



(12) 发明专利

(10) 授权公告号 CN 101687901 B

(45) 授权公告日 2013. 10. 09

(21) 申请号 200880022259. 9

代理人 陈文平 尚继栋

(22) 申请日 2008. 05. 16

(51) Int. Cl.

(30) 优先权数据

60/930, 609 2007. 05. 17 US

C07H 21/02 (2006. 01)

A01H 1/00 (2006. 01)

C12Q 1/68 (2006. 01)

(85) PCT申请进入国家阶段日

2009. 12. 28

(56) 对比文件

US 20060141495 A1, 2006. 06. 29, 全文.

(86) PCT申请的申请数据

PCT/US2008/006321 2008. 05. 16

审查员 王少华

(87) PCT申请的公布数据

W02008/143993 EN 2008. 11. 27

(73) 专利权人 孟山都技术公司

地址 美国密苏里州

(72) 发明人 吴坤生 J·莱德奥克斯

D·巴特鲁伊尔 A·格普塔

R·约翰森 S·伊汀顿 J·布尔

M·爱德华兹

(74) 专利代理机构 北京市金杜律师事务所

11256

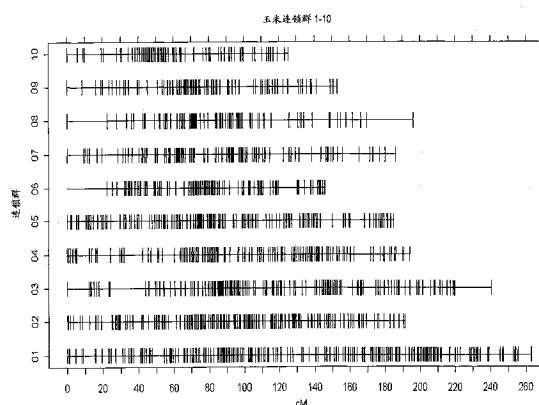
权利要求书9页 说明书36页 附图2页

(54) 发明名称

玉米多态性与基因分型方法

(57) 摘要

本发明提供用于至少两个玉米品种之间基因分型的多态性玉米 DNA 基因座。该基因座的序列可用于为设计用于检测玉米 DNA 多态性的引物和探针寡核苷酸提供基础。多态性可用于玉米中的基因分型应用。多态性标记可用于建立标记 / 性状关联, 例如, 在连锁不平衡作图和关联研究、定位克隆和转基因应用、标记辅助的育种和标记辅助的选择、杂种预测和血缘一致性研究中。多态性标记也可用于 DNA 克隆文库的作图, 例如, 用于与多态性连锁的玉米 QTLs 和基因。



1. 一种核酸分子文库,所述文库包括至少两组不同的核酸分子,其中,所述不同组核酸分子中的每一组允许对表 3 中确定的相应的玉米基因组 DNA 多态性进行分型。

2. 如权利要求 1 所述的文库,其中,所述不同组的核酸分子排列在至少一个固体载体上或至少一个微量滴定板上。

3. 如权利要求 2 所述的文库,其中,所述不同组核酸分子中的每一组位于所述微量滴定板的单独的和不同的孔中。

4. 如权利要求 2 所述的文库,其中,所述不同组核酸分子中的每一组位于所述固体载体上的不同的探测位置处。

5. 如权利要求 1 所述的文库,其中,所述核酸分子组合在单一混合物中。

6. 如权利要求 1 所述的文库,其中,所述不同组核酸分子中的每一组包括至少 12 个连续核苷酸的核酸分子,该核苷酸包括或直接邻近表 3 中确定的相应的多态性,并且其中至少 12 个连续核苷酸的该序列与包括或直接邻近所述多态性的玉米 DNA 片段任一链中相同数目核苷酸的序列至少 90% 相同。

7. 如权利要求 6 所述的文库,其中,所述核酸分子是至少 15 个连续核苷酸的核酸分子。

8. 如权利要求 7 所述的文库,其中,所述核酸分子是至少 18 个连续核苷酸的核酸分子。

9. 如权利要求 1 所述的文库,其中,所述核酸分子进一步包括可检测的标记或提供可检测的标记的掺入。

10. 如权利要求 9 所述的文库,其中,所述可检测的标记选自同位素、荧光团、氧化剂、还原剂、核苷酸和半抗原。

11. 如权利要求 10 所述的文库,其中,所述可检测的标记通过化学反应添加到核酸上,或通过酶促反应掺入。

12. 如权利要求 1 所述的文库,其中,所述每个不同组的核酸分子包括:

a. 一对寡核苷酸引物,其中,所述寡核苷酸引物中的每一个包括至少 15 个核苷酸碱基,并允许 PCR 扩增包含表 3 中确定的所述相应多态性之一的 DNA 片段,和

b. 至少一种检测核酸分子,其允许检测 (a) 中所述扩增片段中的多态性。

13. 如权利要求 12 所述的文库,其中,所述检测核酸包含至少 12 个核苷酸碱基,或包含至少 12 个核苷酸碱基和可检测的标记,并且其中,所述检测核酸分子的序列与包括所述多态性的权利要求 1 的基因座中玉米 DNA 片段任一链中相同数目连续核苷酸的序列至少 95% 相同。

14. 如权利要求 1 所述的文库,其中,所述文库包括至少 8 组不同的核酸分子,其中,所述每组分子允许对表 3 中确定的相应的不同玉米基因组 DNA 多态性进行分型。

15. 如权利要求 14 所述的文库,其中,所述文库包括至少 24 组不同的核酸分子,其中,所述每组分子允许对表 3 中确定的相应的不同玉米基因组 DNA 多态性进行分型。

16. 如权利要求 15 所述的文库,其中,所述文库包括至少 96 组不同的核酸分子,其中,所述每组分子允许对表 3 中确定的相应的不同玉米基因组 DNA 多态性进行分型。

17. 如权利要求 16 所述的文库,其中,所述文库包括至少 384 组不同的核酸分子,其中,所述每组分子允许对表 3 中确定的相应的不同玉米基因组 DNA 多态性进行分型。

18. 如权利要求 1 所述的文库,其中,表 3 中确定的所述相应的不同玉米基因组 DNA 多态性选自 SEQ ID NO:2468、5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、

2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279。

19. 如权利要求 1 所述的文库,其中,选择所述不同组的核酸分子,用于鉴定在待基因分型的群体中预测为多态性的相应的不同玉米基因组 DNA 多态性。

20. 一种分离的核酸分子,其用于检测代表玉米 DNA 中的多态性的分子标记,其中,所述核酸分子包含至少 15 个核苷酸,所述核苷酸包括或直接邻近所述多态性,并且其中,所述核酸分子与包括或直接邻近所述多态性的 DNA 任一链中相同数目连续核苷酸的序列至少 90% 相同,并且其中,所述多态性是在表 3 中所确定的。

21. 如权利要求 20 所述的分离的核酸,其中,所述核酸进一步包含可检测的标记或提供可检测的标记的掺入。

22. 如权利要求 21 所述的分离的核酸,其中,所述可检测的标记选自同位素、荧光团、氧化剂、还原剂、核苷酸和半抗原。

23. 如权利要求 22 所述的分离的核酸,其中,所述可检测的标记通过化学反应添加到核酸上,或者通过酶促反应掺入。

24. 如权利要求 20 所述的分离的核酸,其中,所述表 3 中的多态性选自 SEQ ID NO: 2468、5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279。

25. 一组寡核苷酸,其包含:

a. 一对寡核苷酸引物,其中,所述引物中的每一个包含至少 12 个连续核苷酸,并且其中,所述引物对允许 PCR 扩增包含表 3 中确定的玉米基因组 DNA 多态性的 DNA 片段;和

b. 至少一种检测寡核苷酸,其允许检测所述扩增片段中的多态性,其中,所述检测寡核苷酸的序列与包括或直接邻近步骤 (a) 的所述多态性的玉米 DNA 片段任一链中相同数目连续核苷酸的序列至少 95% 相同。

26. 如权利要求 25 所述的一组寡核苷酸,其中,所述检测核酸包含至少 12 个核苷酸,并且提供可检测标记的掺入或进一步包含可检测的标记。

27. 如权利要求 26 所述的一组寡核苷酸,其中,所述可检测的标记选自同位素、荧光团、氧化剂、还原剂、核苷酸和半抗原。

28. 一种对玉米植物进行基因分型以选择具有有利性状的植物的方法,所述方法包括以下步骤:

a. 检测来自至少两个具有等位基因 DNA 的玉米植物的组织中的 DNA 或 mRNA;

b. 对于来自步骤 (a) 的所述组织,确定表 3 中确定的至少一种玉米基因组 DNA 多态性的等位基因状态;和

c. 选择具有更有利的相关性状组合的植物。

29. 如权利要求 28 所述的基因分型方法,其中,所述玉米基因组 DNA 多态性是表 3 中确定的定位的多态性。

30. 如权利要求 28 所述的方法,其中,通过允许鉴定单核苷酸多态性的试验确定所述玉米基因组 DNA 多态性的所述等位基因状态。

31. 如权利要求 30 所述的方法,其中,所述试验选自单碱基延伸 (SBE)、等位基因特异性引物延伸测序 (ASPE)、DNA 测序、RNA 测序、基于微阵列的分析、通用 PCR、等位基因特异性延伸、杂交、质谱法、连接、延伸-连接和瓣核酸内切酶介导的试验。

32. 如权利要求 28 所述的方法,其中,确定表 3 中确定的至少 8 种不同的玉米基因组 DNA 多态性的等位基因状态。

33. 如权利要求 32 所述的方法,其中,确定表 3 中确定的至少 48 种不同的玉米基因组 DNA 多态性的等位基因状态。

34. 如权利要求 33 所述的方法,其中,确定表 3 中确定的至少 96 种不同的玉米基因组 DNA 多态性的等位基因状态。

35. 如权利要求 34 所述的方法,其中,确定表 3 中确定的至少 384 种不同的玉米基因组 DNA 多态性的等位基因状态。

36. 如权利要求 35 所述的方法,进一步包括利用步骤 (b) 的所述等位基因状态确定情况来选择用于育种的亲本植物、后代植物或测试植物的步骤。

37. 如权利要求 28 所述的方法,进一步包括在计算机可读介质上存储所述一种或多种等位基因状态确定情况产生的基因型数据的步骤。

38. 如权利要求 37 所述的方法,进一步包括将一个玉米植物与另一个玉米植物的所述基因型数据进行比较的步骤。

39. 如权利要求 37 所述的方法,进一步包括将至少一种所述玉米植物的所述基因型数据与表型性状数据或表型性状指数数据进行比较的步骤。

40. 如权利要求 37 所述的方法,进一步包括将至少两种所述玉米植物的基因型数据与表型性状数据或表型性状指数数据进行比较,并确定所述基因型数据和所述表型性状数据之间的一种或多种关联的步骤。

41. 如权利要求 42 所述的方法,其中,确定所述表型性状数据或表型性状指数数据与所述基因型性状数据之间的关联,并且其中,所述基因型性状数据包括至少 10 种定位的表 3 中确定的玉米基因组 DNA 多态性的等位基因状态确定情况。

42. 一种选择亲本、后代或测试植物的方法,包括以下步骤:

a) 在至少第一和第二玉米近交系中,确定多种表 3 中确定的多态性与多种性状之间的关联;

b) 确定亲本、后代或测试植物中的一种或多种多态性的等位基因状态;

c) 选择具有更有利的相关性状组合的亲本、后代或测试植物。

43. 如权利要求 42 所述的方法,其中,所述亲本、后代或测试植物是玉米近交系。

44. 如权利要求 42 所述的方法,其中,所述有利的相关性状组合提供提高的杂种优势。

45. 一种对玉米植物进行基因分型以选择具有有利性状的植物的方法,包括以下步骤:

a. 从至少一个玉米植物的组织获得 DNA 或 RNA 样品;和

b. 对于步骤 (a) 的所述样品,确定一组包含表 3 中确定的至少两种多态性的玉米基因组 DNA 多态性的等位基因状态,其中,用一组提供对所述玉米基因组 DNA 多态性进行分型的核酸分子确定所述等位基因状态;

c. 利用步骤 (b) 的所述等位基因状态确定情况,来选择具有有利的相关性状组合的植物。

46. 如权利要求 45 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组包括至少 5 种在表 3 中确定的多态性。

47. 如权利要求 45 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组包括至少 10 种在表 3 中确定的多态性。

48. 如权利要求 45 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组包括至少 20 种在表 3 中确定的多态性。

49. 如权利要求 45 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组包括至少 2 种选自下组的多态性:SEQ ID NO:5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279 及 SEQ ID NO:2468。

50. 如权利要求 49 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组包括至少 2 种选自下组的多态性:SEQ ID NO:2468、5407、287、574、3407、5367、4566、2457、5295 和 4548。

51. 如权利要求 50 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组包括至少 2 种选自下组的多态性:SEQ ID NO:2468、5407、287、574 和 3407。

52. 如权利要求 51 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组包括 SEQ ID NO:2468 和 SEQ ID NO:5407 的多态性。

53. 如权利要求 45 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组与对于产量、倒伏、成熟度、株高、耐旱性和冷发芽中的至少一种确定的性状值相关。

54. 如权利要求 53 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组与产量的性状值相关。

55. 如权利要求 49 所述的玉米植物基因分型方法,其中,所述玉米基因组 DNA 多态性组与产量的性状值相关。

56. 如权利要求 48 所述的方法,其中,所述至少 20 种玉米基因组 DNA 多态性的组鉴定分布于玉米基因组中的多态性。

57. 如权利要求 56 所述的方法,其中,所述至少 20 种玉米基因组 DNA 多态性的组鉴定分布于玉米的一条染色体上的多态性。

58. 如权利要求 56 所述的方法,其中,所述至少 20 种玉米基因组 DNA 多态性的组鉴定分布于玉米的至少两条染色体上的多态性。

59. 如权利要求 56 所述的方法,其中,所述至少 20 种玉米基因组 DNA 多态性的组鉴定分布于玉米的全部染色体上的多态性。

60. 如权利要求 59 所述的方法,其中,所述至少 20 种玉米基因组 DNA 多态性的组鉴定分布于玉米的全部染色体上的多态性,使得所述组中的至少 1 种所述多态性定位于每条染色体的每个染色体臂上。

61. 如权利要求 60 所述的方法,其中,所述组中的至少 10 种所述玉米基因组 DNA 多态性定位于每个染色体臂上。

62. 如权利要求 60 所述的方法,其中,所述组中的至少 20 种所述玉米基因组 DNA 多态性定位于每个染色体臂上。

63. 如权利要求 60 所述的方法,其中,至少 50 种所述玉米基因组 DNA 多态性定位于每个染色体臂上。

64. 如权利要求 59 所述的方法,其中,定位于染色体臂 1S 上的至少一种多态性选自

SEQ ID NO :381、2339、4410、239、1311、4683、4071、3141、5061、2972、1246、5114、3716、57、58、1114、5495、5476、1323、2451、765、845、5339、5363、1141、4137、3332、3775、1776、2213、3954、1389、870、5441、161、1791、5455、5296、783、3868、5230、5156、4709、5163、66、1766、4779、2672、5262、589、925、2909、4450、5118、669、4979、1553、3927、198、2593、5364、1261、4006、111、5090、4740、2699、2666、4357、4738、5036、697、901、230、5267、939、1219、5356、2290、4283、3062、5320、655、2261、5374、1559、1174、2300、3308、4176、3694、3035、3030、3990、4080、5526、316、3578、900、2384、5050、5344、2768、167、4939、2931、5315、1844、1020、5150、1547、707、1156、4993、1742、5158、5251、1441、5071、105、3425、3426、3817、5504、3918、5227、5152、2950、3877、4675、5214、15、2951、4517、5213、4241、4172、5413、1235、4482、3489、5311、3363、3562、4145、728、3395、5225、4449、4914、1308、4500 和 1543。

65. 如权利要求 59 所述的方法,其中,定位于染色体臂 1L 上的至少一种多态性选自 SEQ ID NO :2835、1301、1374、3766、2624、4571、927、4559、5420、3328、1702、5219、606、4124、3100、5223、4091、3292、3900、4814、5383、4354、4533、5355、2119、3574、5200、1513、732、5026、2326、4478、2099、1229、1443、2944、2325、5326、2669、4973、5142、5078、2645、3112、2194、3021、2986、4936、1577、4004、88、3913、610、4248、4895、4891、489、747、5134、4879、5235、1659、5187、5263、3127、5055、1556、4316、660、5431、1348、2900、133、269、3355、2243、2991、4584、3686、5047、1843、5272、592、4501、5002、1505、1066、549、236、2731、1973、2831、1539、5177、4522、5508、4951、2086、120、1466、10、1238、402、263、89、2811、4013、4015、3944、2706、430、639、4983、211、3919、5、5182、146、955、3339、2817、3485、3587、4171、5416、1627、2093、4093、2217、1956、5310、3261、4753、317、1110、4014、5489、5254、5154、3407、1980、5290、563、1073、3833、3512、5367、4156、3782、5498、4468、929、4676、3468、3754、4077、5333、1903、1771、2043、5490、4168、487、2426、4250、4648、2142、3058、3449、595、3107、3794、2844、1018、2140、5083、507、2299、5524、1871、1885、933、1455 和 3440。

66. 如权利要求 59 所述的方法,其中,定位于染色体臂 2S 上的至少一种多态性选自 SEQ ID NO :185、3347、5302、4102、4852、802、821、1668、5206、5402、4908、2432、3491、1568、4603、5049、2432、4585、4702、3068、4789、4398、4853、4890、621、1506、5039、5029、5179、4907、1204、4669、5451、3872、3390、2649、3325、3982、5481、1447、1726、5130、4322、4149、5104、4994、2979、4643、5328、2870、2861、1084、5115、11、2684、4586、5063、417、2320、5092、4492、2164、2725、4900、4997、5314、1058、3121、5112、4976、5405、4026、5492、2537、1491、4791、434、4580、1032、1352、2563、4003、1226、3697、1859、2635、3080、3110、420、5013、3026、5175、4659、5239、4020、938、1813、2313、1223、314、3258、3981、1090、4721、5018、4136、3084、1415、4417、2983、3695、2849、1393、2279、5427、1634、885、1826、4563、4697、5183、2827 和 4822。

67. 如权利要求 59 所述的方法,其中,定位于染色体臂 2L 上的至少一种多态性选自 SEQ ID NO :375、4781、4929、3474、3497、4579、5008、1008、3825、4220、913、2708、3698、275、4048、596、4002、1431、5377、4875、2942、5207、5064、3527、1339、4292、1690、2806、4115、4602、4746、5258、5418、4838、3789、5173、3783、809、3890、4213、4442、4231、2506、283、3349、1194、4703、4647、3631、951、4402、3356、3803、5245、3805、4236、28、4565、5493、1914、1317、4355、5037、724、1253、1388、5464、4307、5249、123、5048、2210、2434、4062、1796、

2054、1384、4671、2801、1595、1865、2691、3589、3624、2178、4568、550、2734、2303、4808、594、2046、1588、324、668、2977、4086、4173、5308、431、1994、2294、4674、3405、3404、3708、491、241、2524、4299、1210、3010、1062、2710、5271、4416、4170、4453、4399、2678、4446、4327、3540、4521、952、1089、5164、3965、4487、737 和 1121。

68. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 3S 上的至少一种多态性选自 SEQ ID NO :762、3024、1349、5525、4574、3078、2608、4553、4114、1160、3717、1399、1936、2787、5159、4047、3756、5470、3636、2846、4288、5457、2543、4649、668、658、1893、4938、1786、5376、3953、4105、5447、3006、4679、5081、4493、1151、1333、1887、3551、4162、1823、2688、1179、2732、2547、4942、2492、5358、1708、5102、3069、1074、1479、2687、5515、3735、1322、4911 和 4615。

69. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 3L 上的至少一种多态性选自 SEQ ID NO :1153、1497、3616、1022、2324、5006、2715、1712、3721、5269、469、2398、5188、5497、5140、1493、1778、1270、3085、4860、2912、3736、1093、3730、201、2370、4260、3655、4405、2065、1805、2215、4481、3504、3102、4259、4827、4067、3306、4667、5277、4269、3327、55、702、2404、3264、4555、4849、5506、2642、896、4751、4340、3891、1279、505、4017、5040、3461、3495、1993、93、5088、4556、285、4367、2959、877、1643、1456、5289、5467、4856、5473、5387、116、3849、5099、4949、1071、2226、2964、3510、3758、4154、5502、1511、1063、5132、5111、4689、436、4813、4952、1218、3586、5294、990、4655、2409、4651、522、4421、4096、2020、2090、2366、3482、1953、3133、4893、5395、3383、1350、210、4892、1459、2489、5138、5292、5362、1485、2038、3492、243、4519、1312、2594、4972、3706、773、4918、3647、573、991、5323、3970、4452、2823、3930、4869、5319、4281、3848、4965、4959、831、2003、2073、4100、5015、63、2781、4654、4962、5434、4024、356、4199、357、5161、5285、5166、1499、2343、390、4345、5432、2123、3555、5192、5208、2836、3013、3943、3976、580、4297 和 1631。

70. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 4S 上的至少一种多态性选自 SEQ ID NO :4232、1449、2901、3966、4054、3657、4541、752、3419、1621、1086、5221、5384、4085、1923、5453、4434、5077、5298、1571、3669、5283、5360、987、1411、4690、3348、722、5014、4244、4371、5369、3921、5281、5357、2394、3277、5256、2032、3577、1945、3256、5153、1839、872、4382、2523、1146、422、5462、2151、4960、2932、5203、4009、4490、5244、2838、1997、4948、1728、2830、4228、5260、2601、3270、4750、5216、5475、4021、5385、1130、4108、4582 和 4629。

71. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 4L 上的至少一种多态性选自 SEQ ID NO :1057、2619、1798、2017、3894、4641、3672、5000、1508、2754、908、1259、3928、5170、2176、638、1867、426、4273、5426、458、4576、306、1598、23、5056、5423、2486、2427、346、2630、4775、371、2301、4368、4486、2677、4401、4947、2955、4294、2770、1292、2087、177、2771、4984、4437、619、4747、2615、4588、5409、439、4225、2805、3793、5415、5429、611、635、1446、2341、3082、4219、5181、5195、760、4147、4188、4957、5388、142、4030、631、4280、1785、4314、5523、2887、4955、803、4937、5021、3066、4923、169、3159、148、5512、5024、237、4331、5389、3595、4772、1636、1996、3064、1808、3710、2465、5057、2168、2898、5445、1425、2317、3952、3033、5252、5334、4423、4444、409、5122、5341、4261、2796、4339、3858、3807、1329、

5149、4135、2591、4980、4558、1131、5273、4611、3768、5155、5379、3779、156、399、1592 和 1790。

72. 如权利要求 59 所述的方法,其中,定位于染色体臂 5S 上的至少一种多态性选自 SEQ ID NO :141、4609、5309、2637、3146、1365、1207、4242、4455、3529、5378、2154、454、2522、3041、5107、2079、5242、1205、1542、4798、4023、3045、5284、2832、4940、5196、1518、1324、4157、5229、318、5332、3995、1132、3487、5004、5471、4818、443、1014、5435、4699、4670、4666、2129、3950、5119、2935、2284、3922、2592、5141、5430、5069、4566、2610、3152、4832、1963、1866、2256、2692、2457、2933、4943、1958、2461、941、4289、1535、3511、4431、4691、4207、4218、2829、3749、2952、1574、4079、492、1404、1976、5232、179、520、3269、5191、3905、298、3544、251、2761、3370、2729、4321、2586、1529、853、1126、3759、3831、4502、5279、2424、3346、3569、4877、4360、2014、2820、2891、3342、1461、3763、157、2611、4701、5259、29、3118、1258、2767、1360、4295、1689、3627、4473、5190、4634、5321、532、4597、815、3910、3446、4140 和 4950。

73. 如权利要求 59 所述的方法,其中,定位于染色体臂 5L 上的至少一种多态性选自 SEQ ID NO :2292、2800、3386、4183、4989、5124、2698、4304、463、5500、3354、4462、4046、836、4971、4164、2714、2726、4146、3906、4165、1946、2006、1369、3936、3566、945、4025、1762、528、1465、5211、4652、2621、2812、5176、581、4109、1846、2528、2295、5436、2075、3451、287、3300、3399、3095、5297、597、1330、64、574、328、1252、2663、4810、667、3734、780、1091、2311、1899、1760、2748、4864、2002、106、3483、4660、2675、5307、295、3765、3822、2885、4403、4326、4591、2696、4301、2545、4293、2733、5454、1464、4365、2143、413、325、3857、2314、389、385、4523、3505、2271、3787、4692、5075、98、99、1334、1358、3361、4419、2402、3770、4894 和 5299。

74. 如权利要求 59 所述的方法,其中,定位于染色体臂 6S 上的至少一种多态性选自 SEQ ID NO :1639、2378、3516、4479、4771、138、1094、1878、2348、180、4378 和 3901。

75. 如权利要求 59 所述的方法,其中,定位于染色体臂 6L 上的至少一种多态性选自 SEQ ID NO :406、1755、1026、1985、225、1538、1661、2400、3053、4041、4082、4469、5117、5147、5168、5212、3732、5128、1247、22、2851、3275、3046、394、2535、2588、2788、3861、3884、3273、2527、3888、2155、3162、5074、2380、4144、5414、4344、2374、2441、2491、3583、5220、3582、3644、2016、3254、4313、4257、215、5275、4990、3387、4118、4512、4857、716、5127、4862、3844、488、4361、5288、4333、5265、4825、152、3338、694、3777、5340、743、4296、415、1149、1584、2742、4389、3851、1955、2585、5139、2381、2456、2456、2519、3816、4511、3039、506、1731、1775、5359、2643、4870、4996、4828、4886、2549、348、2804、4968、448、1419、1075、1968、5488、1675、3509、3500、4831、656、119、87、4364、3876、4777、5007、1117、4491、3018、2616、4608、51、2852、4792、2609、3924、2629、570、1510、898、3693、4619、5053、5370、4422、3898、1974、4549、3297、5469、4650、1995、4637、5424、1800、3089、5032、514、4087、4841、165、482、794、1198、2221、3892、835、2550、3288、5113、2175、5145、3170、4441 和 2288。

76. 如权利要求 59 所述的方法,其中,定位于染色体臂 7S 上的至少一种多态性选自 SEQ ID NO :1562、4982、590、1245、5466、195、2177、4613、4267、2089、127、3417、4604、5482、5518、1609、5417、3654、1314、4735、5365、4022、1401、1784、2004、2364、3098、2705、5460、

3079、5146、734、2249、5253、5143、1147、1684、5228、534、1306、1544、4987、452、557、1037、3815、5336、4628、4031、2333、4373、3637、1977、4854、651、1534、1901、4059、2507、2589、4445、5178、591、738、1099、2172、2453、5066、4466 和 4958。

77. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 7L 上的至少一种多态性选自 SEQ ID NO : 2995、3122、3524、4848、2798、4569、115、2372、2373、3059、1434、5084、1602、1484、351、2252、3801、1580、2008、3311、2084、5022、1267、2413、4184、600、3576、429、1081、2794、1024、1608、4266、4672、377、820、3984、1536、2436、5076、5327、423、424、2997、5380、1819、5499、2660、4415、2841、5247、2357、2228、5343、4465、5301、1420、4846、1137、1152、4884、1124、509、1277、3824、2428、4967、1162、2328、726、38、499、208、3856、1921、4927、5035、4599、2727、5098、1228、2908、483、4723、5391、4485、1065、2721、2135、663、3882、163、2819、2147、3542、94、1942 和 95。

78. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 8S 上的至少一种多态性选自 SEQ ID NO : 4383、4426、5268、4985、4988、4632、2562、3360、5479、3651、3550、1630、1965、3635、5348、4276、1209、2868、3475、4830、693、3379、5446、5210、3473、4881、2653、3557、975、865、566、5261、584、3570、5106、5456、2105、2280 和 2845。

79. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 8L 上的至少一种多态性选自 SEQ ID NO : 3875、4536、4757、5510、18、1276、5238、5496、273、2078、3357、4922、868、3429、5017、3563、3842、2468、1874、2160、640、4099、2477、2626、5407、2963、3457、4790、1569、5237、2007、3796、5110、3973、4622、5031、1072、1429、3674、5291、1002、4595、4358、344、3685、2724、3004、2778、2469、264、3139、4192、1332、3798、1611、4944、5016、3855、985、4113、1302、44、1982、2664、5067、5400、1725、2793、4303、1978、2719、4324、3546、4673、4392、4040、3865、2455、2797、2883、5516、3469、935、5062、1483、1184、1428、4334、847、4513、775、884、540、376、2704、755、1981、882、5503、338、3818、3960、4057、1591、1896、917 和 4084。

80. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 9S 上的至少一种多态性选自 SEQ ID NO : 404、4166、3980、3899、2982、1164、1013、3937、2270、4456、5329、2923、4323、5038、4963、3031、5129、3853、4748、2452、3400、4435、818、3478、5373、1991、311、3860、4741、4755、5046、5480、4995、5520、1088、2459、4132、2150 和 891。

81. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 9L 上的至少一种多态性选自 SEQ ID NO : 187、1824、3507、4684、2408、4705、2765、3280、272、340、1774、2361、2605、2636、3014、3077、5108、5428、1934、3285、3994、889、4210、4286、2617、4472、5030、2360、3499、4524、4902、5513、5459、2766、1637、2483、3086、3978、4531、4767、1596、2921、4055、4915、1653、4732、4677、5452、2928、5372、4974、5350、3733、4195、2131、2976、3545、5033、2329、91、4768、2039、90、257、2371、3431、1587、2777、4552、4706、5366、562、468、4347、2614、498、3135、4966、4888、4328、3155、1769、1288、2274、4904、4766、4845、65、1128、2067、2049、25、3108、4773、5160、200、5085、1737、4341、2940、4909、256、1952、5051、531、708、2096、5419、5521、4451、326、5338、4526、5100、2053、2869、2848、3757、5121、2867、1326、4506、5483、1275、3568、930、373、4494、5487、1021 和 3983。

82. 如权利要求 59 所述的方法, 其中, 定位于染色体臂 10S 上的至少一种多态性选自 SEQ ID NO : 2281、3571、3724、3939、3521、4776、2792、4010、5392、1926、1930、4532、2138、

4899、4921、5352、5093、4128、4657、696、366、493、3136 和 5345。

83. 如权利要求 59 所述的方法,其中,定位于染色体臂 10L 上的至少一种多态性选自 SEQ ID NO :13、2145、2234、5478、242、5443、2780、3052、5458、3130、839、1069、3374、4724、5060、5303、1412、1331、1583、2895、5368、2113、4429、274、2602、4711、5257、4498、5501、1116、1294、4460、2206、2240、2444、4377、2735、2741、3009、3649、3850、2494、4507、4717、5422、852、2122、3337、4211、1614、2557、4001、4043、5082、1809、2516、2475、3946、1452、1201、2214、3795、3813、5194、5318、2471、3496、1701、3776、3895、4262、5280、5522、3573、1371、5241、3786、1779、3302、4408、5337、2873、3925、1573、1200、2356、2520、5295、2209、1157、2554、5137、3063、858、3908、4548、4338、2816、4876、4285、4961、478、4306、5151、4642、3902、1575、2919、3885、3870、3762、3164、5065、4161、4572、5226、1933、3025、3812、4999、1607、4005、411、3687、2536、5042、3420、5394、2570、2813 和 4903。

84. 一种组合物,其包含用于检测代表玉米 DNA 中的多态性的分子标记的至少两种分离的核酸分子,其中,所述组合物的第一核酸分子包括包含多态性核苷酸残基和至少 8 个直接邻近所述多态性核苷酸残基 3' 端的核苷酸的寡核苷酸,其中,所述组合物的第二核酸分子包括包含多态性核苷酸残基和至少 8 个直接邻近所述多态性核苷酸残基 5' 端的核苷酸的寡核苷酸,并且其中所述多态性是在表 3 中所确定的。

玉米多态性与基因分型方法

[0001] 相关申请的交叉引用

[0002] 无。

[0003] 关于联邦政府资助的研究或开发的声明

[0004] 不适用。

[0005] 序列列表和表格的引入

[0006] 本申请在 CD-ROM 中提供了序列列表和序列列表的计算机可读形式 (CRF), 它们各自包含于 2007 年 5 月 15 日创建的、包含 7174544 个字节 (在 MS-Windows 中统计) 的、文件名为 “46-21 (54824). SEQLIST. txt” 的文件, 本文全部引入。本申请在 CD-ROM 中也提供了表 1 和表 3 两份拷贝, 包含 15945592 个字节 (在 MS-Windows 中统计) 的名称为 “表 1” (拷贝 1 和拷贝 2) 的文件, 和 115305 个字节 (在 MS-Windows 中统计) 的名称为 “表 3” (拷贝 1 和拷贝 2) 的文件, 这些文件都是于 2007 年 5 月 16 日创建的, 本文全部引入。

[0007] 发明背景

发明领域

[0008] 本文公开了玉米多态性、与这些多态性相关的核酸分子及使用这些多态性和分子作为标记分子 (例如在基因分型中) 的方法。

[0009] 相关技术

[0010] 多态性可用作分子标记, 也称为遗传标记, 在农业领域中, 例如, 在植物遗传研究和商业育种中, 用于与基因分型有关的应用。多态性的这些应用在美国专利 5, 385, 835、5, 437, 697、5, 385, 835、5, 492, 547、5, 746, 023、5, 962, 764、5, 981, 832 和 6, 100, 030 中有描述。

[0011] 特别地, 与只是基于表型数据所获得的结果相比, 在育种程序中使用分子标记加速了有价值的性状在种质中的遗传积累。本文中, “种质” 包括育种种质、育种群体、优良近交系的收集、随机交配个体的群体和双亲杂交。分子标记等位基因 (“等位基因” 是基因座处的替代序列) 用于鉴定在多个基因座处包含所需的基因型并且预期向其后代转移所需的基因型以及需要的表型的植物。分子标记等位基因可以用于鉴定在一个标记基因座、几个基因座或单元型处包含所需的基因型并且预期向其后代转移所需的基因型以及所需的表型的植物。

[0012] DNA 的高度保守性, 与稳定的多态性罕见的出现率相结合, 提供了既是可预测性的又可以辨别不同的基因型的分子标记。在现有的分子标记的种类中有多种指示遗传变异的多态性, 包括限制性片段长度多态性 (RFLP)、扩增片段长度多态性 (AFLP)、简单序列重复 (SSR)、单特征多态性 (SFP)、单核苷酸多态性 (SNP) 和插入 / 缺失多态性 (Indel)。

[0013] 分子标记在稳定性和基因组丰度方面不同。SNP 作为分子标记是特别有用的, 因为它们比其它多态性更稳定, 且在植物基因组是丰富的 (Bi 等人 Crop Sci. 46 :12-21 (2006), Kornberg, DNA Replication, W. H. Freeman & Co., San Francisco (1980))。因为植物物种的分子标记的数量是有限的, 发现另外的分子标志对于基因分型应用来说是至关重要的, 包

括标记性状关联研究、基因作图、基因发现、标记辅助选择和标记辅助育种。用作分子标记的多态性的发现和鉴定需要大量的测序和生物信息学工作,需要对两个或更多的进化分支系或群体进行大规模的测序。

[0014] 不断发展的技术使得某些分子标记更适合于快速、大规模地使用。具体地,如用于 SNP 检测的高通量筛选的技术表明,SNP 是优选的分子标记。

[0015] 发明概述

[0016] 正是鉴于上述问题,发展了本发明。本发明提供了一系列用于玉米的分子标记。这些分子标记包括通过对玉米基因组 DNA 进行测序并通过计算机分析确定多态性而发现的玉米 DNA 基因座。这些分子标记可用于多种基因分型应用。本发明的多态性玉米基因座包括至少 12 个连续核苷酸,该核苷酸包括或邻近本文中确定的多态性,例如在表 1 或表 3 中确定的多态性。如表 1 所示,SEQ ID NO :1 至 SEQ ID NO :6552 的核酸序列包括一种或多种多态性,例如,单核苷酸多态性 (SNP) 和插入 / 缺失多态性 (Indel)。如表 3 所示,本文确定的某些多态性也已定位于某些玉米染色体上。

[0017] 本发明首先提供核酸分子文库,其包括至少两组不同的核酸分子,其中,所述不同组核酸分子中的每一组允许对表 1 或表 3 中确定的相应的玉米基因组 DNA 多态性进行分型。在本发明此方面的某些实施方案中,文库包括两组或多组不同的核酸分子,它们排列在至少一个固体载体上或至少一个微量滴定板上。不同组的核酸分子可以位于微量滴定板的单独的和不同的孔中。不同组的核酸也可以位于固体载体上的不同的探针位置。

[0018] 也涉及其中核酸分子组合在单一混合物中的文库。在本发明的其它实施方案中,文库可以包含至少 8、至少 24、至少 96 或至少 384 组不同的核酸分子,其中,每组核酸分子允许对表 1 或表 3 中确定的相应的玉米基因组 DNA 多态性进行分型。也涉及包含允许对表 3 中确定的玉米基因组 DNA 多态性进行分型的几组核酸分子的文库,所述多态性选自 SEQ ID NO :2468、5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279。

[0019] 文库中的不同组核酸分子可以包括至少 12 个连续核苷酸的核酸分子,所述核苷酸包括或直接邻近表 1 中确定的相应的多态性,并且其中至少 12 个连续核苷酸的序列与包括或直接邻近所述多态性的玉米 DNA 片段任一链中的相同数目核苷酸的序列至少 90% 相同。在其它实施方案中,核酸分子是至少 15 个连续核苷酸或至少 18 个连续核苷酸的核酸分子。核酸分子可以进一步包含可检测的标记或提供可检测的标记的掺入。这种可检测的标记可以选自同位素、荧光团、氧化剂、还原剂、核苷酸和半抗原。可检测的标记可以通过化学反应添加到核酸上,或通过酶促反应掺入。

[0020] 不同组的核酸分子还可以包括:(a) 一对寡核苷酸引物,其中,所述寡核苷酸引物中的每一个包含至少 15 个核苷酸碱基,并且允许 PCR 扩增包含表 1 或表 3 中确定的所述相应多态性之一的 DNA 片段,和 (b) 至少一种检测核酸分子,其允许检测 (a) 的所述扩增片段中的多态性。在这些不同组的核酸中,检测核酸包含至少 12 个核苷酸碱基,或者包含至少 12 个核苷酸碱基和可检测的标记,并且其中所述检测核酸分子的序列与包含所述多态性的权利要求 1 之基因座中的玉米 DNA 片段任一链中相同数目核苷酸的序列至少 95% 相同。

[0021] 本发明还提供计算机可读介质,在其上记录有表 1 或表 3 中确定的至少两种玉米基因组 DNA 多态性。在其它实施方案中,至少 8 种表 1 或表 3 中确定的玉米基因组 DNA 多

态性记录在计算机可读介质上。也提供在其上记录有表 1 和表 3 中确定的至少两种玉米基因组 DNA 多态性和所述玉米基因组 DNA 多态性中每一种的相应遗传图谱位置的计算机可读介质。在其它实施方案中,至少 8 种玉米基因组 DNA 多态性和相应的遗传图谱位置记录在计算机可读介质上。

[0022] 本发明还提供用于读取、分类或分析玉米基因型数据的基于计算机的系统,该系统包括以下构件:(a) 数据存储装置,包括计算机可读介质,其上记录有至少 2 种表 1 或表 3 中确定的玉米基因组 DNA 多态性;(b) 搜索装置,用于将来自至少一种测试玉米植物的玉米基因组 DNA 序列与步骤 (a) 的数据存储装置的所述多态性序列进行比较,以鉴定同源或非同源序列;和 (c) 检索装置,用于鉴定步骤 (b) 的所述测试玉米基因组序列的所述同源或非同源序列。在其它实施方案中,至少 96 种表 1 或表 3 中确定的玉米基因组 DNA 多态性记录在基于计算机的系统的计算机可读介质上。在另外其它实施方案中,数据存储装置可以进一步包括计算机可读介质,其上记录有来自所述测试玉米植物中的至少一个的表型性状数据。数据存储装置还可以进一步包括计算机可读介质,其上记录有等位基因状态与亲本、后代或测试玉米植物的关联数据。还涉及基于计算机的系统,其中,多种表 3 中确定的定位的玉米基因组 DNA 多态性记录在计算机可读介质上,并且其中,计算机可读介质进一步包括所述定位的多态性中的每一种的遗传图谱位置数据。

[0023] 也提供了用于检测表 1 和表 3 中确定的玉米基因组 DNA 中的多态性的分离的核酸分子。涉及用于检测分子标记的分离的核酸分子,该分子标记代表在表 1 或表 3 中确定的玉米 DNA 中的多态性,所述核酸分子包含至少 15 个核苷酸,所述核苷酸包括或直接邻近多态性,并且与包括或直接邻近所述多态性的 DNA 任一链中的相同数目连续核苷酸的序列至少 90% 相同。本发明的分离核酸可以进一步包含可检测的标记或提供可检测的标记的掺入。可检测的标记可以选自同位素、荧光团、氧化剂、还原剂、核苷酸和半抗原。可检测的标记可以通过化学反应添加到核酸上,或者通过酶促反应掺入。分离的核酸可以检测选自下组的表 3 中的多态性:SEQ ID NO:2468、5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279。

[0024] 包括一种以上的分离核酸的其它分离寡核苷酸组合物可用于对表 1 或表 3 的玉米多态性进行分型。这样的分离寡核苷酸组合物可以用于通过 Taqman® 试验或瓣核酸内切酶 (Flap Endonuclease) 介导的 (Invader®) 试验对 SNP 多态性进行分型。在一个实施方案中,分离核酸组合物是一组寡核苷酸,包括:(a.) 一对寡核苷酸引物,其中,所述引物中的每一个包含至少 12 个连续核苷酸,并且其中,所述引物对允许 PCR 扩增包含表 1 或表 3 中确定的玉米基因组 DNA 多态性基因座的 DNA 片段;和 (b) 至少一种检测寡核苷酸,其允许检测所述扩增片段中的多态性,其中,所述检测寡核苷酸的序列与包括或直接邻近步骤 (a) 的所述多态性的玉米 DNA 片段任一链中相同数目连续核苷酸的序列至少 95% 相同。在这组寡核苷酸中,检测核酸包含至少 12 个核苷酸,并且提供可检测标记的掺入或进一步包含可检测的标记。可检测的标记可以选自同位素、荧光团、氧化剂、还原剂、核苷酸和半抗原。还提供了用于用瓣核酸内切酶介导的 (Invader®) 试验对公开的多态性进行分型的分离多核苷酸组合物。这种用于瓣核酸内切酶介导的试验的组合物包含至少两种用于检测代表玉米 DNA 中的多态性的分子标记的分离核酸分子,其中,该组合物的第一核酸分子包括包含多态性核苷酸残基和至少 8 个直接邻近所述多态性核苷酸残基 3' 端的核苷酸的寡核苷

酸,其中,该组合物的第二核酸分子包括包含多态性核苷酸残基和至少 8 个直接邻近所述多态性核苷酸残基 5' 端的核苷酸的寡核苷酸,并且其中,所述多态性在表 1 或表 3 中确定。

