

Article

Not peer-reviewed version

Evolutionary Adaptation of Genes Involved in Galactose Derivatives Metabolism in Oil-Tea Specialized *Andrena* Species

[Gonghua Lin](#) , Bo He , Kai Jiang , [Tianjuan Su](#) , [Fang Zhao](#) * , [Zuhao Huang](#) *

Posted Date: 8 May 2023

doi: 10.20944/preprints202305.0462.v1

Keywords: *Andrena camellia*; genome; RNASeq; molecular evolution; gene expression; galactose derivatives; NAGA-like



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Evolutionary Adaptation of Genes Involved in Galactose Derivatives Metabolism in Oil-Tea Specialized *Andrena* Species

Gonghua Lin, Bo He, Kai Jiang, Tianjuan Su, Fang Zhao * and Zuhao Huang *

School of Life Sciences, Jinggangshan University, Ji'an 343009, China

* Correspondence: zf_lgh@163.com (F.Z.), hzhzhou@163.com (Z.H.)

Abstract: Oil-tea (*Camellia oleifera*) is a woody oil crop whose nectar includes galactose derivatives that are toxic to honey bees. Interestingly, some mining bees of the genus *Andrena* can entirely live on nectar (and pollen) of oil tea and are able to metabolize these galactose derivatives. We present the first next-generation genomes for five and one *Andrena* species that are respectively specialized and non-specialized oil-tea pollinators, and combining with published genomes of six other *Andrena* species which did not visit oil-tea, we performed molecular evolution analyses on genes involved in metabolism of galactose derivatives. Six genes (*NAGA*, *NAGA-like*, *galM*, *galK*, *galT*, and *galE*) involved in galactose derivatives metabolism were identified in the five oil-tea specialized species, but only five (with the exception of *NAGA-like*) were discovered in the other *Andrena* species. Molecular evolution analyses revealed that *NAGA-like*, *galK* and *galT* in oil-tea specialized species appeared under positive selection. RNASeq analyses showed that *NAGA-like*, *galK*, and *galT* were significantly up-regulated in the specialized pollinator *A. camellia* compared to the non-specialized pollinator *A. chekiangensis*. Our study demonstrated that the genes *NAGA-like*, *galK*, and *galT* had played an important role in the evolutionary adaptation of the oil-tea specialized *Andrena* species.

Keywords: *Andrena camellia*; genome; RNASeq; molecular evolution; gene expression; galactose derivatives; *NAGA-like*

1. Introduction

Bees are a monophyletic lineage within a clade Anthophila in the superfamily Apoidea [1]. There are around 20,000 different species of bees, and thanks to their effective pollination for both crops and natural flora, bees are now essential parts of practically all terrestrial ecosystems [2,3]. Bees can be broadly divided into two functional groups based on floral specificity: oligolectic species, which forage on one or a small number of closely related plant species, and polylectic species, which collect nectar and/or pollen from many different plant species [4]. The most well-known bee species include the polylectic honey bees (*Apis spp.*) and bumble bees (*Bombus spp.*). Numerous biological and ecological investigations of the polylectic species have been conducted [2,5]. In contrast, oligolectic bees have received significantly less research despite making up a sizable fraction of the world's bee fauna [6]. For oligolectic bee exploitation and conservation, therefore, a greater understanding of their biology and ecology is required.

Andrena Fabricius (Andrenidae) is a large bee genus of around 1,600 species with a wide distribution mainly throughout the Holarctic. They are a crucial pollinator in both natural and agricultural contexts, and they are a particularly important aspect of northern temperate ecosystems [7]. *Andrena* species exhibit a spectrum of diet breadth, from polylectic to oligolectic, which makes this genus a superb group to study the evolution of diet specialization [8,9]. Oil-tea (*Camellia oleifera*) is an important woody oil crop in many countries, including China, the Philippines, India, Brazil and South Korea [10]. This plant only blooms in November and December [11], when there are few wild pollinators present. As a result, crop yields are very limited due to significant pollinator constraint [12,13]. Local farmers have attempted utilizing domestic honey bees (*Apis mellifera* and *Apis cerana*) to increase pollination efficiency, however, both species are badly damaged by the nectar's toxicity

[14]. Interestingly, some wild bees such as *Andrena camellia* and *Colletes gigas* primarily rely on oil-tea nectar to survive, suggesting these species have coevolved to become an expert at oil-tea [15–17].

Direct observation on floral visiting and microscope examination on pollen from pollen basket showed that *A. camellia* near exclusively collects nectar and pollen from oil-tea blossom [12]. *A. camellia* emergences in the middle of October and keep activities (mating, oviposition, and larval development) mainly during November and December, well synchronizing the blossom period of oil-tea tree [11,12]. Besides *A. camellia*, at least three other *Andrena* bees (*A. chekiangensis*, *A. hunanensis* and *A. striata*) are also reported visiting oil tea flowers [18,19]. *A. chekiangensis* is much larger than *A. camellia*, this makes it easier to identify between the two species. Our field observations find that, *A. chekiangensis* is strictly not an oil-tea specialist, because it also frequently visits tea tree blossoms (*Camellia sinensis*). In some sympatric habitats of *C. oleifera* and *C. sinensis*, *A. chekiangensis* individuals were more frequently observed in *C. sinensis* blossoms. In contrary, *A. hunanensis* and *A. striata* are almost indistinguishable from *A. camellia* in terms of morphology features and feeding habits (oil-tea specialization). It should be noted that *A. hunanensis* and *A. striata* are frequently mistakenly classified as *A. camellia* in many amateur fieldwork reports. Due to the morphology similarity among *A. camellia* and its close relatives, the researchers believe there may even be hidden unknown species that have yet to be discovered.

