人类 DNA 指纹分析

中文摘要

本研究聚焦于遗传标记和 DNA 指纹技术，深入探讨了遗传标记在遗传学和分子生物学领域的发展历程。特别关注 DNA 分子标记，其演变经历了形态标记、细胞学标记、生化标记和 DNA 分子标记四个阶段。DNA 分子标记的核心是 DNA 指纹技术，通过该技术可以生成 DNA 指纹图谱，从而实现对生物个体或种群间基因组差异的分析。

人类 DNA 指纹分析成为研究重点，其中可变数目串联重复序列（VNTR）和短串联重复序列（STR）是主要关注对象。VNTR 和 STR 在人类基因组 DNA 中展现高度多态性，通过 PCR 扩增和电泳检测等方法，可以获取有关基因型的信息。本实验选择了三个 VNTR 基因座（D1S80、D17$30 和 ApoB3’）和四个 21 号染色体上的 STR 基因座（D21S11、D21S1432、D21S2054 和 D21S1446）进行研究。实验方法包括全血 DNA 提取、PCR 扩增、电泳与检测以及数据记录与分析。

研究的目的在于深入了解 DNA 指纹分析技术的原理和应用，掌握可变数目串联重复序列和短串联重复序列多态性的检测和分析方法。通过实验，旨在了解串联重复序列多态性的含义，掌握相关技术操作，并为遗传学和分子生物学领域的研究提供实质性的数据支持。

关键词：遗传标记, DNA 指纹技术, 可变数目串联重复序列（VNTR）, 短串联重复序列（STR）, PCR 扩增, 电泳检测, 人类 DNA 指纹分析, 实验方法, 多态性, Hardy-Weinberg 平衡

Human DNA Fingerprint Analysis

Abstract

This study delves into the field of genetic markers and DNA fingerprinting technology, tracing the evolutionary stages from morphological markers to DNA molecular markers. Emphasis is placed on DNA fingerprinting, a technique that generates DNA profiles reflecting genomic differences between individuals or populations.

Human DNA fingerprint analysis, with a focus on variable number tandem repeats (VNTR) and short tandem repeats (STR), is investigated. Three VNTR loci (D1S80, D17$30, and ApoB3') and four 21st chromosome STR loci (D21S11, D21S1432, D21S2054, and D21S1446) are selected for the experiment. Methods include whole blood DNA extraction, PCR amplification, gel electrophoresis, and data analysis.

The purpose of the study is to comprehensively understand the principles and applications of DNA fingerprint analysis, master the detection and analysis methods of variable number tandem repeats and short tandem repeats polymorphisms. The experiment aims to reveal the significance of tandem repeat polymorphism, provide hands-on experience in relevant techniques, and contribute substantial data support to the fields of genetics and molecular biology.

Keywords: Genetic markers , DNA fingerprinting technology , Variable number tandem repeats (VNTR) , Short tandem repeats (STR) , PCR amplification , Gel electrophoresis , Human DNA fingerprint analysis , Experimental methods , Polymorphism , Hardy-Weinberg equilibrium

目　录

图 目 录

表 目 录

第一章 绪 论

随着分子生物学和遗传学领域的飞速发展，遗传标记和 DNA 指纹技术成为研究基因组和个体差异的关键工具。遗传标记是在基因组中标识和检测特定基因或 DNA 区域的分子标记，其演变历程反映了科学技术的进步。DNA 指纹技术作为遗传标记的一种重要形式，具有高度敏感性和准确性，广泛应用于法医学、人类学、亲权鉴定等领域。

遗传标记的发展经历了多个阶段，从形态标记、细胞学标记、生化标记到当前的 DNA 分子标记。DNA 分子标记以其高效的分辨率和稳定性，逐渐取代了前几代标记方法，成为研究个体间基因组差异的主流手段。其中，DNA 指纹技术通过生成 DNA 指纹图谱，展现了生物个体或种群之间的基因差异，成为遗传学研究的重要工具。

人类 DNA 指纹分析是 DNA 指纹技术的一个重要应用领域。可变数目串联重复序列（VNTR）和短串联重复序列（STR）是 DNA 指纹分析的核心内容，它们在人类基因组中呈现出丰富的多态性。本研究选择了具有代表性的 VNTR 和 STR 基因座，通过 PCR 扩增和电泳检测等技术手段，旨在深入了解串联重复序列多态性的含义，并掌握其检测和分析方法。