[0025] 也提供了对玉米植物进行基因分型以选择用于育种的亲本植物、后代植物或测试植物的各种方法。在一个实施方案中,对玉米植物进行基因分型以选择用于育种的亲本植物、后代植物或测试植物的方法包括以下步骤:a. 从至少一个玉米植物的组织获得 DNA 或 RNA 样品;b. 对于来自步骤 (a) 的所述样品,确定表 1 或表 3 中确定的至少一种玉米基因组 DNA 多态性的等位基因状态;和 c. 利用步骤 (b) 的所述等位基因状态确定情况,来选择用于育种的亲本植物、后代植物或测试植物。可以进行这种基因分型方法,以对表 3 中确定的定位的多态性进行分型。在这种基因分型方法中,通过允许鉴定单核苷酸多态性的试验可以确定多态性的等位基因状态。这种方法中使用的单核苷酸多态性试验可以选自单碱基延伸 (SBE)、等位基因特异性引物延伸测序 (ASPE)、DNA 测序、RNA 测序、基于微阵列的分析、通用 PCR、等位基因特异性延伸、杂交、质谱法、连接、延伸-连接和瓣核酸内切酶介导的试验。在这种方法的某些实施方案中,确定在表 1 或表 3 中确定的至少 8、至少 48、至少 96 或至少 384 种不同多态性的等位基因状态。

[0026] 基因分型方法还可以进一步包括在计算机可读介质上存储所述一种或多种等位状态基因状态确定情况产生的基因型数据的步骤,和/或进一步包括比较一个玉米植物与另一个玉米植物的基因型数据的步骤。在包括这些其它步骤的方法的某些实施方案中,也可以比较至少一种所述玉米植物的基因型数据与表型性状数据或表型性状指数数据。在包括这些其它步骤的方法的某些实施方案中,也可以比较至少两种所述玉米植物的基因型数据与表型性状数据或表型性状指数数据,并确定所述基因型数据与所述表型性状数据之间的一种或多种关联。在其中确定所述表型性状数据或表型性状指数数据与所述基因型性状数据之间的关联的这些方法的其它实施方案中,基因型性状数据包括确定至少 10 种定位的表 3 中确定的多态性的等位基因状态。

[0027] 也涉及培育玉米植物的方法。培育玉米植物的方法包括以下步骤:(a) 对于至少两个玉米植物的育种群体,确定与至少两个最多 10 厘摩的基因组窗口中的至少两个单元型相关的至少一种性状的性状值;(b) 在所述育种群体中,培育两个玉米植物,以产生后代种子群体;(c) 在所述后代种子中,确定至少一种表 1 或表 3 中确定的多态性在每个所述窗口中的等位基因状态,以确定所述单元型的存在;和 (d) 在所述后代种子中选择对于至少一种与确定的单元型相关的性状而言具有较高性状值的后代种子,从而培育玉米植物。在这些育种方法的某些实施方案中,对与基本上每条染色体整体上的每个相邻基因组窗口中的至少两个单元型相关的至少一种性状,确定其性状值。这种性状值可以确定选自以下的性状:除草剂耐受性、抗病性、昆虫或虫害抗性、改变的脂肪酸、蛋白质或碳水化合物代谢、增加的谷物产量、增加的油、增加的营养成分含量、提高的生长速度、提高的应激耐受性、优选的成熟度、增强的感官特性、改变的形态特征、其它农艺学性状、用于工业应用的性状、或对消费者有提高的吸引力的性状、或作为多性状指数的性状组合。在这些育种方法的其它实施方案中,对于每条染色体中最多 10 厘摩的基因组窗口中的单元型,选择具有较高的产量性状值的后代种子。在性状值是产量性状值并且对每个窗口中的单元型的性状值进行排序的方法中,可以选择窗口中的产量性状值高于所述窗口中的平均产量性状值的后代种子。在该方法的其它实施方案中,单元型中的多态性在包括 SEQ ID NO:1 至 SEQ ID NO:

6552 的全部 DNA 序列的 DNA 序列组中。

[0028] 也提供了选择用于育种的亲本、后代或测试植物的方法。这些选择用于植物育种的亲本、后代或测试植物的方法包括以下步骤：a) 在至少第一和第二玉米近交系中，确定表 1 或表 3 中确定的多种多态性与多种性状之间的关联；b) 确定亲本、后代或测试植物中的一种或多种多态性的等位基因状态；c) 选择具有更有利的相关性状组合的亲本、后代或测试植物。在某些实施方案中，亲本、后代或测试植物是玉米近交系。在亲本、后代或测试植物中选择的相关性状的有利组合可以是提供改进的杂种优势的亲本、后代或测试植物。

[0029] 也提供提高杂种优势的方法。提高杂种优势的方法包括以下步骤：(a) 在两个以上的玉米近交系中，确定表 1 或表 3 中确定的多种多态性与多种性状之间的关联；(b) 将选自步骤 (a) 的近交系的两个近交系分配至杂种优势群；(c) 在来自步骤 (b) 的至少两个近交系之间进行至少一次杂交，其中，每个近交系来自不同的和互补的杂种优势群，并且其中对于提高杂种优势的遗传特征，优化所述互补杂种优势群；和 (d) 通过步骤 (c) 的所述杂交获得杂种后代植物，其中，相对于与未经选择的近交系杂交产生的后代，所述杂种后代植物显示提高的杂种优势。

[0030] 也提供了对玉米进行基因分型以选择用于育种的亲本植物、后代植物或测试植物的方法，其中，利用多组不同的核酸对定位于多个基因组基因座上的多种不同的多态性进行分型。这些对玉米植物进行基因分型以选择用于育种的亲本植物、后代植物或测试植物的方法包括以下步骤：(a) 从至少一个玉米植物的组织获得 DNA 或 RNA 样品；(b) 对于所述步骤 (a) 的样品，确定一组包含表 1 或表 3 中确定的至少两种多态性的玉米基因组 DNA 多态性的等位基因状态，其中，用一组提供对所述玉米基因组 DNA 多态性进行分型的核酸分子确定所述等位基因状态；和 c. 利用步骤 (b) 的所述等位基因状态确定情况，来选择用于育种的亲本植物、后代植物或测试植物。但是，本方法的其它实施方案提供确定至少 5、至少 10 或至少 20 种在表 1 或表 3 中确定的多态性的等位基因状态。玉米基因组 DNA 多态性组可以包括至少 2 种选自以下的多态性：SEQ ID NO：5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279 和 SEQ ID NO：2468。玉米基因组 DNA 多态性组也可以包括至少 2 种选自以下的多态性：SEQ ID NO：2468、5407、287、574、3407、5367、4566、2457、5295 和 4548。或者，玉米基因组 DNA 多态性组也可以包括选自至少 2 种以下的多态性：SEQ ID NO：2468、5407、287、574 和 3407。在一个实施方案中，玉米基因组多态性组包括多态性 SEQ ID NO：2468 和 5407。在这种方法中，玉米基因组 DNA 多态性组可以与对产量、倒伏、成熟度、株高、耐旱性和冷发芽中的至少一种确定的性状值相关联。特别涉及其中玉米基因组 DNA 多态性组与产量性状值相关联的基因分型方法。在一个实施方案中，与性状值相关的多态性选自 SEQ ID NO：5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279 及 SEQ ID NO：2468。选自 SEQ ID NO：5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279 及 SEQ ID NO：2468 的多态性与产量性状值相关。

[0031] 也提供了对玉米植物进行基因分型以选择用于育种的亲本植物、后代植物或测试植物的方法，其中，利用多组不同的核酸对定位于玉米基因组上分布的多个基因组基因座

上的多种不同的多态性进行分型。在这些方法中,一组至少 20 种玉米基因组 DNA 多态性鉴定分布于玉米基因组中的多态性。在这种方法的某些实施方案中,对其进行分型的至少 20 种玉米基因组 DNA 多态性的组鉴定分布于玉米的一条染色体或分布于玉米的至少两条染色体中的多态性。在这种方法的其它实施方案中,至少 20 种玉米基因组 DNA 多态性的组鉴定分布于玉米的全部染色体中的多态性。当 20 种玉米基因组 DNA 多态性分布于玉米的全部染色体中时,它们可以分布为使得这组中的至少 1 种所述多态性定位于每条染色体的每个染色体臂上,从而使得所述组中的至少 1 种所述多态性定位于每个染色体臂上。但是,该方法也可以采用更多的多态性,使得该组中的至少 10 种玉米基因组 DNA 多态性定位于每个染色体臂上。在其它实施方案中,该组中至少 20 种或至少 50 种玉米基因组 DNA 多态性定位于每个染色体臂上。在该方法的某些实施方案中,至少一种多态性定位于染色体臂 1S 上,且可以选自 SEQ ID NO:381、2339、4410、239、1311、4683、4071、3141、5061、2972、1246、5114、3716、57、58、1114、5495、5476、1323、2451、765、845、5339、5363、1141、4137、3332、3775、1776、2213、3954、1389、870、5441、161、1791、5455、5296、783、3868、5230、5156、4709、5163、66、1766、4779、2672、5262、589、925、2909、4450、5118、669、4979、1553、3927、198、2593、5364、1261、4006、111、5090、4740、2699、2666、4357、4738、5036、697、901、230、5267、939、1219、5356、2290、4283、3062、5320、655、2261、5374、1559、1174、2300、3308、4176、3694、3035、3030、3990、4080、5526、316、3578、900、2384、5050、5344、2768、167、4939、2931、5315、1844、1020、5150、1547、707、1156、4993、1742、5158、5251、1441、5071、105、3425、3426、3817、5504、3918、5227、5152、2950、3877、4675、5214、15、2951、4517、5213、4241、4172、5413、1235、4482、3489、5311、3363、3562、4145、728、3395、5225、4449、4914、1308、4500 和 1543。在该方法的其它实施方案中,定位于染色体臂 1L 上的至少一种多态性选自 SEQ ID NO:2835、1301、1374、3766、2624、4571、927、4559、5420、3328、1702、5219、606、4124、3100、5223、4091、3292、3900、4814、5383、4354、4533、5355、2119、3574、5200、1513、732、5026、2326、4478、2099、1229、1443、2944、2325、5326、2669、4973、5142、5078、2645、3112、2194、3021、2986、4936、1577、4004、88、3913、610、4248、4895、4891、489、747、5134、4879、5235、1659、5187、5263、3127、5055、1556、4316、660、5431、1348、2900、133、269、3355、2243、2991、4584、3686、5047、1843、5272、592、4501、5002、1505、1066、549、236、2731、1973、2831、1539、5177、4522、5508、4951、2086、120、1466、10、1238、402、263、89、2811、4013、4015、3944、2706、430、639、4983、211、3919、5、5182、146、955、3339、2817、3485、3587、4171、5416、1627、2093、4093、2217、1956、5310、3261、4753、317、1110、4014、5489、5254、5154、3407、1980、5290、563、1073、3833、3512、5367、4156、3782、5498、4468、929、4676、3468、3754、4077、5333、1903、1771、2043、5490、4168、487、2426、4250、4648、2142、3058、3449、595、3107、3794、2844、1018、2140、5083、507、2299、5524、1871、1885、933、1455 和 3440。

[0032] 在该方法的其它实施方案中,定位于染色体臂 2S 上的至少一种多态性选自 SEQ ID NO:185、3347、5302、4102、4852、802、821、1668、5206、5402、4908、2432、3491、1568、4603、5049、2432、4585、4702、3068、4789、4398、4853、4890、621、1506、5039、5029、5179、4907、1204、4669、5451、3872、3390、2649、3325、3982、5481、1447、1726、5130、4322、4149、5104、4994、2979、4643、5328、2870、2861、1084、5115、11、2684、4586、5063、417、2320、5092、4492、2164、2725、4900、4997、5314、1058、3121、5112、4976、5405、4026、5492、2537、1491、4791、

434、4580、1032、1352、2563、4003、1226、3697、1859、2635、3080、3110、420、5013、3026、5175、4659、5239、4020、938、1813、2313、1223、314、3258、3981、1090、4721、5018、4136、3084、1415、4417、2983、3695、2849、1393、2279、5427、1634、885、1826、4563、4697、5183、2827 和 4822。

[0033] 在该方法的其它实施方案中,定位于染色体臂 2L 上的至少一种多态性选自 SEQ ID NO :375、4781、4929、3474、3497、4579、5008、1008、3825、4220、913、2708、3698、275、4048、596、4002、1431、5377、4875、2942、5207、5064、3527、1339、4292、1690、2806、4115、4602、4746、5258、5418、4838、3789、5173、3783、809、3890、4213、4442、4231、2506、283、3349、1194、4703、4647、3631、951、4402、3356、3803、5245、3805、4236、28、4565、5493、1914、1317、4355、5037、724、1253、1388、5464、4307、5249、123、5048、2210、2434、4062、1796、2054、1384、4671、2801、1595、1865、2691、3589、3624、2178、4568、550、2734、2303、4808、594、2046、1588、324、668、2977、4086、4173、5308、431、1994、2294、4674、3405、3404、3708、491、241、2524、4299、1210、3010、1062、2710、5271、4416、4170、4453、4399、2678、4446、4327、3540、4521、952、1089、5164、3965、4487、737 和 1121。

[0034] 在该方法的其它实施方案中,定位于染色体臂 3S 上的至少一种多态性选自 SEQ ID NO :762、3024、1349、5525、4574、3078、2608、4553、4114、1160、3717、1399、1936、2787、5159、4047、3756、5470、3636、2846、4288、5457、2543、4649、668、658、1893、4938、1786、5376、3953、4105、5447、3006、4679、5081、4493、1151、1333、1887、3551、4162、1823、2688、1179、2732、2547、4942、2492、5358、1708、5102、3069、1074、1479、2687、5515、3735、1322、4911 和 4615。

[0035] 在该方法的其它实施方案中,定位于染色体臂 3L 上的至少一种多态性选自 SEQ ID NO :1153、1497、3616、1022、2324、5006、2715、1712、3721、5269、469、2398、5188、5497、5140、1493、1778、1270、3085、4860、2912、3736、1093、3730、201、2370、4260、3655、4405、2065、1805、2215、4481、3504、3102、4259、4827、4067、3306、4667、5277、4269、3327、55、702、2404、3264、4555、4849、5506、2642、896、4751、4340、3891、1279、505、4017、5040、3461、3495、1993、93、5088、4556、285、4367、2959、877、1643、1456、5289、5467、4856、5473、5387、116、3849、5099、4949、1071、2226、2964、3510、3758、4154、5502、1511、1063、5132、5111、4689、436、4813、4952、1218、3586、5294、990、4655、2409、4651、522、4421、4096、2020、2090、2366、3482、1953、3133、4893、5395、3383、1350、210、4892、1459、2489、5138、5292、5362、1485、2038、3492、243、4519、1312、2594、4972、3706、773、4918、3647、573、991、5323、3970、4452、2823、3930、4869、5319、4281、3848、4965、4959、831、2003、2073、4100、5015、63、2781、4654、4962、5434、4024、356、4199、357、5161、5285、5166、1499、2343、390、4345、5432、2123、3555、5192、5208、2836、3013、3943、3976、580、4297 和 1631。

[0036] 在该方法的其它实施方案中,定位于染色体臂 4S 上的至少一种多态性选自 SEQ ID NO :4232、1449、2901、3966、4054、3657、4541、752、3419、1621、1086、5221、5384、4085、1923、5453、4434、5077、5298、1571、3669、5283、5360、987、1411、4690、3348、722、5014、4244、4371、5369、3921、5281、5357、2394、3277、5256、2032、3577、1945、3256、5153、1839、872、4382、2523、1146、422、5462、2151、4960、2932、5203、4009、4490、5244、2838、1997、4948、1728、2830、4228、5260、2601、3270、4750、5216、5475、4021、5385、1130、4108、4582 和 4629。

[0037] 在该方法的其它实施方案中,定位于染色体臂 4L 上的至少一种多态性选自 SEQ ID NO :1057、2619、1798、2017、3894、4641、3672、5000、1508、2754、908、1259、3928、5170、2176、638、1867、426、4273、5426、458、4576、306、1598、23、5056、5423、2486、2427、346、2630、4775、371、2301、4368、4486、2677、4401、4947、2955、4294、2770、1292、2087、177、2771、4984、4437、619、4747、2615、4588、5409、439、4225、2805、3793、5415、5429、611、635、1446、2341、3082、4219、5181、5195、760、4147、4188、4957、5388、142、4030、631、4280、1785、4314、5523、2887、4955、803、4937、5021、3066、4923、169、3159、148、5512、5024、237、4331、5389、3595、4772、1636、1996、3064、1808、3710、2465、5057、2168、2898、5445、1425、2317、3952、3033、5252、5334、4423、4444、409、5122、5341、4261、2796、4339、3858、3807、1329、5149、4135、2591、4980、4558、1131、5273、4611、3768、5155、5379、3779、156、399、1592 和 1790。

[0038] 在该方法的其它实施方案中,定位于染色体臂 5S 上的至少一种多态性选自 SEQ ID NO :141、4609、5309、2637、3146、1365、1207、4242、4455、3529、5378、2154、454、2522、3041、5107、2079、5242、1205、1542、4798、4023、3045、5284、2832、4940、5196、1518、1324、4157、5229、318、5332、3995、1132、3487、5004、5471、4818、443、1014、5435、4699、4670、4666、2129、3950、5119、2935、2284、3922、2592、5141、5430、5069、4566、2610、3152、4832、1963、1866、2256、2692、2457、2933、4943、1958、2461、941、4289、1535、3511、4431、4691、4207、4218、2829、3749、2952、1574、4079、492、1404、1976、5232、179、520、3269、5191、3905、298、3544、251、2761、3370、2729、4321、2586、1529、853、1126、3759、3831、4502、5279、2424、3346、3569、4877、4360、2014、2820、2891、3342、1461、3763、157、2611、4701、5259、29、3118、1258、2767、1360、4295、1689、3627、4473、5190、4634、5321、532、4597、815、3910、3446、4140 和 4950。

[0039] 在该方法的其它实施方案中,定位于染色体臂 5L 上的至少一种多态性选自 SEQ ID NO :2292、2800、3386、4183、4989、5124、2698、4304、463、5500、3354、4462、4046、836、4971、4164、2714、2726、4146、3906、4165、1946、2006、1369、3936、3566、945、4025、1762、528、1465、5211、4652、2621、2812、5176、581、4109、1846、2528、2295、5436、2075、3451、287、3300、3399、3095、5297、597、1330、64、574、328、1252、2663、4810、667、3734、780、1091、2311、1899、1760、2748、4864、2002、106、3483、4660、2675、5307、295、3765、3822、2885、4403、4326、4591、2696、4301、2545、4293、2733、5454、1464、4365、2143、413、325、3857、2314、389、385、4523、3505、2271、3787、4692、5075、98、99、1334、1358、3361、4419、2402、3770、4894 和 5299。

[0040] 在该方法的其它实施方案中,定位于染色体臂 6S 上的至少一种多态性选自 SEQ ID NO :1639、2378、3516、4479、4771、138、1094、1878、2348、180、4378 和 3901。

[0041] 在该方法的其它实施方案中,定位于染色体臂 6L 上的至少一种多态性选自 SEQ ID NO :406、1755、1026、1985、225、1538、1661、2400、3053、4041、4082、4469、5117、5147、5168、5212、3732、5128、1247、22、2851、3275、3046、394、2535、2588、2788、3861、3884、3273、2527、3888、2155、3162、5074、2380、4144、5414、4344、2374、2441、2491、3583、5220、3582、3644、2016、3254、4313、4257、215、5275、4990、3387、4118、4512、4857、716、5127、4862、3844、488、4361、5288、4333、5265、4825、152、3338、694、3777、5340、743、4296、415、1149、1584、2742、4389、3851、1955、2585、5139、2381、2456、2456、2519、3816、4511、3039、506、1731、1775、

5359、2643、4870、4996、4828、4886、2549、348、2804、4968、448、1419、1075、1968、5488、1675、3509、3500、4831、656、119、87、4364、3876、4777、5007、1117、4491、3018、2616、4608、51、2852、4792、2609、3924、2629、570、1510、898、3693、4619、5053、5370、4422、3898、1974、4549、3297、5469、4650、1995、4637、5424、1800、3089、5032、514、4087、4841、165、482、794、1198、2221、3892、835、2550、3288、5113、2175、5145、3170、4441 和 2288。

[0042] 在该方法的其它实施方案中,定位于染色体臂 7S 上的至少一种多态性选自 SEQ ID NO :1562、4982、590、1245、5466、195、2177、4613、4267、2089、127、3417、4604、5482、5518、1609、5417、3654、1314、4735、5365、4022、1401、1784、2004、2364、3098、2705、5460、3079、5146、734、2249、5253、5143、1147、1684、5228、534、1306、1544、4987、452、557、1037、3815、5336、4628、4031、2333、4373、3637、1977、4854、651、1534、1901、4059、2507、2589、4445、5178、591、738、1099、2172、2453、5066、4466 和 4958。

[0043] 在该方法的其它实施方案中,定位于染色体臂 7L 上的至少一种多态性选自 SEQ ID NO :2995、3122、3524、4848、2798、4569、115、2372、2373、3059、1434、5084、1602、1484、351、2252、3801、1580、2008、3311、2084、5022、1267、2413、4184、600、3576、429、1081、2794、1024、1608、4266、4672、377、820、3984、1536、2436、5076、5327、423、424、2997、5380、1819、5499、2660、4415、2841、5247、2357、2228、5343、4465、5301、1420、4846、1137、1152、4884、1124、509、1277、3824、2428、4967、1162、2328、726、38、499、208、3856、1921、4927、5035、4599、2727、5098、1228、2908、483、4723、5391、4485、1065、2721、2135、663、3882、163、2819、2147、3542、94、1942 和 95。

[0044] 在该方法的其它实施方案中,定位于染色体臂 8S 上的至少一种多态性选自 SEQ ID NO :4383、4426、5268、4985、4988、4632、2562、3360、5479、3651、3550、1630、1965、3635、5348、4276、1209、2868、3475、4830、693、3379、5446、5210、3473、4881、2653、3557、975、865、566、5261、584、3570、5106、5456、2105、2280 和 2845。

