动植物基因组重测序
遗传图谱结题报告

2024 年 01 月 15 日

目录

1 项目信息

1.1 项目信息表

项目名称			
动植物基因组重测序-遗传进化			
合同编号			
MJ20231204118			
项目样本信息			
物种信息	枣 (Ziziphus jujuba)		
样本个数	142		
合同指标			
备注			
客户信息（以上信息分析员填写，以下信息售后填写）			
单位名称			
项目联系人		电话	
		邮箱	
售后服务热线			
结题报告审核人		电话	021-51875086
		邮箱	DNA@majorbio.com
项目总监审批			
签名： 日期：			

1.2 项目研究背景

遗传图谱（Genetic Map）（Vision et al., 2000）是指分子标记在染色体上的相对位置与遗传距离的线性排列，其构建的理论基础是染色体的交换与重组。重组率的高低取决于交换的频率，而两个基因的交换频率取决于它们之间的物理距离，因此，重组率用来表示图距，单位厘摩（centi-Morgan，cM），1cM 表示 1% 的重组率。QTL（Quantitative Trait Locus）定位就是分析分子标记和数量性状表型值之间的关系，将数量性状位点逐一定位到连锁群的相应位置上，并估计其遗传效应。

本项目利用全基因组测序技术（Whole genome sequencing，WGS），在已知物种基因组信息的情况下，对物种内的不同个体进行基因组重测序（Re-sequencing），开发全基因组范围内的 SNP 和 InDel 分子标记，并利用分子标记进行遗传图谱构建。

1.3 材料基本信息

内容	基本信息
物种名	枣
拉丁文名	Ziziphus_jujuba
美吉基因组编号	GM2809
基因组大小（Mb）	383.84
参考基因组组装水平	Chromosome
参考基因组链接	10.1016/j.xplc.2023.100662

1.4 项目服务内容

按照合同约定，对 142 个检测合格的样本进行以下实验及分析：

1. 全基因组重测序，每个样本测序量达到合同标准，Q30 ≥ 80%。
2. 比对参考基因组进行变异检测分析，具体内容包括：SNP 检测和注释、InDel 检测和注释；
3. 遗传图谱构建：分子标记筛选过滤，binmarker 构建，遗传图谱构建，图谱质量评估。

1.5 分析结果概述

本项目共获得 203.05G reads 数据，测序 Q30 为 100.00%，GC 含量为 34.85%。通过生物信息学分析，共获得 4,752,991 个 SNP。

2 项目流程

2.1 全基因组重测序实验流程

样品基因组 DNA 检测合格后，利用超声波将 DNA 序列片段化形成随机片段，对片段化的 DNA 依次进行末端修复、3′ 端加 A、连接测序接头后，再利用磁珠吸附富集长度为 350bp 左右的片段，经过 PCR

扩增形成测序文库。建好的文库先进行文库质检，质检合格的文库用 Illumina NovaSeq™ 平台进行测序，测序策略为 Illumina PE150，总测序读长为 300bp。建库流程见图 **??**。

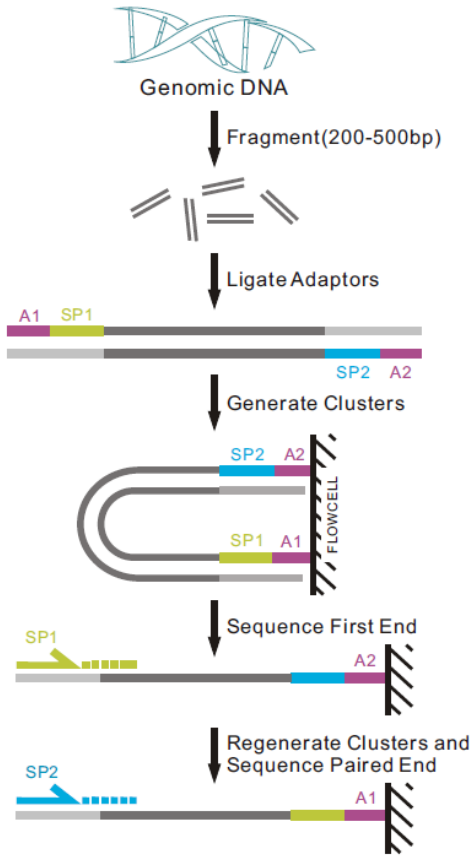

图 2.1 全基因组重测序实验建库流程

2.2 生物信息分析流程

在 Illumina NovaSeq™ 测序数据（Raw Data）下机之后，对下机数据进行质量控制，过滤其中低质量的数据，获得高质量的数据（Clean Data）。利用 BWA-MEME 软件（Jung and Han 2022）将 Clean Data 比对到参考基因组序列上，获得序列的位置归属（即 BAM 文件）。利用 GATK 软件（McKenna A *et al.* 2010）的 Best Practices 流程对 BAM 文件进行校正，并进行 SNP 标记的检测。利用 SNPEff 软件（Cingolani *et al.* 2012）和参考基因组的基因预测信息进行变异功能注释，得到 SNP 的功能注释信息。基于获得的 SNP 分子标记进一步进行图谱构建及 QTL 定位分析。分析流程见图**??**。

地址/Addr: 上海市浦东新区国际医学园区康新公路3399弄3号楼 电话/Tel: 021-5187
网址/Web: www.majorbio.com 服务热线: 400-660

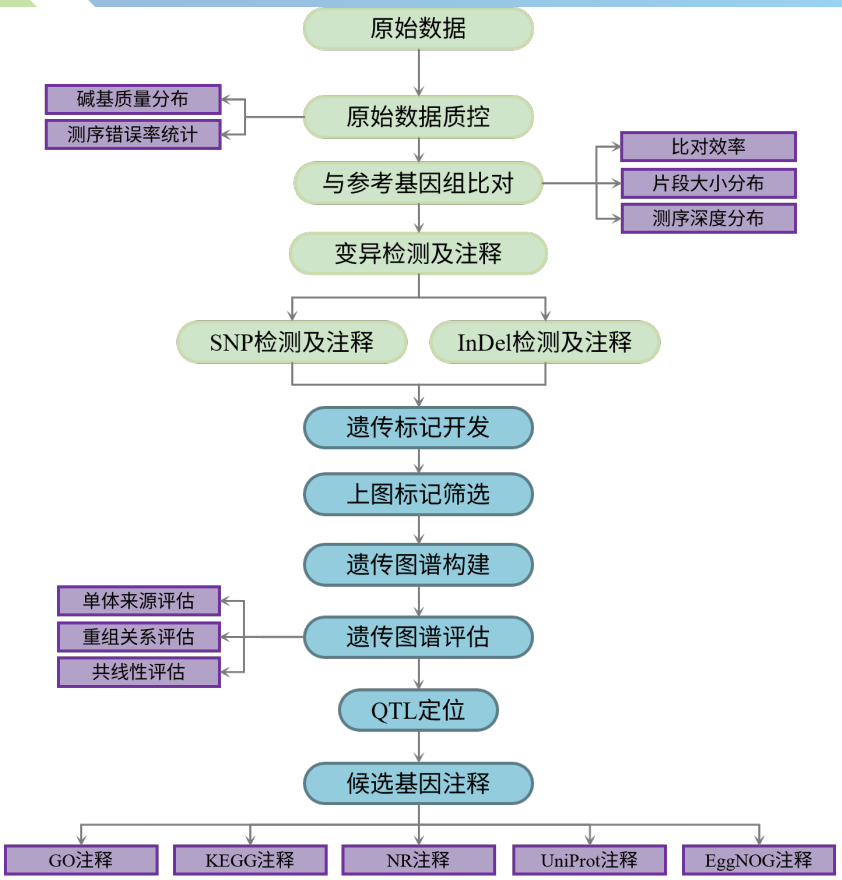

图 2.2 遗传图谱构建流程图

3 生物信息学分析方法和结果

3.1 原始数据质控和过滤

3.1.1 原始测序数据说明

为方便测序数据的分析、发布和共享，Illumina NovaSeq™ 平台测序得到的原始图像数据经过 Base Calling 转化为序列数据，得到最原始的测序数据文件。一般原始数据利用 FASTQ 格式进行储存。FASTQ 格式文件可记录所测读段（Read）的碱基及其质量分数。如图**??**所示，FASTQ 格式以测序读段为单位进行存储，每条 Reads 在 FASTQ 格式文件中占四行，其中第一行和第三行由文件识别标志（Sequence Identifiers）和读段名（ID）组成（第一行以 “@” 开头而第三行以 “+” 开头；第三行中 ID 可以省略，但 “+” 不能省略），第二行为碱基序列，第四行为对应位置碱基的测序质量分数。

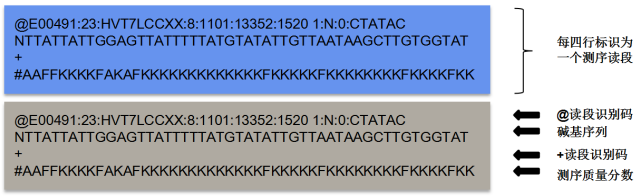

图 3.1 读段 FASTQ 数据格式示例

Illumina NovaSeq™ 测序仪一个 Run 有 2 个 Flowcell，一个 Flowcell 中包含 8 个 Lane，其中一个 Lane 包含 2 列，每一列又包含 60 个 Tile，每一个 Tile 又会包含不同的 Cluster，其产生的测序文件识别标志（Sequence Identifiers）中的详细信息如表**??**所示：

表 3.1 测序文件识别标志详细信息

标识	英文描述
E00491	Unique instrument name
23	Run ID
HVT7LCCXX	Flowcell ID
8	Flowcell lane
1101	Tile number within the flowcell lane
13352	’x’-coordinate of the cluster within the tile
1520	’y’-coordinate of the cluster within the tile
1	Member of a pair, 1 or 2 (paired-end or mate-pair reads only)
N	Y if the read fails filter (read is bad), N otherwise
0	0 when none of the control bits are on, otherwise it is an even number
CTATAC	Index sequence

Reads 的质量分数以不同的字符来表示，其中每个字符对应的 ASCII 值减去 33，即为对应的测序质量值。一般地，碱基质量从 0 到 40，对应的 ASCII 码为从 “!”（0+33）到 “I”（40+33），碱基质量越大，可信度越高。用 E 表示测序错误率，用 Q 表示 Illumina NovaSeq™ 的碱基质量值，则有下列关系：

$$Q = -10log_{10}(E)$$

表 3.2 测序错误率与测序质量值简明对应关系

测序错误率	测序质量值	对应 ASCII 码
5%	13	.
1%	20	5
0.1%	30	?
0.01%	40	I

Illumina 测序属于第二代测序技术，单次运行能产生数百万级的 Reads，如此海量的数据无法逐个展示每条 Read 的质量情况；运用统计学的方法，对所有测序 Reads 的每个 Cycle 进行碱基分布和质量波动的统计，可以从宏观上直观地反映出样本的测序质量和文库构建质量。我们针对每一个样本的原始测序数据进行测序相关质量评估，包括 A/T/G/C 碱基含量分布统计和碱基错误率分布统计。

表 3.3 产出数据统计结果

Sample ID	Raw Reads	Raw Bases (bp)	Raw GC (%)	Raw Q30 (%)
FJMS	45,151,054	6,762,077,180	34.58	100.00
H1	10,155,004	1,520,723,054	35.31	100.00
H3	6,398,326	958,307,115	34.41	100.00
H3J1	8,901,064	1,333,225,139	35.03	100.00
H3J2	9,306,464	1,393,857,776	35.09	100.00
H3J3	7,620,842	1,141,364,534	34.88	100.00
H3J4	8,923,024	1,336,401,487	35.90	100.00
H3J5	11,688,162	1,750,688,952	35.01	100.00
H3J6	10,244,740	1,534,416,755	34.61	100.00
H3J7	9,379,938	1,404,638,088	34.47	100.00
H3J8	8,680,212	1,300,238,221	34.76	100.00
H3J10	8,599,024	1,287,741,769	34.98	100.00
H3J11	8,615,504	1,290,235,870	34.98	100.00
H3J12	9,525,568	1,426,439,169	34.79	100.00
H3J13	8,852,826	1,326,006,951	34.65	100.00
H3J14	9,503,698	1,423,373,983	34.85	100.00
H3J15	8,971,650	1,343,601,836	34.40	100.00
H3J16	11,082,252	1,659,832,584	34.68	100.00
H3J18	8,503,574	1,273,426,048	34.74	100.00
H3J19	8,429,026	1,262,104,630	34.66	100.00
H3J20	9,607,148	1,438,941,535	34.28	100.00
H3J21	7,721,986	1,156,413,604	35.06	100.00

表 3.3 产出数据统计结果 (续)

Sample ID	Raw Reads	Raw Bases (bp)	Raw GC (%)	Raw Q30 (%)
H3J22	9,075,498	1,359,104,642	35.26	100.00
H3J23	6,731,090	1,007,878,505	34.96	100.00
H4	9,515,152	1,425,017,606	35.23	100.00
H5	8,783,882	1,315,594,927	34.89	100.00
H8	7,871,818	1,179,088,243	34.81	100.00
H9	7,927,218	1,187,184,778	34.87	100.00
H10	10,371,526	1,553,541,189	34.55	100.00
H11	10,027,798	1,501,571,220	35.06	100.00
H13	10,203,598	1,528,513,658	34.75	100.00
H15	9,077,466	1,359,198,281	35.52	100.00
H16	10,074,934	1,508,644,424	34.71	100.00
H17	8,364,552	1,252,475,762	34.55	100.00
H18	10,597,118	1,586,685,911	34.40	100.00
H20	8,717,102	1,305,164,494	34.89	100.00
H21	8,768,526	1,313,089,189	34.87	100.00
H22	10,611,832	1,589,518,145	34.29	100.00
H23	7,582,404	1,135,519,100	34.52	100.00
H25	7,623,418	1,141,627,887	34.80	100.00
H26	9,024,440	1,351,395,898	35.02	100.00
H28	10,535,954	1,578,215,020	34.70	100.00
H30	9,513,590	1,424,738,234	34.46	100.00
H31	9,828,554	1,472,232,259	34.40	100.00
H36	11,699,752	1,752,490,161	34.82	100.00
H37	8,149,110	1,220,434,722	34.52	100.00
H38	10,389,144	1,555,769,286	34.54	100.00
H39	7,708,506	1,154,643,031	34.82	100.00
H41	12,309,124	1,843,598,046	34.70	100.00
H42	10,378,202	1,553,781,871	34.58	100.00
H43	8,385,286	1,256,056,917	35.03	100.00
H44	10,724,812	1,606,372,394	34.79	100.00
H45	9,433,246	1,412,686,257	34.83	100.00
H46	8,059,902	1,206,938,916	34.75	100.00
H47	9,245,140	1,384,656,039	34.77	100.00
H49	8,647,976	1,295,292,633	34.86	100.00
H50	9,209,082	1,379,348,066	34.22	100.00
H51	8,999,432	1,347,653,855	34.81	100.00
H52	9,971,156	1,493,435,918	34.57	100.00
H54	8,140,806	1,219,067,465	34.95	100.00

