

مطالعه بیان ژن افتراقی زنبور عسل ملکه، نر و کارگر با استفاده از داده‌های RNA-seq

محمد رضا عطاری^۱، حسین مرادی شهر بابک^{۲*}، غلامعلی نهضتی پاقلعه^۲، محمد حسین بنابازی^۳ و محمد رضا هاشمی^۱
 ۱ و ۲. دانشجوی سابق کارشناسی ارشد و استادیار، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران
 ۳. استادیار، مؤسسه تحقیقات علوم دامی ایران، کرج، ایران
 (تاریخ دریافت: ۱۳۹۷/۸/۱۲ - تاریخ پذیرش: ۱۳۹۸/۱/۲۶)

چکیده

این پژوهش با هدف مطالعه پروفایل بیان ژن و تعیین ژن‌های شاخص در تمایز و تکامل ملکه، نر و کارگر با مقایسه تفریقی آن‌ها در سن ۵ لاروی یا همان سن تمایز انجام شد. لذا ترانسکریپتوم (توالی کل mRNA) ۱۵ نمونه از زنبور عسل نژاد ایتالیایی (*A. m. ligustica*) شامل ۵ زنبور نر، ۵ کارگر و ۵ ملکه از طریق هم‌ردیفی و مکان‌یابی خوانش‌های RNA-Seq بر روی ژنوم مرجع زنبور عسل نسخه Amel_4.5 مرتب شد. سپس آنالیز بیان افتراقی ژن صورت گرفت و در نهایت ۱۵۹۶۲ ژن و ۳۱۲۹۷ ایزوفرم بر روی ترانسکریپتوم این نمونه‌ها مشخص شد و نهایتاً ۴۶۵ ژن با بیان متفاوت بین تیمارهای نر و ملکه، ۴۹۵ ژن بین تیمارهای کارگر و ملکه و ۷۶۴ ژن بین تیمارهای نر و کارگر بود ($p < 0.000005$)، ($p < 0.00001$) و بیشترین تفاوت بیان ژنی بین تیمارهای کارگر و نر، ژن‌های GB42053 و GB45614 با لگاریتم ۲ fold change به ترتیب با مقادیر ۱۰- و ۱۱/۵ به دست آمدند و در مقایسه تیمارهای نر و ملکه ژن‌های GB45614 و GB48020 به ترتیب با مقادیر ۱۱/۷ و ۱۱/۸ و در مقایسه تیمارهای ملکه و کارگر ژن‌های GB43508 و GB42053 به ترتیب با مقادیر ۶/۶ و ۹- با بیشترین تفاوت بیان ژنی مشاهده شدند. آنالیز ماهیت ژن (GO) و مسیرهایی که در آن‌ها درگیر هستند نشان داد، که بررسی دقیقی و جامعی بر روی تعداد زیادی از این ژن‌ها انجام نشده است، ولی تعداد زیادی از این ژن‌های شاخص در مقایسه تیمارهای: ملکه و نر مرتبط با اجزای ساختاری و جدایی‌ناپذیر غشا، پردازش متابولیسم کیتین، بازدارنده‌های تریپسین، فعالیت گیرنده گاسترین، باندشونده با هورمون جوانی، پاسخ‌های دفاعی نسبت به باکتری‌ها، پروتئین‌های کوتیکول سفیره هستند در صورتی که در مقایسه ژن‌های شاخص متمایز در تیمارهای ملکه و کارگر مرتبط با مسیرهای متابولیسمی، متابولیسم داروها و سایر آنزیم‌ها، متابولیسم لیپیدها، باندشونده با نوکلئیک اسید، انتقال‌دهنده کلسترول داخل سلولی، پردازش‌های متابولیسمی کیتین، بیوسنتز یوئیکتون و سایر ترپنوئید کتون‌ها و ژن‌های متمایز در مقایسه کارگر و نر مرتبط با اجزای ساختاری و جدایی‌ناپذیری از غشای متابولیسم و انتقال آمینو اسیدها، باند شونده با یون کلسیم، باندشونده با فلز یون، انتقال‌دهنده کلسترول کیتین درون سلول، متابولیسم بودند.

واژه‌های کلیدی: بیان افتراقی ژن، ترانسکریپتوم، زنبور عسل، ماهیت ژن، RNA-Seq.

Study of differential gene expression in queen, drone and worker honey bee using RNA-seq data

Mohamadreza Attari¹, Hossein Moradi Shahre Babak^{2*}, Gholamali Nehzati Paghalee², Mohamad Hossein Banabazi³ and Mohamadreza Hashemi¹

1, 2. Former M. Sc. Student and Assistant Professor, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran
 3. Assistant Professor, Animal Science Research Institute of Iran (ASRI), Karaj, Iran
 (Received: Nov. 3, 2018 - Accepted: Apr. 15, 2019)

ABSTRACT

The aim of this study was to investigate the gene expression profile and find the important genes in the differentiation and evolution of the queen, worker and drone honey bee at Stage 5 Larvae. Therefore, transcripts (total mRNA sequence) of 15 samples of Italian honey bee (*A. m. ligustica*) including 5 drone, 5 worker and 5 queen bees were aligned to the reference genome of the honey bee. In gene expression analysis of RNA-Seq data, 15962 genes and 31297 isoforms were identified. In our differential gene expression analyses, 465 genes between drone and queen bees, 495 genes between worker and queen bees and 764 genes between drone and worker bees, were expressed differently ($P < 0.000005$). The largest difference in expression of genes was observed between drone and workers were for GB45614 and GB42053, with log2 fold change that was -10 and 11.5, respectively. In drone and queen bees comparison, GB45614 and GB48020 genes with log2 fold change 7.11 and 8.11, and in queen and worker bees comparison, GB43508 and GB42053 with log2 fold change 6.6 and -9, had the largest difference in gene expression, respectively. The analysis of the gene ontology (GO) and the pathways involved showed that the function of many of these genes has yet to be found. However, a large number of the genes expressed defiantly in drone and queen bees were related to integral component of membrane, calcium ion binding, carboxypeptidase, cholecystokinin receptor, chitin metabolic process, chymotrypsin inhibitor, haemolymph juvenile hormone binding and pupal cuticle protein, while differentially expressed genes in queen and worker bees comparison were related to metabolic pathways, enzymes metabolism, Pyrimidine metabolism, lipid metabolism, protein kinase, ATP binding and nucleic acid, intracellular cholesterol transport, chitin metabolic process, nitrogen compound metabolic process, hydrolase activity and chitin binding. The genes expressed at different levels in worker bees and drones were related to structural elements, the metabolism membrane and transfer of amino acids, calcium-ion-bound, ion-binder, intracellular cholesterol transport and chitin metabolism.

Keywords: Deferential gene expression, gene ontology, honey bee, RNA-Seq, transcriptome.