[0045] 在该方法的其它实施方案中,定位于染色体臂 8L 上的至少一种多态性选自 SEQ ID NO :3875、4536、4757、5510、18、1276、5238、5496、273、2078、3357、4922、868、3429、5017、3563、3842、2468、1874、2160、640、4099、2477、2626、5407、2963、3457、4790、1569、5237、2007、3796、5110、3973、4622、5031、1072、1429、3674、5291、1002、4595、4358、344、3685、2724、3004、2778、2469、264、3139、4192、1332、3798、1611、4944、5016、3855、985、4113、1302、44、1982、2664、5067、5400、1725、2793、4303、1978、2719、4324、3546、4673、4392、4040、3865、2455、2797、2883、5516、3469、935、5062、1483、1184、1428、4334、847、4513、775、884、540、376、2704、755、1981、882、5503、338、3818、3960、4057、1591、1896、917 和 4084。

[0046] 在该方法的其它实施方案中,定位于染色体臂 9S 上的至少一种多态性选自 SEQ ID NO :404、4166、3980、3899、2982、1164、1013、3937、2270、4456、5329、2923、4323、5038、4963、3031、5129、3853、4748、2452、3400、4435、818、3478、5373、1991、311、3860、4741、4755、5046、5480、4995、5520、1088、2459、4132、2150 和 891。

[0047] 在该方法的其它实施方案中,定位于染色体臂 9L 上的至少一种多态性选自 SEQ ID NO :187、1824、3507、4684、2408、4705、2765、3280、272、340、1774、2361、2605、2636、3014、3077、5108、5428、1934、3285、3994、889、4210、4286、2617、4472、5030、2360、3499、4524、4902、5513、5459、2766、1637、2483、3086、3978、4531、4767、1596、2921、4055、4915、

1653、4732、4677、5452、2928、5372、4974、5350、3733、4195、2131、2976、3545、5033、2329、91、4768、2039、90、257、2371、3431、1587、2777、4552、4706、5366、562、468、4347、2614、498、3135、4966、4888、4328、3155、1769、1288、2274、4904、4766、4845、65、1128、2067、2049、25、3108、4773、5160、200、5085、1737、4341、2940、4909、256、1952、5051、531、708、2096、5419、5521、4451、326、5338、4526、5100、2053、2869、2848、3757、5121、2867、1326、4506、5483、1275、3568、930、373、4494、5487、1021 和 3983。

[0048] 在该方法的其它实施方案中,定位于染色体臂 10S 上的至少一种多态性选自 SEQ ID NO :2281、3571、3724、3939、3521、4776、2792、4010、5392、1926、1930、4532、2138、4899、4921、5352、5093、4128、4657、696、366、493、3136 和 5345。

[0049] 在该方法的其它实施方案中,定位于染色体臂 10L 上的至少一种多态性选自 SEQ ID NO :13、2145、2234、5478、242、5443、2780、3052、5458、3130、839、1069、3374、4724、5060、5303、1412、1331、1583、2895、5368、2113、4429、274、2602、4711、5257、4498、5501、1116、1294、4460、2206、2240、2444、4377、2735、2741、3009、3649、3850、2494、4507、4717、5422、852、2122、3337、4211、1614、2557、4001、4043、5082、1809、2516、2475、3946、1452、1201、2214、3795、3813、5194、5318、2471、3496、1701、3776、3895、4262、5280、5522、3573、1371、5241、3786、1779、3302、4408、5337、2873、3925、1573、1200、2356、2520、5295、2209、1157、2554、5137、3063、858、3908、4548、4338、2816、4876、4285、4961、478、4306、5151、4642、3902、1575、2919、3885、3870、3762、3164、5065、4161、4572、5226、1933、3025、3812、4999、1607、4005、411、3687、2536、5042、3420、5394、2570、2813 和 4903。

[0050] 下面参考附图,详述了本发明的进一步的特征和优点,以及本发明的各个实施方案的结构和操作。

[0051] 附图简述

[0052] 并入并构成说明书一部分的附图,说明了本发明的实施方案,并与说明书一起,用来解释本发明的原理。在附图中:

[0053] 图 1 是显示定位的本发明多态性的密度的玉米遗传图谱。

[0054] 图 2 是说明基因分型试验结果的等位图 (allelagram)。

[0055] 定义:

[0056] 如下定义本文使用的一些术语和短语。

[0057] “等位基因”指特定的基因座处的替代序列;等位基因的长度可以小到 1 个核苷酸碱基,但通常较大。等位基因序列可以是氨基酸序列或核酸序列。“基因座”是一种短序列,其通常是独特的,且通常在参照点附近在基因组中的一个特定位置处被发现;例如,作为基因或基因部分或基因间区域的短 DNA 序列。本发明的基因座可以是在基因组上特定位置处的独特 PCR 产物。本发明的基因座包含一种或多种多态性;即,在一些个体中存在的替代的等位基因。

[0058] “等位基因状态”指存在于包含基因组多态性的核酸分子中的核酸序列。例如,包含单核苷酸多态性的 DNA 分子的核酸序列可以在多态性位置处包括 A、C、G 或 T 残基,使得通过该多态性位置处存在哪个残基来定义等位基因状态。例如,包含单核苷酸多态性的 RNA 分子的核酸序列可以在多态性位置处包括 A、C、G 或 U 残基,使得通过该多态性位置处存在哪个残基来定义等位基因状态。同样地,包含 Indel 的核酸分子的核酸序列可以在多态性

位置处包括核酸序列的插入或缺失,使得通过在该多态性位置处是否存在插入或缺失来定义等位基因状态。

[0059] “关联”,在用于多态性和表型性状或性状指数时,指在多态性基因座的给定等位基因的存在与表型性状或性状指数值之间的任何统计显著性的相关,其中,该值可以是定性的或定量的。

[0060] “不同组的核酸分子”指一个或多个与包括、直接邻近或在给定玉米基因组多态性的 5' 或 3' 端的大约 1000 个碱基对之内的 DNA 序列杂交的核酸分子。在某些实施方案中,不同组的核酸分子包括一个核酸序列,该核酸序列包括或直接邻近于给定的多态性。在其它实施方案中,不同组的核酸分子包括一个或多个包括或直接邻近于多态性的核酸序列,以及在多态性的 5' 或 3' 端的大约 1000 个碱基对之内的其它核酸序列。

[0061] “基因型”指在个体生物内在一个或多个基因座处的等位基因组合。在二倍体生物的情况下,在每个基因座处有两个等位基因;当等位基因相同时,二倍体基因被称为是纯合的,而当等位基因不同时,二倍体基因被称为是杂合的。

[0062] “单元型”指往往是作为单元遗传的基因组 DNA 的等位基因片段;这种单元型可以用一个或多个多态性分子标记来表征,且可以限定为不大于 10 厘摩的大小。通过用更高的多态性密度提供更高的精度,单元型可以用基因组窗口来表征,例如,在 1-5 厘摩的范围内。

[0063] 短语“直接邻近”,当用来描述与包含多态性的 DNA 杂交的核酸分子时,指与直接邻接多态性核苷酸碱基位置的 DNA 序列杂交的核酸。例如,可用于单碱基延伸试验的核酸分子与多态性“直接邻近”。

[0064] “探询位置”是指固体载体上的物理位置,可以对其进行查询以获取一个或多个预定的基因组多态性的基因分型数据。

[0065] “共有序列”是指构建的 DNA 序列,其确定基因座处等位基因的 SNP 和 Indel 多态性。共有序列可以基于基因座处 DNA 的任一链,并且表示基因座中的每个 SNP 的任一个的核苷酸碱基及基因座中的所有 Indel 的核苷酸碱基。因此,虽然共有序列可能不是一个实际的 DNA 序列的拷贝,但共有序列可用于精确设计用于基因座中的实际多态性的引物和探针。

[0066] “表型”指作为基因表达的表现的细胞或生物体的可检测的特征。

[0067] “表型性状指数”指至少两个表型性状的复合值,其中可以给每个表型性状赋予权重,以反映对选择而言的相对重要性。

[0068] 本文中所用的“标记”或“分子标记”,是显示同一物种的两个或多个植物之间的多态性的 DNA 序列(例如,基因或基因的部分),其可以通过简单的试验鉴定或分型。有用的多态性包括单核苷酸多态性(SNP)、DNA 序列中的插入或缺失(Indel)、单特征多态性(SFP)和 DNA 序列的简单序列重复(SSR)。

[0069] “标记试验”指使用特定方法检测特定基因座处的多态性的方法。检测多态性的方法包括但不限于:限制性片段长度多态性(RFLP)、单碱基延伸、电泳、序列比对、等位基因特异性寡核苷酸杂交(ASO)、RAPD、等位基因特异性引物延伸测序(ASPE)、DNA 测序、RNA 测序、基于微阵列的分析、通用 PCR、等位基因特异性延伸、杂交、质谱法、连接、延伸-连接、核酸内切酶介导的染料释放试验和核酸内切酶介导的试验。美国专利 6,013,431 公开了示

例性的单碱基延伸试验。美国专利 5,538,848 公开了示例性的用于确定 SNP 的等位基因状态的核酸内切酶介导的染料释放试验,其中,核酸内切酶活性从杂交探针上释放报告染料。

[0070] “连锁”是指杂交产生配子类型的相对频率。例如,如果基因座 A 具有基因“A”或“a”,基因座 B 具有基因“B”或“b”,具有 AABB 的亲本 I 与具有 aabb 的亲本 B 之间的杂交将产生 4 种可能的配子,其中基因分离为 AB、Ab、aB 和 ab。空预期是会独立相等地分离成 4 个可能的基因型中的每一个,即,如果没有连锁,每个基因型将会有 1/4 的配子。配子向基因型的分离不同于 1/4 是由于连锁。

[0071] “连锁不平衡”定义为在一代的许多个体的群体中配子类型的相对频率。如果等位基因 A 的频率是 p , a 是 p' , B 是 q , b 是 q' , 那么基因型 AB 的预期频率(没有连锁不平衡)是 pq , Ab 是 pq' , aB 是 $p'q$, ab 是 $p'q'$ 。相对于预期频率的任何偏差被称为连锁不平衡。当两个基因座处于连锁不平衡时,它们被称为是“遗传连锁的”。

[0072] “数量性状基因座 (QTL)”指在一定程度上控制通常连续分布的并且可定量表示的性状的基因座。

[0073] 如本文所用,“序列同一性”指两个最佳比对的多核苷酸或肽序列在例如核苷酸或氨基酸的元件的整个比对窗口中不变的程度。测试序列和参考序列的比对区段的“同一性分数”是两个比对序列共有的相同元件数目除以参考序列区段中的元件总数,即整个参考序列或参考序列的较小的确定部分。“同一性百分比”是同一性分数乘以 100。

[0074] 如本文所用,“分型”(“typing”)指确定给定的玉米基因组多态性的特定等位基因形式的任何方法。例如,通过确定存在哪种核苷酸(即, A、G、T 或 C),对单核苷酸多态性(SNP)进行分型。通过确定是否存在 Indel 来确定插入/缺失(Indel)。通过包括但不限于标记分析的多种试验可以对 Indel 分型。

[0075] 优选实施方案详述

[0076] 下面的详细说明涉及用于玉米植物基因分型的分离的核酸组合物及相关方法。一般来说,这些组合物和方法可用于对玉米属的玉米植物进行基因分型。更具体地说,使用这些组合物和方法可以对玉米(*Zea mays*)种和亚种 *Zea mays* L. ssp. *Mays* 的玉米植物进行基因分型。在另外的一方面,玉米植物来自 *Zea mays* L. subsp. *mays* *Indentata* 群,也被称为马齿种玉米(dent corn)。在另一方面,玉米植物来自 *Zea mays* L. subsp. *mays* *Indurata* 群,也被称为硬质玉米(flint corn)。在另一方面,玉米植物来自 *Zea mays* L. subsp. *mays* *Saccharata* 群,也被称为甜玉米。在另一方面,玉米植物来自 *Zea mays* L. subsp. *mays* *Amylacea* 群,也被称为面粉玉米(fLOUR corn)。在另外的一方面,玉米植物来自 *Zea mays* L. subsp. *mays* *Everta* 群,也被称为爆裂种玉米(pop corn)。可以用本文所述的组合物和方法进行基因分型的玉米或玉米植物包括杂种、近交系、部分近交系或定义或未定义的群体的成员。

[0077] 分离的核酸分子-基因座、引物和探针

[0078] 本发明的玉米基因座包括一系列分子标记,其包括至少 20 个连续核苷酸,并包括或邻近于表 1 或表 3 中确定的一种或多种多态性。这些玉米基因座的核酸序列与包括或邻近多态性的玉米 DNA 片段任一链中相同核苷酸数的序列有至少 90% 的序列同一性,更优选至少 95%,或甚至更优选对于某些等位基因至少为 98%,在许多情况下至少为 99% 的序列同一性。可以在 SEQ ID NO:1 至 SEQ ID NO:6552 的序列中找到这样的玉米 DNA 片段的一

条链的核苷酸序列。根据多态性的性质可以理解,对于至少某些等位基因,与公开的多态性本身没有同一性。因此,对于除公开的多态性序列外的序列,可以确定序列同一性。换句话说,预计对于本文公开的多态性的其它等位基因可能存在,可以容易地通过测序方法表征,且可以用于基因分型。例如,本领域的技术人员可以理解,对于其中仅仅公开了两个多态性残基(例如,“A”或“G”)的单核苷酸多态性也可以包括其它多态性残基(例如,“T”和/或“G”)。

[0079] 每个基因座中的多态性更具体地在表 1 或表 3 中确定。SNP 特别可以用作遗传标记,因为它们比其它种类的多态性更稳定,且在玉米基因组中是丰富的。SNP 可以由插入、缺失和点突变产生。在本发明中,SNP 可以代表一个可能由一个或多个碱基对组成的插入与缺失(indel)事件,或单核苷酸多态性。两个或多个个体共有的多态性可能产生于源自共同祖先的个体。这种“来源同一性”(IBD)表征由两个或多个个体携带且全部来自同一祖先的两个 DNA 基因座/片段。“状态同一性”(IBS)表征由两个或多个个体携带并且在那些基因座处具有可检测到的相等等位基因的两个 DNA 基因座/片段。当考虑一大组作物系,并且多个系在标记基因座处具有相同的等位基因时,有必要确定标记基因座处的 IBS 是否是标记基因座周围的染色体区域处的 IBD 的可靠预测。一个片段中的大量标记基因座足以表征该片段的 IBD 的一个指示是,它们能够预测该片段内其它标记基因座处存在的等位基因。除了它们很少独立出现这一事实外,SNP 的稳定性和丰富性使它们可以用于确定 IBD。

[0080] 对于许多基因分型应用,采用来自一个以上的基因座的多态性作为标记是有用的。因此,本发明的一方面提供了核酸分子的集合,其允许对不同基因座的多态性进行分型。在这样的集合中的基因座的数目可以不同,但将是有限的数值,例如,少至 2 或 5 或 10 或 25 个基因座或更多,例如最多达 40 或 75 或 100 个或更多的基因座。

[0081] 本发明的另一方面提供能够与本发明的多态性玉米基因座杂交的分离的核酸分子。在本发明的某些实施方案中,例如,提供 PCR 引物的实施方案中,这样的分子包括至少 15 个核苷酸碱基。可用作引物的分子可以在高严格条件下与本发明的多态性基因座中的 DNA 片段的一条链杂交。用于扩增 DNA 的引物成对提供,即正向引物和反向引物。一条引物与基因座中的 DNA 的一条链互补,而另一条引物与基因座中的 DNA 的另一条链互补,即引物序列与一条链中相同核苷酸数目的序列优选地至少 90% 相同,更优选地至少 95% 相同。可以理解,这样的引物可以与远离多态性(例如,距多态性至少 5、10、20、50、100、200、500 或最大约 1000 个核苷酸碱基)的基因座中的序列杂交。本发明的引物的设计取决于本领域内熟知的因素,例如,避免或重复序列。

[0082] 本发明的分离的核酸分子的另一方面是用于多态性试验的杂交探针。在本发明的一方面,这样的探针是包含至少 12 个核苷酸碱基和可检测的标记的寡核苷酸。这种分子的目的是,例如在高严格条件下,与包括或邻近于多态性基因座扩增部分中的目标多态性的核苷酸碱基片段中的 DNA 一条链杂交。这样的寡核苷酸与多态性基因座中玉米 DNA 一条链中相同核苷酸数目的片段的序列优选地至少 90% 相同,更优选地至少 95% 相同。该可检测的标记可以是放射性元素或染料。在本发明的优选方面,杂交探针进一步包括荧光标记和猝灭剂,例如,用于可从 AB Biosystems 获得的被称为 Taqman® 试验的类型的杂交探针试验。

[0083] 本发明的分离的核酸分子在一定条件下能够与包括但不限于玉米基因组 DNA、克

隆的玉米基因组 DNA 和扩增的玉米基因组 DNA 的其它核酸分子杂交。如本文所用,如果两个核酸分子能够形成反平行双链核酸结构,那么这两个分子被称为能够彼此杂交。如果两个核酸分子表现出“完全的互补性”,即,一个序列中的每个核苷酸都与另一序列中的碱基配对核苷酸互补,则称一个核酸分子与另一核酸分子“互补”。如果两个分子在至少常规的“低严格”的条件下能互相杂交,并且具有足够的稳定性,从而允许它们保持彼此退火,则称这两个分子是“最低限度互补的”。类似地,如果两个分子在常规的“高严格”的条件下能互相杂交,并且具有足够的稳定性,从而允许它们保持彼此退火,则称这两个分子是“互补的”。例如至少在低严格条件下与其它核酸分子杂交的核酸分子被称为该其它核酸分子的“可杂交同族物”。Sambrook 等人, *Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York (1989) 和 Haymes 等人, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985) 描述了常规的严格条件, 本文引入以上文献作为参考。因而偏离完全互补性是允许的, 只要这种偏离没有完全消除该分子形成双链结构的能力。因此, 为了使核酸分子用作引物或探针, 只需要在序列上充分互补, 以在所采用的特定的溶剂和盐浓度下能够形成稳定的双链结构。

[0084] 促进 DNA 杂交的合适的严格条件, 例如, 大约 45°C、6.0x 氯化钠 / 柠檬酸钠 (SSC), 接着 50°C、2.0xSSC 洗涤, 是本领域的技术人员已知的, 或可以在 *Current Protocols in Molecular Biology*, John Wiley & Sons, N. Y., 1989, 6.3.1-6.3.6 (本文引入作为参考) 中找到。例如, 洗涤步骤中的盐浓度可以选自从大约 2.0xSSC、50°C 的低严格条件到大约 0.2xSSC、50°C 的高严格条件。另外, 洗涤步骤中的温度可以从低严格条件的室温大约 22°C 增加到高严格条件的大约 65°C。温度和盐都可以变化, 或者, 温度或盐浓度可以保持不变, 而另一个变量发生变化。

[0085] 在一个优选的实施方案中, 在中度严格条件下, 例如, 大约 2.0xSSC 和大约 65°C, 更优选地, 在高严格条件下, 例如 0.2xSSC 和大约 65°C, 本发明的核酸分子与具有 SEQ ID NO:1 至 SEQ ID NO:6552 所示核酸序列的玉米 DNA 片段的一条链特异性杂交。

[0086] 对于其中分子被设计为与通过如标记的双脱氧核苷酸的单碱基延伸检测的多态性邻近地杂交的试验, 这些分子在序列中可以包含至少 15 个、更优选至少 16 或 17 个核苷酸碱基, 所述序列与多态性玉米 DNA 片段的任一链中相同数目连续核苷酸的序列至少 90% 相同, 优选地至少 95% 相同。用于单碱基延伸试验的寡核苷酸可以从 Orchid Biosystems 获得。

[0087] 可作为杂交探针用于检测玉米 DNA 中的多态性的分离的核酸分子可设计用于不同的试验。对于其中探针用于杂交包括多态性的片段的试验, 这些分子可以包含至少 12 个核苷酸碱基和可检测的标记。核苷酸碱基序列与本发明的多态性基因座中的玉米 DNA 片段任一链中的相同数目连续核苷酸的序列优选地至少 90% 相同, 更优选地至少 95% 相同。该可检测的标记是位于分子一端的染料。在优选的方面, 分离的核酸分子在其末端包括染料和染料猝灭剂。对于 SNP 检测试验, 成对提供这样的染料和染料猝灭剂是有用的, 例如, 其中每个分子在 5' 端具有不同的荧光染料, 并且具有除单核苷酸多态性之外的相同的核苷酸序列。本领域公知如何为了报告的目的设计与 DNA 靶片段退火的寡核苷酸 PCR 探针对, 其中, 靶标的序列是已知的, 如本发明提供的多态性标记序列。

[0088] 对于其中分离的核酸分子被设计为与通过单碱基延伸检测的多态性邻近地杂交

的试验,这些分子在序列中可以包含至少 15 个、更优选至少 16 或 17 个核苷酸碱基,所述序列与多态性玉米 DNA 片段任一链中的相同数目连续核苷酸的序列至少 90% 相同,优选地至少 95% 相同。在这种情况下,分离的核苷酸提供可检测的标记的掺入。这种可检测的标记可以是同位素、荧光团、氧化剂、还原剂、核苷酸和半抗原。

[0089] 对于涉及使用瓣核酸内切酶的试验(即,Invader®试验)。在某些实施方案中,组合物包含至少两种用于检测代表玉米 DNA 多态性的分子标记的分离的核酸分子,其中,组合物的第一核酸分子包括寡核苷酸,该寡核苷酸包括多态性核苷酸残基和至少 8 个直接邻近所述多态性核苷酸残基的 3' 端的核苷酸,其中,组合物的第二核酸分子包括寡核苷酸,该寡核苷酸包括多态性核苷酸残基和至少 8 个直接邻近所述多态性核苷酸残基的 5' 端的核苷酸,并且其中,所述多态性在表 1 或表 3 中确定。在某些实施方案中,适于用瓣核酸内切酶对表 1 或表 3 的多态性进行分型的分离的核酸分子组合物包括至少一个具有“通用”5' 瓣序列的第一探针、至少一个第二或“Invader®”探针和至少一个包含标记的碱基和猝灭碱基的“FRET”盒(cassettes),该盒包含与一经裂解即从第一探针上释放的“通用瓣序列”互补的序列。

