数量性状基因型选择与基因型值选择潜力的比较

吴为人 周元昌 李维明

(福建农林大学作物科学学院, 福州 350002. E-mail: zjweiren@pub5.fz.fj.cn)

摘要 根据选择依据的不同, 数量性状标记辅助选择的方法可分为基因型选择和基因型值选择两种. 用计算机模拟的方法, 对两种选择方法的潜力进行了比较. 结果显示, 两种方法的基本规律相似, 除了在 QTL 数目很多且效应相等的情况下基因型选择表现较优之外, 在其他情况下, 两种方法的效果差不多. 从应用的角度看, 基因型选择与基因型值选择相结合应是今后的发展方向.

关键词 数量性状 标记辅助选择 基因型选择 基因型值选择 潜力

选择是决定育种成效的一个关键环节. 选择的目的是从分离群体中选出具有符合要求的基因型的个体, 因此, 最有效的选择方法应是直接依据个体的基因型进行选择. 然而, 要鉴定个体的基因型(特别是数量性状的基因型)常常是困难的, 因而传统育种中一般都只是依据个体的表现型进行选择. DNA 分子标记为直接依据基因型进行选择提供了可能, 这就是标记辅助选择(marker-assisted selection). 原则上, 不论是质量性状还是数量性状, 都可以借助相邻连锁的分子标记推断目标基因或数量性状基因座(QTL)的基因型, 从而实现依据基因型的选择(以下简称为基因型选择)^[1]. 对于数量性状, 除了可以进行基因型选择之外, 还可以依据基因型的期望表型——基因型值进行选择(以下简称为基因型值选择). 传统的以个体表型值为依据的选择方法其实就是一种近似的基因型值选择(因为: 基因型值 = 表型值 + 环境差值). 利用分子标记无需复杂的交配设计即可估计个体的基因型值, 因而可以比较容易地实现基因型值选择. 需要注意的是, 由于只有加性效应值是可稳定遗传的, 亦即只有加性效应是产生选择遗传进度的有效成分(事实上, 在经典数量遗传学中, 选择的遗传进度是依据狭义遗传率来预测的), 因此, 更确切地说, 这里需要估计的个体基因型值应该是加性基因型值. Lande 和 Thompson^[2]首先提出了两种利用分子标记估计的(加性)基因型值选择指标, 一种是标记分值(marker score), 它是由标记基因型预测的个体加性效应值, 另一种是选择指数(selection index), 由表型值和标记分值线性构成.

既然数量性状可以有两种选择依据(即基因型和

基因型值), 那么人们自然想知道哪种依据更有效. 就目前情况而言, 由于 QTL 定位的准确性不高, 且许多效应微小的 QTL 尚难以定位, 因此还无法做到依据完整的 QTL 基因型对数量性状进行选择. 相比之下, 依据近似的基因型值估值进行选择的方法显得更为实用. 也许正是因为这个原因, 近十余年来, 有关数量性状标记辅助选择的理论研究主要都集中在基因型值选择方法上^[2~9]. 随着基因组和 QTL 研究的不断深入和发展, 随着研究资料的不断积累, 也许将来有一天能够做到对目标数量性状遗传基础(包括有关基因的数目、位置和效应)的全面和精确的了解. 在这种情况下, 不论是基因型选择还是基因型值选择, 都可以做到高度精确, 达到两种选择方法的最大潜力. 所以, 我们感兴趣的问题是: 到底哪种方法的潜力更大? 回答这个问题, 对引导数量性状标记辅助选择方法学研究方向有重要的参考意义. 为此, 我们进行了计算机模拟研究. 本文报道研究成果.

1 材料与方法

用计算机来模拟选择过程. 为了比较基因型选择和基因型值选择的潜力, 必须考察理想情形, 即所有 QTL 的精确位置和效应都已知道, 并且每个 QTL 皆与一个共显性标记完全连锁或者本身就是一个可识别的共显性标记. 在这种理想情况下, 个体基因型可准确判断, 相应的加性基因型值也可直接计算, 不受环境效应和非加性效应的影响. 因此, 在模拟中可以不考虑环境效应和非加性效应.

