



基于两种对虾参考基因组的中国明对虾性别分化区域初步探查

王琼^{1,2}, 李健^{1,2}, 何玉英^{1,2*}

(1. 中国水产科学研究院黄海水产研究所, 农业农村部海洋渔业可持续发展重点实验室, 山东 青岛 266071;
2. 青岛海洋科学与技术国家实验室, 海洋渔业科学与食物产出过程功能实验室, 山东 青岛 266071)

摘要: 中国明对虾是我国最重要的养殖虾类之一。性成熟后的中国明对虾在生长和体色上表现出性别二态性。为了探查基因组上潜在的性别分化相关区域, 对中国明对虾进行了全基因组重测序研究。分别使用中国明对虾基因组和其近缘物种凡纳滨对虾的基因组作为参考基因组, 计算了雌雄中国明对虾之间的遗传分化指数 F_{st} , 以筛查基因组上的性别分化相关区域。当使用中国明对虾参考基因组时, 有约 89% 的中国明对虾测序读段全部或部分比对到中国明对虾参考基因, 覆盖度大约为 84%。基因组上 F_{st} 出现较多杂乱的峰, 其中 LG5 的 F_{st} 值整体呈升高的趋势, 另外 LG13、LG17 及 LG35 均有 F_{st} 指数相对较高的区域。而当使用凡纳滨对虾参考基因组时, 有约 60% 的中国明对虾测序读段全部或部分比对到凡纳滨对虾参考基因, 覆盖度约为 52%。Scaffold 2550 和 Scaffold 3683 上分别有 2 个最显著的峰。将雌雄中国明对虾转录组测序数据与凡纳滨对虾的参考基因组比对后, 获得了中国明对虾的粗略表达谱, 挖掘出了丰富的差异表达基因。分析了前述 2 个 Scaffold 上的基因表达情况。在 Scaffold 2550 上, 有 2 个肌肉中的差异表达基因, 有 3 个性腺中的差异表达基因; 在 Scaffold 3683 上, 有 4 个性腺中的差异表达基因, 全部都是雄性高表达, 且都位于 F_{st} 峰值区域, 其中 2 个基因编码外周型苯二氮卓受体相关蛋白 1 (peripheral-type benzodiazepine receptor-associated protein 1), 参与类固醇激素生物合成, 推测这两个雄性高表达基因参与了雄虾的性别特异性生理机制, 与雄性生殖功能相关, 在性别分化过程中发挥一定作用。本研究初步筛选到了与中国明对虾性别分化相关的区域和基因, 为中国明对虾性别二态性的进一步研究提供了依据。

关键词: 中国明对虾; 性别分化; 基因组; 重测序; 转录组

中图分类号: Q 344; S 968.22

文献标志码: A

中国明对虾 (*Fenneropenaeus chinensis*) 属于节肢动物门 (Arthropoda) 甲壳动物亚门 (Crustacea) 软甲纲 (Malacostraca) 十足目 (Decapoda) 对虾科 (Penaeus) 对虾属 (*Penaeus*), 又称东方对虾, 俗

称对虾、明虾。主要分布于黄渤海 (包括朝鲜半岛西岸), 以及东海北部的嵎泗列岛和舟山群岛一带, 是黄渤海地区的主要经济虾类, 是我国 20 世纪 60—90 年代最主要的养殖和出口对象之一, 在国

收稿日期: 2021-07-13 修回日期: 2021-08-25

资助项目: 国家虾蟹产业技术体系专项 (CARS-48); 国家自然科学基金 (31902367)

第一作者: 王琼 (照片), 从事甲壳动物遗传育种研究, E-mail: wangqiong@ysfri.ac.cn

通信作者: 何玉英, 从事甲壳动物遗传育种研究, E-mail: heyy@ysfri.ac.cn



际市场上享有较高的声誉。对虾经济价值很高, 受利润和市场需求的驱动, 世界对虾渔业发展迅速, 对虾养殖业的发展已成为海水养殖业中最具有代表性的产业之一^[1]。

中国明对虾的性别二态现象比较突出, 在养殖过程中, 性成熟后的雌性中国明对虾体色偏青, 而雄性则偏黄, 有研究表明, 中国明对虾在雌性成体平均长度达到 15~22 cm, 体重 50~120 g 时, 雄体只有 14~17 cm, 体重 30~60 g, 雌虾平均比雄虾重 1 倍以上^[2]。上市规格、等级的不同, 会极大地影响养殖经济效益。随着虾蟹动物养殖业的快速发展, 对其性别决定和分化生物学问题的研究显得愈发重要。在物种进化过程中, 虾蟹等动物处于承前启后的重要位置, 与高等脊椎动物相比, 其性别决定机制具有原始性、多样性和可塑性^[3]。分子生物学和基因组学技术的快速发展为性别决定和分化机制研究提供了新的研究思路和方法。甲壳动物的性别决定机制和性别分化发育机制非常复杂, 不同种类间的差别很大^[4-5], 其研究具有重要的理论意义及应用价值。十足目的许多种类中已经发现了异形性染色体^[6], 然而迄今为止的研究无法证明或否定对虾有性染色体存在。不过即使没有异形性染色体, 也可能存在性染色体遗传机制, 一些研究表明, 对虾性别决定机制是属于遗传决定型的, 属等位基因控制^[7]。性别决定和分化是有性生殖生物雌雄性别形成的两个阶段, 虾蟹动物的性别决定和分化涉及遗传、发育和进化等多个领域的科学问题^[3]。部分虾蟹动物雌雄个体之间在生长速率、个体大小、品质等经济性状方面表现出明显的差异。因此, 探究虾蟹等的性别决定和分化机制, 不仅有助于丰富甲壳动物性别决定的基础理论知识, 同时还可以指导经济虾蟹动物单性育种和养殖技术研发, 促进水产养殖业发展。