第二章 研究背景

2.1 遗传标记和 DNA 指纹技术

遗传标记在遗传学的建立和发展过程中有着举足轻重的作用，随着遗传学的进一步发展和分子生物学的异军突起，遗传标记先后相应地经历了形态标记、细胞学标记、生化标记和 DNA 分子标记四个发展阶段。DNA 分子标记本质上是指能反映生物个体或种群间基因组中某种差异的特异性 DNA 片段。DNA 分子标记大多以电泳谱带的形式表现生物个体之间的 DNA 差异，通常也称为 DNA 的指纹图谱。产生 DNA 指纹图谱的过程叫作 DNA 指纹分析。DNA 指纹技术的发展日新月异，第一代的分子标记是以 Southem 杂交为基础的限制性片段长度多态（restriction fragment length polymorphism,RFLP），第二代分子标记是以 PCR [1] 为基础的各种 DNA 指纹标记，如 RAPD、AFLP、短串联重复序列（shorttandem repeats,STR）和可变数目的重复序列（variable number of tandem repeat，VNTR），第三代分子标记是以单核苷酸多态性为基础的 SNP。一种理想的分子标记应具有以下特点：多态性高，重复性和稳定性好，带型清晰，容易统计，在染色体上均匀分布，共显性，简单快速，易自动化，开发和使用成本低廉等。

2.2 人类 DNA 指纹分析

人类基因组 DNA 中存在一类串联重复序列，其核心序列的长度为 10 70bp，该串联重复单位（核心序列）数目在人群中存在较大差异，具有高度多态性，称为可变数目串联重复序列（VNTR）或小卫星（mini-satellite）DNA。短串联重复序列（STR）形成多态性的原理与 VNTR 基本相同。STR 的核心序列短，为 2 7bp，片段长度为 100 500bp。STR 位点广泛地分布在人类基因组中。据估计，在人类基因组中，每 20kb 就有一个包含 3 或 4 个核苷酸重复序列的 STR 位点。因 STR 具有高度多态性及遗传稳定性，已逐渐取代 RFLP、VNTR 而被广泛应用于遗传疾病的诊断和法医学个人识别。

2.3 基因座选择

本实验选择人类基因组中的三个 VNTR 基因座（D1S80, D17$30 和 ApoB3’）和 21 号染色体上的四个 STR 基因座（D21S11,D21S1432,D21S2054 和 D21S1446）作为研究对象，通过对人基因组 DNA 的提取、多态性片段的 PCR 扩增以及电泳检测，了解串联重复序列多态性的含义和原理，掌握其检测和分析方法。

第三章　研究目的与方法

3.1　实验目的

- 了解 DNA 指纹分析技术的原理和应用。
- 级悉可發效国事联正复序 3 YNTB）列知事联重复序列（SPR）多态性的含义。
- 掌握可变数目串联重复序列和短串联重复序列多态性的检测和分析方法。

3.2　实验方法

3.2.1　实验材料与数据收集

- 实验数据

  参试者口腔上皮细胞
- 实验试剂

  1. 口腔上皮细胞 DNA 提取
     - 0.4% 生理盐水。
     - 裂解液：25 mmol/L NaOH,0.2 mmoI/L EDTA（乙二胺四乙酸）。
     - Tris-HCI（40 mmol/L,pH 5.0）。
     - 无水乙醇。
     - TE: 7 10 mmol/L Tris-HCI (pH8.0) 1 mmol/L EDTA.
  2. 全血 DNA 提取
     - 5% Chelex100 溶液：
     - Chelex 100 (Bio-Rad) 0.5 g, 50 mmol/L. Tris-HCI 10 μL, $4mol/LNaOHpH11.03$
  3. PCR 试剂
     - 2ŒTaq PCR Master Mix, ddIl.O，上下游引物，模板 DNA。
  4. 电泳检测试剂
     - 2.0% 琼脂糖凝胶：称取 2.0g 琼脂糖放人 250 μL.$1XTAE100\mu L60$°C 左右加入终浓度为 1wg/ μL.EB

- 10ŒTAE 电泳缓冲液：Tris 48.4g，冰醋酸 11.42 μL, $0.5mol/LEDTApH8.020\mu L1000\mu L$
- 0.5 mo/L EDTA （pH8.0）：Naz EDTA 18.61 g，NaOH 2.0g，蒸馏水定容至 100 μL
- 溴化乙锭（EB）：用无菌水配制 5 mg/ $\mu L1wg/ml$

- 10x 上样缓冲液（Loading dye）：溴酚蓝 0.25g，二甲苯腈蓝 0.25g，蔗糖 50.0g（或甘油 50 μL60 μL49 μL100 μL
- 0.16 mol/L 硝酸溶液。
- 10 mmo/L 硝酸银溶液。
- 0.28 mol/L 碳酸钠溶液。
- 1.67 mol/L 乙酸。
- 聚丙烯酰胺，DNA 相对分子质量标记。

• 实验仪器

微量移液器、高速冷冻离心机、恒温水浴锅、旋涡振荡器、PCR 扩增仪、电泳仪、电泳槽、凝胶成像系统、冰箱、制冰机、无菌枪头、无菌离心管（1.5 μL, 0.5 μL

3.2.2 可变数目串联重复序列 (VNTR)