* Corresponding author E-mail: hmoradis@ut.ac.ir

مقدمه

زنبور عسل افزون بر تولید محصولات گوناگون مهمترین نقش خود را در طبیعت با دخالت در عمل گرده‌افشانی و افزایش تولید محصولات کشاورزی و احیای محیط زیست ایفا می‌کند. بررسی‌ها نشان می‌دهد در کشورهای مختلف نقش زنبور عسل در افزایش تولیدات کشاورزی ۶۹ تا ۱۴۳ برابر تولیدات مستقیم آن‌ها است (Tahmasebi *et al.*, 2000). بنابراین بررسی این حشره مفید از جهات مختلف فیزیولوژی، رفتارشناسی، ژنتیک و ... حائز اهمیت است. یک کلنی زنبور عسل از زنبورهای نر، کارگر و ملکه تشکیل شده است. تفاوت تغذیه در تخم لقاح یافته باعث ایجاد دو فنوتیپ ملکه و کارگر از یک ژنوتیپ یا سلول تخم میشود و لاروهای نر نیز از تخم‌های لقاح‌نیافته به وجود می‌آیند (Barchuk *et al.*, 2007). از این‌رو با بررسی ژن‌های متفاوت بیان شده می‌توان روند ایجاد این دو فنوتیپ از یک ژنوتیپ را مشخص نمود. بیان ژن مکانیسمی حساب‌شده، پیچیده و تنظیم‌گر است که باعث می‌شود ژنوتیپ به صورت فنوتیپ ظاهر شود (Banabazi *et al.*, 2012a). در واقع کدهای ژنتیکی که در مولکول DNA ذخیره شده‌اند به وسیله بیان ژن‌ها تفسیر می‌شوند و خصوصیات و نحوه بیان ژن باعث بوجود آمدن فنوتیپ در ارگانیسم خواهد شد (Banabazi *et al.*, 2012b). مطالعات بیان ژن به تولید حجم زیادی از داده‌ها در این زمینه منجر شده است در میان انواع متعدد داده‌های بیان ژن داده‌های RNA-seq به سبب پربروندا بودن و شامل شدن توالی نواحی رونویسی شده و پوشش دادن تمام این نواحی دارای مزیت عمده بر سایر انواع داده‌ها می‌باشد. در سال‌های اخیر امکانات رایانه‌ای آنالیز این حجم از داده‌ها را در زمان و با هزینه کم امکان‌پذیر ساخته است. از مزایای استفاده از داده‌های RNA-seq شناسایی ژن‌های جدید می‌باشد (Gan *et al.*, 2010; Sultan *et al.*, 2008). اساس این روش بر برش کل mRNA (ترانسکریپتوم) به قطعات کوچک و سپس توالی‌یابی همه این قطعات می‌باشد (Marguerat & Bahler, 2010; Haas & Zody, 2010). از این‌رو از روی تعداد خوانش‌های مکان‌یابی شده برای یک ناحیه خاص می‌توان یک معیار دقیقی از میزان رونویسی آن ناحیه رسید (Ekblom & Galindo, 2010).

(Wilhelm & Landry, 2009). شاخص بیان ژن در این روش محاسبه تعداد خوانش یا قطعات (کیلوباز) به ازای هریک میلیون خوانش مکان‌یابی شده (FPKM)^۱ می‌باشد (Mortazavi *et al.*, 2008). با توجه به این که تمایز فنوتیپی حاصل شده برای کارگر و ملکه تنها تفاوت محیطی (تغذیه یا عدم تغذیه با ژل رویال) می‌باشد و همان‌طور که مشخص است علی‌رغم خصوصیات اجتماعی کندو، عملکرد و راندمان کندو تا حد زیادی به ملکه وابسته می‌باشد. لذا شناسایی مجموعه ژن‌های مرتبط با این تمایز در سن ۳ تا ۶ روزگی لارو بسیار حائز اهمیت است، به این دلیل در این مطالعه به بررسی پروفایل بیان ژن زنبور عسل ایتالیایی نر، ملکه و کارگر پرداخته شد تا علاوه بر بررسی ژن‌های دخیل در تمایز جنسیت به کشف ژن‌های جدید بر روی ژنوم زنبور عسل بپردازیم که در واقع با تعیین نقش بسیاری از نواحی ژنوم و همچنین با بررسی ژن‌های دخیل در تمایز جنسیت قادر به تعیین برنامه‌های اصلاح نژادی و تسریع در روند آن شد.

ژنوم مرجع زنبور عسل توسط مرکز Baylor College of Medicine با استفاده از روش شات گان کل ژنوم (shotgun whole genome) انجام شد. توالی، اسمبلی و تجزیه و تحلیل ژنوم زنبور عسل توسط کنسرسیوم توالی‌یابی ژنوم زنبور عسل در اکتبر ۲۰۰۶ در نشریه نیچر به چاپ رسید. این توالی‌یابی توسط روش سنگر، دستگاه ۴۵۴ شرکت ROSCH و SOLiD انجام و برای نقشه‌یابی بهتر بهبود یافت. جدیدترین نسخه ژنوم مرجع زنبور عسل Amel_4.5 می‌باشد که در فوریه ۲۰۱۱ منتشر شده است. ژنوم مرجع زنبور عسل از ۲۲۹ میلیون جفت باز تشکیل شده است که دارای ۱۵۳۱۴ ژن کدکننده (Coding genes)، ۴۰۰ ژن غیر کدکننده^۲ (Non coding genes)، ۳۹۶ ژن کوچک غیرکدکننده^۳ (Small non coding genes)، ۴ ژن متفرقه غیرکدکننده^۴ (Misc non coding genes) و

1. Fragments Per Kilobase of transcript per Million mapped reads

۲. RNAهای دارای وظایف مشخصی هستند که به پروتئین ترجمه نمی‌شوند؛ شامل: pRNA، lncRNA، eRNA، gsRNA
 ۳. RNAهای دارای وظایف مشخصی هستند که به پروتئین ترجمه نمی‌شوند. معمولاً طولی کمتر از ۲۰۰ باز دارند و شامل miRNA، siRNA، snoRNA، piRNA
 ۴. ژنی که شبیه ژنی است که پروتئینی را کد می‌کند اما به علت frameshift یا کدون توقف موجب عدم توانایی در کدکنندگی آن می‌شود.

برای غیرفعال نمودن آنزیم RNAase آب استفاده می‌شود) رسوب داده شد. آماده‌سازی کتابخانه شامل: اتصال آداپتور، رونویسی معکوس و PCR است، که با پروتکل آماده‌سازی کتابخانه RNAها انجام شد (E7300S/L, NEB#). تأیید محصول و انتخاب اندازه کتابخانه cDNA مربوطه با استفاده از ژل اکریل آمید ۶ درصد انجام شد. نوارهای روی ژل با اندازه باندهای ۱۴۰-۱۵۰ جفت باز که برای اتصالات آداپتور RNA با قطعات ۲۱-۳۰ نوکلئوتیدی در نظر گرفته شده بودند از ژل جدا و RNA با انکوباسیون نوارها به مدت یک شب با محلول NaCl ۰/۳ مولار و دمای ۴ درجه سانتی‌گراد استخراج شدند. سازه کتابخانه به وسیله سدیم استات ۳ مولار، ایزوپروپانول سرد، قبل از این که در اتانول ۷۵ درصد شسته شود نهایتاً با پلیت در ۱۰ میکرولیتر بافر Tris-HCl با pH ۸/۵ حل شد. کتابخانه در بیوانالایزر ۲۱۰۰ با استفاده از DNA LabChip فوق حساس تعیین و اعتبارسنجی شد (Agilent Technologies, Integrated Sciences,) (Chatswood, Australia) و پس از انجام تمامی مراحل، برای انجام توالی‌یابی mRNAها از دستگاه illumina Hiseq 2500 استفاده شد. داده‌های خام در تاریخ ۶ سپتامبر ۲۰۱۵ در پایگاه اطلاعاتی NCBI در دسترس قرار گرفته است، و با فرمت sra به صورت رایگان و عمومی از لینک زیر قابل دریافت است.