[0090] 鉴定多态性

[0091] SNP 是序列变异的结果,新的多态性可以通过对随机基因组或 cDNA 分子进行测序而检测。在一方面,可以通过比较不同系的 cDNA 序列来确定基因组中的多态性。尽管通过比较 cDNA 序列检测多态性相对方便,但是 cDNA 序列的评估无法获得关于相应基因组 DNA 中的内含子位置的信息。此外,不能从 cDNA 确定非编码序列中的多态性。这是一个缺点,例如,当使用源自 cDNA 的多态性作为对基因组 DNA 进行基因分型的标记时。如果多态性的范围内包括那些存在于非编码独特序列中的多态性,则可以设计更有效的基因分型试验。

[0092] 基因组 DNA 序列在鉴定和检测多态性方面比 cDNA 更有用。可以通过比较不同系的基因组 DNA 序列来确定基因组中的多态性。然而,高等真核生物的基因组 DNA 一般包含高比例的重复序列和转座子。如果通过减去或消除重复序列来富集编码/独特部分,则可以更有效地对基因组 DNA 测序。

[0093] 有许多本领域熟知的可以用来富集编码/独特序列的策略。这方面的例子包括使用对胞嘧啶甲基化敏感的酶,使用 McrBC 核酸内切酶以切割重复序列,以及印刷基因组文库的微阵列,其然后与重复序列探针杂交。

[0094] 在优选的实施方案中,通过利用甲基化模式的差别来富集编码 DNA;高等真核生物的 DNA 往往非常高度甲基化,但它不是均一地甲基化。事实上,重复序列比编码序列更高度甲基化。关于对 CG 岛(CG islands)中的 DNA 甲基化模式进行作图和评价的方法,参见美国专利 6,017,704。简单地说,一些限制性核酸内切酶在其识别位点对甲基化的胞嘧啶残基的存在敏感。如果在重叠的 5' -CG-3' 或重叠的 5' -CNG-3' 中的胞嘧啶残基被甲基化,那么这些甲基化敏感性限制性核酸内切酶不可能在其识别位点处切开。为了富集编码/独特序列,可以由用 Pst I(或其它甲基化敏感性酶)消化的基因组 DNA 构建玉米文库,并通过琼脂糖凝胶电泳进行大小分级分离。

[0095] 一种用于减少重复 DNA 的方法包括通过以下步骤构建简化的代表文库:分离重复序列与物种的至少两个变种的基因组 DNA 片段,基于核苷酸序列的大小对分离的基因组 DNA 片段进行分级分离,并比较级分中的片段的序列以确定多态性。更具体地,这些鉴定基

基因组 DNA 中的多态性的方法包括用甲基化敏感性核酸内切酶消化来自真核物种的至少两个变种的总基因组 DNA, 以提供消化的 DNA 片段的集合。对于特征为较低的 5- 甲基化胞嘧啶的 DNA 区域, 片段的平均核苷酸长度较小。这些片段是基于核苷酸长度可分开的, 例如通过凝胶电泳分开。从消化的 DNA 的集合中分离具有小于平均核苷酸长度的 DNA 级分。比较级分中的 DNA 序列, 以鉴定多态性。与编码序列相比, 重复序列更可能包含 5- 甲基化的胞嘧啶, 例如, 在 -CG- 和 -CNG- 序列片段中。在该方法的一个实施方案中, 用选自 Aci I、Apa I、Age I、Bsr FI、BssH II、Eag I、Eae I、Hha I、HinP1I、Hpa II、Msp I、MspM II、Nar I、Not I、Pst I、Pvu I、Sac II、SmaI、Stu I 和 Xho I 的甲基化敏感性核酸内切酶消化来自一种作物植物的至少两个不同近交品种的基因组 DNA, 以提供消化的 DNA 的集合, 该集合可以通过如凝胶电泳物理分离。从所述每个品种的消化的 DNA 获得大小相当的 DNA 级分。将来自相当的级分的 DNA 分子插入载体, 以构建简化的基因组 DNA 克隆的代表性文库, 对其测序并比较以鉴定多态性。

[0096] 一种用于富集编码区 DNA 序列的替代方法使用 McrBC 限制性核酸内切酶, 其切开含有甲基化胞嘧啶的 DNA。可以使用基因组 DNA 片段构建简化的代表文库, 该基因组 DNA 片段通过物理剪切或用任何限制性酶消化来切开。

[0097] 另一种富集编码 / 独特序列的方法包括: 构建简化的代表文库 (使用甲基化敏感性的或非甲基化敏感性的酶), 在尼龙膜上印刷文库的微阵列, 接着与由已知存在于文库中的重复元件制成的探针杂交。鉴定重复序列元件, 并通过只挑选阴性克隆重新排列文库。这种方法提供了来自植物的简化代表性基因组 DNA 的片段, 所述植物具有包括有相对较高水平的甲基化胞嘧啶的 DNA 区域和具有相对较低水平的甲基化胞嘧啶的 DNA 区域的基因组 DNA。本发明的简化的代表性片段包括来自具有相对较低水平的甲基化胞嘧啶的 DNA 区域的基因组 DNA, 并且在特征为所述片段的核苷酸大小在例如 500-3000bp 范围内的级分中提供。

[0098] 对玉米基因组 DNA 样品中的多态性的分型

[0099] 通过本领域熟知的多种有效方法可以检测 DNA 序列中的多态性或对其进行分型, 这些方法包括但不限于那些在美国专利 5, 468, 613 和 5, 217, 863、5, 210, 015、5, 876, 930、6, 030, 787、6, 004, 744、6, 013, 431、5, 595, 890、5, 762, 876、5, 945, 283、5, 468, 613、6, 090, 558、5, 800, 944 和 5, 616, 464 中公开的方法, 本文完整引入这些专利作为参考。然而, 本发明的组合物和方法可以与任一种多态性分型方法结合使用, 以对玉米基因组 DNA 样品中的多态性进行分型。所用的这些玉米基因组 DNA 样品包括但不限于直接从玉米植物中分离出的玉米基因组 DNA、克隆的玉米基因组 DNA 或扩增的玉米基因组 DNA。

[0100] 例如, 如美国专利 5, 468, 613 和 5, 217, 863 所公开的, 通过与等位基因特异性的寡核苷酸 (ASO) 探针杂交可以检测 DNA 序列中的多态性。美国专利 5, 468, 613 公开了等位基因特异性的寡核苷酸的杂交, 其中, 通过以下程序可以对核酸检测核酸序列中的一个或多个核苷酸变异, 在该程序中, 扩增含有核苷酸变异的序列, 点样到膜上, 并用标记的序列特异性寡核苷酸探针进行处理。

[0101] 也可以通过美国专利 5, 800, 944 公开的探针连接方法检测靶核酸序列, 其中, 扩增目标序列, 并将其与探针杂交, 接着进行连接以检测该探针的标记的部分。

[0102] 微阵列也可用于多态性检测, 其中, 以重叠的方式组装寡核苷酸探针组以代表

一种序列,这样,靶序列在一个点的差异会导致部分探针杂交 (Borevitz 等人, Genome Res. 13 :513-523 (2003) ;Cui 等人, Bioinformatics 21 :3852-3858 (2005))。在任何一个微阵列上,预计会有多个靶序列,它们可以代表基因和 / 或非编码区,其中每个靶序列由一系列重叠的寡核苷酸,而不是由一个探针所代表。该平台允许高通量筛选多种多态性。单特征多态性 (SFP) 是通过寡核苷酸阵列中的单探针检测的多态性,其中,特征是阵列中的探针。美国专利 6,799,122、6,913,879 和 6,996,476 公开了通过基于微阵列的方法对靶序列的分型。

[0103] 也可以通过美国专利 5,616,464 公开的探针连接方法检测靶核酸序列,该方法采用至少一对探针,该探针具有与靶核酸序列的相邻部分同源的序列且具有侧链,所述侧链在所述探针与所述靶核酸序列碱基配对时非共价结合以形成茎。至少一个侧链具有光可活化的基团,该基团可以与茎的其它侧链成员形成共价交联。

[0104] 检测 SNP 和 Indel 的其它方法包括单碱基延伸 (SBE) 方法。SBE 方法的例子包括但不限于:在美国专利 6,004,744、6,013,431、5,595,890、5,762,876 和 5,945,283 中公开的方法。SBE 方法基于直接邻近多态性的核苷酸引物的延伸,以在引物延伸后掺入可检测的核苷酸残基。在某些实施方案中,SBE 方法使用三种合成寡核苷酸。其中两种寡核苷酸作为 PCR 引物,且与位于含有待测多态性的区域侧翼的玉米基因组 DNA 的基因座序列互补。在对含有多态性的玉米基因组区域扩增后,PCR 产物与第三寡核苷酸 (称为延伸引物) 混合,所述第三寡核苷酸设计为在 DNA 聚合酶和两种差异标记的双脱氧核苷三磷酸的存在下,与直接邻近多态性的扩增的 DNA 杂交。如果模板上存在多态性,则可以在单碱基链延伸中将一个标记的双脱氧核苷三磷酸添加到引物中。然后通过确定两个差别标记中的哪一个被添加到延伸引物中来推断存在的等位基因。纯合的样品将导致两个标记的碱基中只有一个被掺入,因此两个标记中只有一个被检测到。杂合的样品存在两个等位基因,因此直接掺入两个标记 (进入延伸引物的不同的分子中),因此这两个标记都被检测到。

[0105] 在一种优选的检测多态性的方法中,可以通过美国专利 5,210,015、5,876,930 和 6,030,787 中公开的方法检测 SNP 和 Indel,其中,具有 5' 荧光报告染料和 3' 猝灭染料的寡核苷酸探针与探针的 5' 和 3' 端共价连接。当探针完整时,报告染料接近猝灭染料导致报告染料荧光被抑制,例如通过福斯特型能量转移 (Forster-type energy transfer)。在 PCR 正向和反向引物与位于多态性侧翼的靶 DNA 的特定序列杂交过程中,杂交探针与位于扩增的 PCR 产物中的含有多态性的序列杂交。在随后的 PCR 循环中,具有 5' → 3' 核酸外切酶活性的 DNA 聚合酶切割探针,并分离报告染料与猝灭染料,导致报告染料的荧光增强。

[0106] 一种有用的试验是可从 AB Biosystems 获得的 Taqman® 试验,其在一个反应中采用 4 种合成寡核苷酸,其同时扩增玉米基因组 DNA,区别存在的等位基因,并直接提供用于区别和检测的信号。4 种寡核苷酸中的 2 种作为 PCR 引物,并产生包含待检测的多态性的 PCR 产物。另外两个是等位基因特异性的荧光共振能量转移 (FRET) 探针。在这种试验中,使用两种携带不同的荧光报告染料的 FRET 探针,其中,将独特的染料掺入可以与两个等位基因中的仅仅一个高特异性退火的寡核苷酸中。有用的报告染料包括但不限于 6-羧基 -4,7,2',7'-四氯荧光素 (TET)、2' - 氯 -7' - 苯基 -1,4-二氯 -6-羧基荧光素 (VIC) 和 6-羧基荧光素亚磷酰胺 (FAM)。一种有用的猝灭剂是 6-羧基 -N,N,N',N'-四甲基罗丹明 (TAMRA)。此外,化学封闭每个 FRET 探针的 3' 末端,使得它不能作为 PCR 引物发挥作

用。也存在用作被动参比的第三荧光团,例如罗丹明 X (ROX),以帮助有关的荧光值在随后的标准化(校正反应装配中的空间误差)。启动基因组 DNA 的扩增。在每一个 PCR 循环中,FRET 探针以等位基因特异性的方式与模板 DNA 分子退火。当酶遇到退火探针的 5' 端时,退火的(不是非退火的)FRET 探针被 TAQ DNA 聚合酶降解,从而从猝灭剂附近释放出荧光团。PCR 之后,使用荧光计确定两个荧光团各自的荧光,以及被动参比的荧光。两种染料各自的荧光的标准化强度与样本中最初存在的每种等位基因的量成比例,因此,可以推断出样本的基因型。

[0107] 为了设计用于该试验的引物和探针,首先掩蔽基因座序列,以防止这三个引物中的任何一个被设计到匹配已知的玉米重复元件(例如,转座子)或者具有非常低的序列复杂性(二-或三-核苷酸重复序列)的位点。引物向这些重复元件的设计,通过多个基因座的扩增或 FRET 探针与多个位点的退火,导致低特异性的试验。

[0108] PCR 引物设计为:(a) 在多态性基因座中具有 15-25 个碱基大小的长度和匹配序列,(b) 具有 57-60°C 的计算解链温度,例如,对应于 52-55°C 的最佳 PCR 退火温度,(c) 产生包含多态性位点并且通常具有 75-250 个碱基对大小的长度的产物。但是,在美国专利 6,410,277 中也公开了允许扩增长度高达 1000 个或更多碱基对的片段的 PCR 技术。PCR 引物优选位于基因座上,使得多态性位点离开每个 PCR 引物的 3' 端至少一个碱基。但是,应当理解,PCR 引物可以远离多态性多达 1000 个碱基对,并仍然提供 1000 个或更多碱基对的相应 DNA 片段的扩增,所述 DNA 片段包含多态性,并可以用于分型试验。PCR 引物不能包含广泛地自身或相互互补的区域。

[0109] 设计 FRET 探针以跨越多态性位点的序列,优选地多态性位于寡核苷酸的 3' 的 2/3 处。在优选的实施方案中,FRET 探针在其 3' 末端掺入化学部分,当探针与模板 DNA 退火时,所述化学部分与 DNA 的小沟结合,从而提高探针-模板复合物的稳定性。探针应该具有 12-17 个碱基的长度,并具有 3' MGB,具有比 PCR 引物高 5-7°C 的计算解链温度。美国专利 5,538,848、6,084,102 和 6,127,121 公开了探针设计。

[0110] 也涉及通过使用瓣核酸内切酶介导的试验(**Invader®**, ThirdWave Technologies, Madison Wisconsin)对单核苷酸多态性进行分型的寡核苷酸探针。在这些试验中,瓣核酸内切酶(裂解酶)切开由两个重叠的寡核苷酸与分型的序列杂交产生的三股螺旋(Lyamichev 等人, Nat. Biotechnol., 17:, 292-296, 1999)。该分型的序列可以是玉米基因组 DNA、克隆的玉米基因组 DNA 或扩增的玉米基因组 DNA。切割一个与待分型的序列杂交的寡核苷酸释放瓣(flapp),所述瓣又与“FRET 盒”寡核苷酸形成三股螺旋,导致释放荧光共振能量转移(FRET)标记的第二切割反应。已描述了使用一种 FRET 标记对多态性的一个等位基因进行分型的实施方案(Mein C. A., 等人 Genome Res., 10:, 330-343, 2000)。在这种方法的其它实施方案中,可以通过使用不同类型的 FRET 标记同时对多态性的两个等位基因进行分型(Lyamichev 等人, 同上)。也描述了瓣核酸内切酶介导的高通量试验,其适于产生用于对多种多态性进行分型的核苷酸组(Olivier, 等人, Nucleic Acids Res. 30(12): e53, 2002)。

[0111] 适于用裂解酶对表 1 或表 3 的多态性进行分型的分离的核酸分子可以包括至少一个具有“通用”5' 瓣序列的第一探针、至少一个第二或“**Invader®**”探针和至少一个包含标记碱基和猝灭碱基的“FRET”盒,该盒包含与一经裂解即从第一探针上释放的“通用瓣核

酸序列”互补的序列。当扩增分型的玉米基因组 DNA 序列时,也可以使用类似于上文所述的侧翼 PCR 引物。这些探针的设计只需要在表 1 或表 3 中指出的多态性碱基的任一侧提供大约 40-50 个核苷酸。在“SingleNucleotide Polymorphisms”(Methods and Protocols) Volume 212, Chapter 16, V. Lyamichev and B. Neri pp. 229-240 Humana Press. 2002 中描述了设计用于核酸内切酶试验的探针的一般方面。

[0112] 应用多态性建立标记 / 性状关联

[0113] 本发明的基因座中的多态性可用于标记 / 性状关联的鉴定,这种关联是从群体成员的基因型和表型的统计分析推断出的。这些成员可以是单个生物体,例如玉米,密切相关的个体的家族、近交系、密切相关的个体的加倍的单倍体或其它群体。这样的玉米群体被称为“系”,表示起源系。群体可以起源于两个个体或两个系(例如,定位的群体)之间的单个杂交,或者,它可以由具有多个起源系的个体组成。每个个体或系的特征在于单个或平均的性状表型和在一个或多个标记基因座处的基因型。

[0114] 可以利用几种类型的统计学分析从表型 / 基因型数据推断标记 / 性状的关联,但一个基本的想法是检测分子标记,即多态性,对于多态性,可替代的基因型具有显著不同的平均表型。例如,如果给定的标记基因座 A 具有 3 个可替代的基因型(AA、Aa 和 aa),且如果那 3 类个体具有显著不同的表型,那么可以推断基因座 A 与该性状相关。可以通过几种类型的标准统计学检验测试表型的差异的显著性,如分子标记基因型对表型的线性回归或方差分析(ANOVA)。通常用来进行这种类型的分析的市售统计软件包包括 SAS Enterprise Miner(SAS Institute Inc., Cary, NC) 和 Splus(Insightful Corporation, Cambridge, MA)。当同时测试许多分子标记时,在宣布关联所需的显著性水平上进行如 Bonferonni 修正的调整。

[0115] 为了 QTL 作图,包括的标记应当是来源特征性的,以对随后的群体作出推断。基于 SNP 的分子标记对于作图是理想的,因为特定的 SNP 等位基因源自特定物种的现存群体中的独立来源的可能性极低。因此,SNP 标记可用于示踪和协助 QTL 的渗入,特别是在单元型的情况下。

[0116] 通常,关联研究的目标不只是检测标记 / 性状关联,而且评估直接影响性状的基因(即, QTL)相对于标记位置的位置。在实现该目标的一个简单的方法中,在标记基因座之间比较替代基因型之间的差异大小或差异显著性的水平。推断性状基因位于最接近具有最大相关的基因型差异的标记处。可以通过基因作图模型建立另外的标记分子的遗传连锁,所述基因作图模型例如,但不限于, Lander 等人(Lander 等人, Genetics, 121:185-199(1989))报道的侧翼标记模型,和区间作图(interval mapping),其基于其中所述的最大似然法,并用软件包 MAPMAKER/QTL 执行(Lincoln 和 Lander, Mapping Genes Controlling Quantitative Traits Using MAPMAKER/QTL, Whitehead Institute for Biomedical Research, Massachusetts, (1990))。另外的软件包括 Qgene, Version 2.23(1996), Department of Plant Breeding and Biometry, 266 Emerson Hall, Cornell University, Ithaca, NY。使用 Qgene 软件是一种特别优选的方法。

[0117] 对于标记的存在,计算最大似然估计值(MLE),以及假设没有 QTL 效应的 MLE,以避免假阳性。然后计算优势率的 $\log_{10}(\text{LOD}) : \text{LOD} = \log_{10}(\text{假定没有相关的 QTL, 对于 QTL/MLE 的存在的 MLE})$ 。LOD 得分基本上指示,假定存在 QTL,相对于不存在 QTL,获得数

据的可能性增大多少。为了避免比如 95% 的给定置信度时的假阳性的 LOD 阈值取决于标记的数量和基因组的长度。Lander 等人 (1989) 说明了显示 LOD 阈值的曲线图, 且 Arús 和 Moreno-González, Plant Breeding, Hayward, Bosemark, Romagosa (编) Chapman & Hall, London, 第 314-331 页 (1993) 进一步对其描述。

[0118] 可以使用另外的模型。已经报导了对于区间作图的许多修改和可供选择的方法, 包括使用非参数法 (Kruglyak 等人, Genetics, 139:1421-1428 (1995))。也可以使用多元回归方法或模型, 其中, 在大量标记上对性状进行回归 (Jansen, Biometrics in Plant Breed, van Oijen, Jansen (编) Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, 第 116-124 页 (1994); Weber 和 Wricke, Advances in Plant Breeding, Blackwell, Berlin, 16 (1994))。Jansen 等人 (Jansen 等人, Genetics, 136:1447-1455 (1994)) 和 Zeng (Zeng, Genetics 136:1457-1468 (1994)) 报道了组合了区间作图与回归分析的程序, 由此将表型回归到给定的标记区间的单个假定的 QTL 上, 且同时回归到作为 '辅助因素' 的许多标记上。一般来说, 辅助因素的使用减少了估计的 QTL 位置的偏差和抽样误差 (Utz 和 Melchinger, Biometrics in Plant Breeding, van Oijen, Jansen (编) Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, 第 195-204 页 (1994)), 从而提高 QTL 作图的精度和效率 (Zeng 1994)。这些模型可以扩展到多环境实验, 以分析基因型 - 环境的相互作用 (Jansen 等人, Theor. Appl. Genet. 91:33-3 (1995))。

