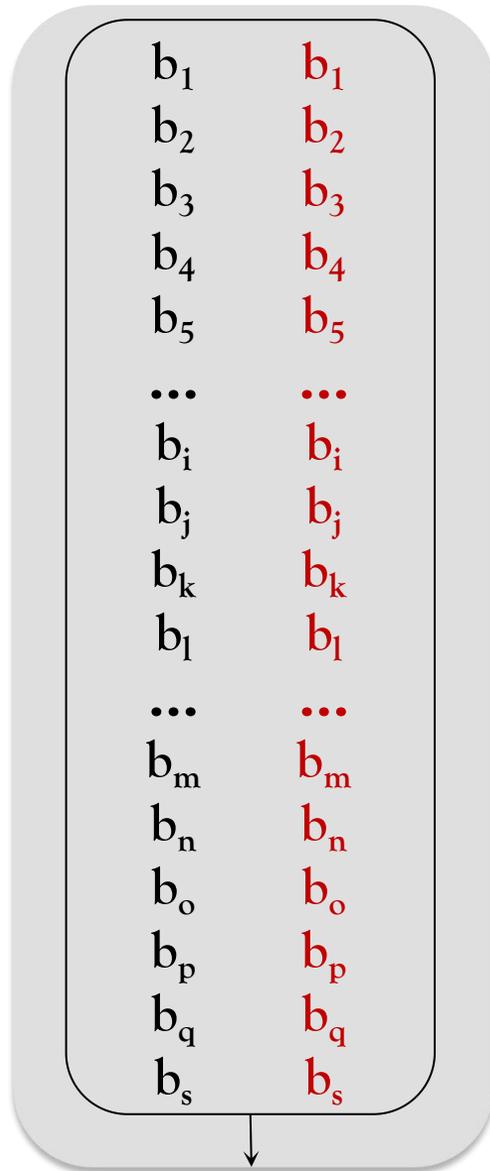
## 精子 卵子



个体

- 班组=连锁不平衡区 ( LD ) :  $i$ 、 $j$ 、 $k$ 三个岗位职责 ( 影响性状 ) 可能不同 , 但在排兵布阵时却被作为一个班组—张三、李四、王二麻子可能总在一起。
- 知道任意一人出现 , 即可推测班组中责任人的战力。
- 配子内同一LD内的碱基组合→单倍体型 (  $h$  )
- 杜、长、大等引进品种的基因组的LD平均约400kb , 故大约有7000个LD

## 精子 卵子



个体

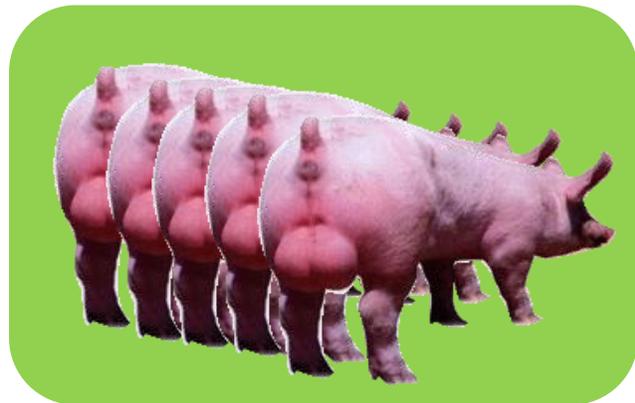
- 通过大量演习知道各种环境之下每个性状每个相关位点或其连锁位点（标记） $b_t$  每种情况（**A T G C O**）时的战力（效应） $g_t$ ，进而建立一个完备的参考群体（演习档案）
- 那么，再有一个个体，只要知道它的排兵布阵（**碱基位点或其标记状态**，**A T G C O**）或与前期演习的相似程度，就能知道其整体的性能！

Meuwissen et al., ( 2001 )

Genome-wide Selection ( GWS )

Genomic Selection ( GS )

参考群体



生产性能测定

遗传变异检测

参考数据  
Training  
Data

候选群体



遗传变异检测

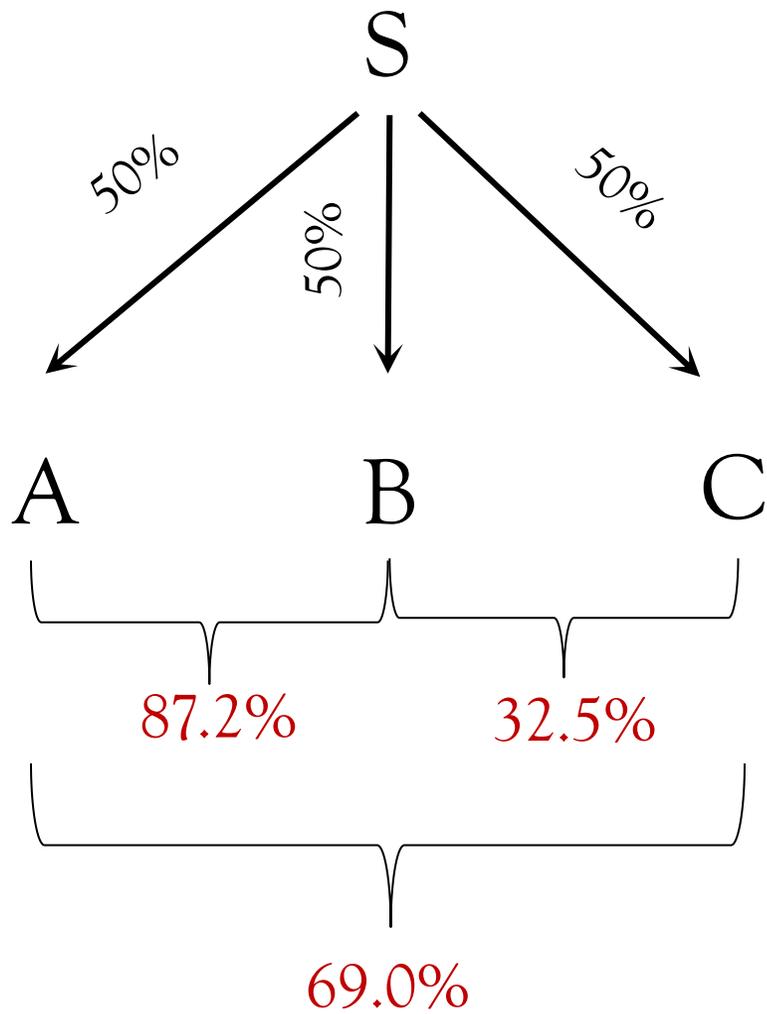
遗传评估

GS基本流程：两大步骤；三大环节

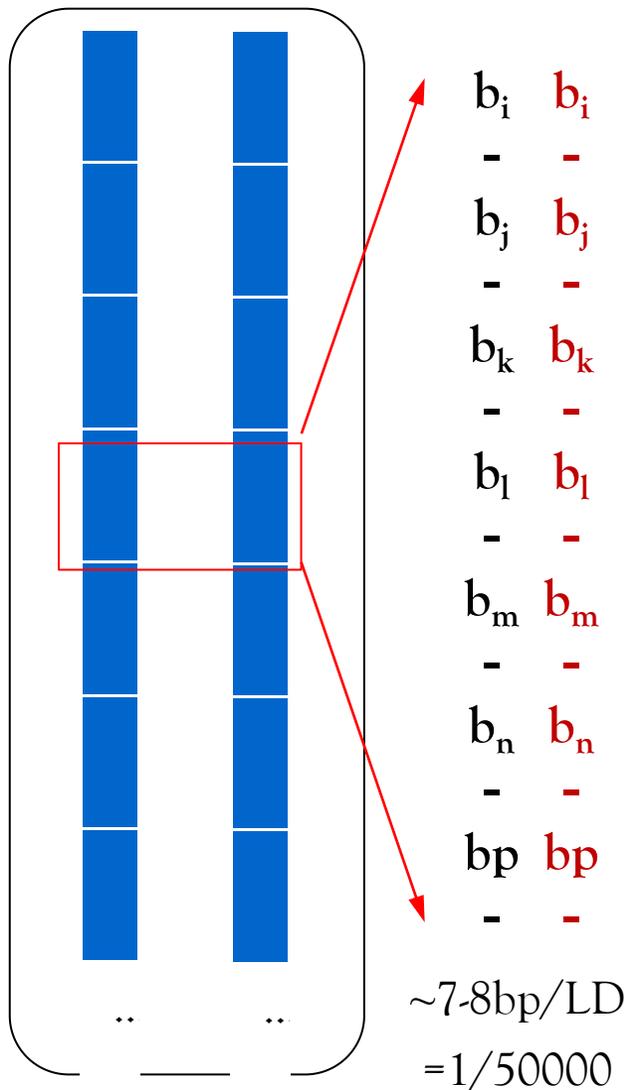


# 遗传变异检测

遗传基础



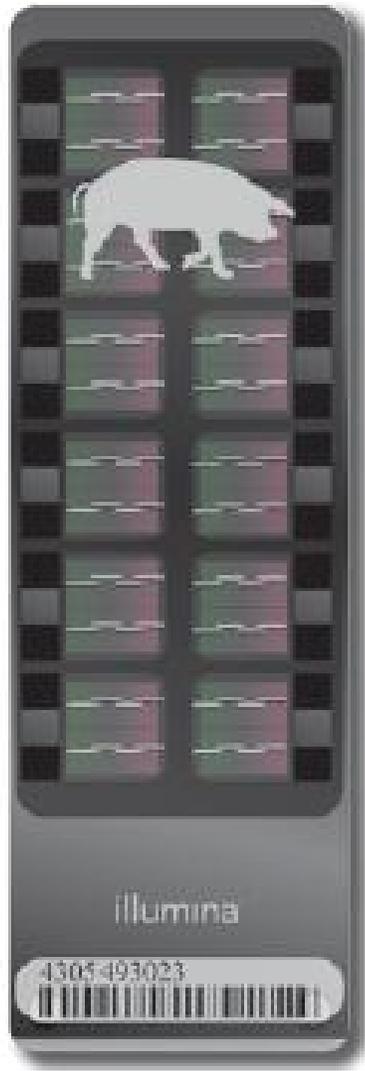
~7000个LD, 400kb/LD



# SNP检测芯片

## Blotter Card

Complete the space on the blotter card for Owner, Breed, Ear Notch and Registration Number immediately to assure that sample is properly identified. Complete the breeder information on the NSR Testing Form with the sample ID and Breed, and indicate the blood banking and/or diagnostic tests required for each corresponding sample.



