Transcriptomic profiles of 33 opium poppy samples in different tissues, growth phases, and cultivars

Yucheng Zhao¹, Zhaoping Zhang², Mingzhi Li³, Jun Luo¹, Fang Chen², Yongfu Gong², Yanrong Li², Yujie Wei², Yujie Su² & Lingyi Kong¹

Opium poppy is one of the most important medicinal plants and remains the only commercial resource of morphinan-based painkillers. However, little is known about the regulatory mechanisms involved in benzylisoquinoline alkaloids (BIAs) biosynthesis in opium poppy. Herein, the full-length transcriptome dataset of opium poppy was constructed for the first time in accompanied with the 33 samples of Illumina transcriptome data from different tissues, growth phases and cultivars. The long-read sequencing produced 902,140 raw reads with 55,114 high-quality transcripts, and short-read sequencing produced 1,923,679,864 clean reads with an average Q30 rate of 93%. The high-quality transcripts were subsequently quantified using the short reads, and the expression of each unigene among different samples was calculated as reads per kilobase per million mapped reads (RPKM). These data provide a foundation for opium poppy transcriptomic analysis, which may aid in capturing splice variants and some non-coding RNAs involved in the regulation of BIAs biosynthesis. It can also be used for genome assembly and annotation which will favor in new transcript identification.

Background & Summary

Opium poppy (Papaver somniferum) is one of the most important medicinal plants in the world and remains the only commercial resource of morphinan-based painkillers. Its main active ingredient, BIAs, also displays potential pharmacological activity in relieving cough, muscle relaxation, anticancer, and so on. Although approximately 100,000 hectares (ha) of opium poppy are cultivated annually worldwide, this insufficient to meet the demand for managing moderate or severe pain. Engineered microbes could be used to produce BIAs such as opiates and noscapine, however, major hurdles, such as low yield and unclear biosynthetic pathways, make it different to scale-up this method of production for most BIAs. Hence, how to guarantee the source of BIAs to meet the medical applications, such as in pain relief or palliative care, has been recognized as a major issue that needs to be resolved.

Investigation into the BIA biosynthesis mechanism began in the 1960s with radiotracer technology. The emergence of transcriptomics, proteomics, and metabolomics, coupled with recent genome analysis tools, accelerated the discovery of new BIA biosynthetic genes that could facilitate metabolic engineering reconstitution of commercial source of valuable BIAs in microbes. However, some key steps in the BIA biosynthesis pathway are yet to be identified, and there has been a lack of investigation into the molecular mechanisms of gene regulation in the pathway. In addition, there are few reports major in the processes of BIAs biosynthesis, such as compound dynamic accumulation, tissue-specific distribution, enzyme interactions and metabolism, compartmentalization and transport. Therefore, investigation into the regulation mechanism and understanding the compound accumulation process is a key way to improve the yield of BIAs.

Of all the strategies in metabolic regulation, over-expression and silencing of genes involved in the metabolism of target compounds are the most widely used method. However, non-coding RNAs, alternative splicing/translation/polyadenylation (AS/AT/APA), formation of heterodimers, and gene fusion have also been shown to increase the flexibility of the transcriptome, functional complexity of plants, and the trend of metabolic flow,

¹Jiangsu Key Laboratory of Bioactive Natural Product Research and State Key Laboratory of Natural Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, No. 24 Tongjiaxiang, Nanjing, 210009, China. ²China Agriculture Research System (CARS-21), No. 234 Xinzhen Road, Huangyang town, Liangzhou District, Wuwei, Gansu, 733006, China. ³Genepioneer Biotechnologies Co. Ltd., No. 9 Weidi Road, Qixia District, Nanjing, 210014, China. Correspondence and requests for materials should be addressed to L.K. (email: cpu_lykong@126.com)
tissue-specific accumulation or product yield\textsuperscript{8,9,12,13}. However, little research has been conducted on the regulatory mechanisms in opium poppy despite the recently release of its genome\textsuperscript{1}. The reason for this oversight may be the lack of the transcriptional information on opium poppy in different growth periods/status or tissues. In addition, the short sequences created by third-generation sequencing and the gene information in DNA standard could not capture the AS/AT/APA of transcripts\textsuperscript{1,3}.