Understanding the poisoning and detoxifying processes used by various bee species could be crucial to enhancing oil-tea flower pollination. According to chemical tests, the primary toxins affecting western honey bees are the galactose derivatives: raffinose, manninotriose, and stachyose [20,21]. Typically, these oligosaccharides must be broken down in two steps: first, the alpha-galactosidine linkages are hydrolyzed by galactosidases to produce sucrose and galactose, and then the galactose molecules are converted to UDP-glucose via the Leloir route [22–24]. It should be noted that little is known about the genes involved in the metabolism of galactose derivatives in *Andrena* species. Here, we present the first next-generation sequencing of *Andrena* species consuming oil-tea nectars. By combining these data with genomic information of other *Andrena* species from GenBank, we performed bioinformatic analyses of genes involved in the metabolism of galactose derivatives. The objective is to determine whether the oil-tea specialized *Andrena* species differ from the other *Andrena* species in terms of evolutionary specialization.

2. Materials and Methods

All pollinators were live-trapped from oil-tea blossoms in Jiangxi, Anhui, Sichuan, Zhejiang, Guangdong, and Hunan Provinces, China (Figure 1). In order to pick out *Andrena* samples the specimens were identified based on morphological characteristics [18]. One female sample for each species was chosen for genome sequencing. Total genomic DNA was extracted from the thorax of each individual using the QIAGEN DNeasy Blood and Tissue kit (Germany), following the manufacturer's protocols. DNA libraries with ~350 bp insertions were constructed and were then sequenced with both directions of 150-bp reads using the Illumina HiSeq 2000 sequencing platform (Illumina Inc. San Diego). Quality control for raw reads data was performed using fastp 0.20.0 with default settings and parameters [25].

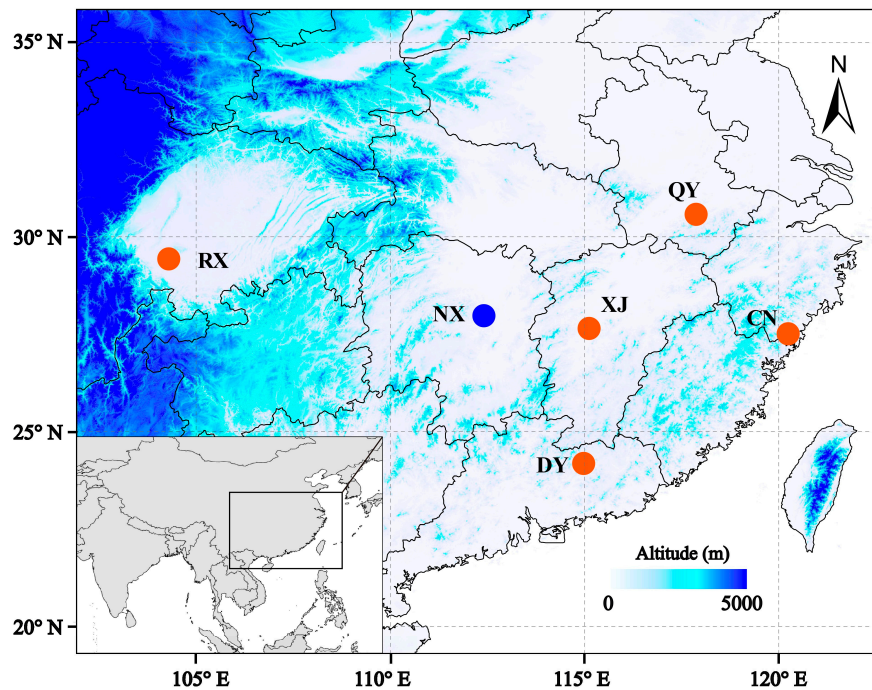


Figure 1. Sampling sites of the six *Andrena* species that pollinating oil-tea (red dots, oil-tea specialized species; blue dot, non-specialized oil-tea pollinator).

The clean reads were used for de novo assembly with MEGAHIT [26]. The mitochondrial COI sequences were extracted by local blast using NCBI-BLAST+ program v2.13.0 [27] and were used as queries to search in BOLD system (www.boldsystems.org) to determine their taxonomic information. Previously published six *Andrena* genomes downloaded from GenBank were used as comparison objects: *A. dorsata*, GCA_929108735.1; *A. fulva*, GCA_946251845.1; *A. haemorrhoea*, GCA_910592295.1; *A. hattorfiana*, GCA_944738655.1; *A. minutula*, GCA_929113495.1; *A. bucephala*, GCA_947577245.1. All 13 mitochondrial coding sequences in the genomes sequenced in this study and those from GenBank were extracted and were concatenated to reconstruct phylogenetic trees using IQ-TREE [28].

The galactose metabolism pathway of Hymenoptera in KEGG (ko00052) was used to identify candidate genes involved in galactose derivatives metabolism. With honey bee (*A. mellifera*) candidate genes as query sequences, the exonerate program v2.4.0 [29] was used to find homologous genes in *Andrena* genomes. MEGA v10 [30] was used to combine and align the coding sequences from all samples for each gene. DNAsP v6 [31] was used to identify the genetic variation information for each gene. The sequence similarity information among the genes as well as their translated protein sequences are calculated using clustal omega [32].

The molecular evolution analyses were performed in PAML program package [33]. The branch model was used to estimate the dN/dS (nonsynonymous / synonymous mutation) ratios of the foreground clade (oil-tea specialized species). The likelihood ratio tests (LRTs) between M0 (null model) and the branch model were performed by comparing twice the difference in log-likelihood values ($2\Delta\ln L$) against a chi-square distribution ($df = 2$). We also used the branch-site model called Model A to test for positive selections in the foreground clade oil-tea specialized species. The null model for Model A is Model A1, which is a modification of Model A, but with $\omega_2 = 1$ fixed [34,35]. The putative positive selection sites were deduced with Bayes Empirical Bayes (BEB) analysis. It should be noted that, the seven non-specialized *Andrena* species lacked *NAGA-like* gene. Because of the high sequence similarity between *NAGA* and *NAGA-like* genes, we arbitrarily used *NAGA* genes of these species instead.