<https://www.ncbi.nlm.nih.gov/sra/SRR1571716/> پس از اخذ داده‌ها، تمامی مراحل آماده‌سازی و تجزیه تحلیل آنها در محیط گرافیکی (GNU) سیستم عامل Linux Mint 17.3 Cinnamon نسخه ۶۴ بیت بر روی یک کامپیوتر شخصی غیر متصل به سرور انجام شدند. مراحل آنالیز شامل: تبدیل فرمت sra به fastq با استفاده از نرم‌افزار sratoolkits نسخه ۲,۸,۲ و سنجش کیفیت داده‌ها با استفاده از FastQC (Andrews, 2010). نسخه ۰,۱۱,۵ بود. سپس براساس نتایج سنجش کنترل کیفیت ویرایش خوانش‌ها با استفاده از نرم‌افزار Trimmomatic (Bolger et al., 2014). نسخه ۰,۳۶ انجام شد. این ویرایش شامل حذف آداپتورها و خوانش‌های بی‌کیفیت بود. این برنامه در محیط جاوا اجرا می‌شود و از منوی اصلی،

۱۵۷۲۰ رونوشت ژن Gene transcripts است ژنوم زنبور حاوی ۳۲ کروموزوم (n=۱۶) است.

مواد و روش‌ها

در این مطالعه از داده‌های RNA-Seq مربوط به ۱۵ نمونه زنبور *Apis mellifera ligustica* در مرحله پنچ لاروی (۱+/۹۱) شامل ۵ نمونه زنبور نر، ۵ نمونه زنبور کارگر و ۵ نمونه زنبور ملکه (از ۱۵ کندوی متفاوت) استفاده شد. این داده‌ها در مطالعه Regan Ashby et al. (2016) به دست آمده است. به منظور تولید داده‌های مذکور نمونه‌های لاروی با ۵۰۰ میکرولیتر معرف تریزول (تریزول به نمونه با نسبت ۳ به ۱) هموژنیزه شدند و قبل از این که نمونه‌ها سانتریفیوژ شوند مجدد ۱۰۰ میکرولیتر کلروفرم به محلول اضافه شد و نمونه‌ها اجازه داده شد که به مدت ۳ دقیقه در دمای اتاق انکوبه شدند و سپس نمونه‌ها در ۱۲۰۰۰g به مدت ۱۰ دقیقه در دمای ۴ درجه سانتی‌گراد سانتریفیوژ شدند و سپس فاز بالایی مایع جمع‌آوری و کل RNA با استفاده از ایزوپروپانول رسوب داده شد و محتوای پلیت با اتانول ۷۵ درصد شست‌وشو و در آب مقطر دو بار تقطیر با انکوباسیون در دمای ۸۰ درجه سانتی‌گراد به مدت دو تا سه دقیقه حل شد. خلوص RNA با ژل الکتروفورز تعیین و میزان کمی آن با استفاده از اسپکتروفتومتری نانودراپ تعیین شد. کتابخانه RNA با استفاده از کیت آماده NEBNext Multiplex برای دستگاه توالی‌یابی ایلومینا ساخته شد. (E7300S/L, New England BioLabs (NEB)) (Genesearch, Arundel, Qld, Australia). به طور خلاصه RNAها بر اساس اندازه موردنظر (با استفاده از ۱۰ میکروگرم RNA کل (n=5) تعداد نمونه) و استفاده از ژل اکریل آمید ۱۵ درصد، ۸ مولار اوره و مارکر میکرو RNA خالص‌سازی شدند (N2102S, NEB#). RNAها براساس اندازه موردنظر و به وسیله انکوباسیون در محلول NaCl 0.3 مولار در مدت یک شب در دمای ۴ درجه سانتی‌گراد از ژل شسته شدند و بعد با استفاده از ایزوپروپانول/ اتانول رسوب داده شدند. محتوای پلیت در ۶/۵ میکرولیتر آب مقطر بدون RNAase دارای دی‌اتیل دی‌کربنات (DEPC-treated)

نوکلئوتیدها راجع به کیفیت طول کل خوانش تصمیم‌گیری می‌شود. به عبارت دیگر نمودار گفته شده، جمع‌بندی کیفیت داده‌ها بر اساس هر ۱۰ آزمون دیگر است. از نتایج هر یک از ۱۰ آزمون به صورت انفرادی نیز برای تعیین چگونگی انجام ویرایش استفاده شد. تعیین ترانسکریپتوم برای هر یک از نمونه‌ها براساس هم‌ردیفی خوانش‌ها و مکان‌یابی آن‌ها بر ژنوم مرجع به ترتیب با نرم‌افزار Bowtie2 و TopHat2 انجام گردید (Langmead & Salzberg, 2012). ژنوم رفرنس مورد استفاده برای تشکیل ترانسکریپتوم Amel_4.5 بود. اطلاعات حاشیه‌نویسی به کاررفته نیز مربوط به این ویرایش نیز موجود بود. ژنوم مرجع و حاشیه‌نویسی آن از بانک اطلاعاتی Ensemble دریافت شد.

ftp://ftp.ensemblgenomes.org/pub/metazoa/release-32/fasta/apis_mellifera/dna

ftp://ftp.ensemblgenomes.org/pub/metazoa/release-32/gtf/apis_mellifera

اندازه فاصله درون کتابخانه‌ای و انحراف معیار آن برای استفاده از کد TopHat2 با استفاده از (Li & Durbin, 2009) و به روش نمونه‌گیری از داده‌ها انجام شد و 50 ± 200 برآورد شد. خروجی فایل BAM حاصل از TopHat2 با استفاده از بسته SAMtools مرتب و Index شد. تشکیل پروفایل بیان ژن برای هر یک از تیمارها با استفاده از بسته نرم‌افزاری Cufflinks انجام شد. در این مجموعه نرم‌افزاری تلفیق دو ترانسکریپتوم با استفاده از Cuffdiff و آنالیز بیان افتراقی ژن با استفاده از Cuffdiff انجام شد. تیمارها به صورت دو به دو با یکدیگر مقایسه شدند و ۳ فایل بیان افتراقی برای ژن‌ها به دست آمد مقدار p-value تصحیح نشده تست آماری انجام شده برابر با 0.0005 بود از آنجاکه این مقدار به تنهایی قابل اعتماد نیست از مقدار "FDR-adjusted p-value" یا q-value نیز استفاده کردیم که مقدار آن برابر با 0.0001 بود برای تست معنی‌داری اگر مقدار p پس از تصحیح بنیامین-هوچبرگ برای تست چندگانه بیشتر از FDR بود آن‌گاه آن را معنی‌دار در نظر گرفتیم. مصورسازی نتایج در قالب گراف با استفاده از پکیج CummeRbund (Loyal et al., 2014) تحت نرم‌افزاری R انجام شد. آنالیز gene ontology به صورت آنلاین و با استفاده از بانک

سنجش کیفیت داده‌ها برای خوانش‌های جفتی یا دوسویه (رفت و برگشت) به طور جداگانه با ۱۰ آزمون مختلف انجام می‌شود. این آزمون‌ها عبارتند از کیفیت توالی به ازای tile^۱، امتیاز کیفیت به ازای توالی^۲، محتوای نوکلئوتیدی توالی به ازای هر باز^۳، محتوای GC به ازای توالی^۴، محتوای باز خوانده نشده (N) به ازای هر باز^۵، توزیع طول توالی‌ها^۶، تکراری بودن توالی^۷، توالی‌هایی که بیش از حد وجود دارند^۸، محتوای آداپتوری^۹ و محتوای Kmer^{۱۰}. نتایج هر آزمون بطور جداگانه و به صورت گرافیکی ارائه می‌گردد. وضعیت قبول، رد یا هشدار برای آزمون‌های ۱۰ گانه ملاک سنجش کیفیت داده‌ها قرار گرفت. از آنجایی که اهمیت نسبی هر یک از آزمون‌ها با هم برابر نیست، نمی‌توان بر اساس تعداد آزمون‌های قبول، رد و یا در وضعیت هشدار نسبت به کیفیت کلی داده‌ها قضاوت نمود. از این رو یک آزمون دیگر با عنوان کیفیت توالی به ازای هر باز^{۱۱} نیز به همراه گزارشی حاوی آماره‌های پایه^{۱۲} درباره داده‌ها ارائه می‌گردد. جمع‌بندی کیفیت داده‌ها با آزمون گفته شده انجام می‌شود. در این آزمون، دامنه تعیین کیفیت بر اساس استاندارد کدگذاری سنجش کیفیت برای پلتفرم^{۱۳} توالی‌یابی نسل جدید مورد استفاده، که ممکن است برای هر مدل آن پلتفرم جداگانه و به طور اختصاصی تعریف شده باشد، به سه منطقه سبز، نارنجی و قرمز تقسیم می‌شود و سپس برای هر نوکلئوتید در طول هر خوانش یک میانگین و دامنه کیفیت به صورت انفرادی تعیین می‌شود و قرار گرفتن میانگین و دامنه مذکور در مناطق سه‌گانه کیفیت نوکلئوتید مربوطه را تعیین می‌کند. همچنین براساس نتیجه کیفیت همه

