

粗皮桉近红外光谱差异与其遗传差异间的关系

王楚彪^{1,2}, 杨艳³, 白卫国⁴, 林彦¹, 谢耀坚¹, 卢万鸿^{1*}, 罗建中¹

1. 国家林业和草原局桉树研究开发中心遗传育种研究室, 广东 湛江 524022
2. 南京林业大学林学院, 江苏 南京 210037
3. 湛江科技学院经济与金融学院, 广东 湛江 524094
4. 广西国有东门林场, 广西 崇左 532199

摘要 摸清粗皮桉(*Eucalyptus pellita*)群体的遗传亲缘关系,对研究桉树杂交育种理论,开发优良新品种具有重要的意义。研究意在通过对比粗皮桉种源遗传差异与其光谱差异间的关系,探索近红外光谱(NIRs)技术用于粗皮桉遗传亲缘关系分析的可行性。以粗皮桉天然种源材料为对象,每个种源采集8~12个家系叶样。通过全基因组重测序,基于核苷酸序列差异计算种源间的遗传距离。同时,每个家系选择4~6片健康叶片烘至绝干后,将其粉碎装于透明自封口塑料袋。用手持式近红外仪Phazir Rx(1624)采集样品的NIRs信息。以簇类独立软模式(SIMCA)判别分析统计对比种源到目标种源的光谱距离,并基于NIRs欧氏距离对种源进行层级聚类。以NIRs的PCA因子得分图分析种源间的遗传亲缘关系及其遗传变异。结果显示,粗皮桉新几内亚岛种源间的平均遗传距离为0.186,昆士兰州种源间的平均为0.157,新几内亚岛种源与昆士兰州种源间的平均遗传距离为0.295,明显大于区域内种源间的遗传距离。粗皮桉2大区域种源间的NIRs光谱距离与其种源间遗传距离基本呈正相关关系。基于NIRs的层级聚类在一定程度上印证了种源遗传距离、光谱距离的大小关系,但与其地理距离非完全对应,说明基因交流对粗皮桉群体的遗传亲缘关系有较大的影响。PCA聚类显示,遗传或光谱距离大的种源样本因子得分图存在严重重叠,而遗传或光谱距离小的种源样本因子得分反而会清晰聚类,这表明NIRs信息区分异质样本的敏感性,同时也反映了粗皮桉种源内家系间遗传变异的大小。研究结果表明,NIRs技术能够真实反映粗皮桉种源间的遗传差异,可用于桉树群体遗传亲缘关系及群体内的遗传变异分析,可辅助桉树群体的世代改良研究。

关键词 遗传距离;光谱距离;层级聚类;簇类独立软模式(SIMCA)

中图分类号: S722.34 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2021)11-3399-06

引言

粗皮桉天然分布于新几内亚岛南部和澳大利亚昆士兰州北部,其不同种源的生物学特征和生长适应性均具有很大的遗传差异^[1]。在炎热潮湿的热带地区,粗皮桉良好的抗病虫害能力使其逐渐取代金合欢属(*Acacia*)树种,成为地区重要的造林树种^[2]。巴西、东南亚等国对粗皮桉进行了大量的引种和高世代改良^[3]。国内系统的粗皮桉引种改良始于20世纪80年代,并于2011年在两广多地建立了其高世代育种群体^[4]。粗皮桉在国内主要作为杂交亲本用于开发桉树优良无性系,并针对其开展选育、抗病虫害等杂交育种理论研究。

摸清育种群体的遗传亲缘关系是推动粗皮桉群体世代改良的基础工作。目前常用的表型数据分析方法很难精确估算家系的遗传和环境分量。DNA生物信息学分析专业要求高,程序复杂且成本高。鉴于此,本研究计划通过对比基于DNA分析的遗传距离和基于NIRs分析的光谱距离间的关系,探索近红外光谱技术用于粗皮桉群体遗传基础分析的精度和可行性。