In order to analyze the relative expression level of each gene, we also carried out RNAseq sequencing for *A. camellia* and *A. chekiangensis*, representing specialized and non-specialized oil-tea pollinators, respectively. Total mRNA was isolated from whole specimens of each individual (4 individuals were analyzed for each species). Following the manufacturer's instructions, 150 bp reads

were sequenced bidirectionally by the Illumina platform (Illumina, San Diego, CA). The obtained clean reads of a randomly selected individual of each species were used for de novo assembly using the Trinity program [36]. The transcripts were processed by CD-HIT-EST [37] to remove the redundant sequences, and the generated unigenes were then used to predict coding sequences (CDSs) with the GeneMarkS-T program [38].

Orthologous genes were identified using OrthoFinder v2.3.11 [39]. The salmon program v1.0.0 [40] was used to calculate the expected read counts and transcripts per million (TPM) value for every orthologous gene, which were then used to identify differentially expressed genes (DEGs) with the DEBrowser program [41]. The genes with posterior fold changes (FC) of *A. camellia* against *A. chekiangensis* over two (i.e., $FC > 2$ or $FC < 0.5$) and with highly significant posterior probabilities of differential expression ($Padj < 0.05$) were considered to be DEGs. It should be noted that one of our candidate genes for the galactose derivatives metabolism had two closely related copies (*NAGA* and *NAGA-like*, see below), which made it challenging to pinpoint the true source of their mapped reads. In order to differentiate the relative expression levels between the copies, we initially extracted all reads that map to the two copies using bowtie2 program [42] and samtools [43]. We then randomly selected five 50 bp variable segments (in each segment, at least seven variable sites occurred between the two gene copies) as baits, and directly counted the number of reads that match the baits using grep module of seqkit program [44].

3. Results

Based on morphological characteristics and DNA barcoding using mitochondrial COI sequences, six distinct *Andrena* species were discovered. Four species were recognized: *Andrena camellia*, *Andrena hunanensis*, *Andrena striata*, and *Andrena chekiangensis*. Since neither GenBank blasting nor BOLDSYSTEMS searching produced any COI hits for the two remaining species, they were temporarily designated as *Andrena sp. 1* and *Andrena sp. 2* (Table 1). A total of 62.74 giga bases (Gb) of WGS clean reads were obtained for the six *Andrena* species (five oil-tea specialized species and *A. chekiangensis*). After assembly 345–389 Mb of contigs were generated, with N50 contig size of 8.8–15.6 Kb (Table 2). Phylogenetic analysis showed that, the known three species (*A. camellia*, *A. hunanensis* and *A. striata*) and the unknown two species formed to a single clade, while *A. chekiangensis* and *A. haemorrhhoa* formed another clade (Figure 2).

Table 1. Sample information of six *Andrena* species collected from oil-tea blossoms.

Sample	Species	Location	Longitude	Latitude
XJ01	<i>Andrena camellia</i>	Xiajiang, Jiangxi	115.1285	27.6546
QY01	<i>Andrena hunanensis</i>	Qingyang, Anhui	117.8796	30.5977
RX01	<i>Andrena striata</i>	Rongxian, Sichuan	104.2913	29.4377
CN02	<i>Andrena sp. 1</i>	Cangnan, Zhejiang	120.2556	27.4591
DY03	<i>Andrena sp. 2</i>	Dongyuan, Guangdong	114.9792	24.1905
NX04	<i>Andrena chekiangensis</i>	Ningxiang, Hunan	112.4206	27.9832

Table 2. Short reads and assembly of next-generation genome of six *Andrena* species.

Species	Reads		Assembly	
	Length (Gb)	Accession	Length (Mb)	N50 (Kb)
<i>Andrena camellia</i>	10.38	SRR23869504	369.7	11.1
<i>Andrena hunanensis</i>	11.01	SRR23869503	384.3	8.2
<i>Andrena striata</i>	9.79	SRR23869502	393.1	8.5
<i>Andrena sp. 1</i>	10.75	SRR23869501	363.2	9.6
<i>Andrena sp. 2</i>	9.74	SRR23869500	353.5	14.7
<i>Andrena chekiangensis</i>	10.02	SRR23869499	365.7	11.6

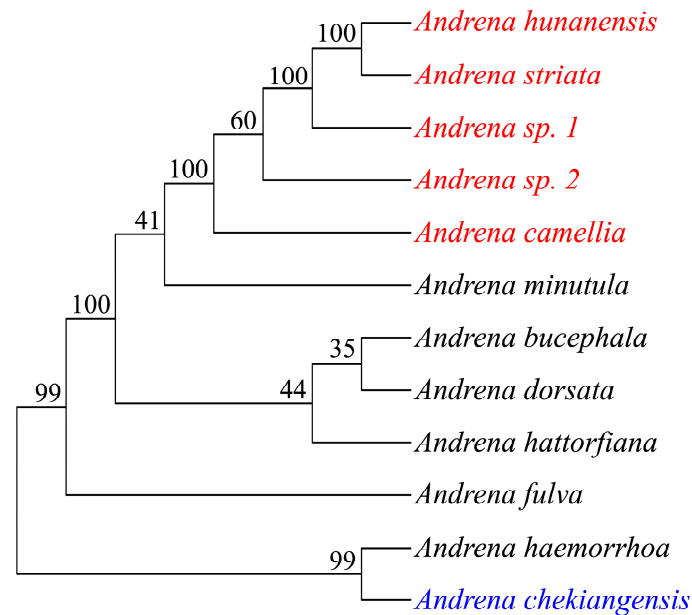


Figure 2. Phylogenetic relationship of 12 *Andrena* species analyzed in this study (red, specialized pollinator of oil-tea; blue, non-specialized pollinator of oil-tea; numbers beside each node represent percentages of bootstrap values).