表 3.3 产出数据统计结果 (续)

Sample ID	Raw Reads	Raw Bases (bp)	Raw GC (%)	Raw Q30 (%)
H55	11,373,972	1,703,591,590	34.92	100.00
H56	8,752,638	1,310,956,477	34.74	100.00
H58	7,937,916	1,188,802,333	34.55	100.00
H59	8,752,856	1,310,802,940	34.93	100.00
H61	7,701,034	1,153,218,387	34.85	100.00
H62	9,037,514	1,353,273,357	34.77	100.00
H63	9,694,688	1,452,103,613	34.62	100.00
H64	7,614,940	1,139,911,971	34.87	100.00
H70	9,560,600	1,431,845,110	34.51	100.00
H74	8,542,670	1,279,257,117	34.67	100.00
H76	8,614,514	1,289,687,778	35.26	100.00
H77	8,845,346	1,324,653,193	34.81	100.00
H80	8,075,084	1,209,394,226	35.03	100.00
H81	9,905,376	1,483,069,380	34.31	100.00
H82	8,330,764	1,247,241,967	34.65	100.00
H83	10,059,616	1,505,721,894	34.98	100.00
H84	9,406,078	1,408,826,710	34.82	100.00
H85	9,198,730	1,377,284,469	34.95	100.00
H87	8,920,732	1,335,589,305	35.23	100.00
H89	8,293,480	1,242,253,277	34.46	100.00
H91	7,435,980	1,113,669,081	34.60	100.00
H92	9,856,270	1,476,352,197	34.81	100.00
H93	7,630,890	1,142,655,506	34.64	100.00
H94	12,243,364	1,833,998,581	35.53	100.00
H95	8,461,766	1,266,493,753	34.68	100.00
H96	9,307,888	1,394,029,728	34.94	100.00
H97	8,315,308	1,245,382,415	35.05	100.00
H100	10,394,320	1,556,674,380	34.63	100.00
H102	9,891,700	1,481,070,979	34.68	100.00
H104	10,363,396	1,552,553,098	34.58	100.00
H105	7,443,108	1,114,588,155	35.01	100.00
H106	9,370,378	1,403,497,904	34.32	100.00
H108	8,313,452	1,244,916,936	34.20	100.00
H109	8,442,638	1,264,291,681	34.86	100.00
H112	12,163,126	1,821,886,417	35.56	100.00
H113	10,585,108	1,585,187,080	34.97	100.00
H115	11,786,208	1,764,929,603	34.71	100.00
H116	9,335,174	1,398,177,236	35.23	100.00

表 3.3 产出数据统计结果 (续)

Sample ID	Raw Reads	Raw Bases (bp)	Raw GC (%)	Raw Q30 (%)
H117	10,474,574	1,568,944,541	34.78	100.00
H120	8,282,458	1,240,217,362	33.96	100.00
H121	8,978,486	1,344,812,205	34.74	100.00
H124	7,749,802	1,160,469,743	34.46	100.00
H127	8,139,828	1,219,020,585	34.80	100.00
H128	9,772,568	1,463,703,697	34.51	100.00
H129	6,886,274	1,031,065,447	34.67	100.00
H132	9,509,644	1,424,134,930	34.93	100.00
H135	8,083,578	1,210,412,745	34.52	100.00
H137	6,493,300	972,062,816	36.32	100.00
H138	9,240,716	1,383,657,705	34.70	100.00
H140	8,047,410	1,205,076,199	34.63	100.00
H142	6,249,442	935,855,185	34.90	100.00
H143	9,277,392	1,388,765,316	34.79	100.00
H144	7,824,922	1,171,907,299	34.63	100.00
H145	9,306,654	1,393,677,385	35.17	100.00
H146	7,289,368	1,091,741,105	34.88	100.00
H148	6,599,092	988,369,703	34.67	100.00
H149	8,072,010	1,208,867,327	34.51	100.00
H150	10,151,600	1,520,370,257	34.34	100.00
H151	10,397,040	1,557,370,608	34.63	100.00
H152	8,218,834	1,230,892,418	35.35	100.00
H157	9,946,306	1,489,389,199	35.03	100.00
H158	8,436,628	1,263,161,223	35.51	100.00
H159	9,089,804	1,361,290,878	35.26	100.00
H160	8,389,996	1,256,505,138	36.48	100.00
H161	10,163,162	1,522,304,934	35.01	100.00
H162	10,948,948	1,639,300,602	34.90	100.00
H164	11,545,164	1,729,192,162	34.82	100.00
H166	9,672,324	1,448,835,471	34.89	100.00
H167	8,970,604	1,343,670,722	36.39	100.00
H168	9,559,198	1,431,700,401	34.88	100.00
H169	7,990,816	1,196,646,811	34.57	100.00
H170	7,951,082	1,191,076,748	34.92	100.00
H171	10,698,960	1,602,615,056	34.92	100.00
H172	6,998,498	1,048,277,174	36.46	100.00
H173	7,296,598	1,092,804,407	35.15	100.00
H174	9,095,666	1,362,552,895	35.12	100.00

表 3.3 产出数据统计结果 (续)

Sample ID	Raw Reads	Raw Bases (bp)	Raw GC (%)	Raw Q30 (%)
H175	8,087,960	1,211,553,230	34.39	100.00
H176	10,435,718	1,563,061,811	35.09	100.00
H177	7,247,366	1,085,315,692	34.20	100.00
H178	8,523,520	1,276,548,718	35.07	100.00
H179	7,408,978	1,109,624,905	34.88	100.00
MJ5	44,209,662	6,620,859,826	35.50	100.00

注：

Sample ID：样本编号；

Raw Reads：原始的 Reads 数；

Raw Bases (bp)：原始测序数据总碱基数；

Raw GC (%)：原始测序数据中的 GC 碱基占所有碱基的比例；

Raw Q30 (%)：原始测序数据中质量值大于或等于 30 的碱基所占百分比。

3.1.2 测序碱基含量分布统计

碱基含量分布检查一般用于检测有无 A 与 T、G 与 C 分离现象。鉴于序列的随机性和碱基互补配对的原则，理论上每个测序循环上的 GC 含量相等、AT 含量相等，且在整个测序过程基本稳定不变，呈水平线。N 为测序仪无法判断的碱基类型。

在实际测序中，首先会将文库 DNA 模板固定到芯片上，使每个 DNA 分子形成一个簇，即一个测序位点，在固定过程中极少量的簇与簇之间物理位置会发生重叠。测序时仪器首先通过前 4 轮测序循环对这些重叠的点进行分析和识别，将这些重叠点位置分开，保证每个点测到的是一个 DNA 分子，因此前几个碱基的错误率可能偏高、碱基含量可能存在着一定波动，属于正常情况，后续的数据质控会对此进行过滤。

本项目中样品的碱基含量分布图如图**??**所示。

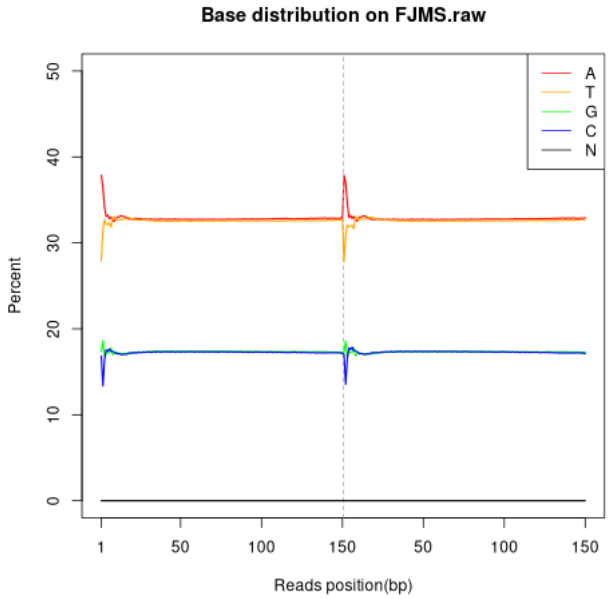

图 3.2 样本碱基组成分布图

注：横坐标是 Reads 碱基坐标，坐标表示 Reads 上从 5’ 到 3’ 端依次碱基的排列；纵坐标是所有 Reads 在该测序位置 A、C、G、T、N 碱基分别占的百分比，不同碱基用不同颜色表示。序列的起始位置与测序的引物接头相连，因此 A、C、G、T 在起始端会有所波动，后面会趋于稳定。模糊碱基 N 所占比例越低，说明未知碱基越少，测序样本受系统 AT 偏好影响越小。虚线之前为 Read1 的统计，虚线之后为 Read2 的统计结果。

3.1.3 测序碱基错误率分布统计

测序错误率会随着测序序列长度的增加而升高，这是由测序过程中化学试剂的消耗导致的，另外，由于 Illumina NovaSeq^{TM} 测序技术特点，测序片段前端几个 Cycles 和末端的错误率会偏高。本项目中所有样品的测序错误率分布图如图**??**所示。

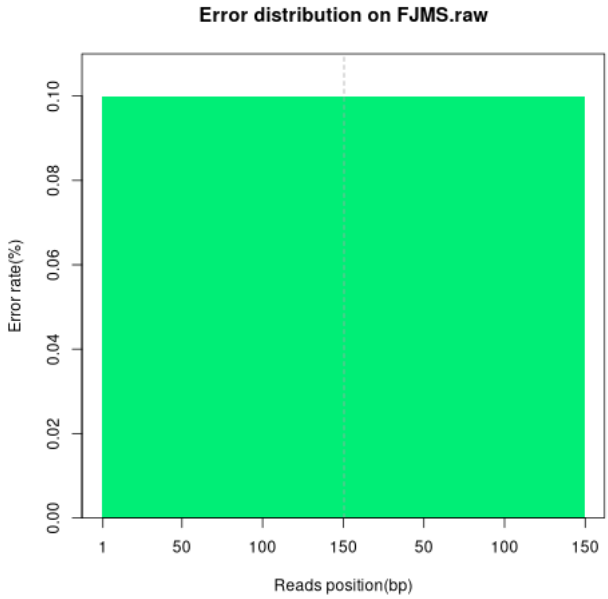

图 3.3 样本碱基错误率分布图

注：横坐标是 Reads 碱基坐标，表示 Reads 上从 5’ 到 3’ 端依次碱基的排列；纵坐标是所有 Reads 在该位点处碱基的平均错误率。前 150bp 为双端测序序列的第一端测序 Reads 的错误率分布情况，后 150bp 为另一端测序 Reads 的错误率分布情况。

3.1.4 原始测序数据过滤

利用 Illumina 的建库测序平台，构建插入片段大小为 350bp 左右的测序文库，按照项目合同要求进行测序，对原始数据进行质量评估，具体步骤如下：

Step1：去除 reads 中的 adapter 序列；

Step2：剪切前去除 5‘端含有非 AGCT 的碱基；

Step3：修剪测序质量较低的 reads 末端（测序质量值小于 Q20）；

Step4：去除含 N 的比例达到 10% 的 reads；

Step5：舍弃去 adapter 及质量修剪后长度小于 25bp 的小片段。

对质量剪切后的数据分别进行测序 Reads 数、总碱基数、GC 含量和 Q30 比例的统计，详细结果见表**??**：

表 3.4 测序质量统计表

Sample ID	Clean Reads	Clean Bases (bp)	Clean GC (%)	Clean Q30 (%)
FJMS	45,140,754	6,755,813,140	34.60	100.00
H1	10,152,458	1,519,220,837	35.32	100.00
H3	6,396,978	957,446,321	34.50	100.00
H3J1	8,899,098	1,331,964,178	35.06	100.00

表 3.4 测序质量统计表 (续)

Sample ID	Clean Reads	Clean Bases (bp)	Clean GC (%)	Clean Q30 (%)
H3J2	9,304,530	1,392,643,281	35.10	100.00
H3J3	7,618,952	1,140,183,157	34.91	100.00
H3J4	8,920,620	1,335,092,618	35.94	100.00
H3J5	11,685,584	1,749,070,423	35.03	100.00
H3J6	10,242,554	1,533,022,227	34.64	100.00
H3J7	9,377,556	1,403,192,664	34.50	100.00
H3J8	8,678,384	1,299,148,818	34.79	100.00
H3J10	8,597,232	1,286,639,442	35.03	100.00
H3J11	8,613,384	1,288,954,551	35.00	100.00
H3J12	9,523,128	1,424,969,262	34.81	100.00
H3J13	8,850,952	1,324,778,893	34.68	100.00
H3J14	9,501,708	1,422,104,065	34.87	100.00
H3J15	8,969,904	1,342,414,615	34.41	100.00
H3J16	11,079,606	1,658,191,242	34.70	100.00
H3J18	8,501,838	1,272,289,549	34.78	100.00
H3J19	8,426,934	1,260,826,271	34.66	100.00
H3J20	9,605,180	1,437,693,473	34.31	100.00
H3J21	7,720,030	1,155,204,662	35.08	100.00
H3J22	9,073,194	1,357,634,556	35.28	100.00
H3J23	6,729,490	1,006,932,657	35.00	100.00
H4	9,512,958	1,423,696,890	35.26	100.00
H5	8,782,104	1,314,442,505	34.92	100.00
H8	7,870,052	1,177,999,125	34.83	100.00
H9	7,925,408	1,186,046,129	34.88	100.00
H10	10,369,304	1,552,147,216	34.58	100.00
H11	10,025,526	1,500,127,440	35.11	100.00
H13	10,201,542	1,527,224,016	34.79	100.00
H15	9,075,662	1,358,064,963	35.55	100.00
H16	10,072,608	1,507,245,286	34.74	100.00
H17	8,362,594	1,251,292,991	34.59	100.00
H18	10,594,506	1,584,983,605	34.42	100.00
H20	8,715,316	1,303,981,449	34.94	100.00
H21	8,766,548	1,311,854,233	34.89	100.00
H22	10,609,638	1,587,983,632	34.30	100.00
H23	7,580,778	1,134,534,547	34.55	100.00
H25	7,621,810	1,140,601,234	34.83	100.00
H26	9,022,580	1,350,116,166	35.05	100.00
H28	10,533,600	1,576,737,094	34.72	100.00