Third-generation single-molecule real-time (SMRT, Pacific Biosciences) sequencing has increasingly been used to detect AS/AT/APA, to identify novel isoforms, to predict non-coding RNA, and for gene fusion studies due to its long reads\textsuperscript{14,15}. However, to date, there has been no report on a full-length transcriptome dataset of opium poppy, despite the recent release of its genome\textsuperscript{1}. Herein, we construct the full-length transcriptome dataset of opium poppy using twenty-one pooled RNA samples from three different tissues and five different growth phases (Fig. 1). This produced 902,140 post-filter polymerase reads and 660,418 circular consensus sequence (CCS) reads (Table 1 and Fig. 2). After data processing, 566,746 full-length (FL) reads, 180,511 non-redundant isoforms and 61,856 unigenes (59,144 protein-coding unigenes and 2,712 non-coding unigenes) were obtained for functional annotation (Fig. 2). Gene expression levels were then determined using RSEM and converted into fragments per kb per million fragments (FPKM) value (Fig. 4)\textsuperscript{16,17}. The dataset reported here, provides an overview of the gene expression
levels, AS/AT/APA, and full-length transcript/unigene/mRNA of the opium poppy during key statuses. It can also be used to analyze the regulation mechanisms of BIAs biosynthesis according to its tissue-specific distribution, dynamic accumulation in different growth phases as well as BIAs diversity in different germplasm resources.

**Methods**

**Plant material and experimental design.** All the opium poppy for this study was cultured in our experimental plot. The original cultivar was named B1 and represents the original state of opium poppy, known as wild opium poppy. Other germplasm resources resulting from directed breeding in our institute were named B2, T, G and Z. Their characteristics and major BIAs content are summarized in Tables 3 and 4, respectively. Samples...
were divided into different tissues, cultivars, and growth phases (Fig. 1). Samples from different tissues in B1 (root, stem, and leaves) and germplasm resources (B1, B2, T, G, ZB) were collected during filling stage on Jul 4, 2017. Samples in different growth phases (seedling, jointing, bolting, florescence, and filling) were collected from March to July of 2017. For each sampling point (eleven in all), three independent samples were collected. Unless otherwise mentioned, samples collected from opium poppy are all leaves. This produced 33 samples in all and their definitions of sampling are also listed in Table 3. The leaves from B1 at the filling stage were used as a shared sample for subsequent analysis: they were used as the filling stage sample when studying the growth rhythm of opium poppy, and also as a leaf sample in different tissues or a B1 sample in different cultivars. After washing and cleaning, the samples were immediately inundated with liquid nitrogen for 10 minutes and then stored in −80 °C freezers until use.

**RNA extraction, library preparation, and sequencing.** For each sample, TRIzol Reagent (Tiangen Biotech Co., Ltd., Beijing, China) was used to extract RNA following the protocol provided by the manufacturer. After cDNA synthesis, samples were subjected to phosphorylation, "A" base addition, and end-repair according to library construction protocol. Sequencing adapters were then added to both sizes of the cDNA fragments. After PCR amplification of cDNA fragments, the 150–250 bp targets were cleaned up. We then performed paired-end sequencing on an Illumina HiSeq X Ten platform (Illumina Inc, CA, USA) (Fig. 1). For PacBio Sequel sequencing, RNA from 21 samples of B1 was mixed in equal amounts for reverse transcription using the Clontech SMARTer PCR cDNA Synthesis Kit (TaKaRa, Dalian, China). In order to determine the optimal amplification cycle number for the downstream large-scale PCR reactions, PCR cycle optimization was employed (PrimeSTAR® GXL DNA polymerase). Then the optimized cycle number was used to generate double-stranded cDNA. Large-scale PCR was performed for SMRTbell library construction (Pacific Biosciences). This include DNA damage repair, end repair, ligating sequencing adapters and removing fragments that failed to connect. Finally, the SMRTbell template was annealed to the sequencing primer, bound to polymerase, and sequenced on the PacBio Sequel platform using V2.1 chemistry (Pacific Biosciences) with 10-hour movies (Fig. 1).