相对于鱼类及其他模式物种, 虾蟹动物性别决定和分化的研究还处于相对落后的阶段。此外, 由于物种本身的特殊性(受精、胚胎生物学特性及孵化方式等), 一些可以在模式生物上广泛应用的分子生物学技术如显微注射、基因编辑等还难以应用于虾蟹动物。随着二代、三代高通量测序技术的快速发展, 对功能基因的研究由单一、小规模逐渐向大规模、批量化的组学研究转变。目前对于中国明对虾的研究主要集中于生产养殖等方面, 随着测序技术的发展, 对虾基因组学研究逐

渐发展。2019 年, 与中国明对虾同属对虾属的凡纳滨对虾 (*Litopenaeus vannamei*) 完成了基因组测序和染色体水平的组装, 获得了国际首个高质量对虾基因组参考图谱^[8], 为对虾基因组育种和分子改良工作提供了重要的基础平台。2021 年, 中国明对虾基因组被破译^[9], 加速了中国明对虾基因组学的研究。

目前关于对虾性别分化的研究多集中于凡纳滨对虾^[10-12], 而中国明对虾基因组性别分化区域的研究报道很少。本研究对雌雄中国明对虾进行了全基因组重测序与参考基因组比对, 分析其肌肉和性腺转录组测序数据, 检测分化区域的性别差异表达的基因, 计算群体分化指数, 寻找潜在的性别分化区域, 为后续的性别分化研究提供线索。

1 材料与方法

1.1 样品采集

随机选择 11 尾雄性和 10 尾雌性 4 月龄的“黄海 1 号”中国明对虾, 用于基于 BGISEQ 平台的二代基因组测序。采集肌肉组织, 使用传统的酚-氯仿方法提取 DNA。DNA 质量通过 NanoDrop 2000 分光光度计 (Thermo Fisher, 美国) 和琼脂糖凝胶电泳进行评估。使用 BGISEQ 平台分别对 21 尾虾的 DNA 样本进行双端测序, 插入片段为 150 bp。实验过程中操作人员严格遵守实验动物相关伦理规范。

1.2 基于 BGISEQ 平台的全基因组测序数据分析

通过 NGSQCToolkit 软件^[13] 的默认参数对测序原始数据进行质控。使用 BWA 软件^[14] 将 DNA 测序的高质量数据比对到中国明对虾参考基因组 (ASM1692082v1) 和凡纳滨对虾参考基因组 (ASM378908v1)。使用 Picard 工具包 (<https://github.com/broadinstitute/picard>) 对得到的 BAM 文件进行排序并去掉重复的读段。GATK 软件^[15] 用于查找 SNP, 使用 PoPoolation 2 软件^[16] 通过测序数据估算群体遗传参数 F_{st} , 该指数用于衡量由于遗传结构导致的种群分化^[17], 所用参数: 滑动窗口大小为 10 kb, 滑动长度 5 kb, 最小等位基因计数为 4 个, 最小碱基质量 20, 最小覆盖深度为 10×, 最大覆盖深度为 200×。

1.3 RNA-Seq 数据处理

RNA-seq 数据与已发表的中国明对虾全长转

录组研究^[18]中所用到的二代转录组数据相同。样品采集与测序过程: 选择 5 月龄的 15 尾雌虾和 15 尾雄虾用于基于 Illumina 平台的二代转录组测序。采集它们的性腺和肌肉, 使用 Trizol (Invitrogen, 美国) 按照制造商的标准流程提取总 RNA。RNA 质量通过 NanoDrop 2000 分光光度计 (Thermo Fisher, 美国) 和琼脂糖凝胶电泳进行评估。将相同性别和相同组织的每 5 个 RNA 样本混合到一个池中, 每个性别同一组织的 3 个样本池视为 3 个生物学重复。按照基因表达样品制备试剂盒 (Illumina, 美国) 的方案, 构建 12 个由 2 个组织和 2 个性别样品组成的文库, 并通过 Illumina NovaSeq S4 平台进行双端测序, 插入片段长度为 150 bp。

Fastq 格式的原始测序数据首先通过 perl 脚本进行处理。从原始数据中删除包含接头或 ploy-N 的读段以及低质量的读段。同时, 计算高质量数据的 Q20、Q30、GC 含量和序列重复水平。所有下游分析均基于高质量的测序数据。通过使用 STAR 工具 (v2.5.3a)^[19] 的默认参数, 将每个 RNA-seq 文库的高质量数据与凡纳滨对虾的参考基因组 (ASM378908v1) 进行比对, 并使用 Feature-count 软件^[20] 将基因表达情况进行定量, 采用 FPKM (fragments per kilobase million, 每百万片段中来自某基因每千碱基长度的片段数) 作为衡量转录本或基因表达水平的指标。读段计数由 edgeR 程序包 (v3.22.0)^[21] 进行校正并筛选差异表达基因。采用 FDR (false discovery rate, 错误发现率) 作为差异表达转录本筛选的指标, 在差异表达转录本检测过程中, 将 FDR<0.05 作为筛选标准。