不失一般性, 模拟中做了如下假设: (1) 研究材

<https://engine.scichina.com/doi/10.1360/csb2002-47-18-1408>

料为某种雌雄同株、能够自由传粉(即随机交配)的二倍体植物,其单倍基因组含有 10 条染色体,每条长皆为 100 cM. (2) 控制某目标数量性状的 QTL 的总数为 40 或 100, 均匀地排列在整个基因组中. (3) 全体 QTL 的加性效应总和为 100, 各 QTL 的加性效应大小相等或不等(对于“不等”的情况,假设 QTL 的加性效应大小表现为几何级数^[2], 最大效应与最小效应之间相差可达 2~3 个数量级,但 QTL 在染色体上的排列顺序与效应大小无关). (4) 选择的出发群体为两个纯系杂交的 F₂ 代. (5) 相邻 QTL 之间最初(即在 F₁ 中)的连锁方式表现为两种极端情况,或者都是相引连锁的(即增效基因彼此连锁,减效基因彼此连锁),或者都是相斥连锁的(即总是增效基因与减效基因相邻连锁). (6) 重组率与图距之间的关系符合 Haldane 作图函数,不存在交换干扰. (7) 每个个体的繁殖能力相同.

选择朝着性状增值的方向进行,到第 20 代停止,每代群体大小皆为 500, 中选个体间随机交配,繁殖下一代. 在基因型值选择中,先计算各个个体的加性基因型值 $y = \sum a_i x_i$, 式中 a_i 为第 i 个 QTL 的加性效应; x_i 为第 i 个 QTL 的基因型指示变量,对应于基因型 QQ , Qq 和 qq 分别取值 1, 0 和 -1; $\sum$ 表示对所有 QTL 求和. 然后,根据加性基因型值进行排序,再按照所要求的选择率从大到小选出相应数目的个体. 共考察了 3 种选择率,即 10%, 5% 和 1%, 各代选择率保持不变. 在基因型选择中,先统计各个个体中所含的 QTL 增效等位基因的数目(N , 不考虑 QTL 的效应大小),然后根据 N 值进行排序并从大到小逐个入选个体,直到中选个体所构成的子群体中包含了全部 QTL 的增效等位基因为止(以保证没有某个 QTL 的增效等位基因在选择过程中丢失). 因此,基因型选择中所用的选择率是动态的. 对每种情况均进行了 100 次模拟.

2 结果与分析

模拟结果(图 1)显示,基因型选择和基因型值选择的基本规律是相似的: (1) 由(a)与(e), (b)与(f), (c)与(g)及(d)与(h)之间的比较可以看出,选择的 QTL 数目越多则遗传进度越慢. 例如,在情形(a) (QTL 数目 = 100)中,基因型选择的累计遗传进度需到第 11 代才接近最大值,而在情形(e) (QTL 数目 = 40)中,到第 6

代已接近最大值. (2) 由(a)与(c), (b)与(d), (e)与(g)及(f)与(h) 之间的比较可以看出, QTL 效应相等情况下的遗传进度比不等情况下的慢. 例如,在情形(c)(效应不等)中,基因型选择的累计遗传进度到第 8 代已接近最大值,比情形(a) (效应相等)早 3 代. (3) 由(a)与(b), (c)与(d), (e)与(f)及(g)与(h)之间的比较可以看出, QTL 相斥连锁情况下的遗传进度比相引连锁情况下的慢. 例如,在情形(b)(相斥连锁)中,基因型选择的累计遗传进度需到第 14 代才接近最大值,比情形(a)(相引连锁)迟 3 代. 以上结论与前人有关基因型值选择的研究^[2-5]是一致的, 尽管我们这里模拟的是 QTL 的数目、位置和效应皆已知的“理想情形”, 而前人模拟的是 QTL 的数目、位置和效应皆未知, 因而必须通过性状对标记逐步回归的方法进行估计的“实际情形”.

