



(12)发明专利

(10)授权公告号 CN 107419017 B

(45)授权公告日 2020.09.08

(21)申请号 201710610369.8

C12N 15/11(2006.01)

(22)申请日 2017.07.25

G16B 20/20(2019.01)

(65)同一申请的已公布的文献号

申请公布号 CN 107419017 A

(56)对比文件

CN 104212886 A,2014.12.17

US 6544730 B1,2003.04.08

(43)申请公布日 2017.12.01

江丽等.27-plex SNP 种族推断方法的优化及验证.《遗传Hereditas (Beijing)》.2017,第39卷(第2期),第166-173页.

(73)专利权人 公安部物证鉴定中心

地址 100038 北京市西城区木樨地南里17号楼

审查员 马琪

(72)发明人 刘京 李彩霞 赵雯婷 江丽

郝伟琪

(74)专利代理机构 北京纪凯知识产权代理有限公司

11245

代理人 关畅 张立娜

(51)Int.Cl.

C12Q 1/6888(2018.01)

权利要求书5页 说明书12页

序列表18页 附图4页

(54)发明名称

对未知来源个体进行五大洲际族群来源推断的方法和系统

(57)摘要

本发明公开了一种用于五个洲际人群区分的SNP体系及其应用。本发明所提供的应用具体为28个SNP位点在如下任一中的应用:(a)构建五大洲际人群基因分型数据库;(b)区分五大洲际人群。本发明所提供的体系不仅能对五大洲际人群进行区分,而且对混合人群也有一定的区分能力,对已知来源样本的祖先主成分和人群匹配与其来源信息一致。并能够推断个体祖先来源成分组成,在实际检案中可以进行推广应用。

1. 28个SNP位点组合在如下任一中的应用:

- (a) 构建五大洲际人群基因分型数据库;
- (b) 区分五大洲际人群;

所述28个SNP位点分别为:rs10483251、rs12142199、rs1229984、rs12402499、rs12498138、rs12594144、rs1426654、rs1557553、rs16891982、rs17822931、rs1871534、rs2080161、rs2139931、rs2789823、rs2814778、rs3751050、rs3827760、rs4657449、rs4749305、rs4792928、rs6054465、rs6437783、rs715605、rs8072587、rs8137373、rs9522149、rs9809818和rs9908046。

2. 用于检测人基因组中28个SNP位点的引物对组;所述28个SNP位点分别为:rs10483251、rs12142199、rs1229984、rs12402499、rs12498138、rs12594144、rs1426654、rs1557553、rs16891982、rs17822931、rs1871534、rs2080161、rs2139931、rs2789823、rs2814778、rs3751050、rs3827760、rs4657449、rs4749305、rs4792928、rs6054465、rs6437783、rs715605、rs8072587、rs8137373、rs9522149、rs9809818和rs9908046;所述引物对组由如下(1)-(28)组成:

(1) 用于检测rs10483251的引物对1,由序列表中序列1和序列2所示的两条单链DNA组成;

(2) 用于检测rs12142199的引物对2,由序列表中序列4和序列5所示的两条单链DNA组成;

(3) 用于检测rs1229984的引物对3,由序列表中序列7和序列8所示的两条单链DNA组成;

(4) 用于检测rs12402499的引物对4,由序列表中序列10和序列11所示的两条单链DNA组成;

(5) 用于检测rs12498138的引物对5,由序列表中序列13和序列14所示的两条单链DNA组成;

(6) 用于检测rs12594144的引物对6,由序列表中序列16和序列17所示的两条单链DNA组成;

(7) 用于检测rs1426654的引物对7,由序列表中序列19和序列20所示的两条单链DNA组成;

(8) 用于检测rs1557553的引物对8,由序列表中序列22和序列23所示的两条单链DNA组成;

(9) 用于检测rs16891982的引物对9,由序列表中序列25和序列26所示的两条单链DNA组成;

(10) 用于检测rs17822931的引物对10,由序列表中序列28和序列29所示的两条单链DNA组成;

(11) 用于检测rs1871534的引物对11,由序列表中序列31和序列32所示的两条单链DNA组成;

(12) 用于检测rs2080161的引物对12,由序列表中序列34和序列35所示的两条单链DNA组成;

(13) 用于检测rs2139931的引物对13,由序列表中序列37和序列38所示的两条单链DNA

组成；

(14) 用于检测rs2789823的引物对14,由序列表中序列40和序列41所示的两条单链DNA组成；

(15) 用于检测rs2814778的引物对15,由序列表中序列43和序列44所示的两条单链DNA组成；

(16) 用于检测rs3751050的引物对16,由序列表中序列46和序列47所示的两条单链DNA组成；

(17) 用于检测rs3827760的引物对17,由序列表中序列49和序列50所示的两条单链DNA组成；

(18) 用于检测rs4657449的引物对18,由序列表中序列52和序列53所示的两条单链DNA组成；

(19) 用于检测rs4749305的引物对19,由序列表中序列55和序列56所示的两条单链DNA组成；

(20) 用于检测rs4792928的引物对20,由序列表中序列58和序列59所示的两条单链DNA组成；

(21) 用于检测rs6054465的引物对21,由序列表中序列61和序列62所示的两条单链DNA组成；

(22) 用于检测rs6437783的引物对22,由序列表中序列64和序列65所示的两条单链DNA组成；

(23) 用于检测rs715605的引物对23,由序列表中序列67和序列68所示的两条单链DNA组成；

(24) 用于检测rs8072587的引物对24,由序列表中序列70和序列71所示的两条单链DNA组成；

(25) 用于检测rs8137373的引物对25,由序列表中序列73和序列74所示的两条单链DNA组成；

(26) 用于检测rs9522149的引物对26,由序列表中序列76和序列77所示的两条单链DNA组成；

(27) 用于检测rs9809818的引物对27,由序列表中序列79和序列80所示的两条单链DNA组成；

(28) 用于检测rs9908046的引物对28,由序列表中序列82和序列83所示的两条单链DNA组成。

3. 根据权利要求2所述的引物对组,其特征在于:所述引物对组中,所述引物对1、所述引物对2、所述引物对3、所述引物对4、所述引物对5、所述引物对6、所述引物对7、所述引物对8、所述引物对9、所述引物对10、所述引物对11、所述引物对12、所述引物对13、所述引物对14、所述引物对15、所述引物对16、所述引物对17、所述引物对18、所述引物对19、所述引物对20、所述引物对21、所述引物对22、所述引物对23、所述引物对24、所述引物对25、所述引物对26、所述引物对27和所述引物对28的摩尔比为0.8:0.6:1.5:2.7:3:0.8:2:1.3:1:1:1.5:1:2.7:0.5:0.8:2.5:0.4:1.6:2.5:4:3:0.8:0.8:3:3:6:0.6:0.6;

每个引物对中两条引物的摩尔比均为1:1。

4. 用于检测人基因组中28个SNP位点的单链DNA组；所述28个SNP位点分别为：rs10483251、rs12142199、rs1229984、rs12402499、rs12498138、rs12594144、rs1426654、rs1557553、rs16891982、rs17822931、rs1871534、rs2080161、rs2139931、rs2789823、rs2814778、rs3751050、rs3827760、rs4657449、rs4749305、rs4792928、rs6054465、rs6437783、rs715605、rs8072587、rs8137373、rs9522149、rs9809818和rs9908046；所述单链DNA组由如下(1)-(28)组成：

(1) 用于检测rs10483251的引物对1和延伸引物1；所述引物对1由序列表中序列1和序列2所示的两条单链DNA组成；所述延伸引物1为序列表中序列3所示单链DNA；

(2) 用于检测rs12142199的引物对2和延伸引物2；所述引物对2由序列表中序列4和序列5所示的两条单链DNA组成；所述延伸引物2为序列表中序列6所示单链DNA；

(3) 用于检测rs1229984的引物对3和延伸引物3；所述引物对3由序列表中序列7和序列8所示的两条单链DNA组成；所述延伸引物3为序列表中序列9所示单链DNA；

(4) 用于检测rs12402499的引物对4和延伸引物4；所述引物对4由序列表中序列10和序列11所示的两条单链DNA组成；所述延伸引物4为序列表中序列12所示单链DNA；

(5) 用于检测rs12498138的引物对5和延伸引物5；所述引物对5由序列表中序列13和序列14所示的两条单链DNA组成；所述延伸引物5为序列表中序列15所示单链DNA；

(6) 用于检测rs12594144的引物对6和延伸引物6；所述引物对6由序列表中序列16和序列17所示的两条单链DNA组成；所述延伸引物6为序列表中序列18所示单链DNA；

(7) 用于检测rs1426654的引物对7和延伸引物7；所述引物对7由序列表中序列19和序列20所示的两条单链DNA组成；所述延伸引物7为序列表中序列21所示单链DNA；

(8) 用于检测rs1557553的引物对8和延伸引物8；所述引物对8由序列表中序列22和序列23所示的两条单链DNA组成；所述延伸引物8为序列表中序列24所示单链DNA；

(9) 用于检测rs16891982的引物对9和延伸引物9；所述引物对9由序列表中序列25和序列26所示的两条单链DNA组成；所述延伸引物9为序列表中序列27所示单链DNA；

(10) 用于检测rs17822931的引物对10和延伸引物10；所述引物对10由序列表中序列28和序列29所示的两条单链DNA组成；所述延伸引物10为序列表中序列30所示单链DNA；

(11) 用于检测rs1871534的引物对11和延伸引物11；所述引物对11由序列表中序列31和序列32所示的两条单链DNA组成；所述延伸引物11为序列表中序列33所示单链DNA；

(12) 用于检测rs2080161的引物对12和延伸引物12；所述引物对12由序列表中序列34和序列35所示的两条单链DNA组成；所述延伸引物12为序列表中序列36所示单链DNA；

(13) 用于检测rs2139931的引物对13和延伸引物13；所述引物对13由序列表中序列37和序列38所示的两条单链DNA组成；所述延伸引物13为序列表中序列39所示单链DNA；

(14) 用于检测rs2789823的引物对14和延伸引物14；所述引物对14由序列表中序列40和序列41所示的两条单链DNA组成；所述延伸引物14为序列表中序列42所示单链DNA；

(15) 用于检测rs2814778的引物对15和延伸引物15；所述引物对15由序列表中序列43和序列44所示的两条单链DNA组成；所述延伸引物15为序列表中序列45所示单链DNA；

(16) 用于检测rs3751050的引物对16和延伸引物16；所述引物对16由序列表中序列46和序列47所示的两条单链DNA组成；所述延伸引物16为序列表中序列48所示单链DNA；

(17) 用于检测rs3827760的引物对17和延伸引物17；所述引物对17由序列表中序列49

和序列50所示的两条单链DNA组成;所述延伸引物17为序列表中序列51所示单链DNA;

(18) 用于检测rs4657449的引物对18和延伸引物18;所述引物对18由序列表中序列52和序列53所示的两条单链DNA组成;所述延伸引物18为序列表中序列54所示单链DNA;

(19) 用于检测rs4749305的引物对19和延伸引物19;所述引物对19由序列表中序列55和序列56所示的两条单链DNA组成;所述延伸引物19为序列表中序列57所示单链DNA;

(20) 用于检测rs4792928的引物对20和延伸引物20;所述引物对20由序列表中序列58和序列59所示的两条单链DNA组成;所述延伸引物20为序列表中序列60所示单链DNA;

(21) 用于检测rs6054465的引物对21和延伸引物21;所述引物对21由序列表中序列61和序列62所示的两条单链DNA组成;所述延伸引物21为序列表中序列63所示单链DNA;

(22) 用于检测rs6437783的引物对22和延伸引物22;所述引物对22由序列表中序列64和序列65所示的两条单链DNA组成;所述延伸引物22为序列表中序列66所示单链DNA;