## 现有芯片

| Chip Name   | No. of SNPs | Company         | Technology                  |
|---|-------------|-----------------|-----------------------------|
| PorcineSNP60 BeadChip v2 array                            | 61,565      | Illumina        | Illumina Infinium Chemistry |
| GeneSeek Genomic Profiler for Porcine LD (GGP-Porcine LD) | 10,241      | GeneSeek/Neogen | Illumina Infinium Chemistry |
| GeneSeek Genomic Profiler for Porcine HD (GGP-Porcine HD) | 68,528      | GeneSeek/Neogen | Illumina Infinium Chemistry |
| GGP Porcine BeadChip                                      | 51,000      | GeneSeek/Neogen |                             |
| Compass porcienSNP55                                      | 55,000      | 康普森             |                             |
| WensSNP55   | 55,000      | 温氏              |                             |
| Axiom Genome-Wide Pig genotyping Array                    | ~650,000    | Affymetrix      | Axiom assay                 |



中国畜牧兽医学动物遗传育种学分会第九次全国会员代表大会  
**第十九次全国动物遗传育种学术讨论会**  
19th Chinese National Conference on Animal Breeding and Genetics

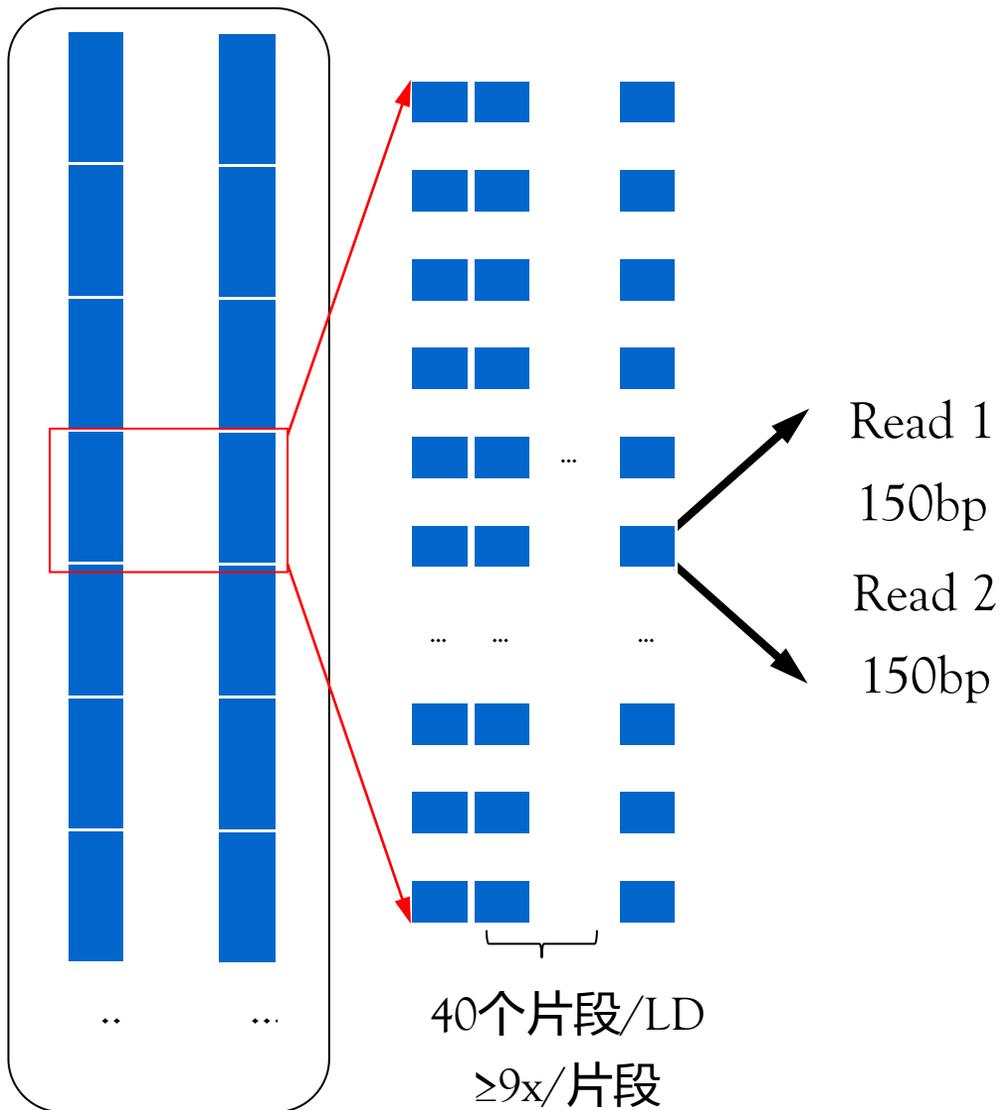
# Omics 时代的动物遗传育种

主办单位：中国畜牧兽医学动物遗传育种学分会  
承办单位：南京农业大学 扬州大学



 **中芯号** 

~7000个LD, 400kb/LD



## 简化基因组测序

28万 片段

56万 reads

8400万 bp

Coverage rate

≈3%

Illumina Hiseq x10

## 基因组重测序

### 遗传基础

|       |       |
|-------|-------|
| $b_1$ | $b_1$ |
| $b_2$ | $b_2$ |
| $b_3$ | $b_3$ |
| $b_4$ | $b_4$ |
| $b_5$ | $b_5$ |
| ...   | ...   |
| $b_i$ | $b_i$ |
| $b_j$ | $b_j$ |
| $b_k$ | $b_k$ |
| $b_l$ | $b_l$ |
| ...   | ...   |
| $b_m$ | $b_m$ |
| $b_n$ | $b_n$ |
| $b_o$ | $b_o$ |
| $b_p$ | $b_p$ |
| $b_q$ | $b_q$ |
| $b_s$ | $b_s$ |

一个人的  
测序  
成本已  
由30亿  
美元降  
到100  
美元



- 自从Meuwissen (2001) 首次提出以来，已有许多统计分析方法产生：
  - 基于Bayes理论的方法  
Bayes A、 Bayes B、 Bayes C $\pi$ 、 Bayes R.....
  - 基于BLUP理论的方法  
RR-BLUP、 GBLUP、 ssBLUP ( single-step GBLUP ) .....
- 没有哪种方法能在所有情况下都具有绝对优势，但目前应用较多的是GBLUP及在其基础上改进的ssBLUP。

➤ GBLUP (VanRaden and Tooker, 2007)

- BLUP with genomic relationship matrix (G matrix)

Phenotype of reference animals

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

GBV of all genotyped animals

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{Xb} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; V \begin{bmatrix} \mathbf{V} & \mathbf{ZG}\sigma_a^2 & \mathbf{I}\sigma_e^2 \\ \mathbf{G}\sigma_a^2 & \mathbf{0} & \mathbf{0} \\ \text{sym.} & & \mathbf{I}\sigma_e^2 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + k\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

GEBV

## ➤ Single-step BLUP (SSBLUP)

- Use individuals **with pedigree and phenotype** but **without genotypes**

Phenotype of all animals

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

GBV of all animals

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{Xb} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; V \begin{bmatrix} \mathbf{V} & \mathbf{ZH}\sigma_a^2 & \mathbf{I}\sigma_e^2 \\ & \mathbf{H}\sigma_a^2 & \mathbf{0} \\ \text{sym.} & & \mathbf{I}\sigma_e^2 \end{bmatrix}$$

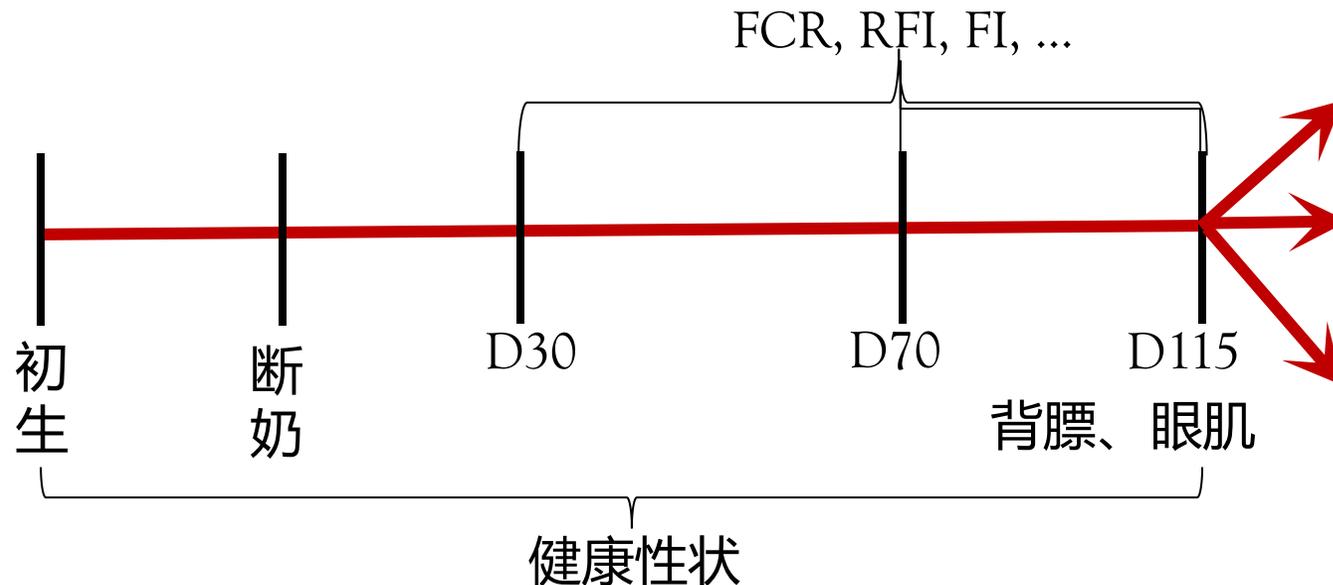
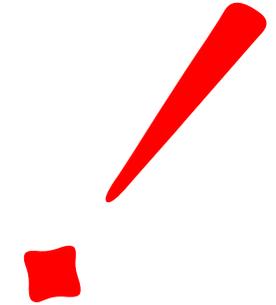
$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + k\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

GEBV

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}_{\omega}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} + \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}_{\omega} \\ \mathbf{G}_{\omega}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G}_{\omega} \end{bmatrix}$$

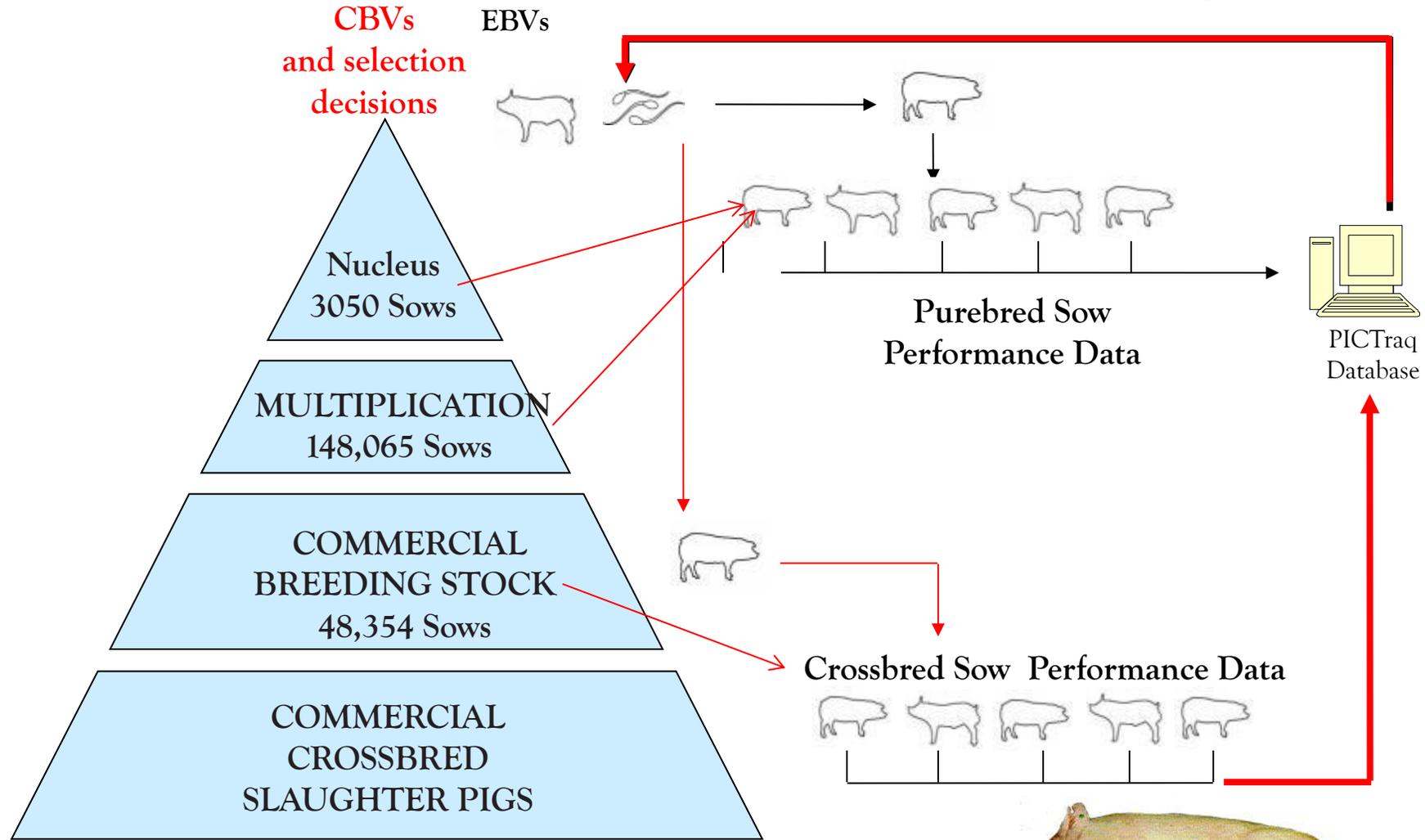
$$\mathbf{G}_{\omega} = (1-\omega)\mathbf{G} + \omega\mathbf{A}_{22}; \quad \mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{\omega} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