关于基因型选择与基因型值选择的比较,如图 1 所示,几乎在所有情形中,基因型选择的遗传进度都快于选择率为 10% 和 5% 的基因型值选择(惟一的例外是情形(d),在那里基因型选择的遗传进度仅比选择率为 10% 的基因型值选择快),这是因为基因型选择采用的动态的选择率数值上始终小于 5%. 比较(a)与(c), (b)与(d), (e)与(g)及(f)与(h), 可以看出,在 QTL 效应相等的情形中,基因型选择与选择率为 10% 和 5% 的基因型值选择之间在遗传进度上的差距较大. 因此看来, QTL 效应相等对基因型值选择的影响要比对基因型选择的影响来得大. 这种现象是可以理解的. 在基因型选择中,各个 QTL 是平等看待的,因此,不论 QTL 效应相等还是不等,各 QTL 受到的选择压是相同的. 在基因型值选择中,在 QTL 效应相等的情况下,各 QTL 受到的选择压是相同的,而在 QTL 效应不等的情况下,各 QTL 受到的选择压是不同的,效应大的 QTL 将受到更大的选择压而优先被选择,从而产生更大的遗传进度.

从图 1 还可看出,在所研究的情形中,选择率为 1% 的基因型值选择的遗传进度总是略快于或相当于基因型选择,但它往往会使得少数有利等位基因丢失,从而最终无法到达累计遗传进度的上限(即本例中假设的加性效应总和 100),而基因型选择则不存在这个问题. 看来,这里 1% 的选择率对基因型值选择来说似乎是过高了. 不过,这种因选择率过高而造成最终

遗传进度损失的程度仅在 QTL 数目较多且效应相等时才比较严重(情形(a)和(b)), 而在其他情形中则不太明显. 这是可以理解的. 事实上, 当 QTL 数目较少时, 选择过程中有利等位基因丢失的机会就会减少;

当 QTL 效应不等时, 由于效应较大的 QTL 的有利等位基因总是优先被选择, 因而选择过程中丢失的往往是一些效应较小的 QTL 的有利等位基因. 这样, 最终遗传进度的损失就不会那么严重.

图 1 基因型选择和基因型值选择的计算机模拟结果

(a) QTL 数目 = 100, 效应相等, 相引连锁, (b) QTL 数目 = 100, 效应相等, 相斥连锁, (c) QTL 数目 = 100, 效应不等, 相引连锁, (d) QTL 数目 = 100, 效应不等, 相斥连锁, (e) QTL 数目 = 40, 效应相等, 相引连锁, (f) QTL 数目 = 40, 效应相等, 相斥连锁, (g) QTL 数目 = 40, 效应不等, 相引连锁, (h) QTL 数目 = 40, 效应不等, 相斥连锁. × 示基因型选择, ■ 示基因型值选择(10%), ▲ 示基因型值选择(5%), ● 示基因型值选择(1%). 每条曲线均是 100 次模拟结果的平均

<https://engine.scichina.com/doi/10.1360/csb2002-47-18-1408>

3 讨论

选择的目的是获得符合要求的基因型. 基因型值是基因型的表达结果. 因此, 基因型选择是直接的, 而基因型值选择是间接的. 从这一点考虑, 直觉上可以预期基因型选择应比基因型值选择更有效. 但从本研究的结果看, 除了在 QTL 数目很多且效应相等的情况下, 两种方法的潜力并没有明显的差别. 而已有的 QTL 定位研究结果显示, QTL 效应一般都不等的, 从大到小排列通常表现为几何级数或 L 形分布的形式^[10], 而且从效应角度看, 一个数量性状实际牵涉到的 QTL 数目可能并不太多^[11]. 因此看来, 在所有 QTL 的位置和效应都准确已知的理想情况下, 基因型选择和基因型值选择的效果相差不大.