According to the KEGG database, alpha-galactosidase (EC 3.2.1.22, also known as alpha-galactosidase A), which is prevalent in chordates, plants, and bacteria, is not found in arthropods such as bees (Hymenoptera) and other insects. As an alternative, bees have the equivalent alpha-N-acetylgalactosaminidase (*NAGA*, EC 3.2.1.49, also called alpha-galactosidase B) which is a homologous gene of alpha-galactosidase. Similar to other bees, there was only one *NAGA* gene in *A. chekiangensis* genome. Intriguingly, two closely similar copies of *NAGA* gene were discovered in the genome of the five oil-tea specialized species. One was the conventional *NAGA*, while the other appeared to be a novel copy of the conventional *NAGA*. For ease of use, we refer to the novel copy as a *NAGA-like* gene. The *NAGA* and *NAGA-like* genes were highly similar, taking *A. camellia* as an example, there were 86% and 78% identity sites between them at nucleotide and amino acid sequence levels, respectively. It was worth noting that *NANA-like* had a termination mutation in the last exon (the sixth exon) which resulted in a shortened protein (Figure 3). All genomes of the five oil-tea specialized species and the other seven species (including *A. chekiangensis*) contained four of the Leloir pathway genes [22], which most organisms use to metabolize galactose: aldose 1-epimerase (*galM*, EC 5.1.3.3), galactokinase (*galK*, EC 2.7.1.6), galactose-1-phosphate uridylyltransferase (*galT*, EC 2.7.7.12), and UDP-galactose 4-epimerase (*galE*, EC 5.1.3.2).

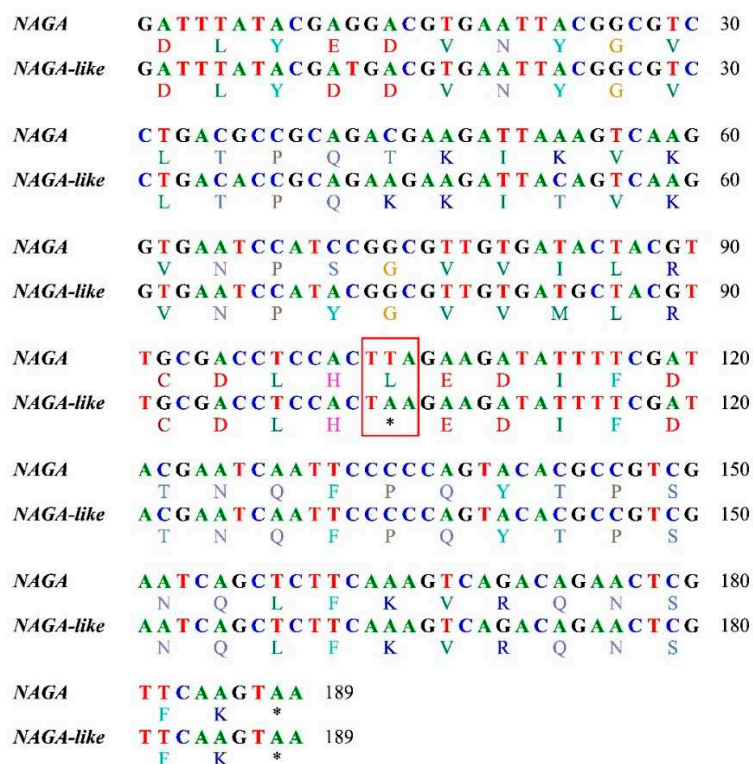


Figure 3. Alignment of nucleotide and amino acid sequences of the last exons of *NAGA* and *NAGA-like* genes of *Andrena camellia* (note the termination mutation in *NAGA-like*).

Genetic variations were surveyed within the five oil-tea specialized species. The coding sequence of *NAGA* was 1,320 bp in length, with 16 (1.21%) variable sites among the five species. The *NAGA-like* gene was shorter but interspecifically much more variable (2.66%) than *NAGA*. The four Leloir pathway genes were shorter than *NAGA* and *NAGA-like* and the number of variable sites ranked as *galM* > *galK* > *galT* > *galE* (Table 3). The sequences of the six genes of all the 12 *Andrena* species analyzed in this study were shown in the supplementary Table S1. Branch model analyses were executed with the five oil-tea specialized species as foreground clade and the remaining seven species as background clade. The results showed that, *NAGA-like*, *galK* and *galT* had significant greater *dN/dS* ratio in the foreground clade than in the background clade (χ^2 test, *df* = 1, *p* < 0.001). However, no significant divergence was seen for *NAGA*, *galM* and *galE* (*p* > 0.2) (Table 4). We also assess the putative positive selection sites in the two genes using the branch-site model test. With the Bayes Empirical Bayes (BEB) analysis, twelve sites in *NAGA-like*, four sites in *galK*, and one site in *galT* were found under positive selection with posterior probabilities > 0.95.

Table 3. Sequence length and genetic variation of genes involved in galactose derivatives metabolism within five oil-tea specialized *Andrena* species.

Gene	Length	Variable sites	Percent of variable sites
<i>NAGA</i>	1,320	16	1.21
<i>NAGA-like</i>	1,239	33	2.66
<i>galM</i>	1,077	29	2.69
<i>galK</i>	1,182	24	2.03
<i>galT</i>	1,152	15	1.30
<i>galE</i>	1,098	8	0.73

Table 4. Branch model analyses of genes involved in galactose derivatives metabolism.