表 3.4 测序质量统计表 (续)

Sample ID	Clean Reads	Clean Bases (bp)	Clean GC (%)	Clean Q30 (%)
H30	9,511,302	1,423,225,341	34.46	100.00
H31	9,826,712	1,471,045,628	34.43	100.00
H36	11,697,444	1,751,007,262	34.85	100.00
H37	8,147,388	1,219,396,021	34.55	100.00
H38	10,386,840	1,554,374,876	34.57	100.00
H39	7,706,940	1,153,624,737	34.84	100.00
H41	12,306,578	1,841,880,684	34.72	100.00
H42	10,375,710	1,552,305,188	34.61	100.00
H43	8,383,702	1,254,989,504	35.06	100.00
H44	10,722,598	1,605,018,207	34.81	100.00
H45	9,431,222	1,411,404,726	34.86	100.00
H46	8,057,922	1,205,830,517	34.77	100.00
H47	9,243,218	1,383,432,588	34.79	100.00
H49	8,646,202	1,294,189,208	34.88	100.00
H50	9,207,260	1,378,169,004	34.25	100.00
H51	8,997,436	1,346,422,653	34.84	100.00
H52	9,969,242	1,492,113,419	34.59	100.00
H54	8,138,986	1,217,876,898	34.97	100.00
H55	11,371,662	1,701,967,034	34.94	100.00
H56	8,750,766	1,309,821,246	34.77	100.00
H58	7,936,232	1,187,674,751	34.58	100.00
H59	8,750,862	1,309,510,713	34.96	100.00
H61	7,699,294	1,152,128,433	34.88	100.00
H62	9,035,418	1,351,994,426	34.81	100.00
H63	9,692,740	1,450,861,561	34.64	100.00
H64	7,613,154	1,138,741,429	34.88	100.00
H70	9,558,656	1,430,545,134	34.54	100.00
H74	8,540,682	1,278,010,668	34.69	100.00
H76	8,612,366	1,288,385,839	35.28	100.00
H77	8,843,366	1,323,399,738	34.84	100.00
H80	8,073,386	1,208,309,404	35.07	100.00
H81	9,902,808	1,481,456,550	34.33	100.00
H82	8,328,758	1,245,977,003	34.68	100.00
H83	10,056,990	1,504,103,695	35.02	100.00
H84	9,404,404	1,407,645,958	34.84	100.00
H85	9,196,522	1,375,921,020	34.95	100.00
H87	8,918,526	1,334,279,442	35.26	100.00
H89	8,291,774	1,241,182,510	34.47	100.00

表 3.4 测序质量统计表 (续)

Sample ID	Clean Reads	Clean Bases (bp)	Clean GC (%)	Clean Q30 (%)
H91	7,434,422	1,112,660,905	34.62	100.00
H92	9,854,586	1,475,140,391	34.85	100.00
H93	7,629,038	1,141,510,656	34.68	100.00
H94	12,240,580	1,832,236,431	35.57	100.00
H95	8,459,492	1,265,016,542	34.69	100.00
H96	9,305,902	1,392,705,111	34.96	100.00
H97	8,313,474	1,244,227,080	35.06	100.00
H100	10,391,928	1,555,123,916	34.65	100.00
H102	9,889,248	1,479,620,961	34.71	100.00
H104	10,361,438	1,551,243,651	34.59	100.00
H105	7,441,408	1,113,612,931	35.04	100.00
H106	9,368,472	1,402,253,377	34.34	100.00
H108	8,311,702	1,243,758,434	34.27	100.00
H109	8,440,978	1,263,228,877	34.89	100.00
H112	12,160,434	1,820,135,470	35.61	100.00
H113	10,582,644	1,583,603,132	34.99	100.00
H115	11,783,496	1,763,218,331	34.73	100.00
H116	9,333,264	1,396,925,087	35.27	100.00
H117	10,472,374	1,567,560,265	34.80	100.00
H120	8,280,800	1,239,003,022	33.99	100.00
H121	8,976,734	1,343,670,342	34.77	100.00
H124	7,747,946	1,159,383,770	34.49	100.00
H127	8,137,924	1,217,873,821	34.83	100.00
H128	9,770,584	1,462,438,599	34.54	100.00
H129	6,884,656	1,030,023,449	34.68	100.00
H132	9,507,626	1,422,929,371	34.96	100.00
H135	8,081,732	1,209,239,802	34.55	100.00
H137	6,491,818	971,128,370	36.34	100.00
H138	9,238,582	1,382,335,396	34.72	100.00
H140	8,045,708	1,203,957,536	34.65	100.00
H142	6,247,994	934,955,155	34.92	100.00
H143	9,274,956	1,387,302,549	34.82	100.00
H144	7,823,204	1,170,837,736	34.66	100.00
H145	9,304,756	1,392,501,774	35.20	100.00
H146	7,287,908	1,090,754,965	34.90	100.00
H148	6,597,718	987,471,976	34.69	100.00
H149	8,070,276	1,207,703,281	34.52	100.00
H150	10,149,446	1,518,988,176	34.38	100.00

表 3.4 测序质量统计表 (续)

Sample ID	Clean Reads	Clean Bases (bp)	Clean GC (%)	Clean Q30 (%)
H151	10,394,830	1,555,973,805	34.66	100.00
H152	8,217,034	1,229,761,325	35.37	100.00
H157	9,943,834	1,487,756,340	35.05	100.00
H158	8,434,152	1,261,711,237	35.53	100.00
H159	9,087,786	1,360,073,789	35.27	100.00
H160	8,388,386	1,255,447,170	36.49	100.00
H161	10,160,868	1,520,849,267	35.05	100.00
H162	10,946,162	1,637,427,732	34.91	100.00
H164	11,542,740	1,727,637,403	34.84	100.00
H166	9,670,560	1,447,614,948	34.91	100.00
H167	8,968,558	1,342,483,242	36.43	100.00
H168	9,557,078	1,430,338,006	34.90	100.00
H169	7,988,952	1,195,452,507	34.58	100.00
H170	7,949,462	1,190,137,720	34.96	100.00
H171	10,696,708	1,601,197,622	34.95	100.00
H172	6,996,912	1,047,300,959	36.47	100.00
H173	7,294,958	1,091,743,854	35.19	100.00
H174	9,093,880	1,361,411,757	35.16	100.00
H175	8,086,392	1,210,539,765	34.42	100.00
H176	10,433,180	1,561,613,423	35.12	100.00
H177	7,245,758	1,084,255,028	34.23	100.00
H178	8,521,384	1,275,268,886	35.09	100.00
H179	7,407,594	1,108,637,615	34.90	100.00
MJ5	44,199,780	6,614,788,643	35.53	100.00

注:
Sample ID：样本编号；
Clean Reads：高质量的 Reads 数；
Clean Bases (bp)：原始数据过滤后的高质量测序数据总碱基数；
Clean GC (%)：原始数据过滤后的 GC 碱基占所有碱基的比例；
Clean Q30 (%)：原始数据过滤后质量值大于或等于 30 的碱基所占百分比。

3.2 基因组比对

3.2.1 基因组比对效率

样本基因组比对率反映了样本测序数据与参考基因组的相似性，该结果可以帮助判断参考基因组的选择是否合理以及排除异常样本。在本项目中，我们以枣（Ziziphus_jujuba）的基因组序列作为参考基因组。利用 BWA-MEME 软件将质控后的测序片段（Clean Reads）比对参考基因组，比对方法为 MEM。表**??**为比对结果的数据统计表。

表 3.5 比对结果数据统计表

Sample ID	Mapped Ratio(%)	Proper Ratio(%)	Real Depth	Insert Size	Coverage(%) (>=1x)	Coverage(%) (>=4x)
FJMS	99.49	91.16	17.56	413	95.08	91.14
H1	99.46	90.70	4.50	416	83.43	36.65
H3	99.46	91.26	3.18	416	74.43	19.00
H3J1	99.46	91.88	4.02	394	81.97	31.75
H3J2	99.38	89.65	4.25	431	80.97	31.57
H3J3	99.45	92.69	3.60	301	78.16	26.71
H3J4	99.40	90.95	4.19	398	78.66	30.58
H3J5	99.39	91.58	5.00	403	86.36	44.77
H3J6	99.43	90.16	4.53	430	83.58	37.02
H3J7	99.33	93.42	4.12	324	84.13	38.15
H3J8	99.46	91.87	4.00	381	80.25	30.63
H3J10	99.43	90.70	3.86	428	82.29	29.63
H3J11	99.43	91.44	3.93	413	81.10	31.10
H3J12	99.39	93.09	4.20	304	83.72	37.25
H3J13	99.43	91.97	4.02	392	81.35	32.38
H3J14	99.41	90.28	4.17	444	84.13	34.70
H3J15	99.45	91.47	3.94	421	84.19	33.28
H3J16	99.48	91.74	4.80	404	85.45	42.47
H3J18	99.45	89.80	3.81	448	82.47	28.18
H3J19	99.47	92.44	3.84	321	81.17	31.12
H3J20	98.89	92.23	4.20	334	84.20	38.30
H3J21	99.47	92.34	3.64	389	78.54	26.77
H3J22	99.46	91.80	4.06	409	82.66	33.44
H3J23	99.41	90.20	3.25	435	76.44	19.79
H4	99.40	90.40	4.28	423	82.22	33.30
H5	99.38	91.00	3.91	429	82.93	31.57
H8	99.47	89.98	3.65	437	79.67	25.21
H9	99.48	91.82	3.59	411	81.68	27.00
H10	99.44	91.31	4.46	411	85.92	39.82
H11	99.46	90.55	4.37	425	84.92	36.41
H13	99.42	91.35	4.47	404	84.32	38.74
H15	99.40	90.37	4.02	451	83.52	32.14
H16	99.43	91.11	4.36	420	85.31	38.26
H17	99.41	91.23	3.73	430	82.78	29.65
H18	99.44	91.07	4.50	443	87.02	41.10
H20	99.34	90.84	3.87	428	83.11	30.73
H21	99.42	91.06	3.92	421	82.65	30.99

表 3.5 比对结果数据统计表 (续)

Sample ID	Mapped Ratio(%)	Proper Ratio(%)	Real Depth	Insert Size	Coverage(%) (>=1x)	Coverage(%) (>=4x)
H22	99.50	92.92	4.49	327	87.47	43.83
H23	99.46	91.01	3.52	416	79.75	24.54
H25	99.41	90.74	3.50	436	80.41	24.86
H26	99.38	90.54	4.00	428	83.26	31.55
H28	99.40	90.86	4.57	423	85.12	40.07
H30	99.47	92.31	4.19	422	84.04	37.79
H31	99.34	91.32	4.29	417	84.67	37.83
H36	99.44	90.81	4.97	451	87.12	46.17
H37	99.43	90.95	3.70	417	81.36	27.63
H38	99.35	90.38	4.49	426	85.41	38.95
H39	99.44	91.30	3.57	400	79.93	25.02
H41	99.49	90.85	5.17	443	88.01	48.36
H42	99.44	91.97	4.45	422	86.28	42.04
H43	99.41	90.70	3.78	422	81.90	28.51
H44	99.39	91.11	4.62	414	85.83	41.19
H45	99.44	90.60	4.16	422	83.88	33.60
H46	99.41	91.14	3.67	413	81.06	27.33
H47	99.40	91.80	4.06	415	84.18	34.99
H49	99.46	90.82	3.88	422	82.33	29.97
H50	99.40	90.71	4.05	435	83.94	33.68
H51	99.45	91.05	3.97	436	83.89	32.66
H52	99.48	92.79	4.26	333	86.53	40.39
H54	99.37	91.19	3.69	418	81.56	28.17
H55	99.44	90.84	4.83	432	87.10	43.63
H56	99.46	91.13	3.88	420	83.43	30.72
H58	99.48	92.24	3.56	421	82.56	28.42
H59	99.48	90.55	3.89	443	83.26	30.30
H61	99.42	90.86	3.54	424	80.31	24.94
H62	99.47	91.19	3.99	422	83.68	32.93
H63	99.32	90.73	4.25	422	84.22	35.73
H64	99.39	90.21	3.49	457	80.50	24.21
H70	99.45	91.13	4.17	423	84.68	35.64
H74	99.41	90.96	3.80	426	82.96	30.27
H76	99.48	91.77	3.85	415	82.68	31.14
H77	99.40	91.84	3.94	402	82.99	32.70
H80	99.45	90.19	3.69	440	80.89	26.35
H81	99.45	93.33	4.20	342	87.25	41.46

表 3.5 比对结果数据统计表 (续)