[0119] 传统的 QTL 作图的替代方法包括通过相对于各个标记对单元型作图以实现更高的分辨率 (Fan 等人 2006 Genetics 172:663-686), 因为传统的 QTL 作图研究的一种局限性是推断仅限于作图群体的特定亲本和这些亲本变种的基因或基因组合的事实。这种方法跟踪被称为单元型的 DNA 模块, 该 DNA 模块由多态性标记所定义, 假定它们在作图群体中来源相同。长期以来一直认为, 基因和基因组序列可以是状态同一的 (即, 独立起源相同) 或来源同一的 (即, 通过来自共同祖先的历史遗传), 这对于连锁不平衡研究、且最终对作图研究具有巨大的意义 (Nordberg 等人 2002 Trends Gen.). 从历史上看, 遗传标记不适合区分状态同一或来源同一。然而, 标记的新类型, 如 SNP (单核苷酸多态性), 更能说明起源。特定 SNP 等位基因源自特定物种的现存群体中的独立来源的可能性非常低。连锁基因中出现的多态性以缓慢但可预测的速度随机分类, 由连锁不平衡的衰减或连锁平衡的方法描述。这种良好建立的科学发现的后果是由多态性的特定组合确定的一长段编码 DNA 是非常独特的, 并且除了通过连锁不平衡以外, 非常不可能重复存在, 这指示来自共同祖先的最近的共祖先。由一些等位基因组合定义的特定基因组区域表示整个间插遗传序列的绝对同一性的可能性取决于该基因组区域中的连锁的多态性的数量, 除非在该区间中最近发生了突变。在此, 这样的基因组区域被称为单元型窗口。该窗口内的每个单元型由等位基因的特定组合定义; 等位基因的数量越大, 潜在的单元型的数目就越大, 而且该区域中状态同一性是来源同一性的结果的确确定性就越大。在新系的开发过程中, 祖先单元型通过这一过程保留, 并且一般被认为是作为单元通过谱系遗传的 '连锁块'。此外, 如果特定的单元型具有已知的效应, 或者表型, 可以推断它在具有相同单元型的其它系中的效应, 这可以使用对于该单元型窗口的一个或多个诊断标记来确定。

[0120] 这一假设导致较大的有效样本量, 提供更大的 QTL 分辨率。用于确定表型和基因型 (在这种情况下为单元型) 之间相关性的统计学显著性的方法可以通过本领域已知的并且具有任何公认的所需统计学显著性阈值的任何统计学检验来确定。特定的方法和显著性阈值的应用是本领域的普通技术人员所熟知的。

[0121] 遗传图谱的构建

[0122] 在本发明的另一方面, 本发明的基因座中的多态性定位于玉米基因组上, 例如, 作为玉米基因组的遗传图谱, 其包括如表 1 所示的、更优选地如表 3 所示的两种或多种多态性的图谱位置。这种遗传图谱如图 1 所示。遗传图谱数据也可以记录在计算机可读介质上。本发明的优选实施方案提供高密度的 (例如在玉米基因组图谱上有至少 150 种或更多, 例如至少 500 或 1000 种多态性) 多态性遗传图谱。特别有用的遗传图谱包括在连锁群上的平均距离不超过 10 厘摩 (cM) 的多态性。

[0123] 连锁不平衡作图和关联研究

[0124] 另一种确定性状基因位置的方法是分析其中个体的性状和标记基因座都不同的群体中的标记 / 性状关联。在该群体中, 由于群体的遗传过程, 如突变的独特起源、建立者事件 (founder events)、随机漂变和群体结构, 某些标记等位基因可能与某些性状基因座等位基因相关联。这种关联被称为连锁不平衡。

[0125] 在植物育种群体中, 连锁不平衡 (LD) 是离开群体中两个或多个基因座之间的随机关联的水平, 且 LD 往往存在于大的染色体片段上。虽然有可能关注该片段中每个基因单独效应, 但是对于实际的植物育种来说, 一般强调当区域存在于系、杂种或变种中时对目标性状的平均影响。在连锁不平衡作图中, 比较在标记基因座处具有不同基因型的个体的性状值。通常, 显著的性状差别表明标记基因座与一个或多个性状基因座之间非常接近。如果标记密度适当地高, 且连锁不平衡只在染色体上非常紧密连锁的位点之间发生, 那么性状基因座的位置可以非常精确。

[0126] 标记辅助的育种和标记辅助的选择

[0127] 当数量性状基因座 (QTL) 已被定位于分子标记的附近时, 这些标记可以用来针对提高的性状值进行选择, 而无需在每个选择循环时进行表型分析。在标记辅助的育种和标记辅助的选择中, 首先通过遗传作图分析建立 QTL 与标记之间的关联 (如在 A. 1 或 A. 2 中)。在同样的过程中, 确定哪些分子标记等位基因与有利的 QTL 等位基因连锁。随后, 在群体中选择与有利的 QTL 等位基因相关联的标记等位基因。如果在标记与 QTL 之间有足够紧密的连锁, 此过程将提高性状值。所需的连锁程度取决于选择的代数, 因为在每一代, 有机会通过重组打破关联。

[0128] 特定标记等位基因与有利的 QTL 等位基因之间的关联还可以用于预测哪些类型的后代可以从给定的杂交中分离。这种预测可以允许选择适合于产生群体的亲本, 从该群体中装配有利的 QTL 等位基因的新组合以产生新的近交系。例如, 如果系 A 在基因座 1、20 和 31 处具有以前已知与有利的 QTL 等位基因相关联的标记等位基因, 而系 B 在基因座 15、27 和 29 处具有与有利的效应相关联的标记等位基因, 那么可以通过杂交 AxB 并选择在全 6 个 QTL 处具有有利的等位基因的后代来开发新系。

[0129] 分子标记用于加速转基因向新的遗传背景中的渗入 (即, 进入不同范围的种质中)。简单的基因渗入包括使转基因系与优良近交系杂交, 然后使该杂种与优良 (轮回) 亲

本反复回交,同时针对转基因的保持进行选择。经过多代回交,通过重组和分离,初始转基因系的遗传背景逐渐被优良近交系的遗传背景所取代。通过根据源自回交亲本的分子标记等位基因进行选择,可以加速这个过程。

[0130] 此外,近交系的指纹是在一组两个或多个标记基因座处等位基因的组合。高密度指纹可以用来建立和追踪种质的身份,种质身份可用于建立标记-性状关联的数据库,以益于整个作物育种程序,以及种质所有权的保护。

[0131] 选择用于植物育种的亲本亲本、后代或测试植物的方法

[0132] 也想到本文提供的多态性可以用于为植物育种选择亲本、后代或测试植物。从表型上无法区分的植物的群体中选择这些植物的能力可以加速植物育种并减少因进行表型性状分析而导致的费用。选择用于育种的植物的方法包括以下步骤:a) 确定表 1 或表 3 中确定的多种多态性与至少第一和第二玉米近交系中的多种性状之间的关联;b) 确定亲本、后代或测试植物中的一种或多种多态性的等位基因状态;和 c) 选择具有更有利的相关性状组合的亲本、后代或测试植物。在某些应用中,通过这种方法选择的亲本、后代或测试植物是玉米近交系。在其它实施方案中,相关性状的有利组合提供了改善的杂种优势。

[0133] 在一个实施方案中,确定至少两种多态性的基因型有助于选择用于育种杂交的亲本。这种确定给育种者提供了产生杂交的优势,其中针对至少两个优选的基因组区域,以产生具有至少两个优选的基因组区域的后代。在另一方面,确定至少两种多态性的基因型可以为在后代中作出选择决定提供基础,其中,那些包含优选的基因组区域的后代在育种计划中被选出。在另外一方面,可以选择用于评估近交系在杂种组合中的组合能力的测试系,纳入基于存在或不存在至少两个基因组区域的近交测试计划中,以确保在不同的种质库(即不同的杂种优势群)之间进行杂交。

[0134] 杂种预测

[0135] 通过在两个属于不同的“杂种优势群”的优良近交系之间进行杂交而产生商品玉米种子。这些群在遗传学上足够不同,使得它们之间的杂种显示高水平的杂种优势(即,相对于亲本系性能提高)。通过分析优良杂种的标记组成,可以鉴定在良好组合产生杂种优势的雄系和雌系中的不同基因座处的等位基因组。认识这些模式并了解不同近交系的标记组成,可以预测不同对品系之间的杂种优势的水平。这些预测可以减少应使用相反杂种优势群的哪些品系测试新近交系的性能的可能性。

[0136] 本发明提供了用于提高杂交玉米的杂种优势的方法。在这些方法中,在与本发明多态性基因座连锁的多种多态性与两个以上的玉米近交系中的性状之间建立关联。选择两个这样的具有预测可提高杂种优势的互补性杂种优势群的近交系用于育种。提高杂种优势的方法包括以下步骤:(a) 确定表 1 或表 3 中确定的多种多态性与两个以上的玉米近交系中的多种性状之间的关联;(b) 将选自步骤(a)的近交系的两个近交系分配至杂种优势群;(c) 在步骤(b)的至少两个近交系之间进行至少一次杂交,其中,每个近交系来自不同的和互补的杂种优势群,并且其中,对于提高杂种优势的遗传特征优化互补杂种优势群;和(d) 通过步骤(c)的所述杂交获得杂种后代植物,其中,相对于与未经选择的近交系杂交产生的后代,所述杂种后代植物显示提高的杂种优势。这些方法还可以在步骤(c)中包括传统的单杂交(即,两个近交系之间,理想地来自不同的杂种优势群)、三元杂交(单杂交后,与第三近交系杂交)和双杂交(也称为四元杂交,即两个单杂交的后代杂交)。可以通

过在选择雄性不育亲本之间进行手工杂交或通过使用雄性不育杂交系统实现杂交。在 Bernardo, *Breeding for Quantitative Traits in Plants*, Stemma Press, Woodbury, MN, 2002 中描述了优良近交系的开发和选择、这些系的杂交和选择优良杂种杂交鉴定新的优良玉米杂种。

[0137] 遗传来源同一性

[0138] 杂种优势的一种理论预测,在用于产生杂种的雄性和雌性系之间的遗传来源同一性 (IBD) 区域会降低杂种性能。可以从不同系中的标记等位基因的模式推断遗传来源同一性。如果在一系列邻近的基因座处的一串相同标记不可能偶然地独立发生,则可以认为它们是遗传来源同一的。雄性和雌性系中的标记指纹分析可以鉴定 IBD 区域。对这些区域的知识有助于选择杂种亲本,因为在杂种中避免 IBD 可能提高性能。这种知识也有助于育种计划,其中可以设计杂交以产生显示很少或没有 IBD 的近交系对 (一雄一雌)。

[0139] 用于基因分型的核酸分子文库

[0140] 本发明提供的核酸文库可用于与玉米种质改良相关的活动,包括但不限于使用植物进行育种杂交,对植物的进一步的遗传或表型测试,植物通过自体受精的改进,使用植物或其部分进行转化,以及使用植物或其部分进行诱变。可以对文库中的不同组核酸采样,访问,或者对其任何组、亚组或组合单独进行查询,以对本文表 1 或 3 中提供的任何玉米基因组 DNA 进行分型。一般来说,文库包括至少两组不同的核酸分子,其中所述不同核酸分子组中的每一个允许对表 1 或表 3 中确定的相应的玉米基因组 DNA 多态性进行分型。

[0141] 在一个实施方案中,允许对表 1 或表 3 中确定的相应的玉米基因组 DNA 多态性进行分型的不同组核酸分子分布于微量滴定板的各个孔中。在某些实施方案中,微量滴定板的每个孔中包含一种或多种允许对表 1 或表 3 中确定的仅一种玉米多态性进行分型的核酸分子。但是,也涉及其它实施方案,其中,微量滴定板的每个孔中包含一种或多种允许对表 1 或表 3 中确定的一种以上的玉米多态性进行分型的核酸分子。微量滴定板可以具有至少 8 个孔,或多达 24、96、384、1536 或 3456 个孔。微量滴定板可以由以下材料制造,包括但不限于,聚苯乙烯、聚丙烯或环-烯烃塑料。每个孔中的核酸分子可以在溶液或是干燥的 (即,冻干形式)。通常,核酸分配到微量滴定板的孔中,使得微量滴定板每孔中的核酸是已知的。但是,在核酸分子与独特的标识物 (如独特的染料或其它独特的识别标记) 相关联的其它实施方案中,核酸可以随机地分配到微量滴定板的孔中。从本说明书中可以清楚地看出,也涉及包括分配在微量滴定板孔中的、固定于固体载体 (如珠子) 上的核酸的文库。

[0142] 在其它实施方案中,允许对表 1 或表 3 中确定的玉米基因组多态性进行分型的核酸固定 (即,共价连接) 于固体载体上。固体载体包括但不限于珠子、芯片、阵列或过滤器。

[0143] 用作固体载体的珠子可以是磁珠,以帮助杂交复合物的纯化。或者,珠子可以包含独特的识别标记。特别地,用可以根据其分光光度或荧光性质进行区分的荧光染料染色的珠子,可以偶联到用于对多态性进行分型的核酸分子上。这些用于对多态性进行分型的基于珠子的系统已有描述 (美国专利 5,736,330)。染料标记的珠子、分析试剂和用于对多态性进行分型的装置也已有描述 (美国专利 6,649,414、6,599,331 和 6,592,822),并且可从 Luminex Corporation (Austin, Texas, USA) 获得。如上所述,与珠子连接的文库核酸分子也可以是

[0144] 芯片、阵列或过滤器还可以用于固定对表 1 或表 3 的多态性进行分型的核酸分子。

在某些实施方案中,用于对给定的多态性进行分型的核酸标记将固定于阵列上规定的物理位置,使得可以产生并记录来自对应于给定多态性的位置的分型数据,用于随后的分析。制造及使用用于对多态性进行分型的阵列的方法包括但不限于在美国专利 5,858,659(基于杂交的方法)和美国专利 6,294,336(单碱基延伸方法)中所描述的方法。

[0145] 应用多态性分析对 DNA 克隆文库进行作图

[0146] 由本发明的分子标记代表的多态性和基因座可用于鉴定和定位与分子标记连锁的 QTL 和基因的 DNA 序列。例如,可以使用与性状连锁的分子标记查询 BAC 或 YAC 克隆库,以找到包含与性状相关的特定 QTL 和基因的克隆。例如,多种(如数百种或数千种)大的多基因序列中的 QTL 和基因可以通过与寡核苷酸探针杂交来鉴定,所述寡核苷酸探针能够与定位的和/或连锁的分子标记杂交,其中,可以检测一个或多个分子标记。通过在高密度阵列中提供克隆序列可以改进这种杂交筛选。该筛选方法更优选地通过采用汇集策略来改进,以明显减少鉴别包含分子标记的克隆所需要的杂交数。当对分子标记作图时,筛选能够有效地将克隆作图。

[0147] 例如,在数千个克隆排列于规定的阵列中例如在 96 孔板中的情况下,这些板可以任意地排列,形成三维排列的孔的堆叠,每个孔包括独特的 DNA 克隆。每个堆叠中的孔可以表示为行、列和板的三维阵列中的独立要素。在发明的一方面,堆叠数目和每个堆叠中板的数目大致相等,以使试验次数减至最少。板的堆叠允许构建克隆 DNA 池。

[0148] 对于三维排列的堆叠,可以为以下要素创建克隆 DNA 池:(a) 每一行的所有要素,(b) 每一列的所有要素,和 (c) 每块板的所有要素。用可与针对一个克隆独特的分子标记杂交的寡核苷酸探针杂交筛选该池将为一个列的池、一个行的池和一块板的池提供阳性指示,从而指示包含目标克隆的孔单元(要素)。

[0149] 在多堆叠的情况下,每个堆叠中所有克隆 DNA 的其它池允许指示具有目标克隆的行-列-板坐标的堆叠。例如,4608 个克隆的组可以排列于 48 块 96 孔板中。48 块板可以安排在 8 组各 6 块板的堆叠中,提供 6x12x8 三维阵列的要素,即每个堆叠包括 8 行和 12 列的 6 个堆叠。对于整个克隆组,有 36 个池,即 6 个堆叠的池、8 个行的池、12 个列的池和 8 个堆叠的池。因此,需要最多 36 个杂交反应以找到包含与每个作图分子标记相关或连锁的 QTL 或基因的克隆。

[0150] 一旦鉴定了克隆,从分子标记的基因座设计的寡核苷酸引物就可以用于连锁 QTL 和/或基因的定位克隆。

[0151] 计算机可读介质和数据库

[0152] 本发明的核酸分子的序列可以在多种介质中“提供”,以方便使用,例如,数据库或计算机可读介质,它们也可以以允许熟练技术人员检查或查询序列并获取有用信息的形式包含描述性注释。在本发明的一个实施方案中,可以准备包含核酸序列的计算机可读介质,这些核酸序列中至少 10%或以上,如至少 25%或甚至至少 50%或以上的基因座和核酸分子的序列代表本发明的分子标记。例如,这样的数据库或计算机可读介质可以包括本发明的基因座组或用于检测本发明分子标记的引物和探针组。此外,这样的数据库或计算机可读介质可以包括本发明的作图或未作图的分子标记的图或表和遗传图谱。

[0153] 本文所用的“数据库”是指任何可检索的所收集数据的任何表现形式,包括计算机文件,如文本文件、数据库文件、电子表格文件和图像文件、印刷表格和图形表示及数字和

图像数据集合的组合。在本发明的一个优选方面，“数据库”是指可以存储计算机可搜索的信息的存储系统。目前，优选的数据库应用程序包括由 DB2、Sybase 和 Oracle 提供的数据库应用程序。

[0154] 本文所用的“计算机可读介质”是指任何可以由计算机直接读取并访问的介质。这样的介质包括但不限于：磁性存储介质，如软盘、硬盘、存储介质和磁带；光存储介质，如 CD-ROM；电子存储介质，如 RAM、DRAM、SRAM、SDRAM 和 ROM；和 PROM (EPROM、EEPROM、Flash EPROM)，以及这些类别的杂合体，例如磁 / 光存储介质。熟练技术人员可以容易地了解如何使用任何目前已知的计算机可读介质创建包括在其上记录有本发明核苷酸序列的计算机可读介质的产品。

[0155] 本文所用的“记录”是指在可检索的数据库或计算机可读介质中存储信息的过程的结果。例如，熟练技术人员可以容易地采用目前已知的任何一种方法在计算机可读介质上记录信息，以产生包含本发明的作图的多态性和其它核苷酸序列信息的介质。熟练技术人员可以获得多种数据存储结构用于创建计算机可读介质，其中，数据存储结构的选择一般基于所选择的访问存储信息的手段。此外，多种数据处理程序和格式可用于在计算机可读介质上存储本发明的多态性和核苷酸序列信息。

[0156] 计算机软件可以公开获得，其允许熟练技术人员获得计算机可读介质提供的序列信息。下面的例子证明了如何利用在 Sybase 系统上执行搜索算法如 BLAST 算法 (Altschul 等人, J. Mol. Biol. 215 :403-410 (1990), 本文引入作为参考) 和 BLAZE 算法 (Brutlag 等人, Comp. Chem. 17 :203-207 (1993), 本文引入作为参考) 的软件鉴别与具有高水平同一性的本发明基因座序列同源的 DNA 序列。可以对高同一性的序列进行比较，以找到对于玉米品种有用的多态性标记。

[0157] 本发明还提供了包含本文所述的序列信息的系统，特别是基于计算机的系统。这些系统设计为用来鉴定商业上重要的本发明核酸分子的序列片段。本文所用的“基于计算机的系统”是指用于分析核苷酸序列信息的硬件、软件和存储器。熟练技术人员可以容易地了解，任何一种现有的基于计算机的系统适用于本发明。

[0158] 如上所述，本发明的基于计算机的系统包括存储有本发明的多态性标记、遗传图谱和 / 或核酸分子序列的数据库和必要的支持和实现基因分型应用的硬件和软件，这样的基于计算机的系统可以用于读取、分类或分析玉米基因型数据。基于计算机的系统的关键部件包括：(a) 数据存储装置，包括计算机可读介质，其上记录有至少 2 种表 1 或表 3 中确定的玉米基因组 DNA 多态性；(b) 搜索装置，用于将来自至少一种测试玉米植物的玉米基因组 DNA 序列与步骤 (a) 的数据存储装置的多态性序列进行比较，以鉴定同源或非同源序列；和 (c) 检索装置，用于鉴定步骤 (b) 的测试玉米基因组序列的同源或非同源序列。用于进行 DNA 数据库查询的基于计算机的方法和系统（如装置）在美国专利 6,691,109 中描述。

[0159] 在本发明一个有用的方面，来自表 1 或表 3 的多态性玉米基因座的数据集记录在计算机可读介质上。在本发明的一方面，玉米基因组多态性在一个或多个 DNA 序列数据集中提供，即，数据集包括多达有限数目的记录在计算机可读介质上的多态性基因座的不同序列。在记录的数据集中的有限数目的多态性基因座可少至 2 种或多达 1000 种或更多，例如 5、8、10、25、40、75、96、100、384 或 500 种表 1 或表 3 的玉米基因组多态性。这些数据集可以用于基因分型应用，其中，1) 查询确定分布于玉米基因组上的多态性的多种多态性；

2) 查询聚类于区间内的多种多态性 ; 和 / 或当在大量植物中查询多种多态性时。记录在计算机可读介质上的数据集也可包括对于记录在其上的每种玉米基因组 DNA 多态性的相应的图谱位置。在其它实施方案中, 表型性状或表型性状指数数据记录在计算机可读介质上。在另外其它实施方案中, 等位基因状态与亲本、后代或测试玉米植物相关联的数据记录在计算机可读介质上。

[0160] 育种方法

[0161] 也涉及培育玉米植物的方法。培育玉米植物的方法包括以下步骤 : (a) 对于至少两个玉米植物的育种群体, 确定至少两个最多 10 厘摩的基因组窗口中的至少两个单元型的性状值 ; (b) 在所述育种群体中, 培育两个玉米植物, 以产生后代种子群体 ; (c) 在所述后代种子中, 确定至少一种表 1 或表 3 中确定的多态性在每个所述窗口中的等位基因状态, 以确定所述单元型的存在 ; 和 (d) 在所述后代种子中选择对于确定的单元型而言具有较高性状值的后代种子, 从而培育玉米植物。在这些育种方法的某些实施方案中, 确定基本上每条染色体整体上的每个相邻基因组窗口中的至少两个单元型的性状值。可以理解, 单元型区域是在多代育种中保持并且由一个或多个育种系携带的染色体片段。这些片段可以用包含在该片段中的多个连锁的标记基因座鉴定, 并且两个系中这些基因座处的共同的单元型同一性给出了这些种系携带的整个相应染色体片段的遗传来源同一性的高置信度。这些育种方法需要使用分布于玉米基因组中的多种玉米基因组多态性。