**Fig. 3** Correlation analysis of repeated samples and saturation curve of transcripts and genes. (a) Correlation analysis of repeated samples. Heatmap displaying the similarities among all samples based on Poisson distances. (b) Saturation curve of transcripts and genes. The saturation curve shows with the incense of FL reads, the number of genes tends to remain flat, while the number of transcripts rises rapidly.
Data filtering, processing and yield. After sequencing, the raw reads were classified and clustered into a transcript consensus using the SMRT Link 5.1 pipeline (http://www.pacb.com/products-and-services/analytical-software/smrt-analysis/). After adaptor removal and elimination of low quality regions, we obtain 902,140 post-filter polymerase reads (19.77 GB) with an average length of 22 kb (Fig. 2 and Table 1). In order to improve the accuracy of sequencing, CCS reads were extracted from the subreads BAM file, which produced a total of 660,418 CCS reads with an average insert length of 2.61 kb (Fig. 2a,b). Briefly, CCS reads were extracted out of subreads.bam file with minimum full pass of 1 and a minimum read score of 0.8. CCS reads were then classified into full-length (FL) non-chimeric (NC), non-full-length (NFL), chimeras (C), and short reads based on cDNA primers and poly-A tail signal. Reads shorter than 50 bp were discarded. CCS reads with 5′ primer, 3′ primer and polyA tails were identified as FL reads, and 566,746 FL sequences ranging from 300 bp to 25,247 bp were obtained (Fig. 2c,d). Subsequently, the full-length non-chimeric (FLNC) reads were clustered by Iterative Clustering for Error Correction (ICE) software to generate the cluster consensus isoforms18. NFL reads were used by Arrow software to polish the obtained cluster consensus isoforms to obtain the final 55,114 FL polished high quality consensus sequences (accuracy ≥ 99%)18. Lordec was used to correct FL transcripts and CD-HIT was used to remove redundant sequences according to sequence similarity of high-quality transcripts19,20. Finally, 180,511 non-redundant isoforms and 61,856 unigenes were obtained for functional annotation (Fig. 2e,f). For Illumina paired-end RNA-seq, the low-quality reads (reads containing sequencing adaptors, reads containing sequencing primers, nucleotide with q quality score lower than 20) were removed. After that, a total of 1,923,679,864 clean, paired-end reads were produced (Table 2 and Fig. 3a). Illumina clean data was mapped onto our SMRT sequencing data using hisat2 v2.0521. RSEM was used to identify gene expression levels, which were then converted into an FPKM value16,17. DESeq R package was used to analyze differential expression22. Fold change ≥ 2 and adjusted P-value < 0.05 were set as threshold for significance of gene expression differences between the two samples (Fig. 4).

| Sample | Index | Clean Bases | Q30 Rate (%) | RPKM |
|--------|-------|-------------|--------------|------|
|        |       |             |              | 0.0–0.1 | 0.1–3.75 | 3.75–15 | >15  |
| 131    | TTAGGC | 8713642719  | 95           | 16,032 | 29,642   | 9,636   | 6,546 |
| 132    | TGACCA | 9564232510  | 96           | 16,532 | 29,732   | 9,285   | 6,307 |
| 133    | ACAGTG | 8107787693  | 96           | 16,599 | 29,566   | 9,314   | 6,377 |
| 231    | GCAAT  | 8956735582  | 96           | 14,703 | 28,829   | 11,103  | 7,221 |
| 331    | GATC   | 8796177410  | 96           | 14,664 | 28,591   | 11,303  | 7,298 |
| 332    | TGGCTT | 1071069418  | 96           | 15,936 | 29,443   | 9,956   | 6,521 |
| 333    | GGGCAT | 9433838049  | 96           | 15,573 | 29,564   | 10,177  | 6,542 |
| 334    | CTTGTA | 8512383596  | 96           | 15,180 | 28,368   | 10,846  | 7,462 |
| 335    | AGTCAA | 8863911611  | 96           | 15,682 | 27,710   | 10,627  | 7,837 |
| 336    | AGTCC  | 9205308102  | 96           | 14,991 | 27,757   | 11,208  | 7,900 |
| 431    | GCCAA  | 7080192660  | 94           | 15,480 | 26,114   | 12,024  | 8,238 |
| 432    | AGTCAA | 6668054022  | 96           | 15,392 | 26,004   | 11,811  | 8,649 |
| 433    | GCCAA  | 8814911815  | 96           | 15,470 | 28,361   | 10,500  | 7,525 |
| 513    | CAGTC  | 8857730898  | 96           | 15,831 | 26,107   | 11,330  | 8,588 |
| 512    | ACAGTG | 8354907591  | 96           | 16,108 | 26,313   | 10,792  | 8,643 |
| 511    | GATCCA | 6919453352  | 96           | 17,249 | 28,572   | 8,458   | 7,577 |
| 510    | TGGCTT | 8282182945  | 96           | 15,611 | 28,124   | 10,408  | 7,713 |
| 512    | GGCTAC | 6896148722  | 95           | 15,771 | 28,693   | 9,866   | 7,426 |
| 513    | TGGCTA | 8631197062  | 95           | 14,860 | 28,393   | 10,683  | 7,920 |
| 521    | GTGTA  | 7080192660  | 94           | 15,480 | 26,114   | 12,024  | 8,238 |
| 522    | AGTCAA | 6668054022  | 96           | 15,392 | 26,004   | 11,811  | 8,649 |
| 523    | GCCAA  | 8814911815  | 96           | 15,470 | 28,361   | 10,500  | 7,525 |
| 534    | GGCTAC | 6896148722  | 95           | 15,771 | 28,693   | 9,866   | 7,426 |
| 535    | TGGCTA | 8631197062  | 95           | 14,860 | 28,393   | 10,683  | 7,920 |
| 532    | GGCTAC | 6896148722  | 95           | 15,771 | 28,693   | 9,866   | 7,426 |
| 533    | GGCTAC | 8631197062  | 95           | 14,860 | 28,393   | 10,683  | 7,920 |
| 631    | GGCTAC | 8631197062  | 95           | 14,860 | 28,393   | 10,683  | 7,920 |
| 632    | GGCTAC | 8631197062  | 95           | 14,860 | 28,393   | 10,683  | 7,920 |
| 633    | GGCTAC | 8631197062  | 95           | 14,860 | 28,393   | 10,683  | 7,920 |

