InTransBo: An Integrative Transcript Library to Enable Genome-Free Systematic Exploration of *Bougainvillea*

Qi Luo
The R&D Group of the General Engineer Office, The Road Development Center of Xiamen, Xiamen, 361000
Ziliang Chen
The Wanyin Environmental Technology Limited Company, Xiamen, 361000
Tingting Xu
School of Life Sciences, Xiamen University, Xiamen, 361102
Dangzheng Huang
The R&D Group of the General Engineer Office, The Road Development Center of Xiamen, Xiamen, 361000
Haitao Hou
The R&D Group of the General Engineer Office, The Road Development Center of Xiamen, Xiamen, 361000
Chenjie Hong
The R&D Group of the General Engineer Office, The Road Development Center of Xiamen, Xiamen, 361000
Fulin Zhan
The Wanyin Environmental Technology Limited Company, Xiamen, 361000
Hangqi Guo
The Wanyin Environmental Technology Limited Company, Xiamen, 361000
Zhe Lin
School of Life Sciences, Xiamen University, Xiamen, 361102
Xiaoling Guo
College of the Environment & Ecology, Xiamen University, Xiamen, 361102
Liang Chen
School of Life Sciences, Xiamen University, Xiamen, 361102
Zhi-Liang Ji (✉ appo@xmu.edu.cn)
School of Life Sciences, Xiamen University, Xiamen, 361102

Research Article

Keywords: Bougainvillea, InTransBo, RNA library, Transcriptome

Posted Date: December 31st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-134688/v1

Abstract

Members of the genus *Bougainvillea* are rich sources of natural dyes, pigments, and traditional medicines. They are also commonly used as ornamentals in roadside landscape construction. However, the horticultural development of *Bougainvillea* flowers with extended growth periods and coloration is not always feasible. One reason is limited molecular knowledge and no genomic information for *Bougainvillea*. Here, we compiled an expressed transcript sequence library for *Bougainvillea* by integrating 20 Illumina-sequencing RNA transcriptomes. The library consisted of 97,623 distinct transcripts. Of these, 47,006 were protein-coding, 31,109 were lncRNA, and 19,508 were unannotated. We also confirmed that the library is an alternative genomic reference for accurate transcriptome assembly and its performance was substantially better than that of the *de novo* method. We also curated the Integrative Transcript Library database for *Bougainvillea* known as InTransBo (http://www.bio-add.org/InTransBo/index.jsp). To the best of our knowledge, the present study is the first large scale genomic resource for *Bougainvillea*. Overall, the library helps fill the genomic gap and elucidate the transcriptional nature of *Bougainvillea*. It may also advance progress in the precise regulation of flowering in horticulture. The same strategy can be readily applied toward the systematic exploration of other plant species lacking complete genomic information.

Introduction

Bougainvillea sp. are native to the Amazonian rainforests of South America but are globally distributed. They are used mainly as ornamental and landscaping plants (Tsai, Su, Liao, & Hung, 2005). *Bougainvillea* sp. belongs to the family Nyctaginaceae and comprises ~ 18 species (*B. berberidifolia, B. buttiana, B. campanulata, B. glabra, B. herzogiana, B. infesta, B. lehmanniana, B. lehmannii, B. malmeana, B. modesta, B. pachyphylla, B. peruviana, B. pomacea, B. precox, B. spectabilis, B. spinosa, B. stipitata*, and *B. trollii*) (Abarca-Vargas & Petricevich, 2018). To date, over 100 cultivars and three major hybrids have been recognized (Abarca-Vargas & Petricevich, 2018; Tsai et al., 2005). However, only four species (*B. buttiana, B. glabra, B. spectabilis*, and *B. peruviana*) have been commercially exploited (Choudhary, Kapoor, & Lodha, 2008).

Bougainvillea sp. have been extensively investigated since 1970 as potential sources of traditional medicine (Hernandez-Martinez, Estevez, Vargas, Quintanilla, & Rodriguez, 2011; Heuer et al., 1994). Aqueous *Bougainvillea* extracts and decoctions are used for fertility control by the tribal people of several countries (Ghogar & Jiraungkoorskul, 2017). *Bougainvillea* may also have anticancer, antidiabetic, antihepatotoxic, anti-inflammatory, antihyperlipidemic, antimicrobial, antioxidant, and anti-ulcer properties (Arteaga Figueroa, Barbosa Navarro, Patino Vera, & Petricevich, 2015). The alkaloids, essential oils, flavonoids, glycosides, oxalates, phenolics, phlobotannins, quinones, saponins, tannins, and terpenoids in *Bougainvillea* sp. might account for their putative medicinal properties (Saleem et al., 2019). Bougainvinones, pinitol, quercetagetin, quercetin, and terpinolene may also contribute to the therapeutic efficacy of *Bougainvillea* (Ghogar & Jiraungkoorskul, 2017; Hernandez-Martinez et al., 2011).