扩大参考  
群体外延



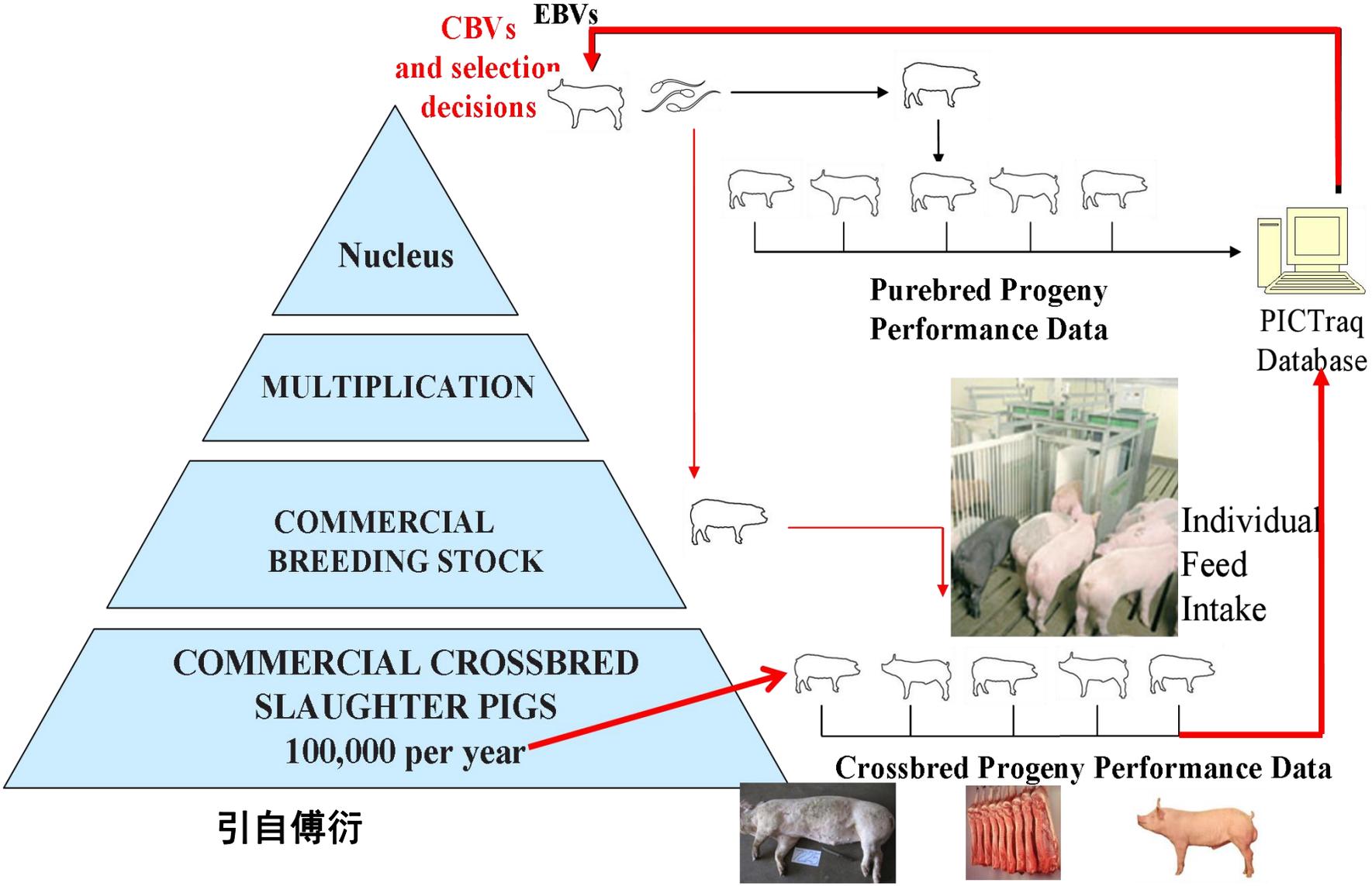
# Dam-Line Programs

## 信息管理平台



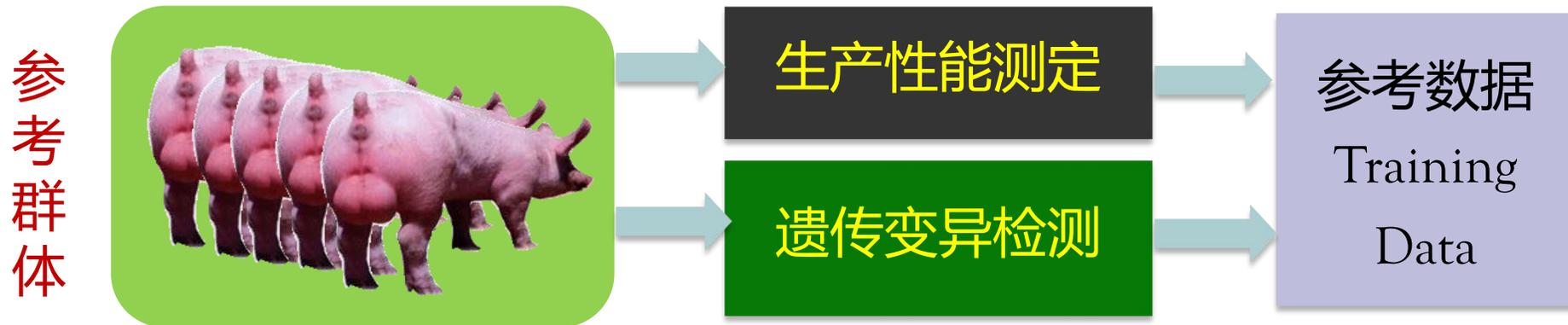
引自傅衍

# Sire-Line Program



引自傳衍

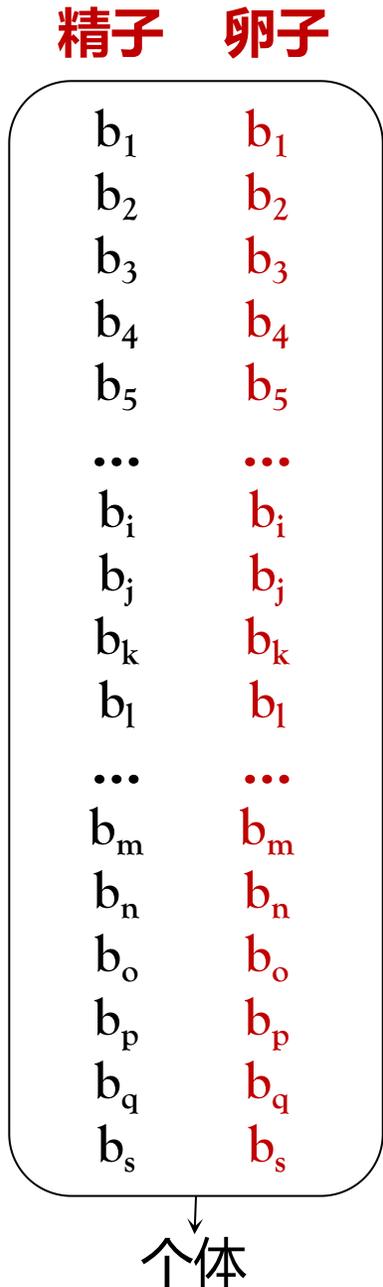
# □ 参考群体建立



- 个体遴选：数量、来源
- 性能测定：类别、方法
- 测定环境：多样/类似

} 代表性





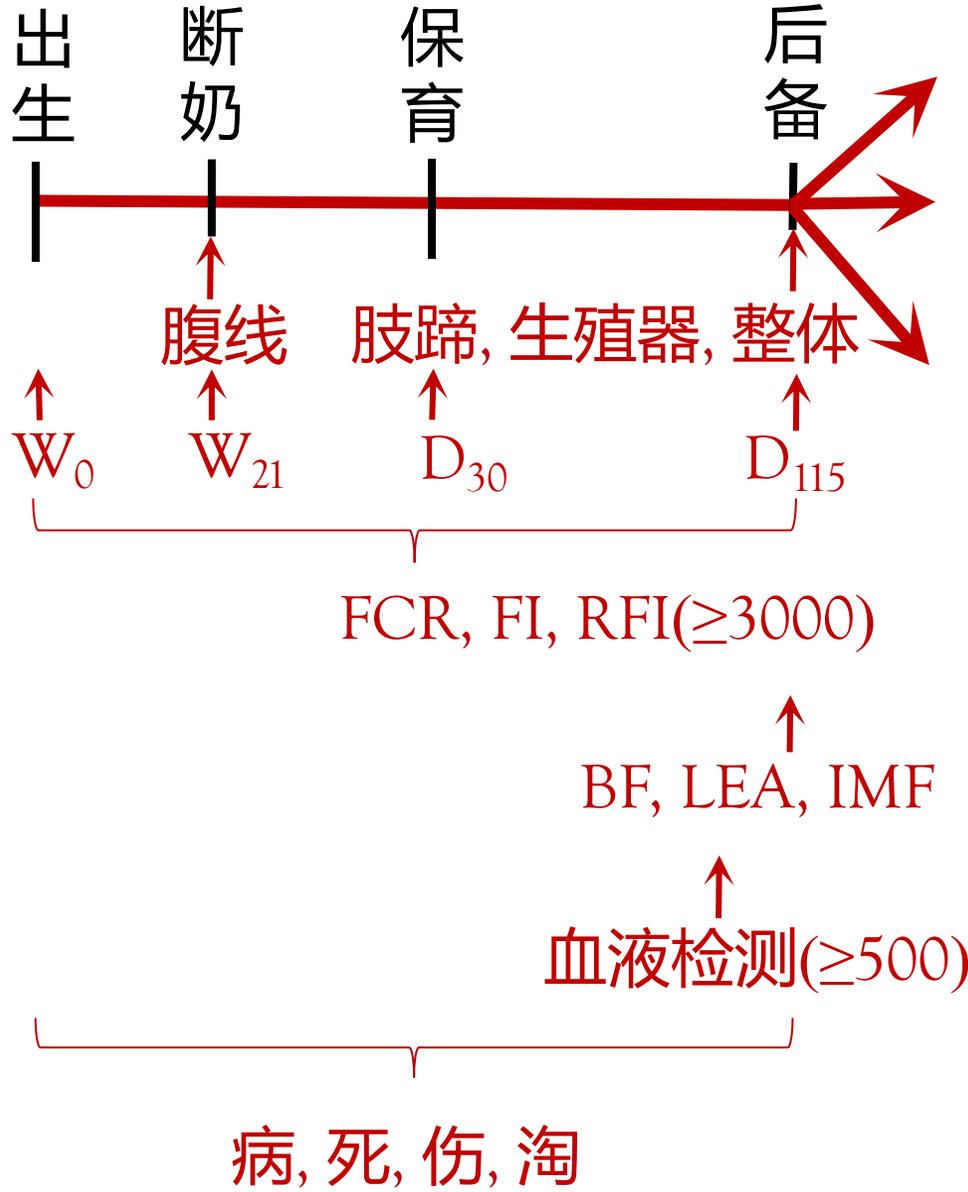
- 理论上，  
~~528亿种精子；528亿种卵子  
 → → 528亿 × 528亿种个体~~
- 按单倍体型  
 ~~$N^{7000}$ 种精子； $N^{7000}$ 种卵子  
 → →  $N^{7000} \times N^{7000}$ 种个体~~
- 个体多多益善  
 至少  $\geq 2000$ ；最好  $\geq 6000$
- 尽量包含各种变异  
 SNP；单倍体
- 与应用对象关系越近越好

# 性能测定

6000

- 体型外貌
- 生长性能
- 饲料转化
- 胴体肉质
- 健康性状

+1000



- 屠宰/肉质 ( $\geq 1000$ )
- 精液品质 ( $\geq 1000$ )
- 繁殖性能 ( $\geq 2500$ )
- 3500后续

精子 卵子

|       |       |
|-------|-------|
| $b_1$ | $b_1$ |
| $b_2$ | $b_2$ |
| $b_3$ | $b_3$ |
| $b_4$ | $b_4$ |
| $b_5$ | $b_5$ |
| ...   | ...   |
| $b_i$ | $b_i$ |
| $b_j$ | $b_j$ |
| $b_k$ | $b_k$ |
| $b_l$ | $b_l$ |
| ...   | ...   |
| $b_m$ | $b_m$ |
| $b_n$ | $b_n$ |
| $b_o$ | $b_o$ |
| $b_p$ | $b_p$ |
| $b_q$ | $b_q$ |
| $b_s$ | $b_s$ |

↓  
个体

测定环境

组建参考群体时的测定环境  
应与用种环境尽量一致

环境的一致性与多样性

表型(P) = 遗传(G) + 环境(E) + 互作(G×E)

$$P = A + D + I + E + G \times E$$

参考群体供谁参考

→ 参照群体、标准群体



# □ 国外状况

➤ 自2012年起，在一些大的育种公司开始应用

- PIC
- Topigs Norsvin
- Hypor
- Danbred
- Genesis
- 温氏
- .....

