

DOI: 10.3724/SP.J.1005.2012.00050

利用全基因组连锁不平衡估计中国荷斯坦牛有效群体大小

尼桂琰, 张哲, 姜力, 马裴裴, 张勤, 丁向东

中国农业大学动物科技学院, 农业部动物遗传育种与繁殖重点实验室, 北京 100193

摘要: 有效群体大小是群体遗传学研究的一个重要内容, 有助于我们更清楚地了解群体的遗传变异、进化和复杂性状的遗传机制等。随着高密度 SNP 标记的出现, 越来越多的研究利用 SNP 标记间连锁不平衡估计有效群体大小。文章采集北京地区中国荷斯坦牛 2 093 个样本, 并利用牛 SNP 芯片(Illumina BovineSNP50, 含 5 4001 SNPs)进行基因型测定, 估计不同世代中国荷斯坦牛的有效群体大小。质量控制标准设定为 SNP 检出率 0.95, 最小等位基因频率 >0.05, 样本检出率 0.95, 哈代温伯格平衡检验显著性水平 $P < 0.0001$ 。经过质量控制, 共 1 968 个样本和 38 796 个 SNPs 用于连锁不平衡分析。文章选取 SNP 间距 0.1、0.2、0.5、1、2、5、10、15 (Mb), 估计中国荷斯坦牛在 4 世代之前有效群体大小。结果表明, 中国荷斯坦牛的有效群体呈逐代下降趋势, 至 4 世代前, 中国荷斯坦牛平均有效群体为 45 头左右。

关键词: 中国荷斯坦牛; 有效群体; 连锁不平衡

Chinese Holstein Cattle effective population size estimated from whole genome linkage disequilibrium

NI Gui-Yan, ZHANG Zhe, JIANG Li, MA Pei-Pei, ZHANG Qin, DING Xing-Dong

Key Laboratory of Animal Genetics and Breeding of the Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

Abstract: Effective populations size (N_e) is an important population parameter that helps to explain genetic variation, population evolution and understanding of the genetic architecture underlying complex traits. With the availability of high-dense SNP panels, more and more researches focus on estimating of N_e using linkage disequilibrium (LD) between SNPs. In this study, we estimated the effective population size from 2093 Chinese Holstein Cattle genotyped with Illumina BovineSNP50 BeadChip. After removal of individual with call rate < 95%, SNPs with call rate < 95%, minor allele frequency < 5% and Hardy-Weinberg Equilibrium test with $P < 0.0001$, 1 968 individuals with 38 796 SNPs were remained. Eight kinds of SNP pairs with the distances 0.1, 0.2, 0.5, 1, 2, 5, 10, and 15 Mb were respectively chosen to estimate the effective population size of Chinese Holstein cattle from 4 generations ago. It is demonstrated from the results of this study

收稿日期: 2011-03-23; 修回日期: 2011-04-27

基金项目: 国家自然科学基金项目(编号: 30800776)和国家高技术研究发展计划(863 计划)项目(编号: 2008AA101002)资助

作者简介: 尼桂琰, 硕士研究生, 专业方向: 动物分子数量遗传学。Tel: 010-62734408; E-mail: nigy1111@163.com

通讯作者: 丁向东, 博士, 副教授, 研究方向: 动物分子数量遗传学。Tel: 010-62734277; E-mail: xd Ding@cau.edu.cn

网络出版时间: 2011-10-18 8:50:29

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20111018.0850.001.html>

that the effective population size of Chinese Holstein is decreased in the past generations, and the corresponding effective population size at ~4 generations ago is only around 45.

Keywords: Chinese Holstein Cattle; effective population; linkage disequilibrium

有效群体大小(Effective population size, N_e)指与实际群体有相同基因频率方差或相同杂合度衰减率的理想群体含量, 通常小于绝对的群体大小。它决定了群体平均近交系数增量的大小, 反映了群体遗传结构中基因的平均纯合速度, 是群体遗传学研究的一个重要内容。有效群体大小的研究有利于我们更清楚地了解群体进化历史, 了解复杂性状的遗传机制^[1], 也在畜禽育种中有重要的指示作用^[2], 有效群体太小意味着群体遗传变异很小, 进而影响到遗传进展, 同时有效群体规模过小难以避免过早和过高的近亲交配, 进而影响品种内的异质性和广泛的使用价值。

有效群体大小估计方法最早由Wright提出^[3], 后来根据群体结构、群体类型的不同也相应发展了诸多方法^[2]。传统的有效群体大小估计方法是建立在个体间相互独立的随机交配群体基础上的, 这点在商业畜禽群体中很难达到, 畜禽中普遍采用的利用系谱信息估计有效群体大小, 但由于畜禽的家系结构往往都很复杂, 常常由于系谱错误会对有效群体的估计产生很大影响^[4]。随着分子标记的出现, 利用DNA序列的多态信息也可以估计有效群体大小, 该方法原理比较简单, 但有其局限性。因为在进化和育种过程中, 有效群体规模不是一成不变的, 而利用DNA序列的多态信息只能估计出过去很长一段时间内有效群体大小的平均值^[5]。近年来, 随着高密度分子标记尤其是海量SNP(Single nucleotide polymorphism)的出现, 利用标记间的连锁不平衡(Linkage disequilibrium, LD)可以更准确估计有效群体大小, 长距离LD可以用来估计近世代的有效群体大小, 短距离LD可以用来估计远世代的有效群体大小^[6]。