Analyses of the metabolites (Sangthong, Suksabye, & Thiravetyan, 2016), natural dyes and pigments (Sangthong et al., 2016), medicinal uses, and species diversity of *Bougainvillea* have been conducted. In contrast, there have been few molecular studies of this genus. No genome of *Bougainvillea* has ever been sequenced up-to-date; and it won't be done in the coming few years due to technical and economic difficulties. Limited omics research has been performed to elucidate the molecular basis of the aforementioned properties of *Bougainvillea* especially at the systematic level. As no genome has been clarified for *Bougainvillea*, current molecular research on this plant is often compared against or referred to *Arabidopsis thaliana*. Consequently, its gene behavior is uncertain, ambiguous, or even misunderstood as there are genomic gaps between organisms. Therefore, an alternative genomic resource is required that can complement or fill the no-genome gap. Here, our objectives were to use multiple Illumina RNA sequencing (RNA-seq) transcriptomes determined for various *Bougainvillea* tissues and generate a sequence library consisting of all expressed transcripts. This library could serve as an alternative genomic reference for the molecular exploration of *Bougainvillea* and will be presented as an online interactive database.

Materials And Methods

Sample Collection

Tissue samples of pot-grown *Bougainvillea glabra* L. Choisy Tonganhong) in its normal flowering period were collected in October 2018 at Xiamen City, P.R. China. Samples comprised thorns, flower thorns, small buds, bracteoles, leaf sprouts, flower sprouts, lobules, stems under buds, stems under bracteoles, and flowers. Two biological replicates per tissue type were used (Table 1). The tissues were excised either from different parts of the same plant or from different plants, washed with distilled water, and briefly air-dried in a clean environment. The tissues were mixed, randomly divided into two replicates, packed in silver paper, frozen in liquid nitrogen, and stored in the School of Life and Science, Xiamen University, Chen's lab.

Table 1
Illumina RNA-sequencing datasets used in this study.

Tissue	Raw reads	Clean reads	Raw base (G)	Clean base (G)	Q20 (%)	Q30 (%)	GC content (%)
thorn	33,613,973	31,358,904	10.08	9.41	97.68	93.60	43.13
thorn replicate 1	28,729,693	27,644,873	8.62	8.29	97.47	93.09	43.16
flower thorn	26,343,968	25,698,944	7.90	7.71	96.26	90.50	42.95
flower thorn replicate 1	34,229,176	32,719,720	10.27	9.82	97.72	93.73	42.91
small bud	31,318,836	23,553,700	9.39	7,07	97.47	93.24	42.84
small bud replicate 1	29,759,552	21,379,446	9.93	8.63	93.56	85.86	42.55
bracteole	31,247,574	30,110,664	9.37	9.03	97.78	93.90	42.92
bracteole replicate 1	32,190,609	30,967,796	9.66	9.29	97.66	93.62	42.38
leaf sprout	30,698,035	30,294,725	9.21	9.09	97.71	93.72	42.41
leaf sprout replicate 1	29,239,014	28,026,332	8.77	8.41	97.66	93.62	42.80
flower sprout	32,856,663	31,772,572	9.86	9.53	97.35	92.94	42.85
flower sprout replicate 1	33,433,464	32,960,179	10.03	9.89	97.70	93.65	42.87
lobule	30,645,057	30,328,864	9.19	9.10	96.39	90.81	42.80
lobular replicate 1	29,613,243	29,302,625	8.88	8.79	97.71	93.74	42.50
stem under bud	30,698,035	30,294,725	9.21	9.09	97.71	93.72	42.41
stem under bud replicate 1	33,460,411	31,701,317	10.04	9.51	98.04	94.24	42.99
stem under bracteole	30,982,149	29,649,261	9.29	8.89	96.03	90.09	42.89
stem under bracteole replicate 1	33,115,209	31,568,381	9.93	9.47	96.42	90.82	42.69
flower	32,141,549	30,546,299	9.64	9.16	97.59	93.45	42.90
flower replicate 1	31,700,976	30,410,918	9.51	9.12	97.36	92.93	42.42

RNA Library Construction and Deep Sequencing

The sample mixtures were lysed with 1 mL TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Total RNA was prepared according to the manufacturer's instructions. RNA purity was evaluated with a NanoPhotometer® spectrophotometer (Implen USA, Westlake Village, CA, USA). The RNA concentrations

were measured with a Qubit® RNA assay kit and a Qubit® 2.0 fluorometer (Life Technologies, Waltham, MA, USA). RNA integrity was evaluated with the RNA Nano 6000 assay kit in an Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). RNA samples passing the quality control test were stored at -20 °C until later use.

The RNA library was constructed by Novogene Co. Ltd. (Beijing, China) following the standard operation procedure. Three micrograms RNA per sample was used as the input material for library preparation. Polyadenylated (poly(A)) RNA was purified from total RNA with poly T oligo-attached magnetic beads. RNA sequencing was performed in an Illumina HiSeq 4000 system (Illumina, San Diego, CA, USA) using the 125-bp, strand-specific, paired-end mode.

RNA-Seq Data Preprocessing

Before proceeding to the transcriptome assembly, the RNA-seq raw data were filtered with Trimmomatic (Bolger, Lohse, & Usadel, 2014) to exclude low-quality reads and those with adaptors or N content > 10%. The proportion of bases with sQ ≤ 5 occurred in > 50% of all reads.