2 结果

2.1 中国明对虾基因组测序数据与凡纳滨对虾参考基因组比对结果

中国明对虾测序数据质控后, 总共获得了 219.9 Gb 的高质量数据, 其中包含 104.5 Gb 的雌

性和 115.4 Gb 的雄性测序数据。质控后的测序数据约 90% 的读段被全部或部分比对到中国明对虾的参考基因组中 (表 1), 平均分布深度约为 55×, 覆盖度约为 84%; 约 60% 的读段被全部或部分比对到凡纳滨对虾的参考基因组中, 平均分布深度约为 20×, 覆盖度约为 52%。

2.2 通过选择信号筛选性别分化相关区域

比对结果用于评估不同性别中国明对虾在整个基因组上的遗传分化情况。2 个参考基因组的比对结果分别计算了雌雄之间的遗传分化指数 F_{st} , 以筛查基因组上的性别分化相关区域。当以中国明对虾基因组为参考时, 出现较多杂乱的峰, 其中连锁群 LG5 的 F_{st} 值整体呈升高的趋势 (图 1-a)。当以凡纳滨对虾基因组为参考时, Scaffold 2550 (登录号: NW_020870010.1) 和 Scaffold 3683 (登录号: NW_020871268.1) 上分别有 2 个最显著的峰 (图 1-b)。Scaffold 2550 的长度约为 880 kb, F_{st} 峰位于 250~450 kb 区域。Scaffold 3683 的长度约为 800 kb, 而 F_{st} 峰位于 450~730 kb 的区域。

2.3 RNA-seq 数据发现大量性别偏好表达的基因

将中国明对虾 RNA-seq 数据与凡纳滨对虾的参考基因组比对, 获得了中国明对虾的粗略表达谱。挖掘出了丰富的差异表达基因。在肌肉中, 有 543 个基因在雄性中高表达, 1 043 个基因在雌性中高表达; 在性腺中, 雄性高表达的基因数量为 1 004, 雌性高表达的基因数量为 1 112。

筛选出前述 2 个 Scaffold 上的性别差异表达基因 (表 2)。在 Scaffold 2550 上, 基因 *LOC113814658* (位置: 423 402~432 118 bp) 和 *LOC113814660* (位置: 456 470~481 711 bp) 在肌肉中差异表达, 其中, *LOC113814658* 是雌性高表达, 而 *LOC113814660* 是雄性高表达 (图 2)。在性腺中, 基因 *LOC113814640* (位置: 125 412~131 933 bp)、*LOC113814649* (位

表 1 中国明对虾基因组重测序数据比对情况

Tab. 1 Mapping profile of the *F. chinensis* DNA resequencing data

参考基因组 reference genome	性别 sex	比对比例/% fraction of mapped reads	平均深度 average depth	覆盖度/% coverage (>0×)	覆盖度/% coverage (≥4×)	覆盖度/% coverage (≥10×)
中国明对虾 <i>F. chinensis</i>	雌 female	89.95	55.08	83.89	78.54	71.77
	雄 male	89.37	55.62	83.92	78.49	71.79
凡纳滨对虾 <i>L. vannamei</i>	雌 female	60.33	19.30	52.22	37.42	26.40
	雄 male	57.72	20.77	52.68	37.53	26.74