Gene	Foreground	Background	$2\Delta\ln L$	p ($df = 1$)
NAGA	0.049	0.023	1.412	0.234
NAGA-like	0.680	0.021	145.673	<1.000e-10
galM	0.251	0.360	1.283	0.257
galK	0.864	0.161	24.279	8.336e-7
galT	0.387	0.088	12.761	3.540e-4
galE	0.129	0.063	0.913	0.339

A total of 33.7 Gb and 35.1 Gb of RNASeq clean reads were obtained for *A. camellia* and *A. chekiangensis*, respectively (Table 5). The assembly of *A. camellia* generated 23,087 unigenes, with a N50 value of 1,668 bp. For *A. chekiangensis*, 18,387 unigenes were produced with a N50 value of 1,692 bp. A total of 7,151 orthologs were shared by the two species, with a total length of 7,847,865 bp and N50 of 1,632 bp. The average value of relative expression level (TPM) of these orthologs in each sample were 140. The TPM values of the six candidate galactose metabolism genes are shown in Table 6. Taking the *A. chekiangensis* samples as the control group, 1,987 differentially expressed genes were detected, including 1,155 up-regulated ($FC > 2$) and 853 down-regulated ($FC < 0.5$) genes in *A. camellia*. NAGA-like, galK, and galT were significantly up-regulated in *A. camellia* ($FC > 2$, $P_{adj} < 0.05$), while NAGA, galM and galE were not deviated in expression levels between *A. camellia* and *A. chekiangensis* ($P_{adj} > 0.05$) (Figure 4).

Table 5. RNASeq clean reads of *Andrena camellia* and *Andrena chekiangensis*.

Sample	Accession	Length (Gb)	Q30 (%)	GC (%)
Acam1	SRR8335252	7.54	91.55	45.84
Acam2	SRR8335251	9.50	92.41	46.35
Acam3	SRR8335254	7.64	91.76	46.23
Acam4	SRR8335253	9.04	92.35	45.97
Ache1	SRR23869498	9.03	95.82	41.95
Ache2	SRR23869497	8.88	96.10	42.46
Ache3	SRR23869496	7.82	96.16	41.61
Ache4	SRR23869495	9.36	95.81	43.69

Table 6. Gene expression analyses of genes involved in galactose derivatives metabolism.

Gene	TPM (Mean \pm SD)		FC	P _{adj}
	<i>Andrena camellia</i>	<i>Andrena chekiangensis</i>		
NAGA	574.88 \pm 162.33	740.45 \pm 470.32	0.639	0.387
NAGA-like	69,465.93 \pm 7690.83	0	$+\infty$	$-\infty$
galM	537.64 \pm 78.69	210.82 \pm 211.28	2.14	0.118
galK	172.87 \pm 47.92	41.77 \pm 34.05	3.35	0.016
galT	987.77 \pm 139.04	318.55 \pm 57.03	2.99	1.781E-12
galE	32.45 \pm 10.07	21.82 \pm 14.73	1.23	0.624

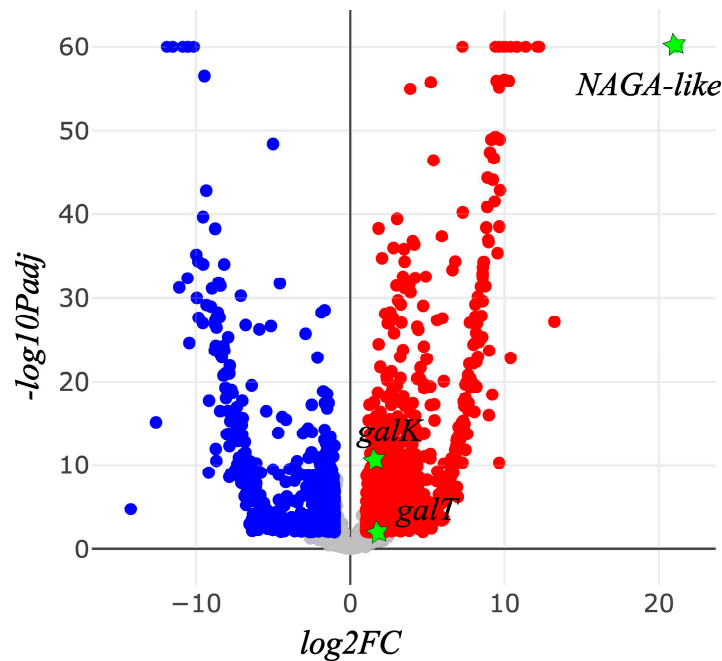


Figure 4. Differentially expressed genes of *Andrena camellia* against *Andrena chekiangensis* (blue dots, down-regulated; red dots, up-regulated; green stars, the three up-regulated genes involved in galactose derivatives metabolism).

4. Discussion

Oil-tea is an important woody edible and industrial oil tree species [45]. Its products, tea oil, was categorized by the FAO (Food and Agriculture Organization of the United Nations) as a premium health-grade edible oil [46]. This plant presents a low oil yield because of self-incompatibility. Previous studies showed that the oil yield can be improved by an increase in pollinating insects [19,47]. However, blooms occur in late autumn and winter (from October to January), when bee pollinators are few due to cold temperatures. Additionally, some compounds in the nectar are toxic to most bees, including managed honey bees [13]. It's interesting to note that the poisonous elements in oil-tea nectar can be detoxified by both adults and larvae of several *Andrena* species. Unfortunately, despite the significant attention these species have received [47–50], little is known about the molecular mechanisms of detoxification. In this study, we present the first genome and transcriptome sequencing on oil-tea specialized *Andrena* species and carried out bioinformatic analyses on genes involved in galactose derivatives metabolism.