Construction of the Integrative Transcript Library

An integrative transcript library was constructed with TransIntegrator (Qin et al., 2020). TransIntegrator required four continuous steps for transcript library construction. Here, the computational workflow was specified as follows: (1) 20 RNA-seq datasets were *de novo* assembled separately with Trinity software (version 1.0) by contig length > 200 bp, K-mer > 5, and max reads of per graph > 200 bp (Haas et al., 2013). (2) All expressed transcripts were clustered with CD-HIT-EST (version 4.5.4) (Fu, Niu, Zhu, Wu, & Li, 2012) using a sequence identity threshold -c > 90% and an alignment coverage -aS > 80%. The clustered sequences were identified, the longest sequences were selected as representatives, and the shorter sequences were discarded. (3) The representative transcripts were bridged with CAP3 (version 12/21/07, default parameters) to form longer sequences (Huang & Madan, 1999) when two sequences had > 40 bp overlap (> 90% sequence identity). (4) The library was refined by removing sequences < 300 bp, and the mature *Bougainvillea* transcript library was obtained.

A conventional procedure was followed to annotate the integrative transcript library. The coding transcripts were identified with TransDecoder v3.0.0 (parameter: default) (Haas et al., 2013) by referring to the SwissProt and Pfam databases. The rRNA was annotated with RNAmmer (version 1.0.0, default parameters) (Lagesen et al., 2007). The tRNA was annotated with tRNAscan-SE (version 2.1.3, default parameters) (Lowe & Eddy, 1997). The ncRNAs and miRNA precursors were identified by using the BLAST tool (version 1.3.1, e-value:1e-5, identify:90%) against the NONCODE database (Xie et al., 2014) and miRBase (Kozomara & Griffiths-Jones, 2013). All annotations were then integrated and those with low confidence (subject to the default threshold of corresponding annotation tools) were discarded.

Library-Based Transcriptome Assembly and Performance Comparison

To make a genomic reference for the transcriptome assembly, the library was annotated and preformatted as a gff file with Annoscript (version 1.1.3, default parameters) (Musacchia, Basu, Petrosino, Salvemini, & Sanges, 2015). For library-based transcriptome assembly, we used HISAT software (version 1.3.1, min-intronlen:200 bp, pen-noncansplice:20, no-spliced-alignment) and quantified by HTSeq (version 3.2, default parameters) (Anders, Pyl, & Huber, 2015). For *de novo* transcriptome assembly, we used Trinity software V1.0 as described above.

The transcriptome assemblies of the library-based and *de novo* methods were compared by using two external *Bougainvillea* RNA-seq datasets derived from public resources (SRA: SRR10097433 and SRR2912684). The comparison implemented several statistical parameters such as the reads mapping ratio, N_{50}, and the unique gene (unigene) number (Moreton, Izquierdo, & Emes, 2015). Of them, N_{50} is a common indicator to evaluate genome assembly quality, the larger the better. In this study, N_{50} was used to indicate general transcript assembly quality. A large N_{50} suggested more long transcripts were assembled. The completeness and fragmentation ratio of the transcripts were evaluated with BUSCO (version 2.1.3, default parameters) (Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). The potential bridge and assembly score were determined with TransRate (version 1.3, default parameters) (Smith-Unna, Boursnell, Patro, Hibberd, & Kelly, 2016).

Database Construction

For user convenience, the transcript library was presented as an online interactive database called InTransBo, which was constructed on the Linux-Apache-JSP platform. MySQL software was used to manage data storage, access, and maintenance. Efficient and friendly user interfaces were designed with JavaScript for interactive transcript search and retrieval.

Results And Discussion

Construction of the Integrative Transcript Library for the*Bougainvillea*

After quality control, the 20 RNA-seq datasets generated ~ 590.30 million clean reads and 172.23 gigabase pairs (Gbp). Detailed information on these datasets was summarized in Table 1. Integration of these datasets yielded a *Bougainvillea* transcript library comprising 97,623 non-redundant sequences. Transcript integrity was supported by multiple mapping reads.

The average library transcript length was 724.11 bp. The average library GC content was 40.12. About 77.03% of the transcripts were in the range of 300–885 bp long which agreed with a previous plant gene length estimation of 183–2,000 bp (Cheng et al., 2017). Hence, most of the library transcripts were correctly assembled. Of the 97,623 sequences, ~ 48.15% (47,006) were annotated as protein-coding transcripts, ~ 31.87% (31,109) were annotated as potential long non-coding RNAs (lncRNAs) and other non-coding RNAs (ncRNAs), and the remaining 19.98% (19,508) were not annotated (Fig. 1a).