Table 2. Statistics of Illumina-based RNA-seq data and quantification of gene expression.
To analysis the major BIAs content, some 0.5 g dry leaves in each sample of different germplasm lines at filling stag were employed according to our previous publicized method. For details, the sample was sequentially extracted three times in methanol, with the help of ultrasonication, for 30 min at room temperature. Then, methanol extracts were combined and concentrated under reduced pressure conditions to a volume of 2 mL. At last, 10 μL concentrated extracts was subject to analysis using HPLC equipped with a reversed phase C18 column (XDB-C18, 5 mm; Agilent, USA) according to our publicized method.

**Table 3.** Description and definition of the main characteristics of opium poppy in each germplasm lines, growth phase, and tissue.

| Groups          | Sample | Description                                                                                                                                                                                                 |
|-----------------|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Germplasm lines | B1     | Four white petals; white seed; plant height up to 115 cm and have an average leaves of 15; white flower with only one fruit                                                                                      |
|                 | B2     | Two green sepals with 4–6 white petals; multi-branching; plant height up to 106 cm and usually have 13 leaves; white seed                                                                                         |
|                 | T      | It has an average plant height of 110 cm and 12 leaves; White flower with purple spots; juice in white and filament in white; multi-branching which could produce at least three fruit |
|                 | Z      | It has an average plant height of 105 cm and 11 leaves; White flower with purple spots; juice in light red and filament in white; multi-branching                                                                 |
|                 | G      | It has an highest plant height of 125 cm; 15 leaves; pink flower and golden anther; grey seed; juice in light red and filament in white                                                                        |
| Growth phases   | Seedling | Twenty days after sprouting                                                                                                                                                                                |
|                 | Jointing | Sixty days after sprouting. In this time, a height of 50 cm stem could be observed                                                                                                                        |
|                 | Bolting  | Seventy days after sprouting. In this time, a small bud was first observed at the tips of the stem                                                                                                          |
|                 | Florescence | The date of opium poppy first bloom                                                                                                                                                                       |
|                 | Filling stag | Fourteen days after falling flowers                                                                                                                                                                        |
| Tissues         | Root    | The root in underground 5–8 cm part                                                                                                                                                                        |
|                 | Stem    | The middle position of the stem with a length of 3 cm                                                                                                                                                    |
|                 | Leaf    | The top leaf of opium poppy                                                                                                                                                                               |

**Major BIAs content measurement.** To analysis the major BIAs content, some 0.5 g dry leaves in each sample of different germplasm lines at filling stag were employed according to our previous publicized method. For details, the sample was sequentially extracted three times in methanol, with the help of ultrasonication, for 30 min at room temperature. Then, methanol extracts were combined and concentrated under reduced pressure conditions to a volume of 2 mL. At last, 10 μL concentrated extracts was subject to analysis using HPLC equipped with a reversed phase C18 column (XDB-C18, 5 mm; Agilent, USA) according to our publicized method.
The BIAs content was shown as weight (mg) in one kilogram dry leaves (mg/kg) and listed in Table 4, and its raw data of individual measurements are available at Figshare²⁴.