1. 基因组 DNA 的提取

(a) 漱口，用无菌棉签刮取口腔脱落上皮细胞

(b) 将富集口腔脱落細胞的口監拭于放人盤有人 m 生和不的 1.3 μL

(c) 将拭子置于振荡器上，振荡 1min 左右，少心地娶這相金，再用适最的生理盐水冲洗。

(d) 13000rmin 离心 Jmin、小入批将主余的消被全部取起。记能物即为脱答上皮細胞，

(e) 在流淀中加人 25～50ML 裂解液，振满 10s。

(f) 98 と孵育 20min，振满后加入等体根 Tis-HICI （40 mmol/L, pH5.0）。

(g) 13 000 t/min 离心 10 min，取上清液。

(h) 加入无水乙醇，−20 °C 放置 15 min。

(i) 13000 r/min 离心 15 mino

(j) 弃上清，晾干后即得到口腔脱落上皮细胞 DNA。

(k) 加入适量 TE 溶解 DNA，4 °C 或 −20 °C 保存待用。

为提高 DNA 的产率和纯度，可选用 TANamp SwabDNA Kit（口腔拭子基因组 DNA 提取试剂盒，离心柱型，目录号：DP322）。

2. PCR 扩增

3 个 VNTR 基因座的引物序列和 PCR 扩增体系及扩增参数分别见表3.1和表3.2。

位点	引物序列	变性	退火	延伸	循环数
Marker	Primer Sequences	95 °C, 60 s	65 °C, 60 s	72 °C, 60 s	Cycle

DIS80	5'-GAAACTGGCCTCCAAACACTGCCCGCCG-3' 5'-GTCTTGTTGGAGATGCACGTGCCCCTTGC-3'	95 °C, 60 s	65 °C, 60 s	72 °C, 60 s	30
D17S30	5'-GGAAGAGTGAAGTGCACAGG-3' 5'-CACAGTCTTTATTCTTCAGCG-3'	94 °C, 30 s	55 °C, 30 s	72 °C, 80 s	30
ApoB3'	5'-ATGGAAACGGAGAAATTATG-3' 5'-CCTTCTCACTTGGCAAATAC-3'	94 °C, 60 s	63 °C, 60 s	72 °C, 120 s	26

表 3.1　3 个 VNTR 基因座的引物序列和 PCR 扩增条件

名称	体积/微升
上游引物（5wm）	1.0
下游弓物（5wm）	1.0
2xTaq PCR Master Mix	12.5
ddH_2O	0.5
模板 DNA	10.0
总体积	25.0

表 3.2　3 个 VNTR 位点 PCR 扩增体系

3. 电泳与检测

采用 2% 琼脂糖凝胶 60V 电泳，30 40 min 后取出凝胶用清水漂洗 5～10 min。凝胶成像系统观察和记录每个个体的 DNA 条带数目及其位置

4. 数据记录与分析

读取各个体样本的基因型，计算基因型频率及等位基因频率，用 x 检验进行 Hardy-Neinberg 平衡吻合度检验。

3.2.3　短串联重复序列 (STR)

1. 基因组 DNA 的提取

(a) 取 3 mL to 10 mL 全血加入 1.5 mL 离心管中，再加入 500 μL 纯水，剧烈振荡，室温下放置 15 min。

(b) 13 000 rpm 离心 3 min，弃去上清，收集沉淀。若需要，可使用蒸馏水反复清洗沉淀物，直至无色或血色素很少。

(c) 沉淀中加入 200 μL 5 % Chelex-100 溶液（5 % Chelex-100 为悬浊液，使用前要充分振摇，使 Chelex-100 颗粒悬浮），在振荡器上反复振荡后，放入 56 °C 水浴保温 30 min 以上。

(d) 取出后振荡，100 °C 保温 8 min，再振荡后，13 000 rpm 离心 3 min，上清用于 PCR

扩增，或放入 4 °C 保存备用。此 DNA 样本可在 4 °C 或 −20 °C 保存，必要时在使用前再次加热并离心，使管内物质分层。

2. 位点选择及其特性 (表3.3)

位点	重复序列	片段大小 (bp)	染色体位置 (Mb)
D21S11	TCTG	202-260	21q22.11
D21S1432	ATAG/TAGA	127-155	21q22.2
D21S2054	TCTA	162-182	21q22.11
D21S1446	TCTA/ATCT	160-187	21q22.3

表 3.3　4 个 21 号染色体 STR 位点的基本参数

3. PCR 扩增

3 个 VNTR 基因座的引物序列和 PCR 扩增体系及扩增参数分别见表3.4和表3.5。

位点	引物序列	变性	退火	延伸	循环数
D21S11	5’-TATGTGAGTCAATTCCCCAAG-3’ 5’ -GTTGTATTAGTCAATGTTCTCC-3’	95 °C,15s	58 °C,60s	60 °C,60s	30
D21S1432	5’-CTTAGAGGGACAGAACTAATAGGC-3’ 5’-AGCCTATTGTGGGTTTGTGA-3’	95 °C,15s	60 °C,60s	60 °C,60s	30
D21S2054	5’-GAGTAAATGTCATGAAACAAGG-3’ 5’-ATGATAGGTAGATGGATCAATTggAGA-3’	95 °C,40s	56 °C,40s	72 °C,30s	32
D21S1446	5’-ATGTACGATACGTAATACTTGAGAA-3’ 5’-GTCCCAAAGGACCTGCTC-3’	94 °C,40s	56 °C,50s	72 °C,50s	35