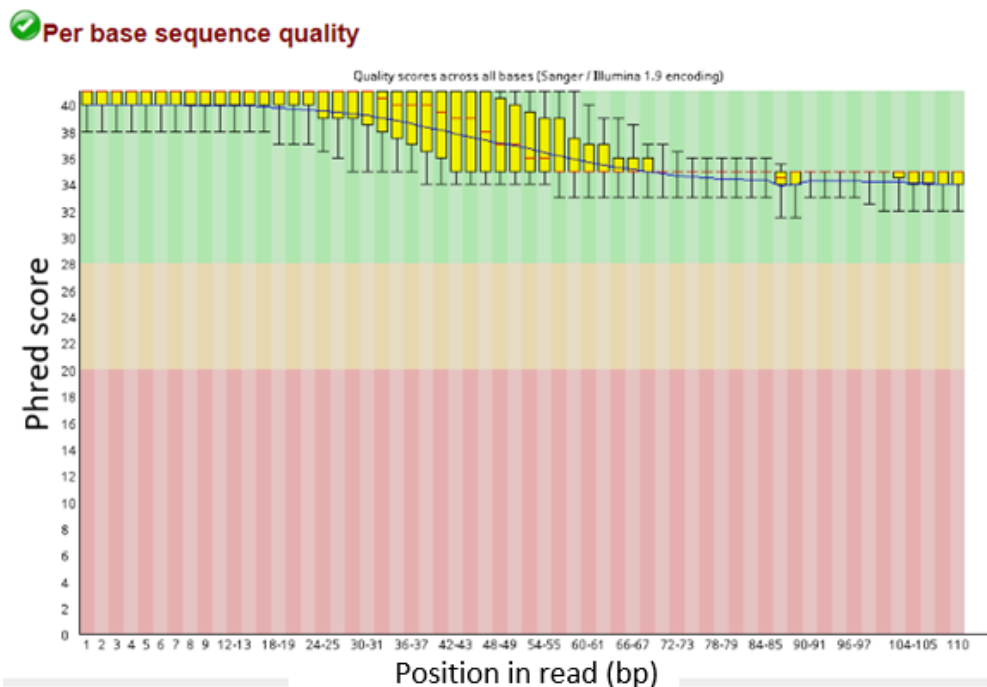
1. Per tile sequence quality
2. Per sequence quality scores
3. Per base sequence content
4. Per sequence GC content
5. Per base N content
6. Sequence Length Distribution
7. Sequence Duplication Levels
8. Overrepresented sequences
9. Adapter Content
10. Kmer Content
11. Per base sequence quality
12. Basic Statistics
13. Platform

توجه به اطلاعات به‌دست‌آمده از تمامی نرم‌افزارهای مورد استفاده در این مطالعه در مقایسه بین تیمار نر و ملکه از بین ۱۵۹۶۲ ژن شناسایی شده روی ژنوم زنبور عسل ۵۰۰۶ ژن با بیان متفاوت مشاهده شد. $p < 0.05$ و با $p < 0.000005$ تنها ۲۹۰۵ ژن با بیان متفاوت به‌دست آمد. مقایسه تیمارهای نر و کارگر حاکی از ۶۴۶۹ ژن با بیان متفاوت بود و با $p < 0.000005$ تعداد ۴۰۷۱ ژن با بیان متفاوت مشخص می‌شود. در مقایسه بین تیمارهای کارگر و ملکه با سطح معنی‌داری $(p < 0.05)$ ۶۲۹۶ ژن بیان متفاوت داشتند که با $(p < 0.000005)$ ۳۹۵۴ ژن با بیان متفاوت یافت شد به‌علت زیاد بودن تعداد ژن‌های با بیان متفاوت و با سطح معنی‌داری $(p < 0.000005)$ برای رسیدن به ژن‌های شاخص تر از معیار $\log_2(\text{fold_change})$ با مقادیر بیشتر از ۲ و کمتر از ۲- استفاده شد که نهایتاً ۴۶۵ ژن با بیان متفاوت در مقایسه تیمارهای نر و ملکه، ۴۹۵ ژن بین تیمارهای کارگر و ملکه و ۷۶۴ ژن بین تیمارهای نر و کارگر بود.

اطلاعاتی دیوید (<https://david.ncifcrf.gov>) صورت گرفت.

نتایج و بحث

پس از تبدیل فرمت بین ۱۰ تا ۱۵ میلیون خوانش کوتاه جفتی (از هر دو انتها توالی‌یابی شده) با میانگین طول ۱۵۰ جفت باز برای نمونه‌ها به‌دست آمد. نتایج سنجش کیفیت برای داده‌ها در شکل زیر نشان داده شده است. شاخص دوازده‌گانه کنترل کیفیت نمونه‌های مورد مطالعه حاکی از کیفیت مناسب داده‌ها به‌ازای هر باز و خوانش صحیح همه آنها براساس میزان باز N (خوانده‌نشده) به‌ازای هر توالی بود. با در نظر گرفتن حداقل طول خوانش ۸۰ جفت باز در هنگام ویرایش، ۹ تا ۱۴ میلیون خوانش با کیفیت مناسب برای نمونه‌ها باقی ماند و کمتر از ۱۰ درصد از کل خوانش‌ها کنار گذاشته شد. درصد مکان‌یابی خوانش‌ها به‌طور میانگین ۶۵ تا ۷۰ درصد بود و خوانش‌های با محل هم‌ردیفی چندگانه کمتر از یک درصد بودند. با



شکل ۱. آزمون کنترل کیفیت تک نوکلئوتیدی مربوط به نمونه "SRR157171". در این آزمون نمونه با فرد اسکور بالاتر از ۲۸ به‌عنوان نمونه با کیفیت بالا، فرد اسکور بین ۲۰ تا ۲۸ به‌عنوان نمونه با کیفیت متوسط و نمونه‌های با فرد اسکور زیر ۲۰ نیز به‌عنوان نمونه‌های نیازمند ویرایش شناخته می‌شوند.