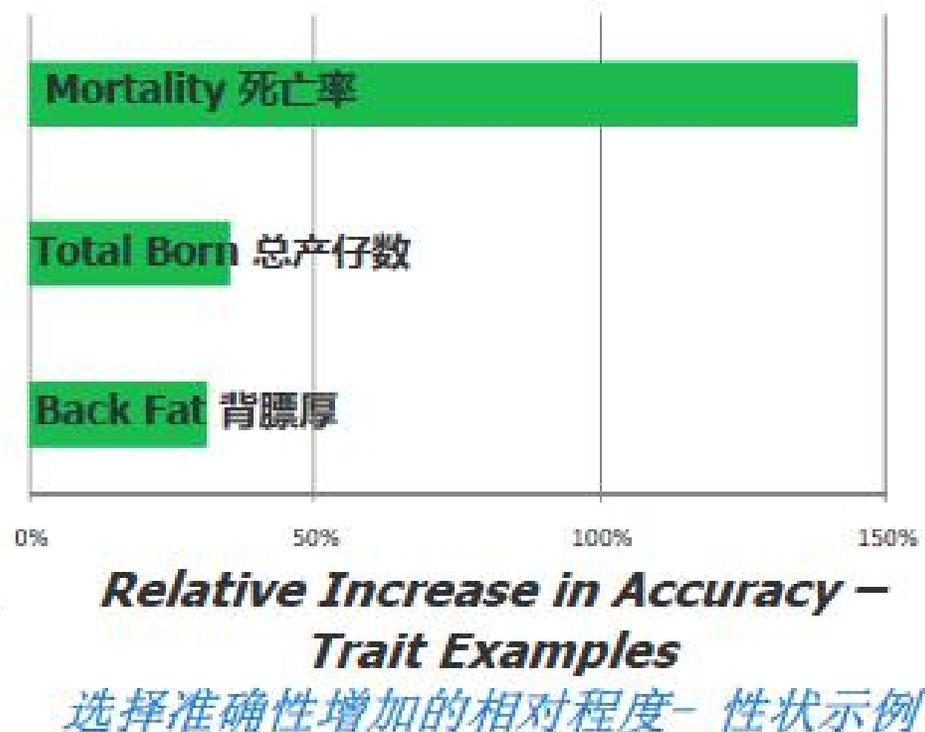
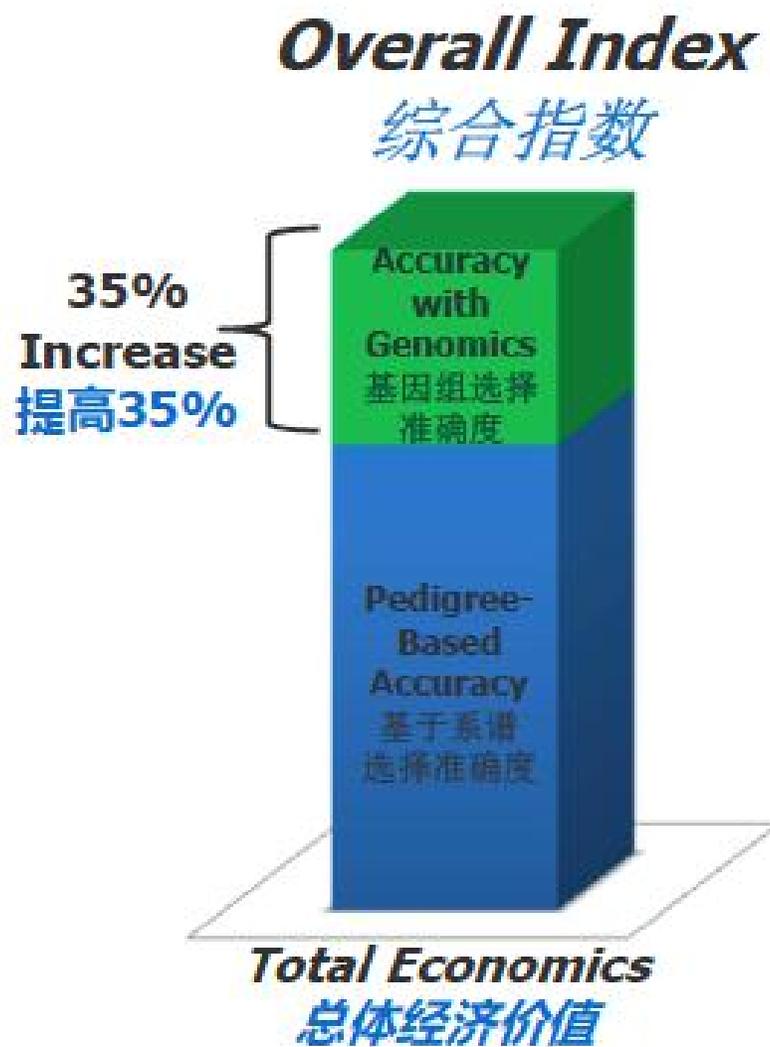


## ➤ PIC的应用情况

- 2013年开始应用
- 应用于所有育种目标性状
- 用GeneSeek芯片检测
- 平均每年检测约10万个个体
- 核心群所有公猪都检测
- 用一步法估计基因组育种值
- 2015年11月，启动了与罗斯林研究所合作的基因组测序项目，计划测定14,000头PIC个体的全部基因组信息