As stated in Introduction, in order to degrade galactose derivatives, the alpha-galactosidine bonds need to be firstly hydrolyzed to release galactose residues. In most organisms such as chordates, plants, and bacteria this process is accomplished by alpha-galactosidase A. However, honey bees and other insects have not yet been found to contain such a gene. Instead, *NAGA*, a homologous gene of *alpha-galactosidase A*, was commonly present in insect genomes. Although the protein produced by *NAGA* has been initially thought to be an isozyme of α -galactosidase and given the name α -galactosidase B, it was actually an exoglycosidase acting on N-acetylgalactosamine [51]. There is no proof that it can replace alpha-galactosidase A's role, which is why honey bees (such as *Apis mellifera* and *Apis cerana*) are unable to process oil-tea nectar. According to our molecular evolution analyses, there was no discernible selective differentiation of *NAGA* between the five oil-tea specialized species and the other *Andrena* species. According to gene expression assessments, *NAGA* in *A. camellia* was not significantly deviated from that in *A. chekiangensis* ($p > 0.05$). As a result, we hypothesize that the conventional *NAGA* makes no contribution to the detoxification of oil-tea specialized *Andrena* species.

The most intriguing discovery in this study might be the novel copy of *NAGA*, say *NAGA-like* gene, in the five oil-tea specialized *Andrena* species. Such gene duplication pattern was not found in the other seven *Andrena* species, including *A. chekiangensis* which also consume oil-tea nectar, although unspecifically. We also examined the genomes and transcriptomes of *Colletes gigas*, another crucial oil-tea pollinator [17,19], and no novel copy was found. Considering that the five oil-tea specialized *Andrena* species formed to a monophyletic group, we speculated that the *NAGA-like* gene was created from a particular gene duplicating event occurred in the common ancestor of these species. Molecular evolution analyses with branch models and branch-site models indicated that *NAGA-like* was under strong positive selections, a sign that a new phenotype on this gene would arise for these species [52]. The seqkit grep counts showed that the majority of reads mapping on *NAGA* were actually from *NAGA-like*. As a result, it is possible to estimate that the expression level of *NAGA-like* in *A. camellia* is ~120 times of *NAGA* in the same species, or ~94 times of *NAGA* in *A. chekiangensis*. Additionally, given that each of the 7,151 orthologs had an average expression level (TPM) of 140, *NAGA-like* had a TPM that was around 496 times of the average value. Such a high degree of *NAGA-like* expression suggested that it is essential for the oil-tea specialization in *A. camellia*, and maybe the other four oil-tea specialized species. Since the novel *NAGA-like* gene was highly similar (86% DNA identity) with conventional *NAGA*, it was logical to assume that *NAGA-like* protein could likewise catalyze N-acetylgalactosamine residue. In other words, although more research is required to confirm this idea, we propose that *NAGA-like* may have acquired a new role to break the alpha-galactosidine linkages from galactose derivatives.

There are four steps in the classic Leloir pathway. First, β -d-galactose (natural galactose) is epimerized to α -d-galactose by *galM*. Second, α -d-galactose is phosphorylated to yield α -d-galactose 1-phosphate by *galK*. Third, *galT* catalyzes the transfer of a UMP group from UDP-glucose to galactose 1-phosphate, thereby generating glucose 1-phosphate and UDP-galactose. Finally, UDP-galactose is converted to UDP-glucose by *galE* [22]. Our results of the branch model and branch-site model tests showed that, *galK* and *galT*, but not *galM* nor *galE*, showed significant larger *dN/dS* ratios in the five oil-tea specialized species than in the background *Andrena* species. Moreover, the gene expression analyses showed that *galT* and *galT*, but not *galM* nor *galE*, was significantly upregulated in *A. camellia* than in *A. chekiangensis*. These findings suggested that *galK* and *galT* may have been crucial in helping the oil-tea specialized *Andrena* species adapt.

Galactokinase (GALK) and galactose-1-phosphate uridylyltransferase (GALT) are the most important proteins for galactose metabolism in humans. GALT deficiency causes the inborn error of metabolism known as "classic galactosemia" (type I galactosemia); while deficiency causes type II galactosemia, which is the most clinically severe form of the disease [53]. The galactosemia disease manifests soon after birth if the newborn is exposed to galactose, with acute symptoms such as failure to thrive, hepatocellular damage, bleeding, and possible death from bacterial sepsis [54]. Catalyzing efficiency of GALK and GALT are sensitive to its amino acid mutations [53]. It should be mentioned that mutations themselves are inherently directionless. For instance, in human, some mutations in *galK* gene may result in significant decreases in catalyzing efficiency and lead to diseases, while others may, in the opposite direction, result in an increase in catalyzing function [55]. We hypothesize that an improvement in catalytic efficiency may result from positive selection in *galK* and *galT* of the oil-tea specialized species. In contrast, although *galM* and *galE* were also involved in galactose metabolism, no significant deviations in molecular evolution and gene expression were detected between *A. camellia* than in *A. chekiangensis*, suggesting that the universal activity and quantity of these two epimerases are sufficient to deal with the catalytic demand in the oil-tea specialized species.

5. Conclusions

Our study clearly demonstrated that the genes involved in galactose derivatives metabolism were crucial in the evolution of the oil-tea specialized *Andrena* species. A novel *NAGA-like* gene was created to aid in the hydrolysis of the galactose residue from galactose derivatives, while the *galK* and *galT* genes were functionally improved to speed up the metabolism of the hydrolyzed galactose. It should be noted that despite the fact that these species can handle poisonous oligosaccharides, their

too-small population densities make it appear as though they are unable to meet the pollination needs of oil-tea. Our findings would provide insight into the poisoning and detoxify processes of various bee species. We propose that genetic engineering of relevant genes in cultivated species such as *Apis mellifera* may finally assist in meeting the enormous pollination needs.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.zenodo.org/deposit/7898872#>, DOI: 10.5281/zenodo.7898872. Table S1: Genes involved in galactose derivatives metabolism used in this study.