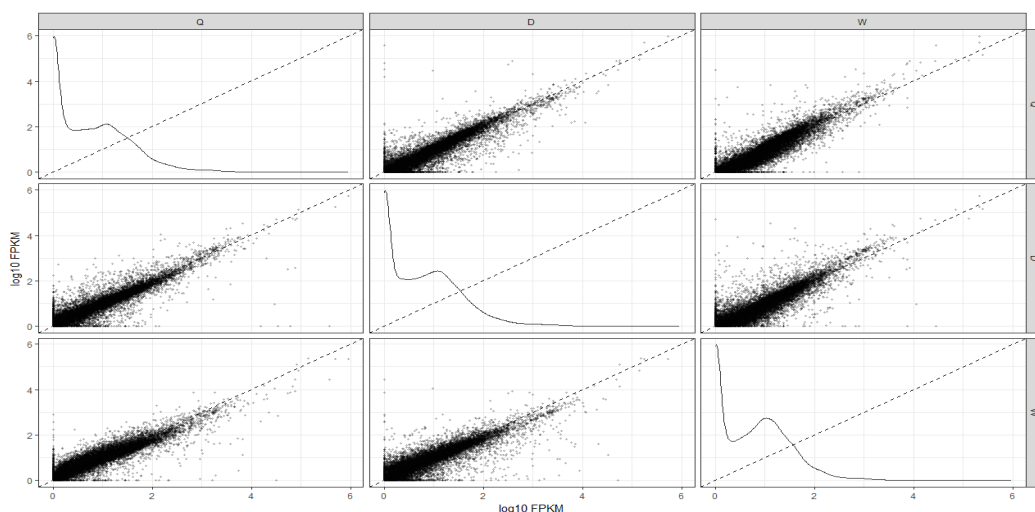
Figure 1. Per base sequence quality related to "SRR157171" sample, In this test, samples with Phred scores over 28 as a high quality sample, Phred scores between 20 and 28 as a medium-quality sample, and samples with Phred scores below 20 are also examples that need to trim.

۴ و ۵) که مقایسه دو به دو تیمارها با یکدیگر رسم شده است فقط ژن‌ها با fold change بیشتر از ۴ و کمتر از -۴ آورده شده است و از مقدار بیان ژن‌ها در هر تیمار لگاریتم ۱۰ گرفته شده است تا در نمودار بهتر نمایش داده شوند. همان‌طور که در این نمودارها مشاهده می‌شود بیان برخی ژن‌ها در بعضی تیمارها چند برابر تیمار دیگر است که نشان‌دهنده شاخص بودن آن ژن در تیمارهای مورد مقایسه می‌باشد. ژن‌هایی که دارای عملکرد شناخته شده بودند و در پایگاه‌های اطلاعاتی وظایف و عملکرد آن‌ها بررسی شده بود در جدول ۱ آورده شده است. در شکل ۳ ژن‌های GB42053 (XLOC_009493) در مسیر epididymal secretory protein E1-like، GB40136 (XLOC_004480)، ASF1 در (XLOC_011456) و GB45614 (XLOC_006751) در مسیر chymotrypsin inhibitor نقش داشتند که نسبت به باقی ژن‌ها تفاوت بیان زیادی داشتند. ژن XLOC_004877 کاملاً ناشناخته است و هیچ اطلاعاتی از آن در دسترس نیست و در نر اصلاً بیان نشده و به مقدار بسیار زیاد و قابل توجهی در کارگر بیان شده است. ژن‌های XLOC_011252، XLOC_013629، XLOC_013661، XLOC_015220 و XLOC_015331 نیز کاملاً ناشناخته‌اند و فقط با مقدار بسیار زیاد در زنبور نر بیان شده‌اند.

جهت رسیدن به تصویر مناسبی از تغییرات بیان در این تیمارها تنها ژن‌هایی با $\log_2(\text{fold_change})$ بیشتر از ۴ و کمتر از -۴ برای مطالعه انتخاب شدند. در مجموع از ۱۵۳۱۴ ژن تنها ۴۸ ژن در بین تیمار نر و ملکه بیان متفاوت داشتند و از آن‌ها ۲۷ ژن در نر بیان بیشتری نسبت به نمونه‌های ملکه داشت، ۳۲ ژن بین تیمارهای کارگر و ملکه بیان متفاوت داشتند و از آن‌ها ۲۵ ژن در ملکه بیان بیشتری نسبت به نمونه‌های کارگر داشتند و ۷۲ ژن بین نر و کارگر با بیان متفاوت بودند که از این تعداد ۵۳ ژن در تیمار نر بیان بیشتری نسبت به تیمار کارگر داشتند ($q < 0.001$) و ($\text{fold change} > 4$) و ($\text{fold change} < -4$) در نمودارهای ۳، ۴ و ۵ ID این ژن‌ها به‌همراه میزان بیان آن‌ها آمده است.

نمودار نقطه‌ای زیر (شکل ۲) میزان انحراف هر ژن، از میانگین بیان ژن در هر سه تیمار را نشان می‌دهد. میزان فاصله بیشتر از خط میانگین نشان‌دهنده بیان متفاوت‌تر می‌باشد. ژن‌های کاملاً متفاوت بیان شده روی محورهای X و Y نمودار داخل کادر مشخص شده‌اند.

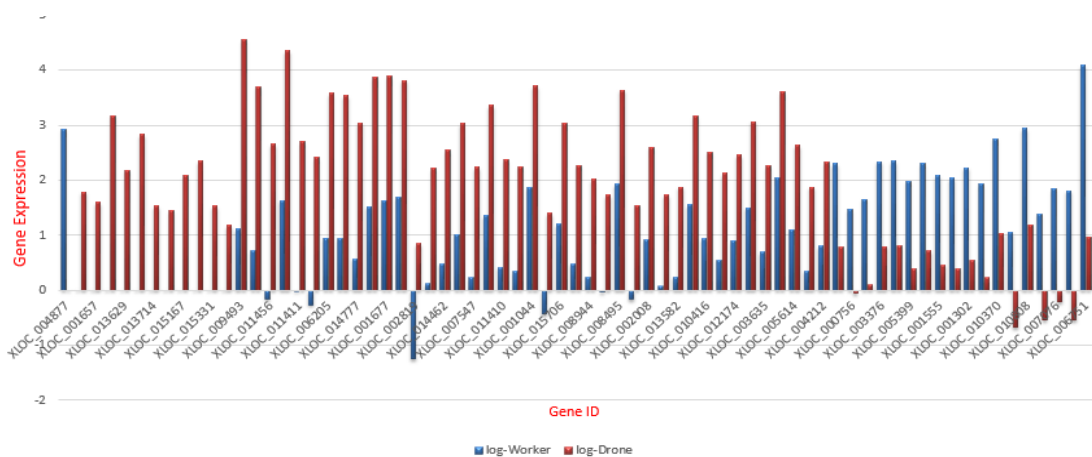
همچنان‌که از نمودار هیستوگرام برمی‌آید بخشی از تفاوت بیان ژن ناشی از آن است که برخی از ژن‌ها فقط در یکی از تیمارها بیان شده است و بیان آنها در تیمارهای دیگر صفر بوده است. در نمودارهای میله‌ای (۳)،



شکل ۲. نمودار نقطه‌ای تفاوت بیان ژن‌ها بین سه تیمار (کارگر: W، نر: D، ملکه: Q) (هر چه میزان FPKM در این نمودار کمتر باشد بیانگر میزان بیان کمتر ژن‌ها و در صورت نزدیکی ژن‌های مشابه در دو گروه به خط میانه میزان تفاوت بیان نیز کمتر می‌باشد).
Figure 2. Gene expression scatter plot in three treatments (Q: Queen, D: Drone and W: Worker) (The lower level FPKM in this chart, indicated lower level of the genes expression and the Closer to the middle line of the genes in the two groups, indicate the lower difference in expression).