Sample ID	Mapped Ratio(%)	Proper Ratio(%)	Real Depth	Insert Size	Coverage(%) (>=1x)	Coverage(%) (>=4x)
H82	99.38	91.18	3.73	421	82.42	29.14
H83	99.38	93.36	4.29	340	86.57	42.38
H84	99.40	89.62	4.18	434	83.10	32.03
H85	99.45	91.52	4.04	413	84.13	33.39
H87	99.43	90.86	4.00	418	82.51	31.16
H89	99.41	91.56	3.76	404	81.60	29.50
H91	99.44	91.58	3.44	403	80.01	24.22
H92	99.39	89.57	4.34	436	84.02	34.18
H93	99.47	91.64	3.51	404	80.49	25.49
H94	99.44	89.61	5.31	455	85.27	45.23
H95	99.43	91.98	3.73	420	83.67	30.96
H96	99.40	90.44	4.15	444	82.91	33.34
H97	99.40	90.44	3.80	416	80.90	27.19
H100	99.46	92.21	4.46	416	86.23	41.55
H102	99.46	91.46	4.29	416	85.17	37.41
H104	99.43	90.57	4.49	443	85.28	39.20
H105	99.44	91.18	3.50	406	78.62	24.15
H106	99.46	90.98	4.14	452	83.71	35.25
H108	99.43	90.42	3.71	456	82.79	29.65
H109	99.40	89.80	3.82	449	81.67	28.21
H112	99.42	90.42	5.23	439	86.03	45.88
H113	99.37	91.10	4.55	432	85.98	40.88
H115	99.47	92.08	4.90	411	88.91	48.26
H116	99.45	90.76	4.21	443	81.93	34.30
H117	99.46	92.16	4.52	427	85.77	41.57
H120	99.29	94.05	3.62	338	84.34	33.36
H121	99.41	89.80	4.02	439	82.46	30.74
H124	99.43	90.79	3.55	425	80.65	25.57
H127	99.45	92.27	3.67	327	82.04	29.39
H128	99.47	91.20	4.23	427	85.54	37.01
H129	99.42	90.47	3.29	432	77.33	20.22
H132	99.41	90.45	4.18	431	84.09	34.15
H135	99.36	91.57	3.63	427	82.22	28.53
H137	99.41	89.80	3.26	443	73.64	17.99
H138	99.41	90.88	4.06	424	84.20	33.28
H140	99.47	90.45	3.63	456	81.87	26.71
H142	99.42	90.40	3.06	446	75.50	17.51

表 3.5 比对结果数据统计表 (续)

Sample ID	Mapped Ratio(%)	Proper Ratio(%)	Real Depth	Insert Size	Coverage(%) (>=1x)	Coverage(%) (>=4x)
H143	99.49	91.57	4.04	427	84.79	34.62
H144	99.41	91.64	3.55	419	81.41	27.28
H145	99.47	90.51	4.11	424	83.65	32.52
H146	99.44	90.89	3.44	414	78.28	22.74
H148	99.46	91.11	3.16	438	77.22	20.01
H149	99.43	91.26	3.67	412	81.40	27.09
H150	99.37	92.04	4.33	422	86.64	40.97
H151	99.37	90.69	4.54	418	84.54	38.43
H152	99.43	90.47	3.83	422	79.33	27.26
H157	99.45	91.44	4.38	415	83.99	36.60
H158	99.43	91.98	3.87	398	80.62	29.47
H159	99.50	91.74	4.12	385	81.67	32.14
H160	99.38	88.51	3.91	478	79.23	26.62
H161	99.47	91.55	4.48	393	83.91	37.49
H162	99.41	91.86	4.72	414	85.60	42.20
H164	99.39	90.79	4.98	416	85.59	43.76
H166	99.41	89.55	4.31	440	82.97	33.35
H167	99.40	90.45	4.28	392	77.39	29.67
H168	99.43	91.01	4.25	421	83.06	34.97
H169	99.43	92.04	3.57	417	82.65	28.68
H170	99.46	91.94	3.79	382	77.56	27.03
H171	99.44	91.03	4.68	412	84.60	39.98
H172	99.37	90.61	3.55	400	72.87	21.05
H173	99.44	91.45	3.48	416	77.60	24.03
H174	99.33	91.23	4.22	391	79.63	32.10
H175	99.43	90.81	3.75	418	79.73	27.22
H176	99.44	91.00	4.60	403	83.93	37.90
H177	99.37	93.26	3.28	344	81.52	26.63
H178	99.49	91.77	3.90	400	80.80	30.09
H179	99.39	91.26	3.44	427	79.53	24.69
MJ5	99.44	89.97	17.24	426	94.81	89.75

表 3.5 比对结果数据统计表 (续)

Sample ID	Mapped Ratio(%)	Proper Ratio(%)	Real Depth	Insert Size	Coverage(%) (>=1x)	Coverage(%) (>=4x)

注：
Sample ID：样品编号；
Mapped Ratio(%)：定位到基因组的 Clean Reads 数占所有 Clean Reads 数的百分比；
Properly Mapped(%)：双端均定位到基因组上且距离符合测序片段长度的 Reads 数百分比；
Real Depth：相对于整体基因组中覆盖度大于 1 部分序列的平均覆盖深度；
Insert Size：样品平均插入片段长度；
Coverage(%) (>=1x)：至少有一条 Reads 覆盖的碱基占基因组长度的百分比；
Coverage(%) (>=4x)：至少有四条 Reads 覆盖的碱基占基因组长度的百分比。

3.2.2 插入片段长度统计

通过检测双端序列在参考基因组上的起止位置，可以得到样品 DNA 打断后得到的测序片段的实际大小，即插入片段大小（Insert Size），是生物信息分析时的一个重要参数。插入片段大小的分布一般符合正态分布，且只有一个单峰。样品的插入片段长度分布如图**??**所示，插入片段长度分布符合正态分布，中心值在 350 bp 左右，说明测序数据文库构建无异常。

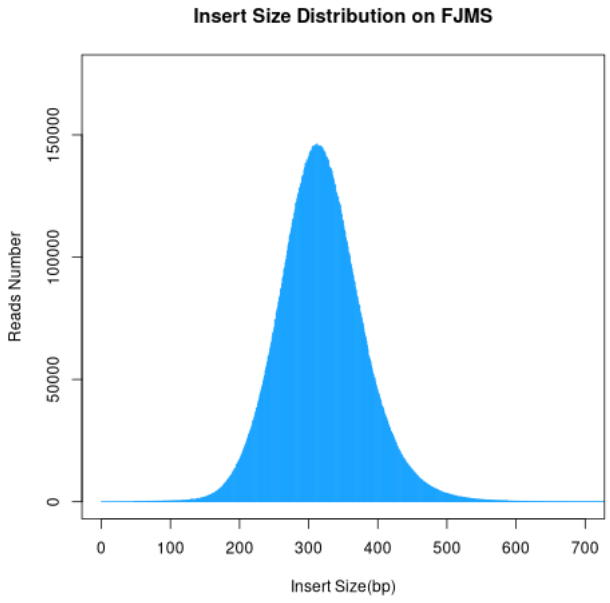

图 3.4 样品的插入片段长度分布图

注：横坐标为 Reads 对应的插入片段大小，纵坐标为相应插入片段大小所对应的 Reads 数。

3.2.3 深度分布统计

Reads 定位到参考基因组后，可以统计参考基因组上碱基的覆盖情况。参考基因组上被 reads 覆盖到的碱基数占基因组总长度的百分比称为基因组覆盖度；碱基上覆盖的 reads 数为覆盖深度，样品的碱基深度分布图见图**??**。覆盖深度和覆盖度能够直接反应测序数据的均一性及与参考序列的同源性。基因组覆盖度可以反映参考基因组上变异检测的完整性，覆盖到的区域越多，可以检测到的变异位点也越多。基因组的覆盖深度会影响变异检测的准确性，在覆盖深度较高的区域（非重复序列区），变异检测的准确性也越高。另外，若基因组上碱基的覆盖深度分布较均匀，也说明测序随机性较好。碱基在基因组上的覆盖深度分布如图**??**所示。

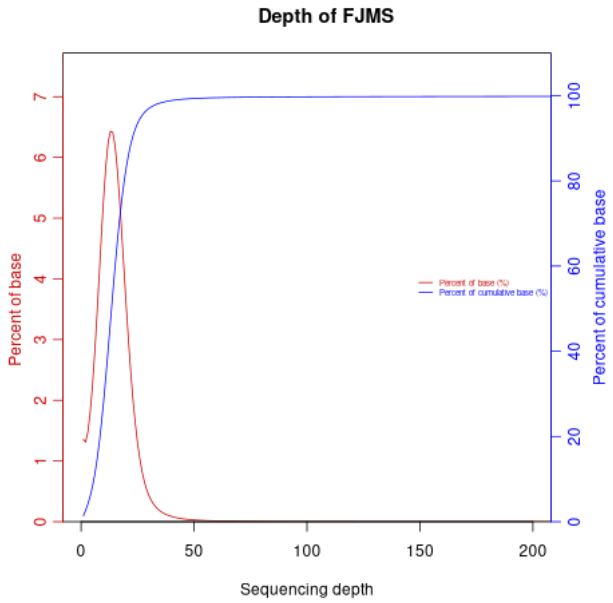

图 3.5 样品的深度分布图

注：横坐标表示测序深度，图中左侧的纵坐标轴（红色）对应红色曲线，表示对应深度的位点占全基因组的百分比，图中右侧的纵坐标（蓝色）对应蓝色曲线，表示小于或等于该深度的位点占全基因组的百分比。

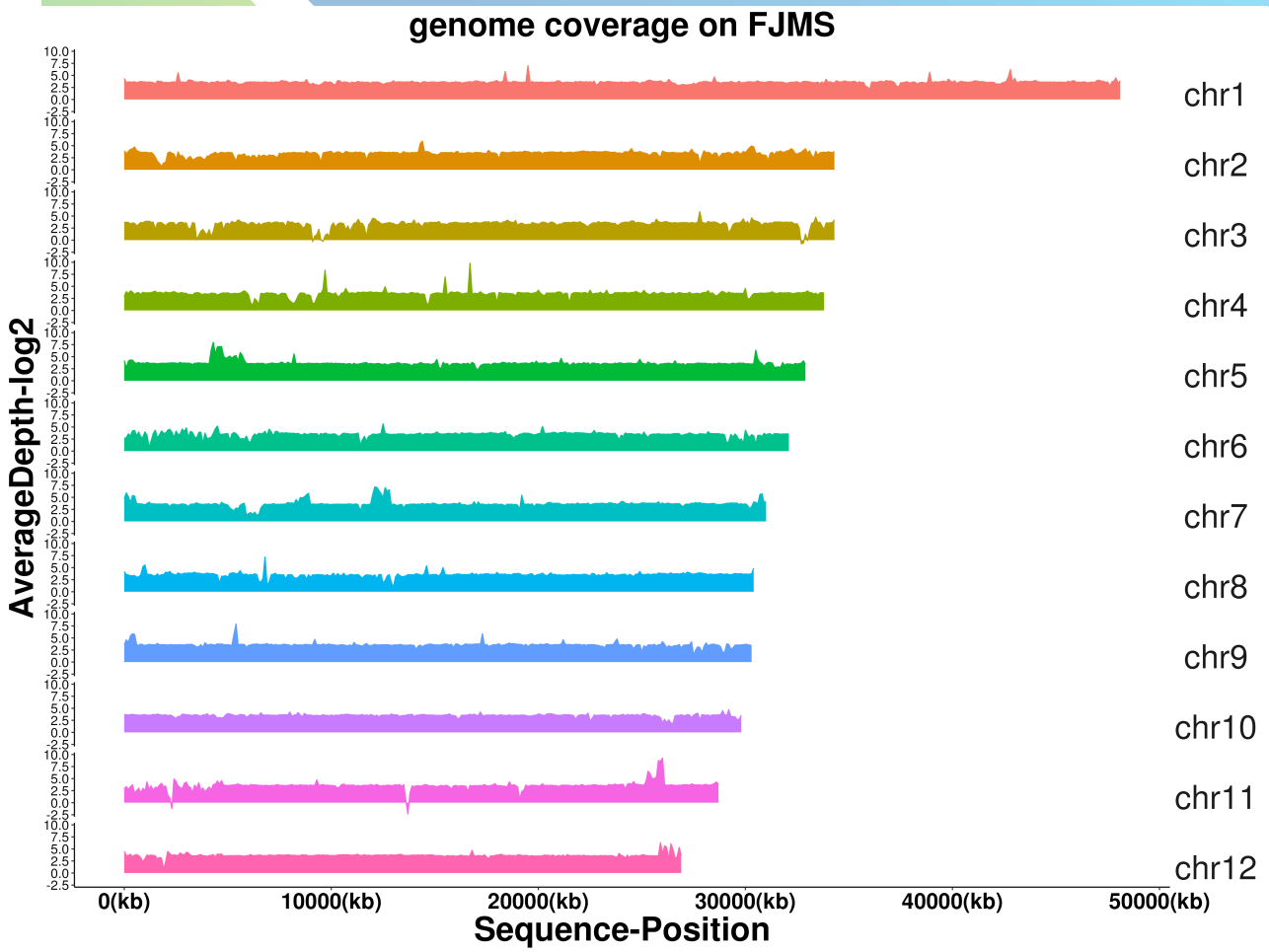

图 3.6 样品的染色体覆盖深度分布图

注：横坐标为染色体位置，纵坐标为染色体上对应位置的覆盖深度取对数（log2）得到的值。基因组被覆盖的较均匀，说明测序随机性较好。图上深度不均一的地方可能是由于重复序列、PCR 偏好性、或着丝粒部分引起的。

3.3 SNP 检测和注释

单核苷酸多态性（Single Nucleotide Polymorphism，SNP），是指基因组中由单个核苷酸的变异所引起的 DNA 序列多态性，是基因组上多态性最高的遗传变异之一。SNP 变异类型分为转换和颠换两种，同种类型碱基（嘌呤与嘌呤、嘧啶与嘧啶）之间的突变称为转换（Transition）；不同类型碱基（嘌呤与嘧啶）之间的突变称为颠换（Transversion）。一般转换比颠换更容易发生，所以转换/颠换（Ts/Tv）的比例一般大于 1，具体比值和所测物种有关。

3.3.1 SNP 检测

利用 GATK 的 Best Practices 流程处理比对结果（BAM 文件），利用 GATK 的 Haplotyper 方法进行 SNP 检测，过滤条件按照 GATK 推荐的参数进行，具体可见https://software.broadinstitute.org/gatk/documentation/article.php?id=3225。样品 SNP 统计结果见表**??**：