Author Contributions: Conceptualization, G.L., F.Z. and Z.H.; methodology, G.L., B.H. and F.Z.; software, G.L.; validation, formal analysis, investigation, resources, data curation, visualization, B.H., T.S., K.J.; writing—original draft preparation, writing—review and editing, G.L. and F.Z.; supervision, project administration and funding acquisition, F.Z. and Z.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Jiangxi, China (No. 20212BAB215024, 20212ACB205006), the Jiangxi “Double Thousand Plan” (No. jxsq2020101050), and the Science and Technology Foundation of Jiangxi Provincial Department of Education (No. GJJ170655, GJJ190538).

Data Availability Statement: All data are presented in the text.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Peters, R.S.; Krogmann, L.; Mayer, C.; Donath, A.; Gunkel, S.; Meusemann, K.; Kozlov, A.; Podsiadlowski, L.; Petersen, M.; Lanfear, R.; et al. Evolutionary history of the Hymenoptera. *Curr. Biol.* **2017**, *27*, 1013–1018.
- Michener, C.D. *The Bees of the World*, 2nd ed. The Johns Hopkins University Press: Baltimore, USA, 2007.
- Potts, S.G.; Imperatriz-Fonseca, V.; Ngo, H.T.; Aizen, M.A.; Biesmeijer, J.C.; Breeze, T.D.; Dicks, L.V.; Garibaldi, L.A.; Hill, R.; Settele, J.; Vanbergen, A.J. Safeguarding pollinators and their values to human well-being. *Nature* **2016**, *540*(7632), 220–229.
- Cane, J.H.; Sipes, S. Characterizing floral specialization by bees: analytical methods and revised lexicon for oligolecty. In *Plant-pollinator Interactions: from Specialization to Generalization*; Waser, N.M., Ollerton, J., Eds. The University of Chicago Press: Chicago, USA, 2006; pp. 99–121.
- Goulson, D. *Bumblebees: Behaviour, Ecology, and Conservation*. Oxford University Press Inc.: New York, USA, 2010.
- Minckley, R.L.; Roulston, T.H. Incidental mutualisms and pollen specialization among bees. In *Plant-pollinator Interactions: from Specialization to Generalization*; Waser, N.M., Ollerton, J., Eds. The University of Chicago Press: Chicago, USA, 2006; pp. 69–98.
- Bossert, S.; Wood, T.J.; Patiny, S.; Michez, D.; Almeida, E.A.B.; Minckley, R.L.; Packer, L.; Neff, J.L.; Copeland, R.S.; Straka, J.; et al. Phylogeny, biogeography and diversification of the mining bee family Andrenidae. *Syst. Entomol.* **2022**, *47*(2), 283–302.
- Larkin, L.L.; Neff, J.L.; Simpson, B.B. The evolution of pollen diet: host choice and diet breadth of *Andrena* bees (Hymenoptera: Andrenidae). *Apidologie* **2008**, *39*, 133–145.
- Wood, T.J.; Roberts, S.P.M. An assessment of historical and contemporary diet breadth in polylectic *Andrena* bee species. *Biol. Conserv.* **2017**, *215*, 72–80.
- Luan, F.; Zeng, J.; Yang, Y.; He, X.; Wang, B.; Gao, Y.; Zeng, N. Recent advances in *Camellia oleifera* Abel: A review of nutritional constituents, biofunctional properties, and potential industrial applications. *J. Funct. Food.* **2020**, *75*, 104242.
- Wang, X.N. Research on phenology and blossom biology of oil-tea *Camellia*. Master Thesis, Central South University of Forestry and Technology, Changsha, China, 2011.
- Huang, D.Y.; Ding, L.; Zhang, Y.Z.; Huang, H.R.; Yu, J.F.; Hao, J.S.; Zhu, C.D. Life history and relevant biological features of *Andrena camellia* Wu (Hymenoptera: Andrenidae). *Acta Entomol. Sin.* **2008**, *51*(7), 778–783.
- Xie, Z.; Chen, X.; Qiu, J. Reproductive failure of *Camellia oleifera* in the plateau region of China due to a shortage of legitimate pollinators. *Int. J. Agric. Biol.* **2013**, *15*, 458–464.
- Zhao, S.W. Management measure for honey colony in flowering period of *Camellia oleifera*. *Apicult. China* **1993**, *5*, 19–20.
- Ding, L.; Huang, D.Y.; Zhang, Y.Z.; Huang, H.R.; Li, J.; Zhu, C.D. Observation on the nesting biology of *Andrena camellia* Wu (Hymenoptera: Andrenidae). *Acta Entomol. Sin.* **2007**, *50*(10), 1077–1082.