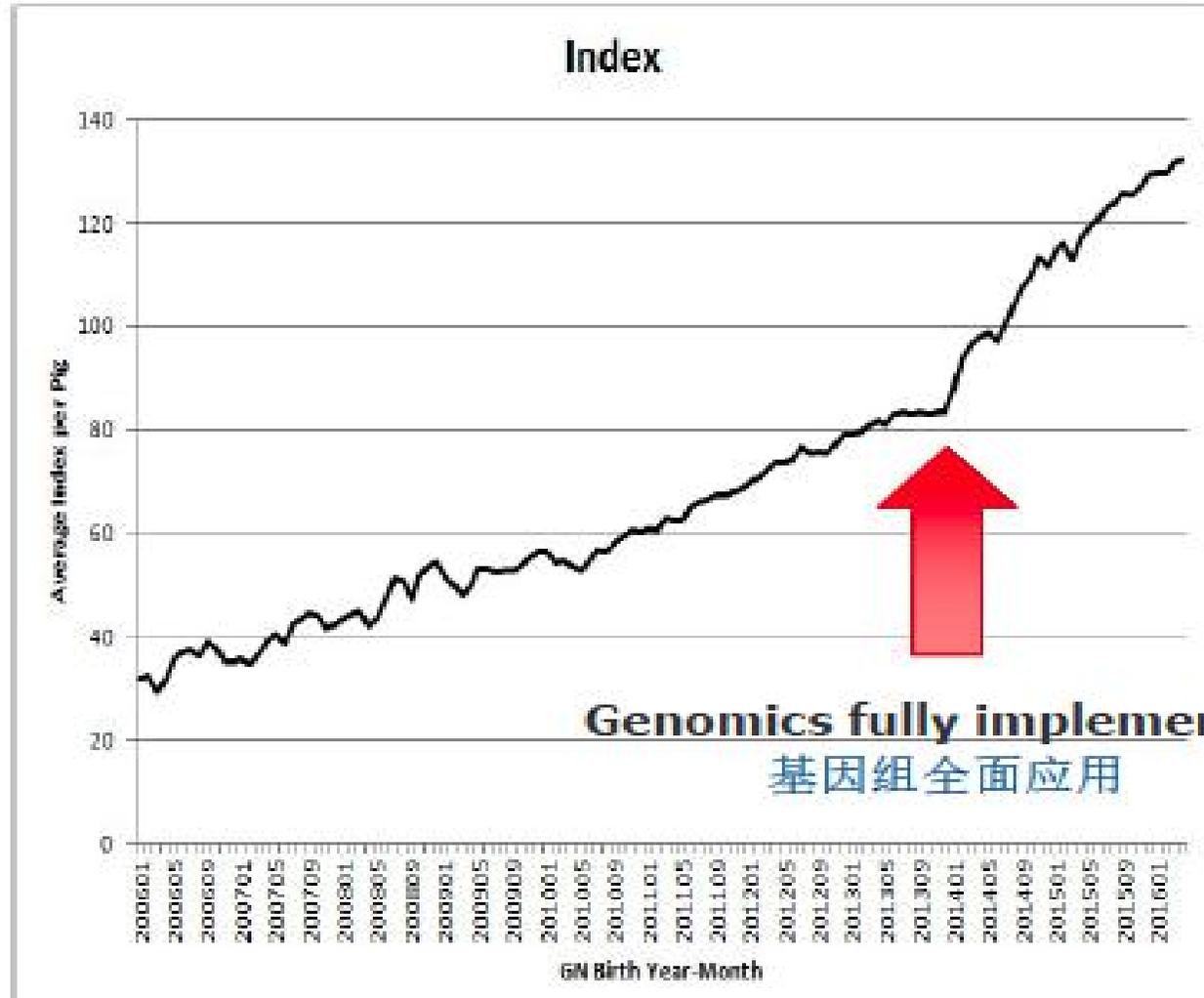


# 应用效果：提高选择准确性



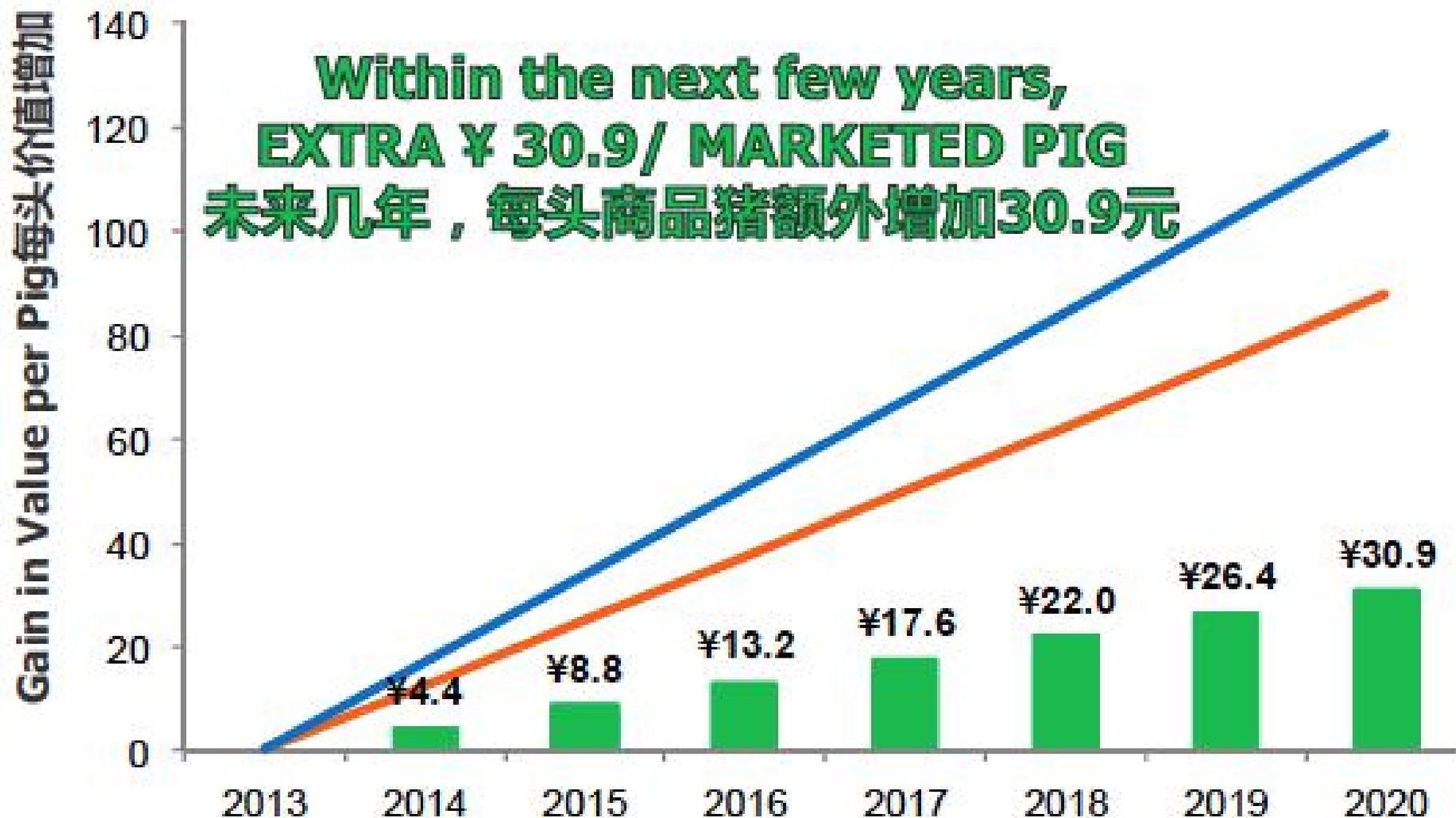
From David (PIC)

# 应用效果：加快遗传进展



|   | <u>1 Year Avg</u> |
|---|-------------------|
| <b>Index</b><br>指数                            | 19.5              |
| <b>Pigs weaned/sow/year</b><br>断奶猪/母猪/年       | 1.1               |
| <b>kg weaned/sow/year</b><br>断奶重/母猪/年         | 8.03              |
| <b>Pigs marketed/sow/year</b><br>商品猪/母猪/年     | 1.0               |
| <b>Total kg marketed/sow/year</b><br>出栏重/母猪/年 | 186.3             |
| <b>PROFIT PER PIG, CNY/pig</b><br>利润/商品猪      | 25.6              |

From David (PIC)



From David (PIC)