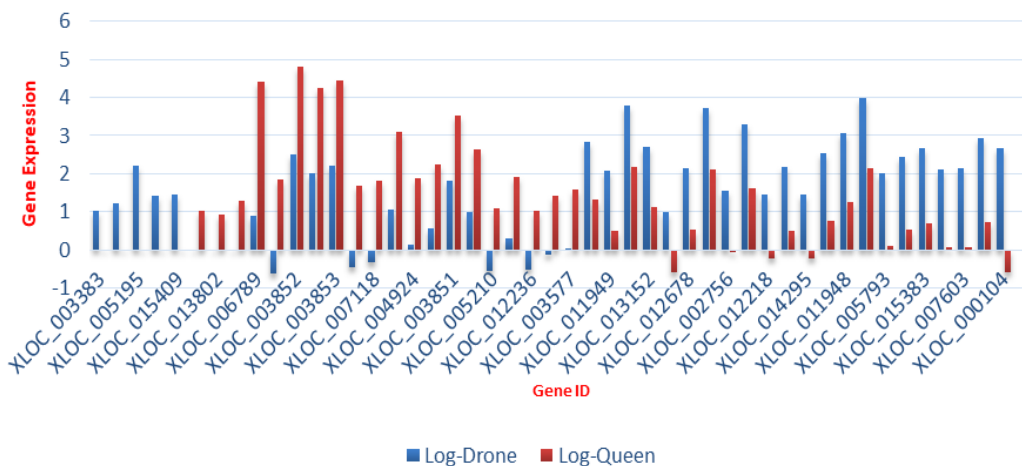
مسیر epididymal secretory protein E1-like XLOC_010282 در ملکه بیشتر بیان را داشتند. از ژن‌های XLOC_011148 و XLOC_013524 هیچ اطلاعاتی در دسترس نیست در ملکه بیان بسیار بالایی داشتند و در کارگر بیان آنها صفر است. از ژن‌های XLOC_007616، XLOC_004844، XLOC_001811 و XLOC_010197 نیز هیچ اطلاعاتی در دسترس نیست و فقط در کارگر به مقدار بسیار زیاد بیان شده‌اند.

در این نمودار ژن XLOC_006789 در ملکه و XLOC_000104 در نر ژن‌های مهم با بیشترین تفاوت بیان بودند. از ژن‌های XLOC_000104 در نر و XLOC_003745 در ملکه هیچ اطلاعاتی در دسترس نیست ولی هر دو در تیمار مقابل بیانشان کاملاً صفر بود. ژن‌های XLOC_003351 در مسیر lipase member H-A-like و XLOC_009027 در مسیر diacylglycerol kinase theta در کارگر و ژن‌های XLOC_009392



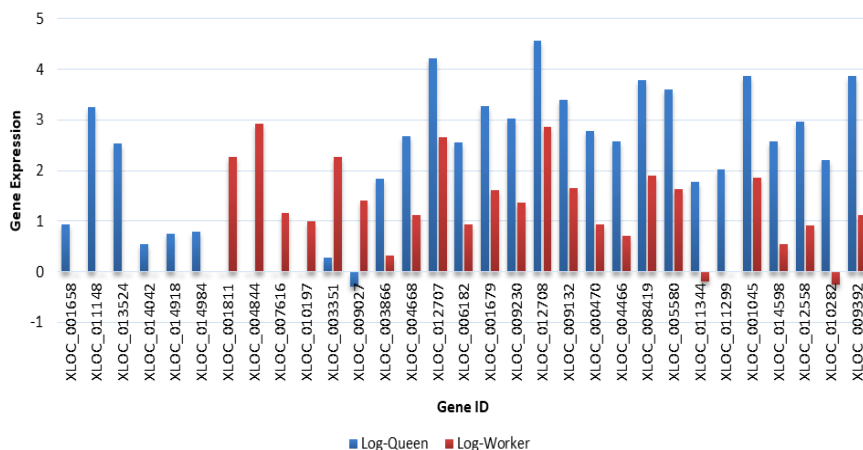
شکل ۳. مقایسه تفاوت بیان ژن‌ها در تیمارهای کارگر و نر (محور افقی نشان‌دهنده ژن‌های با تفاوت بیان در دو گروه کارگر و نر می‌باشد و محور عمودی مقدار بیان ژن‌ها بر اساس لگاریتم ۱۰ FPKM می‌باشد و در این نمودار ژن‌هایی که تنها در یکی از دو تیمار نر و کارگر بیان شده‌اند نیز مشخص شده‌اند).

Figure 3. Differential gene expression in drone vs worker (Horizontal axis represents genes with difference in expression in both drone and worker groups. The vertical axis of the expression of genes is based on the 10 logarithm of FPKM, and in this chart the genes expressed in only one of the two treatments are males and workers also specified).



شکل ۴. مقایسه تفاوت بیان ژن‌ها در تیمارهای ملکه و نر (محور افقی نشان‌دهنده ژن‌های با تفاوت بیان در دو گروه ملکه و نر می‌باشد و محور عمودی مقدار بیان ژن‌ها بر اساس لگاریتم ۱۰ FPKM می‌باشد و در این نمودار ژن‌هایی که تنها در یکی از دو تیمار نر و ملکه بیان شده‌اند نیز مشخص شده‌اند).

Figure 4. Differential gene expression in drone vs queen (Horizontal axis represents genes with difference in expression in both drone and worker groups. The vertical axis of the expression of genes is based on the 10 logarithm of FPKM, and in this chart the genes expressed in only one of the two treatments are drone and queen also specified).



شکل ۵. مقایسه تفاوت بیان ژن‌ها در تیمارهای کارگر و ملکه (محور افقی نشان‌دهنده ژن‌های با تفاوت بیان در دو گروه کارگر و ملکه می‌باشد و محور عمودی مقدار بیان ژن‌ها بر اساس لگاریتم ۱۰ FPKM می‌باشد و در این نمودار ژن‌هایی که تنها در یکی از دو تیمار ملکه و کارگر بیان شده‌اند نیز مشخص شده‌اند).

Figure 5. Differential gene expression in queen vs worker (Horizontal axis represents genes with difference in expression in both male and worker groups. The vertical axis of the expression of genes is based on the 10 logarithm of FPKM, and in this chart the genes expressed in only one of the two treatments are worker and queen also specified).

است در ملکه به میزان قابل قبولی بیان شده و بیان آن در تیمار کارگر و نر صفر می‌باشد که دلیل آن فعالیت تخم‌گذاری قابل توجه ملکه می‌باشد که نیازمند مسیرهای متابولیک پیشرفته‌تری نسبت به تیمارهای دیگر است. ژن GB42797 در تیمار نر بسیار بیشتر از ملکه و کارگر بیان شده است و مرتبط با پروتئین کنترل‌شونده با نور روزانه می‌باشد. یکی از دلایل این تفاوت بیان این است که زنبورهای نر نیاز بیشتری به این ژن برای تنظیم فعالیت تولید مثلی خود نسبت به سایر تیمارها دارند.

از دیگر نتایج این پژوهش ثبت ژن Attari 1 در زنبور عسل نژاد *Apis mellifera ligustica* در پایگاه اطلاعاتی NCBI با شماره بانک ژن MK085892.1 می‌باشد. ثبت این ژن به صورت partial cds (کدون پایان آن شناخته‌شده نیست) می‌باشد. این ژن در تیمار نر شناسایی شد و از ۵۶۵ بازآلی تشکیل شده است. ژن Attari 2 نیز با شماره بانک ژن MN586260 با ۳۴۲ بازآلی در زنبور ملکه در این پایگاه اطلاعاتی به ثبت رسید. ژن Attari 3 هم در دست بررسی توسط این پایگاه است و به زودی منتشر خواهد شد.