[0162] 在这种育种方法的各方面, 确定基本上每条染色体的整体上每个相邻基因组窗口中的至少两个单元型的性状值。在本方法的另一个有用的方面, 在每条染色体中高达 10 厘摩的基因组窗口中, 针对单元型产量的较高的性状值, 选择后代种子。在本发明的另一方面, 培育方法涉及提高产量, 其中, 性状值是产量性状的值, 其中, 对每个窗口中的单元型的性状值进行排序 ; 并且选择在窗口中的产量性状值高于所述窗口中的平均产量性状值的后代种子。在育种方法的某些方面, 单元型使用表 1 中确定的多态性定义, 或定义为在包括 SEQ ID NO : 1 至 SEQ ID NO : 6552 的所有 DNA 序列的分子标记组中, 或定义为与那些多态性之一连锁不平衡。

[0163] 为了利用这种方法促进育种, 计算每种性状的值或性状组合的值 (例如多性状指数) 是有用的。在多性状指数中分配给各种性状的权重可以根据育种目标而变化。例如, 如果产量是关键目标, 则在多性状指数中, 产量值可以 50-80% 加权, 成熟、倒伏、株高或抗病性可以以较低的百分比加权。

[0164] 下文阐述用于培育本发明植物的选择的、非限制性的方法。对任何杂交后代使用标记辅助的选择 (MAS) 可以加强育种计划。可以理解 : 本发明核酸标记可以用于 MAS (育种) 计划。进一步可以理解 : 可以在育种计划中使用任何商业和非商业的栽培种。例如, 如发芽活力、植物生长活力、应激抗性、抗病性、分枝、开花、结实、种子的大小、种子密度、直立性 (standability) 和脱粒性 (threshability) 等因素通常指导选择。

[0165] 对于高度可遗传的性状, 选择在一个位置评估的优良单株植物是有效的, 而对于具有低遗传性的性状, 选择应该基于从相关植物的家族的重复评估获得的平均值。流行的选择方法通常包括谱系选择、改良的谱系选择、混合选择 (mass selection) 和轮回选择。在一个优选的方面, 采用回交或轮回育种计划。

[0166] 遗传的复杂性影响育种方法的选择。回交育种可以用来将对于高度可遗传性状的

一个或几个有利的基因转移到理想的栽培种中。这种方法已被广泛用于培育抗病栽培种。各种轮回选择技术用于改善受许多基因控制的数量遗传性状。

[0167] 育种系可以在代表商业目标地区的环境中测试并与适当的标准比较两代或更多代。最好的株系是新的商业栽培种的候选株系；那些仍缺乏性状的株系可用作亲本，以产生新的群体用于进一步选择。

[0168] 对于杂种作物，新的优良杂种的开发需要开发和选择优良的近交系、杂交这些系以及选择优良的杂种。可以通过在选择的雄性能育亲本之间进行手工杂交或通过使用雄性不育系统产生杂种种子。关于亲本系的其它数据以及杂种的表型影响育种员关于是否继续进行特定杂种杂交的决定。

[0169] 谱系育种和轮回选择育种方法可用于从育种群体中开发栽培种。育种计划将来自两个或多个栽培种或各种广泛来源的理想性状组合为育种池，由该育种池通过自交和选择所需的表型来开发栽培种。可以评价新栽培种以确定哪些具有商业潜力。

[0170] 回交育种已被用来将简单遗传的、高度可遗传的性状的基因转移到作为轮回亲本的理想的纯合栽培种或近交系中。待转移的性状的来源称为供体亲本。在初始杂交后，选择具有供体亲本的表型的个体，并与轮回亲本反复杂交（回交）。产生的植物预计具有轮回亲本（例如，栽培种）的大部分属性，此外，还具有从供体亲本转移来的理想的性状。

[0171] 严格意义上来说，单种子遗传程序指种植分离群体，每株植物收获一个种子的样品，并使用一个种子样品种植下一代。当群体已从 F_2 发展到理想的近交水平时，产生该株系的植物均回溯到不同的 F_2 个体。由于一些种子不能发芽或者一些植物不能产生至少一个种子，群体中的植物数量每一代都在下降。结果，当完成了世代进步时，并非全部原先在群体中取样的 F_2 植物都有后代代表。

[0172] 加倍单倍体 (DH) 方法在较短的时间内获得等基因植物。DH 植物为植物育种员提供了宝贵的工具，特别是对于产生近交系和数量遗传研究而言。对于育种员，DH 群体特别可以用于 QTL 作图、细胞质转换和性状渗入。此外，在测试和评价用于植物育种计划的纯合系方面具有价值。所有的遗传变异都在育种杂交的后代中，这提高了选择增益。

[0173] 大多数研究和育种应用依赖于人工 DH 生产方法。最初的步骤包括植物的单倍化，这导致包括单倍体种子的群体的产生。用诱导亲本与非纯合系杂交，导致单倍体种子的产生。具有单倍体胚、但具有正常三倍体胚乳的种子进展到第二阶段。即，单倍体种子和植物是任何具有单倍体胚的植物，与胚乳的倍性水平无关。

[0174] 在从群体中选择单倍体种子后，选定的种子进行染色体加倍以产生加倍的单倍体种子。细胞谱系中自发的染色体加倍会导致正常配子产生或从单倍体细胞谱系产生未减少的配子。使用化学化合物，如秋水仙碱，可以用来增加二倍化率。秋水仙碱结合微管蛋白并阻止其聚合为微管，从而在中期停止有丝分裂，可以用来增加二倍化率，即加倍染色体数目。这些嵌合植物是自花授粉以产生二倍体（加倍的单倍体）种子。种植此 DH 种子，随后评价并用于杂种测交生产。常用于不同性状和作物的其它育种方法的描述可在以下几本参考书之一中找到 (Allard, "Principles of Plant Breeding," John Wiley & Sons, NY, U. of CA, Davis, CA, 50-98, 1960; Simmonds, "Principles of crop improvement," Longman, Inc., NY, 369-399, 1979; Sneep 和 Hendriksen, "Plant breeding perspectives," Wageningen (编), Center for Agricultural Publishing and Documentation, 1979; Fehr, In : Soybeans :

Improvement, Production and Uses, 第2版, Monograph., 16:249, 1987; Fehr, "Principles of variety development," Theory and Technique, (卷1) 与 Crop Species Soybean (卷2), Iowa State Univ., Macmillan Pub. Co., NY, 360-376, 1987)。

[0175] 用单分子标记进行基因分型的方法

[0176] 用单分子标记(例如, 玉米基因组多态性)进行基因分型的方法也可以用于将玉米植物的表型性状与基因型相关联。检测来自至少两个具有等位基因 DNA 的玉米植物的组织中的 DNA 或 mRNA, 以确定是否存在本发明提供的作为分子标记的多态性。鉴定分子标记与表型性状之间的关联, 其中所述标记是在表 1 或表 3 中确定的。在另一方面, 在染色体的特定基因座中具有等位基因 DNA 的玉米植物分离群体中, 将性状与基因型相关联, 所述基因座对目标性状具有表型效应, 并且其中分子标记定位于该基因座之中或附近。

[0177] 用单分子标记(例如, 玉米基因组多态性)进行基因分型的方法也可以用来选择用于育种的亲本植物、后代植物或测试植物。在这种情况下, 多态性与赋予一种或多种理想的表型特状的染色体区域遗传连锁。选择包含与表型性状相关联的特定等位基因状态的亲本、后代或测试玉米植物提供了加速的和较低成本的育种。

[0178] 预期本文在表 1 或表 3 中公开的某些玉米基因组多态性可以与给定的表型性状直接相关, 因为它们包括某些改变赋予性状或有助于性状表达的基因的调控或编码序列的等位基因状态。这些性状包括产量、倒伏、成熟、株高、抗病性, 例如对色二孢属 (diploia)、灰斑病、赤霉菌、炭疽病、镰刀霉 (fusarium) 以及其它穗和茎腐病、枯萎病、锈病、细菌性病害、昆虫病害等的抗性, 非生物应激耐受性, 例如, 耐旱性、耐寒性、耐热性、耐风暴性、营养缺乏等, 和质量性状, 例如, 提高的淀粉含量、提高的油含量、减少的饱和脂肪酸含量、提高的蛋白质含量、增加的赖氨酸含量等。当玉米基因组多态性以这种方式与性状直接关联时, 它在旨在将该性状引入许多不同的玉米遗传背景内的玉米育种计划中是非常有用的。

[0179] 可以通过使用多个标记以使与可能不提供农艺学优良性质的基因组区域相关的连锁阻力减至最小, 来加速与此单标记相关的基因组区域的渗入。可以通过使用多个直接位于单标记侧翼的标记, 以使可能与密切相关的基因组区域相关的连锁阻力减至最小, 来加速与此单标记密切关联的基因组区域的渗入。因此, 使用聚类的一组 2、5、10 或 20 个位于单标记近端和远端 10、5、2 或 1cm 的标记, 可以提供所需要的与单标记相关的基因组区域的渗入, 同时不需要的直接侧翼区域的渗入减至最少。也可以通过使用分布于基因组中的多个标记以使可能与位于同一染色体远端区域上和其它染色体上的基因组区域密切关联的任何连锁阻力减至最小, 来加速与此单标记密切相关的基因组区域的渗入。这组多个标记可以包括另外 10 个标记, 每个染色体臂有至少一个标记。然而, 在优选的实施方案中, 标记密度是每个染色体臂至少大约 10 个标记, 更优选地每个染色体臂至少大约 100 个标记, 以有效地区分来自供体和接受体亲本的基因组区域。因而, 使用与单标记直接连锁或分布于基因组上的多个侧翼标记可以提供在选择的杂交后代中最大回收接受体亲本。

[0180] 用玉米基因组 DNA 多态性组进行基因分型的方法

[0181] 本发明尤其涉及采用可以对多种不同的多态性进行分型的核酸分子组的基因分型方法。在这样的方法中, 对有限数量的至少两种玉米基因组多态性进行分型。查询的这种有限数量的玉米基因组多态性可以包含至少 2、5、10 或 20 种不同的基因型, 它们在表 1 或 3 中表示为 2、5、10 或 20 种不同的 SEQ ID NO。这些基因分型方法必然需要使用可以对

玉米基因组多态性组进行分型的核酸分子组。

[0182] 在某些应用中,这些基因分型方法使用在给定染色体区间集中的多个分子标记(即玉米基因组多态性)。用于建立和追踪种质身份的高密度指纹可以通过进行基因分型方法来获得,所述方法利用在特定染色体区间和/或赋予某些性状的某些基因座周围集中或群集的多个分子标记。高密度指纹信息可以用于评估种质多样性,行使遗传质量保证功能,开发罕见的等位基因,评估外来种质库和评估遗传纯度。这些高密度指纹可以用来建立标记-性状关联数据库,有益于整个作物育种计划。高密度指纹也可以用来建立和保护种质所有权。可以从表3提供的定位的玉米多态性中选择聚集在需要的染色体区间或遗传性状周围的标记组。

[0183] 这些用多个分子标记进行基因分型的方法也可用于将玉米植物的表型性状与基因型相关联。检测来自至少两个具有等位基因DNA的玉米植物的组织中的DNA或mRNA,以确定是否存在本发明提供的作为分子标记的一组有限系列的多态性。确定这组分子标记与这组表型性状之间的关联,其中,这组分子标记至少包括2种、至少5种、或至少10种与本发明的多态性基因座连锁的分子标记,例如至少10种与定位的多态性连锁的分子标记,例如,如表3中确定的那些。在一个更优选的方面,在对目标性状赋予表型效应的染色体基因座中具有等位基因DNA的玉米植物分离群体中,性状与基因型相关联,其中分子标记之间和多态性与性状之间的关联程度允许确定多态性和性状基因座的线性次序。在这样的方法中,至少5个分子标记与允许基因座不平衡作图的基因座连锁。

[0184] 在其它应用中,这些基因分型方法使用分布于玉米基因组中的分子标记。在这些方法中,分子标记可以分散在一个染色体上、位于多条染色体上、位于所有染色体上或位于每条染色体的每个臂上。在一个具体的实施方案中,在使用多个标记的基因分型方法中使用的至少1种分子标记定位于所有10条玉米染色体的每个染色体臂上,因此必须对至少20种玉米基因组DNA多态性进行分型。但是,也涉及该方法的其它实施方案,其中至少10种玉米基因组DNA多态性定位于每个染色体臂上,因此必须对至少200种玉米基因组DNA多态性进行分型。同样,也涉及其它实施方案,必须对每个染色体臂上的至少20种玉米基因组DNA多态性进行分型(必须对至少400种多态性进行分型),或对每个染色体臂上的至少50种玉米基因组DNA多态性进行分型(必须对至少1,000种多态性进行分型)。分布于玉米基因组上的标记组可以选自用于这些方法的表3提供的定位的玉米多态性。

[0185] 使用分布于玉米基因组上的分子标记的基因分型方法可以用于多种应用。在一种应用中,基因分型方法用于选择用于育种的亲本植物、后代植物或测试植物。涉及这些基因分型方法在玉米育种计划中的多种应用。这些基因分型方法可用于促进一种或多种性状、基因组基因座的渗入和/或转基因从一个遗传背景向不同的遗传背景中的插入。一般来说,查询来自远交(out-crossed)群体的后代植物中选择的标记组,以鉴定并选择包含所需的性状、基因组基因座和/或转基因插入、而仍包含尽可能多的来自远交的不同遗传背景的等位基因的个体后代。这些方法可以通过几代加速所需的性状、基因组基因座的渗入和/或转基因向新遗传背景中的插入。

[0186] 这些方法还通过探询玉米遗传图谱上平均密度小于大约10cM的分子标记如SNP的集合提供性状筛选。可以在一种或多种表型性状的范围内,分析与表1或表3的多态性基因座连锁的分子标记的存在与否,以鉴定在与一种或多种所述性状相关的一个或多个基

因组区域处的一种或多种特定分子标记等位基因。在本发明的另一方面,利用分子标记鉴定单元型,该单元型是基因组 DNA 的等位基因片段,其特征在于处于连锁不平衡的至少两种多态性,并且所述多态性在不超过 10 厘摩长度的基因组窗口中,例如,不超过大约 8 厘摩或更小的窗口中,例如,在 1-5 厘摩的范围内。在这些方法的某些实施方案中,这样的分子标记的组在每个玉米染色体中的一系列相邻的基因组窗口中鉴定多种单元型,例如,用这些窗口提供基本上完全的基因组覆盖。使用足够大的和多样性的玉米育种群体,可以在每个窗口中鉴定大量的单元型,从而提供可与一种或多种性状相关的等位基因 DNA,以允许聚焦的标记辅助的育种。因此,本发明的玉米分析的一方面进一步包括以下步骤:对所述玉米植物群体表征一种或多种性状,并将所述性状与所述等位基因 SNP 或 Indel 多态性进行关联,优选地进行组织以定义单元型。这样的性状包括产量、倒伏、成熟、株高、抗病性,例如对色二孢属、灰斑病、赤霉菌、炭疽病、镰刀霉以及其它穗和茎腐病、枯萎病、锈病、细菌性病害、昆虫病害等的抗性,非生物应激耐受性,例如,耐旱性、耐寒性、耐热性、耐风暴性、营养缺乏等,和质量性状,例如,提高的淀粉含量、提高的油含量、减少的饱和脂肪酸含量、提高的蛋白质含量、增加的赖氨酸含量等。

[0187] 实施例

[0188] 本文包括下面的实施例来证明本发明的优选的实施方案。本领域的技术人员应该理解,实施例中公开的技术代表了本发明人发现的在实施本发明中运用良好的技术,因此可被认为构成了实施本发明的优选方式。然而,本领域的技术人员根据本公开内容应该理解,在不背离本发明的概念、精神和范围的情况下,可以对公开的具体实施方案进行许多改变,而仍然获得相同或类似的结果。更具体地说,显然某些化学和生理学相关的试剂可以代替本文中所述的试剂,而将获得相同或相似的结果。所有这些对于本领域技术人员明显的类似的代替和修改被认为是在所附的权利要求书限定的本发明的精神、范围和概念之内。

[0189] 实施例 1

[0190] 本实施例说明为了富集独特 / 编码序列基因组 DNA,使用对甲基化胞嘧啶残基敏感的酶制备简化形式的文库。

[0191] 基因组 DNA 提取方法是本领域众所周知的。一种使产量和方便性都达到最佳的优选的方法是用来自 Life Technologies(Grand Island, NY) 的“植物 DNAzol 试剂”提取 DNA。简单地说,用研钵和研杵在液氮中研磨冷冻的叶组织。然后用 DNAzol 试剂提取磨碎的组织。这除去了细胞蛋白质、细胞壁物质和其它碎片。用该试剂提取后, DNA 经沉淀,洗涤,再悬浮,并用 RNA 酶处理以除去 RNA。再次沉淀 DNA,并再悬浮于适当体积的 TE 中(使浓度为 $1 \mu\text{g}/\mu\text{l}$)。该基因组 DNA 备用于文库构建。

[0192] 分别用 Pst I 限制性内切核酸酶消化来自为了检测多态性而比较的两个玉米系的基因组 DNA, Pst I 限制性内切核酸酶为 DNA 片段的末端提供粘性末端,该粘性末端可以连接到具有相同限制性位点的质粒中。举例来说,将 100 单位 Pst I 添加到 20 微克的 DNA 中,并在 37°C 温育 8 小时。在 1% 的低熔点琼脂糖凝胶上通过电泳分离消化的 DNA 产物,以按大小分离 DNA 片段。将来自两个玉米系的消化的 DNA 并排加样到凝胶上(之间用一个电泳泳道作为间隔)。将 1-KB DNA 阶梯分子量标准和 100-bp DNA 阶梯分子量标准加样到两个玉米 DNA 泳道的各一侧。这些分子量标准作为消化的玉米 DNA 的大小分级分离的指导。依照 500-600bp、600-700bp、700-800bp、800-900bp、900-1100bp、1100-1500bp、1500-2000bp、

2000–2500bp 和 2500–3000bp 的大小部分,从凝胶上逐步切下 500–3000bp 范围的片段。使用 β -琼脂糖酶纯化每个部分中的 DNA,并连接到 pUC18 的 Pst I 克隆位点中。质粒连接产物通过电穿孔转化到 DH10B 大肠杆菌细菌宿主中,以产生简化形式的文库。例如,大约 500ng 大小选择的 DNA 与 50ng 去磷酸化的 pUC18 载体连接。

[0193] 通过电穿孔进行转化,简化形式的 Pst I 文库的转化效率是 1 微升连接产物大约 50,000–300,000 个转化体,或 1000–6000 个转化体/ngDNA。

[0194] 评价质量的基本试验包括平均插入大小、叶绿体/线粒体 DNA 含量以及重复序列的分数。

[0195] 在文库构建过程中评估该文库的平均插入大小。通过每个连接测定 10–20 个克隆检测每个连接,以确定平均插入大小。使用标准的微量制备方案从重组克隆中分离 DNA,用 Pst I 消化,以使插入片段从载体上释放出来,然后使用 1% 的琼脂糖凝胶电泳进行大小分离 (Maule, Molecular Biotechnology 9:107–126(1998), 本文引入其全文作为参考)。

[0196] 通过对克隆 (400) 小样本进行测序,并相对各种序列数据库交互核对获得的序列,评估叶绿体/线粒体 DNA 含量和重复序列在文库中的百分比。一些重复元件不存于数据库中,但通常可以通过相同序列的大量拷贝来鉴定。例如,在对一组 400 个克隆进行测序之后,没有通过重复元件数据库过滤但在样品中以超过 10 倍存在的任何序列被认为是重复元件。

[0197] 通过插入从以下玉米系获得的富含编码区的 DNA 构建本发明的玉米简化形式文库:01INL1、7051、5750、17INI20、LH185、7797、WDHQ11、LH172、5CM1、LH82、B73 及 M017。

[0198] 实施例 2

[0199] 本实施例说明从实施例 1 制备的简化形式文库中的克隆确定玉米基因组 DNA 序列。两种基本方法可以用于 DNA 测序:Sanger 等人, Proc. Natl. Acad. Sci. USA 74:5463–5467(1977) 的链终止法,以及 Maxam 和 Gilbert, Proc. Natl. Acad. Sci. USA 74:560–564(1977) 的化学降解法。技术的自动化和进步,例如用基于荧光的测序代替放射性同位素,减少了 DNA 测序所需的工作 (Craxton, Methods, 2:20–26(1991), Ju 等人, Proc. Natl. Acad. Sci. USA 92:4347–4351(1995) 及 Tabor 和 Richardson, Proc. Natl. Acad. Sci. USA 92:6339–6343(1995))。自动化测序仪可以获自,例如,Applied Biosystems, Foster City, California (ABI Prism® systems); Pharmacia Biotech, Inc., Piscataway, New Jersey (Pharmacia ALF), LI-COR, Inc., Lincoln, Nebraska (LI-COR 4,000) 和 Millipore, Bedford, Massachusetts (Millipore BaseStation)。

[0200] 通过可从 CodonCode Corporation, Dedham, MA 获得的 PHRED 分配来自跟踪文件 (trace files) 的序列碱基识别 (base calling) 和质量得分, Brent Ewing, 等人 “Base-calling of automated sequencer traces using phred”, 1998, Genome Research, 第 8 卷, 第 175–185 和 186–194 页描述了该 PHRED, 本文引入该文献作为参考。