表 3.4　4 个 VNTR 基因座的引物序列和 PCR 扩增条件

名称	体积/微升
上游引物 (S wm)	2.2
下游引物 (S wm)	1.2
2XTaq PCR MasterMix	6.0
ddH_2O	2.1
模板 DNA	2.0
总体积	12.5

表 3.5　4 个 21 号染色体 STR 位点 PCR 扩增体系

4. 变性凝胶电泳

取扩增产物 2.0L 与 2.0L 变性加样缓冲液混合均匀，95 °C 变性 2 min，立即置于冰浴，

采用 6 % 变性聚丙烯酰胺凝胶垂直电泳，先预电泳 1 h，上样后恒功率 40 W，电泳 3 h。完毕后将凝胶取下，用双蒸水冲洗 1 次，置于 0.16 氮酸溶液中，轻摇反应 10 min，双蒸水冲洗，再置于 10 m 氮酸银溶液中，轻摇反应 20 min，双蒸水冲洗，然后置于 0.28 碳酸钠溶液中，轻摇反应，至条带清晰后加入 1.67 μL 乙酸终止反应并固定显色。

5. DNA 序列的测定

挑选每个 SIR 位点中至少 2 个以上不同片段长度的纯合子样本，经 PCR 扩增及纯化试剂盒纯化后进行 DNA 序列测定，根据测序结果推测其余不同片段长度等位基因的重复数。

6. 统计学分析

运用直接计数法观察位点的基因频率，通过 PowerStats 软件进行数据统计分析，计算杂合度（heterozygosity，H）、多态信息量（polymorphism information content, PIC）及个体识别率（average power of discrimination, PD）。用 χ^2 进行 Hardy-Weinberg 平衡吻合度检验。

$$H = 1 - \sum_{i=1}^{n} p_i^2$$

其中: n 为等位基因的数目, p_i 为单位基因的频率。

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i-1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

其中: n 为等位基因的数目, p_i 为第一个等位基因在群体中的频率。

第四章　研究结果与分析

4.1　电泳结果

在我们的实验中，21 级生科院的同学们对基因进行了分析并得到了多个结果。其中，图 4.1 展示了其中两组数据使用 Gelanalysis 分析后的分布情况, 这两个图具有代表性。

(a) dataimg2-1　　(b) dataimg2-2

图 4.1　电泳分析结果

如图A.1a所示，我们可以看到，第八条跑道为 marker 有 11 条带 (从上到下分别为: 1500bp, 1000bp, 900bp, ..., 100bp), 其余跑到均为样本, 大多样本有 2 个条带, 代表该样本为杂合子, 有两个等位基因, 也有少数样本只有一个条带, 代表该样本为纯合子, 只有一个等位基因。

图A.1b类似, 只是数据量较少, 且第六条跑道为 marker.

由于实验涉及多组数据，其他的一些结果图被在附录中供读者参考。

4.2　电泳条带数据

我们使用 Gelanalysis 对电泳结果进行分析，得到了每个样本的电泳条带数据，表4.1展示了图A.1b中的各个条带数据。

Lane #	Band #	Peak Rf	Peak value	Raw volume	Cal. volume	MW
1	1	0.382	52	405	-	663
1	2	0.466	74	560	-	441
1	3	0.741	251	3355	-	114
2	1	0.329	129	763	-	859
2	2	0.39	187	1281	-	639

Lane #	Band #	Peak Rf	Peak value	Raw volume	Cal. volume	MW
2	3	0.749	28	410	-	110
3	1	0.382	218	1704	-	663
4	1	0.359	196	1316	-	741
4	2	0.405	203	1226	-	593
5	1	0.367	144	767	-	714
5	2	0.405	132	804	-	593
5	3	0.764	34	376	-	102
6	1	0.214	96	592	-	1500
6	2	0.298	103	255	-	1000
6	3	0.321	140	517	-	900
6	4	0.336	125	352	-	800
6	5	0.367	102	370	-	700
6	6	0.405	98	422	-	600
6	7	0.443	178	869	-	500
6	8	0.489	70	379	-	400
6	9	0.558	68	470	-	300
6	10	0.634	62	474	-	200
6	11	0.733	43	449	-	100

表 4.1 电泳条带数据

可以看到跑道六有 11 个条带, 为 marker 与图A.1b中的第六条跑道对应, 而其他跑道有 1-3 个条带, 对应的 MW 为通过 Gelanalysis 计算得到的分子量, 其中小于 177 的条带都被认为是噪声, 将在数据处理中被过滤掉。

其他的数据结果请见附件

4.3 数据处理与分析

1. 批量导入数据

   如代码4.1, 导入../data/目录下 summary.csv 以外的所有.csv 文件, 并将其存储在 data_list 中, data_list 中的每个元素都是一个 pandas.DataFrame 对象, 代表一个.csv 文件的数据