(23) 用于检测rs715605的引物对23和延伸引物23;所述引物对23由序列表中序列67和序列68所示的两条单链DNA组成;所述延伸引物23为序列表中序列69所示单链DNA;

(24) 用于检测rs8072587的引物对24和延伸引物24;所述引物对24由序列表中序列70和序列71所示的两条单链DNA组成;所述延伸引物24为序列表中序列72所示单链DNA;

(25) 用于检测rs8137373的引物对25和延伸引物25;所述引物对25由序列表中序列73和序列74所示的两条单链DNA组成;所述延伸引物25为序列表中序列75所示单链DNA;

(26) 用于检测rs9522149的引物对26和延伸引物26;所述引物对26由序列表中序列76和序列77所示的两条单链DNA组成;所述延伸引物26为序列表中序列78所示单链DNA;

(27) 用于检测rs9809818的引物对27和延伸引物27;所述引物对27由序列表中序列79和序列80所示的两条单链DNA组成;所述延伸引物27为序列表中序列81所示单链DNA;

(28) 用于检测rs9908046的引物对28和延伸引物28;所述引物对28由序列表中序列82和序列83所示的两条单链DNA组成;所述延伸引物28为序列表中序列84所示单链DNA。

5. 根据权利要求4所述的单链DNA组,其特征在于:所述单链DNA组中,所述引物对1、所述引物对2、所述引物对3、所述引物对4、所述引物对5、所述引物对6、所述引物对7、所述引物对8、所述引物对9、所述引物对10、所述引物对11、所述引物对12、所述引物对13、所述引物对14、所述引物对15、所述引物对16、所述引物对17、所述引物对18、所述引物对19、所述引物对20、所述引物对21、所述引物对22、所述引物对23、所述引物对24、所述引物对25、所述引物对26、所述引物对27和所述引物对28的摩尔比为0.8:0.6:1.5:2.7:3:0.8:2:1.3:1:1:1.5:1:2.7:0.5:0.8:2.5:0.4:1.6:2.5:4:3:0.8:0.8:3:3:6:0.6:0.6;每个引物对中两条引物的摩尔比均为1:1;

所述延伸引物1、所述延伸引物2、所述延伸引物3、所述延伸引物4、所述延伸引物5、所述延伸引物6、所述延伸引物7、所述延伸引物8、所述延伸引物9、所述延伸引物10、所述延伸引物11、所述延伸引物12、所述延伸引物13、所述延伸引物14、所述延伸引物15、所述延伸引物16、所述延伸引物17、所述延伸引物18、所述延伸引物19、所述延伸引物20、所述延伸引物21、所述延伸引物22、所述延伸引物23、所述延伸引物24、所述延伸引物25、所述延伸引物26、所述延伸引物27和所述延伸引物28的摩尔比为0.45:0.35:1.2:1.7:4:1:3:0.8:1:0.8:1.8:1.1:1.1:1.1:0.8:1.4:1:0.5:1.3:1.8:1.6:0.9:0.9:2:2.3:3:1:0.6。

6. 用于区分五大洲际人群的试剂盒,含有权利要求2或3所述的引物对组或权利要求4

或5所述的单链DNA组,以及如下物质中的至少一种:dNTP、DNA聚合酶、碱性磷酸酶。

7. 用于检测28个SNP位点组合的物质在如下任一中的应用:

- (a) 构建五大洲际人群基因分型数据库;
- (b) 区分五大洲际人群;

所述28个SNP位点分别为:rs10483251、rs12142199、rs1229984、rs12402499、rs12498138、rs12594144、rs1426654、rs1557553、rs16891982、rs17822931、rs1871534、rs2080161、rs2139931、rs2789823、rs2814778、rs3751050、rs3827760、rs4657449、rs4749305、rs4792928、rs6054465、rs6437783、rs715605、rs8072587、rs8137373、rs9522149、rs9809818和rs9908046。

8. 根据权利要求7所述的应用,其特征在于:所述用于检测28个SNP位点组合的物质为权利要求2或3所述的引物对组或权利要求4或5所述的单链DNA组或权利要求6所述的试剂盒。

9. 一种构建五大洲际人群基因分型数据库的方法,包括如下步骤:

(a1) 从千人基因组项目和人类基因组多样性计划中选取五大洲际人群的28个SNP位点分型形成原始分型库;

(a2) 将原始分型库里所有样本进行structure聚类分析,从中选取祖先主成分大于90%的部分即构成五大洲际人群基因分型数据库;

所述28个SNP位点分别为:rs10483251、rs12142199、rs1229984、rs12402499、rs12498138、rs12594144、rs1426654、rs1557553、rs16891982、rs17822931、rs1871534、rs2080161、rs2139931、rs2789823、rs2814778、rs3751050、rs3827760、rs4657449、rs4749305、rs4792928、rs6054465、rs6437783、rs715605、rs8072587、rs8137373、rs9522149、rs9809818和rs9908046。

10. 一种区分五大洲际人群的方法,包括如下步骤:

(b1) 按照权利要求9所述方法构建五大洲际人群基因分型数据库;

(b2) 提取待测者的基因组DNA,并进行28个SNP位点的检测,获得所述待测者在所述28个SNP位点上的原始基因型数据;

(b3) 将所述待测者在所述28个SNP位点上的原始基因型数据与所述五大洲际人群基因分型数据库进行比对,从而确定所述待测者属于五大洲际人群中的哪一种;

所述28个SNP位点分别为:rs10483251、rs12142199、rs1229984、rs12402499、rs12498138、rs12594144、rs1426654、rs1557553、rs16891982、rs17822931、rs1871534、rs2080161、rs2139931、rs2789823、rs2814778、rs3751050、rs3827760、rs4657449、rs4749305、rs4792928、rs6054465、rs6437783、rs715605、rs8072587、rs8137373、rs9522149、rs9809818和rs9908046。

11. 根据权利要求10所述的方法,其特征在于:进行所述28个SNP位点的检测时,采用的是权利要求2或3所述的引物对组或权利要求4或5所述的单链DNA组或权利要求6所述的试剂盒;进行的是28重PCR扩增,退火温度为55℃。

对未知来源个体进行五大洲际族群来源推断的方法和系统

技术领域

[0001] 本发明属于生物技术领域,涉及一种对未知来源个体进行五大洲际族群来源推断的方法和系统。

背景技术

[0002] 检测人群间分布差异大的DNA多态性位点即祖先信息位点(Ancestry informative marks,AIMs)可以推断犯罪现场DNA供者的族群地域来源。短串联重复序列(STRs)、单核苷酸多态性(SNPs)、插入/缺失多态性(Indels)等均可作为祖先信息位点(AIMs)用于族群推断,发挥了良好的作用。而SNPs有HapMap Project、1000Genomes等数据库支持,成为近年来筛选AIMs位点的重要遗传标记。目前报道了大量洲际大人群众体区分的AIMs体系,例如本项目组的27-SNPs等用于非洲、东亚和欧洲三大人群的推断。在现行法医DNA实验室检测技术条件下,表现良好的SNPs族群推断体系应在确保人群区分效能的前提下,各人群之间的区分能力要保持均衡;位点信息量尽量高,位点数目尽量少;检测方法简单易行。

发明内容

[0003] 本发明的目的是提供一种对未知来源个体进行五大洲际族群来源推断的方法和系统。

[0004] 本发明保护28个SNP位点在如下任一中的应用:

[0005] (a) 构建五大洲际人群基因分型数据库;

[0006] (b) 区分五大洲际人群;

[0007] 所述28个SNP位点分别为:rs10483251、rs12142199、rs1229984、rs12402499、rs12498138、rs12594144、rs1426654、rs1557553、rs16891982、rs17822931、rs1871534、rs2080161、rs2139931、rs2789823、rs2814778、rs3751050、rs3827760、rs4657449、rs4749305、rs4792928、rs6054465、rs6437783、rs715605、rs8072587、rs8137373、rs9522149、rs9809818、rs9908046(表2)。下文所述28个SNP位点与此相同。

[0008] 本发明保护用于检测人基因组中28个SNP位点的引物对组。

[0009] 本发明所提供的用于检测人基因组中28个SNP位点的引物对组由如下(1)-(28)组成:(1)用于检测rs10483251的引物对1,由序列表中序列1和序列2所示的两条单链DNA组成;(2)用于检测rs12142199的引物对2,由序列表中序列4和序列5所示的两条单链DNA组成;(3)用于检测rs1229984的引物对3,由序列表中序列7和序列8所示的两条单链DNA组成;(4)用于检测rs12402499的引物对4,由序列表中序列10和序列11所示的两条单链DNA组成;(5)用于检测rs12498138的引物对5,由序列表中序列13和序列14所示的两条单链DNA组成;(6)用于检测rs12594144的引物对6,由序列表中序列16和序列17所示的两条单链DNA组成;(7)用于检测rs1426654的引物对7,由序列表中序列19和序列20所示的两条单链DNA组成;(8)用于检测rs1557553的引物对8,由序列表中序列22和序列23所示的两条单链DNA组成;

(9) 用于检测rs16891982的引物对9,由序列表中序列25和序列26所示的两条单链DNA组成;
(10) 用于检测rs17822931的引物对10,由序列表中序列28和序列29所示的两条单链DNA组成;
(11) 用于检测rs1871534的引物对11,由序列表中序列31和序列32所示的两条单链DNA组成;
(12) 用于检测rs2080161的引物对12,由序列表中序列34和序列35所示的两条单链DNA组成;
(13) 用于检测rs2139931的引物对13,由序列表中序列37和序列38所示的两条单链DNA组成;
(14) 用于检测rs2789823的引物对14,由序列表中序列40和序列41所示的两条单链DNA组成;
(15) 用于检测rs2814778的引物对15,由序列表中序列43和序列44所示的两条单链DNA组成;
(16) 用于检测rs3751050的引物对16,由序列表中序列46和序列47所示的两条单链DNA组成;
(17) 用于检测rs3827760的引物对17,由序列表中序列49和序列50所示的两条单链DNA组成;
(18) 用于检测rs4657449的引物对18,由序列表中序列52和序列53所示的两条单链DNA组成;
(19) 用于检测rs4749305的引物对19,由序列表中序列55和序列56所示的两条单链DNA组成;
(20) 用于检测rs4792928的引物对20,由序列表中序列58和序列59所示的两条单链DNA组成;
(21) 用于检测rs6054465的引物对21,由序列表中序列61和序列62所示的两条单链DNA组成;
(22) 用于检测rs6437783的引物对22,由序列表中序列64和序列65所示的两条单链DNA组成;
(23) 用于检测rs715605的引物对23,由序列表中序列67和序列68所示的两条单链DNA组成;
(24) 用于检测rs8072587的引物对24,由序列表中序列70和序列71所示的两条单链DNA组成;
(25) 用于检测rs8137373的引物对25,由序列表中序列73和序列74所示的两条单链DNA组成;
(26) 用于检测rs9522149的引物对26,由序列表中序列76和序列77所示的两条单链DNA组成;
(27) 用于检测rs9809818的引物对27,由序列表中序列79和序列80所示的两条单链DNA组成;
(28) 用于检测rs9908046的引物对28,由序列表中序列82和序列83所示的两条单链DNA组成。