We performed a gene conservation analysis by mapping the expressed *Bougainvillea* transcripts against the *Arabidopsis thaliana* genome. We found that 44,344 transcripts had homologs in the *Arabidopsis thaliana* genome, 24,870 transcripts mapped to intergenic regions, and 28,409 were unmapped (Fig. 1). The latter were either annotated *Bougainvillea*-specific or wrongly assembled transcripts. We selected ten conserved plant genes to verify sequence completeness. To this end, we compared them with their homologs in *Arabidopsis thaliana* (Fig. 1b**)**. Out of the ten *Bougainvillea* genes, seven had sequence identity and coverage > 90% of their corresponding homologs in *Arabidopsis thaliana.* The remaining three genes failed either criterion. Thus, we confirmed the sequence completeness of the transcript library.

Integrative Transcript Library Database for the Bougainvillea

The InTransBo database is freely accessible at http://www.bio-add.org/InTransBo/ or at its mirror site http://bioinf.xmu.edu.cn/InTransBo/. InTransBo uses keyword search and sequence BLAST to retrieve data interactively. The keyword search function allows both accurate and fuzzy transcript search via the input of complete or partial gene symbols, gene names, protein names, and abbreviated protein names (Fig. 2a). The search engine feeds back the hit terms in alphabet order along with the gene symbols, protein names, transcript lengths, and transcript ID. Clicking on the transcript ID redirects to the transcript information page containing various data listed in order including transcript annotations, sequence information, and transcript expression profiles (Fig. 3).

For newly identified sequences and unnamed sequences, InTransBo enables data access via an alternative BLAST method. The database supports BLASTn and tBLASTn to identify nucleic acid and protein sequences, respectively. The input may either be the typing sequence in text form or an uploaded file in FASTA format (Fig. 2b). The embedded BLAST engine responds to all hit sequences meeting the default expectation threshold of E value = 1e-10 and sorted by hit score (Fig. 2b). For each hit, alignment details may be obtained using the “Alignment” hyperlink. Detailed information of the hit transcript may be acquired via the transcript ID hyperlink. The transcript library is free for downloading; however, user registration is required. The library is annotated and formatted as a gff file such that it may be used as an alternative reference to the genome for transcriptome assembly and other molecular applications.

Library-based Transcriptome Assembly

The library was used to demonstrate the reference-based assembly of 20 RNA-seq transcriptomes according to a typical genome-based method. The gene expression profiles in different tissues are accessed in the InTransBo database by specifying gene symbols. We compared transcriptome assembly performance between the library-based and *de novo* Trinity methods. The comparison was tested on two external *Bougainvillea* RNA-seq datasets (Table 1) determined by different research groups under various experimental conditions and using diverse sequencing qualities. Assembly performance was evaluated based on the read mapping ratio, number of unique genes and bridges, sequence completeness, fragmentation ratio, and estimated assembly score. The results are illustrated in Fig. 4. The library-based method substantially outperformed the *de novo* Trinity method. The former had superior N50, lower

fragmentation ratio, and higher assembly score. The Trinity-assembled transcriptomes had relatively higher read mapping ratios and unigene numbers. However, the Trinity method also considered numerous fragmental sequences that could not be correctly assembled.

Bougainvillea sp. require systematic transcriptome research but lack genome support. Whole-genome sequencing is not a viable option to-date as it is costly and technically impracticable. Here, we used the computational method TransIntegrator to construct an integrative transcript library for *Bougainvillea* based on 20 heterogeneous RNA-seq datasets. We curated the InTransBo database on the transcript library for interactive data retrieval. By using this library, we demonstrated reference-based transcriptome assembly for the 20 RNA-seq transcriptomes. To the best of our knowledge, this study is the first to investigate *Bougainvillea* at the molecular level with no genome, and the InTransBo database could be the first genomic sequence source for *Bougainvillea*. A subsequent analysis revealed that the library-based method outperformed the *de novo* Trinity method in terms of transcriptome assembly. Therefore, the library may serve as a reliable alternative reference that replaces the genome and enables the molecular exploration of *Bougainvillea*. The present study showed that the InTransBo database mitigates the genome constraint on the systematic investigation of *Bougainvillea*. We believe that it also helps elucidate the molecular biology of *Bougainvillea* and facilitates precise flower regulation in horticultural practices. The same strategy could be readily applied toward the systematic exploration of other plant species lacking adequate genomic data.

Declarations

ACKNOWLEDGMENTS

Financial support from the Transportation Technology Project of Fujian Province (No. 2018Y002) are gratefully acknowledged.

AVAILABILITY OF DATA AND MATERIALS

The Illumina sequencing datasets determined in this study have been deposited in the BioProject of the BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under the accession number PRJCA003739. The data are publicly accessible at http://bigd.big.ac.cn/gsa.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Q.L., Z.C., and T.X. wrote the manuscript text, demonstrated majority of the bioinformatics analysis and prepared most of the figures and tables. D.H., H.H., C.H., F.Z., H.G., Z.L., and X.G. supplied *Bougainvillea* RNA for sequencing and experimental verification. L.C. reviewed the manuscript. Z.J. presided over the

whole project, designed the experiments, reviewed and revised the manuscript. All co-authors have reviewed the manuscript.

FUNDING

This work was supported by a grant from the Transportation Technology Project of Fujian Province (No. 2018Y002).