# 国内动态

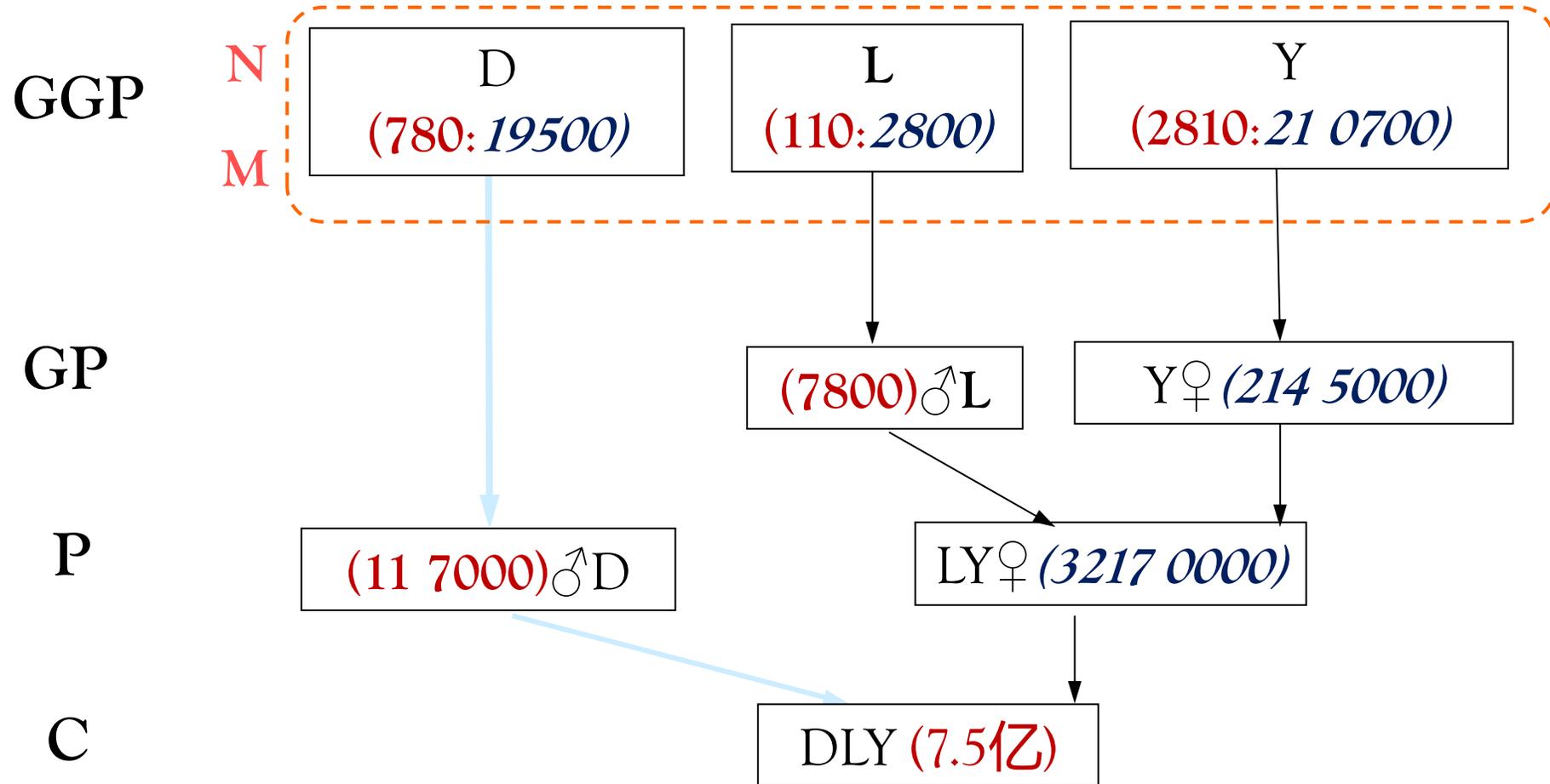
主办：农业部畜牧业司 全国畜牧总站  
承办：中国畜牧兽医学会儿养猪学分会



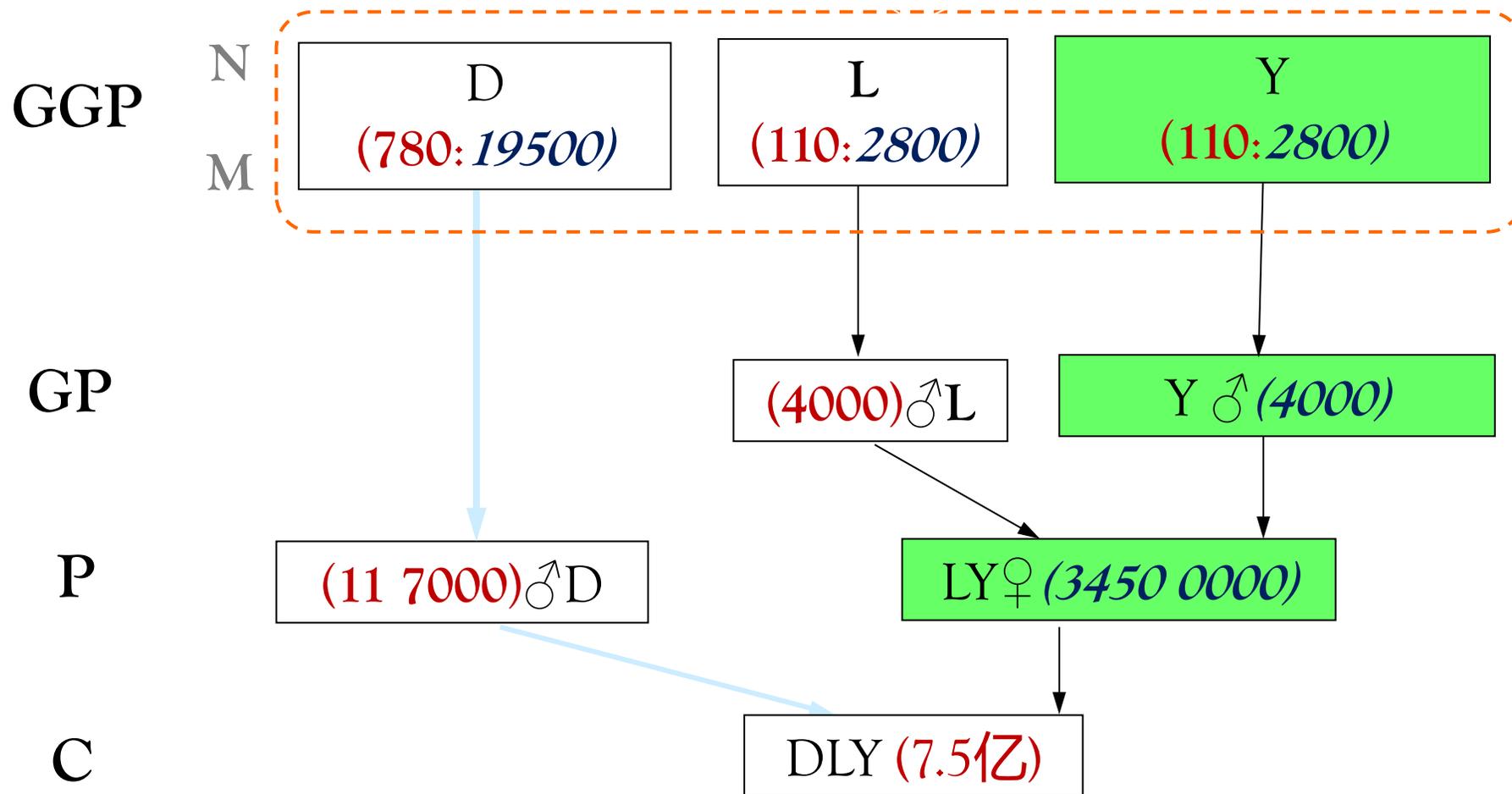


# □ 繁育体系

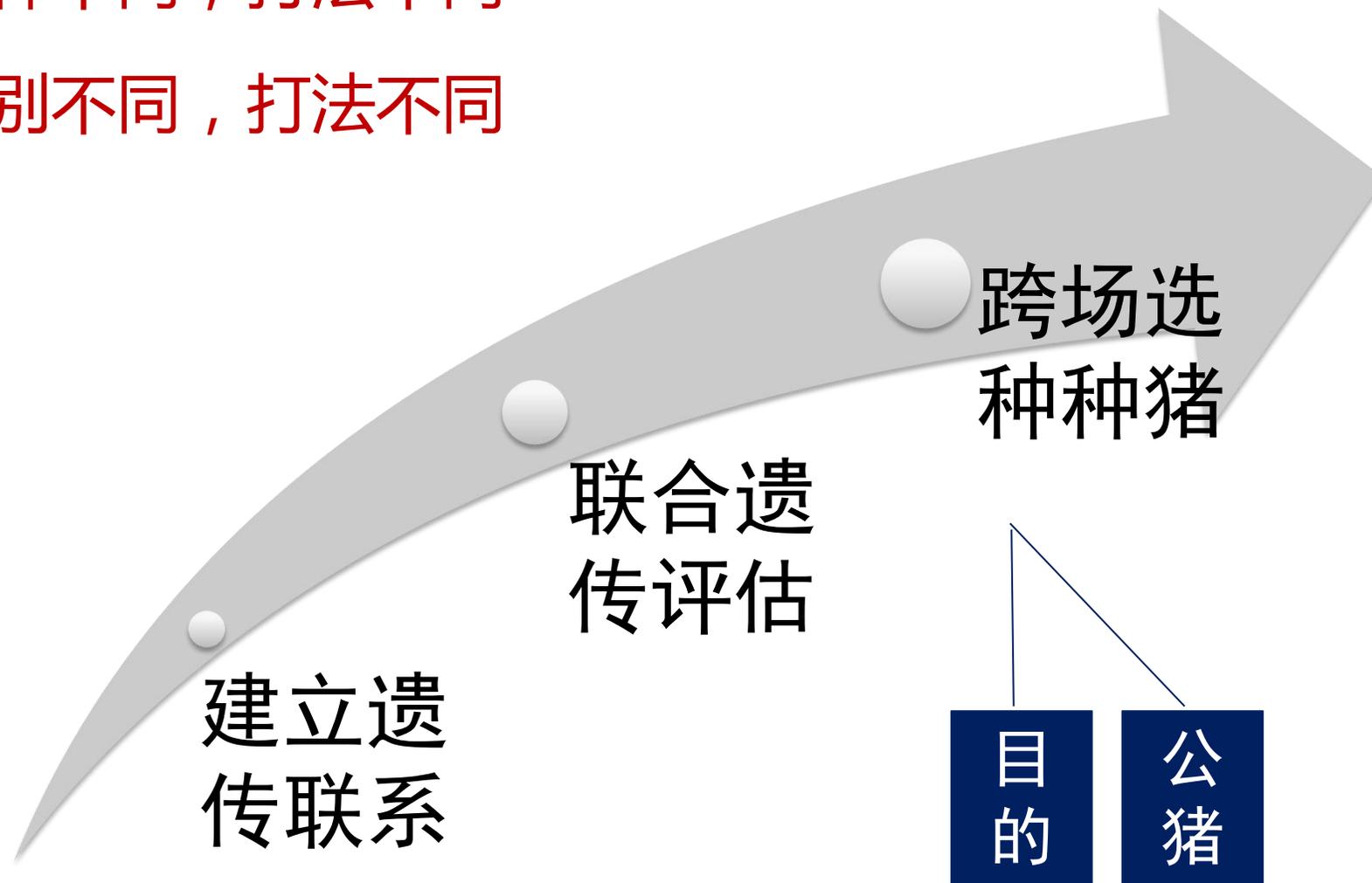
## 年产7.5亿头商品猪繁育体系各代规模



# 轮回杂交繁育体系



- 品种不同，打法不同
- 性别不同，打法不同



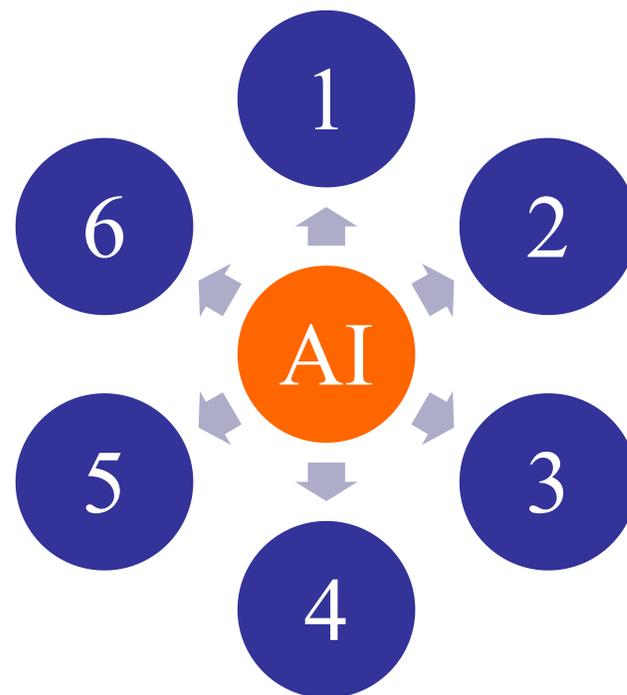
# 大白猪 (Y)

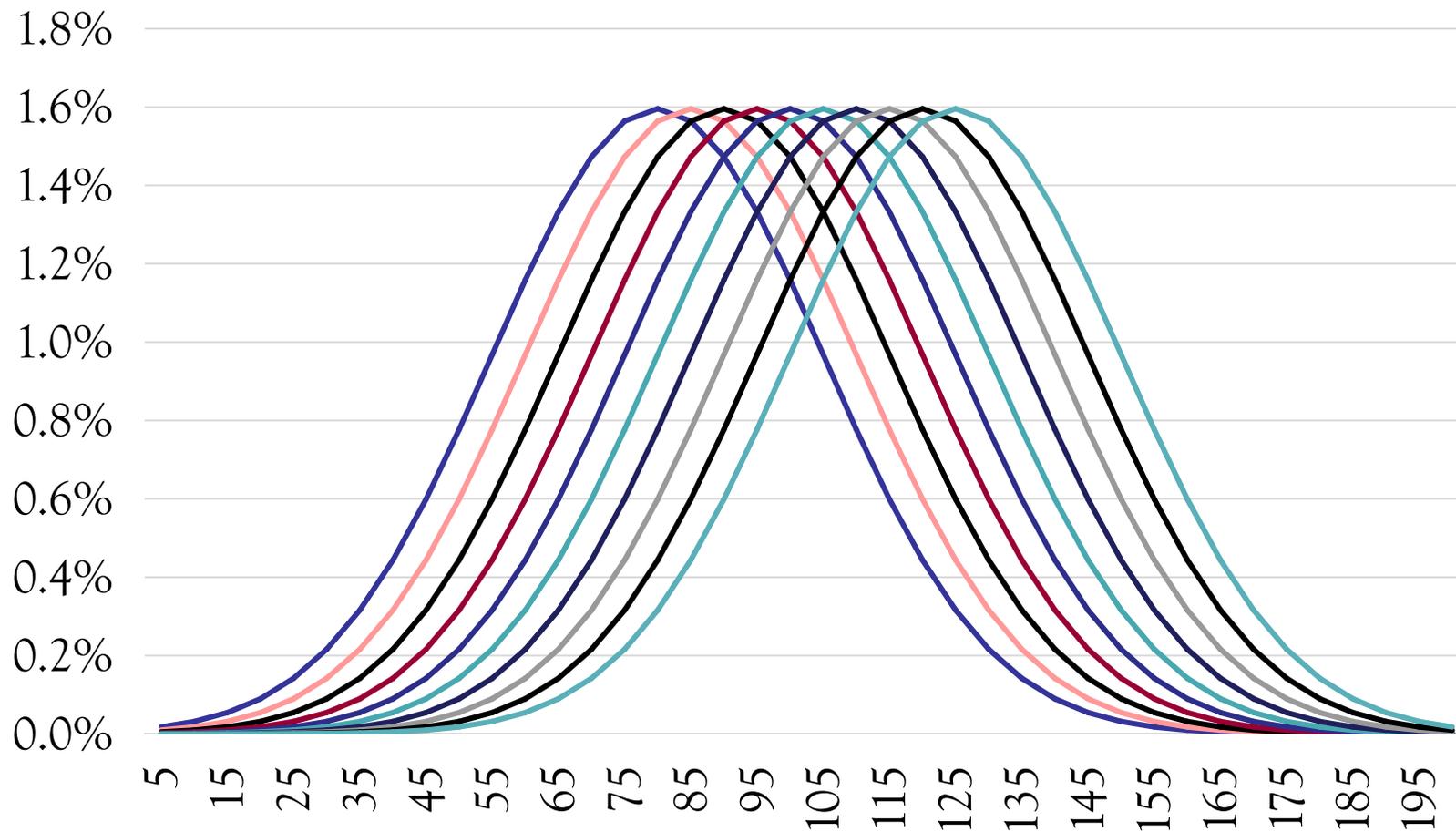


- 群体规模较大 (80000)，但是来源丰富，并且育种群 (N) 与扩繁群 (M) 乃至祖代群 (GP) 混杂
- 重视生物安全，育种群 (N) 种猪倾向自繁自留而非横向联合、跨场选种
- GGP+GP 母猪需求量很大 (19 0000+214 5000)，场家更关心纵向联合 (N→M→GP) 以图销售更多母猪
- GGP 种公猪需求量很少 (2530)，但是作用很大

## 联盟方式

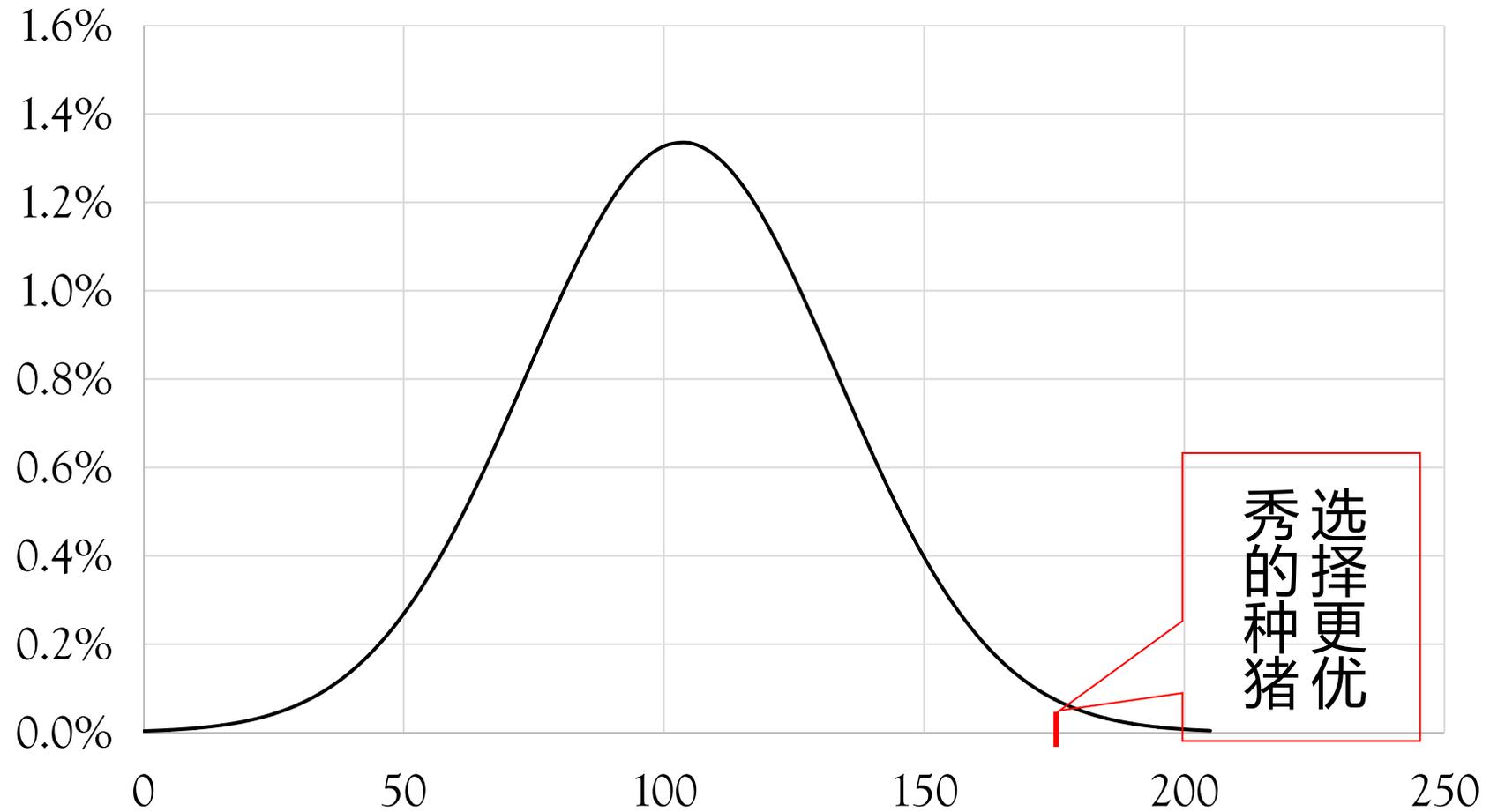
- 区域性或跨区域种源类似（！）的大白猪（Y）核心育种场成立联盟，共建参考群体。
- 联盟内各大白猪育种场（N），开展联合育种（ssBLUP）并且确保生物安全。
- 联盟内各大白猪育种场（N）共享种公猪或者精液，母猪可以共享或者自留，并提供给M、GP。
- 联盟内的育种场应提供其下属M、GP、P场的繁殖性能等信息。