表 3.6 SNP 数据统计表

Sample ID	SNP Number	Transition	Transversion	Ts/Tv	Heterozygosity Number	Homozygosity Number
FJMS	2,897,610	1,854,989	1,046,838	1.77	2,085,117	2,600,923
H1	1,182,424	755,719	427,343	1.77	322,662	3,164,400
H3	898,920	575,697	323,611	1.78	236,228	2,739,714
H3J1	1,044,004	667,847	376,710	1.77	313,483	3,108,844
H3J2	1,046,299	665,488	381,268	1.75	285,990	2,939,314
H3J3	950,124	608,239	342,328	1.78	260,199	2,871,296
H3J4	1,034,047	662,404	372,076	1.78	285,091	2,851,734
H3J5	1,537,420	988,963	549,530	1.80	657,314	3,151,585
H3J6	1,256,080	804,639	452,173	1.78	450,368	3,083,513
H3J7	1,349,327	867,748	482,482	1.80	481,025	3,180,956
H3J8	1,186,816	766,620	420,850	1.82	423,517	2,981,030
H3J10	1,062,649	677,305	385,854	1.76	306,843	3,087,034
H3J11	1,244,928	803,931	441,758	1.82	423,932	2,953,747
H3J12	1,284,529	824,705	460,566	1.79	401,838	3,137,295
H3J13	1,282,927	826,115	457,572	1.81	427,989	2,998,097
H3J14	1,337,068	855,775	482,056	1.78	412,556	3,107,677
H3J15	1,393,718	895,468	499,156	1.79	526,322	3,111,893
H3J16	1,221,793	778,636	443,900	1.75	389,584	3,220,289
H3J18	1,217,350	783,250	434,772	1.80	470,393	2,944,355
H3J19	1,196,395	769,617	427,497	1.80	399,083	3,043,279
H3J20	1,232,090	785,441	447,289	1.76	376,375	3,156,952
H3J21	905,389	578,178	327,557	1.77	228,732	2,876,387
H3J22	1,045,616	669,637	376,494	1.78	305,719	3,014,280
H3J23	1,060,596	684,302	376,787	1.82	304,961	2,741,742
H4	1,138,952	726,699	412,890	1.76	352,502	3,000,358
H5	1,358,280	873,521	485,568	1.80	490,835	3,031,299
H8	1,086,355	694,879	392,005	1.77	302,648	2,860,338
H9	1,029,785	659,523	370,762	1.78	275,220	3,119,215
H10	1,599,479	1,027,101	573,643	1.79	621,465	3,201,220
H11	1,530,196	985,020	546,286	1.80	529,045	3,166,257
H13	1,345,369	860,461	485,748	1.77	434,088	3,165,460
H15	1,106,924	709,633	397,928	1.78	361,418	3,185,015
H16	1,305,955	834,303	472,478	1.77	440,067	3,211,584
H17	1,117,516	712,492	405,588	1.76	327,861	3,123,331
H18	1,354,989	867,866	488,028	1.78	476,859	3,301,422
H20	1,224,318	786,475	438,516	1.79	376,673	3,114,853
H21	1,309,458	843,229	466,998	1.81	475,232	3,044,031

表 3.6 SNP 数据统计表 (续)

Sample ID	SNP Number	Transition	Transversion	Ts/Tv	Heterozygosity Number	Homozygosity Number
H22	1,433,942	922,372	512,633	1.80	605,005	3,338,535
H23	1,149,905	738,689	411,826	1.79	377,140	2,974,030
H25	1,141,232	732,692	409,178	1.79	374,613	2,969,784
H26	1,262,333	807,440	455,593	1.77	384,535	3,027,133
H28	1,368,649	872,873	496,637	1.76	445,805	3,106,650
H30	1,488,511	956,942	532,647	1.80	499,954	3,142,562
H31	1,426,619	909,929	517,635	1.76	453,975	3,086,107
H36	1,487,803	951,111	537,816	1.77	528,896	3,352,036
H37	1,018,347	647,407	371,428	1.74	263,477	3,086,346
H38	1,510,892	963,181	548,726	1.76	464,098	3,153,614
H39	1,191,795	766,955	425,515	1.80	370,418	2,979,622
H41	1,618,262	1,037,306	582,214	1.78	618,248	3,281,891
H42	1,579,348	1,011,710	568,914	1.78	583,739	3,260,081
H43	1,231,572	793,820	438,423	1.81	437,274	3,055,995
H44	1,633,548	1,049,825	584,984	1.79	583,763	3,223,686
H45	1,418,498	911,321	508,191	1.79	545,059	3,096,074
H46	940,553	596,518	344,382	1.73	245,510	2,975,822
H47	1,212,120	774,215	438,578	1.77	347,106	3,224,923
H49	1,238,801	797,959	441,577	1.81	464,113	3,032,916
H50	1,389,228	889,411	500,643	1.78	486,860	3,122,130
H51	1,306,784	840,802	466,772	1.80	478,919	3,145,641
H52	1,414,679	905,109	510,597	1.77	517,664	3,316,993
H54	1,351,109	869,210	482,718	1.80	437,922	2,994,706
H55	1,385,850	884,652	502,141	1.76	512,329	3,319,764
H56	1,220,333	783,375	437,682	1.79	467,132	3,091,957
H58	1,142,225	732,535	410,286	1.79	360,078	3,137,796
H59	1,180,199	757,042	423,780	1.79	376,408	3,132,727
H61	1,129,273	724,776	405,160	1.79	350,438	3,007,840
H62	1,199,147	762,392	437,440	1.74	365,312	3,100,784
H63	1,450,524	930,819	520,653	1.79	510,814	3,133,707
H64	1,133,305	730,226	403,767	1.81	378,624	2,955,751
H70	1,458,231	935,987	523,173	1.79	517,053	3,165,381
H74	1,042,435	663,364	379,539	1.75	277,880	3,022,012
H76	1,254,895	806,303	449,346	1.79	398,550	3,101,153
H77	1,389,381	895,102	495,211	1.81	479,464	3,079,513
H80	1,040,475	667,161	373,831	1.78	317,777	3,051,960
H81	1,625,001	1,042,112	584,175	1.78	663,176	3,204,595

表 3.6 SNP 数据统计表 (续)

Sample ID	SNP Number	Transition	Transversion	Ts/Tv	Heterozygosity Number	Homozygosity Number
H82	1,047,081	666,867	380,710	1.75	307,614	3,038,846
H83	1,683,536	1,085,284	599,627	1.81	645,456	3,165,714
H84	1,147,137	731,188	416,555	1.76	338,709	3,005,174
H85	1,372,920	882,960	490,898	1.80	522,336	3,107,437
H87	1,383,683	890,912	493,702	1.80	471,973	3,050,734
H89	1,347,695	869,666	478,912	1.82	461,297	2,976,018
H91	1,076,363	693,139	383,778	1.81	400,360	2,837,751
H92	1,276,304	812,450	464,526	1.75	363,550	3,091,538
H93	1,093,113	701,281	392,460	1.79	363,892	3,005,324
H94	1,624,591	1,044,470	581,496	1.80	650,986	3,103,655
H95	1,252,707	806,171	447,240	1.80	428,421	3,083,227
H96	1,309,978	840,666	470,071	1.79	424,208	3,159,264
H97	1,114,153	712,392	402,276	1.77	312,429	2,910,565
H100	1,425,646	911,387	515,323	1.77	563,261	3,225,079
H102	1,366,972	877,775	490,053	1.79	520,986	3,175,808
H104	1,548,965	995,146	554,885	1.79	612,385	3,114,988
H105	994,522	637,319	357,691	1.78	275,193	2,957,253
H106	1,281,497	817,151	465,061	1.76	381,772	3,212,426
H108	1,096,957	701,210	396,293	1.77	326,172	3,142,112
H109	1,153,224	735,395	418,425	1.76	329,506	2,986,232
H112	1,470,250	941,043	530,105	1.78	503,579	3,202,837
H113	1,453,042	930,546	523,384	1.78	469,121	3,256,688
H115	1,720,544	1,106,676	615,388	1.80	762,359	3,288,920
H116	1,261,119	807,871	453,980	1.78	377,669	3,074,095
H117	1,437,786	923,182	515,578	1.79	557,123	3,257,956
H120	1,224,851	786,232	439,379	1.79	420,859	3,226,883
H121	1,219,200	777,920	441,926	1.76	353,646	3,009,246
H124	1,248,257	801,444	447,543	1.79	419,573	2,951,628
H127	1,185,708	761,343	425,003	1.79	391,436	3,141,121
H128	1,241,937	794,874	447,780	1.78	466,327	3,184,184
H129	957,898	616,699	341,634	1.81	279,240	2,833,161
H132	1,437,446	924,849	513,538	1.80	557,239	3,095,963
H135	1,045,106	668,389	377,190	1.77	302,418	3,079,892
H137	800,210	515,595	284,907	1.81	177,357	2,604,970
H138	1,170,150	746,082	424,690	1.76	349,740	3,155,380
H140	1,003,083	640,588	362,928	1.77	265,428	3,133,289
H142	946,771	609,850	337,348	1.81	283,838	2,703,634

表 3.6 SNP 数据统计表 (续)

Sample ID	SNP Number	Transition	Transversion	Ts/Tv	Heterozygosity Number	Homozygosity Number
H143	1,316,995	846,144	471,719	1.79	481,778	3,234,321
H144	1,300,357	837,038	464,061	1.80	435,025	2,943,770
H145	1,131,404	724,950	407,027	1.78	348,932	3,174,201
H146	1,084,945	698,225	387,289	1.80	346,763	2,852,541
H148	1,022,150	656,583	366,058	1.79	318,491	2,794,913
H149	1,009,633	645,372	364,723	1.77	275,529	3,103,113
H150	1,645,245	1,056,405	590,067	1.79	675,233	3,108,762
H151	1,456,227	934,931	522,348	1.79	543,444	3,125,172
H152	1,039,824	671,178	369,176	1.82	351,094	2,960,205
H157	1,090,479	696,784	394,252	1.77	357,248	3,065,918
H158	932,887	596,384	336,935	1.77	257,581	2,935,650
H159	1,189,565	766,109	424,203	1.81	401,479	3,105,191
H160	958,159	616,465	342,105	1.80	253,184	2,874,274
H161	1,126,010	717,362	409,199	1.75	343,161	3,065,642
H162	1,244,137	796,828	448,138	1.78	429,092	3,250,928
H164	1,576,207	1,011,742	565,596	1.79	594,579	3,200,078
H166	1,169,015	749,576	420,113	1.78	361,713	3,109,474
H167	1,023,700	657,969	366,177	1.80	275,667	2,732,167
H168	1,283,016	826,518	457,295	1.81	495,039	3,055,720
H169	1,118,635	717,944	401,212	1.79	323,234	3,048,966
H170	890,831	569,611	321,651	1.77	236,495	2,905,653
H171	1,386,263	892,740	494,616	1.80	546,125	3,180,814
H172	847,466	548,852	298,954	1.84	215,804	2,646,352
H173	1,001,126	642,214	359,402	1.79	318,625	2,763,557
H174	1,153,925	737,156	417,303	1.77	297,795	2,897,723
H175	1,027,102	654,840	372,781	1.76	268,203	2,932,255
H176	1,357,660	875,526	483,017	1.81	538,228	3,025,964
H177	1,194,030	767,005	427,692	1.79	362,657	2,980,631
H178	1,050,555	673,119	378,016	1.78	312,352	3,056,580
H179	1,147,366	741,915	406,114	1.83	339,101	2,947,237
MJ5	3,174,875	2,040,899	1,139,682	1.79	2,099,013	2,571,255

表 3.6 SNP 数据统计表 (续)

Sample ID	SNP Number	Transition	Transversion	Ts/Tv	Heterozygosity Number	Homozygosity Number

注：
Sample ID：样品编号；
SNP Number：检测到的单核苷酸多态性位点的数量，表示材料与参考基因组之间的核苷酸变异；
Transition：转换的 SNP 数量；
Transversion：颠换的 SNP 数量；
Ts/Tv：转换型 SNP（Transition）和颠换型 SNP（Transversion）的比值；
Heterozygosity Number：杂合分型的 SNP 位点总数；
Homozygosity Number：纯合分型的 SNP 位点总数。

3.3.2 SNP 功能注释

采用 SnpEff（Cingolani *et al.* 2012）程序结合本项目枣（Ziziphus_jujuba）基因组注释信息，对检测到的 SNP 进行功能注释，SnpEff 会根据基因组的基因和功能区域的分布进行分析，对每个 SNP 所在的位置和功能进行统计，并对每个变异类型的功能进行统计。表**??**为 SNP 位置分布信息统计表，表**??**为 SNP 功效信息统计表。

完整的 *SNP* 功能信息统计详见结题文件夹目录*data_release/01.vcf_filter*下的 *pop.snp_anno.xls* 文件

表 3.7 全基因组区域 SNP 位置分布信息统计表

Sample ID	Start Lost	Stop Lost	Stop Gained	Missense Variant	Synonymous Variant	Intergenic Region
FJMS	268	464	3,275	119,411	79,928	2,288,257
H1	100	207	1,363	53,927	37,337	908,400
H3	93	165	989	40,638	27,228	693,832
H3J1	110	195	1,228	49,106	33,535	799,112
H3J2	113	170	1,058	44,888	31,088	809,061
H3J3	91	172	1,035	43,788	30,206	731,620
H3J4	91	179	1,192	47,169	32,859	794,659
H3J5	161	279	1,791	69,115	46,843	1,194,576
H3J6	109	203	1,291	52,100	36,135	983,375
H3J7	146	238	1,590	61,778	41,894	1,045,455
H3J8	120	201	1,344	53,255	36,150	924,052
H3J10	98	188	1,157	47,443	32,263	819,365
H3J11	146	226	1,491	58,076	39,515	965,573
H3J12	125	225	1,455	55,689	38,248	1,001,469
H3J13	130	238	1,487	57,812	39,202	1,000,128

表 3.7 全基因组区域 SNP 位置分布信息统计表 (续)