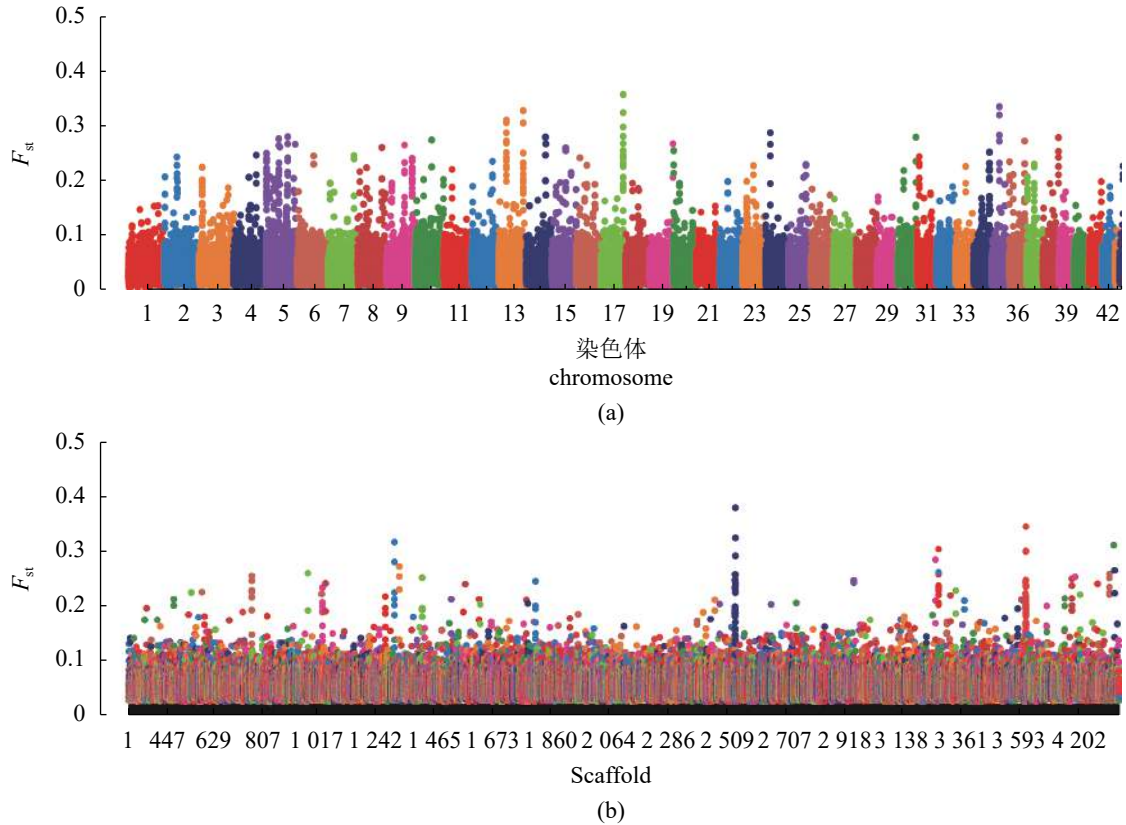


图 1 中国明对虾雌雄群体基因组上的 F_{st} 值

(a) 以中国明对虾基因组为参考, 不同颜色代表不同染色体; (b) 以凡纳滨对虾基因组为参考, 不同颜色代表不同 Scaffold。

Fig. 1 Value of F_{st} between male and female shrimps across the genome of *F. chinensis*

(a) genome of *F. chinensis* was used as reference, different colors representing different chromosomes; (b) genome of *L. vannamei* was used as reference, different colors representing different Scaffolds.

置: 78 831~81 919 bp) 和 *LOC113814660* (位置: 456 470~481 711 bp) 差异表达, 且均为雄性高表达。在 Scaffold 3683 上, 性腺中筛选出 4 个差异表达基因, 分别为 *LOC113825080* (位置: 583 785~644 864 bp)、*LOC113825079* (位置: 650 265~705 188 bp)、*LOC113825087* (位置: 711 662~716 375 bp) 和 *LOC113825086* (位置: 457 150~527 259 bp), 它们都是雄性高表达, 且都位于 F_{st} 峰值区域。此外, 在肌肉中未发现 Scaffold 3683 上的差异表达基因。

3 讨论

本研究同时使用中国明对虾和凡纳滨对虾基因组作为参考。中国明对虾和凡纳滨对虾是近缘物种, 并且都属于甲壳动物亚门对虾属, 因此基因组内有大量的同源序列。中国明对虾的基因组测序数据覆盖了凡纳滨对虾基因组的一半以上, 表明两种对虾存在相当比例的同源序列, 但是同

源性不高, 因此与凡纳滨对虾基因组比对的结果只能作为参考。RNA-seq 数据比对结果与比对到中国明对虾全长转录组的研究相似^[18], 在肌肉中, 雌性高表达的基因明显多于雄性高表达, 这些结果证实, 比对到近缘物种的参考基因组的方法有一定的有效性。在缺乏本物种高质量参考基因组的情况下, 近缘的参考基因组在某些研究中可以暂时进行替代, 但可能会有一些重要的信息被忽略。本研究旨在使用 F_{st} 指数查找基因组内暴露于选择压力下的可疑区间^[22], 即使错过了一些变异位点, 相邻区域中的其他位点也可以反映目标区域的变化, 因此以凡纳滨对虾基因组为参考的分析结果具有一定的参考价值。

由于中国明对虾参考基因组是基于 1 尾雄性个体的测序结果组装完成, 而本研究中雌雄中国明对虾重测序数据比对到参考基因组后, 结果显示雌雄群体不管是重测序数据的比对率还是在基因组上的覆盖度都没有显著差别, 推测其性染色

表 2 基因在凡纳滨对虾 Scaffold 2550 和 Scaffold 3683 上的表达信息

Tab. 2 Expression information of genes on Scaffold 2550 and Scaffold 3683

Scaffold	组织 tissue	基因编号 geneid	log ₂ (FC)	FDR	F1	F2	F3	M1	M2	M3
2550	肌肉 muscle	LOC113814658	-3.13	0.0150	13.58	8.72	14.47	1.36	1.21	1.12
		LOC113814660	3.09	0.0286	7.06	18.60	17.86	104.74	201.63	60.67
		LOC113814650	1.23	0.2757	0.61	2.08	1.43	3.74	5.16	0.74
		LOC113814645	1.46	0.3502	0.00	4.98	1.69	2.95	6.68	7.57
		LOC113814657	0.95	0.3622	0.48	2.41	0.26	2.27	2.12	1.49
		LOC113814651	0.70	0.4099	8.21	61.70	16.42	32.91	47.18	52.36
	性腺 gonad	LOC113814640	8.67	0.0029	0.00	0.00	0.00	8.46	7.10	9.76
		LOC113814649	7.39	0.0102	0.00	0.00	0.00	4.23	2.58	4.32
		LOC113814660	1.61	0.0488	6.40	17.90	12.07	43.68	43.40	28.05
		LOC113814657	2.25	0.2414	0.28	0.15	0.46	3.02	1.13	0.33
		LOC113814645	1.07	0.3373	2.78	0.92	2.75	3.17	3.55	3.77
		LOC113814650	-0.42	0.5201	29.21	14.84	29.19	11.64	23.23	9.09
		LOC113814651	0.33	0.6812	35.89	22.49	32.40	36.27	28.24	49.57
		LOC113814658	0.10	0.8996	23.65	25.24	29.19	30.38	22.27	38.59
3683	肌肉 muscle	LOC113825078	1.95	0.1682	0.00	1.91	0.85	4.54	3.34	3.60
		LOC113825079	-0.69	0.7911	0.61	0.50	1.24	2.04	0.00	0.00
	性腺 gonad	LOC113825080	5.77	0.0020	0.00	0.15	0.46	6.35	17.75	9.54
		LOC113825079	5.13	0.0039	0.00	1.07	2.14	20.10	53.57	38.48
		LOC113825087	4.17	0.0089	0.28	0.15	0.92	2.57	9.52	7.10
		LOC113825086	4.34	0.0407	0.00	0.00	0.15	1.06	1.77	0.89
		LOC113825089	2.66	0.1655	0.00	0.15	1.22	3.63	1.77	4.99
		LOC113825078	-1.33	0.4435	0.28	1.38	3.82	1.06	0.48	0.67