各家猪场单独选种选配

## 跨场选种选配



- 如果12个月龄前的公猪用于核心育种场育种使用，则平均每头公猪可配母猪数

$$=5400/12.5/(2.25 \times 2) = 96\text{头}$$

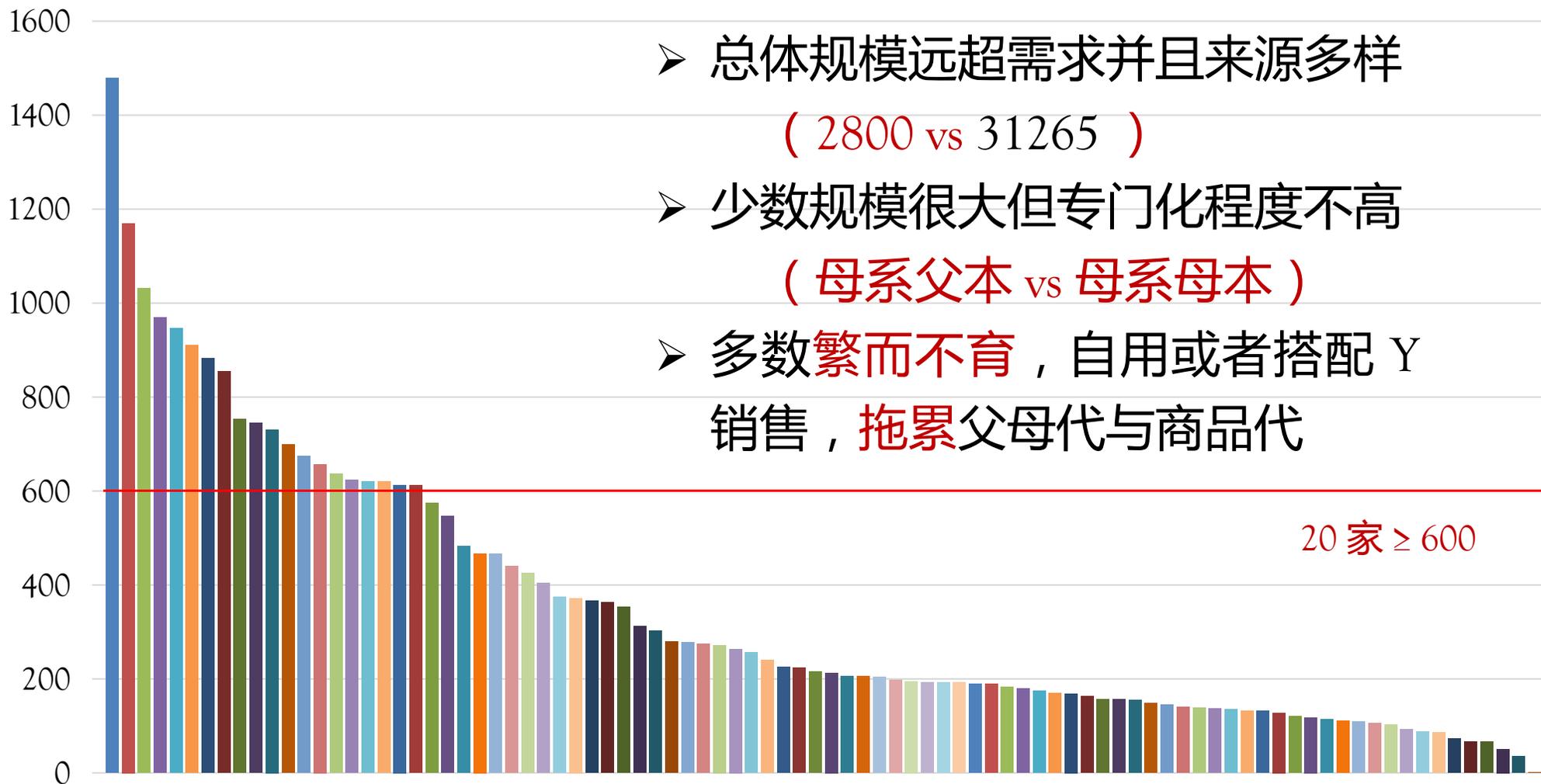
实际生产当中公母比例 1:10-1:25，在此假定1:16，则公猪数累计只需约1/6，选择强度也将提高。加上公猪世代间隔缩短，选择反应将大大提高！

$$\Delta G_t = \frac{\Delta G}{L} = \frac{\sigma_A \cdot i \cdot r_{AI}}{L}$$

同等重要的是，各核心场留种公猪可在12月龄前完成自家配种并为合作场供精后对外销售或服务。

- 鼓励核心育种场间交换精液，种公猪站主要为M服务

# □ 长白猪 (L)



➤ 总体规模远超需求并且来源多样

( 2800 vs 31265 )

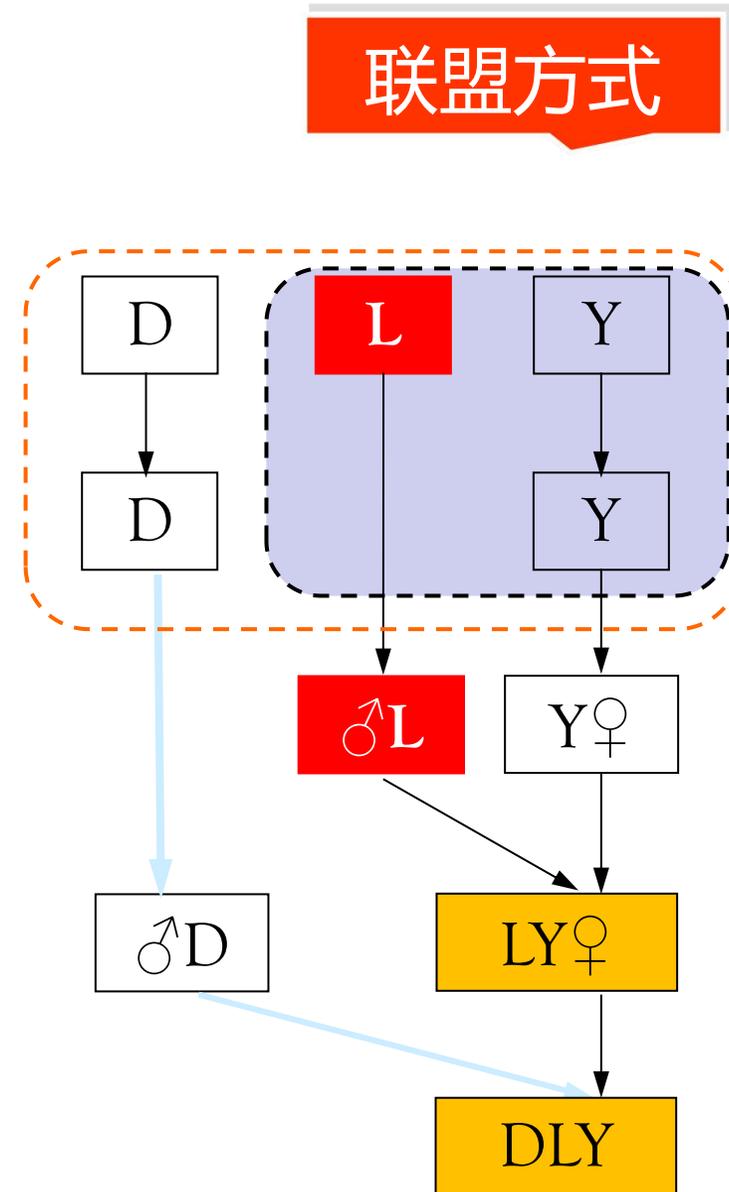
➤ 少数规模很大但专门化程度不高

( 母系父本 vs 母系母本 )