作为世界上产奶量最高的奶牛品种, 荷斯坦牛的有效群体大小受到了众多关注, 已相继开展了利用标记间连锁不平衡估计荷斯坦牛有效群体大小的诸多研究。Hayes等^[6]利用位于牛 20 号染色体上的

16 个微卫星标记估计出澳大利亚荷斯坦牛 5~6 个世代前有效群体规模约为 150 头。随着大规模、高通量测序技术的发展, 越来越多研究利用全基因组高密度 SNP 芯片来估计有效群体大小, Kim 和 Kirpatrick^[7]利用 7 117 个分布于全基因组的 SNP 标记, 估计北美荷斯坦牛有效群体大小, Sargolzaei 等^[8]利用更高密度的 Affymetrix 10K SNP 芯片估计北美荷斯坦牛有效群体大小, Qanbari 等^[9]利用 Illumina 公司出品的 50K 高密度 SNP 芯片估计德国荷斯坦牛群体有效大小。由于基于高密度标记估计有效群体大小受标记间连锁不平衡程度影响很大, 而连锁不平衡又具有群体特异性, 因此针对不同荷斯坦牛群体进行有效群体大小估计十分必要。

中国荷斯坦牛由我国动物育种工作者通过纯种荷斯坦牛与中国地方黄牛的杂交改良, 经过一定世代的级进杂交后, 进而横交固定, 在长时间的育种工作后, 培育而成, 与其他国家荷斯坦牛相比具有特殊性, 其有效群体大小估计将对今后中国荷斯坦牛育种工作具有参考和指示作用。

1 材料和方法

1.1 实验动物

本研究的实验对象为北京地区的中国荷斯坦牛, 由 14 个公牛半同胞家系组成, 母牛头数为 2 093 头, 分别出生于 2001~2006 年期间, 来自北京地区 25 个牛场。每个公牛家系规模从 83 头到 358 头不等, 家系平均大小为 150 头。母牛血样由各个牛场提供, 公牛的精液由北京市奶牛中心提供。

1.2 基因型测定

提取 DNA 后, 用含有 54 001 个 SNP 标记的牛高密度芯片(Illumina BovineSNP50)进行基因型测定, 并利用 Beadstudio 软件(由 Illumina 公司提供)读取芯片数据及进行数据处理。根据以下标准对基因型测定进行质量控制: SNP 检出率(call rate)大于 0.95;

最小等位基因频率(MAF)大于等于 0.05, 样本检出率大于等于 0.95, 哈代-温伯格平衡检验显著性水平 0.0001; 同时剔除染色体位置未知的 SNP。

1.3 有效群体大小估计

Sved^[10]提出了连锁不平衡程度和有效群体大小之间关系的近似公式,

$$E(r^2) \approx 1/(\alpha + kN_e c) \quad (1)$$

其中 r^2 为不同分子标记间的连锁不平衡程度, N_e 为有效群体大小, c 为标记间的遗传距离, 以摩尔根(M)为单位, α 取值表示模型中是否考虑突变, 1 表示不考虑突变^[10], 2 则考虑突变^[11-13]; k 的取值与染色体有关, 常染色体为 4, X染色体为 2。

目前大多数研究采用无突变模型利用高密度 SNP 标记估计有效群体大小^[6, 8, 9], 本研究同样利用此模型估计中国荷斯坦牛有效群体大小。

根据公式(1), 有效群体大小的估计需要已知标记间连锁不平衡程度和遗传距离。

1.4 LD 度量

连锁不平衡程度有多种度量指标, 其中 D' ^[14]和 r^2 ^[15]最常用, 但 D' 受群体大小影响很大, 尤其在小组群和中等大小群体中 D' 估计值偏高, r^2 则对基因频率和有效群体大小的变化不敏感^[16]。本研究因此采用 r^2 作为LD的度量指标, 对于两个SNP标记, 其等位基因分别为 A, a 和 B, b 及相应的等位基因频率为 p_A, p_a, p_B, p_b , 等位基因间形成的 4 种单倍型频率分别为 $p_{AB}, p_{Ab}, p_{Ab}, p_{aB}$, 则两个标记间的 r^2 度量公式为:

$$r^2 = \frac{(p_{AB}p_{ab} - p_{Ab}p_{aB})^2}{p_A \cdot p_a \cdot p_B \cdot p_b} \quad (2)$$

本研究运用Haploview软件^[17]计算各SNP间的 r^2 值。

1.5 标记间遗传距离

对于高密度SNP标记, 目前还无法通过连锁图谱获得SNP间的遗传距离, 通常根据物理距离近似换算遗传距离^[2]。

$$\text{标记间遗传距离}(cM) = \frac{\text{染色体遗传距离}(cM)}{\text{染色体物理距离}(Mb)} \times \text{标记间物理距离}(Mb) \quad (3)$$

染色体的物理距离和标记间的物理距离可由牛的基因组参考序列 Btau4.2(<ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/fasta/Btau20080815/LinearizedChromosomes/>)获得。

以牛 6 号染色体(BTA6)为例, 从 NCBI (<http://www.ncbi.nlm.nih.gov/projects/mapview/>)上获得该染色体的遗传长度为 125.6 cM, 本研究采用的 50K SNP 芯片中该染色体的物理长度为 122.63 Mb, 则根据公式(3), 两个物理距离为 100 kb 的 SNP 间的遗传距离为 0.1019 cM。据此可以计算 BTA6 上所有 SNP 标记间的遗传距离, 其他染色体也如此操作。