References

1. Abarca-Vargas, R., and Petricevich, V. L. (2018). Bougainvillea genus: a review on phytochemistry, pharmacology, and toxicology. *Evid. Based Complement. Alternat. Med.* 2018, 9070927. doi:10.1155/2018/9070927.
2. Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi:10.1093/bioinformatics/btu638.
3. Arteaga Figueroa, L., Barbosa Navarro, L., Patiño Vera, M., and Petricevich, V. L. (2015). Preliminary studies of the immunomodulator effect of the *Bougainvillea xbuttiana* extract in a mouse model. *Evid. Based Complement. Alternat. Med.* 2015, 479412. doi:10.1155/2015/479412.
4. Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170.
5. Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89, 789–804. doi:10.1111/tpj.13415.
6. Choudhary, N., Kapoor, H. C., and Lodha, M. L. (2008). Cloning and expression of antiviral/ribosome-inactivating protein from *Bougainvillea xbuttiana*. *Journal of Biosciences* 33, 91–101. doi:10.1007/s12038-008-0025-8.
7. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi:10.1093/bioinformatics/bts565.
8. Ghogar, A., and Jiraungkoorskul, W. (2017). Antifertility effect of *Bougainvillea spectabilis* or paper flower. *Pharmacogn. Rev.* 11, 19–22. doi:10.4103/phrev.phrev_44_16.
9. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J. et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi:10.1038/nprot.2013.084.
10. Hernandez-Martinez, A. R., Estevez, M., Vargas, S., Quintanilla, F., and Rodriguez, R. (2011). New dye-sensitized solar cells obtained from extracted bracts of *Bougainvillea glabra* and *spectabilis* betalain pigments by different purification processes. *Int. J. Mol. Sci.* 12, 5565–5576. doi:10.3390/ijms12095565.
11. Heuer, S., Richter, S., Metzger, J. W., Wray, V., Nimtz, M., and Strack, D. (1994). Betacyanins from bracts of *Bougainvillea glabra*. *Phytochemistry* 37, 761–767. doi:10.1016/s0031-9422(00)90354-6.

12. Huang, X., and Madan, A. (1999). CAP3: a DNA sequence assembly program. *Genome Res.* 9, 868–877. doi:10.1101/gr.9.9.868.
13. Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi:10.1093/nar/gkt1181.
14. Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi:10.1093/nar/gkm160.
15. Lowe, T. M., and Eddy, S. R. (1997). TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9023104. doi:10.1093/nar/25.5.955.
16. Moreton, J., Izquierdo, A., and Emes, R. D. (2015). Assembly, assessment, and availability of de novo generated eukaryotic transcriptomes. *Front. Genet.* 6, 361. doi:10.3389/fgene.2015.00361.
17. Musacchia, F., Basu, S., Petrosino, G., Salvemini, M., and Sanges, R. (2015). Annocript: a flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. *Bioinformatics* 31, 2199–2201. doi:10.1093/bioinformatics/btv106.
18. Qin, Y., Lin, Z., Shi, D., Zhong, M., An, T., Chen, L. et al. (2020). Integrate heterogeneous NGS and TGS data to boost genome-free transcriptome research. *bioRxiv*, 2020.2005.2027.117796. doi:10.1101/2020.05.27.117796.
19. Saleem, H., Zengin, G., Ahmad, I., Lee, J. T. B., Htar, T. T., Mahomoodally, F. M. et al. (2019). Multidirectional insights into the biochemical and toxicological properties of Bougainvillea glabra (Choisy.) aerial parts: A functional approach for bioactive compounds. *J. Pharm. Biomed. Anal.* 170, 132–138. doi:10.1016/j.jpba.2019.03.027.
20. Sangthong, S., Suksabye, P., and Thiravetyan, P. (2016). Air-borne xylene degradation by Bougainvillea buttiana and the role of epiphytic bacteria in the degradation. *Ecotoxicol. Environ. Saf.* 126, 273–280. doi:10.1016/j.ecoenv.2015.12.017.
21. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351.
22. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., and Kelly, S. (2016). TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26, 1134–1144. doi:10.1101/gr.196469.115.
23. Tsai, C. H., Su, H. J., Liao, Y. C., and Hung, T. H. (2005). First report of *Bougainvillea spectabilis* chlorotic vein-banding virus infecting *Bougainvillea* plants in Taiwan. *Plant Dis.* 89, 1363. doi:10.1094/PD-89-1363C.
24. Xie, C., Yuan, J., Li, H., Li, M., Zhao, G., Bu, D. et al. (2014). NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.* 42, D98–D103. doi:10.1093/nar/gkt1222.

