



(12) 发明专利申请

(10) 申请公布号 CN 112365926 A

(43) 申请公布日 2021.02.12

(21) 申请号 202011244144.3

(22) 申请日 2020.11.10

(71) 申请人 沈阳农业大学

地址 110866 辽宁省沈阳市沈河区东陵路
120号

(72) 发明人 张敖 张学才 阮燕晔 崔震海
李丛 曹慧颖 朱延姝 陈珊
苏玉静 张立军

(74) 专利代理机构 大连智高专利事务所(特殊
普通合伙) 21235

代理人 张钦

(51) Int. Cl.

G16B 30/00 (2019.01)

G16B 25/00 (2019.01)

权利要求书1页 说明书4页 附图1页

(54) 发明名称

一种玉米自交系一般配合力和特殊配合力的全基因组预测方法及应用

(57) 摘要

本发明属于玉米分子辅助育种领域,公开了一种玉米自交系一般配合力和特殊配合力的全基因组预测方法及应用。依据被测玉米自交系基因型信息预测一般配合力和基于自交系与测验种杂种F1基因型预测特殊配合力的方法。降低育种成本。

1. 一种玉米自交系一般配合力的全基因组预测方法,其特征在于,包括如下步骤:

- (1) 获得全部被测定自交系的基因型;
- (2) 将基因型数据数值化;
- (3) 将被测定自交系分为建模自交系和待预测自交系;
- (4) 将建模自交系与测验种杂交,获得杂种F1;
- (5) 田间种植F1获得表型数据;
- (6) 计算建模自交系的一般配合力;

(7) 利用建模自交系的数值化基因型数据和一般配合力数据,建立一般配合力的全基因组预测模型;

(8) 将待预测自交系的数值化基因型数据代入一般配合力全基因组预测模型,预测一般配合力。

2. 一种玉米自交系特殊配合力的全基因组预测方法,其特征在于,包括如下步骤:

- (1) 获得全部被测定自交系和测验种的基因型;
- (2) 将基因型数据数值化;
- (3) 利用自交系和测验种基因型数据虚拟合成杂种F1的基因型;
- (4) 将被测自交系分为建模自交系和待预测自交系;
- (5) 将建模自交系与测验种杂交,获得杂种F1;
- (6) 田间种植F1获得表型值;
- (7) 计算建模自交系的特殊配合力;

(8) 利用建模自交系与测验种的F1虚拟合成基因型信息,与特殊配合力数据,建立特殊配合力的全基因组预测模型;

(9) 将待预测自交系与测验种杂交的杂种F1虚拟合成基因型数据代入特殊配合力全基因组预测模型,预测特殊配合力。

3. 如权利要求1或2所述的方法,其特征在于,所述的测验种包括玉米自交系和杂交种。

4. 如权利要求1或2所述的方法,其特征在于,所述的建模自交系占总自交系的10~80%。

一种玉米自交系一般配合力和特殊配合力的全基因组预测方法及应用

技术领域

[0001] 本发明涉及玉米分子辅助育种领域,具体涉及一种玉米自交系一般配合力和特殊配合力的全基因组预测方法及应用。

背景技术

[0002] 玉米杂交种 (hybrid) 是由基因型不同父、母亲本自交系之间进行杂交产生的第一代种子 (通常称为杂种F1代)。杂交种在生长势、抗逆性和产量方面优于父、母亲本,即,产生杂种优势。然而,不是任何具有遗传差异 (基因型不同) 的自交系之间杂交都能产生适合农业产生需要的优良杂交种。产生优良杂交种的两个自交系之间不仅要有遗传差异,而且还要优良性状互补。所以,玉米杂交育种者的重要工作之一是培育能够产生杂种优势的优良自交系。

[0003] 获得优良自交系是玉米杂交育种成败的基础。玉米自交系由不同遗传背景 (基因型不同的) 的玉米材料进行杂交,在经过多代自交后产生。在选育时,通常用一般配合力 (GCA) 和特殊配合力 (SCA) 衡量自交系的优劣。一般配合力是自交系与一系列测验种 (自交系或杂交种) 杂交产生具有强杂种优势F1的能力。特殊配合力是自交系与特定测验种 (自交系) 杂交产生具有高的杂种优势F1的能力。高的一般配合力是高特殊配合力的基础。在传统杂交育种中,通过将被测自交系与一系列测验种杂交,根据其杂种F1的田间表现确定一般配合力和特殊配合力,两者都需要大量的杂交和测定工作。