1 实验部分

1.1 材料

研究对象为粗皮桉育种群体试验中的23个天然种源

收稿日期: 2020-11-06, 修订日期: 2021-03-12

基金项目: 国家自然科学基金项目(31700599, 31670680)和广东省林业科技创新项目(2017KJCX031, 2019KJCX014)资助

作者简介: 王楚彪, 1982年生, 国家林业和草原局桉树研究开发中心博士研究生 e-mail: scauweb@163.com

* 通讯作者 e-mail: luwanhong@outlook.com

(表 1)。遗传材料基本涵盖了粗皮桉的天然原生地,即澳大利亚昆士兰州(QLD, AUS)北部和新几内亚岛(New Guinea Island)南部(印度尼西亚(Indonesia)和巴布亚新几内亚(Papua New Guinea, PNG))。两地被大洋分隔,最近的海岸线

距离约为 150 km。在粗皮桉大田群体试验中,每家系采集 10~15 片新鲜叶样,每个种源采样 8~12 个家系代表该种源。叶样尽快置于-30 °C 冰箱,用于后续遗传分析及 NIRs 信息采集。

表 1 粗皮桉 23 个天然种源信息

Table 1 Details of 23 natural provenances of *E. pellita* studied

种批	种源	经纬度	海拔/m	家系数
A17854	Bupul-Muting, Indonesia	140°36'E, 7°21'S	40	8
A20659	Meru WP, PNG	141°29'E, 8°27'S	40	8
A18197	S of Kiriwa WP, PNG	141°30'E, 8°25'S	45	10
A18199	Serisa Village WP, PNG	141°26'E, 8°36'S	45	10
S16120	6 km South Keru to Mata, PNG	141°45'E, 8°36'S	30	9
S16121	North Tokwa to Kiriwa, PNG	141°25'E, 8°30'S	45	11
S16122	Between Goe and Kiriwa, PNG	141°32'E, 8°20'S	50	10
S11947	Near Kuranda, QLD, AUS	145°33'E, 16°14'S	450	8
S14211	5~12 km S of Helenvale, QLD, AUS	145°15'E, 15°45'S	500	10
A18773	Kirrama Range, QLD, AUS	145°46'E, 18°12'S	560	9
A18774	Helenvale, QLD, AUS	145°14'E, 15°49'S	220	8
A18599	Julatten, QLD, AUS	145°28'E, 16°37'S	425	11
S14339	14.6 km NE of Coen, QLD, AUS	143°17'E, 13°53'S	580	12
A18314	El Arish Tullybringa, QLD, AUS	146°03'E, 17°50'S	50	8
A18597	9.25 km S Cardwell, QLD, AUS	146°08'E, 18°36'S	15	8
A17860	13 km SSW Kuranda, QLD, AUS	145°36'E, 16°56'S	425	8
A17861	NW Kuranda, QLD, AUS	145°32'E, 16°41'S	440	10
A18749	Daintree, QLD, AUS	145°21'E, 16°15'S	25	8
A18598	Cardwell-Kennedy, QLD, AUS	145°55'E, 18°15'S	30	9
A18596	5 km NW Ingham, QLD, AUS	146°08'E, 18°36'S	15	8
B10	SF 591 Abergowrie, QLD, AUS	145°57'E, 18°25'S	50	9
A18750	Wonga-Daintree, QLD, AUS	145°22'E, 16°16'S	15	12
A18772	N Kuranda, QLD, AUS	145°28'E, 16°38'S	420	8

1.2 光谱仪器

手持式近红外仪 Phazir Rx (1624) (Polychromix, Thermo Scientific, USA)用于叶样 NIRs 的采集。Phazir Rx 波长范围为 1 600~2 400 nm,分辨率 12 nm,单次扫描获得 100 维向量数据,自带背景校正,内置可编程的 MEMS 技术微衍射光栅。

1.3 方法

1.3.1 光谱采集

光谱采集样本与 DNA 分析样本完全对应,每个家系选择 4~6 片外形健康的叶片,置于 ClassicLine 干燥箱(BIND-ER, 德国),设置 60 °C, 72 h, 以至绝干。用中药打粉机粉碎后装入透明自封口塑料袋。用 Phazir Rx (1624)隔袋扫描 5 次,其均值代表该样品的近红外 NIRs 信息^[5],每个种源共获得 8~12 条 NIRs。