1.6 不同世代下有效群体大小

标记间的连锁不平衡程度会随着世代增加而衰减, 不同长度染色体片段的连锁不平衡程度可以反映不同世代的有效群体大小。长染色体片段通常连锁不平衡较低, 可以用来估计近世代的有效群体大小, 而短染色体片段连锁不平衡较高, 则可以估计远世代的有效群体大小, 遗传距离为 c 的染色体片段可以近似反映出前 $1/(2c)$ 世代的有效群体大小^[6]。因此, 根据公式(1)、(2)、(3), 可以通过对具有不同物理距离的SNP标记间的连锁不平衡分析, 估计不同世代下的有效群体大小。如物理距离为 100 kb 的两个SNP, 其遗传距离为 0.001Morgan。根据其连锁不平衡程度可以得到前 500 世代时有效群体大小的一个估计值, 而将所有间距 100 kb 的SNP估计出的有效群体大小加以平均, 则可得到前 500 世代时有效群体大小较为可靠的估计。参照利用牛 50K SNP 芯片估计有效群体大小的其他研究^[2], 本研究选取了 0.1、0.2、0.5、1、2、5、10 和 15 Mb 等不同SNP间距, 通过对这些不同间距SNP间的连锁不平衡程度分析, 估计中国荷斯坦牛在相应前 452、226、91、46、23、10、5、4 世代时的有效群体大小。

2 结果与分析

2.1 质量控制前后 SNP 分布

根据牛基因组参考序列 Btau4.2, 图 1 反映了质量控制前后 SNP 分布变化。样本由 2 093 减少为 1 698 头, SNP 则由 54 001 个减少为 38 796 个, 平均间距为 70 kb, 平均 MAF 为 0.28, 用于后期分析。

2.2 LD 变化规律

图 2 反映了中国荷斯坦牛整个基因组不同间距的SNP间的LD分布情况, 总体而言, 随着物理距离的增加, LD 呈衰减趋势, 但是某些距离较远的 SNP

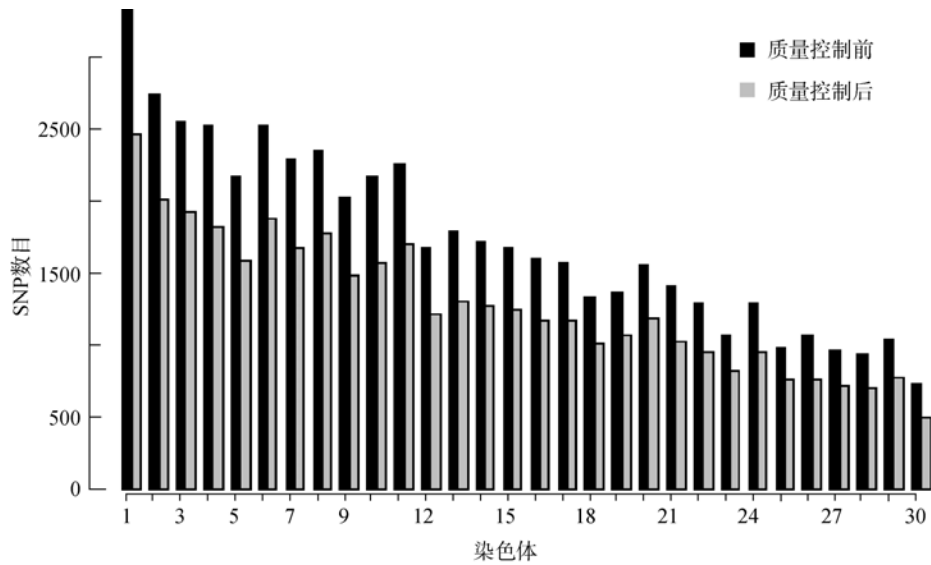


图 1 质量控制前、后, 中国荷斯坦牛 29 条常染色体和 X 染色体上 SNP 数量分布(图中 30 表示 X 染色体)

间也存在强 LD, 这表明牛基因组不仅存在近距离的连锁不平衡, 而且远距离的标记间也以一定的模式存在连锁不平衡, 这与其他研究报道一致^[18-21]。图 2 同时表明距离较近的 SNP 间 LD 的变异程度要高于距离较远的 SNP 间 LD 的变异程度。

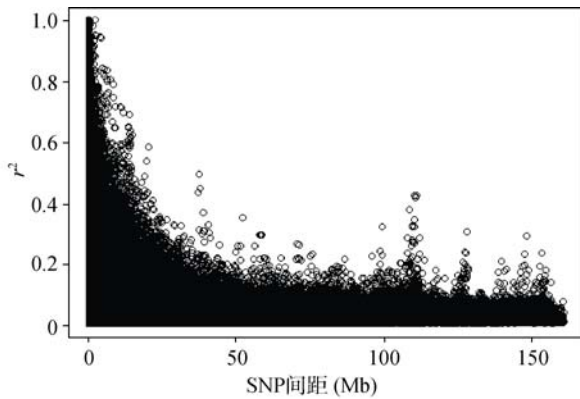


图 2 中国荷斯坦牛全基因组连锁不平衡(r^2)分布

2.3 有效群体大小估计

如表 1 和图 3 所示, 中国荷斯坦牛有效群体大小从 452 世代前的 1 088 头减少至 4 世代前的 45 头, 反映了由于人工选择中国荷斯坦牛有效群体大小下降的趋势, 这与其他国家荷斯坦牛有效群体大小研究比较一致^[7, 8, 22]。同时从图 2 可以看出, 即使相同的 SNP 间距, 不同的 SNP 标记间的连锁不平衡程度