Functional annotation. Functional annotations of the novel genes were performed using BLAST searching against public databases such as Swiss-Prot, GO (Gene Ontology), and KEGG (Kyoto Encyclopedia of Genes and Genomics)²⁵,²⁶.

### Table 4. The major BIAs contents in one Kilogram (mg/kg) dry leaves in different germplasm lines at filling stage. — Is represented as the content of BIAs could not be detected. All data are represented as mean ± SD from three independent plants (n = 3). The unit is mg/kg dry leaves.

| BIAs name | B1 | B2 | G | T | Z |
|-----------|----|----|---|---|---|
| Morphine  | 599.9 ± 139.77 | 273.8 ± 17.68 | 700.0 ± 193.71 | 92.4 ± 14.64 | 807.8 ± 18.93 |
| Codeine   | 44.7 ± 5.85 | 50.5 ± 34.67 | 266.7 ± 97.39 | — | 197.8 ± 79.43 |
| Norcoclaurine | 46.2 ± 13.25 | 14.9 ± 7.75 | 101.6 ± 55.20 | 55.1 ± 9.56 | 153.8 ± 89.71 |
| Thebaine  | — | — | — | 16009.2 ± 605.32 | — |
| Scoulerine | 632.3 ± 116.43 | 629.0 ± 183.91 | 408.0 ± 5.54 | 107.8 ± 4.53 | 866.4 ± 222.46 |
| Noscapine | 58.3 ± 13.80 | — | 2021.9 ± 480.39 | — | — |
| Papaverine | 175.2 ± 44.88 | 118.9 ± 59.13 | — | 99.9 ± 12.25 | — |
| Canadine  | 1150.9 ± 257.82 | 957.9 ± 203.02 | 1463.3 ± 365.45 | 1641.6 ± 372.12 | 1021.3 ± 164.90 |
| Sanguinarine | 507.8 ± 23.82 | 417.4 ± 17.56 | 405.8 ± 9.41 | — | — |

### Table 5. Metadata and description of each of the 33 samples that were sequenced.
Data Records

The Illumina HiSeq X Ten data (different growth phases, cultivars and tissues) and PacBio Sequel sequencing data have been submitted to the Sequence Read Archive (SRA) of NCBI under accession numbers SRP17355127, SRP17356528, SRP17354629, and SRP17372830, respectively (Table 5). The functional annotation and gene expression (RPKM) information of high-quality transcripts and unigenes are deposited in Figshare and Gene Expression Omnibus (GEO) in NCBI24,31. The differential gene expression data among different samples was also deposited in Figshare24.

Technical Validation

RNA quality control. The integrity of RNA sample was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) and agarose gel electrophoresis. The purity and concentration of RNA samples were determined with the Nanodrop microspectrophotometer (Thermo Fisher Scientific, USA). For Illumina sequencing, 33 high-quality RNA samples (OD260/280 = 1.7~2.3, OD260/230 ≥ 2.0, RIN ≥ 7) were used to construct the sequencing library. The OD260/280, OD260/230, and RIN values for all RNA samples are listed in Table 6.

Quality evaluation of raw data. A high quality region finder was used to identify the longest region of a singly-loaded enzyme using a signal-to-noise ratio of 0.8 to filter out low-quality areas. In order to improve the accuracy of sequencing, the same polymerase reads were read multiple times in a closed loop, and the random error correction was then performed on the sequence read multiple times by the same insert fragment. This produced 660,418 CCS reads with a mean of 12.62 passes per read (Fig. 2a,b). FL reads were classified based on the location of and relationships between the 5′ primer, 3′ primer and polyA tail (Fig. 2c,d).