[0201] 碱基识别完成后,通过切去质量较差的末端序列提高序列质量。如果产生的序列小于 50bp,则将它删除。删除整体质量小于 12.5 的序列。而且,除去污染序列,例如大肠杆菌 BAC 和载体序列和亚克隆载体。使用可从 DoubleTwist Inc., Oakland, CA 获得的 PangeaClustering and Alignment Tools,通过比较序列对的重叠碱基,装配叠连群。使用下列高严格性参数确定重叠:字长=8;窗口大小=60;同一性为 93%。使用可从

CodonCode Corporation 获得的 PHRAP 片段装配程序,使用 0.5 或更低的“重复严格性”参数,重新装配群簇(clusters)。最终的装配输出包含序列的集合,其中包括代表重叠聚类序列(叠连群)的共有序列的叠连群序列,和不存在于相关序列(单拷贝序列)的任何群簇中的单拷贝序列(singleton)。总起来说,由 DNA 装配产生的叠连群和单拷贝序列被称为岛(islands)。

[0202] 实施例 3

[0203] 本实施例说明通过比较来自如实施例 2 制备的至少 2 个单独玉米系的叠连群和单拷贝序列的序列对比,来鉴定 SNP 和 Indel 多态性。将来自多个玉米系的序列装配成具有一种或多种多态性、即 SNP 和 / 或 Indel 的基因座。合格的候选多态性具有以下参数:

[0204] 用于共有比对的叠连群或单拷贝序列的最小长度是 200 个碱基。

[0205] 在候选 SNP 每一侧上的 15 个碱基的区域中,观测到的碱基的同一性百分比是 75%。

[0206] 在多态性位点处的每个叠连群中的最小 BLAST 质量是 35。

[0207] 在多态性位点每一侧的 15 个碱基的区域中,最小 BLAST 质量是 20。

[0208] 多个具有合格的多态性的基因座被确定为具有如 SEQ ID NO :1 至 SEQ ID NO :6552 报告的共有序列。表 1 中确定了每个基因座中的合格的 SNP 和 Indel 多态性。更具体地,表 1 如下确定了多态性的类型和位置:

[0209] SEQ_NUM 指多态性玉米 DNA 基因座的 SEQ ID NO. (序列 ID 号)。

[0210] CONSEQ_ID 指多态性玉米 DNA 基因座的任意确定的名称。

[0211] MUTATION_ID 指每种多态性的任意确定的名称。

[0212] START_POS 指在多态性玉米 DNA 基因座的核苷酸序列中多态性开始的位置。

[0213] END_POS 指在多态性玉米 DNA 基因座的核苷酸序列中多态性结束的位置;对于 SNP, START_POS 和 END_POS 是共同的。

[0214] TYPE 指确定多态性为 SNP 或 IND(Indel)。

[0215] ALLELE 和 STRAIN 指特定等位基因玉米品种中的多态性的核苷酸序列。

[0216] 实施例 4

[0217] 本实施例说明利用引物碱基延伸检测 SNP 多态性。

[0218] 使用正向和反向 PCR 引物扩增少量的玉米基因组 DNA(如大约 10ng),该引物设计为具有 55℃ 的与模板的退火温度,即,在特定分子标记的多态性的周围。将 PCR 产物加到新板中,其中延伸引物与 GBA 板中的反应孔表面共价结合。加入含有 DNA 聚合酶、2 种差异标记的 ddNTP 和延伸缓冲液的延伸混合物。GBA 板在 42℃ 下温育 15 分钟,以允许延伸。通过用合适的缓冲液洗涤,从孔中除去反应混合物。对于每种标记,通过相继与第一和第二检测试剂温育来检测这两种标记。通过以下步骤测定特定 ddNTP-FITC 的掺入:与 HRP-抗-FITC 温育,接着洗涤孔,接着在含有 HRP 的发色底物的缓冲液中温育。在适于 HRP 反应产物的波长处,用分光光度法测量每个孔的反应程度。再次洗涤该孔,用 AP-链霉抗生物素蛋白重复该程序,接着在含有 AP 的发色底物的缓冲液中温育,并在适于 AP 反应产物的波长处进行分光光度法测量。

[0219] 结果分析

[0220] 从对检测步骤特异性标记的反应产物测量的吸光度推断每种标记 ddNTP 的掺入

程度,并从这些吸光度相对于已知基因型的标准和无模板对照反应的比例推断样品的基因型。在最常见的做法中,观察到的每个数据点的吸光度相对于彼此在散点图中绘图,产生“等位图”。利用该实施例的单碱基延伸试验的一个成功的基因分型试验提供了如图 2 所示的等位图,其中数据点分为四个群簇:纯合子 1(例如,A 等位基因)、纯合子 2(例如,G 等位基因)、杂合子(每个样品含有两个等位基因)和由无模板对照或失败的扩增或检测产生的“无信号”簇。

[0221] 实施例 5

[0222] 本实施例说明利用标记探针降解试验检测 SNP 多态性。在 5 μ l 的总体积中,一定量的玉米基因组模板 DNA(如大约 2-20ng)与如表 2 所述的 4 种寡核苷酸(即正向引物、反向引物、具有附着在 5' 端的 VIC 报道分子的杂交探针和具有附着在 5' 端的 FAM 报道分子的杂交探针)以及含有被动参比染料 ROX 的 PCR 反应缓冲液相混合。使用 60°C 的退火-延伸温度,进行 PCR 反应 35 个循环。反应后,使用荧光计确定每个荧光团的荧光,以及被动参比的荧光。每个荧光团的荧光值相对于被动参比的荧光值标准化。每一个样品的标准化的值相对于彼此绘图,以产生等位图。使用本实施例的引物和杂交探针的一个成功的基因分型试验提供了如图 2 所示的在可清楚分开的群簇中具有数据点的等位图。

[0223] 表 2. 使用 SNP 多态性的标记探针降解检测的分子标记试验的例子。每个试验提供两个用于扩增横跨多态性的区域的寡核苷酸引物,和两个附着有用于 SNP 等位基因检测的荧光报告分子的寡核苷酸探针。有用的报告染料包括但不限于 6-羧基-4,7,2',7'-四氯荧光素(TET)、2'-氯-7'-苯基-1,4-二氯-6-羧基荧光素(VIC)和 6-羧基荧光素亚磷酰胺(FAM)。一种有用的猝灭剂是 6-羧基-N,N,N',N'-四甲基罗丹明(TAMRA)。

[0224]

标 记 SEQ ID	CONSEQ_ ID	引物 SEQ ID NO:	序列类型	序列	等 位 基 因
381	20030484- CON.1	SEQ ID NO:6553	正向引物	CGTCGCAAGAGTCAAGGATATTACA	
381	20030484- CON.1	SEQ ID NO:6554	探针 1	AAAAGGTGTCAATATCGTCT	A
381	20030484- CON.1	SEQ ID NO:6555	探针 2	AAGGTGTCAATGTCGTCT	G
381	20030484- CON.1	SEQ ID NO:6556	反向引物	CGAGCAGTGGACGACAACA	
346	20030439- CON.1	SEQ ID NO:6557	正向引物	CGGAGTCGTCGTTGGCAAATA	
346	20030439- CON.1	SEQ ID NO:6558	探针 1	AAATCATCAGATCGTTTTT	A
346	20030439- CON.1	SEQ ID NO:6559	探针 2	CATCAGCTCGTTTTT	C
346	20030439- CON.1	SEQ ID NO:6560	反向引物	ACTAGGCGTCCAGTTCAGATTTTG	
708	20031563- CON.1	SEQ ID NO:6561	正向引物	TGGCTTTATTTTTTCGATACTGAAGAT CCT	
708	20031563- CON.1	SEQ ID NO:6562	探针 1	CAATGTTCTTCAATCCATAC	A
708	20031563- CON.1	SEQ ID NO:6563	探针 2	CAATGTTCTTCAGTCCATAC	G
708	20031563- CON.1	SEQ ID NO:6564	反向引物	GGGTGATGAGAGGAATTAGAGCAA A	
134	20030159- CON.1	SEQ ID NO:6565	正向引物	CTTCATGGCAACATTAGCTGTGTTT	
134	20030159- CON.1	SEQ ID NO:6566	探针 1	CAAGATGTTTATGAAAGCG	A
134	20030159- CON.1	SEQ ID NO:6567	探针 2	CAAGATGTTTATGAGAGCG	G
134	20030159- CON.1	SEQ ID NO:6568	反向引物	GCCTGCCCATTAGGCTGAAA	

[0225] 为证实试验产生精确的结果,对代表三种可能的基因型(即两种纯合子等位基因和一种杂合子样品)中的每一种的已知基因型身份的多个重复样品进行每个新的试验。为了成为有效的和有用的试验,它必须产生可清楚分开的数据点簇,使得可以为至少 90%的数据点分配三种基因型中的一种,并且观察到这种分配对于至少 98%的数据点是正确的。在此验证步骤后,对两个高度近交的个体之间的杂交后代进行试验,以获得分离数据,然后利用该分离数据计算多态性基因座的基因图谱位置。

[0226] 实施例 6

[0227] 本实施例说明,对于玉米系 B73 和 Mo17 杂交产生的 274 个相互杂交(intermated)重组近交系(IRI),基于超过 6000 个 SNP 的基因型,本发明基因座中的分子标记的遗传作图。这些基因型与对 IRI 评分的 1320 个公共核心 SSR 和 RFLP 标记的基因型组合。作图之前,除去任何显示扭曲的分离的基因型(对于分离比为 1 : 1 的卡方检验, $P < 1e-5$)。低 α 水平用来说明多重检验的问题。

[0228] 一方面,可以使用由 Stam, P. “Construction of integrated genetic linkage maps by means of a new computer package :JoinMap, The Plant Journal, 3 :

739-744(1993); Stam, P. 和 van Ooijen, J. W. “JoinMapversion 2.0: Software for the calculation of genetic linkage maps(1995) CPRO-DLO, Wageningen. 描述的 JoinMap 2.0 版软件构建图谱。JoinMap 实现对于多点作图的加权最小二乘方法, 图中加入来自所有连锁基因座对(相邻或不相邻)的信息。使用 5.0 的 LOD 阈值形成连锁群。利用 SSR 和 RFLP 公共标记将连锁群分配到染色体上。在构建图谱之前, 连锁群合并入染色体内部。

[0229] 其它高密度标记作图方法是本领域内已知的; 关于 IRI 在较高分辨率作图中的应用, 参见, 例如, Winkler 等人 (Genetics 164:741-745(2003))。另外参见, Jansen 等人 (Theor Appl Genet 102:1113-1122(2001))。在许多条件下, Jansen 等人的方法导致非常近似于最大似然图。此外, 由这种方法评估的图谱与使用 JoinMap 2.0 获得的图谱非常一致。此外, 上述的和本文作为参考引入的方法的组合可以用于在一定范围的群体结构以及计算限制下最有效地调节 (leverage) 标记数据。

[0230] 本发明的另一方面, 利用 Kosambi 作图函数将重组部分转化为图谱距离。定位的 SNP 分子标记在表 3 中确定, 其中, “染色体”和“位置”确定对于根据 “Conseq_ID” 确定的 SNP, 以 cM 为单位测量的距玉米染色体 5' 端的距离。“公开名称”(“Public Name”) 提供了不属于本发明的部分的参比公开标记的公布名称。对于表 3 中所列的某些定位的多态性标记, 多次列出突变 ID, 其表明基于多重基因分型试验进行作图。多重基因分型试验的图谱位置一般用于证实图谱位置, 除了在图谱位置分歧的情况下以外, 例如, 由于试验设计或实践的误差。作图的分子标记的密度和分布如图 1 所示。

[0231] 实施例 7

[0232] 本实施例说明使用表 1 中和 SEQ ID NO:1-6552 的 DNA 序列中公开的分子标记的本发明的方法。

[0233] 使用基于 SEQ ID NO:1-6552 的序列, 如实施例 5 中所述制备的用于表 1 中确定的每种分子标记的引物对和探针对, 分析具有不同的遗传的玉米育种群体。密切连锁的分子标记被确定为玉米基因组中大约 10 厘摩的相邻基因组窗口之内的特征性单元型。代表群体的至少 4% 的单元型与对于玉米群体的每个成员确定的性状值相关, 包括产量、成熟度、倒伏、株高、锈病抗性、耐旱性和冷发芽的性状值。每个单元型的性状值在各为 10 厘摩的窗口内排序。分析来自群体的随机交配成员的后代种子在每个窗口中的单元型的身份。基于在所述种子中确定的单元型的高性状值, 选择后代种子进行种植。

[0234] 实施例 8

[0235] 本实施例说明可用于获得用于育种的、具有优选性状的亲本植物、后代植物或测试植物的多态性的鉴定。在这个具体的实施例中, 为了说明性的目的, 已选择多态性用于鉴定具有优选产量性状的植物。但是, 也预计可以以类似的方式鉴定可用于鉴定其它优选性状的其它标记(即, 通过指出多态性的遗传图谱位置在单元型窗口内的定位)。进一步预期, 本实施例中披露的具体标记除了作为产量性状的标记以外也可以发现其它用途。

[0236] 首先, 如美国专利申请 60/837864 所公开的那样确定与产量相关的单元型窗口。利用表 3 中公开的图谱位置确定本发明的标记, 该标记位于包括优选的产量单元型的单元型窗口中, 并且可以用作这些区域的标记。选择与 13 个单元型窗口一致的 25 种多态性, 这些窗口包括在孟山都 (Monsanto) 的玉米种质中产量排名前 13 位的单元型。因而对于这些产量单元型窗口的大多数提供两 (2) 种标记。可用于鉴定具有优选产量性状的用于育种

的植物的具体标记可以选自 SEQ ID NO :5407、287、574、3407、5367、4566、2457、5295、4548、5182、5489、2714、2726、375、275、1415、885、2067、4773、1708、1479、3507、2765 和 1279 及 SEQ ID NO :2468。

[0237] 从以上描述中可以看出,可以实现并获得本发明的几个优势。

[0238] 选择并描述实施方案,以便最好地解释本发明的原理及其实际应用,从而使本领域的其它技术人员能够在各种实施方案中最好地利用本发明,并且进行各种修改以适于预期的特定应用。

[0239] 本文引用了多篇专利和非专利出版物,它们的公开内容均完整引入本文作为参考。

[0240] 由于在不背离本发明的范围的情况下,可以在本文描述的和说明的结构和方法中进行各种修改,上述说明书包含的或附图显示的所有内容应当解释为示例性的,而非限制性的。本发明的广度和范围不应由上述任何示例性的实施方案限制,而应该只根据下面的权利要求书及其等同物加以限定。

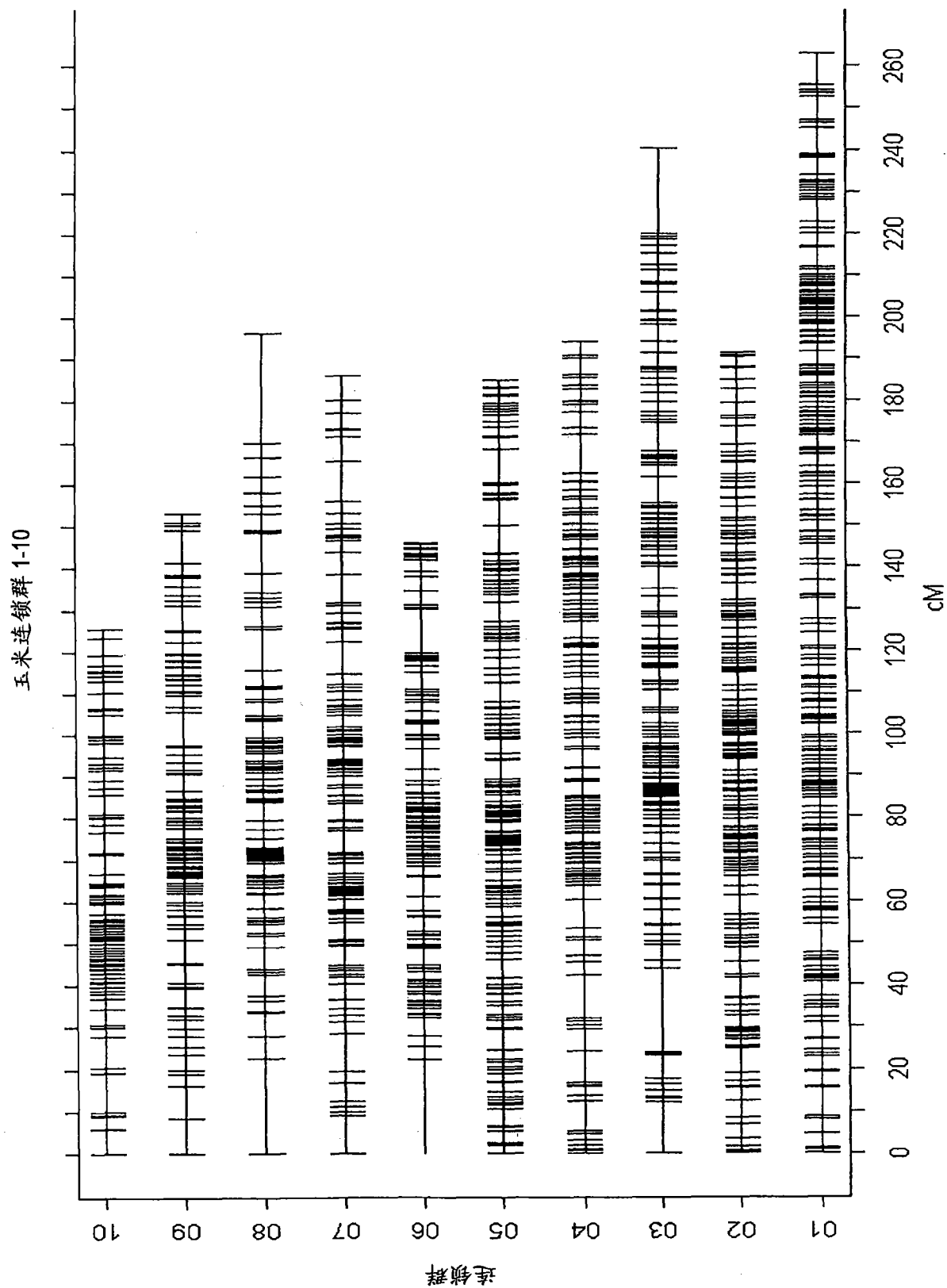


图 1

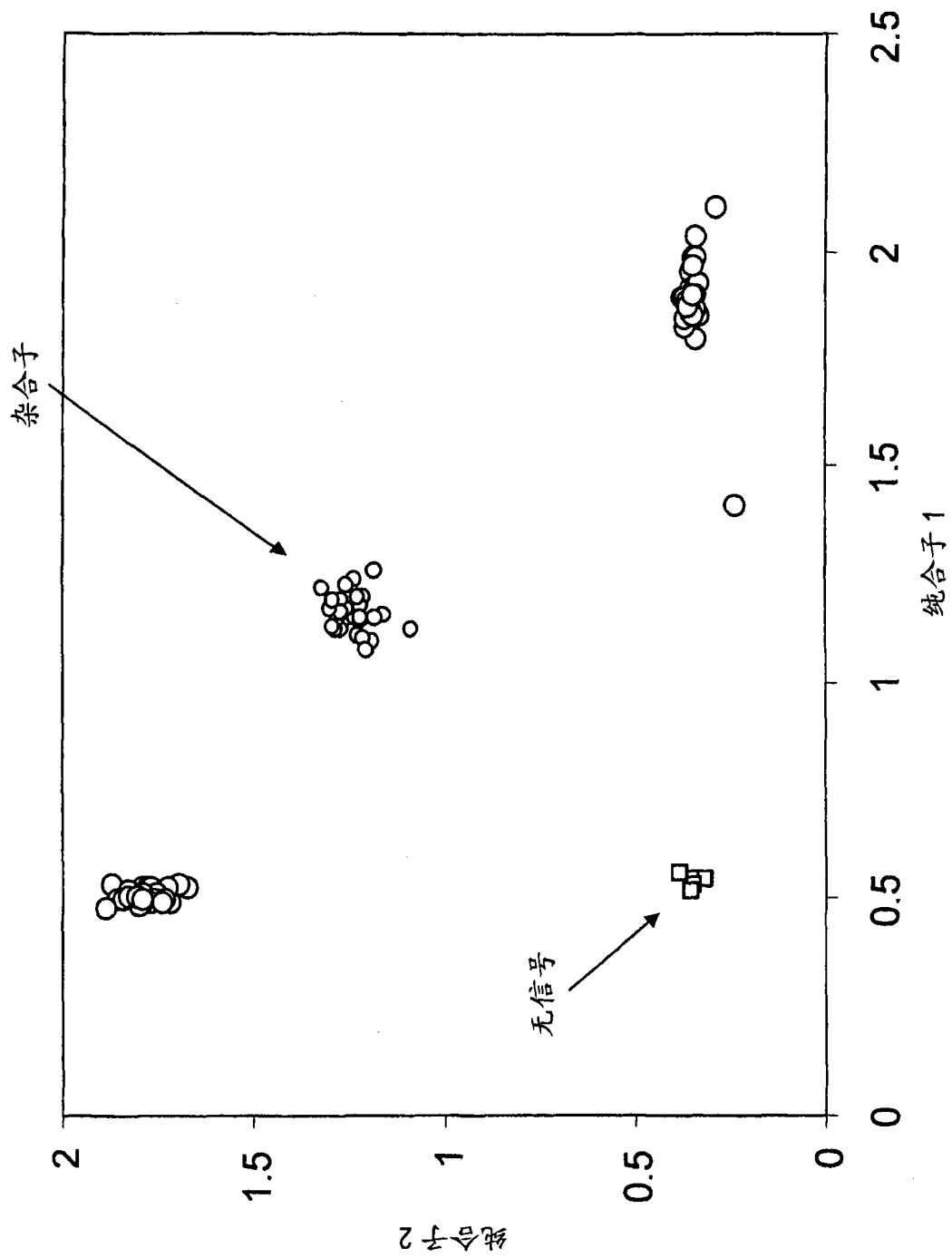


图 2