آنالیز pathway با استفاده از وب سایت <http://amp.pharm.mssm.edu/Enrichr/enrich> انجام شد که بسیاری از ابزار آنلاین مرتبط را در خود جای داده است و می‌توان نتایج آن‌ها را به صورت تجمیع‌شده در این

در مقایسه تیمارهای نر و ملکه ژن‌هایی که تأثیر شناخته‌شده‌ای در تولید پروتئین‌ها و سایر اجزای مرتبط با غشای سلولی داشتند تفاوت بیان قابل توجهی از خود نشان دادند که می‌توان به ژن‌های GB48190، GB48148 و GB46663 که در تیمار نر بیشتر از ملکه و کارگر بیان شده بود اشاره نمود. اما ژن‌های GB49017، GB45613 که آن هم مرتبط با ساختار غشا است در تیمار ملکه بیشتر از کارگر و نر بیان شده بود که تفاوت بیان این ژن‌ها به دلیل تفاوت اندازه بدن، نوع فعالیت و نقش هر یک در کندو می‌باشد. ژن GB54554 که مرتبط با پروتئین کوتیکول شفیره است در تیمار نر بیشتر از ملکه و کارگر بیان شده است دلیل آن می‌تواند بزرگ تر بودن اندازه شفیره نر نسبت به ملکه و کارگر می‌باشد. ژن GB49509 که نقش شناخته‌شده آن بازدارنده کیموتریپسین می‌باشد در ملکه بیشتر از کارگر بیان شده و بیان آن در تیمار نر صفر می‌باشد. ژن GB51373 که در پاسخ‌های مرتبط با استرس نقش دارد در نر بیشتر از ملکه بیان شده است و یکی از دلایل آن فعالیت پر استرس زنبور نر نسبت به ملکه می‌باشد و نقش زنبور نر صرفاً جفت‌گیری است و برای جفت‌گیری با ملکه زنبور نر باید با تعداد زیادی نر رقیب رقابت نماید تا موفق به جفت‌گیری شود.

ژن GB41948 که مرتبط با مسیرهای متابولیک

تفاوت بیان در ژن‌های nr ۲ و ۴ روزه ۵۹۸ ژن شناسایی و در کارگر ۱۱۹۰ ژن و در ملکه ۱۱۸۱ ژن شناسایی شد. در پژوهش Wu et al. (2016) ترانسکریپتوم زنبورهای nr و ملکه که به تازگی از شفیره خارج شده‌اند با تیمار بالغ آن‌ها و با یکدیگر مورد مقایسه قرار گرفت. در این پژوهش با استفاده از آنالیز GO ژن‌های دخیل در فعالیت‌های متابولیکی، فرآیندهای سلولی، فرآیندهای بیولوژیک و ترکیبات ساختاری سلولی بین تیمار ملکه بالغ و تیمار ملکه تازه متولدشده متفاوت بیان شده بودند و بیماری‌های پارکینسون و آلزایمر و واکنش فسفوریلاسیون اکسیداتیو به‌عنوان مهمترین pathwayها معرفی شدند. که در مقایسه با پژوهش حاضر نیز مهمترین pathwayها مربوط به انسان یا سایر پستانداران بودند. نتایج آنالیز GO نیز از آنجاکه مرتبط با رشد و توسعه بافت‌ها و اندام‌ها بود با پژوهش حاضر که در لارو سن ۵ که یکی از مهمترین سنین لاروی در رشد و توسعه اندام‌ها می‌باشد قابل مقایسه است و نتایج نیز یکدیگر را تأیید می‌نمایند.

آنالیز ماهیت ژنی نیز به‌وسیله بانک اطلاعاتی <https://david.ncicrf.gov> انجام شد که نتایج آن در جدول‌های ۱، ۲ و ۳ آمده است.

وب سایت یافت. نتیجه این آنالیز نشان داد که در مقایسه تیمار ملکه و nr مهمترین pathwayها، "تراکم معدنی ستون فقرات"، "ابتلا به بیماری آنفولانزا"، "اندازه نیش در پشه"، "موداری بیش از حد" در مقایسه تیمار ملکه و کارگر "سوخت‌وساز سلنیوم/ سلنوپروتئین"، "اثر کلسینورین در تمایز کراتینوسیت در انسان"، "کانال‌های انتقال یون در انسان" و در مقایسه تیمارهای ملکه و کارگر، "miRNA و OSx در توسعه دندان" و "سندرم مرگ ناگهانی نوزادان (SIDS)" از مهمترین pathwayها بودند. از آنجاکه زنبور nr نیش ندارد ولی در ملکه نیش وجود دارد pathway نیش حائز اهمیت می‌باشد.

در پژوهش Doori Park et al. (2015) با استفاده از داده‌های توالی‌یابی کامل ژنوم و روش de novo assembly در زنبور *A. cerana* توانستند ۱۰۶۵۱ ژن را شناسایی کنند که در مقایسه با مطالعه حاضر که با روش توالی‌یابی ترانسکریپتوم انجام شد و ۱۵۹۶۲ ژن شناسایی شد تعداد کمتری است.

در پژوهش Xu-Jiang He et al. (2017) لاروهای ۲ و ۴ روز، nr، ملکه و کارگر با استفاده از داده‌های RNA-seq بررسی شدند. لاروهای ۲ روزه و ۴ روزه nr و کارگر پروفایل بیان ژن شبیه‌تری نسبت به ملکه داشتند. همچنین با معیار $\log_2(\text{foldchange})$ و $q\text{-value} < 0.01$

جدول ۱. آنالیز هستی‌شناسی ژن‌های متفاوت بیان‌شده در دو گروه ملکه و nr
Table 1. GO analysis of genes differentially expressed in queen vs drone

Gene	Process	Description
GB48190	integral component of membrane,	
GB49809	calcium ion binding,	
GB51783	carboxypeptidase Q-like(LOC410149)	
GB51373	cell wall integrity and stress response component 1-like(LOC100577331)	chitin metabolic process,
GB45613	cholecystokinin receptor-like(LOC410654)	gastrin receptor activity,
GB45614	chymotrypsin inhibitor(LOC100576444)	
GB49509	Trypsin Inhibitor-like, cysteine rich domain,	
GB42797	circadian clock-controlled protein(LOC726981)	Haemolymph juvenile hormone binding,
GB41428	defensin 1(Def1)	defense response to bacterium, innate immune response,
GB49017	facilitated trehalose transporter Tret1-like(LOC412007)	transmembrane transporter activity,
GB48020	flocculation protein FLO11-like(LOC552041)	
GB54554	pupal cuticle protein G1A-like(LOC100576182)	
GB46224	odorant binding protein 15(Obp15)	Pheromone/general odorant binding protein,
GB48492	take-out-like carrier protein(JHBP-1)	
GB41097	trypsin-7(LOC724565)	
GB41707	tubulin beta chain-like(LOC410559)	
GB46663	uncharacterized LOC410375(LOC410375)	integral component of membrane
GB53618	Protein phosphatase 1, regulatory subunit 15A/B, C-terminal	
GB48148	integral component of membrane,	
GB54508	Protein of unknown function DUF227, Protein kinase-like domain, CHK kinase-like,	

1. Excessive hairiness
2. Role of Osx and miRNAs in tooth development WP3971
3. Sudden Infant Death Syndrome (SIDS)

جدول ۲. آنالیز هستی‌شناسی ژن‌های متفاوت بیان‌شده در دو گروه ملکه و کارگر
Table 2. GO analysis of genes differentially expressed in queen vs worker

Gene	Process	Description
GB41948	cytidine deaminase-like(LOC724183)	Pyrimidine metabolism, Drug metabolism - other enzymes, Metabolic pathways,
GB50415	diacylglycerol kinase theta(LOC408809)	Lipid metabolism / General function prediction only,
GB42053	epididymal secretory protein E1-like(LOC724386)	intracellular cholesterol transport,
GB43508	lipase member H-A-like(LOC727193)	
GB42287	peritrophin-1-like(LOC100577577)	chitin metabolic process,
GB50026	transmembrane protease serine 11G-like(LOC725250)	
GB43311	vanin-like protein 1(LOC100578995)	
GB40614	vitamin K-dependent gamma-carboxylase(LOC551869)	Ubiquinone and other terpenoid-quinone biosynthesis,