Sample ID	Start Lost	Stop Lost	Stop Gained	Missense Variant	Synonymous Variant	Intergenic Region
H3J14	121	243	1,442	58,256	40,055	1,037,979
H3J15	141	235	1,574	61,077	41,138	1,087,316
H3J16	88	208	1,322	52,958	36,774	944,117
H3J18	123	211	1,359	52,741	35,715	951,295
H3J19	106	231	1,389	53,970	36,062	928,178
H3J20	113	221	1,345	52,807	36,389	957,101
H3J21	91	152	1,065	42,319	28,999	692,978
H3J22	101	179	1,157	47,066	32,593	807,450
H3J23	98	204	1,222	49,362	33,247	820,965
H4	110	194	1,166	47,533	33,105	882,612
H5	123	241	1,680	62,306	41,677	1,055,223
H8	86	184	1,127	45,123	30,969	849,013
H9	95	196	1,215	47,496	32,367	792,413
H10	159	258	1,837	70,389	47,648	1,251,032
H11	171	249	1,687	66,809	45,316	1,201,473
H13	103	218	1,349	53,207	35,936	1,062,125
H15	114	201	1,292	52,876	36,441	846,028
H16	104	212	1,387	55,508	38,022	1,016,643
H17	94	186	1,152	47,656	33,302	866,468
H18	120	211	1,461	56,326	38,092	1,064,229
H20	99	229	1,415	55,317	37,887	948,800
H21	124	231	1,572	58,614	39,408	1,019,332
H22	144	255	1,711	63,310	42,819	1,119,679
H23	115	221	1,394	52,129	34,927	892,793
H25	115	214	1,258	51,656	34,849	885,633
H26	99	188	1,296	51,896	35,926	988,649
H28	114	221	1,380	54,486	37,287	1,073,779
H30	133	249	1,592	62,536	42,108	1,172,812
H31	113	210	1,395	55,555	38,464	1,124,378
H36	138	262	1,681	65,771	45,114	1,151,981
H37	85	156	1,021	42,676	29,925	790,245
H38	126	221	1,485	58,874	40,619	1,190,341
H39	119	217	1,384	53,761	36,306	927,082
H41	150	247	1,673	64,750	44,378	1,280,024
H42	134	277	1,770	69,166	46,680	1,232,141
H43	121	229	1,505	58,169	39,709	949,572
H44	157	314	1,827	69,721	46,621	1,281,678

表 3.7 全基因组区域 SNP 位置分布信息统计表 (续)

Sample ID	Start Lost	Stop Lost	Stop Gained	Missense Variant	Synonymous Variant	Intergenic Region
H45	144	224	1,589	61,671	41,667	1,110,147
H46	76	172	984	40,565	28,409	727,202
H47	117	224	1,377	56,064	38,320	932,167
H49	123	222	1,423	55,376	38,030	963,240
H50	130	228	1,501	58,680	39,790	1,087,536
H51	136	236	1,533	58,904	40,092	1,014,571
H52	138	254	1,622	63,936	43,221	1,095,092
H54	143	242	1,484	61,887	41,300	1,050,501
H55	138	226	1,591	61,944	41,825	1,071,762
H56	129	196	1,385	54,604	37,357	948,613
H58	103	199	1,306	50,458	34,932	888,029
H59	102	189	1,306	48,847	33,270	926,205
H61	122	215	1,301	51,290	34,628	874,666
H62	99	194	1,201	51,013	35,201	930,060
H63	146	248	1,682	64,636	43,280	1,129,081
H64	117	205	1,337	50,938	33,884	881,401
H70	128	258	1,641	64,620	43,216	1,136,883
H74	76	175	1,052	44,445	31,062	808,158
H76	110	227	1,403	57,547	38,705	972,647
H77	134	224	1,574	61,517	41,820	1,086,662
H80	97	199	1,172	45,042	30,535	810,299
H81	163	303	1,905	71,105	47,993	1,270,451
H82	108	172	1,112	44,366	30,492	813,154
H83	169	289	1,985	75,405	51,075	1,314,510
H84	108	172	1,176	45,737	31,919	898,549
H85	135	242	1,656	60,797	40,620	1,071,240
H87	130	271	1,642	63,267	42,583	1,073,153
H89	120	260	1,529	58,575	39,532	1,057,719
H91	97	195	1,186	47,707	32,614	838,699
H92	109	182	1,185	50,242	34,999	1,003,690
H93	98	207	1,307	49,993	33,875	846,020
H94	164	299	1,914	76,428	51,507	1,251,445
H95	115	217	1,402	54,404	36,862	978,770
H96	128	225	1,533	59,251	39,841	1,014,269
H97	93	181	1,129	45,616	31,293	873,382
H100	124	252	1,622	64,144	43,814	1,101,567
H102	117	237	1,554	59,592	40,642	1,064,380

表 3.7 全基因组区域 SNP 位置分布信息统计表 (续)

Sample ID	Start Lost	Stop Lost	Stop Gained	Missense Variant	Synonymous Variant	Intergenic Region
H104	157	273	1,756	67,362	45,742	1,209,592
H105	89	170	1,108	44,750	30,708	767,769
H106	129	218	1,361	53,429	36,318	1,000,735
H108	103	173	1,203	46,738	31,715	850,738
H109	78	179	1,184	46,733	31,885	901,806
H112	124	244	1,618	63,614	43,197	1,138,986
H113	137	244	1,561	62,136	42,533	1,131,280
H115	169	301	2,002	74,325	49,615	1,350,179
H116	110	227	1,361	54,893	37,385	978,368
H117	139	269	1,746	65,468	43,842	1,114,158
H120	116	198	1,409	50,848	34,185	965,210
H121	96	194	1,213	48,521	33,661	955,552
H124	112	247	1,393	54,967	37,052	972,888
H127	101	213	1,460	54,276	36,929	915,256
H128	108	224	1,403	54,309	37,095	964,937
H129	98	155	1,063	41,357	27,873	748,967
H132	138	262	1,722	64,735	43,816	1,117,229
H135	101	191	1,122	46,504	31,420	810,470
H137	80	161	918	37,636	26,286	614,750
H138	105	200	1,227	48,869	33,447	913,790
H140	102	174	1,098	45,910	30,863	772,077
H142	117	185	1,108	45,289	30,458	729,062
H143	124	221	1,516	57,779	38,912	1,025,923
H144	135	226	1,442	56,867	38,501	1,016,544
H145	110	209	1,274	50,787	34,764	873,071
H146	92	208	1,295	49,304	33,035	842,436
H148	99	192	1,146	46,893	32,012	790,277
H149	84	169	1,109	45,354	30,743	778,565
H150	165	306	1,817	70,737	47,772	1,290,172
H151	140	260	1,602	63,572	42,977	1,136,457
H152	130	210	1,318	51,648	34,965	794,663
H157	85	180	1,121	45,708	31,238	850,228
H158	81	167	1,014	40,781	27,888	721,167
H159	117	214	1,482	56,224	37,952	919,363
H160	97	171	1,116	46,332	32,520	731,468
H161	78	208	1,147	48,298	33,962	873,765
H162	124	233	1,478	55,283	37,411	962,990

表 3.7 全基因组区域 SNP 位置分布信息统计表 (续)

Sample ID	Start Lost	Stop Lost	Stop Gained	Missense Variant	Synonymous Variant	Intergenic Region
H164	149	283	1,850	71,758	48,228	1,223,218
H166	102	209	1,271	49,457	33,839	913,184
H167	104	197	1,230	49,001	34,146	781,487
H168	140	215	1,468	57,148	38,862	1,000,394
H169	102	189	1,223	49,172	33,952	869,262
H170	83	151	941	40,772	27,831	684,126
H171	136	253	1,698	63,697	42,839	1,074,018
H172	106	163	1,158	45,146	31,158	639,656
H173	95	184	1,180	46,331	31,786	768,954
H174	99	180	1,158	47,741	33,407	898,295
H175	76	178	1,033	41,495	28,524	805,942
H176	127	255	1,537	61,148	41,239	1,058,173
H177	102	206	1,339	52,022	35,297	934,161
H178	89	187	1,232	49,254	33,490	806,417
H179	120	218	1,318	51,743	35,308	894,942
MJ5	283	502	3,596	129,507	86,150	2,519,669

注：

Sample ID：样品编号；

Start Lost：由于 SNP 的突变导致启动子缺失的 SNP 位点个数；

Stop Lost：由于 SNP 的突变导致终止子突变的 SNP 位点个数；

Stop Gained：由于 SNP 的突变导致终止子获得的 SNP 位点个数；

Missense Variant：样本在外显子区域的错义突变的 SNP 位点个数；

Synonymous Variant：样本在外显子区域的同义突变的 SNP 位点个数；

Intergenic Region：样本在基因间隔区的 SNP 位点个数。

表 3.8 全基因组区域 SNP 功效信息统计表

Sample ID	High	Moderate	Low	Modifier
FJMS	5,421	119,411	89,142	2,703,739
H1	2,250	53,927	41,574	1,092,056
H3	1,730	40,638	30,358	831,475
H3J1	2,078	49,106	37,326	961,829
H3J2	1,850	44,888	34,678	970,572
H3J3	1,796	43,788	33,619	876,471
H3J4	2,031	47,169	36,593	954,269
H3J5	3,059	69,115	52,197	1,422,494
H3J6	2,212	52,100	40,234	1,168,329

表 3.8 全基因组区域 SNP 功效信息统计表 (续)

Sample ID	High	Moderate	Low	Modifier
H3J7	2,738	61,778	46,629	1,246,455
H3J8	2,306	53,255	40,243	1,098,086
H3J10	2,022	47,443	36,010	983,348
H3J11	2,526	58,076	43,890	1,148,159
H3J12	2,470	55,689	42,513	1,191,538
H3J13	2,533	57,812	43,613	1,186,962
H3J14	2,498	58,256	44,705	1,239,241
H3J15	2,704	61,077	45,784	1,292,316
H3J16	2,255	52,958	41,076	1,132,575
H3J18	2,331	52,741	39,796	1,129,451
H3J19	2,347	53,970	40,210	1,107,419
H3J20	2,305	52,807	40,564	1,143,409
H3J21	1,810	42,319	32,219	834,564
H3J22	1,995	47,066	36,192	966,538
H3J23	2,126	49,362	36,947	978,563
H4	2,058	47,533	37,113	1,058,643
H5	2,773	62,306	46,289	1,255,129
H8	1,930	45,123	34,499	1,010,558
H9	2,077	47,496	35,980	950,098
H10	3,068	70,389	53,071	1,482,942
H11	2,919	66,809	50,390	1,419,084
H13	2,345	53,207	40,341	1,256,450
H15	2,228	52,876	40,492	1,018,324
H16	2,336	55,508	42,457	1,213,288
H17	1,972	47,656	37,265	1,036,968
H18	2,486	56,326	42,602	1,261,383
H20	2,413	55,317	42,105	1,131,862
H21	2,582	58,614	43,876	1,212,232
H22	2,865	63,310	47,703	1,328,887
H23	2,381	52,129	38,935	1,063,407
H25	2,219	51,656	38,760	1,055,394
H26	2,186	51,896	40,123	1,174,972
H28	2,367	54,486	41,739	1,277,352
H30	2,680	62,536	46,858	1,384,937
H31	2,383	55,555	43,045	1,333,367
H36	2,866	65,771	50,253	1,378,099
H37	1,767	42,676	33,409	945,943
H38	2,556	58,874	45,432	1,412,043

表 3.8 全基因组区域 SNP 功效信息统计表 (续)

Sample ID	High	Moderate	Low	Modifier
H39	2,366	53,761	40,375	1,102,661
H41	2,855	64,750	49,611	1,510,217
H42	2,975	69,166	52,008	1,464,909
H43	2,555	58,169	44,153	1,134,453
H44	3,133	69,721	52,056	1,518,641
H45	2,699	61,671	46,452	1,316,154
H46	1,695	40,565	31,612	871,817
H47	2,358	56,064	42,591	1,118,584
H49	2,450	55,376	42,246	1,146,361
H50	2,568	58,680	44,336	1,291,489
H51	2,629	58,904	44,580	1,208,524
H52	2,766	63,936	48,140	1,308,753
H54	2,586	61,887	46,032	1,248,754
H55	2,680	61,944	46,604	1,282,893
H56	2,347	54,604	41,594	1,129,353
H58	2,200	50,458	38,870	1,057,447
H59	2,164	48,847	37,190	1,098,679
H61	2,273	51,290	38,539	1,044,128
H62	2,077	51,013	39,291	1,113,413
H63	2,869	64,636	48,238	1,343,563
H64	2,277	50,938	37,770	1,049,090
H70	2,816	64,620	48,064	1,351,477
H74	1,797	44,445	34,666	967,505
H76	2,446	57,547	43,129	1,159,433
H77	2,656	61,517	46,469	1,287,208
H80	1,992	45,042	34,114	965,197
H81	3,186	71,105	53,436	1,507,338
H82	1,901	44,366	34,015	972,451
H83	3,291	75,405	56,824	1,558,698
H84	1,971	45,737	35,770	1,069,841
H85	2,761	60,797	45,268	1,272,651
H87	2,773	63,267	47,352	1,278,819
H89	2,590	58,575	43,987	1,250,466
H91	2,056	47,707	36,190	996,879
H92	2,117	50,242	39,170	1,191,667
H93	2,194	49,993	37,649	1,010,177
H94	3,318	76,428	57,367	1,498,171
H95	2,364	54,404	41,044	1,162,239

表 3.8 全基因组区域 SNP 功效信息统计表 (续)

Sample ID	High	Moderate	Low	Modifier
H96	2,592	59,251	44,466	1,211,844
H97	1,961	45,616	34,951	1,037,370
H100	2,757	64,144	48,818	1,318,900
H102	2,571	59,592	45,337	1,267,760
H104	2,970	67,362	50,954	1,437,000
H105	1,897	44,750	34,142	919,432
H106	2,329	53,429	40,728	1,192,239
H108	2,080	46,738	35,441	1,019,043
H109	2,029	46,733	35,731	1,074,632
H112	2,767	63,614	48,290	1,364,034
H113	2,675	62,136	47,441	1,349,450
H115	3,346	74,325	55,281	1,598,462
H116	2,370	54,893	41,692	1,169,377
H117	2,939	65,468	48,839	1,329,817
H120	2,357	50,848	38,113	1,140,579
H121	2,068	48,521	37,732	1,137,147
H124	2,389	54,967	41,323	1,157,120
H127	2,383	54,276	41,144	1,095,709
H128	2,368	54,309	41,400	1,151,033
H129	1,830	41,357	31,090	888,995
H132	2,879	64,735	48,723	1,330,144
H135	1,993	46,504	34,991	967,717
H137	1,606	37,636	29,178	736,216
H138	2,126	48,869	37,348	1,088,113
H140	1,921	45,910	34,436	926,575
H142	1,905	45,289	33,815	871,692
H143	2,554	57,779	43,390	1,221,362
H144	2,506	56,867	42,757	1,205,864
H145	2,139	50,787	38,806	1,046,417
H146	2,153	49,304	36,787	1,003,049
H148	1,997	46,893	35,624	943,530
H149	1,871	45,354	34,291	934,152
H150	3,133	70,737	53,253	1,528,104
H151	2,773	63,572	47,901	1,350,679
H152	2,241	51,648	38,874	953,925
H157	1,946	45,708	34,945	1,013,806
H158	1,735	40,781	31,225	864,378
H159	2,481	56,224	42,130	1,096,318