1.3.2 NIRs 数据的预处理和分析

经不同的预处理效果对比后,本研究对 NIRs 原始数据进行 Savitzky-Golay 平滑的二阶导数预处理^[6-7]。采用全交互内部交叉验证算法,建立目标种源的主成分分析(PCA)模型,通过簇类独立软模式(SIMCA)判别分析,统计对比种源到目标种源模型间的 NIRs 光谱距离。基于 NIRs 欧氏距离

(Euclidean distance)按全链接算法(Complete linkage),对测试种源进行层级聚类(Hierarchical clustering)。根据样本 NIRs 数据 PCA 的因子得分图分析种源间的遗传亲缘关系及其遗传变异。NIRs 数据的预处理和分析均由 The Unscrambler x10.4 (CAMO, Oslo, Norway)实现。

1.3.3 样本遗传距离计算

通过全基因组重测序(Whole Genome Resequencing)方法获得粗皮桉种源各家系的 DNA 信息,根据 Kimura 提出的核苷酸序列差异原理,使用 Mega-X 软件估算粗皮桉种源间的遗传距离^[8-9]。

2 结果与讨论

2.1 粗皮桉种源间的遗传距离与其 NIRs 光谱距离

粗皮桉新几内亚岛 7 个种源间的平均遗传距离为 0.186,昆士兰州 16 个种源间的平均为 0.157(数据冗余,未列出)。表 2 为新几内亚岛与昆士兰州 2 大区域种源间的遗传距离。数据显示,2 大区域种源间的最小、最大及平均遗传距离分别为 0.229, 0.370 和 0.295,明显大于区域内种源间的遗传距离。新几内亚岛种源 A20659, S16120, S16121

和 S16122 与昆士兰州各种源间的遗传距离均超过了 0.300。2 大区域内和区域间种源间遗传距离的差异,说明区域分隔(海洋)明显拉大了粗皮桉种源间的遗传亲缘关系。

利用 NIRs 光谱距离可以从不同角度解析、印证样本间的内在遗传差异。表 3 为粗皮桉新几内亚岛种源与昆士兰州种源间的 NIRs 光谱欧式距离。数据显示,2 大区域种源间 NIRs 平均光谱距离大小与其遗传距离的大小趋势基本一致,

如种源 S16120 与昆士兰州种源间的平均遗传距离最大(0.370),两者之间的平均光谱距离(0.220)也明显大于其他种源间的光谱距离。偶有例外,S16122 与昆士兰州种源间的平均光谱距离就与其遗传距离的大小关系相悖。种源 A17854 与昆士兰州种源间的 NIRs 光谱距离偏大可能与其有效样本偏少有关。

表 2 粗皮桉新几内亚岛种源与昆士兰州种源间的遗传距离

Table 2 The genetic distance of provenances between New Guinea Island and Queensland of *E. pellita*

种源	新几内亚岛种源						
	A20659	A17854	A18197	A18199	S16120	S16121	S16122
S11947	0.303	0.251	0.230	0.252	0.359	0.323	0.295
S14211	0.309	0.249	0.224	0.257	0.355	0.335	0.298
A18773	0.323	0.262	0.223	0.260	0.377	0.345	0.304
A18774	0.287	0.230	0.206	0.237	0.337	0.311	0.277
A18599	0.321	0.261	0.236	0.260	0.379	0.345	0.312
S14339	0.317	0.265	0.243	0.274	0.374	0.337	0.313
A18314	0.332	0.268	0.228	0.266	0.384	0.357	0.309
A18597	0.326	0.257	0.220	0.264	0.384	0.349	0.314
A17860	0.310	0.252	0.228	0.253	0.367	0.336	0.303
A17861	0.311	0.252	0.225	0.252	0.368	0.334	0.301
A18749	0.241	0.214	0.224	0.227	0.308	0.262	0.244
A18598	0.331	0.262	0.221	0.264	0.385	0.353	0.311
A18596	0.334	0.271	0.241	0.277	0.394	0.356	0.323
B10	0.321	0.262	0.235	0.267	0.380	0.342	0.306
A18750	0.323	0.263	0.237	0.262	0.380	0.347	0.314
A18772	0.322	0.262	0.238	0.262	0.383	0.349	0.312
均值	0.313	0.255	0.229	0.258	0.370	0.336	0.302
均值				0.295			