注: F1、F2、F3为3尾雌虾的FPKM, M1、M2、M3为3尾雄虾的FPKM。

Notes: F1, F2, and F3 refer to the FPKM of the 3 female shrimps, and similarly, M1, M2, M3 refer to the FPKM of the 3 male shrimps.

体的分化程度较小。根据达尔文的推测,造成自然界中广泛存在的性别二态性的性选择压力是一种不同于自然选择的力量^[23-24]。在性染色体形成的过程中,由于选择压力同时作用于性别决定区域和它附近的性别拮抗突变而抑制了染色体的重组,进而产生了不同形态的性染色体^[25-27]。但是,在许多物种中,性别决定基因座附近的重组抑制尚未遍及整条性染色体,它们不具有异形染色体,并且它们的性染色体仅显示有限的分化水平。

性别特异性进化压力作用于性别拮抗变异,预期会引起该变异位点及其邻近区域的分化。 F_{st} 指数可以反映基因组上这些区域的分化情况。中国明对虾 LG5 上的 F_{st} 值整体偏高,另外 LG13、LG17 及 LG35 均有 F_{st} 指数相对较高的区域;而凡纳滨对虾参考基因组 2 个 Scaffold 上发现了 2 个 F_{st} 峰,表明以上几个区域可能与中国明对虾性别分化相关。

基因通过性别偏好表达可导致雌性和雄性不同的形态^[28]。由于生物基因组中的遗传差异有限,基因表达在表型多样性的产生中起到重要作用,大多数性别二态性是由不同性别之间的基因表达差异引起的^[29-30]。有研究表明,Y 或 W 染色体上的等位基因转录衰退发生的非常迅速^[31-32],性别决定区域差异表达基因的富集发生在 Y 或 W 染色体退化之前^[33]。本研究提取了位于凡纳滨对虾 Scaffold 2550 和 Scaffold 3683 上基因的表达信息。由于性成熟的雌雄中国明对虾在体长和体质量上有明显差别,本研究选取与性别分化直接相关的性腺组织和与生长相关的肌肉组织进行研究。取样时间在 5 月龄,是中国明对虾性成熟的关键时期,雌雄个体在外型上开始出现明显差别。在 Scaffold 2550 上,筛选出了肌肉中 2 个差异表达基因和性腺中 3 个差异表达基因,其中基因 LOC113814660 在两个组织中均差异表达,并且都是雌性高表达,