[0010] 所述引物对组中,所述引物对1、所述引物对2、所述引物对3、所述引物对4、所述引物对5、所述引物对6、所述引物对7、所述引物对8、所述引物对9、所述引物对10、所述引物对11、所述引物对12、所述引物对13、所述引物对14、所述引物对15、所述引物对16、所述引物对17、所述引物对18、所述引物对19、所述引物对20、所述引物对21、所述引物对22、所述引物对23、所述引物对24、所述引物对25、所述引物对26、所述引物对27和所述引物对28的摩尔比为0.8:0.6:1.5:2.7:3:0.8:2:1.3:1:1:1.5:1:2.7:0.5:0.8:2.5:0.4:1.6:2.5:4:3:0.8:0.8:3:3:6:0.6:0.6。每个引物对中两条引物的摩尔比均为1:1。

[0011] 本发明保护用于检测人基因组中28个SNP位点的单链DNA组。

[0012] 本发明所提供的用于检测人基因组中28个SNP位点的单链DNA组由如下(1)-(28)组成:(1) 用于检测rs10483251的引物对1和延伸引物1;所述引物对1由序列表中序列1和序列2所示的两条单链DNA组成;所述延伸引物1为序列表中序列3所示单链DNA;(2) 用于检测rs12142199的引物对2和延伸引物2;所述引物对2由序列表中序列4和序列5所示的两条单链DNA组成;所述延伸引物2为序列表中序列6所示单链DNA;(3) 用于检测rs1229984的引物对3和延伸引物3;所述引物对3由序列表中序列7和序列8所示的两条单链DNA组成;所述延伸引物3为序列表中序列9所示单链DNA;(4) 用于检测rs12402499的引物对4和延伸引物4;所述引物对4由序列表中序列10和序列11所示的两条单链DNA组成;所述延伸引物4为序列表中序列12所示单链DNA;(5) 用于检测rs12498138的引物对5和延伸引物5;所述引物对5由序列表中序列13和序列14所示的两条单链DNA组成;所述延伸引物5为序列表中序列15所示

单链DNA；(6) 用于检测rs12594144的引物对6和延伸引物6；所述引物对6由序列表中序列16和序列17所示的两条单链DNA组成；所述延伸引物6为序列表中序列18所示单链DNA；(7) 用于检测rs1426654的引物对7和延伸引物7；所述引物对7由序列表中序列19和序列20所示的两条单链DNA组成；所述延伸引物7为序列表中序列21所示单链DNA；(8) 用于检测rs1557553的引物对8和延伸引物8；所述引物对8由序列表中序列22和序列23所示的两条单链DNA组成；所述延伸引物8为序列表中序列24所示单链DNA；(9) 用于检测rs16891982的引物对9和延伸引物9；所述引物对9由序列表中序列25和序列26所示的两条单链DNA组成；所述延伸引物9为序列表中序列27所示单链DNA；(10) 用于检测rs17822931的引物对10和延伸引物10；所述引物对10由序列表中序列28和序列29所示的两条单链DNA组成；所述延伸引物10为序列表中序列30所示单链DNA；(11) 用于检测rs1871534的引物对11和延伸引物11；所述引物对11由序列表中序列31和序列32所示的两条单链DNA组成；所述延伸引物11为序列表中序列33所示单链DNA；(12) 用于检测rs2080161的引物对12和延伸引物12；所述引物对12由序列表中序列34和序列35所示的两条单链DNA组成；所述延伸引物12为序列表中序列36所示单链DNA；(13) 用于检测rs2139931的引物对13和延伸引物13；所述引物对13由序列表中序列37和序列38所示的两条单链DNA组成；所述延伸引物13为序列表中序列39所示单链DNA；(14) 用于检测rs2789823的引物对14和延伸引物14；所述引物对14由序列表中序列40和序列41所示的两条单链DNA组成；所述延伸引物14为序列表中序列42所示单链DNA；(15) 用于检测rs2814778的引物对15和延伸引物15；所述引物对15由序列表中序列43和序列44所示的两条单链DNA组成；所述延伸引物15为序列表中序列45所示单链DNA；(16) 用于检测rs3751050的引物对16和延伸引物16；所述引物对16由序列表中序列46和序列47所示的两条单链DNA组成；所述延伸引物16为序列表中序列48所示单链DNA；(17) 用于检测rs3827760的引物对17和延伸引物17；所述引物对17由序列表中序列49和序列50所示的两条单链DNA组成；所述延伸引物17为序列表中序列51所示单链DNA；(18) 用于检测rs4657449的引物对18和延伸引物18；所述引物对18由序列表中序列52和序列53所示的两条单链DNA组成；所述延伸引物18为序列表中序列54所示单链DNA；(19) 用于检测rs4749305的引物对19和延伸引物19；所述引物对19由序列表中序列55和序列56所示的两条单链DNA组成；所述延伸引物19为序列表中序列57所示单链DNA；(20) 用于检测rs4792928的引物对20和延伸引物20；所述引物对20由序列表中序列58和序列59所示的两条单链DNA组成；所述延伸引物20为序列表中序列60所示单链DNA；(21) 用于检测rs6054465的引物对21和延伸引物21；所述引物对21由序列表中序列61和序列62所示的两条单链DNA组成；所述延伸引物21为序列表中序列63所示单链DNA；(22) 用于检测rs6437783的引物对22和延伸引物22；所述引物对22由序列表中序列64和序列65所示的两条单链DNA组成；所述延伸引物22为序列表中序列66所示单链DNA；(23) 用于检测rs715605的引物对23和延伸引物23；所述引物对23由序列表中序列67和序列68所示的两条单链DNA组成；所述延伸引物23为序列表中序列69所示单链DNA；(24) 用于检测rs8072587的引物对24和延伸引物24；所述引物对24由序列表中序列70和序列71所示的两条单链DNA组成；所述延伸引物24为序列表中序列72所示单链DNA；(25) 用于检测rs8137373的引物对25和延伸引物25；所述引物对25由序列表中序列73和序列74所示的两条单链DNA组成；所述延伸引物25为序列表中序列75所示单链DNA；(26) 用于检测rs9522149的引物对26和延伸引物26；所述引物对26由序列表中序列76和序列77所示的两条单链DNA

组成;所述延伸引物26为序列表中序列78所示单链DNA;(27)用于检测rs9809818的引物对27和延伸引物27;所述引物对27由序列表中序列79和序列80所示的两条单链DNA组成;所述延伸引物27为序列表中序列81所示单链DNA;(28)用于检测rs9908046的引物对28和延伸引物28;所述引物对28由序列表中序列82和序列83所示的两条单链DNA组成;所述延伸引物28为序列表中序列84所示单链DNA。

[0013] 所述单链DNA组中,所述引物对1、所述引物对2、所述引物对3、所述引物对4、所述引物对5、所述引物对6、所述引物对7、所述引物对8、所述引物对9、所述引物对10、所述引物对11、所述引物对12、所述引物对13、所述引物对14、所述引物对15、所述引物对16、所述引物对17、所述引物对18、所述引物对19、所述引物对20、所述引物对21、所述引物对22、所述引物对23、所述引物对24、所述引物对25、所述引物对26、所述引物对27和所述引物对28的摩尔比为0.8:0.6:1.5:2.7:3:0.8:2:1.3:1:1:1.5:1:2.7:0.5:0.8:2.5:0.4:1.6:2.5:4:3:0.8:0.8:3:3:6:0.6:0.6;每个引物对中两条引物的摩尔比均为1:1。所述延伸引物1、所述延伸引物2、所述延伸引物3、所述延伸引物4、所述延伸引物5、所述延伸引物6、所述延伸引物7、所述延伸引物8、所述延伸引物9、所述延伸引物10、所述延伸引物11、所述延伸引物12、所述延伸引物13、所述延伸引物14、所述延伸引物15、所述延伸引物16、所述延伸引物17、所述延伸引物18、所述延伸引物19、所述延伸引物20、所述延伸引物21、所述延伸引物22、所述延伸引物23、所述延伸引物24、所述延伸引物25、所述延伸引物26、所述延伸引物27和所述延伸引物28的摩尔比为0.45:0.35:1.2:1.7:4:1:3:0.8:1:0.8:1.8:1.1:1.1:1.1:0.8:1.4:1:0.5:1.3:1.8:1.6:0.9:0.9:2:2.3:3:1:0.6。

[0014] 本发明还保护用于区分五大洲际人群的试剂盒。

[0015] 本发明提供的用于区分五大洲际人群的试剂盒含有前文所述引物对组或单链DNA组。

[0016] 根据需要,所述试剂盒还可含有如下物质中的至少一种:dNTP、DNA聚合酶、碱性磷酸酶。

[0017] 本发明还保护用于检测28个SNP位点的物质在如下任一中的应用:

[0018] (a) 构建五大洲际人群基因分型数据库;

[0019] (b) 区分五大洲际人群。

[0020] 其中,所述用于检测28个SNP位点的物质可为前文所述的引物对组或单链DNA组或试剂盒。

[0021] 本发明还保护一种构建五大洲际人群基因分型数据库的方法。

[0022] 本发明所提供的构建五大洲际人群基因分型数据库的方法,具体可包括如下步骤:

[0023] (a1) 从千人基因组项目和人类基因组多样性计划中选取五大洲际人群的所述28个SNP位点分型形成原始分型库;

[0024] (a2) 将原始分型库里所有样本进行structure聚类分析,从中选取祖先主成分大于90%的部分即构成五大洲际人群基因分型数据库。

[0025] Structure是一款免费的、多平台(Windows,Mac,Linux)、开源的、使用由不连锁的标记组成的多位点基因型数据实施基于模型的聚类方法来推断人群组成结构的经典生物信息软件,并广泛用于人类遗传学、群体遗传学、法医遗传学等领域。所用参数为:

10000burnins,10000repetitions,混合模型;运行结果可得知每个样本的祖先成分比例。

[0026] 本发明还保护一种区分五大洲际人群的方法。

[0027] 本发明所提供的区分五大洲际人群的方法,具体可包括如下步骤:

[0028] (b1) 按照前文所述方法构建五大洲际人群基因分型数据库;

[0029] (b2) 提取待测者的基因组DNA,并进行28个SNP位点的检测,获得所述待测者在所述28个SNP位点上的原始基因型数据;

[0030] (b3) 将所述待测者在所述28个SNP位点上的原始基因型数据与所述五大洲际人群基因分型数据库通过分析方法进行比对,从而确定所述待测者属于五大洲际人群中的哪一种。

[0031] 在上述两种方法中,进行所述28个SNP位点的检测时,采用的是前文所述的引物对组或单链DNA组或试剂盒;进行的是28重PCR扩增,退火温度为55℃。

[0032] 前文所述的五大洲际为东亚、欧洲、非洲、大洋洲和美洲。

[0033] 为实现五大洲际人群(东亚、欧洲、非洲、大洋洲和美洲)的区分,本发明筛选出28个SNP位点,构建复合检测体系,用该体系检测来自16个人群的712份样本,检测结果与千人基因组中的20个人群和CEPH库中的2个人群合并共计2804份个体的分型数据,采用聚类分析的方法和主成分分析方法进行体系效能评价。选取祖先主成分大于90%的样本构建参考人群基因分型数据库,对140份已知祖先来源的个体进行群体匹配概率、个体主成分分析等进行人群来源推断,评估该体系在实际样本中的人群来源区分能力。结果发现,该体系不仅能对五大洲际人群进行区分,而且对混合人群也有一定的区分能力,对已知来源样本的祖先主成分和人群匹配与其来源信息一致。并能够推断个体祖先来源成分组成,在实际检案中可以进行推广应用。

附图说明

[0034] 图1为DNA浓度为5ng/μL样本分型图(28-plex SNP检测结果)。图中,1-28表示28个SNP(与表2中的编号相对应),每个数字后面是对应的是分型数据(由于有的检测的是互补链,所以数字后显示的分型数据为表2中的互补碱基)。