也有差异, 因此有效群体大小估计也有所差异。如前 4 世代时, 间距为 15 Mb 的不同 SNP 对估计的中国荷斯坦牛有效群体大小为 39~51。

连锁不平衡程度的差异也反映在不同染色体上, 表 2 和图 3 显示了根据不同染色体上的 SNP 估计的中国荷斯坦牛有效群体大小。表 2 反映出根据所有染色体和单条染色体估计有效群体大小的差异, 如前 4 世代时, 中国荷斯坦牛有效群体大小范围为 39~51, 而如果根据每条常染色体估计有效群体大小, 则中国荷斯坦牛前 4 世代时有效群体大小范围扩大为 36~62, 这是由于不同染色体相同染色体片段大小连锁不平衡程度差异较大, 有较高连锁不平衡程度的染色体估计的有效群体偏小, 反之, 则较大。如 BTA6、14 估计出的有效群体较小, 其对应的 r^2 值也较高(表 3)。相同世代下用不同染色体估计出的有效群体大小经方差分析表明, 染色体间差异极显著 ($P < 0.001$)。表 2 也反映出每条染色体上 LD 随 SNP 间距增加而减小的趋势, 与图 2 表现的整体衰减趋势一致。每条染色体估计的有效群体大小也随着世代增加而减少, 相同世代下, X 染色体估计出的有效群体大小显著大于其他常染色体(表 2, 图 3), 这可能与根据 X 染色体估计特殊性和与常染色体不同的估计方法有关, 因此本研究在表 1 中仅根据所有常染色体连锁不平衡信息估计中国荷斯坦牛有效群体大小。

表 1 中国荷斯坦牛根据常染色体 SNP 标记估计的各世代有效群体大小

	SNP 间距(kb)							
	100	200	500	1 000	2 000	5 000	10 000	15 000
遗传距离	0.001	0.002	0.006	0.011	0.022	0.055	0.111	0.166
世代	452	226	91	46	23	10	5	4
SNP 对数	5105	4800	4626	43217	40636	34807	28168	22845
N_e (平均数 \pm 标准差)	1087.65 \pm 173.43	767.30 \pm 116.44	410.50 \pm 75.27	233.81 \pm 28.26	137.95 \pm 17.02	76.81 \pm 9.25	53.83 \pm 7.80	44.66 \pm 5.88

表 2 由不同染色体估计中国荷斯坦牛各世代有效群体大小

染色体	世代数							
	452	226	91	46	23	10	5	4
BTA1	848.35	638.44	354.9	210.29	117.656	68.188	45.93	36.606
BTA2	1062.04	664.35	364.02	212.986	112.581	64.93	43.996	38.016
BTA3	1040.6	713.94	423.3	241.374	141.211	75.11	46.511	35.732
BTA4	1232.9	688.24	432.88	221.37	133.65	70.02	49.309	37.515
BTA5	917	757.72	348.87	227.014	135.13	69.499	49.398	42.869
BTA6	904.22	656.71	381.56	220.217	129.08	71.277	54.123	41.29
BTA7	1018.22	767.56	356.34	230.595	136.59	77.51	55.504	47.046
BTA8	935.1	713.9	388.79	235.27	141.46	76.246	51.935	41.79
BTA9	1222.9	832.44	466.32	241.249	146.203	75.977	49.827	43.763
BTA10	955.21	694.96	411.5	187.39	107.77	61.21	44.338	37.371
BTA11	1128.67	757.34	328.81	228.438	126.53	64.551	47.092	41.16
BTA12	1209.6	894.2	392.25	253.465	147.15	90.377	63.404	45.099
BTA13	852.73	619.19	334.27	202.031	131.07	76.301	48.336	48.318
BTA14	900.08	585.7	362.6	183.91	129.126	75.906	49.004	44.913
BTA15	1264.1	800.54	400	266.377	166.392	82.83	61.161	47.457
BTA16	914.8	758.35	439.91	224.196	140.55	75.842	46.925	40.981
BTA17	991	854.1	334.4	220.55	122.66	70.898	50.709	45.9
BTA18	1126.4	780.12	574.9	293.33	181.956	90.976	66.517	54.05
BTA19	1160.6	694.46	543.59	266.114	165.14	94.513	73.043	51.782
BTA20	1058.8	627.19	304.86	179.8	113.64	59.364	44.313	40.8
BTA21	1134	782.21	519.81	259.607	154.34	81.824	60.188	45.768
BTA22	1255	806.95	383.5	218.838	145.784	88.811	59.024	48.799
BTA23	1050	676.31	311.22	235.911	121.78	68.115	46.491	38.15
BTA24	988	724.9	450.38	248.325	143.478	81.488	56.613	45.388
BTA25	1504.2	997.6	392.13	272.28	151.366	86.98	68.057	61.63
BTA26	917	849.5	373.58	217.15	125.35	82.882	57.872	48.585
BTA27	1444.1	1006	563.64	285.37	155.705	88.167	52.423	53.023
BTA28	1361.1	1041.2	527.72	250.57	136.335	74.822	60.411	43.554
BTA29	1145.2	867.6	438.32	246.436	140.81	82.802	58.56	47.64
X	2343	1778	881.6	416.2	245.48	117.37	64.229	55.048