Assessment of sample composition. For Illumina paired-end RNA-seq data, we measured the correlation coefficient and quantitative saturation of gene expression among 33 samples (biological repetition and biological variation, Fig. 3). Correlation of expression levels correlation among samples is an important index to

| Sample | 260/280 | 260/230 | RIN | 28 s/18 s |
|--------|---------|---------|-----|----------|
| 131    | 2.20    | 2.45    | 7.8 | 1.5      |
| 132    | 2.20    | 2.45    | 7.9 | 1.5      |
| 133    | 2.16    | 2.40    | 8.0 | 1.5      |
| 231    | 2.18    | 2.44    | 8.7 | 1.2      |
| 231    | 2.19    | 2.42    | 8.7 | 1.4      |
| 233    | 2.17    | 2.40    | 9.4 | 1.5      |
| 331    | 2.18    | 2.41    | 7.8 | 1.4      |
| 332    | 2.18    | 2.39    | 8.4 | 1.5      |
| 333    | 2.19    | 1.80    | 7.3 | 1.3      |
| 431    | 2.18    | 2.22    | 7.5 | 1.4      |
| 432    | 2.19    | 2.19    | 6.9 | 1.4      |
| 433    | 2.18    | 2.02    | 7.2 | 1.5      |
| 511    | 2.16    | 2.41    | 10  | 2.0      |
| 512    | 2.14    | 2.13    | 10  | 2.1      |
| 513    | 2.12    | 2.08    | 10  | 2.4      |
| 521    | 2.13    | 2.26    | 9.9 | 1.9      |
| 522    | 2.12    | 1.61    | 10  | 3.0      |
| 523    | 2.12    | 2.35    | 9.8 | 2.0      |
| 531    | 2.18    | 2.23    | 8.4 | 1.7      |
| 532    | 2.19    | 2.17    | 8   | 1.6      |
| 533    | 2.21    | 2.12    | 7.3 | 1.4      |
| B2-31  | 2.18    | 2.45    | 7.2 | 1.4      |
| B2-32  | 2.17    | 2.33    | 7.7 | 1.5      |
| B2-33  | 2.18    | 2.35    | 7   | 1.3      |
| T31    | 2.19    | 2.48    | 7.9 | 1.6      |
| T32    | 2.18    | 2.46    | 7.4 | 1.5      |
| T33    | 2.19    | 2.44    | 8   | 1.5      |
| Z31    | 2.19    | 2.40    | 7.9 | 1.5      |
| Z32    | 2.18    | 1.79    | 7.5 | 1.6      |
| Z33    | 2.19    | 2.47    | 7.4 | 1.5      |
| G31    | 2.17    | 2.40    | 8   | 1.7      |
| G32    | 2.14    | 2.35    | 8.1 | 1.8      |
| G33    | 2.14    | 2.45    | 7.9 | 1.7      |

Table 6. RNA sample quality used in this study.
test the reliability of experiments and the rationality of sample selection. Correlation coefficients close to 1 (red) indicate that the samples have a high similarity of expression patterns. If there is biological repetition in the sample, the correlation coefficient of biological repetition is usually higher. By calculating the relationship between the number of different full-length transcripts and the number of genes, we can observe whether the total number of genes has been measured up to the saturation level (Fig. 3b).

**Code Availability**

SMRT Link 5.1 pipeline: [http://www.pacb.com/products-and-services/analytical-software/smrt-analysis/](http://www.pacb.com/products-and-services/analytical-software/smrt-analysis/). CD-HIT: [http://www.bioinformatics.org/cd-hit/](http://www.bioinformatics.org/cd-hit/) (version 4.6.6). Blast: ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/ (version 2.2.31).