表 3 粗皮桉新几内亚岛种源与昆士兰州种源间的 NIRs 欧式距离

Table 3 NIRs Euclidean distance of provenances between New Guinea Island and Queensland of *E. pellita*

种源	新几内亚岛种源						
	A20659	A17854	A18197	A18199	S16120	S16121	S16122
S11947	0.129	0.390	0.161	0.153	0.100	0.099	0.100
S14211	0.163	0.468	0.183	0.174	0.135	0.117	0.116
A18773	0.134	0.236	0.158	0.160	0.236	0.221	0.215
A18774	0.148	0.183	0.136	0.136	0.239	0.192	0.205
A18599	0.129	0.198	0.136	0.134	0.225	0.185	0.185
S14339	0.174	0.390	0.168	0.172	0.178	0.121	0.143
A18314	0.107	0.160	0.075	0.082	0.236	0.115	0.148
A18597	0.203	0.223	0.135	0.135	0.272	0.189	0.220
A17860	0.165	0.114	0.082	0.076	0.231	0.124	0.167
A17861	0.191	0.216	0.123	0.126	0.277	0.167	0.220
A18749	0.156	0.159	0.099	0.090	0.251	0.127	0.161
A18598	0.142	0.150	0.105	0.099	0.229	0.145	0.183
A18596	0.153	0.185	0.107	0.099	0.263	0.160	0.201
B10	0.191	0.429	0.183	0.179	0.200	0.123	0.200
A18750	0.091	0.146	0.116	0.107	0.203	0.186	0.187
A18772	0.150	0.227	0.148	0.172	0.251	0.194	0.203
均值	0.152	0.242	0.132	0.131	0.220	0.154	0.178
均值				0.173			

因实际值太小, 表中所列光谱距离为实际数值放大 100 倍后的结果。

2.2 粗皮桉种源基于 NIRs 光谱距离的层级聚类

图 1 中粗皮桉天然种源的 NIRs 层级聚类显示, 23 个种源的聚类效果在一定程度上印证了其遗传距离、光谱距离的差异。巴布亚新几内亚岛的 S16120, S16121 和 S16122 与昆士兰州种源间的遗传距离和光谱距离均大于其他种源间的两

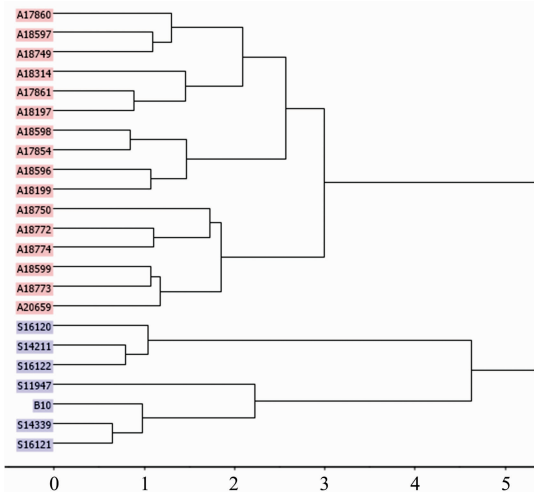


图 1 粗皮桉 23 个种源基于叶样 NIRs 欧氏距离的层级聚类

Fig. 1 Hierarchical clustering by Euclidean distance of foliage NIRs from 23 provenances on *E. pellita*

类距离, 也基本显示了与昆士兰州种源的分别聚类。但 23 个种源并没有完全按照巴布亚新几内亚岛和昆士兰州 2 大地域分别聚类, 来自巴布亚新几内亚的种源 A18197, A17854, A18199 及 A20659 与多数来自昆士兰州的种源聚在了一起。这或与几个种源的海拔较低(<50 m), 且两大地域间的地理距离比较接近, 地域间存在花粉、种子或传粉媒介等某种形式的基因交流有关。