[0004] 随着测序技术的发展,基因测序成本大幅度降低,分子辅助育种技术已经发展成为可以实用的技术,其中的全基因组预测技术,是根据待预测材料的基因型,利用已经建立好的基因型与实测表现型关系的数学模型,来预测表现型的技术。全基因组预测技术在理论上可以用来预测玉米自交系的一般配合力和特殊配合力。这种方法将减少大量的田间种植和性状测定工作,可以节省大量的人力、物力和时间,提高育种效率。目前尚未见到利用全基因组预测技术预测玉米自交系一般配合力和特殊配合力的方法。现在有一般配合力的预测方法,一般是先通过预测表型,再通过预测的表型预测一般配合力。特殊配合力同理。

[0005] 一般配合力是对自交系适配性的评价,当测验种一定时,不同自交系的一般配合力表现决定于自交系的基因型,因此,可以直接用自交系的基因型信息来进行预测。而特殊配合力是个别自交系之间的匹配能力,并不单一决定于被测自交系的基因型,而是决定于被测自交系与测验种基因型的组合。因此,利用全基因组预测技术预测自交系的特殊配合力时,需要利用杂交组合的基因型与表现型建立预测模型。然而,在预测待预测自交系的特殊配合力时,由于没有做真实的杂交,也就没有真实的杂种一代F1种子用来基因测序,因此,必须利用虚拟合成的杂种基因型。

发明内容

[0006] 为提高玉米自交系一般配合力和特殊配合力的测定效率,降低育种成本,本发明

提供一种利用全基因组技术,依据被测玉米自交系基因型信息预测一般配合力和基于自交系与测验种杂种F1基因型预测特殊配合力的方法。将全部被测自交系和测验种(自交系或杂交种)进行测序获得基因型,取部分被测自交系(建模自交系)与所有测验种(自交系或杂交种)杂交,将得到的F1在田间种植和测定GCA和SCA,然后:(1)根据这部分自交系基因型和GCA数据建立GCA的全基因组预测模型,再用待预测自交系的基因型,预测GCA;(2)利用建模自交系与测验种杂交F1的虚拟合成基因型和SCA数据建立SCA的全基因组预测模型,再利用待预测自交系与测验种杂交F1的虚拟合成基因型,预测SCA。

[0007] 本发明的技术方案如下:

[0008] 本发明第一个目的是想要保护一种玉米自交系一般配合力的全基因组预测方法,包括如下步骤(图1):

[0009] (1)对HMP格式的基因型数据进行质量控制;

[0010] (2)计算自交系和测验种的一般配合力(GCA);

[0011] (3)利用建模自交系的基因型数据和一般配合力数据,建立一般配合力的全基因组预测模型,将待预测自交系的数值化基因型数据代入一般配合力全基因组预测模型,预测一般配合力。

[0012] 上述一种玉米自交系一般配合力的全基因组预测方法,具体步骤为:

[0013] (1)获得全部被测定自交系的基因型;

[0014] (2)将基因型数据数值化;

[0015] (3)将被测定自交系分为建模自交系和待预测自交系;

[0016] (4)将建模自交系与测验种杂交,获得杂种F1;

[0017] (5)田间种植F1获得表型数据;

[0018] (6)计算建模自交系的一般配合力;

[0019] (7)利用建模自交系的数值化基因型数据和一般配合力数据,建立一般配合力的全基因组预测模型;

[0020] (8)将待预测自交系的数值化基因型数据代入一般配合力全基因组预测模型,预测一般配合力。

[0021] 本发明第二个目的是想要保护一种玉米自交系特殊配合力的全基因组预测方法,包括如下步骤(图1):

[0022] (1)对HMP格式的基因型数据进行质量控制;

[0023] (2)利用自交系和测验种基因型数据虚拟合成杂种F1的基因型;

[0024] (3)计算建模自交系的特殊配合力;

[0025] (4)利用建模自交系与测验种的F1虚拟合成基因型信息,与特殊配合力数据,建立特殊配合力的全基因组预测模型;将待预测自交系与测验种杂交的杂种F1虚拟合成基因型数据代入特殊配合力全基因组预测模型,预测特殊配合力。

[0026] 上述一种玉米自交系特殊配合力的全基因组预测方法,具体步骤为:

[0027] (1)获得全部被测定自交系和测验种的基因型;

[0028] (2)将基因型数据数值化;

[0029] (3)利用自交系和测验种基因型数据虚拟合成杂种F1的基因型;

[0030] (4)将被测自交系分为建模自交系和待预测自交系;