Listing 4.1 批量导入数据

```
1 import os
2 import pandas as pd
3
```

```
# 设置工作目录
os.chdir("F://Onedrive//study//生物//遗传学//实验报告//3_人类DNA指纹分析//python")

# 导入数据 ############################
csv_files = [file for file in os.listdir("..\data") if file.endswith(".csv") and file != "summary.csv"]
csv_files = [os.path.join("..\data", file) for file in csv_files]
data_list = [pd.read_csv(file) for file in csv_files]
# 显示导入的数据 ######################
#for i, data in enumerate(data_list):
#    print(f"Data from file {csv_files[i]}:\n")
#    print(data)
```

2. 封装并清洗数据

如代码4.2, 将每个炮到数据清洗封装为一个 Lane 对象, 并将所有 Lane 对象存储在 lane_objects 中, lane_objects 中的每个元素都是一个 Lane 对象, 代表一个泳道的数据

Listing 4.2 封装并清洗数据

```

    设定m_edge = 177, 代表电泳图谱中的噪声边界, 小于177的条带都被认为是噪声, 将在数据处理中被过滤掉
class DnaFingerprintLane:
"""
DNA 指纹泳道类
"""
def __init__(self, data, filename):
    self.m_filename = filename
    self.m_data = data

    self.m_edge = 177
    self.m_laneID = data["Lane #"].iloc[0]
    if len(data[" Band #"]) == 11:
        self.m_category = "marker"
        self.m_MW = None
        self.m_repeatNum = None
        self.m_homoORheter = None

    else:
        self.m_category = "sample"
        self.m_MW = data[" MW"][data[" MW"] > self.m_edge].astype(float).tolist()
        self.m_repeatNum = self.calculate_repeat_num()
        if len(self.m_MW) == 1:
            self.m_homoORheter = "homo"
        elif len(self.m_MW) == 2:
            self.m_homoORheter = "heter"
        elif len(self.m_MW) > 2:
            self.m_homoORheter = "contamination"
```

```
29          elif len(self.m_MW) == 0:
30              self.m_homoORheter = "no result"
31          else:
32              self.m_homoORheter = "???"
33
34  lane_objects = []
35
36  for i, data_frame in enumerate(data_list):
37      for j in data_frame["Lane #"].unique():
38          lane = data_frame[data_frame["Lane #"] == j]
39          #DnaFingerprintLane(lane, csv_files[i]).printLane()
40          lane_objects.append(DnaFingerprintLane(lane,
                csv_files[i]))
```

3. 计算重复数

   定义类方法 calculate_repeat_num, 用于计算重复数, 代码4.3

Listing 4.3 计算重复数

```
1  def calculate_repeat_num(self):
2      mw_series = pd.Series(self.m_MW)
3      repeat_num = 1 + (mw_series[mw_series > self.m_edge] -
           161) / 16
4      return repeat_num.astype(float).tolist()
```

4. 数据可视化与保存

   将数据整理, 剔除没有数据与数据被污染的通道, 按照重复数的相似度进行排序, 并进行简单的可视化, 保存可视化结果, 数据保存到../data/summary.csv 中, 具体代码见附录.

   - 最终整理后的部分数据 (表4.2), 以及数据汇总 (表4.3)

File	Lane	MW	RepeatNum	HomoOrHeter
..10-1.csv	16	[468.0, 419.0]	[20.1875, 17.125]	heter
..7-2.csv	11	[472.0]	[20.4375]	homo
..10-2.csv	8	[472.0, 403.0]	[20.4375, 16.125]	heter
..8-1.csv	8	[476.0, 376.0]	[20.6875, 14.4375]	heter
..7-2.csv	8	[482.0]	[21.0625]	homo
..8-1.csv	3	[485.0]	[21.25]	homo
...	...	...	...	...