2.3 粗皮桉巴布亚新几内亚岛种源与昆士兰州种源间的主成分分析

由于巴布亚新几内亚岛种源与昆士兰州种源间的遗传距离普遍较大, 因此, 主要列举了 2 大区域种源间的 PCA 主因子得分聚类结果。图 2 中巴布亚新几内亚岛内种源 S16120 与 A18199 间(a), 及 S16120 与昆士兰州种源 A18750 间(b)相同主因子的因子得分显示, 不仅不同地理区域的种源可以清晰聚类(S16120 与 A18750), 同一地理来源的种源也可以清晰地区分(S16120 与 A18199)。对比种源间的遗传距离大小及其 PCA 主因子的得分图, 遗传距离小的种源间其因子得分有可能将其清晰地区分; 而遗传距离大的种源间, 其因子得分有可能存在严重的重叠, 这可能与种源内家系间的遗传变异较大有关。研究中大多数种源间的 PCA 主因子得分都显示出清晰的聚类, 为免冗余, 未一一展示。

图 2 显示种源 S16120 与 B10(c)和 S14339(d)不能清晰区分, 这与它们之间具有较大的遗传距离不符, 可能是种源 B10和 S14339 内家系间存在较大的遗传变异, 从其非常分

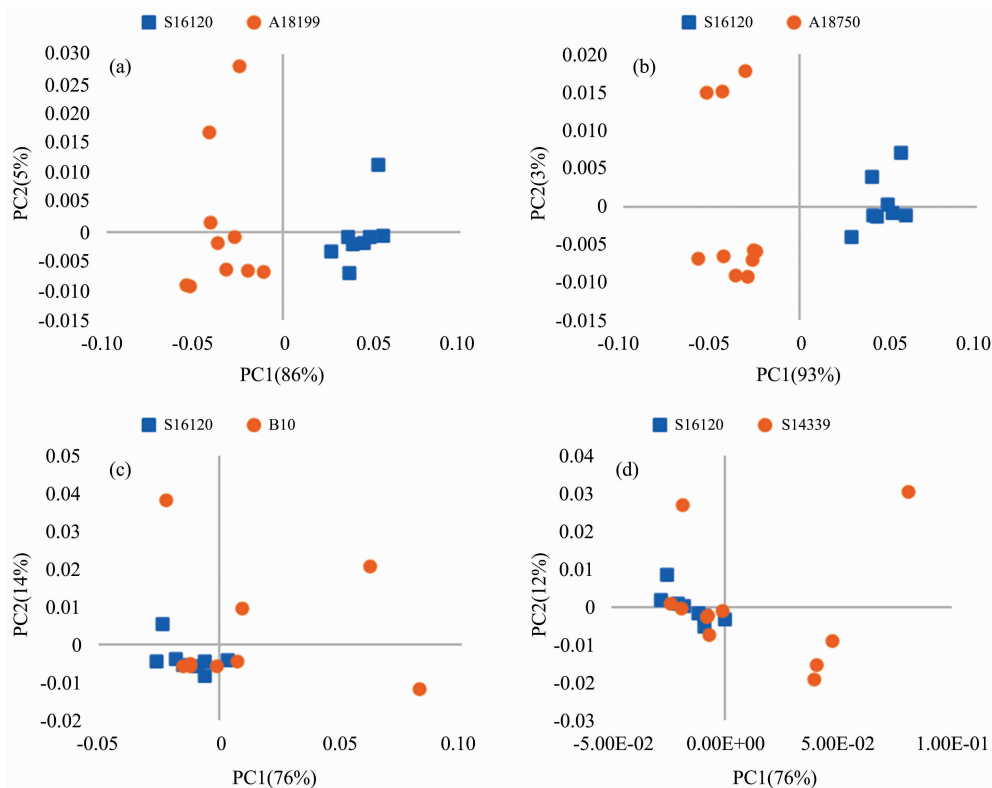


图 2 粗皮桉种源间的 PCA 因子得分图(部分)

(a): PC1 (86%)~PC2 (5%); (b): PC1 (93%)~PC2 (3%); (c): PC1 (76%)~PC2 (14%); (d): PC1 (76%)~PC2 (12%)