在本文研究中, 考虑到在理想情况下非加性效应对个体加性基因型值的计算不会产生影响, 因此模拟中没有考虑非加性效应. 不过, 在非加性效应中, “加性×加性”上位(简称加加上位)效应最终是可以固定的. 因此, 当存在显著的加加上位时, 应考虑它对选择的影响. 对基因型选择来说, 加加上位不会影响其过程, 但会影响最终结果. 由于两个增效基因之间的加加上位效应未必就是增效的, 因此最终选得的所有 QTL 上都是增效基因的基因型未必就是最佳基因型. 所以这时必须综合考虑加性效应和加加上位效应来确定基因型选择的目标. 对基因型值选择来说, 为了最终固定有利的加加上位效应, 在每代选择中就必须对加加上位效应进行估计. 但由于加加上位效应在选择过程中不能稳定遗传, 因此会使选择的可预见性下降. 总的来说, 加加上位使标记辅助选择的问题复杂化, 需要做进一步研究. 但从加加上位会降低基因型值选择预见性这一点来看, 我们预计存在加加上位时基因型值选择的效果可能会明显不如基因型选择.

在实际情况中, 由于以基因型值为依据的选择方法必须对加性基因型值进行估计, 因此非加性效应和环境效应必然会对估计值(包括标记分值和选择指数)产生影响, 尽管模拟研究显示非加性效应的影响并不十分明显^[4]. 相对来说, 基因型选择方法可以避免非加性效应和环境效应的影响, 因此具有优越性. 影响基因型选择效率的因素主要是 QTL 定位的精度. 随着基因组研究的深入, 相信会做到对 QTL 位置的精确定位. 当然, 要做到对一个性状的所有

QTL 都被识别和精确定位是困难的, 因为有些 QTL 的效应可能太小了. 我们针对实际情况的模拟研究显示(结果将另文发表), 在已有部分 QTL 被定位的情况下, 将基因型选择与依据基因型值(以表型值、标记分值或选择指数作为基因型值的近似值)的选择相结合的方法, 要比单纯依据基因型值的方法有效得多. 已定位的 QTL 越多, 则这种相结合的选择方法的效果越好. 因此, 将基因型选择和基因型值选择相结合可能是今后数量性状标记辅助选择方法的主要发展方向. 为此, 在今后相当长的时期内, QTL 定位仍将是摆在遗传育种学家面前的一项重要的基础工作.

致谢 本工作为国家自然科学基金资助项目(批准号: 39893350).

参 考 文 献

- 1 Hospital F, Charcosset A. Marker-assisted introgression of quantitative trait loci. *Genetics*, 1997, 147: 1469 ~ 1485
- 2 Lande R, Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 1990, 124: 743 ~ 756
- 3 Gimelfarb A, Lande R. Simulation of marker assisted selection in hybrid population. *Genet Res Camb*, 1994, 63: 39 ~ 47
- 4 Gimelfarb A, Lande R. Simulation of marker assisted selection for non-additive traits. *Genet Res Camb*, 1994, 64: 127 ~ 136
- 5 Edwards M D, Page N J. Evaluation of marker-assisted selection through computer simulation. *Theor Appl Genet*, 1994, 88: 376 ~ 382
- 6 Gimelfarb A, Lande R. Marker-assisted selection and marker-QTL associations in hybrid populations. *Theor Appl Genet*, 1995, 91: 522 ~ 528
- 7 Hospital F, Moreau L, Lacoudre F, et al. More on the efficiency of marker-assisted selection. *Theor Appl Genet*, 1997, 95: 1181 ~ 1189
- 8 Knapp S J. Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. *Crop Sci*, 1998, 38: 1164 ~ 1174
- 9 Moreau L, Charcosset A, Hospital F, et al. Marker-assisted selection efficiency in populations of finite size. *Genetics*, 1998, 148: 1353 ~ 1365
- 10 Bost B, Vienne D, Hospital F, et al. Genetic and non-genetic bases for the L-shaped distribution of quantitative trait loci effects. *Genetics*, 2001, 157: 1773 ~ 1787
- 11 Otto S P, Jones C D. Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics*, 2000, 156: 2093 ~ 2107

(2002-03-14 收稿, 2002-08-14 收修改稿)