[0035] 图2为基于28SNPs在38人群中分型基因频率的主成分分析。A:非洲个体;B:美洲个体;C:东亚个体;D:欧洲个体;E:混合人群;F:大洋洲个体。

[0036] 图3为28个SNPs对38人群的Structure分析结果。

[0037] 图4为测试样本的人群归类分析图。A:非洲个体;B:美洲个体;C:东亚个体(广西汉);D:东亚个体(河南汉);E:欧洲个体;F:维族;G:大洋洲个体。

[0038] 图5为本发明和文章体系对“9947”和“HWQ”样本的检测图谱。A:“9947”样本文章体系检测结果;B:“HWQ”样本文章体系检测结果;C:“9947”样本本发明体系检测结果;D:“HWQ”样本本发明体系检测结果。A-D中,1-28表示28个SNP(与表6中“编号”一栏的数字相对应),每个数字后面是对应的是分型数据。

具体实施方式

[0039] 下述实施例中所使用的实验方法如无特殊说明,均为常规方法。

[0040] 下述实施例中所用的材料、试剂等,如无特殊说明,均可从商业途径得到。

[0041] 实施例1、本发明五大洲际人群基因分型数据库的构建及应用

[0042] 一、材料和方法

[0043] 1、样本信息

[0044] 本发明选取公共数据库千人基因组 (1000genomes) 里的20个人群2000个样本及人类基因组多样性计划 (HGDP-CEPH) 中2个人群92个样本,检测样本包括16个人群的712份样本,共38个人群2804个个体作为验证样本。详细信息见表1。

[0045] 表1人群样本信息表

	国家地区	人群	缩写	样本数 (N)	来源
	1 非洲(AFR)	Esan in Nigeria	ESN	99	1000 Genomes
	2	Luhya in Webuye	LWK	99	1000 Genomes
	3	Mandingka tribe in Gambia	MAG	113	1000 Genomes
	4	Mende in Sierra Leone	MSL	85	1000 Genomes
	5	Yoruba in Ibadan,Nigeria	YRI	108	1000 Genomes
	6	Mbuti Pygmies	MBU	38	Kidd Lab
	7	Masai	MAS	20	Kidd Lab
	8	Sandawe	SND	10	Kidd Lab
	9	<i>Yoruba</i>	<i>YOR</i>	37(20)	Kidd Lab
	10 美洲(AMR)	America	AMR	64	HGDP-CEPH
	11	<i>Karitiana</i>	<i>KAR</i>	54(20)	Kidd Lab
	12	Pima, Mexico	PMX	5	Kidd Lab
	13	Guihiba speakers	GHB	12	Kidd Lab
	14 东亚(EAS)	Chinese Dai in Xishuangbanna	CDX	93	1000 Genomes
	15	Han Chinese in Beijing	CHB	103	1000 Genomes
	16	Southern Han Chinese	CHS	105	1000 Genomes
	17	Japanese in Tokyo	JPT	104	1000 Genomes
	18	Kinh in Chi Minh City	KHV	99	1000 Genomes
[0046]	19	<i>Han Chinese in Henan</i>	<i>HNH</i>	63(20)	Caixia lab
	20	<i>Han Chinese in Guangxi</i>	<i>GXH</i>	54(20)	Caixia lab
	21	Hakka	HKA	16	Kidd Lab
	22 欧洲(EUR)	Utah residents with European ancestry	CEU	99	1000 Genomes
	23	Finnish in Finland	FIN	99	1000 Genomes
	24	British in England and Scotland	GBR	91	1000 Genomes
	25	Iberian population in Spain	IBS	107	1000 Genomes
	26	Tosceni in Italy	TSI	107	1000 Genomes
	27	<i>EuroAmer (Not Perf.)</i>	<i>EAM</i>	91(20)	Kidd Lab
	28 大洋洲(OCE)	Oceania	OCE	28	HGDP-CEPH
	29	<i>Nasioi Melanesians</i>	<i>NAS</i>	22(20)	Kidd Lab
	30	Papua-New Guinean	PNG	22	Kidd Lab
	31 中亚 (CA)	<i>Uyghurs</i>	<i>UY</i>	237(20)	Caixia lab
	32 中南亚(SCA)	Kachari from Assam	KCH	17	Kidd Lab
	33	Thoti	THT	14	Kidd Lab
	34 南亚(SA)	Bengali in Bangladesh	BEB	86	1000 Genomes
	35	Gujarati Indian in Houston	GIH	103	1000 Genomes
	36	Indian Telugu in the UK	ITU	102	1000 Genomes
	37	Punjabi in Lahore,Pakistan	PJL	96	1000 Genomes
	38	Sri Lankan Tamli in the UK	STU	102	1000 Genomes

[0047] 注:斜体表示测试人群,括号里表示所选用的测试样本的数目。

[0048] 2、SNPs位点的来源

[0049] 通过对“The Global AIMs Nano set”的分析,删去31个AIMs (de la Puente M, Santos C, Fondevila M, et al. The Global AIMs Nano set: A 31-plex SNaPshot assay of ancestry-informative SNPs [J]. Forensic Science International: Genetics, 2016, 22: 81-88.) 中的3个三等位SNP位点,利用剩余的28个二等位基因SNPs位点进行复合体系的构建,位点信息见表2。

[0050] 表2 28个SNP位点的详细信息

编号	dbSNPrs#	染色体	在基因组 37 上的位置	等位基因	世界地区*
1	rs715605	22	30640308	A G	Oceania
2	rs1229984	4	100239319	T C	Asia
3	rs1426654	15	48426484	A G	Europe
4	rs1557553	22	44760984	C T	America
5	rs1871534	8	145639681	G C	Africa
6	rs2080161	7	13331150	A C	America
7	rs2139931	1	84590527	A G	Oceania
8	rs2789823	9	136769888	G A	Africa
9	rs2814778	1	159174683	T C	Africa
10	rs3751050	11	9091244	T C	Oceania
11	rs3827760	2	109513601	A G	Asia
12	rs4657449	1	165465281	G A	Asia
13	rs4749305	10	28391596	A G	Europe
[0051]	rs4792928	17	42105174	T C	Europe
15	rs6054465	20	6673018	T C	Oceania
16	rs6437783	3	108172817	C T	Asia
17	rs8072587	17	19211073	C G	Europe
18	rs8137373	22	41729216	G A	America
19	rs9809818	3	71480566	C A	Oceania
20	rs9522149	13	111827167	T C	Europe
21	rs9908046	17	53563782	C T	Oceania
22	rs10483251	14	21671277	G T	America
23	rs12142199	1	1249187	G A	Europe
24	rs12402499	1	101528954	G A	America
25	rs12498138	3	121459589	G A	America
26	rs12594144	15	64161351	C A	Asia
27	rs16891982	5	33951693	C G	Europe
28	rs17822931	16	48258198	C T	Asia

[0052] 3、PCR复合扩增

[0053] PCR复合扩增反应体系为5 μ L,内含10 \times PCR buffer (含Mg²⁺15mmol/L) 0.6 μ L, 25mmol/L MgCl₂ 0.9 μ L, 10mmol/L dNTP 0.1 μ L, 复合扩增引物0.7 μ L, 5U/ μ L HotStarTaq[®] plus DNA聚合酶 (QIAGEN公司, 德国) 0.1 μ L, 5ng/ μ L模板DNA 1 μ L, 水补足至5 μ L。PCR反应条件: 95 $^{\circ}$ C 10min后; 95 $^{\circ}$ C 30s, 55 $^{\circ}$ C 40s, 72 $^{\circ}$ C 1min, 循环40次; 最后延伸72 $^{\circ}$ C 20min。纯化反应体系为7.5 μ L, 内含扩增产物5 μ L, H₂O 1 μ L, ExoI (10U/ μ L) 0.2 μ L, SAP (1U/ μ L) 1 μ L, 10 \times SAP buffer 0.3 μ L。充分振荡混匀后37 $^{\circ}$ C 孵育45min, 85 $^{\circ}$ C 15min 灭活酶活性。

[0054] 使用SNaPshot[®] Multiplex Kit (ABI公司, 美国) 进行单碱基延伸反应, 采用5.5 μ L

体系,内含纯化后PCR产物2 μ L,复合延伸引物1 μ L,SNaPshot mix 2.5 μ L。PCR反应条件:96 $^{\circ}$ C 10s后;59 $^{\circ}$ C 5s,60 $^{\circ}$ C 30s,循环33次。之后向反应产物中加入1 μ L SAP (1U/ μ L),37 $^{\circ}$ C 孵育 80min,85 $^{\circ}$ C 15min进行纯化处理,去除多余的引物和dNTP。

[0055] 在3130-XL遗传分析仪(ABI公司,美国)上对上述延伸纯化产物进行毛细管电泳检测。检测体系10 μ L,包括1 μ L单碱基延伸纯化后产物、9 μ L甲酰胺和内标GeneScanLiz-120 (38:1体积比)的混合物(ABI公司,美国)。电泳参数设置:进样时间18s,进样电压3kV,电泳电压13.4kV,电泳时间15min。根据Genemapper ID v3.2软件进行基因分型分析。

[0056] 其中,用于检测各SNP位点的引物序列及其浓度详见表3。

[0057] 表3用于检测各SNP位点的引物序列及其浓度

[0058]

位点	标号	扩增引物(F)	扩增引物(R)	扩增引物浓度(μ M)	延伸引物	碱基数	延伸引物浓度(μ M)
rs10483251	EX-28*22	序列1	序列2	0.8	序列3	48	0.45
rs12142199	EX-28*23	序列4	序列5	0.6	序列6	63	0.35
rs1229984	EX-28*2	序列7	序列8	1.5	序列9	73	1.2
rs12402499	EX-28*24	序列10	序列11	2.7	序列12	51	1.7
rs12498138	EX-28*25	序列13	序列14	3	序列15	63	4
rs12594144	EX-28*26	序列16	序列17	0.8	序列18	40	1
rs1426654	EX-28*3	序列19	序列20	2	序列21	67	3
rs1557553	EX-28*4	序列22	序列23	1.3	序列24	93	0.8
rs16891982	EX-28*27	序列25	序列26	1	序列27	80	1
rs17822931	EX-28*28	序列28	序列29	1	序列30	59	0.8
rs1871534	EX-28*5	序列31	序列32	1.5	序列33	105	1.8
rs2080161	EX-28*6	序列34	序列35	1	序列36	48	1.1
rs2139931	EX-28*7	序列37	序列38	2.7	序列39	109	1.1
rs2789823	EX-28*8	序列40	序列41	0.5	序列42	51	1.1
rs2814778	EX-28*9	序列43	序列44	0.8	序列45	89	0.8
rs3751050	EX-28*10	序列46	序列47	2.5	序列48	44	1.4
rs3827760	EX-28*11	序列49	序列50	0.4	序列51	89	1
rs4657449	EX-28*12	序列52	序列53	1.6	序列54	93	0.5
rs4749305	EX-28*13	序列55	序列56	2.5	序列57	67	1.3
rs4792928	EX-28*14	序列58	序列59	4	序列60	99	1.8
rs6054465	EX-28*15	序列61	序列62	3	序列63	84	1.6
rs6437783	EX-28*16	序列64	序列65	0.8	序列66	45	0.9
rs715605	EX-28*1	序列67	序列68	0.8	序列69	113	0.9
rs8072587	EX-28*17	序列70	序列71	3	序列72	96	2
rs8137373	EX-28*18	序列73	序列74	3	序列75	100	2.3
rs9522149	EX-28*19	序列76	序列77	6	序列78	72	3
rs9809818	EX-28*20	序列79	序列80	0.6	序列81	55	1
rs9908046	EX-28*21	序列82	序列83	0.6	序列84	109	0.6

[0059] 二、软件及分析方法

[0060] 1、主成分分析(PCA)