16. He, B.; Su, T.J.; Niu, Z.Q.; Zhou, Z.Y.; Gu, Z.Y.; Huang, D.Y. Characterization of mitochondrial genomes of three *Andrena* bees (Apoidea: Andrenidae) and insights into the phylogenetics. *Int. J. Biol. Macromol.* **2019**, *127*, 118–125.
17. Su, T.J.; He, B.; Zhao, F.; Jiang, K.; Lin, G.; Huang, Z. Population genomics and phylogeography of *Colletes gigas*, a wild bee specialized on winter flowering plants. *Ecol. Evol.* **2022**, *12*(4), e8863.
18. Wu, Y.R. The pollinating bees on *Camellia olifera* with descriptions of 4 new species of the genus *Andrena*. *Acta Entomol. Sin.* **1977**, *20*, 199–204.
19. Li, H.Y.; Luo, A.C.; Hao, Y.J.; Dou, F.Y.; Kou, R.M.; Orr, M.C.; Zhu, C.D.; Huang, D.Y. Comparison of the pollination efficiency of *Apis cerana* with wild bees in oil-seed camellia fields. *Basic Appl. Ecol.* **2021**, *56*, 250–258.
20. Kang, X.D.; Fan, Z.Y. Toxic contents of nectar of oil-tea flowers to honey bees. *J. Bee* **1991**, *1*, 8–10.
21. Li, Z.; Huang, Q.; Zheng, Y.; Zhang, Y.; Li, X.; Zhong, S.; Zeng, Z. Identification of the toxic compounds in *Camellia oleifera* honey and pollen to honey bees (*Apis mellifera*). *J. Agric. Food Chem.* **2022**, *70*, 13176–13185.
22. Holden, H.M.; Rayment, I.; Thoden, J.B. Structure and function of enzymes of the Leloir pathway for galactose metabolism. *J. Biol. Chem.* **2003**, *278*(45), 43885–43888.
23. Vinson, C.C.; Mota, A.P.Z.; Porto, B.N.; Oliveira, T.N.; Sampaio, I.; Lacerda, A.L.; Danchin, E.G.J.; Guimaraes, P.M.; Williams, T.C.R.; Brasileiro, A.C.M. Characterization of raffinose metabolism genes uncovers a wild *Arachis* galactinol synthase conferring tolerance to abiotic stresses. *Sci. Rep.* **2020**, *10*(1), 15258.
24. Elango, D.; Rajendran, K.; Van der Laan, L.; Sebastiar, S.; Raigne, J.; Thaiparambil, N.A.; El Haddad, N.; Raja, B.; Wang, W.; Ferela, A.; Chiteri, K.O.; Thudi, M.; Varshney, R. K.; Chopra, S.; Singh, A.; Singh, A.K. Raffinose family oligosaccharides: friend or foe for human and plant health? *Front. Plant Sci.* **2022**, *13*, 829118.
25. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*(17), i884–i890.
26. Li, D.; Liu, C.M.; Luo, R.; Sadakane, K.; Lam, T.W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*(10), 1674–1676.
27. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421.
28. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*(1), 268–274.
29. Slater, G. S. C., Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **2005**, *6*, 31.
30. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.
31. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **2017**, *34*(12), 3299–3302.
32. Sievers, F.; Higgins, D.G. Clustal omega. *Curr. Protoc. Bioinf.* **2014**, *48*, 3.13.1–3.13.16.
33. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591.
34. Yang, Z.; Wong, W.S.W.; Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118.
35. Zhang, J.; Nielsen, R.; Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **2005**, *22*, 2472–2479.
36. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.D.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*(7), 644–U130.
37. Li, W.Z., Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*(13), 1658–1659.
38. Tang, S.; Lomsadze, A.; Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucl. Acid. Res.* **2015**, *43*(12), e78.
39. Emms, D.M.; Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*(1), 238.
40. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A., Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14* (4): 417–419.
41. Kucukural, A.; Yukselen, O.; Ozata, D.M.; Moore, M.J.; Garber, M. DEBrowser: interactive differential expression analysis and visualization tool for count data. *BMC Genomics* **2019**, *20*(1), 6.
42. Langmead, B.; Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359.
43. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*(16), 2078–2079.

44. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q File manipulation. *PLoS ONE* **2016**, *11*(10), e0163962.
45. Wen, Y.; Su, S.C.; Ma, L.Y.; Yang, S.Y.; Wang, Y.W.; Wang, X.N. Effects of canopy microclimate on fruit yield and quality of *Camellia oleifera*. *Sci. Horticult.* **2018**, *235*, 132–141.
46. Feng, J.L.; Jiang, Y.; Yang, Z.J.; Chen, S.P.; El-Kassaby, Y.A.; Chen, H. Marker-assisted selection in *C. oleifera* hybrid population. *Silvae Genet.* **2020**, *69*(1), 63–72.
47. Deng, Y.; Yu, X.; Liu, Y. The role of native bees on the reproductive success of *Camellia oleifera* in Hunan Province, Central South China. *Acta Ecol. Sin.* **2010**, *30*, 4427–4436.
48. Huang, D.; Su, T.; Qu, L.; Wu, Y.; Gu, P.; He, B.; Xu, X.; Zhu, C. The complete mitochondrial genome of the *Colletes gigas* (Hymenoptera: Colletidae: Colletinae). *Mit. DNA Part A* **2016**, *27*(6), 3878–3879.
49. Huang, D.Y.; He, B.; Gu, P.; Su, T.J.; Zhu, C.D. Discussion on current situation and research direction of pollination insects of *Camellia oleifera*. *J. Environ. Entomol.* **2017**, *39*, 213–220.
50. Zhou, Q.S.; Luo, A.; Zhang, F.; Niu, Z.Q.; Wu, Q.T.; Xiong, M.; Orr, M.C.; Zhu, C.D. The first draft genome of the plasterer bee *Colletes gigas* (Hymenoptera: Colletidae: Colletes). *Genome Biol. Evol.* **2020**, *12*(6), 860–866.
51. Michalski, J.C.; Klein, A. Glycoprotein lysosomal storage disorders: K- and L-mannosidosis, fucosidosis and K-N-acetylgalactosaminidase deficiency. *Biochim. Biophys. Acta* **1999**, *1455*, 69–84.
52. Vallender, E.J.; Lahn, B.T. Positive selection on the human genome, *Human Mol. Genet.* **2004**, *13*, R245–R254.
53. Pasquali, M.; Yu, C.; Coffee, B. Laboratory diagnosis of galactosemia: a technical standard and guideline of the American college of medical genetics and genomics (ACMG). *Genet. Med.* **2018**, *20*, 3–11.
54. Verdino, A.; D'Urso, G.; Tammone, C.; Scafuri, B.; Marabotti, A. Analysis of the structure-function-dynamics relationships of GALT enzyme and of its pathogenic mutant p.Q188R: a molecular dynamics simulation study in different experimental conditions. *Molecules* **2021**, *26*, 5941.
55. Timson, D.J.; Reece, R.J. Functional analysis of disease-causing mutations in human galactokinase. *Eur. J. Biochem.* **2003**, *270*(8), 1767–1774.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.