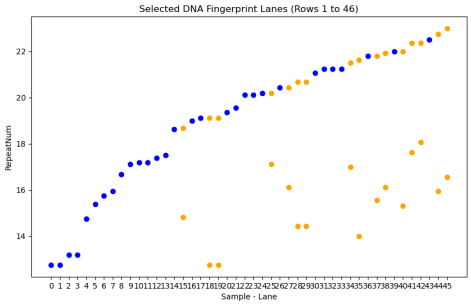
表 4.2 整理后部分数据

type	num
heter	106
homo	82

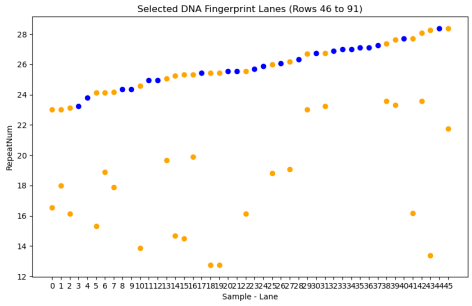
表 4.3　数据统计

杂合率: 0.56383

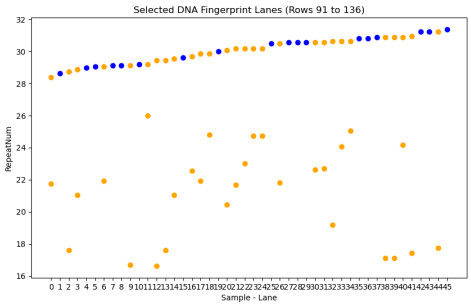
- 可视化结果, 图4.2



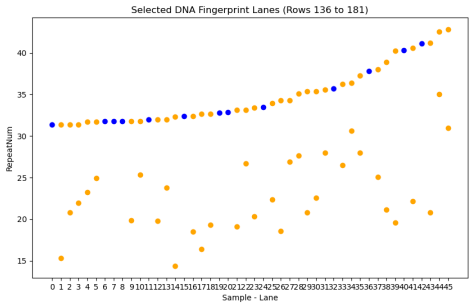
(a) Row1-46



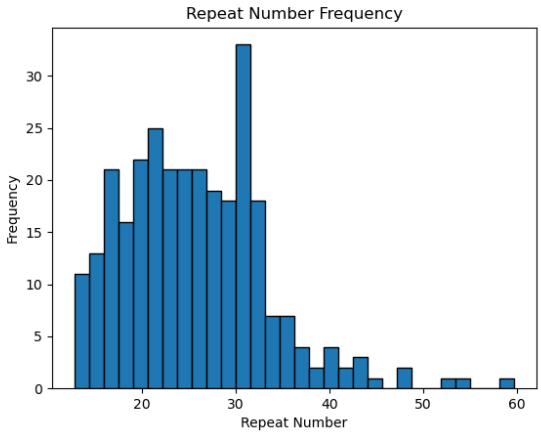
(b) Row47-90



(c) Row91-136



(d) Row137-181



(e) 分布直方图

图 4.2　可视化结果

第五章 讨 论

综上所述，我们通过实验成功进行了人类 DNA 指纹分析，主要关注可变数目串联重复序列（VNTR）和短串联重复序列（STR）的基因座。通过 PCR 扩增和电泳检测等技术手段，我们获得了电泳结果，并通过数据处理和分析，得到了每个样本的电泳条带数据。接着，我们进行了数据的清洗、封装、计算重复数以及最终的整理和可视化。

在数据处理的过程中，我们对电泳图谱中的噪声进行了边界设定，并清理了没有数据或被污染的通道。通过计算重复数，我们对样本的等位基因进行了分析，判断样本是纯合子还是杂合子。最终，我们将整理后的数据进行了可视化，并保存了数据汇总结果。

通过对 103 个杂合样本和 78 个纯合样本的统计，我们得到了杂合率为 0.564。这些结果对于深入了解 DNA 指纹分析技术的原理和应用，以及掌握可变数目串联重复序列和短串联重复序列多态性的检测和分析方法具有重要意义。

但是这一结果与我们取其的结果相差较大, 我们认为这一结果的原因有以下几点:

1. 样本数量不足：实验中选择了一部分样本进行分析，但样本数量相对较少，这可能限制了结果的普适性和可靠性。在今后的研究中，可以考虑增加样本数量，以更全面地了解 DNA 指纹分析的特征。
2. 数据的一致性：在电泳结果中，有时出现了不同样本之间数据的一致性问题，例如重复数的计算可能存在一些差异。这可能是由实验操作中的技术差异或误差引起的。在未来的实验中，需要进一步优化实验步骤，提高数据的一致性。
3. 标准化处理：在数据处理和分析过程中，可能存在一些标准化处理上的不足。例如，对于重复数的计算可能需要更加严格的标准和算法，以确保结果的准确性和可比性。
4. 深层次的统计分析：对于得到的数据，我们进行了一些简单的统计分析，但在深层次的数据挖掘和统计学分析方面，仍有提升空间。可以探索更多的统计学方法，了解不同基因座之间的关系，进一步挖掘数据背后的信息。
5. 实验技术的进一步改进：实验中使用的技术，如 PCR 扩增和电泳检测，是关键的步骤。技术水平的提高可能会改善实验结果的质量。在今后的实验中，可以考虑引入新的技术或改进现有技术，以提高实验的灵敏度和准确性。

参考文献

[1] Saiki R K, Scharf S, Faloona F, et al. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia[J]. Science, 1985, 230(4732):1350–1354.

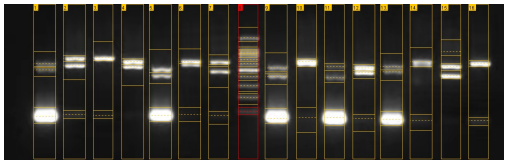
[2] 李文杨. 8 对人类遗传性状的调查和指纹分析——以信阳农林学院为例 [J]. 生物学通报, 2018, 53(41-44).

[3] 史燕顺 . Dna 指纹分析在肿瘤临床诊断中的应用研究进展 [J]. 临床检验杂志, 2005, (310-311).