[0061] 利用Rv3.2.3软件进行主成分分析:a、将包括公共数据库样本和检测样本在内的38个人群,按洲际和地域分为非洲、美洲、东亚、欧洲、大洋洲、混合人群(中亚、中南亚、南亚),并进行基于基因频率的群体主成分分析;b、随机从7个测试人群中(表1)各抽取一个样本进行个体主成分分析(用R v3.3.2进行主成分分析,并用R程序包ggplot2画出种族归类图)。

[0062] 2、聚类分析

[0063] 针对表1中的38个人群,用Structure.v2.3.4软件进行聚类分析(K取3-7),分析各

人群的遗传结构,使用Distruct 1.1绘制人群聚类结果图;并对上述进行个体主成分分析的7个样本进行个体祖先成分的统计。

[0064] Structure是一款免费的、多平台(Windows,Mac,Linux)、开源的、使用由不连锁的标记组成的多位点基因型数据实施基于模型的聚类方法来推断人群组成结构的经典生物信息软件,并广泛用于人类遗传学、群体遗传学、法医遗传学等领域。所用参数为:10000burnins,10000repetitions,混合模型;运行结果可得知每个样本的祖先成分比例。

[0065] 3、随机人群匹配概率

[0066] 用法医智能软件对随机挑选的来自7个人群140份测试样本(表1中已标注)进行人群随机匹配的计算。

[0067] 三、实验结果

[0068] 1、28-plex SNP检测结果

[0069] 结果如图1所示,由图可见:28个AIMs符合以下标准:1)对五大洲际人群的区分能力保持良好的均衡性;2)位点间间隔至少1Mb,减少了连锁遗传的发生几率。该体系采用较为普及的SNaPshot检测分型技术,28个位点等位基因均可明显判断分型。

[0070] 2、对体系区分效能的评价

[0071] (1)主成分分析

[0072] 结果如图2所示,由图可见:主成分1(PC1)和主成分2(PC2)解释了61.3%的差异,28个位点可以把38个人群很明显的区分成六大部分,结合表1我们发现,分布较集中的人群为同一洲际祖先来源,来源于五大洲际(非洲、欧洲、东亚、美洲、大洋洲)的30个人群被很明显的区分为五大部分,在主成分1中美洲、东亚、大洋洲、非洲、欧洲被依次区分开,其中东亚人群和欧洲人群分布相对较集中,混合人群(8个)的人群分布相对较分散,但均位于东亚人群和欧洲之间,该部分属于南亚和欧亚混合人群,其分布的不集中也进一步说明了其遗传结构的复杂,需要进一步研究;在主成分2中,大洋洲和美洲人群被区分开。

[0073] (2)聚类分析

[0074] 使用该28个AIMs,对上述的38个人群共2804个样本进行遗传结构的分析。结果如图3所示,由图可见:当K=3时,人群被聚类为三大部分:非洲、欧洲被区分开,美洲、大洋洲和东亚人群呈现出一致的祖先组成成分,维族等混合人群的遗传成分呈现在欧洲和东亚祖先成分连续分布。随着K值增加,先后在美国、大洋洲出现新的祖先主成分,当K值为6时,所有个体对应于6个人群组:非洲、美洲、大洋洲、东亚、混合人群、欧洲,混合人群出现独立祖先成分,说明维吾尔族和南亚的混合人种经过长期的融合与进化,已经成为遗传成分相对稳定的过渡人种。当K值增加到7时,南亚混合人群出现新的成分,说明其与中亚的混合人群在祖先来源与遗传结构上有差异。当K值继续增加时,东亚人群未出现进一步分层现象,说明该体系不能对东亚局部的亚人群进一步区分。

[0075] (3)单独个体的人群来源推断测试

[0076] 为避免混合成分出现,同时达到最大的区分能力,选择K=6作为该体系较理想的区分能力;同时为了进一步应用于法医实际检案中,提高推断的准确性,本发明选取祖先成分大于90%的个体(共2201个样本)作为参考样本构建参考人群基因分型数据库,并采用以下统计方法进行该体系个体推断能力的评价。

[0077] A.似然比

[0078] 对已知个体来源的140个样本基于参考数据库(不包含140个测试样本)进行随机人群匹配概率进行计算,基于似然比进行其可能洲际人群来源的统计。群体匹配概率即随机匹配概率,简单地说,就是对某位点组合的一个特定分型可能出现在人群中的估计概率,也可以理解为从人群中随机抽取一份样本,会出现特定DNA分型的理论概率。LR值的计算具体为:未知个体概率最大的群体匹配概率为分母,其他群体的匹配概率为分子,依次得到不同人群的似然比值。结果如表4。140个测试样本中祖先来源推断与样本信息一致的有137个;另外3个来源于维族的样本祖先推断来源分别为混合人群、东亚和欧洲,但其似然比均小于100,推断结果不排除样本信息。综上,该体系对测试样本祖先来源推断绝对准确率达97.86%,另有2.14%不能排除样本信息来源。结合表4可以看出该体系在对五大洲际人群来源推断时准确性较高,而在进行混合人群推断时需要综合祖先成分和MP值进行分析。另外,种族和民族不是一致的概念,维族地处亚欧交界,属欧亚混合人种,不同人种之间的融合没有明显的界限,户籍登记中的民族信息也并非完全与种族一致,三个维族样本虽没达到判别标准,但推断结果符合其地理位置分布,在以后对该人群的区分中人要进行多方面的综合考虑。

[0079] 表4测试样本匹配概率结果

世界地区 测试样本	非洲	美洲	东亚	欧洲	大洋洲	混合人群
YOR (20)	20/20					
KAR (20)		20/20				
[0080] GXH (20)			20/20			
HNH (20)			20/20			
EAM (20)				20/20		
NAS (20)					20/20	
UY (20)						17/20

[0081] B. 测试个体的人群来源归类

[0082] 从上述140个测试样本中,每个测试人群随机抽取1个样本基于所构建的参考数据库进行个体主成分分析,推断其人群来源,如图4所示,7个样本的祖先成分计算统计见表5。从个体主成分分析结果(图4)可以看出7个已知个体样本均能落到相对应的祖先人群中,个体祖先成分计算中(表5)我们可以看出7个测试样本的祖先主成分都达到了94%以上。

[0083] 表5七个测试样本祖先来源分析结果

世界地区 测试样本	非洲	美洲	东亚	欧洲	混合人群	大洋洲
非洲(A)	97.80%	0.40%	0.50%	0.40%	0.60%	0.40%
美洲(B)	0.10%	99.30%	0.20%	0.10%	0.20%	0.10%
[0084] 东亚(C)	0.50%	0.30%	97.30%	0.20%	0.80%	0.90%
东亚(D)	0.20%	0.80%	98.20%	0.10%	0.30%	0.30%
欧洲(E)	0.20%	0.30%	0.20%	98.50%	0.50%	0.30%
维吾尔人(F)	0.50%	0.50%	1.10%	0.80%	96.90%	0.30%
大洋洲(G)	0.20%	0.40%	4.90%	0.20%	0.20%	94.20%

[0085] 综合本实施例的结果,可见:本发明所提供的SNP复合体系可以有效进行东亚、欧

洲、非洲、美洲和大洋洲五大人群和混合人群的遗传结构分析和个体的祖先来源推断。考虑到当今社会人口流动频繁,流动范围广,该体系采用较少的位点实现了针对五大洲际人群的区分,相比我们之前建立的针对欧、亚、非三大洲际人群的27-Plex SNP覆盖范围更广,两体系在法医DNA检验中针对未知样本的祖先来源推断可以相互印证,为案件提供更多更准确的侦查线索。

[0086] 实施例2、本发明与现有技术的比较

[0087] 本实施例以标准欧洲人的DNA样本(9947)和东亚人的DNA样本(HWQ,发明人所在实验室的亚洲人)来进一步验证本发明28体系与31个AIMs(de la Puente M,Santos C,Fondevila M,et al.The Global AIMs Nano set:A 31-plex SNaPshot assay of ancestry-informative SNPs[J].Forensic Science International:Genetics,2016,22:81-88.)两者的优劣。

[0088] 本发明28体系组:按照实施例1中的相关步骤进行操作。

[0089] 文章31个AIMs组:按照“de la Puente M,Santos C,Fondevila M,et al.The Global AIMs Nano set:A 31-plex SNaPshot assay of ancestry-informative SNPs [J].Forensic Science International:Genetics,2016,22:81-88.”文中记载进行操作,即扩增引物及延伸引物按照文章中浓度进行配比(除去三个三等位基因)。

[0090] 结果如表6和图5所示。对于“9947”这一标准欧洲人的DNA样本来说,采用本发明28体系和文章体系鉴定结果一致;而对于“HWQ”这一东亚人的DNA样本,采用本发明28体系和文章体系鉴定结果出现两处不一致(详见表6中粗体下划线部分),进一步通过对不一致位点进行单扩和测序,结果证实本发明28体系的检测结果是正确的。可见,本发明28体系,相比于文章的31个AIMs来说,在减少了SNP位点数量的情况下,还进一步提高了检测结果的准确性,本实施例中特别体现为对于东亚样本的检测结果更加准确。

[0091] 由图5可见,扩增引物及延伸引物按照文章中浓度进行配比(除去三个三等位基因),主要出现以下问题:文章31个AIMs组存在位点出峰不全、位点出峰不平衡(杂合子峰高差别较大),以及某些位点之间出峰重叠,对判型有影响。本发明28体系有效的解决了以上问题:1、针对“位点出峰不全”的解决方法:调整引物之间的配比。2、针对“位点出峰不平衡(杂合子峰高差别较大)”的解决方法:保证出峰特异的情况下降低退火温度。3、针对“某些位点之间出峰重叠,对判型有影响”的解决方法:调整延伸引物长度。本发明28体系通过以上调整,最终达到了准确地分型。

[0092] 表6本发明和文章两检测体系对“9947”和“HWQ”的检测结果

编号		本发明调整体系		文章中体系		
		9947	HWQ	9947	HWQ	
	20-M1	rs9809818	AA	<u>AC</u>	AA	<u>CC</u>
	26-K1	rs12594144	GG	TT	GG	TT
	10-R1	rs3751050	AA	AA	AA	AA
	16-Y1	rs6437783	CT	CC	CT	CC
	22-K2	rs10483251	GG	GG	GG	GG
[0093]	6-M2	rs2080161	AA	AC	AA	AC
	24-R2	rs12402499	AG	GG	AG	GG
	8-Y2	rs2789823	TT	TT	TT	TT
	19-Y3	rs9522149	CC	TT	CC	TT
	28-R3	rs17822931	GG	AA	GG	AA
	2-Y4	rs1229984	CC	CT	CC	CT
	23-R4	rs12142199	AA	GG	AA	GG
	25-Y5	rs12498138	CC	CC	CC	CC
	13-R5	rs4749305	AG	AG	AG	AG
	3-Y6	rs1426654	TT	CC	TT	CC
	27-S1	rs16891982	CC	GG	CC	GG
	11-R6	rs3827760	AA	GG	AA	GG
	9-Y7	rs2814778	TT	TT	TT	TT
	4-Y8	rs1557553	CC	CC	CC	CC
	12-R7	rs4657449	GG	AA	GG	AA
[0094]	17-S2	rs8072587	CG	CC	CG	CC
	14-R8	rs4792928	AA	AG	AA	AG
	18-Y9	rs8137373	CT	CC	CT	CC
	5-S3	rs1871534	CC	CC	CC	CC
	21-R9	rs9908046	GG	GG	GG	GG
	7-Y10	rs2139931	TT	CT	TT	CT
	1-R10	rs715605	AA	AA	AA	AA
	15-Y11	rs6054465	TT	<u>CT</u>	TT	<u>CC</u>