表 3 各染色体不同物理间距的 r^2 分布情况

染色体	SNP 间距(kb)							
	100	200	500	1 000	2 000	5 000	10 000	15 000
BTA1	0.211±0.241	0.151±0.179	0.114±0.128	0.098±0.109	0.088±0.096	0.062±0.066	0.047±0.049	0.039±0.039
BTA2	0.176±0.217	0.146±0.171	0.111±0.129	0.096±0.1	0.091±0.102	0.065±0.063	0.049±0.048	0.038±0.034
BTA3	0.179±0.209	0.137±0.164	0.097±0.103	0.086±0.092	0.074±0.081	0.057±0.061	0.047±0.049	0.040±0.04
BTA4	0.156±0.187	0.142±0.166	0.095±0.124	0.093±0.105	0.078±0.09	0.061±0.064	0.044±0.044	0.039±0.036
BTA5	0.199±0.231	0.130±0.175	0.115±0.124	0.091±0.105	0.077±0.085	0.061±0.064	0.044±0.044	0.034±0.033
BTA6	0.201±0.231	0.148±0.181	0.106±0.111	0.094±0.098	0.081±0.086	0.060±0.063	0.040±0.039	0.035±0.033
BTA7	0.182±0.199	0.129±0.169	0.113±0.12	0.090±0.105	0.076±0.082	0.055±0.053	0.039±0.035	0.031±0.03
BTA8	0.196±0.227	0.137±0.159	0.105±0.114	0.088±0.095	0.074±0.077	0.056±0.059	0.042±0.042	0.035±0.029
BTA9	0.157±0.208	0.120±0.165	0.089±0.102	0.086±0.098	0.072±0.077	0.056±0.064	0.044±0.041	0.033±0.029
BTA10	0.192±0.223	0.141±0.153	0.099±0.095	0.108±0.12	0.095±0.106	0.069±0.073	0.049±0.049	0.039±0.038
BTA11	0.168±0.195	0.130±0.149	0.121±0.125	0.090±0.1	0.082±0.086	0.065±0.07	0.046±0.046	0.035±0.035
BTA12	0.158±0.192	0.113±0.14	0.104±0.117	0.082±0.092	0.071±0.078	0.048±0.045	0.035±0.029	0.032±0.029
BTA13	0.210±0.229	0.155±0.169	0.120±0.133	0.101±0.106	0.079±0.086	0.056±0.06	0.045±0.055	0.030±0.025
BTA14	0.202±0.194	0.162±0.172	0.111±0.113	0.110±0.116	0.081±0.079	0.056±0.059	0.044±0.046	0.032±0.028
BTA15	0.152±0.195	0.124±0.145	0.102±0.118	0.079±0.091	0.064±0.067	0.052±0.052	0.036±0.032	0.031±0.026
BTA16	0.199±0.242	0.130±0.15	0.094±0.095	0.092±0.102	0.074±0.082	0.056±0.058	0.046±0.051	0.035±0.031
BTA17	0.187±0.209	0.117±0.155	0.120±0.125	0.0930±0.1	0.084±0.095	0.060±0.064	0.043±0.039	0.032±0.028
BTA18	0.168±0.185	0.127±0.147	0.073±0.064	0.072±0.076	0.059±0.057	0.047±0.049	0.033±0.028	0.027±0.023
BTA19	0.164±0.194	0.141±0.176	0.077±0.084	0.079±0.082	0.064±0.064	0.046±0.047	0.030±0.025	0.028±0.021
BTA20	0.177±0.213	0.153±0.167	0.130±0.146	0.112±0.123	0.091±0.1	0.071±0.078	0.049±0.048	0.036±0.03
BTA21	0.167±0.203	0.127±0.158	0.080±0.077	0.080±0.087	0.072±0.078	0.052±0.051	0.036±0.033	0.032±0.031
BTA22	0.153±0.15	0.123±0.14	0.106±0.133	0.094±0.106	0.072±0.079	0.048±0.049	0.037±0.03	0.030±0.025
BTA23	0.178±0.211	0.144±0.143	0.127±0.157	0.088±0.094	0.085±0.086	0.062±0.072	0.047±0.049	0.038±0.04
BTA24	0.187±0.206	0.136±0.159	0.092±0.107	0.084±0.094	0.073±0.074	0.052±0.05	0.039±0.037	0.032±0.028
BTA25	0.131±0.131	0.102±0.118	0.104±0.122	0.077±0.079	0.070±0.077	0.049±0.043	0.032±0.029	0.024±0.018
BTA26	0.199±0.198	0.118±0.126	0.108±0.145	0.095±0.104	0.083±0.09	0.052±0.056	0.038±0.038	0.030±0.025
BTA27	0.136±0.145	0.101±0.106	0.075±0.078	0.074±0.08	0.068±0.069	0.049±0.049	0.042±0.04	0.028±0.026
BTA28	0.143±0.177	0.098±0.111	0.079±0.096	0.083±0.099	0.077±0.081	0.057±0.05	0.036±0.029	0.033±0.029
BTA29	0.166±0.173	0.116±0.122	0.094±0.104	0.084±0.087	0.074±0.088	0.052±0.049	0.037±0.032	0.031±0.023
X	0.162±0.182	0.113±0.1	0.093±0.115	0.098±0.108	0.084±0.098	0.071±0.078	0.066±0.067	0.052±0.052