جدول ۳. آنالیز هستی‌شناسی ژن‌های متفاوت بیان‌شده در دو گروه کارگر و نر
Table 3. GO analysis of genes differentially expressed in drone vs worker

Gene	Description	Process
GB53306	integral component of membrane,	Niemann-Pick C1 protein-like(LOC408992)
GB45983	Amino acid transport and metabolism,	alpha-methyl dopa hypersensitive protein-like(LOC410639)
GB49809	calcium ion binding,	calmodulin-like(LOC408690)
GB51783	Peptidase M28,	carboxypeptidase Q-like(LOC410149)
GB45613	integral component of membrane,	cholecystokinin receptor-like(LOC410654)
GB45614	Trypsin Inhibitor-like, cysteine rich domain,	chymotrypsin inhibitor(LOC100576444)
GB41361	heme binding, metal ion binding,	cytochrome b5-like(LOC724654)
GB42053	intracellular cholesterol transport,	epididymal secretory protein E1-like(LOC724386)
GB49017	integral component of membrane,	facilitated trehalose transporter Tret1-like(LOC412007)
GB44871	glycine N-methyltransferase activity,	glycine N-methyltransferase(LOC552832)
GB51223	defense response to bacterium, innate immune response,	hymenoptaecin(LOC406142)
GB46224	odorant binding,	odorant binding protein 15(Obp15)
GB43510	extracellular region,	pancreatic triacylglycerol lipase-like(LOC551268)
GB42287	chitin metabolic process,	peritrophin-1-like(LOC100577577)
GB53911	chitin metabolic process,	peritrophin-1-like(LOC725725)
GB49639	Reeler domain,	putative defense protein 3(LOC726072)
GB45855	serine-type peptidase activity,	putative serine protease K12H4.7(LOC410729)
GB50026	serine-type endopeptidase activity,	transmembrane protease serine 11G-like(LOC725250)
GB41097	serine-type endopeptidase activity,	trypsin-7(LOC724565)
GB54159	integral component of membrane,	uncharacterized LOC552735(LOC552735)
GB48148	integral component of membrane,	uncharacterized LOC725175(LOC725175)
GB41912	oxidoreductase activity,	uncharacterized oxidoreductase YrbE-like(LOC552024)
GB43311	nitrogen compound metabolic process,	vanin-like protein 1(LOC100578995)
GB43310	nitrogen compound metabolic process,	vanin-like protein 1(LOC724312)

نتیجه‌گیری کلی

نتایج به‌دست‌آمده از این پژوهش نشان می‌دهد که، در مجموع از ۱۵۳۱۴ ژن تنها ۴۸ ژن در بین تیمار نر و ملکه، ۳۲ ژن بین تیمارهای کارگر و ملکه و ۷۲ ژن بین نر و کارگر با بیان متفاوت بودند ($p < 0.001$) و ($fold\ change < 4$ و $fold\ change > 4$). تعداد زیادی از این ژن‌های شاخص در مقایسه تیمارهای: ملکه و نر مرتبط با اجزای ساختاری و جدایی‌ناپذیر غشا، پردازش متابولیسم کیتین، بازدارنده‌های تریپسین، فعالیت گیرنده گاسترین، باندشونده با هورمون جوانی، پاسخ‌های دفاعی نسبت به باکتری‌ها، پروتئین‌های کوتیکول شفیره هستند در صورتی‌که در مقایسه ژن‌های شاخص متمایز در تیمارهای ملکه و کارگر مرتبط با مسیرهای متابولیسمی، متابولیسم داروها و سایر آنزیم‌ها، متابولیسم لیپیدها،

باندشونده با نوکلئیک اسید، انتقال‌دهنده کلسترول داخل سلولی، پردازش‌های متابولیسمی کیتین، بیوسنتز یوبیکینون و سایر تریپتوفان‌ها و ژن‌های متمایز در مقایسه کارگر و نر مرتبط با اجزای ساختاری و جدایی‌ناپذیری از غشای متابولیسم و انتقال آمینواسیدها، باندشونده با یون کلسیم، باندشونده با فلز یون، انتقال‌دهنده کلسترول کیتین درون سلول، متابولیسم بودند. قابل ذکر است که این تحقیق منجر به شناسایی ژن‌های متمایزکننده بین زنبورهای کارگر و ملکه و نر شد و می‌توان این ژن‌ها را در شرایط آزمایشگاهی نیز چک کرد و در صورت تأیید آزمایشگاهی بتوان مراحل بعدی در ارتباط با استراتژی‌های اصلاح نژادی را پیشبرد و همچنین ژن‌های شناسایی و ثبت‌شده برای اولین بار در زنبور نیز انجام شد.

REFERENCES

1. Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. From: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
2. Banabazi, M. H., Naserkheil, M. & Miraei-Ashtiani, S. R. (2012a). Network regulatory Gene expression of cerevisiae cell cycle with correlation weight method. *The third national conference on agricultural biotechnology Iran*. Ferdowsi University of Mashhad. (in Farsi)
3. Banabazi, M. H., Naserkheil, M. & Miraei-Ashtiani, S. R. (2012b). Algorithm to identify genes expressed differently in the yeast *Saccharomyces cerevisiae* microarray data using software packages R. *The third national conference on agricultural biotechnology Iran*. Ferdowsi University of Mashhad. (in Farsi)
4. Barchuk, A. R., Cristino, A. S., Kucharski, R., Costa, L. F., Simões, Z. L. & Maleszka, R. (2007). Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Developmental Biology*, 7(1), 70.
5. Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
6. Park, D., Jung, J. W., Choi, B. S., Jayakodi, M., Lee, J., Lim, J., ... & Choi, I. Y. (2015). Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. *BMC Genomics*, 16(1), 1.
7. Ekblom, R. & Galindo, J. (2010). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1-15.
8. Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K. & Chen, X. (2010). Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. *Cell Research*, 20(7), 763-783.
9. Haas, B. J. & Zody, M. C. (2010). Advancing RNA-seq analysis. *Nature Biotechnology*, 28(5), 421.
10. He, X.-J., Jiang, W.-J., Zhou, M., Barron, A. B. & Zeng, Z.-J. (2017). A comparison of honeybee (*Apis mellifera*) queen, worker and drone larvae by RNA-Seq. *Insect Science*
11. Langmead, B. & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357-359.
12. Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.
13. Goff, L. A., Trapnell, C. & Kelley, D. (2012). CummeRbund: visualization and exploration of Cufflinks high-throughput sequencing data. *R Package Version*, 2(0).
14. Marguerat, S. & Bähler, J. (2010). RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences*, 67(4), 569-579.
15. Mortazavi, A., Williams, B., McCue, K., Schaeffer, L. & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), 621-628.
16. Ashby, R., Forêt, S., Searle, I. & Maleszka, R. (2016). MicroRNAs in honey bee caste determination. *Scientific Reports*, 6, 18794.
17. Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M. & Schmidt, D. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, 321(5891), 956-960.
18. Tahmasebi, Gh. & Porgharaei, H. (2000). Investigating the role of bee in pollination and increasing the production of agricultural products in Iran
19. Wilhelm, B. T. & Landry, J. R. (2009). RNA-Seq-quantitative measurement of expression through massively parallel RNA-Sequencing. *Methods*, 48, 249-257.
20. Wu, X. B., Wang, Z. L., Gan, H. Y., Li, S. Y. & Zeng, Z. J. (2016). Transcriptome comparison between newly emerged and sexually matured bees of *Apis mellifera*. *Journal of Asia-Pacific Entomology*, 19(3), 893-897.