- [0001] <110> 公安部物证鉴定中心
- [0002] <120> 对未知来源个体进行五大洲际族群来源推断的方法和系统
- [0003] <130> GNCLN171048
- [0004] <160> 84
- [0005] <170> PatentIn version 3.5
- [0006] <210> 1
- [0007] <211> 22
- [0008] <212> DNA
- [0009] <213> 人工序列
- [0010] <220>
- [0011] <223>
- [0012] <400> 1
- [0013] gcacgttctt aaccttggt at 22
- [0014] <210> 2
- [0015] <211> 21
- [0016] <212> DNA
- [0017] <213> 人工序列
- [0018] <220>
- [0019] <223>
- [0020] <400> 2
- [0021] ttctgaatat cccaccaca a 21
- [0022] <210> 3
- [0023] <211> 48
- [0024] <212> DNA
- [0025] <213> 人工序列
- [0026] <220>
- [0027] <223>
- [0028] <400> 3
- [0029] gtgccacgtc gtgaaagtct gacaaggaaa aagttatgtg accagatt 48
- [0030] <210> 4
- [0031] <211> 20
- [0032] <212> DNA
- [0033] <213> 人工序列
- [0034] <220>
- [0035] <223>
- [0036] <400> 4
- [0037] aggccttgat gtgcttgaac 20
- [0038] <210> 5

- [0039] <211> 20
[0040] <212> DNA
[0041] <213> 人工序列
[0042] <220>
[0043] <223>
[0044] <400> 5
[0045] cgagaaggcc aaccactact 20
[0046] <210> 6
[0047] <211> 63
[0048] <212> DNA
[0049] <213> 人工序列
[0050] <220>
[0051] <223>
[0052] <400> 6
[0053] aacaactgac taaactaggt gccacgtcgt gaaagtctga caatcaaaca tgttcctctg 60
[0054] cac 63
[0055] <210> 7
[0056] <211> 24
[0057] <212> DNA
[0058] <213> 人工序列
[0059] <220>
[0060] <223>
[0061] <400> 7
[0062] attctgtaga tgggtggctgt agga 24
[0063] <210> 8
[0064] <211> 21
[0065] <212> DNA
[0066] <213> 人工序列
[0067] <220>
[0068] <223>
[0069] <400> 8
[0070] ctgcctcatg gcctaaaatc a 21
[0071] <210> 9
[0072] <211> 59
[0073] <212> DNA
[0074] <213> 人工序列
[0075] <220>
[0076] <223>
[0077] <400> 9

[0078] caactgacta aactaggtgc cacgtcgtga aagtctgaca aaccacgtgg tcactctgtg 59
[0079] <210> 10
[0080] <211> 20
[0081] <212> DNA
[0082] <213> 人工序列
[0083] <220>
[0084] <223>
[0085] <400> 10
[0086] tgaagggtat tactagtggc 20
[0087] <210> 11
[0088] <211> 20
[0089] <212> DNA
[0090] <213> 人工序列
[0091] <220>
[0092] <223>
[0093] <400> 11
[0094] ttgacagact tctgcttttg 20
[0095] <210> 12
[0096] <211> 51
[0097] <212> DNA
[0098] <213> 人工序列
[0099] <220>
[0100] <223>
[0101] <400> 12
[0102] aggtgccacg tcgtgaaagt ctgacaactg cttttgattt caagtatcag t 51
[0103] <210> 13
[0104] <211> 20
[0105] <212> DNA
[0106] <213> 人工序列
[0107] <220>
[0108] <223>
[0109] <400> 13
[0110] tcttcttcag ggaatcctgt 20
[0111] <210> 14
[0112] <211> 21
[0113] <212> DNA
[0114] <213> 人工序列
[0115] <220>
[0116] <223>

- [0117] <400> 14
[0118] gagttacata ggatttgcca g 21
[0119] <210> 15
[0120] <211> 63
[0121] <212> DNA
[0122] <213> 人工序列
[0123] <220>
[0124] <223>
[0125] <400> 15
[0126] caactgacta aactaggtgc cacgtcgtga aagtctgaca agggaatcct gttattcaca 60
[0127] tta 63
[0128] <210> 16
[0129] <211> 20
[0130] <212> DNA
[0131] <213> 人工序列
[0132] <220>
[0133] <223>
[0134] <400> 16
[0135] cctacaagac cacccaccag 20
[0136] <210> 17
[0137] <211> 20
[0138] <212> DNA
[0139] <213> 人工序列
[0140] <220>
[0141] <223>
[0142] <400> 17
[0143] ggacccatgg tcattccata 20
[0144] <210> 18
[0145] <211> 40
[0146] <212> DNA
[0147] <213> 人工序列
[0148] <220>
[0149] <223>
[0150] <400> 18
[0151] cacgtcgtga aagtctgaca agctcccacc ctgaaaaaga 40
[0152] <210> 19
[0153] <211> 20
[0154] <212> DNA
[0155] <213> 人工序列

- [0156] <220>
[0157] <223>
[0158] <400> 19
[0159] aattcaggag ctgaactgcc 20
[0160] <210> 20
[0161] <211> 20
[0162] <212> DNA
[0163] <213> 人工序列
[0164] <220>
[0165] <223>
[0166] <400> 20
[0167] tgttcagccc ttggattgtc 20
[0168] <210> 21
[0169] <211> 67
[0170] <212> DNA
[0171] <213> 人工序列
[0172] <220>
[0173] <223>
[0174] <400> 21
[0175] ctctctctct ctctctctct ctctctctct ctctctctct ctctctcttt cgctgccatg 60
[0176] aaagttg 67
[0177] <210> 22
[0178] <211> 20
[0179] <212> DNA
[0180] <213> 人工序列
[0181] <220>
[0182] <223>
[0183] <400> 22
[0184] taatacaaga gccgcctgga 20
[0185] <210> 23
[0186] <211> 21
[0187] <212> DNA
[0188] <213> 人工序列
[0189] <220>
[0190] <223>
[0191] <400> 23
[0192] cttgcaagga actgcagcta t 21
[0193] <210> 24
[0194] <211> 93

- [0195] <212> DNA
[0196] <213> 人工序列
[0197] <220>
[0198] <223>
[0199] <400> 24
[0200] taaactaggt gccacgtcgt gaaagtctga caacaactga ctaaactagg tgccacgtcg 60
[0201] tgaaagtctg acaaccctaaa gccctggaa aaa 93
[0202] <210> 25
[0203] <211> 25
[0204] <212> DNA
[0205] <213> 人工序列
[0206] <220>
[0207] <223>
[0208] <400> 25
[0209] gaataaagtg aggaaaacac ggagt 25
[0210] <210> 26
[0211] <211> 25
[0212] <212> DNA
[0213] <213> 人工序列
[0214] <220>
[0215] <223>
[0216] <400> 26
[0217] gtttctcatc tacgaaagag gagtc 25
[0218] <210> 27
[0219] <211> 80
[0220] <212> DNA
[0221] <213> 人工序列
[0222] <220>
[0223] <223>
[0224] <400> 27
[0225] cacgtcgtga aagtctgaca acaactgact aaactaggtg ccacgtcgtg aaagtctgac 60
[0226] aaggttgat gttggggctt 80
[0227] <210> 28
[0228] <211> 20
[0229] <212> DNA
[0230] <213> 人工序列
[0231] <220>
[0232] <223>
[0233] <400> 28

- [0234] cctagagtcc cccaaacctc 20
[0235] <210> 29
[0236] <211> 20
[0237] <212> DNA
[0238] <213> 人工序列
[0239] <220>
[0240] <223>
[0241] <400> 29
[0242] cacttctggg catctgcttc 20
[0243] <210> 30
[0244] <211> 59
[0245] <212> DNA
[0246] <213> 人工序列
[0247] <220>
[0248] <223>
[0249] <400> 30
[0250] aactgactaa actaggtgcc acgtcgtgaa agtctgacaa ctgcattgcc agtgtactc 59
[0251] <210> 31
[0252] <211> 20
[0253] <212> DNA
[0254] <213> 人工序列
[0255] <220>
[0256] <223>
[0257] <400> 31
[0258] acatcctgca gaccttcctg 20
[0259] <210> 32
[0260] <211> 19
[0261] <212> DNA
[0262] <213> 人工序列
[0263] <220>
[0264] <223>
[0265] <400> 32
[0266] cagaccttgg gcgtcagat 19
[0267] <210> 33
[0268] <211> 105
[0269] <212> DNA
[0270] <213> 人工序列
[0271] <220>
[0272] <223>

- [0273] <400> 33
[0274] ctgacaaca ctgactaac taggtgccac gtcgtgaaag tctgacaaca actgactaaa 60
[0275] ctaggtgcca cgctcgtgaaa gtctgacaac ctggcagtggt gtgca 105
[0276] <210> 34
[0277] <211> 27
[0278] <212> DNA
[0279] <213> 人工序列
[0280] <220>
[0281] <223>
[0282] <400> 34
[0283] gagtatgata taattttggtt cctgctg 27
[0284] <210> 35
[0285] <211> 23
[0286] <212> DNA
[0287] <213> 人工序列
[0288] <220>
[0289] <223>
[0290] <400> 35
[0291] tggactttat gggttggtt ttt 23
[0292] <210> 36
[0293] <211> 48
[0294] <212> DNA
[0295] <213> 人工序列
[0296] <220>
[0297] <223>
[0298] <400> 36
[0299] cacgtcgtga aagtctgaca attttttggtt ttttttttgc actcatca 48
[0300] <210> 37
[0301] <211> 22
[0302] <212> DNA
[0303] <213> 人工序列
[0304] <220>
[0305] <223>
[0306] <400> 37
[0307] agtccttggt agggcggttag ta 22
[0308] <210> 38
[0309] <211> 22
[0310] <212> DNA
[0311] <213> 人工序列

- [0312] <220>
[0313] <223>
[0314] <400> 38
[0315] ctcctagtca tggttgatgt gg 22
[0316] <210> 39
[0317] <211> 109
[0318] <212> DNA
[0319] <213> 人工序列
[0320] <220>
[0321] <223>
[0322] <400> 39
[0323] acaacaactg actaaactag gtgccacgtc gtgaaagtct gacaacaact gactaaacta 60
[0324] ggtgccacgt cgtgaaagtc tgacaattcg tggttgatgag aaaatttca 109
[0325] <210> 40
[0326] <211> 20
[0327] <212> DNA
[0328] <213> 人工序列
[0329] <220>
[0330] <223>
[0331] <400> 40
[0332] agagggcttc tgttcacacc 20
[0333] <210> 41
[0334] <211> 20
[0335] <212> DNA
[0336] <213> 人工序列
[0337] <220>
[0338] <223>
[0339] <400> 41
[0340] atgcaccact actgtccaag 20
[0341] <210> 42
[0342] <211> 51
[0343] <212> DNA
[0344] <213> 人工序列
[0345] <220>
[0346] <223>
[0347] <400> 42
[0348] aaactaggtg ccacgtcgtg aaagtctgac aaggaggtga gcttcacggg g 51
[0349] <210> 43
[0350] <211> 21