3 讨论

作为世界上产奶量最高的奶牛品种, 荷斯坦牛经历了长期的较强的人工选择历程, 尤其是人工授精技术的推广使目前商业奶牛的有效群体相对较小, 大概在 50 头左右^[8]。根据系谱和近交系数, Sorensen 等^[4]对出生于 1999~2003 年的丹麦荷斯坦牛群体共 2 238 349 头进行有效群体大小估计, 结果表明丹麦荷斯坦牛的有效群体大小约为 50 头。随着高密度 SNP 芯片的出现, 其越来越被普遍作为估计有效群体大小的有力工具。Kim 和 Kirpatrick^[2]利用 7 117

个分布于全基因组的 SNP 标记, 根据 200 头公牛估计北美荷斯坦牛当前世代的有效群体规模小于 100 头; Sargolzaei 等^[8]利用 Affymetrix 公司出品的 10K SNP 芯片根据 821 头公牛进一步证明了 Kim 和 Kirpatrick 的研究结果。Qanbari 等^[9]利用牛 50K SNP 芯片数据和 810 头牛(469 头公牛和 341 头公牛母亲)估计出德国荷斯坦牛 4 个世代前的有效群体规模约为 103 头。本研究利用牛全基因组高密度 50K SNP 芯片估计出 4 世代前中国荷斯坦牛的有效群体为 39~51 头。由于群体的差异, 各国的有效群体大小有所区别, 但都反映出在近代各国荷斯坦牛有效

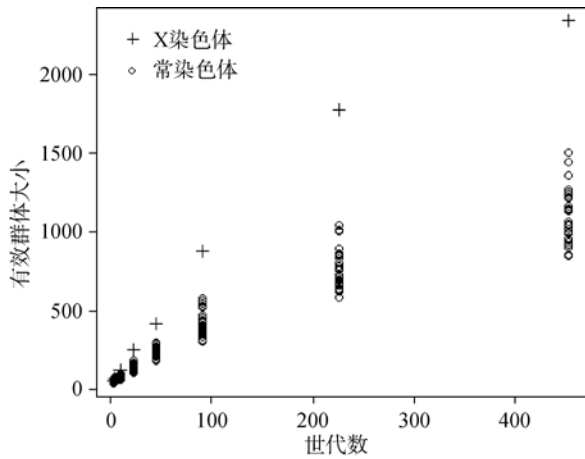


图 3 中国荷斯坦牛历史(前 452~前 4 世代)有效群体大小

群体下降至很小的规模。

有效群体大小的估计方法很多, 根据不同的算法和不同的信息来源, 群体有效规模的估计值可能存在较大的差异^[23]。Hill等^[15]认为, 利用LD信息估计有效群体时, 样本大小与估计的准确度关联显著。本研究利用的北京地区荷斯坦牛群体有一定的局限性, 不能充分代表全国荷斯坦牛群体情况, 有效群体大小估计中使用的 1 698 头母牛来自 14 头公牛家系, 相比其他研究^[7-9]的数百头公牛, 本研究使用的实际有效样本数目偏小, 导致本研究估计的有效群体大小与其他报道相比偏小。目前有效群体大小的估计是基于理想的数学模型进行的, 物种或者品种在进化、育种过程中的实际情况往往要复杂的多, 所以本研究的估计值只是一个尽可能准确的参考值, 要更准确的估计有赖于数学模型的优化和更好的试验群体。

相比其他国家荷斯坦牛群体, 中国荷斯坦牛育成历史并不长, 正式成为一个品种仅几十年的时间, 但为与其他研究进行比较, 本研究根据选取了 8 种不同长度的染色体片段 0.1、0.2、0.5、1、2、5、10 和 15 Mb, 估计前 452 世代到前 4 世代的中国荷斯坦牛有效群体大小, 反映中国荷斯坦牛有效群体的变化趋势。染色体片段的选择基于以下考虑: 本研究所采用的芯片的密度约为 50 kb, 但是标记并不是均匀分布于整个基因组上, 为了保证计算的准确性, 选取 0.1 Mb 作为最小片段; 对距离较远的 SNP 标记估计的 LD 偏差较大, 考虑到估计的准确性, 选取 15 Mb 作为最大片段。

本研究为与其他研究进行比较, 选取了上述的 8 种不同长度的染色体片段长度。实际上, 其他世代的有效群体大小也可以根据标记间的连锁不平衡进行估计, 如根据间距更小的 SNP 标记间的连锁不平衡, 可以估计 2000 世代前的有效群体大小。由于遗传距离为 c 的染色体片段可以近似反映出前 $1/(2c)$ 世代的有效群体大小, 因此给定所要估计的世代, 即可近似计算出相应的遗传距离, 进而根据公式(3)得到需要的染色体片段物理长度, 然后根据 haploview 计算所得的该物理长度下所有 SNP 标记对的 r^2 , 由于 SNP 标记并不是均匀分布, 本研究选取物理长度 $*(1\pm 5\%)$ 作为 SNP 标记对选取标准, 结合公式(1)即可估计出该世代下有效群体大小。