Fig. 2 PCA score plot of NIRs from provenances of *E. pellita* (partial)

(a): PC1 (86%)~PC2 (5%); (b): PC1 (93%)~PC2 (3%); (c): PC1 (76%)~PC2 (14%); (d): PC1 (76%)~PC2 (12%)

散的因子得分图也能看出。另外,图 2 各种源因子得分的聚集度各不相同, B10 和 S14339 的得分比较分散,而 S16120, A18199 和 A18750 的得分则相对集中,影响了种源间的相互区分。从遗传角度也反映了种源内家系间的遗传变异信息,分散可能表明其遗传变异大,集中则表示遗传变异小。

2.4 主成分分析中不同波段的因子载荷

图 3 是种源 S16120—A17861 间 PCA 聚类分析时第一主因子在不同波段的因子载荷。图中标出的 8 处特征峰代表

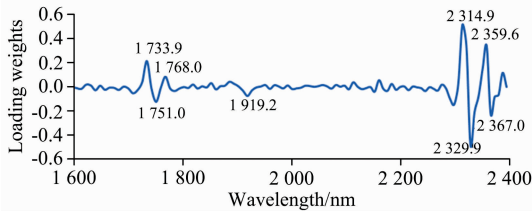


图 3 种源 S16120 与 A17861 的 PCA 分析中第一主因子不同波段的载荷(其他种源间因子载荷与此相似)

Fig. 3 The loading weights for the first latent variable of the PCA clustering between provenance S16120 and A17861. The loading weights for that of other provenances were similar with above

的化学键、化合物或是导致样本差异的来源。参阅文献[10-11]可知,波长 1 731~1 768 nm 表征 C—H 伸缩、C—H 和 S—H 一级倍频吸收峰;1 919 nm 表征水的吸收峰及 C=O 的伸缩吸收峰;2 310 和 2 328~2 332 nm 表征 C—H 伸缩、C—H 变形吸收峰,对应化合物为半纤维素;2 361 nm 左右表征 C—H 伸缩、C—H 变形及 C—H₂ 一级倍频弯曲振动吸收峰,对应化合物为纤维素。特征峰分析的意义在于直接建立其与目标物质间的模型,以优化模型结构和精度。另外,对于林木遗传育种研究,特征峰也为利用 NIRs 更好地辅助分析材料的遗传差异或性状变异来源提供了方向。

3 结 论

对比粗皮桉天然种源间的遗传距离与其 NIRs 光谱距离间的关系发现,种源遗传距离与其光谱距离间普遍存在正相关关系,但个别种源间的遗传距离与其光谱距离呈负相关。种源 NIRs 聚类结果与地理距离间的关系也非完全对应,这印证了粗皮桉群体不同地域间的基因交流对其遗传亲缘关系有较大的影响。NIRs 数据的 PCA 聚类显示,遗传或光谱距离大的种源间存在严重重叠,而遗传或光谱距离小的种源样本反而会清晰聚类,这既表明了 NIRs 信息的敏感性,也在一定程度反映了粗皮桉各种源内存在不同水平的遗传变异。

References

- [1] Thavamanikumar S, Arnold R J, Luo J Z, et al. *G3-Genes Genomes Genetics*, 2020, 10(10): 3751.
- [2] Hung T D, Brawner J T, Meder R, et al. *Annals of Forest Science*, 2015, 72(2): 205.
- [3] Castro C A O, Resende R T, Bhering L L, et al. *Ciência Rural*, 2016, 46(9): 1585.
- [4] LIU Xiao-hua, LUO Jian-zhong, LU Wan-hong, et al (刘晓华, 罗建中, 卢万鸿, 等). *Molecular Plant Breeding(分子植物育种)*, 2017, 15(12): 5103.
- [5] LU Wan-hong, YANG Gui-li, LIN Yan, et al (卢万鸿, 杨桂丽, 林彦, 等). *Scientia Silvae Sinicae(林业科学)*, 2017, 53(5): 16.
- [6] Yang G L, Lu W H, Lin Y, et al. *Journal of Tropical Forest Science*, 2017, 29(1): 121.
- [7] LU Wan-hong, LI Peng, WANG Chu-biao, et al (卢万鸿, 李鹏, 王楚彪, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2020, 40(3): 873.
- [8] Bertheau C, Aurélien S, Rossi J P, et al. *Forest Ecology & Management*, 2018, 258(7): 1619.
- [9] Sun D, Tang X, Zhan M, et al. *Frontiers in Microbiology*, 2020, 11: 589268.
- [10] Schwanninger M, Rodrigues J C, Fackler K. *Journal of Near Infrared Spectroscopy*, 2011, 19: 287.
- [11] YAN Yan-lu(严衍禄). *Near Infrared Spectroscopy Foundations and Applications(近红外光谱分析基础与应用)*. Beijing: China Light Industry Press(北京:中国轻工业出版社), 2005.