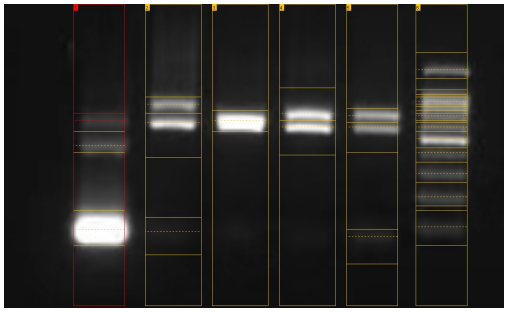
[4] Jeffreys 苏长林. 人类系谱中多个遗传标记的 dna“指纹图”和分离分析 [J]. 国外医学. 遗传学分册, 1987, (138-140).

附　　录

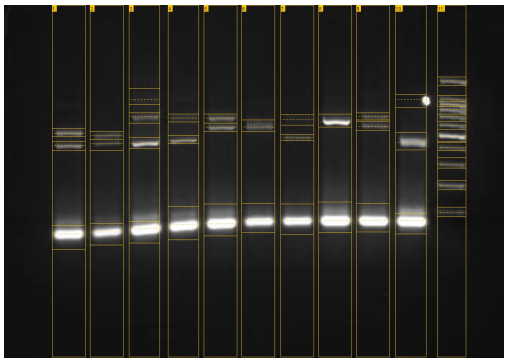
A.1　电泳结果



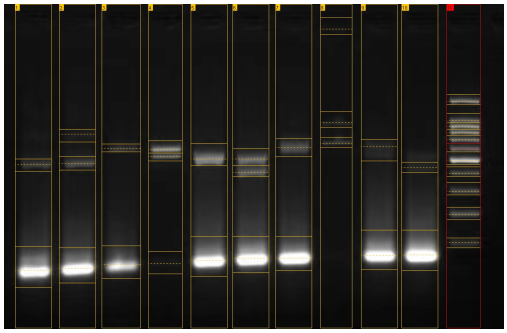
(a) dataimg1-1



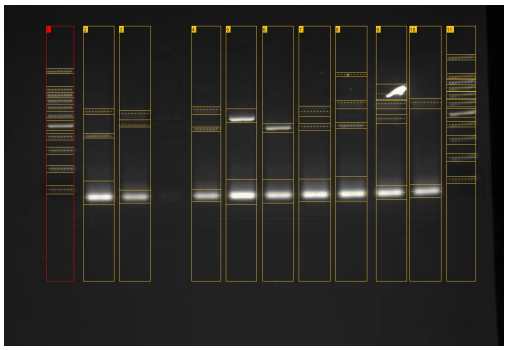
(b) dataimg1-2



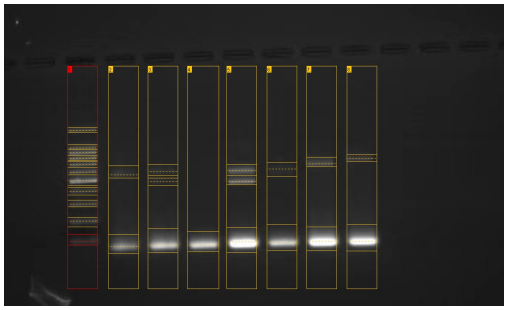
(c) dataimg1-3



(d) dataimg1-4

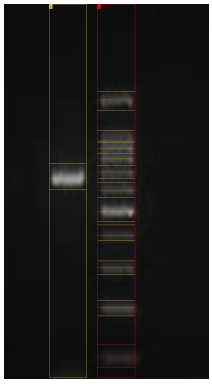


(e) dataimg1-5

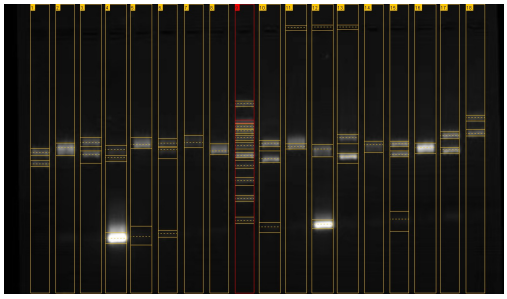


(f) dataimg1-6

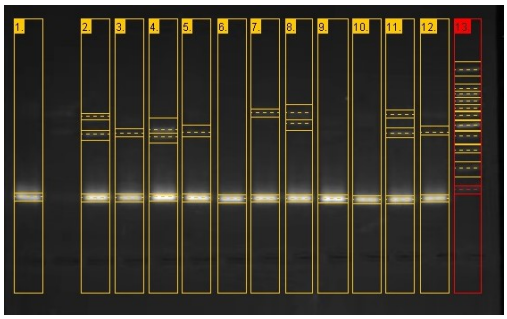
图 A.1　电泳分析结果



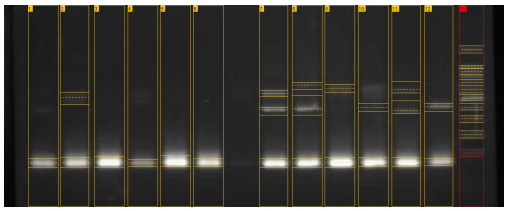
(a) dataimg1-7



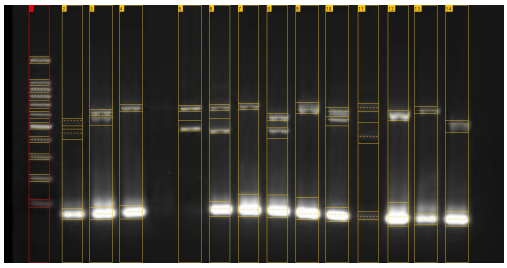