本研究仅估计了 4 世代前的有效群体大小, 对于更有现实意义的当前世代有效群体大小估计没有涉及。主要是当前世代下根据标记连锁不平衡估计有效群体大小准确性较差, 比如 1 世代下标记间遗传距离为 0.5 摩尔根, 标记间几乎为自由重组(重组率近似为 0.50), 处于连锁平衡状态, 连锁不平衡程度几乎为 0, 而利用高密度标记估计有效群体大小主要依赖于标记间的连锁不平衡, 因此极弱的连锁不平衡程度将严重影响到有效群体大小估计的准确性。而 0 世代(当前世代)所要求的标记间距更大, 这时估计的有效群体大小准确性更低。因此考虑到估计的准确性, 本研究选取 15 Mb 作为最大片段长度, 此时中国荷斯坦牛群体平均 r^2 为 0.03, 反映出前 4 世代时中国荷斯坦牛的有效群体大小, 对当前世代有效群体大小也有一定的指示作用。

用不同染色体上的 SNP 估计出的有效群体大小差异显著, 从有效群体大小的计算公式可以看出, SNP 标记间的 LD 程度和遗传距离均会对有效群体大小的估计产生影响, 而 SNP 的遗传距离是一定的, 因此染色体间有效群体估计的显著差异主要是由染色体间不同的 LD 程度造成。由此可以推测出, 影响 LD 的因素也能影响有效群体大小, 例如人工选择、迁移、突变、漂变等。以人工选择为例, 人们在对牛群进行选择时, 不管是按照表型选择还是育种值选择, 会提高群体中优势基因型的比例, 导致人为高强度的 LD, 从而影响有效群体大小。本研究结果显示, 根据 6 和 14 号染色体的连锁不平衡信息估计出的中国荷斯坦牛有效群体较其他染色体小(表 2),

而这两条染色体是目前报道携带数量性状基因座位 (QTL) 数目最多的染色体; 畜禽 QTLdb 数据库 (<http://www.genome.iastate.edu/cgi-bin/QTLdb/BT/browse>) 显示, 到目前为止, 分别有 424 和 361 个 QTL 定位在 6 和 14 号染色体上, 其中在多个群体中得到验证的与乳脂相关的 *DGATI* 基因就位于 14 号染色体。由于 6 和 14 号染色体受到较强的选择, 提高了其连锁不平衡程度, 致使估计的有效群体较小。

本研究也表明根据 X 染色体估计出的有效群体较常染色体大, 这与其他报道一致^[8], 一方面可能由于 50K SNP 芯片 X 染色体上的 SNP 密度相对较低, 导致估计偏差, 需要更高密度的芯片来进一步验证; 另一方面可能与目前 X 染色体估计有效群体大小方法有关, 虽然目前的估计方法已经考虑了 X 染色体与常染色体的差异, 但基于 X 染色体的特殊性, 其估计方法还需要进一步研究^[8]。

估计有效群体大小的方法很多, 传统的有效含量估计方法是建立在个体间相互独立的随机交配群体基础上的, 且需要找到本群体中所有独立的公畜和母畜, 这在实际操作中很难做到。而畜禽基于系谱信息的有效群体大小估计方法常常由于普遍存在的系谱错误导致估计的准确性下降。通过 DNA 序列检测等分子手段往往不适合大规模的检测, 其该方法估计出的有效群体大小很大程度上依赖于染色体区段的选择。相对而言, 利用连锁不平衡信息估计有效群体大小作为一种新方法, 可以不考虑系谱, 仅通过全基因组的高密度标记信息, 即能够更准确估计有效群体大小, 且根据不同的染色体片段长度可以估计出不同世代下有效群体大小, 直观地反映出整个育种过程中有效群体大小的变化趋势。但该方法对连锁不平衡程度依赖较大, 连锁不平衡度量准确性、标记间的连锁不平衡程度及标记密度对该方法估计准确性都有影响, 同时本研究发现的远距离存在较强连锁不平衡是否对有效群体大小估计准确性产生影响, 以及这些标记是否应该在有效群体大小估计中加以考虑都需要在进一步研究中探讨。由于无法获知真实的有效群体大小, 通过数据模拟进行此类研究不失为一种有效研究手段。

本研究表明中国荷斯坦有效群体呈减少的趋势, 在其他研究中, 也发现荷斯坦牛有效群体不断减少的类似趋势^[7-9], 而有效群体的减小意味着遗传变

异的减小, 进而影响到遗传进展, 因此控制合理的有效群体规模十分重要。在动物育种中, FAO 建议群体的有效群体大小最好维持在 50~100 头^[4]。虽然本研究结果尚不能反映中国荷斯坦牛群体的真实情况, 但对今后我国荷斯坦牛的育种提供了一个参考, 奶牛育种工作者在育种规划中应该采取相应的措施避免中国荷斯坦有效群体的进一步减小, 以保持中国荷斯坦牛遗传进展的持续性, 这也是其他国家正在开展的工作。

参考文献(References):