表 3.8 全基因组区域 SNP 功效信息统计表 (续)

Sample ID	High	Moderate	Low	Modifier
H160	1,902	46,332	36,153	879,709
H161	1,996	48,298	37,918	1,044,229
H162	2,538	55,283	41,816	1,152,100
H164	3,115	71,758	53,708	1,457,705
H166	2,163	49,457	37,720	1,086,210
H167	2,122	49,001	38,012	940,648
H168	2,488	57,148	43,233	1,187,527
H169	2,119	49,172	37,688	1,036,006
H170	1,664	40,772	31,044	822,542
H171	2,893	63,697	47,570	1,281,118
H172	1,915	45,146	34,459	771,798
H173	2,020	46,331	35,461	923,351
H174	2,030	47,741	37,318	1,072,999
H175	1,768	41,495	31,928	957,402
H176	2,686	61,148	45,847	1,256,273
H177	2,278	52,022	39,180	1,107,490
H178	2,072	49,254	37,291	968,237
H179	2,276	51,743	39,217	1,061,094
MJ5	5,946	129,507	96,034	2,966,415

注：
Sample ID：样品编号；
High：具有破坏性影响，可能导致蛋白质功能丧失；
Moderate：该类变异可能改变蛋白质的有效性；
Low：该类变异大部分无害，不太可能改变蛋白质；
Modifier：非编码变异或影响非编码基因的变异。

3.3.3 InDel 检测

利用 GATK 的 Best Practices 流程处理比对结果（BAM 文件），利用 GATK 的 Haplotyper 方法进行 InDel 检测及过滤，过滤条件按照 GATK 推荐的参数进行，具体可见：https://software.broadinstitute.org/gatk/documentation/article.php?id=3225。

对项目样品进行 InDel 标记开发，这里的 InDel 指能够明确获得序列组成的 InDel 标记。最终样本获得 Insertion 和 Deletion 详情统计结果如表**??**所示：

表 3.9 InDel 数据统计表

Sample ID	Insert Number	Delete Number	Heterozygosity Number	Homozygosity Number
FJMS	227,289	254,914	356,651	125,552
H1	74,279	84,974	36,806	122,447
H3	56,230	64,698	27,437	93,491
H3J1	64,373	73,749	35,856	102,266
H3J2	63,069	71,846	30,607	104,308
H3J3	54,438	62,859	26,728	90,569
H3J4	58,676	67,736	29,583	96,829
H3J5	98,762	112,867	81,612	130,017
H3J6	78,819	90,431	55,527	113,723
H3J7	86,039	98,931	58,667	126,303
H3J8	72,451	83,901	48,548	107,804
H3J10	65,934	75,451	34,997	106,388
H3J11	75,865	86,337	49,035	113,167
H3J12	78,784	90,492	46,011	123,265
H3J13	79,056	91,333	50,850	119,539
H3J14	85,010	96,636	48,309	133,337
H3J15	90,363	103,381	65,671	128,073
H3J16	77,112	88,780	46,110	119,782
H3J18	77,089	87,573	59,122	105,540
H3J19	75,487	86,599	50,144	111,942
H3J20	78,485	89,445	43,806	124,124
H3J21	52,241	60,025	23,020	89,246
H3J22	60,216	69,121	31,732	97,605
H3J23	63,622	73,488	35,001	102,109
H4	71,101	80,896	40,796	111,201
H5	86,709	99,254	60,577	125,386
H8	67,005	76,203	35,889	107,319
H9	64,222	73,131	31,101	106,252
H10	105,869	120,811	81,264	145,416
H11	98,463	112,096	66,503	144,056
H13	86,330	97,929	55,585	128,674
H15	67,868	77,414	41,145	104,137
H16	84,011	95,371	54,706	124,676
H17	70,980	80,760	39,581	112,159
H18	88,967	100,585	58,867	130,685
H20	75,437	86,527	44,064	117,900
H21	83,961	95,756	59,816	119,901

表 3.9 InDel 数据统计表 (续)

Sample ID	Insert Number	Delete Number	Heterozygosity Number	Homozygosity Number
H22	95,733	110,083	80,464	125,352
H23	72,893	84,049	46,235	110,707
H25	73,109	83,247	46,689	109,667
H26	79,030	90,092	47,032	122,090
H28	88,935	101,085	57,680	132,340
H30	98,952	112,070	67,553	143,469
H31	92,476	104,228	56,856	139,848
H36	100,712	114,118	71,341	143,489
H37	63,876	73,370	30,855	106,391
H38	100,274	113,022	59,287	154,009
H39	75,893	86,512	43,905	118,500
H41	108,447	122,898	82,137	149,208
H42	104,804	119,528	78,029	146,303
H43	76,941	88,078	52,003	113,016
H44	107,632	122,876	77,615	152,893
H45	92,499	105,509	69,761	128,247
H46	57,498	65,324	26,621	96,201
H47	76,212	87,151	39,590	123,773
H49	78,458	90,202	56,616	112,044
H50	92,750	105,570	64,085	134,235
H51	83,956	95,416	59,232	120,140
H52	95,025	108,879	68,547	135,357
H54	85,538	97,330	53,787	129,081
H55	92,758	106,330	67,327	131,761
H56	78,736	89,684	58,705	109,715
H58	73,445	83,596	44,250	112,791
H59	75,721	86,426	46,302	115,845
H61	72,587	83,001	44,026	111,562
H62	76,323	87,658	43,272	120,709
H63	95,166	108,676	65,167	138,675
H64	71,780	81,908	46,344	107,344
H70	96,736	110,458	65,722	141,472
H74	62,209	71,198	29,563	103,844
H76	77,909	89,232	47,659	119,482
H77	87,001	100,122	57,711	129,412
H80	64,896	73,176	37,979	100,093
H81	110,651	125,897	87,565	148,983

表 3.9 InDel 数据统计表 (续)

Sample ID	Insert Number	Delete Number	Heterozygosity Number	Homozygosity Number
H82	65,323	74,278	35,288	104,313
H83	107,428	122,459	80,707	149,180
H84	70,369	80,019	38,663	111,725
H85	88,851	101,793	65,012	125,632
H87	87,138	100,107	58,108	129,137
H89	84,801	96,527	56,816	124,512
H91	66,605	76,431	48,569	94,467
H92	81,780	93,574	44,813	130,541
H93	68,780	79,039	44,829	102,990
H94	103,683	118,314	81,047	140,950
H95	79,354	90,171	52,050	117,475
H96	86,043	98,207	55,233	129,017
H97	68,310	77,884	36,897	109,297
H100	95,312	108,992	75,140	129,164
H102	89,143	101,588	66,507	124,224
H104	102,237	116,512	80,733	138,016
H105	60,509	68,498	30,448	98,559
H106	86,722	98,500	50,713	134,509
H108	69,802	79,399	39,604	109,597
H109	71,804	82,242	39,890	114,156
H112	94,700	107,925	61,817	140,808
H113	96,015	108,814	59,725	145,104
H115	116,206	133,012	101,995	147,223
H116	79,075	89,620	46,041	122,654
H117	95,579	109,021	73,247	131,353
H120	80,234	91,319	53,894	117,659
H121	76,915	87,247	43,873	120,289
H124	80,214	91,301	53,393	118,122
H127	76,607	87,535	50,775	113,367
H128	78,393	89,061	55,464	111,990
H129	58,066	67,175	32,467	92,774
H132	93,270	106,672	68,456	131,486
H135	64,249	73,036	33,550	103,735
H137	42,660	48,867	16,627	74,900
H138	74,194	84,369	40,786	117,777
H140	63,571	71,867	30,999	104,439
H142	56,796	64,629	31,727	89,698

表 3.9 InDel 数据统计表 (续)

Sample ID	Insert Number	Delete Number	Heterozygosity Number	Homozygosity Number
H143	86,253	98,383	61,749	122,887
H144	81,677	93,606	54,535	120,748
H145	69,405	79,384	39,746	109,043
H146	67,440	77,795	41,750	103,485
H148	64,380	73,064	38,244	99,200
H149	62,962	72,070	32,349	102,683
H150	108,771	123,567	90,095	142,243
H151	94,397	107,756	68,234	133,919
H152	62,358	71,525	40,018	93,865
H157	64,942	74,400	39,072	100,270
H158	53,654	61,035	26,089	88,600
H159	74,064	85,138	47,579	111,623
H160	53,246	60,797	24,784	89,259
H161	66,850	76,769	36,888	106,731
H162	78,442	90,168	50,927	117,683
H164	103,321	118,527	75,413	146,435
H166	71,893	82,412	42,201	112,104
H167	54,620	62,594	26,469	90,745
H168	79,906	91,482	59,342	112,046
H169	67,782	77,408	34,522	110,668
H170	50,810	58,570	24,590	84,790
H171	88,020	101,410	68,335	121,095
H172	45,857	52,891	21,199	77,549
H173	60,812	68,469	36,688	92,593
H174	68,293	78,427	32,922	113,798
H175	62,784	70,972	30,962	102,794
H176	83,880	96,861	64,945	115,796
H177	74,687	85,368	42,402	117,653
H178	64,812	74,640	37,182	102,270
H179	70,069	80,129	39,125	111,073
MJ5	239,681	269,722	346,512	162,891

注：

Sample ID：样品编号；

Insert Number：检测到的插入变异的位点个数；

Delete Number：检测到的缺失变异的位点个数；

Heterozygosity Number：杂合分型的 InDel 的位点个数；

Homozygosity Number：纯合分型的 InDel 位点个数。

3.3.4 InDel 功能注释

采用 SnpEff 程序结合本项目枣（Ziziphus_jujuba）基因组注释信息，对检测到的 InDel 进行功能注释，SnpEff 会根据基因组的基因和功能区域的分布进行分析，对每个 InDel 所在的位置和功能进行统计，并对每个变异类型的功能进行统计。表**??**为 InDel 位置分布信息统计表，表**??**为 InDel 功效信息统计表：

完整的 *InDel* 功效信息统计表详见结题文件夹目录下*data_release/01.vcf_filter*文件夹下的 *pop.indel_anno.xls* 文件

表 3.10 全基因组区域 InDel 位置分布信息统计表

Sample ID	Frameshift Variant	Intergenic Region	Intragenic Variant	Start Lost	Stop Gained	Stop Lost
FJMS	6,431	397,716	0	89	186	72
H1	2,053	124,701	0	29	64	33
H3	1,544	94,891	0	20	43	16
H3J1	1,957	107,514	0	26	59	28
H3J2	1,351	106,147	0	15	47	20
H3J3	1,672	91,382	0	13	41	17
H3J4	1,442	98,347	0	15	37	19
H3J5	3,354	166,544	0	38	95	42
H3J6	2,076	134,575	0	19	64	26
H3J7	2,961	145,519	0	33	93	40
H3J8	2,357	123,177	0	24	54	27
H3J10	1,708	111,035	0	22	46	21
H3J11	2,602	127,480	0	33	89	32
H3J12	2,193	133,792	0	27	53	30
H3J13	2,608	134,403	0	29	85	29
H3J14	2,187	143,142	0	30	59	36
H3J15	2,816	153,564	0	34	79	42
H3J16	1,929	131,078	0	31	62	30
H3J18	2,361	130,539	0	22	63	37
H3J19	2,511	127,426	0	26	73	25
H3J20	1,970	132,704	0	29	53	18
H3J21	1,410	87,575	0	22	36	16
H3J22	1,551	101,408	0	27	37	27
H3J23	2,168	107,222	0	27	67	27
H4	1,562	119,399	0	25	45	20
H5	2,835	147,279	0	27	77	36
H8	1,564	114,023	0	20	36	22
H9	1,862	108,077	0	29	44	20
H10	3,328	180,702	0	43	89	35
H11	3,173	167,515	0	47	79	38

表 3.10 全基因组区域 InDel 位置分布信息统计表 (续)

Sample ID	Frameshift Variant	Intergenic Region	Intragenic Variant	Start Lost	Stop Gained	Stop Lost
H13	1,959	147,460	0	29	49	20
H15	2,080	112,835	0	20	58	36
H16	2,039	142,114	0	27	53	19
H17	1,730	120,068	0	31	49	23
H18	2,375	151,884	0	26	60	29
H20	2,363	127,391	0	26	62	30
H21	2,622	142,388	0	30	75	38
H22	3,055	163,802	0	43	95	31
H23	2,363	123,796	0	26	65	26
H25	2,277	123,634	0	28	66	27
H26	1,875	134,502	0	33	56	25
H28	1,862	152,122	0	26	59	30
H30	2,842	169,191	0	38	76	34
H31	1,957	157,619	0	34	68	27
H36	2,841	170,386	0	35	82	38
H37	1,426	108,925	0	14	31	18
H38	2,056	171,420	0	29	56	26
H39	2,455	128,142	0	25	74	37
H41	2,730	186,208	0	31	77	37
H42	3,139	178,656	0	40	75	35
H43	2,630	128,936	0	31	70	28
H44	3,104	183,920	0	43	90	41
H45	2,808	157,510	0	33	69	32
H46	1,160	97,439	0	21	29	19
H47	2,271	128,342	0	23	68	21
H49	2,492	133,078	0	18	63	24
H50	2,640	157,996	0	29	74	27
H51	2,685	141,376	0	34	65	33
H52	2,889	161,502	0	27	81	42
H54	2,815	144,829	0	21	77	26
H55	2,606	157,248	0	43	68	35
H56	2,407	133,218	0	20	72	36
H58	2,169	124,347	0	29	45	20
H59	2,006	129,046	0	31	69	22
H61	2,387	122,889	0	25	66	37
H62	1,718	129,909	0	22	42	23
H63	3,024	161,393	0	36	80	50

表 3.10 全基因组区域 InDel 位置分布信息统计表 (续)