The Relationship Between Genetic Variations and NIRs Differences of *Eucalyptus Pellita* Provenances

WANG Chu-biao^{1, 2}, YANG Yan³, BAI Wei-guo⁴, LIN Yan¹, XIE Yao-jian¹, LU Wan-hong^{1*}, LUO Jian-zhong¹

1. Department of Genetics and Breeding, China Eucalypt Research Centre, Zhanjiang 524022, China

2. College of Forestry, Nanjing Forestry University, Nanjing 210037, China

3. School of Economics and Finance, Zhanjiang University of Science and Technology, Zhanjiang 524094, China

4. Guangxi Dongmen State Forest Farm, Chongzuo 532199, China

Abstract Clarifying the pedigree on *Eucalyptus pellita* populations is of great significance for studying rules of interspecific hybridization of eucalypt and the development of excellent new eucalypt genotypes. The purpose of the present study was to assess the accuracy and reliability of near infrared spectroscopy (NIRs) used in the analysis of the pedigree of *E. pellita* populations by comparing the relationship between genetic variations and NIRs differences that. The genetic materials involved natural provenances from the *E. Pellita* population, fresh leaves of 8~12 families were collected from each provenance. The DNA information of materials was obtained through whole-genome resequencing. Firstly, the genetic distances among provenances were evaluated with the DNA nucleotide sequence differences between samples. Meanwhile, four to six healthy leaves of each sample were placed in a drying oven until completely dry. The dried leaves were milled and then put into a transparent self-sealing plastic bag. A portable NIR device, phazir RX (1 624), was used to take the NIRs information of samples. The NIRs spectral distance between validating provenance and calibrating provenance was estimated with the soft independent modeling of class analogy (SIMCA). Hierarchical clustering was performed for all provenances with NIRs Euclidean distance. PCA scores plots of provenances NIRs demonstrated the pedigree and the genetic variations of provenances. The results showed that the total mean of the genetic distance of provenances from New Guinea Island and Queensland were 0.186 and 0.157 respectively, the total mean of genetic distance between New Guinea Island and Queensland was 0.295, which was higher than that within each separate district significantly. There was a positive correlation between NIRs spectral distance and genetic distance between provenances in two separate districts, but a negative correlation was also found between some provenances of *E. pellita*. The correlation between genetic distance and NIRs spectral distance was also proved by the NIRs Hierarchical clustering of all provenances. However, the clustering did not completely correspond with their geographical distance of provenances, suggesting that gene flow of some forms greatly affects the genetic relationship among separate districts of *E. pellita* populations. The PCA score plots demonstrated that PCs plots of some provenances with large genetic distance or NIRs spectral distance would overlap seriously, and PCs plots of some provenances with close genetic distance or NIRs spectral distance would be clustered, which verified the sensitivity of NIRs in the distinguishing of heterogeneous samples, also showed the genetic variation among families in provenance of *E. pellita*. All the current study results proposed that NIRs could genuinely reflect the genetic differences among provenances of *E. pellita*, and could be used to analyze the genetic relationship and genetic variation within eucalypt populations, and could be used to assist the improvement of eucalypts breeding populations in a generation.

Keywords Genetic distance; Spectral distance; Hierarchical clustering; Soft independent modeling of class analogy (SIMCA)

(Received Nov. 6, 2020; accepted Mar. 12, 2021)

* Corresponding author