**References**

1. Guo, L. et al. The opium poppy genome and morphinan production. Science 362, 343–347 (2018).
2. Dang, T. T., Onoyovwi, A., Farrow, S. C. & Facchini, P. J. Biochemical genomics for gene discovery in benzylisoquinoline alkaloid biosynthesis in opium poppy and related species. Method Enzymol 515, 231 (2012).
3. Beaudoin, G. A. & Facchini, P. Benzylisoquinoline alkaloid biosynthesis in opium poppy. Planta 240, 19–32 (2014).
4. Suya, M. J., Gelders, S. F., Achara, O. U., Milani, B. & Scholten, W. K. A first comparison between the consumption of and the need for opioid analgesics at country, regional, and global levels. J Pain Palliat Care Pharmacother 25, 6–18 (2013).
5. Galanie, S., Thokey, K., Trenchard, L. J., Interrante, M. F. & Smolke, C. D. Complete biosynthesis of opioids in yeast. Science 349, 1095–1100 (2015).
6. Thokey, K., Galanie, S. & Smolke, C. D. A microbial biomanufacturing platform for natural and semi-synthetic opiates. Nat Chem Biol 10, 837 (2014).
7. Li, Y. & Smolke, C. D. Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. Nat Commun 7, 12137 (2016).
8. Winzer, T. et al. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. Science 349, 309–312 (2015).
9. Dasmalchik, M., Chang, L., Torres, M. A., Ng, K. & Facchini, P. J. Codeinone reductase isoforms with differential stability, efficiency and product selectivity in opium poppy. Plant J 95, 631–647 (2018).
10. Winzer, T. et al. A Papyrus somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science 336, 1704–1708 (2012).
11. Dang, T. T., Chen, X. & Facchini, P. J. Acetylation serves as a protective group in noscapine biosynthesis in opium poppy. Nat Chem Biol 11, 104–106 (2015).
12. Suttinant, N. et al. The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in Catharanthus roseus. Plant Physiol 157, 2081–2093 (2011).
13. Schultenhofer, C. & Yuan, L. Regulation of specialized metabolism by WRKY transcription factors. Plant Physiol 167, 295–306 (2015).
14. Abdel-Ghany, S. E. et al. A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun 7, 11706 (2016).
15. Wang, B. et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nat Commun 7, 11708 (2016).
16. Cole, T. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511–515 (2010).
17. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29, 644–652 (2011).
18. John, E. et al. Real-time DNA sequencing from single polymerase molecules. Science 323, 133–138 (2009).
19. Leena, S. & Eric, R. LoRD:EC: accurate and efficient long read error correction. Bioinformatics 30, 3506 (2014).
20. Li, W. Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. Bioinformatics 22, 1658 (2006).
21. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: fast and memory-efficient alignment of shallow and single-molecule long-read sequencing data to reference genomes. Nat Methods 13, 357–360 (2015).
22. Love, M. I., Wolfgang, H. & Simon, A. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 590 (2014).
23. Zhan, Y.-C. et al. Integration of a Decrescent Transcriptome and Metabolomics Dataset of Peucedanum praeruptorum to Investigate the CYP450 and MDR Genes Involved in Courmarins Biosynthesis and Transport. Front Plant Sci 6, 696 (2015).
24. Zhao, Y.-C. Transcriptomic profiles of 33 opium poppy samples in different tissues, growth phases, and cultivars. figshare, https://doi.org/10.6084/m9.figshare.c.4433138.v2 (2019).
25. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 1, 25 (2000).
26. Miomor, K., Susumnu, G., Shiuchi, K., Yasushi, O. & Masahiro, H. The KEGG resource for deciphering the genome. Nucleic Acids Res 32, D277 (2004).
27. NCBIR Sequence Read Archive, http://identifiers.org/ncbi/insdc.sra/SRP173551 (2019).
28. NCBIR Sequence Read Archive, http://identifiers.org/ncbi/insdc.sra/SRP173565 (2019).
29. NCBIR Sequence Read Archive, http://identifiers.org/ncbi/insdc.sra/SRP173546 (2019).
30. NCBIR Sequence Read Archive, http://identifiers.org/ncbi/insdc.sra/SRP173728 (2019).
31. NCBIR Gene Expression Omnibus, http://identifiers.org/geo/GSE129568 (2019).

**Acknowledgements**

This Project was supported by China Postdoctoral Science Foundation (2016M601922, 2018T110577), the National Natural Science Foundation of China (81703637) and Natural Science Fund in Jiangsu Province (BK20170736). This research was also supported by the Program for Changjiang Scholars and Innovative Research Team in University (IRT_15R63), 111 Project from Ministry of Education of China and the State Administration of Foreign Export Affairs of China (B18056).

**Author Contributions**

Y.Z. and L.K. conceived the project and designed the experiment. Z.Z., M.L., J.L., F.C. and Y.G. performed the majority of the experiment work. Y.L. and Y.W. performed the data analyses. Y.Z., Y.S. and L.K. wrote the manuscript. All authors read the final manuscript.

**Additional Information**

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