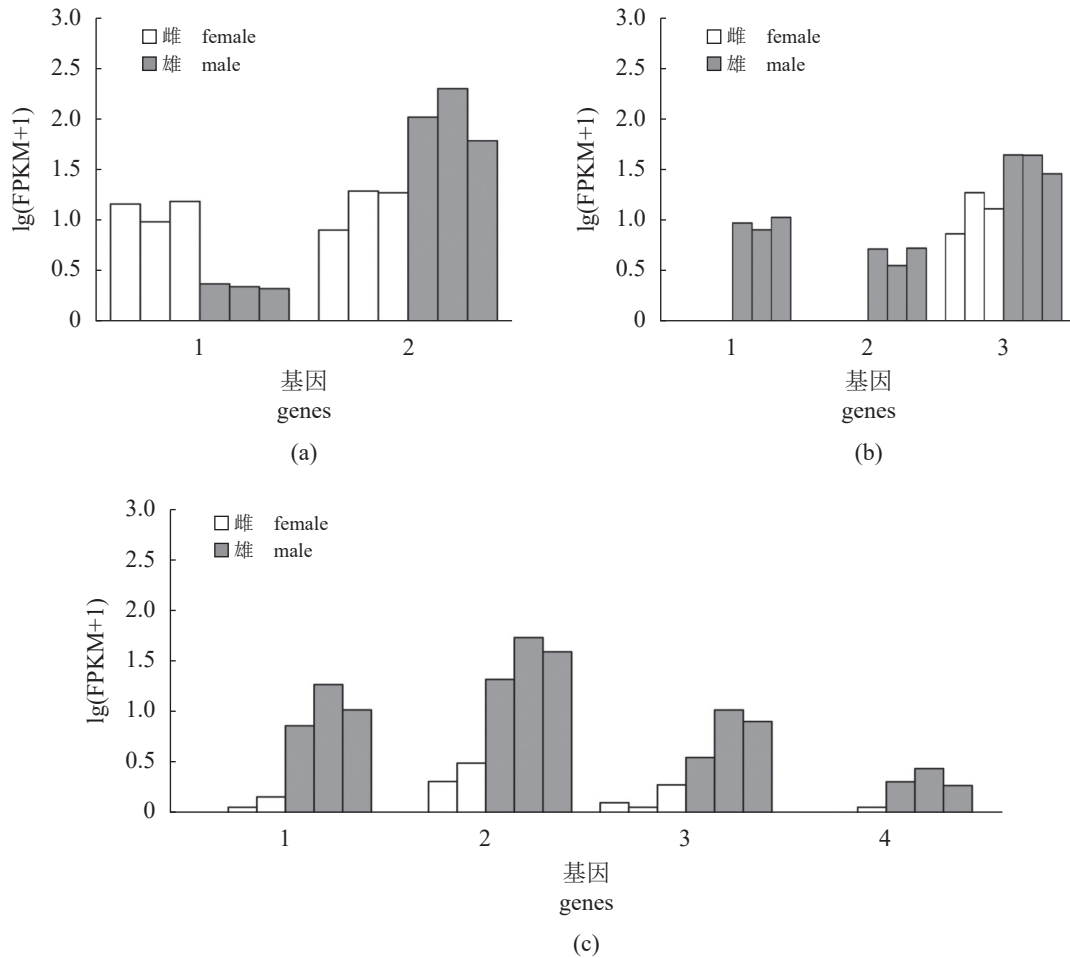


图 2 Scaffold 2550 和 Scaffold 3683 上差异表达基因的表达情况

(a) Scaffold 2550 肌肉, 1. *LOC113814658*, 2. *LOC113814660*; (b) Scaffold 2550 性腺, 1. *LOC113814640*, 2. *LOC113814649*, 3. *LOC113814660*;
(c) Scaffold 3683 性腺, 1. *LOC113825080*, 2. *LOC113825079*, 3. *LOC113825087*, 4. *LOC113825086*。

Fig. 2 Expression of DEGs on Scaffold 2550 and Scaffold 3683

(a) Scaffold 2550 muscle, 1. *LOC113814658*, 2. *LOC113814660*; (b) Scaffold 2550 gonad, 1. *LOC113814640*, 2. *LOC113814649*, 3. *LOC113814660*;
(c) Scaffold 3683 gonad, 1. *LOC113825080*, 2. *LOC113825079*, 3. *LOC113825087*, 4. *LOC113825086*。

能够编码 CBP80/20 依赖性翻译起始因子 (CBP80/20-dependent translation initiation factor)。在性腺中, 基因 *LOC113814640* 和 *LOC113814649* 只在雄性中表达, 雌性中不表达。它们分别编码富含亮氨酸重复序列的蛋白质 DDB_G0290503 (putative leucine-rich repeat-containing protein DDB_G0290503) 和乙酰透明质酸介导的运动受体 (hyaluronan mediated motility receptor), 推测它们是雄性特异性表达的基因。在 Scaffold 3683 上, 性腺中筛选出 4 个差异表达基因, 它们都是雄性高表达, 且都位于 F_{ST} 高峰。值得注意的是, 其中的 *LOC113825079* 和 *LOC113825087* 均编码外周型苯二氮卓受体相关蛋白 1 (peripheral-type benzodiazepine receptor-associated protein 1)。外周型苯二氮卓受体是一种线粒

体蛋白, 参与调节胆固醇从线粒体外膜到内膜的运输, 是类固醇激素生物合成的限速步骤^[34]。类固醇激素包括两大类: 肾上腺皮质激素和性激素。外周型苯二氮卓受体存在于许多外周组织中, 但在肾上腺皮质和睾丸间质细胞中表现出独特的高密度^[35]。因此, 推测这两个雄性高表达基因参与了雄虾的性别特异性生理机制, 并参与了中国明对虾的性别分化。

4 结论

本研究中, 通过对两种性别的中国明对虾进行全基因组重测序, 并与中国明对虾及其近缘物种凡纳滨对虾参考基因组进行比对, 寻找潜在的性别分化区域。根据 F_{ST} 分析定位到中国明对虾的

一个连锁群以及凡纳滨对虾两个片段对应的区域, 发现这些区域可能与性别分化相关。进一步分析雌雄中国明对虾的转录组测序数据, 检测到分化区域性别差异表达的基因, 筛选到疑似与雄虾性别分化相关的基因。本研究为中国明对虾性别相关研究提供了一定的研究思路和理论基础。

(作者声明本文无实际或潜在的利益冲突)

参考文献 (References):