- [1] Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet*, 2001, 17(9): 502-510. DOI
- [2] Falconer DS, CMackay TFC. Introduction to Quantitative Genetics. Harlow: Addison Wesley Longman, 1996. DOI
- [3] Wright S. Size of population and breeding structure in relation to evolution. *Science*, 1938, 87(2263): 430-431. DOI
- [4] Sorensen AC, Sørensen MK, Berg P. Inbreeding in Danish dairy cattle breeds. *J Dairy Sci*, 2005, 88(5): 1865-1872. DOI
- [5] Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM. Recent human effective population size estimated from linkage disequilibrium. *Genome Res*, 2007, 17(4): 520-526. DOI
- [6] Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res*, 2003, 13(4): 635-643. DOI
- [7] Kim ES, Kirkpatrick BW. Linkage disequilibrium in the North American Holstein population. *Anim Genet*, 2009, 40(3): 279-288. DOI
- [8] Sargolzaei M, Schenkel FS, Jansen GB, Schaeffer LR. Extent of linkage disequilibrium in Holstein cattle in North America. *J Dairy Sci*, 2008, 91(5): 2016-2117. DOI
- [9] Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H. The pattern of linkage disequilibrium in German Holstein cattle. *Anim Genet*, 2010, 41(4): 346-356. DOI
- [10] Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol*, 1971, 2(2): 125-141. DOI
- [11] Hill WG. Linkage disequilibrium among multiple neutral alleles produced by mutation in finite population. *Theor Popul Biol*, 1975, 8(2): 117-126. DOI
- [12] McVean GAT. A genealogical interpretation of linkage

- disequilibrium. *Genetics*, 2002, 162(2): 987-991. [DOI](#)
- [13] Weir BS, Hill WG. Effect of mating structure on variation in linkage disequilibrium. *Genetics*, 1980, 95(2): 477-488. [DOI](#)
- [14] Lewontin RC. The interaction of selection and linkage. I. General considerations; Heterotic models. *Genetics*, 1964, 49(1): 49-67. [DOI](#)
- [15] Hill WG. Estimation of linkage disequilibrium in randomly mating populations. *Heredity*, 1974, 33(2): 229-239. [DOI](#)
- [16] Zhao H, Nettleton D, Dekkers JCM. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. *Genet Res*, 2007, 89(1): 1-6. [DOI](#)
- [17] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 2005, 21(2): 263-265. [DOI](#)
- [18] Farnir F, Coppieiers W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M. Extensive genome-wide linkage disequilibrium in cattle. *Genome Res*, 2000, 10(2): 220-227. [DOI](#)
- [19] de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*, 2008, 179(3): 1503-1512. [DOI](#)
- [20] Mueller JC, Löhmußaar E, Mägi R, Remm M, Bettecken T, Lichtner P, Biskup S, Illig T, Pfeufer A, Luedemann J, Schreiber S, Pramstaller P, Pichler I, Romeo G, Gaddi A, Testa A, Wichmann HE, Metspalu A, Meitinger T. Linkage disequilibrium patterns and tagSNP transferability among European populations. *Am J Hum Genet*, 2005, 76(3): 387-398. [DOI](#)
- [21] Smith EM, Wang X, Littrell J, Eckert J, Cole R, Kissebah AH, Olivier M. Comparison of linkage disequilibrium patterns between the HapMap CEPH samples and a family-based cohort of Northern European descent. *Genomics*, 2006, 88(4): 407-414. [DOI](#)
- [22] Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Cavanagh JAL, Barris W, Schnabel RD, Taylor JF, Raadsma HW. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics*, 2008, 9(1): 187. [DOI](#)
- [23] Wang JL. Estimation of effective population sizes from data on genetic markers. *Philos Trans R Soc Lond B Biol Sci*, 2005, 360(1459): 1395-1409. [DOI](#)

•科学新闻•

干细胞研究为遗传肾病治疗带来新希望

当前,社会上对于干细胞研究既有争议又有支持,这完全见仁见智。最近,澳洲蒙纳士大学(Monash University)免疫学和干细胞实验室(MISCL)的科学家证明,他们可以从健康的成年人肾脏细胞中培育出干细胞,从而规避了这个研究的伦理问题。沙龙·里卡多副教授及其研究团队将利用这一研究成果,在实验室中建立遗传性肾脏疾病的研究模型,并梳理出治疗机制,控制这些难以治疗的疾病。

“我们利用人类肾细胞进行研究,并让时光倒流,去找回前体”,里卡多说。这个在美国《肾脏病学杂志》发表的独具挑战性的项目是由蒙纳士大学的研究人员通过把健康成人肾细胞重新编程回到胚胎样状态的肾脏干细胞,然后与现成的胚胎干细胞比较。结果表明,两者都可能形成不同的胚胎组织类型,而其遗传特性不变。“这些肾脏细胞已被重新编程回到胚胎样状态,这使我们转变肾脏细胞前体为各种类型肾脏细胞成为可能。”通过与蒙纳士医学中心的彼得·克尔教授合作,该研究小组现在已从患有多囊肾、遗传性肾炎这两个主要的遗传型肾脏疾病患者的肾脏细胞中产生了4个干细胞株。

“由于这些干细胞可以在培养皿中无限地分裂,我们可产生病人特异性干细胞的无限来源,并利用这些肾细胞来产生囊肿并筛选药物,”副教授里卡多说。“我们的最终目标是提供现成的,成熟的肾脏细胞,以进行药物检测和疾病建模。”

里卡多副教授正在与制药公司接洽,希望通过肾脏干细胞来筛选药物,她认为这种个性化的医药验证方法将在未来产生更安全的药物。但在短期,澳洲蒙纳士大学研究人员计划继续研究环境因素对肾脏干细胞性能的影响,测试不同药物在减少肾囊肿的形成和细胞增殖的效能,并更好地了解遗传肾病是如何发展的。

英文原稿及图片请阅: <http://www.monash.edu.au/news/show/winding-back-the-clock-with-kidney-stem-cells>

(吕欢)