

伴性基因遗传的计算机模拟

赵冬生 (首都师范大学计算机系 100037)

赵银荣 毛盛贤 (首都师范大学生物系 100037)

摘要:本文用 VISUAL BASIC 语言,在群体水平上对伴性基因在两性中的遗传传递规律和达到稳定平衡的过程进行了计算机模拟程序设计,为遗传学的科研和应用提供了一快捷的方法。

关键词:伴性遗传 遗传平衡 随机函数 动态数据链接

对伴性遗传的规律进行研究,是发展生产和预防人类遗传疾病的重要途径。

本文目的在于:阐述利用 VISUAL BASIC 语言,在群体水平上对伴性基因的遗传进行计算机模拟的过程,从而可得出有效的研究依据、掌握变化规律,为遗传学的研究交流和应用提供一新的手段。

1. 伴性基因的遗传模型

位于性染色体上的基因为伴性基因。以人为例,女性和男性分别称为同配性别(XX)和异配性别(XY)。假定在一开始群体中性染色体上一基因座有两种基因 A 和 a,且女性中的基因型 AA、Aa 和 aa 的频率分别为 P_f 、 H_f 和 Q_f ,男性中的基因型 A(Y)和 a(Y)的频率分别为 P_m 和 Q_m ,则该群体中女性的基因 A 和 a 的频率分别为:

$$P_f = P_f + \frac{1}{2}H_f; \quad q_f = 1 - P_f$$

男性的基因 A 和 a 的频率分别为:

$$P_m = P_m; \quad q_m = 1 - P_m$$

两相邻世代(n)代和(n+1)代的基因频率之间的关系为:

$$P_{f(n+1)} = \frac{1}{2}(P_{f(n)} + P_{m(n)}); \quad q_{f(n+1)} = 1 - P_{f(n+1)}$$

$$P_{m(n+1)} = P_{f(n)}; \quad Q_{m(n+1)} = 1 - P_{m(n+1)}$$

群体达到平衡时的基因 A 和 a 的频率:

$$\bar{P} = \hat{P}_f = \hat{P}_m = \frac{2}{3}P_f + \frac{1}{3}P_m; \quad \bar{q} = \hat{q}_f = \hat{q}_m = 1 - \bar{P}$$

群体平衡时,在女性和男性中的基因型频率分别为:
 $P^2(AA), 2pq(Aa), q^2(aa); P[A(Y)], q[a(Y)]$.

2. 模拟过程

对上述伴性遗传模型的模拟,我们主要通过以下步骤完成:

①数据输入:分两种情况:一是根据测得被研究群体中的雌性和雄性的基因型频率,通过对话框输入;二是对任一群体,通过随机函数模拟产生群体中的雌性和雄性

基因型频率。随机函数的使用,使我们可以对群体中产生各种可能的基因频率在世代传递过程的变化进行观察和分析。(注:本文图例数据由随机函数产生。)

②数据计算:以输入群体的基因型频率为基础,模拟随机交配过程。根据前述的遗传模型,可实现计算各世代雌、雄性的基因型频率、基因频率,以及群体达到平衡(平衡两性等位基因频率差 ≤ 0.0005)所需时间。通过数组存储模拟结果,在窗体中显示输出。这种模拟效果可供研究者全面、准确地掌握两性基因频率及基因型频率在各世代中的数值及其变化规律。

③数据传送:利用动态数据链接,将计算出的结果,通过“LINKPOKE”和“LINKREQUEST”命令把数据传送到 EXCEL 的工作簿中的相应位置,然后进行数据处理和绘图。

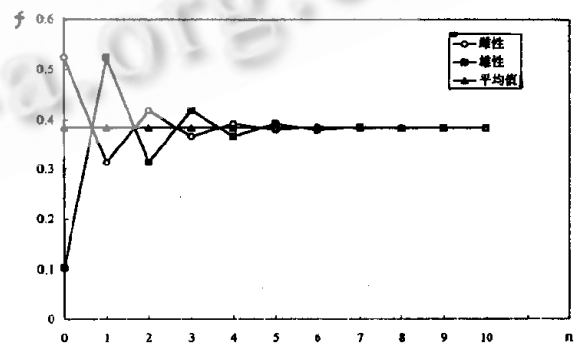


图 1 伴性基因频率 f 在两性中的世(n)代变化

④绘图:这方面的工作,主要是利用在模拟过程中产生的数据,绘制成伴性基因频率在两性中的世代变化折线图。通过图 1,我们可以看到群体内雌雄两性基因频

率在世代间的波动。由于两性中的基因频率差逐代减半,可以看到波动的幅度逐代递减,几代以后渐趋平衡,达到平衡后,两性的基因频率相等,从而处于稳定状态。图中的水平线代表着整个群体平均基因频率。从中我们可以看出,在两性基因频率变化的过程中,雄性基因频率总是追随雌性基因频率,时间晚一代,但就整个群体平均基因频率而言,始终没有变化。