- [1] 邓景耀. 对虾渔业生物学研究现状[J]. 生命科学, 1998, 10(4): 191-194,197.
Deng J Y. Research status of fishery biology of penaeus[J]. Chinese Bulletin of Life Sciences, 1998, 10(4): 191-194,197 (in Chinese).
- [2] 王兵. 中国对虾免疫相关基因和性别相关核酸片段研究 [D]. 北京: 中国科学院研究生院 (海洋研究所), 2003.
Wang B. The study of immune function related genes and sex linked nucleic acid marker of Chinese shrimp (*Fenneropenaeus chinensis*)[D]. Beijing: University of Chinese Academy of Sciences (The Institute of Oceanology), 2003 (in Chinese).
- [3] 张银, 林帆, 石西, 等. 虾蟹动物性别决定和分化机制研究现状与展望[J]. 汕头大学学报(自然科学版), 2018, 33(4): 3-14.
Zhang Y, Lin F, Shi X, *et al.* Research progress and prospect of sex determination and differentiation mechanism of shrimps and crabs[J]. Journal of Shantou University (Natural Science Edition), 2018, 33(4): 3-14 (in Chinese).
- [4] 王兵, 樊拥军, 张晓军, 等. 中国对虾性别相关片段的初步研究[J]. 高技术通讯, 2003, 13(5): 82-86.
Wang B, Fan Y J, Zhang X J, *et al.* The preliminary study on the sexual related DNA fragment in Chinese shrimp[J]. High Technology Letters, 2003, 13(5): 82-86 (in Chinese).
- [5] 吴长功, 相建海. 甲壳动物性别分化的遗传决定及外界因素的影响(英文)[J]. 发育与生殖生物学学报, 2002, 11(2): 137-144.
Wu C G, Xiang J H. Genetic determination and exogenous influence in sex differentiation in crustacean[J]. Developmental & Reproductive Biology, 2002, 11(2): 137-144 (in Chinese).
- [6] 康现江, 王所安. 高等甲壳动物性别决定机制及其性逆转[J]. 动物学杂志, 1998, 33(3): 43-46.
Kang X J, Wang S A. Sex-determination mechanism and sex-reversion of higher crustacean[J]. Chinese Journal of Zoology, 1998, 33(3): 43-46 (in Chinese).
- [7] 杨丛海, 王清印, 孔杰. 高温处理中国对虾受精卵对性比结构的影响[J]. 海洋科学, 1993(4): 1-2.
Yang C H, Wang Q Y, Kong J. Effects of high temperature treatment on sex ratio structure of fertilized eggs of *Penaeus chinensis*[J]. Marine Science, 1993(4): 1-2 (in Chinese).
- [8] Zhang X J, Yuan J B, Sun Y M, *et al.* Penaeid shrimp genome provides insights into benthic adaptation and frequent molting[J]. Nature Communications, 2019, 10(1): 356.
- [9] Yuan J B, Zhang X J, Wang M, *et al.* Simple sequence repeats drive genome plasticity and promote adaptive evolution in penaeid shrimp[J]. Communications Biology, 2021, 4(1): 186.
- [10] Zhang L S, Yang C J, Zhang Y, *et al.* A genetic linkage map of Pacific white shrimp (*Litopenaeus vannamei*): sex-linked microsatellite markers and high recombination rates[J]. Genetica, 2007, 131(1): 37-49.
- [11] Yu Y, Zhang X J, Yuan J B, *et al.* Identification of Sex-determining loci in pacific white shrimp *Litopenaeus vannamei* using linkage and association analysis[J]. Marine Biotechnology, 2017, 19(3): 277-286.
- [12] Wang Y, Yu Y, Li S H, *et al.* Sex-specific transcriptome sequencing of Zoea I larvae and identification of sex-linked genes using bulked segregant analysis in pacific white shrimp *Litopenaeus vannamei*[J]. Marine Biotechnology, 2020, 22(3): 423-432.
- [13] Patel R K, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data[J]. PLoS One, 2012, 7(2): e30619.
- [14] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform[J]. Bioinformatics, 2010, 26(5): 589-595.
- [15] McKenna A, Hanna M, Banks E, *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data[J]. Genome Research, 2010, 20(9): 1297-1303.
- [16] Kofler R, Pandey R V, Schlotterer C. PoPoolation2: identifying differentiation between populations using

- sequencing of pooled DNA samples (Pool-Seq)[J]. *Bioinformatics*, 2011, 27(24): 3435-3436.
- [17] Holsinger K E, Weir B S. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} [J]. *Nature Reviews Genetics*, 2009, 10(9): 639-650.
- [18] Wang Q, He Y Y, Li J. Conjoint analysis of SMRT- and Illumina-Based RNA-sequencing data of *Fenneropenaeus chinensis* provides insight into sex-biased expression genes involved in sexual dimorphism[J]. *Frontiers in Genetics*, 2019, 10: 1175.
- [19] Dobin A, Davis C A, Schlesinger F, *et al.* STAR: ultra-fast universal RNA-seq aligner[J]. *Bioinformatics*, 2013, 29(1): 15-21.
- [20] Liao Y, Smyth G K, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features[J]. *Bioinformatics*, 2014, 30(7): 923-930.
- [21] Robinson M D, McCarthy D J, Smyth G K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data[J]. *Bioinformatics*, 2010, 26(1): 139-140.
- [22] Weir B S, Cockerham C C. Estimating F -statistics for the analysis of population structure[J]. *Evolution*, 1984, 38(6): 1358-1370.
- [23] Tipton J A. Darwin's beautiful notion: sexual selection and the plurality of moral codes[J]. *History and Philosophy of the Life Sciences*, 1999, 21(2): 119-135.
- [24] Lande R. Sexual dimorphism, sexual selection, and adaptation in polygenic characters[J]. *Evolution*, 1980, 34(2): 292-305.
- [25] Bergero R, Charlesworth D. The evolution of restricted recombination in sex chromosomes[J]. *Trends in Ecology & Evolution*, 2009, 24(2): 94-102.
- [26] Bull J J. Sex determining mechanisms: an evolutionary perspective[J]. *Experientia*, 1985, 41(10): 1285-1296.
- [27] Wright A E, Dean R, Zimmer F, *et al.* How to make a sex chromosome[J]. *Nature Communications*, 2016, 7(1): 12087.
- [28] Mank J E, Nam K, Brunstrom B, *et al.* Ontogenetic complexity of sexual dimorphism and sex-specific selection[J]. *Molecular Biology and Evolution*, 2010, 27(7): 1570-1578.
- [29] Ellegren H, Parsch J. The evolution of sex-biased genes and sex-biased gene expression[J]. *Nature Reviews Genetics*, 2007, 8(9): 689-698.
- [30] Grath S, Parsch J. Sex-biased gene expression[J]. *Annual Review of Genetics*, 2016, 50: 29-44.
- [31] Papadopoulos A S T, Chester M, Ridout K, *et al.* Rapid Y degeneration and dosage compensation in plant sex chromosomes[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(42): 13021-13026.
- [32] Bachtrog D, Hom E, Wong K M, *et al.* Genomic degradation of a young Y chromosome in *Drosophila miranda*[J]. *Genome Biology*, 2008, 9(2): 30.
- [33] Pucholt P, Wright A E, Conze L L, *et al.* Recent Sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*[J]. *Molecular Biology and Evolution*, 2017, 34(8): 1991-2001.
- [34] Papadopoulos V. In search of the function of the peripheral-type benzodiazepine receptor[J]. *Endocrine Research*, 2004, 30(4): 677-684.
- [35] Snyder S H, Verma A, Trifiletti R R. The peripheral-type benzodiazepine receptor: a protein of mitochondrial outer membranes utilizing porphyrins as endogenous ligands[J]. *The FASEB Journal*, 1987, 1(4): 282-288.