(b) dataimg1-8



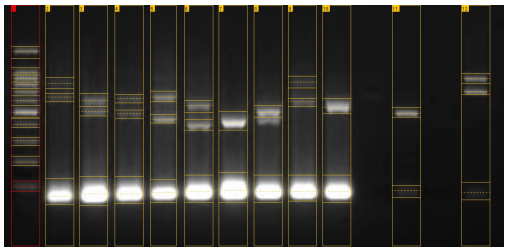
(c) dataimg1-9



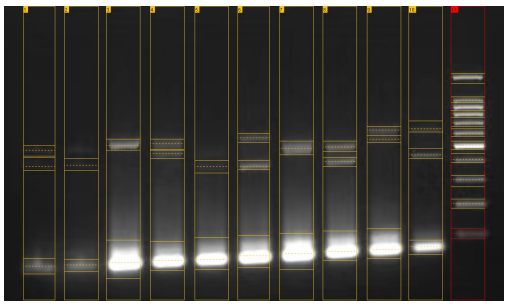
(d) dataimg1-10



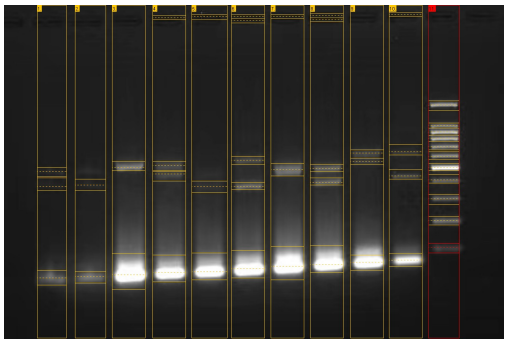
(e) dataimg1-11



(f) dataimg1-12

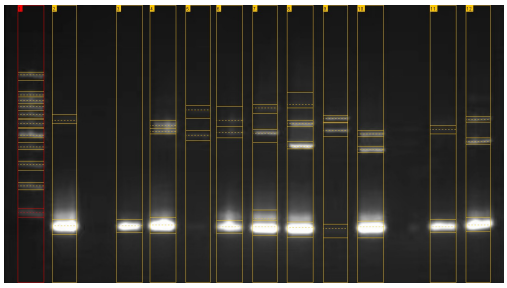


(g) dataimg1-13

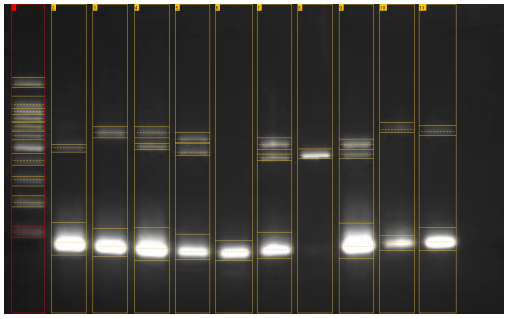


(h) dataimg1-14

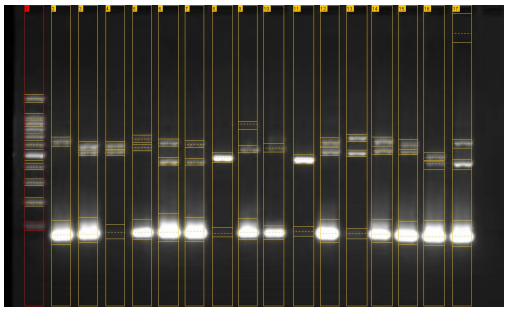
图 A.2　电泳分析结果



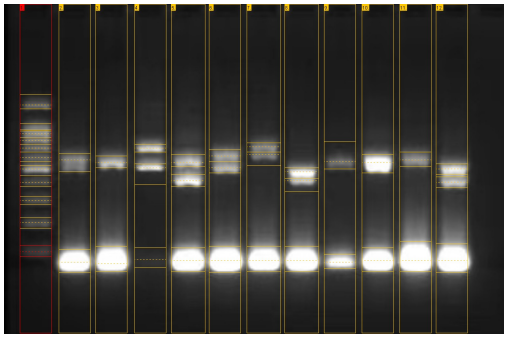
(a) dataimg1-15



(b) dataimg1-16



(c) dataimg1-17



(d) dataimg1-18

图 A.3　电泳分析结果

A.2　重复数总表

见附录 csv 文件, 位于/data/summary.csv(点我打开链接)

A.3　数据处理与分析代码

见附录 py 文件, 位于/python/main.py(点我打开链接)