- [0351] <212> DNA
[0352] <213> 人工序列
[0353] <220>
[0354] <223>
[0355] <400> 43
[0356] aacctgatgg ccctcattag t 21
[0357] <210> 44
[0358] <211> 19
[0359] <212> DNA
[0360] <213> 人工序列
[0361] <220>
[0362] <223>
[0363] <400> 44
[0364] atggcaccgt ttggttcag 19
[0365] <210> 45
[0366] <211> 89
[0367] <212> DNA
[0368] <213> 人工序列
[0369] <220>
[0370] <223>
[0371] <400> 45
[0372] agtctgacaa ctaggtgcca cgctcgtgaaa gtctgacaac taggtgccac gtcgtgaaag 60
[0373] tctgacatct cattagtcct tggctctta 89
[0374] <210> 46
[0375] <211> 20
[0376] <212> DNA
[0377] <213> 人工序列
[0378] <220>
[0379] <223>
[0380] <400> 46
[0381] gaaggctccc aactcgttag 20
[0382] <210> 47
[0383] <211> 21
[0384] <212> DNA
[0385] <213> 人工序列
[0386] <220>
[0387] <223>
[0388] <400> 47
[0389] gtcattaaag tcaacctagg c 21

- [0390] <210> 48
[0391] <211> 44
[0392] <212> DNA
[0393] <213> 人工序列
[0394] <220>
[0395] <223>
[0396] <400> 48
[0397] ctctctctct ctctctctct cttgttttagg agagttgaga catc 44
[0398] <210> 49
[0399] <211> 20
[0400] <212> DNA
[0401] <213> 人工序列
[0402] <220>
[0403] <223>
[0404] <400> 49
[0405] tgctcagctc cacgtacaac 20
[0406] <210> 50
[0407] <211> 19
[0408] <212> DNA
[0409] <213> 人工序列
[0410] <220>
[0411] <223>
[0412] <400> 50
[0413] ctcttcaggc cgaagctct 19
[0414] <210> 51
[0415] <211> 89
[0416] <212> DNA
[0417] <213> 人工序列
[0418] <220>
[0419] <223>
[0420] <400> 51
[0421] actaggtgcc acgtcgtgaa agtctgacaa caactgacta aactaggtgc cacgtcgtga 60
[0422] aagtctgaca atggcgccac gttttcaca 89
[0423] <210> 52
[0424] <211> 20
[0425] <212> DNA
[0426] <213> 人工序列
[0427] <220>
[0428] <223>

- [0429] <400> 52
[0430] cccctcggga gaaaacatag 20
[0431] <210> 53
[0432] <211> 24
[0433] <212> DNA
[0434] <213> 人工序列
[0435] <220>
[0436] <223>
[0437] <400> 53
[0438] ttctagagtt gaatgagggt caga 24
[0439] <210> 54
[0440] <211> 93
[0441] <212> DNA
[0442] <213> 人工序列
[0443] <220>
[0444] <223>
[0445] <400> 54
[0446] aaactaggtg ccacgtcgtg aaagtctgac aacaactgac taaactaggt gccacgtcgt 60
[0447] gaaagtctga caagagctaa ggaaagatac gtg 93
[0448] <210> 55
[0449] <211> 20
[0450] <212> DNA
[0451] <213> 人工序列
[0452] <220>
[0453] <223>
[0454] <400> 55
[0455] cagcccaacc tactcctctg 20
[0456] <210> 56
[0457] <211> 20
[0458] <212> DNA
[0459] <213> 人工序列
[0460] <220>
[0461] <223>
[0462] <400> 56
[0463] tcctacaaa gtggcaaacc 20
[0464] <210> 57
[0465] <211> 67
[0466] <212> DNA
[0467] <213> 人工序列

- [0468] <220>
[0469] <223>
[0470] <400> 57
[0471] caacaactga ctaaactagg tgccacgtcg tgaaagtctg acaacagtaa atagtaactc 60
[0472] catcttc 67
[0473] <210> 58
[0474] <211> 21
[0475] <212> DNA
[0476] <213> 人工序列
[0477] <220>
[0478] <223>
[0479] <400> 58
[0480] tctctcagga tatecctttg g 21
[0481] <210> 59
[0482] <211> 25
[0483] <212> DNA
[0484] <213> 人工序列
[0485] <220>
[0486] <223>
[0487] <400> 59
[0488] aaaatcttga ttctgtatcg cagtc 25
[0489] <210> 60
[0490] <211> 101
[0491] <212> DNA
[0492] <213> 人工序列
[0493] <220>
[0494] <223>
[0495] <400> 60
[0496] caactgacta aactaggtgc cacgtcgtga aagtctgaca acaactgact aaactaggtg 60
[0497] ccacgtcgtg aaagtctgac aacgcagtct actagttgtc c 101
[0498] <210> 61
[0499] <211> 20
[0500] <212> DNA
[0501] <213> 人工序列
[0502] <220>
[0503] <223>
[0504] <400> 61
[0505] tatggcctca ggttctccac 20
[0506] <210> 62

- [0507] <211> 21
[0508] <212> DNA
[0509] <213> 人工序列
[0510] <220>
[0511] <223>
[0512] <400> 62
[0513] cacatgatct caccgtttcc t 21
[0514] <210> 63
[0515] <211> 113
[0516] <212> DNA
[0517] <213> 人工序列
[0518] <220>
[0519] <223>
[0520] <400> 63
[0521] tgacaacaac tgactaaact aggtgccacg tcgtgaaagt ctgacaaca ctgactaac 60
[0522] taggtgccac gtcgtgaaag tctgacaaca catgcaaat caggataata atg 113
[0523] <210> 64
[0524] <211> 22
[0525] <212> DNA
[0526] <213> 人工序列
[0527] <220>
[0528] <223>
[0529] <400> 64
[0530] gcaatgagat tagttgcact gg 22
[0531] <210> 65
[0532] <211> 20
[0533] <212> DNA
[0534] <213> 人工序列
[0535] <220>
[0536] <223>
[0537] <400> 65
[0538] attatatgcc caccctgctc 20
[0539] <210> 66
[0540] <211> 44
[0541] <212> DNA
[0542] <213> 人工序列
[0543] <220>
[0544] <223>
[0545] <400> 66

- [0546] tgccacgtcg tgaaagtctg acaactgggtt gaggcacact atta 44
[0547] <210> 67
[0548] <211> 20
[0549] <212> DNA
[0550] <213> 人工序列
[0551] <220>
[0552] <223>
[0553] <400> 67
[0554] cccagctagg gctagacacc 20
[0555] <210> 68
[0556] <211> 20
[0557] <212> DNA
[0558] <213> 人工序列
[0559] <220>
[0560] <223>
[0561] <400> 68
[0562] tcaaagactg agccatgcac 20
[0563] <210> 69
[0564] <211> 113
[0565] <212> DNA
[0566] <213> 人工序列
[0567] <220>
[0568] <223>
[0569] <400> 69
[0570] gaaagtctga caacaactga ctaaactagg tgccacgtcg tgaaagtctg acaacaactg 60
[0571] actaaactag gtgccacgtc gtgaaagtct gacaaccacc ctaaggggac aga 113
[0572] <210> 70
[0573] <211> 20
[0574] <212> DNA
[0575] <213> 人工序列
[0576] <220>
[0577] <223>
[0578] <400> 70
[0579] tggcaacctc acatggtaga 20
[0580] <210> 71
[0581] <211> 20
[0582] <212> DNA
[0583] <213> 人工序列
[0584] <220>

- [0585] <223>
[0586] <400> 71
[0587] ccaggggagg tagaaagagg 20
[0588] <210> 72
[0589] <211> 97
[0590] <212> DNA
[0591] <213> 人工序列
[0592] <220>
[0593] <223>
[0594] <400> 72
[0595] aactgactaa actaggtgcc acgtcgtgaa agtctgacaa caactgacta aactaggtgc 60
[0596] cacgtcgtga aagtctgaca acagtctcct gcccggc 97
[0597] <210> 73
[0598] <211> 20
[0599] <212> DNA
[0600] <213> 人工序列
[0601] <220>
[0602] <223>
[0603] <400> 73
[0604] ccagagcttt gcagcacttt 20
[0605] <210> 74
[0606] <211> 19
[0607] <212> DNA
[0608] <213> 人工序列
[0609] <220>
[0610] <223>
[0611] <400> 74
[0612] caaggacgca gctctctca 19
[0613] <210> 75
[0614] <211> 101
[0615] <212> DNA
[0616] <213> 人工序列
[0617] <220>
[0618] <223>
[0619] <400> 75
[0620] acaactgact aaactaggtg ccacgtcgtg aaagtctgac aacaactgac taaactaggt 60
[0621] gccacgtcgt gaaagtctga caagagtgtt ttgtgggcct c 101
[0622] <210> 76
[0623] <211> 20

- [0624] <212> DNA
[0625] <213> 人工序列
[0626] <220>
[0627] <223>
[0628] <400> 76
[0629] agaaaggaga ggaaacaccg 20
[0630] <210> 77
[0631] <211> 20
[0632] <212> DNA
[0633] <213> 人工序列
[0634] <220>
[0635] <223>
[0636] <400> 77
[0637] tcagcaactt ctagtctctcg 20
[0638] <210> 78
[0639] <211> 55
[0640] <212> DNA
[0641] <213> 人工序列
[0642] <220>
[0643] <223>
[0644] <400> 78
[0645] ctctctctct ctctctctct ctctctctct gacaatctga ggtccttgca gctcc 55
[0646] <210> 79
[0647] <211> 20
[0648] <212> DNA
[0649] <213> 人工序列
[0650] <220>
[0651] <223>
[0652] <400> 79
[0653] tgtgtggttt tctcagcgac 20
[0654] <210> 80
[0655] <211> 20
[0656] <212> DNA
[0657] <213> 人工序列
[0658] <220>
[0659] <223>
[0660] <400> 80
[0661] agcatggtat gagcactgag 20
[0662] <210> 81

- [0663] <211> 40
[0664] <212> DNA
[0665] <213> 人工序列
[0666] <220>
[0667] <223>
[0668] <400> 81
[0669] tctctctctc tctctctctc tcctcctaataagagctggc 40
[0670] <210> 82
[0671] <211> 20
[0672] <212> DNA
[0673] <213> 人工序列
[0674] <220>
[0675] <223>
[0676] <400> 82
[0677] ccttgcatg ttctctctc 20
[0678] <210> 83
[0679] <211> 24
[0680] <212> DNA
[0681] <213> 人工序列
[0682] <220>
[0683] <223>
[0684] <400> 83
[0685] tcagaggaat tagaaaggcc taaa 24
[0686] <210> 84
[0687] <211> 109
[0688] <212> DNA
[0689] <213> 人工序列
[0690] <220>
[0691] <223>
[0692] <400> 84
[0693] agtctgacaa caactgacta aactaggtgc cacgtcgtga aagtctgaca acaactgact 60
[0694] aaactaggtg ccacgtcgtg aaagtctgac aaggaggtag gagcaccca 109