同样,我们利用雌雄两性的隐性基因为表现型时的数据生成相应分布图,如图2所示。从该图中我们可以清楚地看到,由于雄性为异配性别(XY),而Y染色体上又很少携带基因,所以群体内2/3的性连锁基因由雌性个体(同配性别)所携带。因此,即使平衡时雌雄两性的基因频率相同,呈现隐性表现型的个体在雌雄性中的比例也是不同的。在雌性中该比例是 q^2 ,而在雄性中则是 q 。由于 $0 \leq q \leq 1$, q 一般总是大于 q^2 , q 值越小,二者的差异越大。在该图,雌性的基因型为 $q^2 = 0.14$,而雄性却为 $q = 0.38$ 。隐性伴性遗传的红绿色盲、血友病等表现男性比女性多得多,就是这个原因。

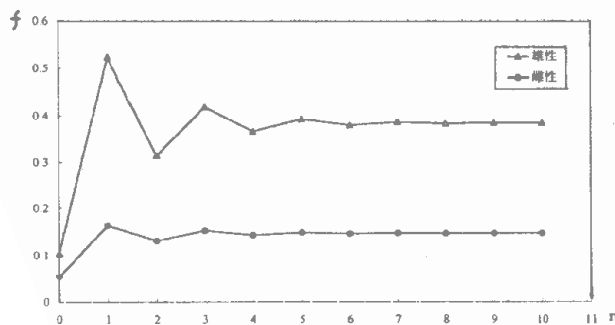


图2 雌性和雄性隐性基因型频率 f 随世代(n)的变化

3. 模拟环境与特点

该程序设计是在 Windows 下使用 VISUAL BASIC 进行编程,运用动态数据链接和 EXCEL 进行数据及图形交换,由 VB 窗体显示执行结果。运用鼠标控件的点、拖、拉功及非动态链接技术激活共享资源。利用“SEND-KEYS”命令,使我们通过程序设计实现对其他应用程序的调用,既节省了内存,又保持界面清晰、完整。再一点是我们在程序设计中完成了对数据和图形的编辑功能,例如要保证对数据的准确传送,我们在程序中设计了拷贝、粘贴等操作;对图形的图例、标题及数轴等的设计,也

是通过程序设计,使之一步到位,生成图形界面,达到使实验结果一目了然的效果。利用了 VB 语言中的控件功能,通过击点各窗体或命令钮来控制程序的执行及界面的切换。

4. 该模拟系统的应用

(1) 只要知道初始世代两性中的基因频率,就可知道群体中各世代两性基因频率和基因型频率的变化情况,以及知道要经过几个世代群体才能达到平衡和平衡时的基因频率、基因型频率。

(2) 平衡时若为隐性伴性遗传,具隐性特性的雄性基因型频率(或雄性个体数)明显大于相应的雌性基因型频率(或雌性个体数),如图2所示。由于人类红绿色盲和血友病菌均属隐性伴性遗传,如果色盲基因频率在男性中为0.07,则女性中为 $(0.07)^2 = 0.0049$,即平衡时男性色盲人数约为女性的14倍,我国实际调查的结果正是如此。该模拟系统的使用,可供有关研究人员参考各类相关数据及其分布。

(3) 利用该系统也可模拟显性伴性遗传。如前述的人类抗维生素D佝偻病,属此类遗传。如果模拟,它将表明显性伴性遗传疾病,女性患者应约为男性患者的2倍。

(4) 可追溯群体的历史。假定一个物种的群体C是由同一物种的群体A和B组成的,但不知是从何时开始组建的。这时,可假定C群体某伴性基因的频率(假定为 p);再从群体A和B群体测得相应的伴性基因的频率,并输入该模拟系统,就可发现某伴性基因频率达到 p 时所需要的世代数(假定为30个世代)。假设该物种一个世代为20年,则群体C是600年前建立起来的。也就是说,伴性基因遗传可用来研究生物进化史。

5. 结论

通过计算机模拟使得遗传学上的实验变为可视,打破了在生命现象研究的时、空界限。执行该程序的重要意义是在观察两性的伴性基因的世代变化的同时,也可对由此产生的一系列问题进行更深入的研究。

参考文献

- [1] 毛盛贤、黄远樟. 群体遗传及其程序设计. 北京师范大学出版社
- [2] VISUAL BASIC FOR WINDOWS 程序设计 林启隆编 清华大学出版社
- [3] EXCEL 5.0 应用开发技术精解 ERIC WELLS 著 清华大学出版社

(来稿时间:1998年4月)