- [0031] (5) 将建模自交系与测验种杂交,获得杂种F1;
- [0032] (6) 田间种植F1获得表型值;
- [0033] (7) 计算建模自交系的特殊配合力;
- [0034] (8) 利用建模自交系与测验种的F1虚拟合成基因型信息,与特殊配合力数据,建立特殊配合力的全基因组预测模型;
- [0035] (9) 将待预测自交系与测验种杂交的杂种F1虚拟合成基因型数据代入特殊配合力全基因组预测模型,预测特殊配合力。
- [0036] 一种玉米自交系一般配合力的全基因组预测方法和一种玉米自交系特殊配合力的全基因组预测方法中所述的测验种包括玉米自交系和杂交种。
- [0037] 一种玉米自交系一般配合力的全基因组预测方法和一种玉米自交系特殊配合力的全基因组预测方法中所述的建模自交系占总自交系的10~80%。
- [0038] 育种中通过一般配合力选择优良的自交系,通过特殊配合力选择杂交种。自交系的一般配合力可以使用Line by tester组配设计,即一个群体和另一个杂种优势群的少量有代表性的自交系进行杂交,然后计算群体中每个个体的平均表现。利用该方法,可以只对群体中的一部分进行表型鉴定,其余的部分通过全基因组预测获得结果。可以在节约大量土地和人工的情况下找到可能性最高的自交系用于接下来的育种循环。利用一个小样本群体的GCA预测结果见图2。预测值与计算值之间的Person相关系数为0.65,最高和最低值的区域重合性较高,可以应用于生产实践。
- [0039] 本发明的有益效果如下:
- [0040] 现有技术需要利用双列杂交组配设计将不同杂种优势群的自交系分别一一组配成杂交种,再在多个地点鉴定表型数据,才能计算一般配合力(GCA)和特殊配合力(SCA),通常情况下群体规模较大,没有足够的土地去进行这样的测试,需要每年测试一部分,但年份之间的差异无法去除。通过本发明提供的方法,只取部分有代表性的材料进行多地点的田间鉴定,可以大大减少土地、人工和成本,同时消除不同环境对杂交种的影响。

附图说明

- [0041] 图1是依据被测自交系与测验种杂交产生的F1基因型信息预测SCA的技术路线图,其中GCA为一般配合力;SCA为特殊配合力。
- [0042] 图2是GCA的预测表现图。

具体实施方式

- [0043] 实施例1
- [0044] 依据被测自交系全基因组精确序列信息利用生物信息学技术预测GCA,包括以下步骤:
- [0045] (1) 用TASSEL软件对基因型文件进行质量控制。
- [0046] a) 打开Hapmap格式的基因型文件,选择File→Open,在打开窗口中选择HMP格式的基因型文件,点打开。本例有53个自交系和36959个碱基座。Filter→Filter Genotype Table Sites,根据杂合率、最小等位基因频率和缺失率,去除不符合实际情况的材料和SNP标记,去掉缺失率高的分子标记,本例缺失率≤20%,杂合率≤15%,最小等位基因频率≥

0.05.筛选过后,剩下4566个碱基座,保存筛选后的文件。

[0047] (2) 使用AGD-R计算GCA,本例采用Line by Tester组配设计和 α 拉丁方实验设计。

[0048] a) 打开AGD-R,选择Line by Tester。

[0049] b) 选择表型文件并打开。

[0050] c) 用REML方法计算GCA和SCA。结果在“Output_AGD-R\GCA”目录下。

[0051] d) 打开“Output_AGD-R\GCA”目录下的“GY_12_2017Balanced REML LxT Multi-Env Lattice GY.xlsx”文件,将“GCA LINE ACROSS SITES”工作表中的LINE列和GCA列复制到表型文件的LINE和GCA中,再将“GCA TESTER ACROSS SITES”工作表中的TESTER和GCA列复制到表型文件中的LINE和GCA中。

[0052] (3) 利用R语言预测GCA的表现。

[0053] a) 安装R语言和RStudio(略)。

[0054] b) 在Rstudio的Console写“predict_type<-"Line/GCA"”并回车。

[0055] c) 在RStudio中打开“rrBLUP杂交种预测.R”,Ctrl+R全选,Ctrl+Enter运行。依次选择基因型文件和表型文件。

[0056] d) 程序运行完成后,“GenoInfo\GCA”目录下会生成“Accuracy-k-fold-5-GCA.csv”文件,里面记录了每个循环的预测精度和平均预测精度,本例设置循环为10。平均预测精度为58.4143%。由于生产上预测受基因型环境互作的影响,同时本例样本太小,导致预测精度有限。GCA的估计值在“GCA-k-fold-5-Predictive breeding value.csv”文件中。

[0057] e) GCA的预测20次的平均结果见图2。Person相关系数为0.65,最高为0.76。由于本例使用的材料数量较少,只有8个材料和3个测验种,因此,当使用材料足够多时,预测准确性还有上升空间。

[0058] 实施例2

[0059] 依据被测自交系全基因组SNP信息利用生物信息学技术预测SCA,包括以下步骤:

[0060] (1) 该步骤与实施例1中的(1)相同;

[0061] (2) 利用杂交种虚拟合成技术,将自交系的基因型与测验种的基因型合并成可用于全基因组预测的数值化的杂交种基因型。

[0062] (3) 该步骤与实施例1中的(2)基本相同,唯一的不同是实施例1(2)(d)“GCA TESTER ACROSS SITES”工作表中的LINE、TESTER和SCA复制到表型文件中的LINE和GCA中。

[0063] (4) 基本同实施例1中的(3)(b),“predict_type<-"hybrid/SCA"”并回车。预测结果在“GenoInfo\SCA”目录的“SCA-k-fold-5-Predictive breeding value.csv”文件中,由于SCA的预测对材料数量的要求很高,对每个材料的种植环境要求也很高,本例只是演示该方法可以运行,实际预测需要采用双列杂交组配设计才能得到理想的结果。

[0064] 上述实施例只是用于对本发明的举例和说明,而非意在将本发明限制于所描述的实施例范围内。此外本领域技术人员可以理解的是,本发明不局限于上述实施例,根据本发明的教导还可以做出更多种的变型和修改,这些变型和修改均落在本发明所要求保护的范围内。

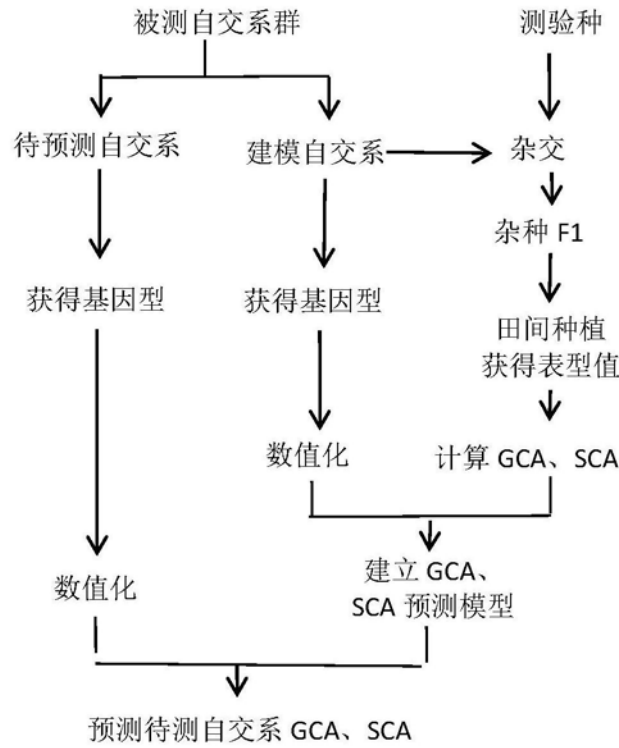


图1

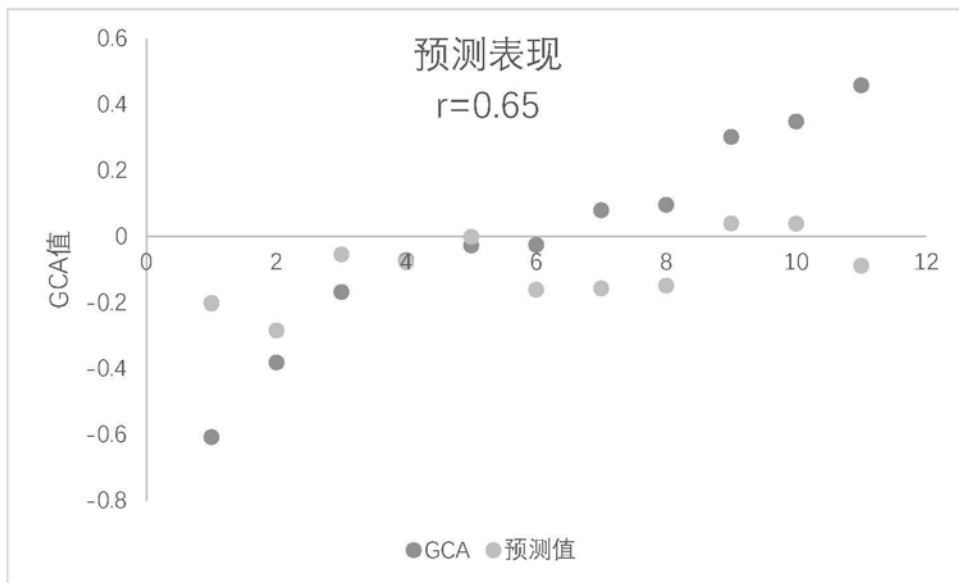


图2