Sample ID	Frameshift Variant	Intergenic Region	Intragenic Variant	Start Lost	Stop Gained	Stop Lost
H64	2,249	121,302	0	27	53	30
H70	2,984	164,361	0	43	76	39
H74	1,264	105,843	0	24	36	16
H76	2,485	131,293	0	37	66	33
H77	2,874	148,264	0	33	78	31
H80	1,749	108,928	0	14	60	31
H81	3,318	188,292	0	46	87	29
H82	1,355	111,023	0	20	40	24
H83	3,515	182,357	0	44	89	45
H84	1,372	120,335	0	21	44	26
H85	2,776	151,236	0	31	79	30
H87	2,989	147,500	0	39	84	38
H89	2,605	144,525	0	28	80	33
H91	2,052	113,492	0	27	58	26
H92	1,532	140,662	0	21	42	29
H93	2,087	116,544	0	24	43	24
H94	3,620	172,856	0	35	97	40
H95	2,247	134,348	0	24	72	23
H96	2,722	145,281	0	33	77	35
H97	1,448	116,480	0	19	39	24
H100	2,876	161,536	0	34	88	37
H102	2,544	151,090	0	33	69	34
H104	3,240	174,115	0	39	77	45
H105	1,578	101,415	0	26	55	18
H106	2,226	147,770	0	34	54	26
H108	1,842	118,106	0	21	48	21
H109	1,485	122,580	0	17	44	17
H112	2,579	159,045	0	33	87	30
H113	2,597	162,435	0	36	72	29
H115	3,624	199,410	0	48	98	41
H116	2,120	132,557	0	24	57	32
H117	3,127	161,841	0	36	86	42
H120	2,210	138,010	0	18	57	30
H121	1,565	131,200	0	19	43	25
H124	2,506	135,868	0	26	76	22
H127	2,384	128,954	0	31	66	22
H128	2,241	132,353	0	24	58	27

表 3.10 全基因组区域 InDel 位置分布信息统计表 (续)

Sample ID	Frameshift Variant	Intergenic Region	Intragenic Variant	Start Lost	Stop Gained	Stop Lost
H129	1,695	99,383	0	17	41	22
H132	3,049	157,950	0	42	81	33
H135	1,866	108,360	0	25	59	23
H137	1,151	71,113	0	11	32	16
H138	1,735	126,315	0	18	44	28
H140	1,797	106,490	0	27	47	27
H142	1,955	95,012	0	21	53	26
H143	2,698	146,361	0	26	78	38
H144	2,526	139,233	0	37	74	38
H145	1,873	116,769	0	26	51	26
H146	2,173	114,691	0	29	53	21
H148	2,061	108,181	0	27	61	27
H149	1,736	106,314	0	29	38	21
H150	3,226	185,391	0	41	81	36
H151	2,933	161,022	0	34	92	38
H152	2,395	103,449	0	24	56	28
H157	1,530	109,981	0	16	47	19
H158	1,239	90,035	0	20	39	17
H159	2,478	124,619	0	28	68	29
H160	1,461	87,957	0	22	50	21
H161	1,512	113,445	0	25	49	19
H162	2,304	132,870	0	37	58	33
H164	3,395	175,333	0	37	85	39
H166	1,887	122,303	0	25	56	22
H167	1,446	90,028	0	18	44	30
H168	2,656	134,950	0	28	80	38
H169	1,900	114,655	0	27	44	29
H170	1,380	85,253	0	11	34	16
H171	3,061	149,327	0	31	93	37
H172	1,837	74,837	0	25	45	26
H173	1,855	100,547	0	25	48	28
H174	1,473	115,342	0	19	49	31
H175	1,373	106,568	0	18	38	24
H176	2,802	142,385	0	35	70	44
H177	2,225	127,242	0	26	64	25
H178	1,963	108,928	0	27	67	29
H179	2,323	118,105	0	27	55	24

表 3.10 全基因组区域 InDel 位置分布信息统计表 (续)

Sample ID	Frameshift Variant	Intergenic Region	Intragenic Variant	Start Lost	Stop Gained	Stop Lost
MJ5	7,066	422,538	0	97	207	92

注：
Sample ID：样品编号；
Frameshift Variant：导致移码突变的 InDel 个数
Intergenic Region：在基因间区的 InDel 的个数；
Intragenic Variant：在基因内非功能区的 InDel 个数；
Start Lost：由于 InDel 的突变导致启动子缺失的 InDel 位点个数及所占比例；
Stop Gained：由于 InDel 的突变导致终止子获得的 InDel 位点个数及所占比例；
Stop Lost：由于 InDel 的突变导致终止子缺失的 InDel 位点个数及所占比例。

表 3.11 全基因组区域 InDel 功效信息统计表

sampleID	HIGH	MODERATE	LOW	MODIFIER
MJ5	7,445	3,231	1,638	516,080
FJMS	6,776	3,021	1,525	485,691
H13	2,074	985	624	181,108
H157	1,608	803	490	136,547
H158	1,321	676	390	112,209
H159	2,620	1,217	612	155,030
H160	1,569	762	475	111,137
H161	1,602	819	485	140,719
H162	2,438	1,107	592	164,838
H164	3,558	1,569	824	216,932
H166	2,018	957	547	151,066
H167	1,550	798	428	114,461
H168	2,807	1,227	627	167,271
H15	2,199	1,060	554	141,685
H169	2,004	902	514	141,855
H170	1,480	717	421	106,692
H171	3,217	1,370	693	185,038
H172	1,925	903	440	95,403
H173	1,960	920	498	125,896
H174	1,573	804	520	143,854
H175	1,459	677	437	131,093
H176	2,936	1,339	631	176,452
H177	2,335	1,008	547	156,330
H178	2,065	984	534	136,074

表 3.11 全基因组区域 InDel 功效信息统计表 (续)

sampleID	HIGH	MODERATE	LOW	MODIFIER
H16	2,154	1,040	633	176,088
H179	2,449	1,148	572	146,203
H3J1	2,085	960	485	134,711
H3J2	1,438	660	410	132,451
H3J3	1,762	815	452	114,300
H3J4	1,541	780	470	123,588
H3J5	3,516	1,515	798	206,814
H3J6	2,188	1,007	598	165,934
H3J7	3,111	1,378	671	180,398
H3J8	2,480	1,137	535	152,362
H3J10	1,798	872	512	138,325
H17	1,831	893	507	148,701
H3J11	2,761	1,247	635	157,976
H3J12	2,294	1,041	605	165,674
H3J13	2,735	1,220	625	166,183
H3J14	2,325	1,108	611	178,047
H3J15	2,964	1,327	674	189,481
H3J16	2,037	982	574	162,657
H3J18	2,489	1,113	595	160,860
H3J19	2,631	1,205	597	158,003
H3J20	2,080	1,034	567	164,461
H3J21	1,486	771	387	109,503
H18	2,499	1,075	634	186,033
H3J22	1,634	769	504	126,367
H50	2,793	1,301	660	194,166
H51	2,832	1,305	642	175,235
H52	3,042	1,365	714	199,699
H54	2,958	1,290	668	178,496
H4	1,655	827	526	149,155
H55	2,738	1,259	699	195,179
H56	2,538	1,178	613	164,575
H58	2,292	1,058	538	153,401
H59	2,126	941	527	158,764
H61	2,492	1,027	563	151,794
H62	1,818	887	535	161,018
H63	3,190	1,444	732	199,303
H64	2,360	1,068	573	149,948
H70	3,134	1,443	740	202,511

表 3.11 全基因组区域 InDel 功效信息统计表 (续)

sampleID	HIGH	MODERATE	LOW	MODIFIER
H74	1,332	680	420	130,817
H5	3,000	1,297	668	181,607
H76	2,610	1,151	610	163,140
H77	3,000	1,354	662	182,764
H80	1,841	840	509	134,983
H81	3,493	1,665	806	232,073
H82	1,452	663	462	137,132
H83	3,696	1,653	790	225,388
H84	1,463	724	483	147,871
H85	2,920	1,304	652	186,483
H3J23	2,270	998	547	133,330
H20	2,501	1,198	570	157,940
H21	2,763	1,289	637	175,546
H22	3,215	1,451	714	201,531
H23	2,460	1,159	578	152,989
H25	2,405	1,118	532	152,464
H1	2,167	1,060	584	155,623
H26	1,978	882	560	166,089
H28	1,980	946	589	187,174
H30	2,978	1,324	665	207,096
H31	2,078	948	638	193,737
H36	3,018	1,333	726	210,995
H37	1,506	736	456	134,619
H38	2,185	1,027	642	210,220
H39	2,580	1,148	599	158,253
H41	2,879	1,313	721	227,889
H42	3,311	1,495	745	220,048
H3	1,636	826	461	117,888
H43	2,767	1,309	612	160,622
H44	3,286	1,397	774	226,379
H45	2,948	1,426	675	193,699
H46	1,230	624	382	120,333
H47	2,376	1,123	609	159,596
H49	2,618	1,226	569	164,523
H87	3,137	1,407	706	182,641
H89	2,743	1,262	630	177,279
H8	1,653	739	489	140,395
H91	2,164	991	513	139,574

表 3.11 全基因组区域 InDel 功效信息统计表 (续)

sampleID	HIGH	MODERATE	LOW	MODIFIER
H92	1,642	790	561	172,702
H93	2,192	1,075	510	144,291
H94	3,787	1,709	872	216,991
H95	2,355	1,071	583	165,879
H96	2,852	1,326	646	179,977
H97	1,542	721	463	143,600
H100	3,041	1,378	700	200,293
H102	2,693	1,304	674	186,747
H104	3,389	1,480	750	214,179
H9	1,958	905	488	133,974
H105	1,679	803	487	125,988
H106	2,349	1,080	606	181,737
H108	1,938	864	484	146,143
H109	1,567	801	482	151,317
H112	2,733	1,266	748	198,759
H113	2,759	1,237	686	200,779
H115	3,809	1,722	843	244,684
H116	2,236	1,063	570	165,172
H117	3,294	1,470	748	200,140
H120	2,330	1,064	538	168,068
H10	3,487	1,560	769	222,113
H121	1,662	855	479	161,598
H124	2,628	1,184	599	167,507
H127	2,511	1,175	613	160,221
H128	2,357	1,081	521	163,846
H129	1,785	813	421	122,313
H132	3,210	1,419	714	195,285
H135	1,951	927	479	134,083
H137	1,210	602	360	89,229
H138	1,834	882	514	155,411
H140	1,896	837	482	132,155
H11	3,325	1,475	712	205,929
H142	2,075	913	434	117,954
H143	2,835	1,268	644	180,596
H144	2,647	1,196	636	171,284
H145	1,986	960	525	145,435
H146	2,285	1,012	525	141,569
H148	2,172	973	508	133,906

表 3.11 全基因组区域 InDel 功效信息统计表 (续)

sampleID	HIGH	MODERATE	LOW	MODIFIER
H149	1,818	881	447	131,977
H150	3,370	1,518	778	228,111
H151	3,072	1,332	674	197,913
H152	2,517	1,077	539	129,972

注：
Sample ID：样品编号；
High：具有破坏性影响，可能导致蛋白质功能丧失；
Moderate：该类变异可能改变蛋白质的有效性；
Low：该类变异大部分无害，不太可能改变蛋白质；
Modifier：非编码变异或影响非编码基因的变异。

3.4 遗传图谱构建

3.4.1 遗传标记筛选及分析

基于遗传学基本原理，对所有的 SNP 和 InDel 标记进行筛选，获得符合遗传图谱构建的分子标记，筛选和处理标准如下：

1. 保留两个亲本测序深度均超过 5X 的变异位点，如果某个子代在当前位点测序深度低于 2x，则将子代基因型在该位点基因型定义为缺失
2. 针对 F2、BC 或其他近交群体，选择双亲纯和且差异（aaxbb 型）的变异位点，并对变异位点进行分型重编码，用于后续图谱构建
3. 针对 F1 群体，选择至少有一个亲本杂合（abxcc，ccxab，abxcd，efxeg，hkxhk，nnxnp，lmxll）的变异位点，并对变异位点进行分型重编码用于后续图谱构建
4. 在完成上述处理的基础上，对于子代缺失超过 30% 或偏离孟德尔分离比例（偏分离，$p< 0.05$）的变异位点进行过滤，保留下来的数据作为图谱构建的分子标记

基于上述过滤标准，获得可以用于构建遗传图谱的分子标记，用于后续的遗传分析，在各个基因组序列中的分子标记数量如**??**所示。

表 3.12 标记筛选统计表

CHROM	SNP Number	INDEL Number
chr1	1,049	39
chr2	1,983	101
chr3	1,115	51
chr4	917	26
chr5	498	31
chr6	549	19
chr7	1,048	51

表 3.12 标记筛选统计表 (续)

CHROM	SNP Number	INDEL Number
chr8	831	26
chr9	1,001	21
chr10	961	23
chr11	564	26
chr12	718	27

注：
CHROM：染色体编号
SNP Number：检测到的单核苷酸多态性位点的数量；
INDEL Number：检测到的插入/缺失位点的数量；

3.4.2 连锁分群和图谱构建

根据标记之间的连锁关系和物理位置，一般确定连锁群个数与染色体个数一致，以 LOD > 5 为指标，将标记划分为对应的连锁群，使用统计学软件，将标记按照重组率线性排列，构建遗传图谱; 遗传图谱构建结果如图**??**所示:

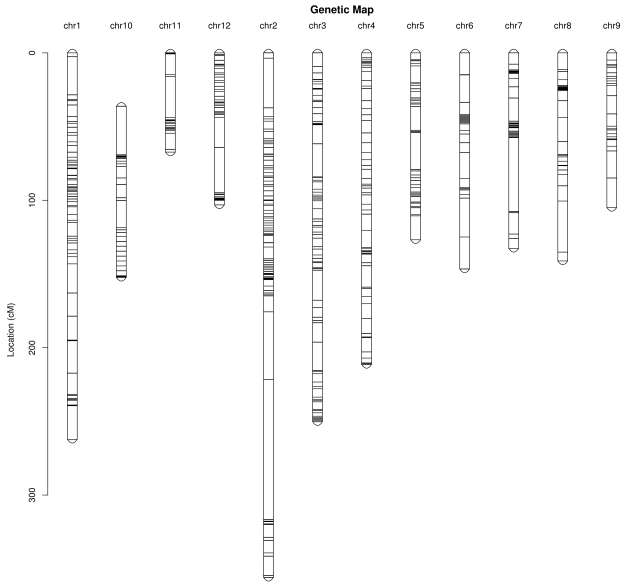

图 3.7 遗传图谱示意图

根据构建遗传图谱构建结果，进行遗传图谱质量统计，最终确定图谱评估，如**??**所示.