Figures

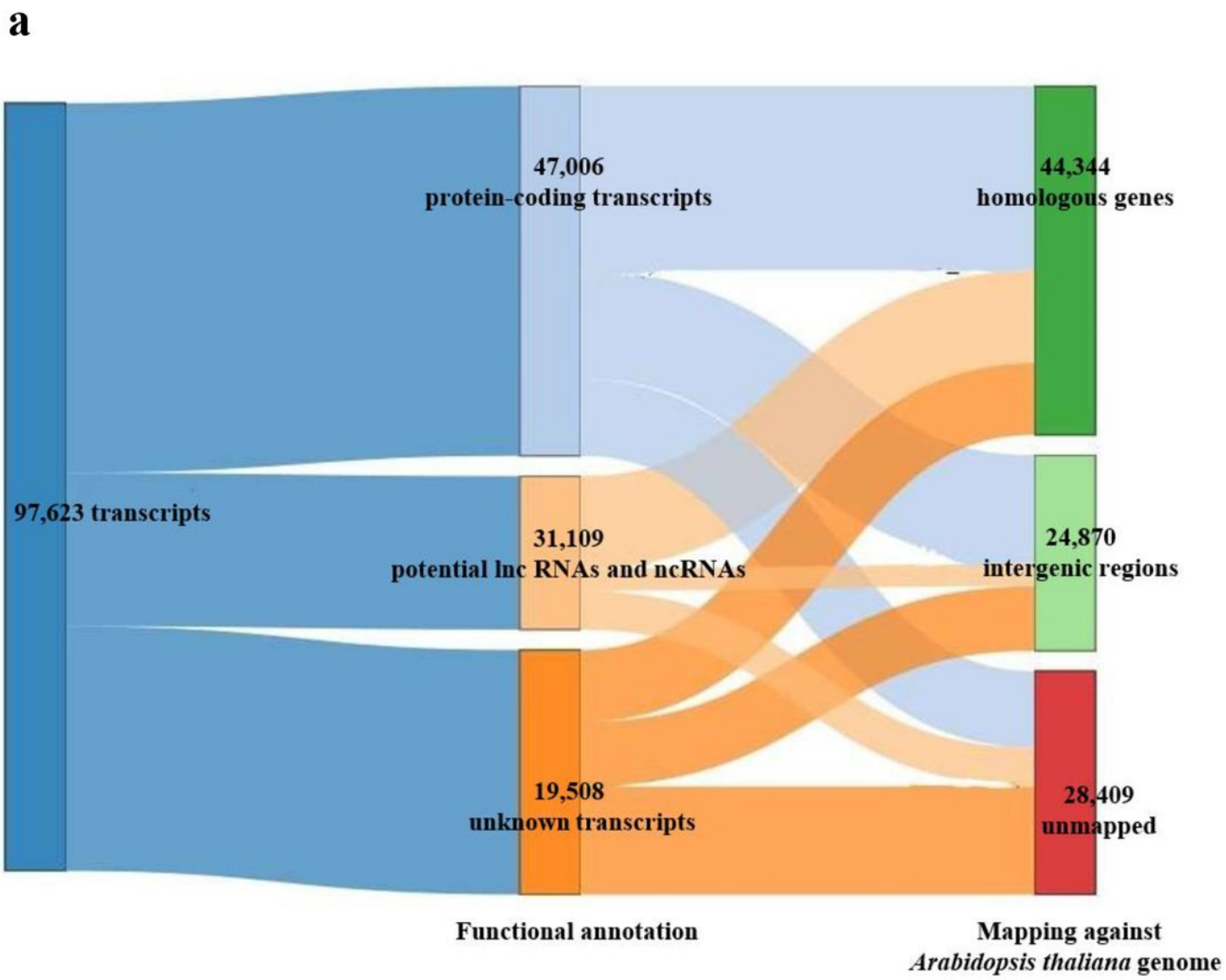

Figure 1

Overview of library transcripts. (a) Functional annotation of library transcripts. (b) Sequence completeness comparison.