Preliminary investigation of sex-linked loci of Chinese shrimp (*Fenneropenaeus chinensis*) based on the reference genome of two *Penaeus* shrimps

WANG Qiong^{1,2}, LI Jian^{1,2}, HE Yuying^{1,2*}

(1. Key Laboratory for Sustainable Utilization of Marine Fisheries Resources, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China;

2. Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China)

Abstract: *Fenneropenaeus chinensis* is one of the most commercially important cultured shrimps in China. It belongs to the family Penaeidae of Crustacea. *F. chinensis* exhibits a significant sexual dimorphism, the color of adult female shrimps tends to be blue, while the males tend to yellow, and the females were noticeably bigger in body size than males. However, no sex-determining region has been reported. A closely related species, Pacific white shrimp, *Litopenaeus vannamei*, which also belongs to the *Penaeus*, has been completely sequenced, and an assembled genome of *F. chinensis* was published recently, which accelerated the progress of genome research on shrimp culture. Sex-differentiation study has been carried out only in recent years, and this study aspires to screening out the potential sex-linked regions on the genome of *F. chinensis*. A total of 11 male and 10 female *F. chinensis* shrimps at 4-months-old were picked randomly, and the DNA samples of these 21 shrimps were extracted from muscle and sequenced using the BGISEQ platform individually. 104.5 Gb clean data of female and 115.4 Gb clean data of male were obtained, and the reference genome of *F. chinensis* (assembly ASM1692082v1) and *L. vannamei* (assembly ASM378908v1) were both used. The mapped result was used to evaluate the genetic differentiation between two sexes across the genome, and fixation index (F_{st}) was calculated across the genome. According to the results using reference genome of *F. chinensis*, there were many disorderly peaks of F_{st} on the genome, and the F_{st} value showed an overall increasing trend on LG5. According to the results using reference genome of *L. vannamei*, there were 2 most significant peaks located on scaffold 2550 (Accession: NW_020870010.1) and scaffold 3683 (Accession: NW_020871268.1), respectively. Many variants were found showing genetic differentiation between the two sexes. The 2 peaks of the F_{st} value on 2 scaffolds indicate that the corresponding regions on the genome of *F. chinensis* may be responsible for sex differentiation. RNA-seq data of gonad and muscle of female and male *F. chinensis* shrimps at 5-months-old were aligned to the reference genome of *L. vannamei* to obtain the expression information of genes on these two scaffolds. On scaffold 2550, two genes expressed differentially in muscle, while three genes expressed differentially in gonad. On scaffold 3683, no DEG was found in muscle. Four DEGs were filtered out in gonad, all of which had higher expression levels in males. Coincidentally, they all located on the range of F_{st} peak. Two DEGs encoded peripheral-type benzodiazepine receptor-associated protein 1, which is a key factor in steroid hormone biosynthesis. The peripheral-type benzodiazepine receptor was detected in many peripheral tissues but showed uniquely high densities in the adrenal cortex and in Leydig cells of the testes. Therefore, it is speculated that these male-biased expression genes were involved in the sex-specific physiological mechanism of male shrimps and participated in the sex differentiation of *F. chinensis*. In this study, we preliminarily located the sex-linked regions in genome of *F. chinensis* and filtered candidate genes related to the sex differentiation. The study provides a basis for further research of sexual dimorphism in *F. chinensis*.

Key words: *Fenneropenaeus chinensis*; sex differentiation; genome; resequencing; RNA-seq

Corresponding author: HE Yuying. E-mail: heyuying@ysfri.ac.cn

Funding projects: China Agriculture Research System (CARS-48); National Natural Science Foundation of China (31902367)