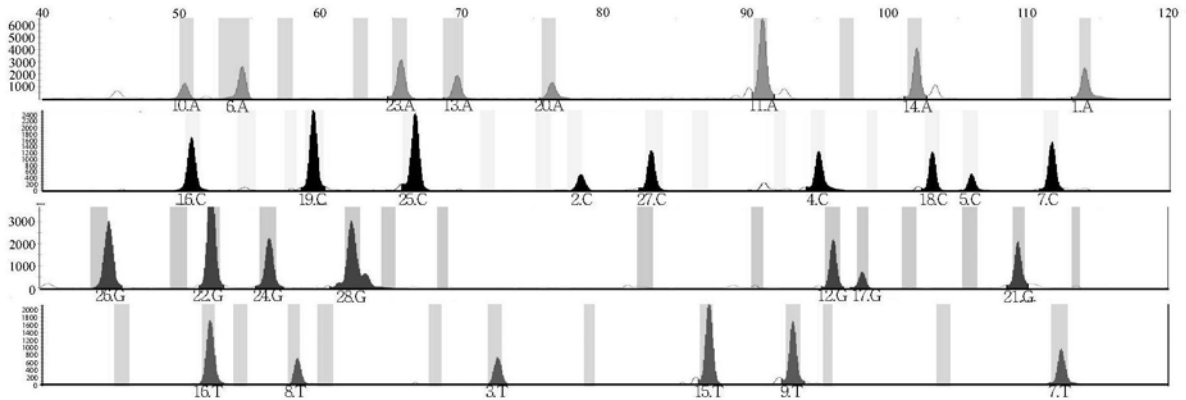


图1



图2

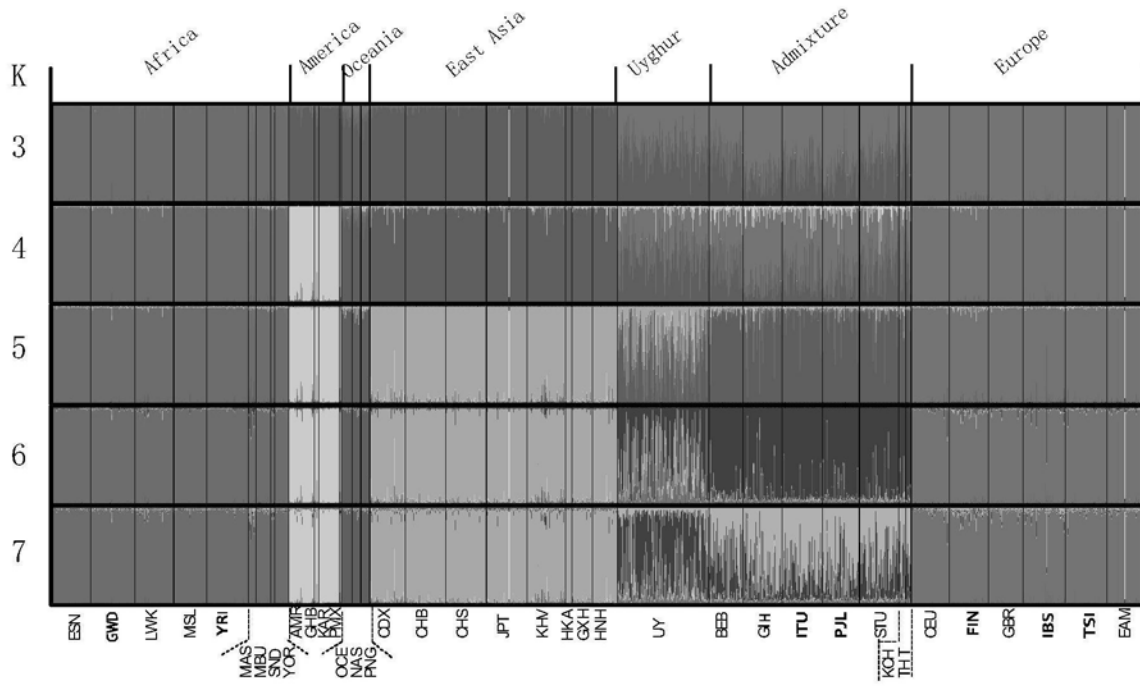


图3

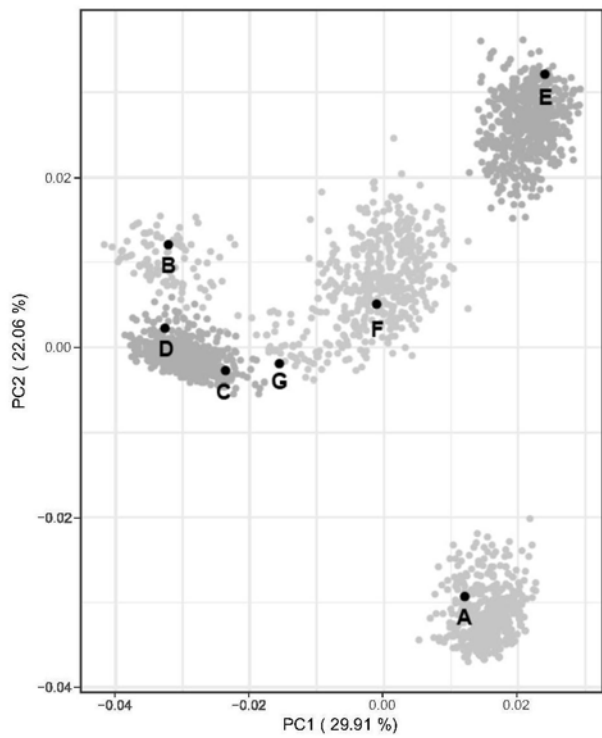
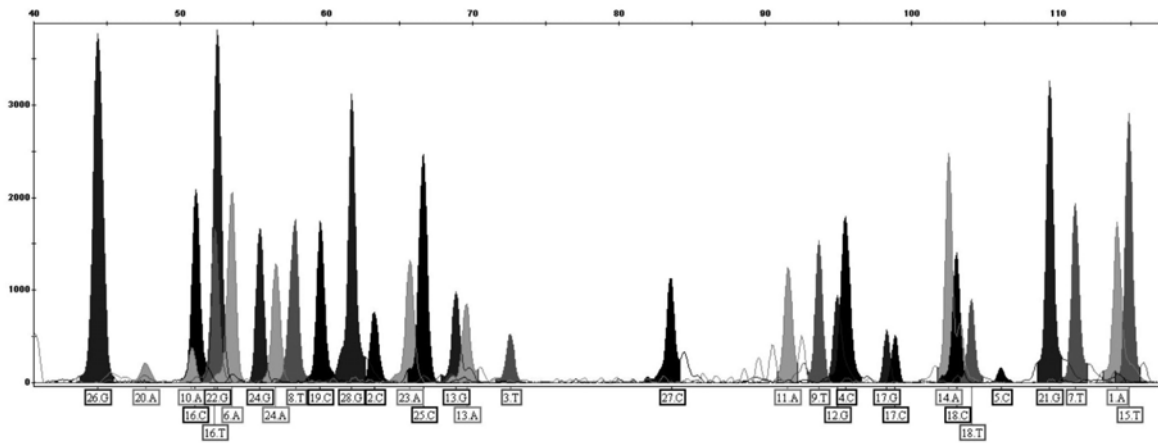
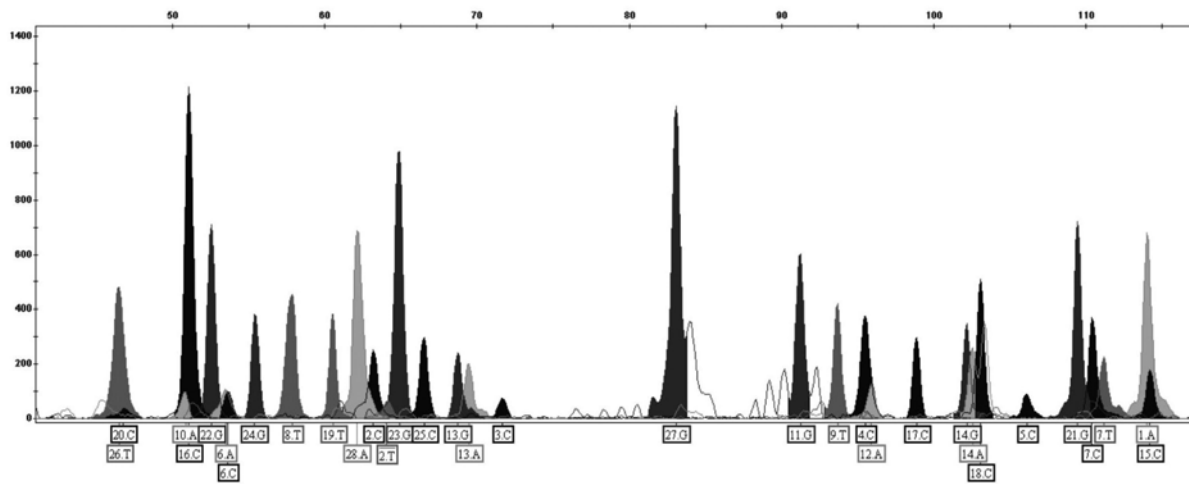


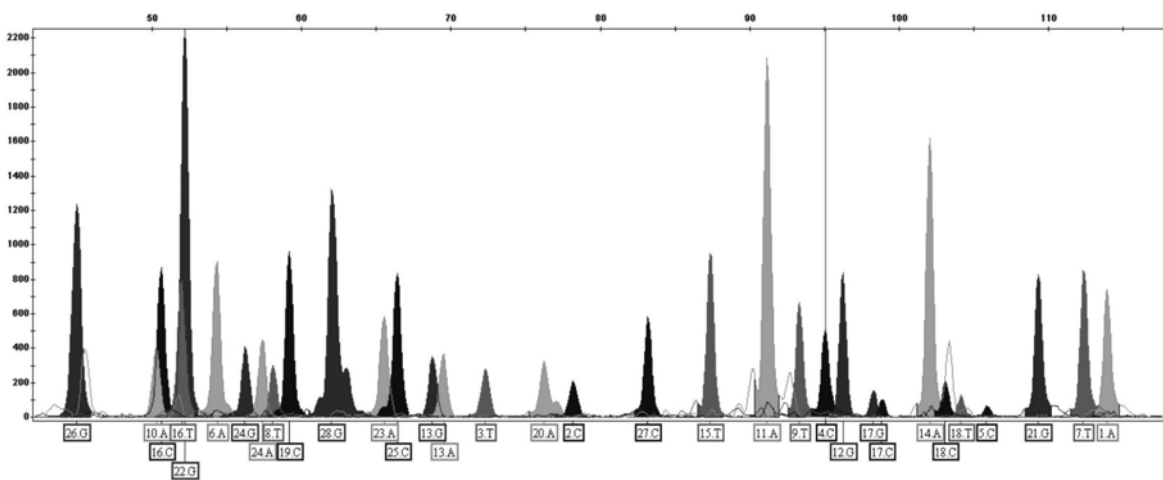
图4



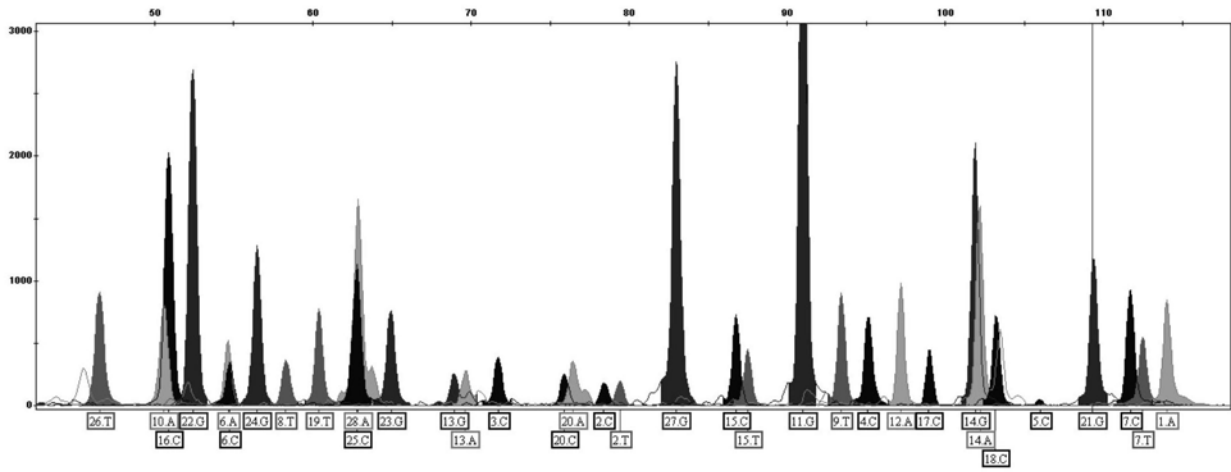
A



B



C



D

图5