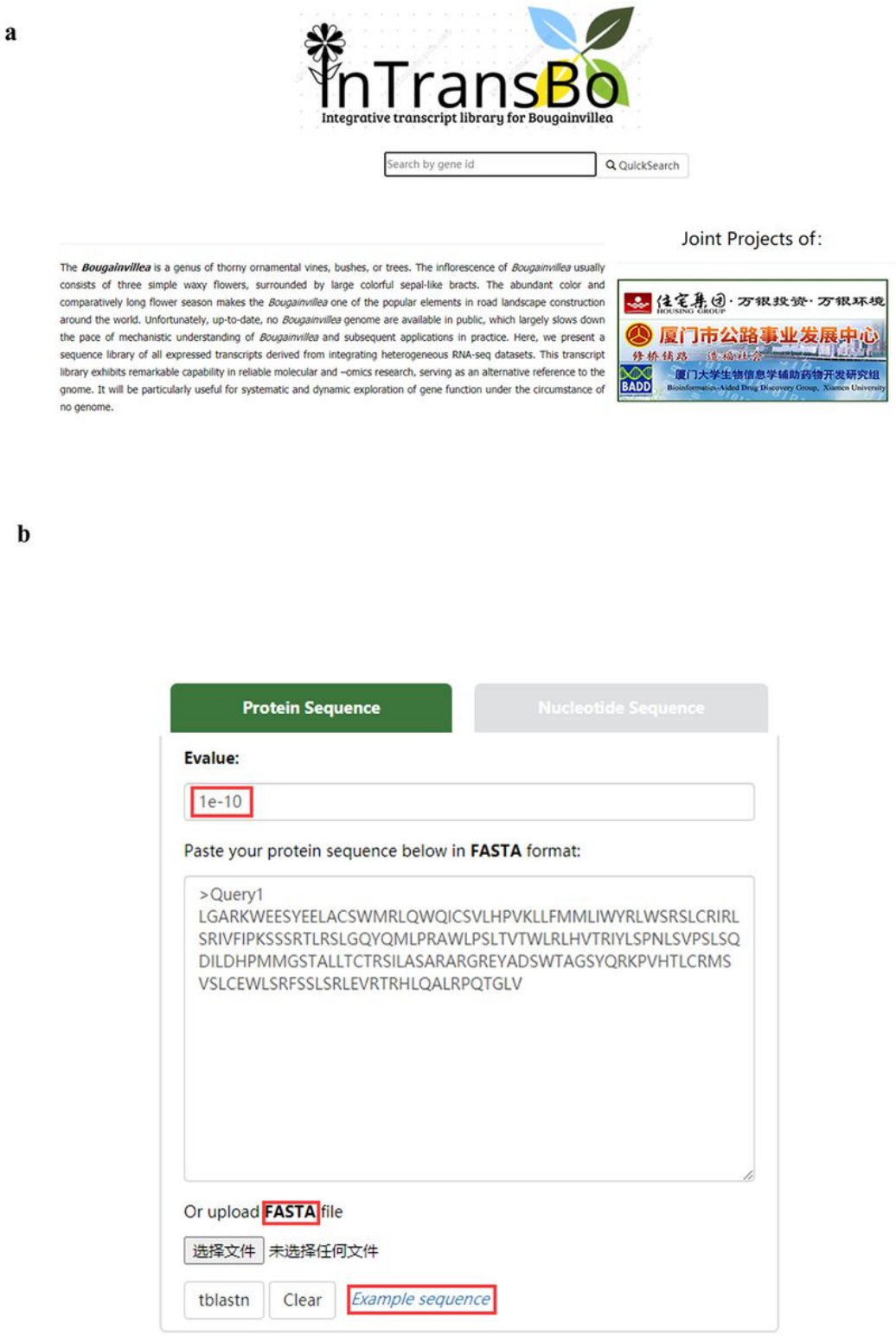

Figure 2

Snapshots of InTransBo user interfaces. (a) Homepage and keyword search form. (b) Blast search form.