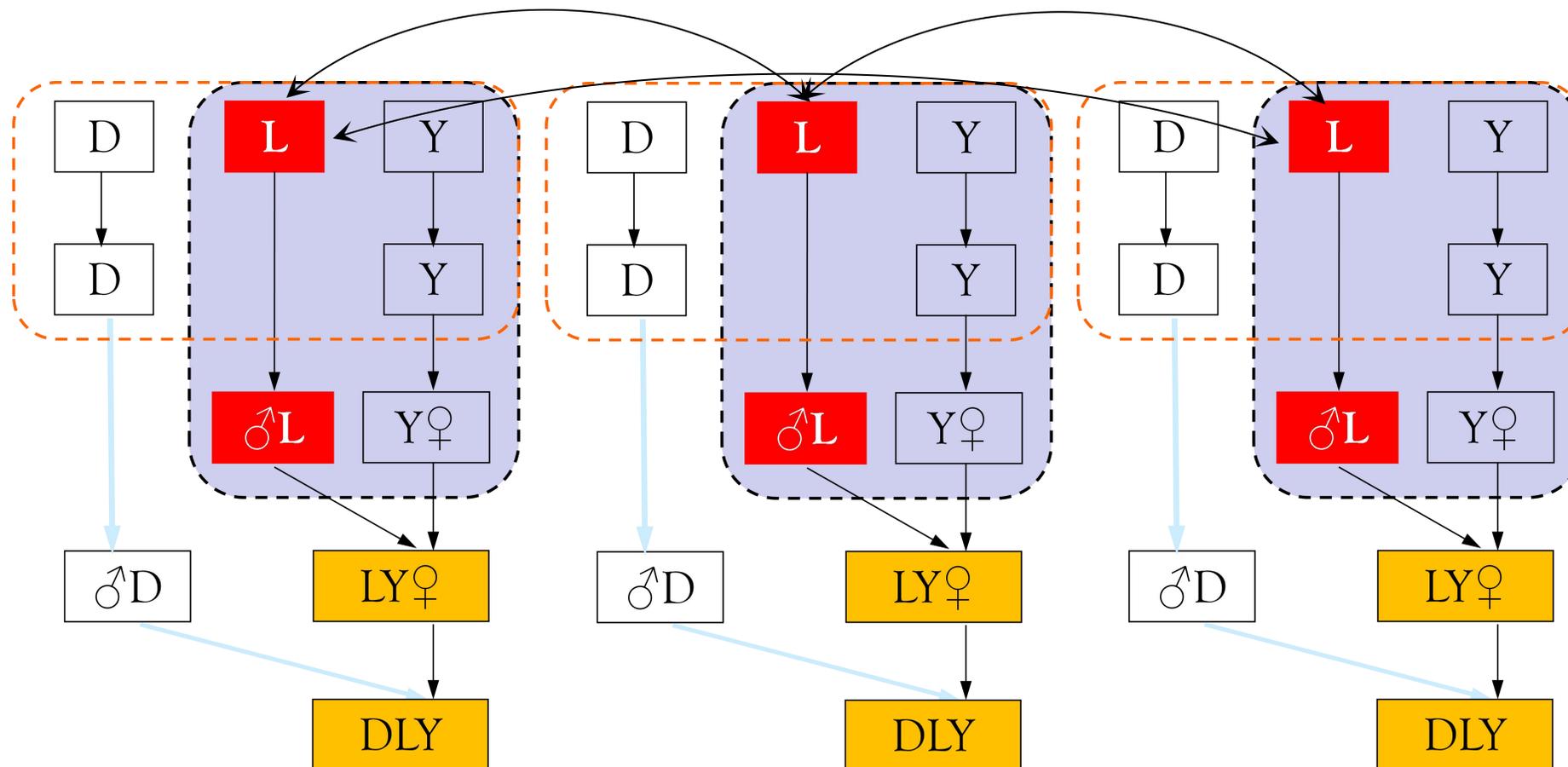
➤ 多数繁而不育，自用或者搭配 Y 销售，拖累父母代与商品代

2016年末国家级核心育种场长白猪规模

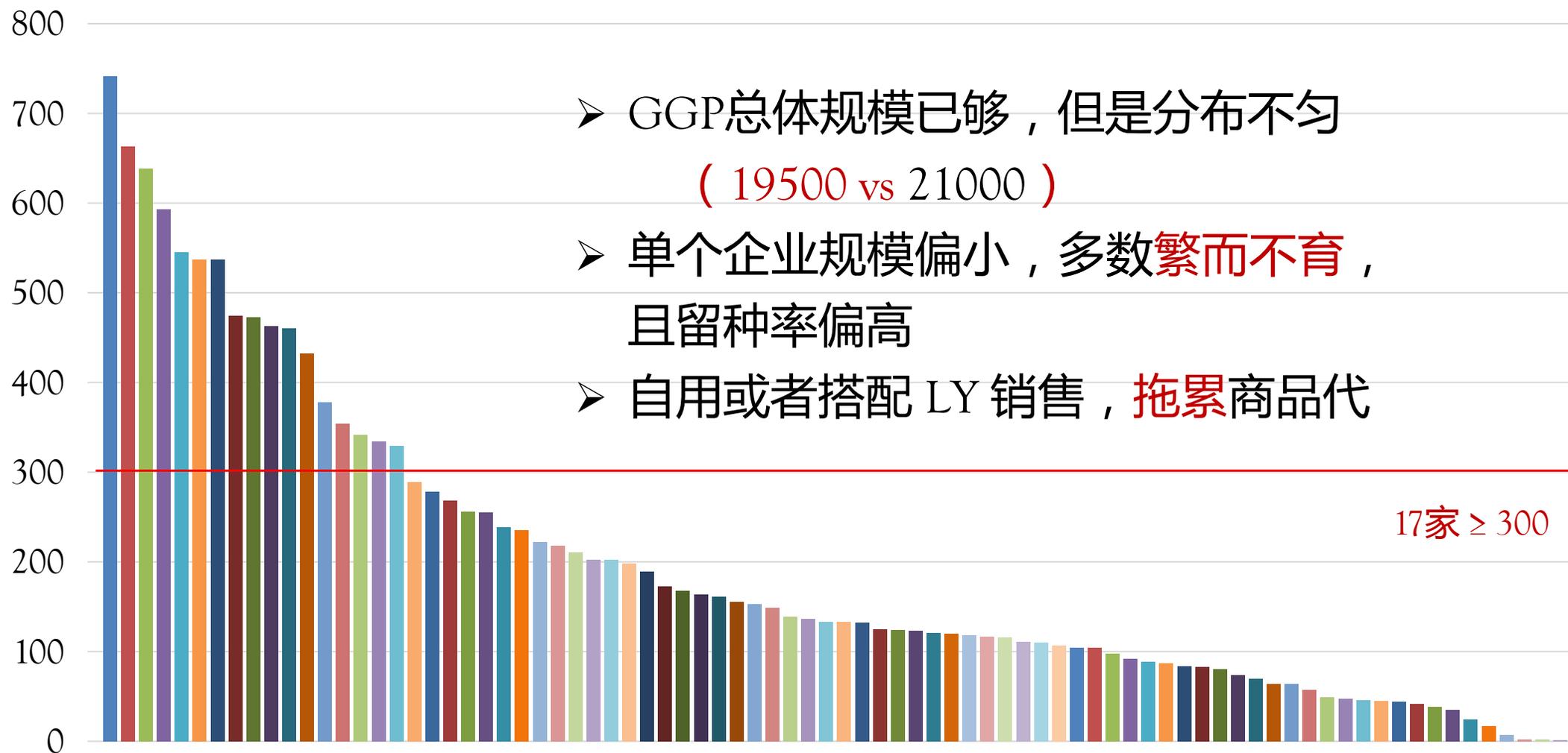
- 区域性的核心育种场成立联盟，共建参考群体，建好后不再饲养长白母猪生产种公猪。
- 联盟合建专门化的长白猪育种场（N），开展“常规+分子”育种并且确保生物安全。
- 专门化的长白猪育种场（N）直接向联盟下属的及区域内其他的祖代场（GP）提供种公猪，或由联盟再建一个种公猪站提供精液。
- 联盟的GP与P场应向长白猪育种场提供LY的繁殖性能等信息。



- 国家应高度重视长白猪育种场的区域布局及其联合育种（共享种公猪或精液+母猪）
- 国家应支持依托长白猪育种场的种公猪站建设，约束GGP场对长白猪的只繁不育，并使GP场直接使用精液。
- 鼓励探索不同杂交配套组合。例如，DLY→DY<sub>Y</sub>（美美法）



# □ 杜洛克 (D)

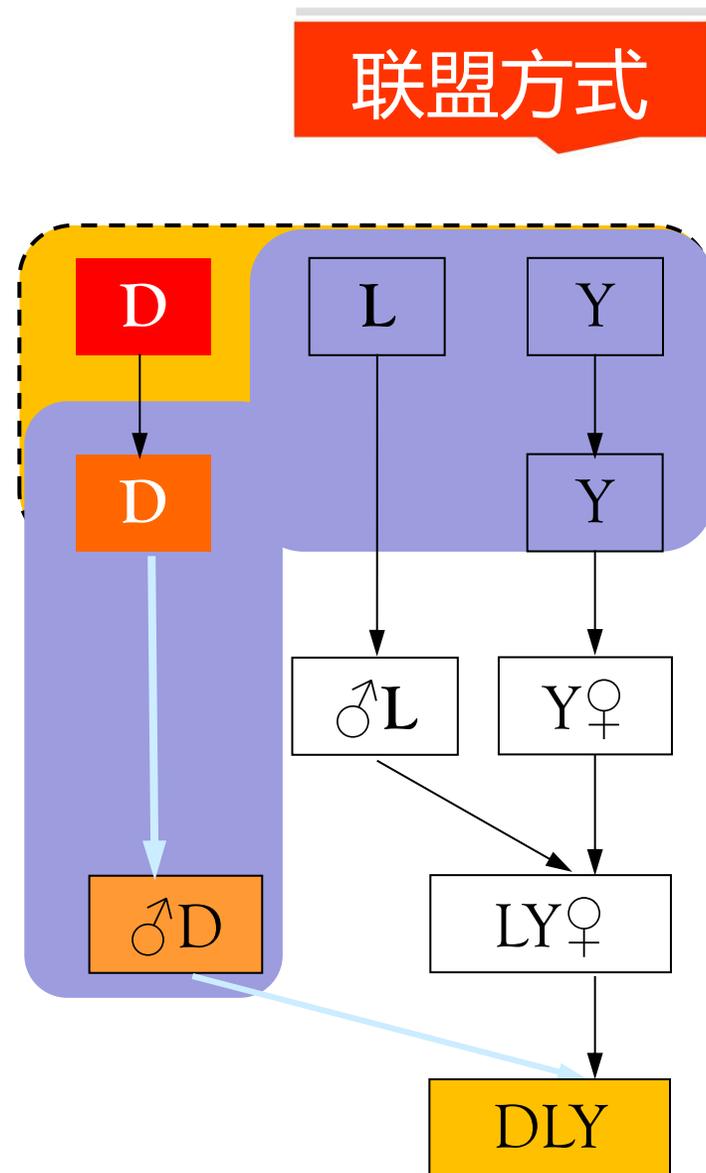


- GGP总体规模已够，但是分布不匀  
( 19500 vs 21000 )
- 单个企业规模偏小，多数**繁而不育**，且留种率偏高
- 自用或者搭配 LY 销售，**拖累**商品代

2016年末国家级核心育种场杜洛克规模

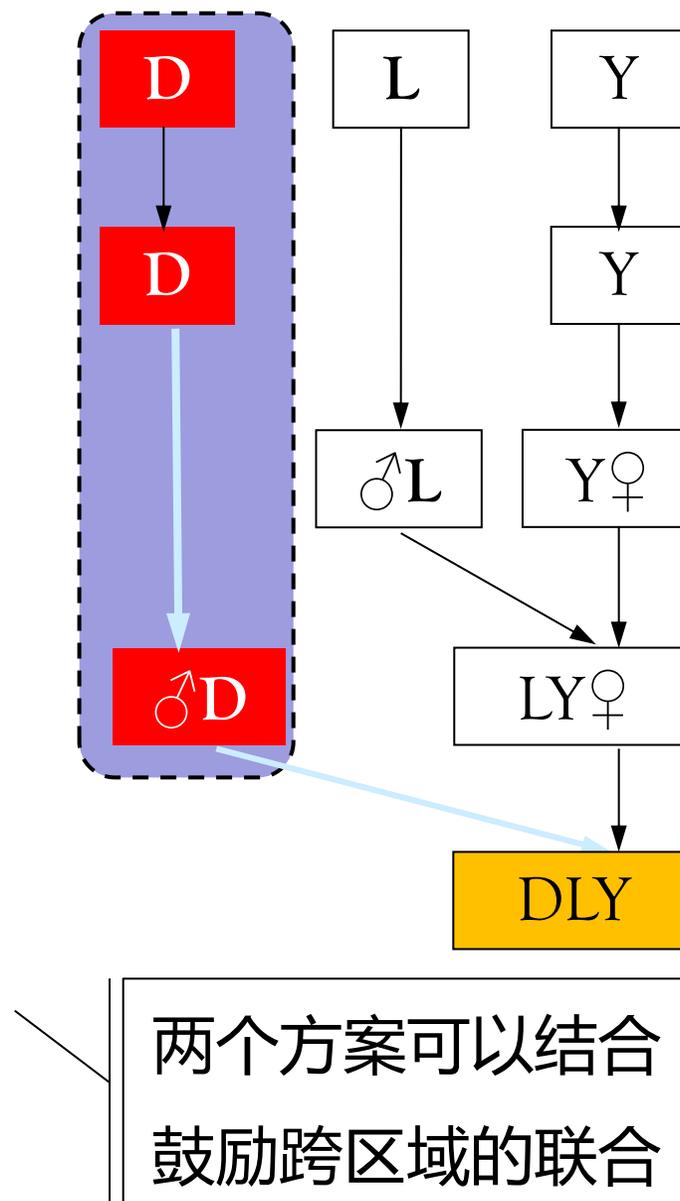
## ➤ 第一方案

- 区域内国家级核心育种场建立联盟，共建参考群体。
- 联盟设立专门化杜洛克猪育种场（N），开展“常规+分子”育种，并且确保生物安全，向扩繁群（M）和父母代场（P）提供种公猪或精液。
- 各核心场建扩繁群，只饲养母猪且少量自行更新、公猪自己不留，向P提供种公猪或精液。
- 联盟下属商品场向N提供信息。



## ➤ 第二方案

- 区域内由专门化杜洛克猪GGP场 (N+M) 自行构建参考群体, 开展“常规+分子”育种, 并且确保生物安全。
- GGP场直接向区域内父母代场 (P) 提供种公猪, 或建种公猪站提供精液 (首选)。
- 商品猪场 (C) 向GGP场提供生产性能信息。
- 区域内的核心育种场不再饲养杜洛克猪。



# 恳请多提宝贵意见！

潘玉春

浙江大学动物科学学院

13701869463

[panyc@zju.edu.cn](mailto:panyc@zju.edu.cn)

动物科学学院