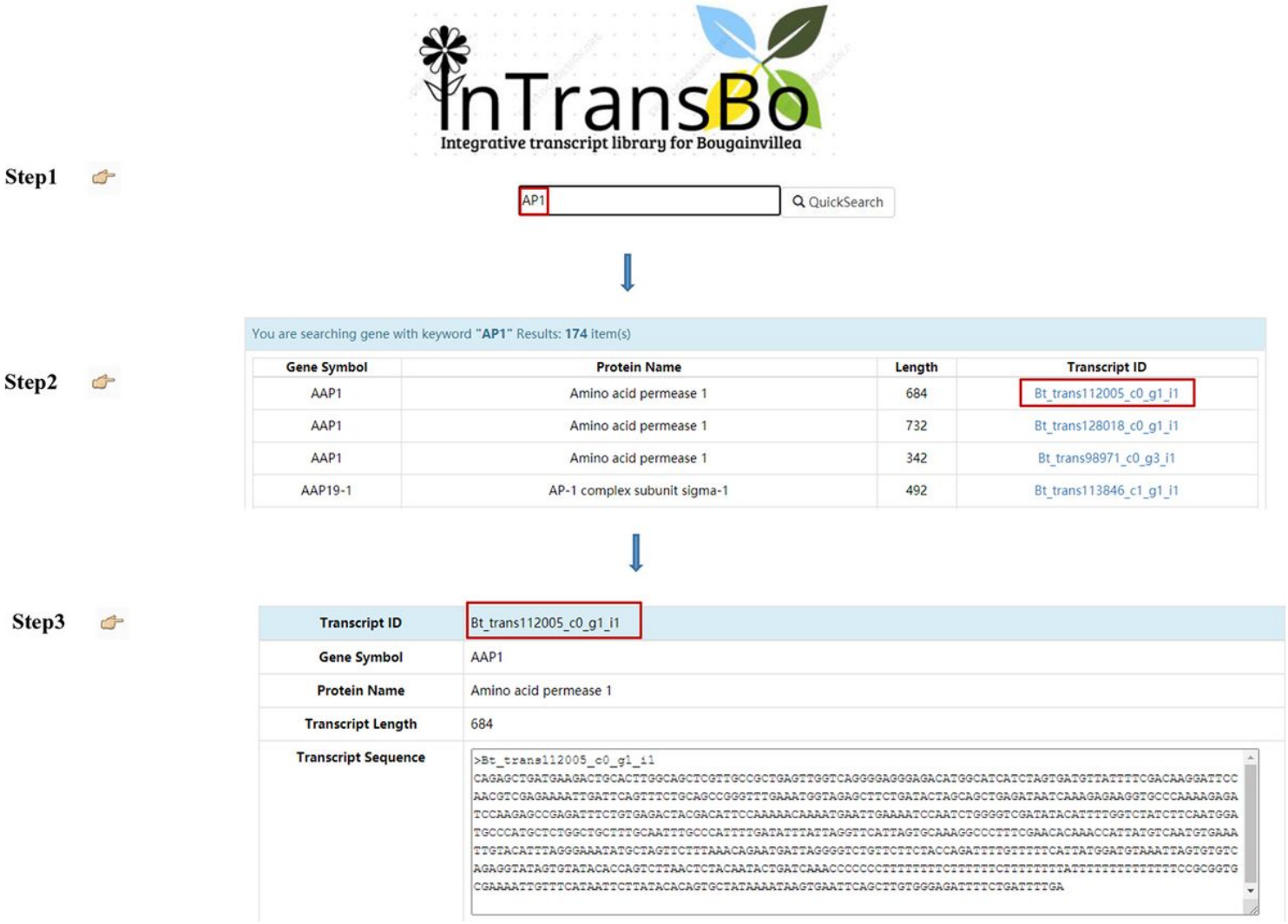

Figure 3

Data retrieval by keyword search method

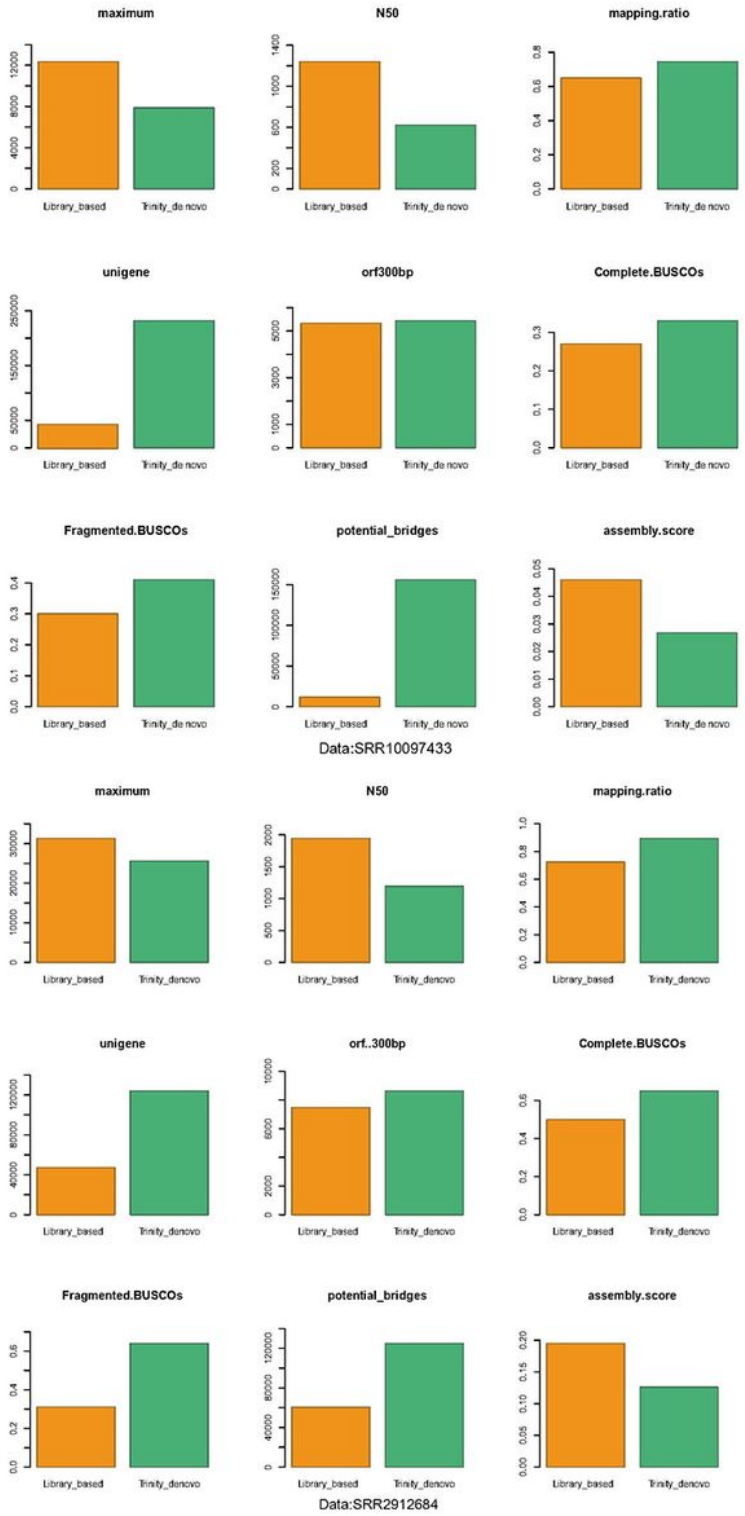

Figure 4

Comparison of transcriptome assembly performance between library-based and de novo Trinity methods