Article

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

Targeted transcriptomics of the Mexican prickly poppy (*Argemone mexicana* L.) reveals diverse proteins related to benzylisoquinoline alkaloid biosynthesis

José Germán Serrano-Gamboa[1,§], Jorge Froylán Xool-Tamayo[1,2,§], Lloyd Loza-Muller[1], Yahaíra J. Tamayo-Ordóñez[1,3], Felipe Vázquez-Flota[1]*

[1]Unidad de Biología Integrativa, Centro de Investigación Científica de Yucatán, Calle 43 No. 130 Chuburná 97205, Mérida, Yucatán, México.
[2]Current address: División de Biología Experimental y Aplicada, Centro de Investigación Científica y de Educación Superior de Ensenada, Carretera Ensenada-Tijuana No. 3918 Zona Playitas 22860, Ensenada BC México.
[3]Current address: Facultad de Ciencias Químicas, Universidad Autónoma de Coahuila, Ing J. Cárdenas Valdez S/N República 25280, Saltillo Coahuila México.

[§]*These authors contributed equally*

***Corresponding author:** Felipe Vázquez-Flota, email: felipe@cicy.mx

Received February 19th, 2024; Accepted November 11th, 2024.

DOI: http://dx.doi.org/10.29356/jmcs.v69i4.2223

Abstract. A transcriptomic approach was employed to describe a set of putatively protein-coding sequences involved in the biosynthesis of berberine and sanguinarine, the two major benzylisoquinoline alkaloids (BIA) from *Argemone mexicana* (L.; Papaveraceae). A robust *de novo* assembled transcriptome was obtained from developing seedlings. Initial screening identified 514 unigenes, from eight different Pfam domains, such as Cyt-P450 dependent proteins, which are recurrently involved in BIA biosynthesis. Additional annotation by KEGG Orthology and Gene Ontology supported putative participation of the selected proteins in alkaloid biosynthesis. Moreover, *in silico* structure prediction of sanguinarine reductase (SanR), dihydrobenzophenantridine oxidase (DHBO) and tetrahydroprotoberberine oxidase (STOX), involved in the last reactions of sanguinarine and berberine biosynthesis, fitted to those of previously characterized proteins from related species, and thus, further supporting proper annotation. Hence, the pipeline analysis presented can provide a comprehensive description of the biosynthetic potential of this plant through functionality associated to its transcripts.
Keywords: *Argemone mexicana*; benzylisoquinoline alkaloids; cytochrome P450; dihydrobenzophenantridine oxidase; sanguinarine reductase; tetrahydroprotoberberine oxidase; transcriptomics.

Resumen. Un acercamiento transcriptómico se empleó para predecir un conjunto de secuencias condificantes para proteínas presuntamente involucradas en la biosíntesis de dos de los principales alcaloides bencilisoquinlínicos (ABI) en *Argemone mexicana* (L.; Papaveraceae). Un transcriptoma robusto, se obtuvo de plántulas en desarrollo. Un cribaje inicial identificó 514 unigenes de ocho dominios Pfam diferentes, incluyendo proteínas dependientes del CitP450, que participan en muchas reacciones en la biosíntesis de ABI. Anotaciones adicionales siguiendo la ortología KEGG y Gene Ontology sugirió la participación de un grupo de proteínas en la biosíntesis de alcaloides. Más aún, el modelaje estructural *in silico* de la sanguinarina reductasa (SanR), dihidrobenzophenantridina oxidasa (DHBO) y tetrahidroprotoberberine oxidase (STOX), responsables de las últimas reacciones de la síntesis de sanguinarina y berberina encajaron los previamente descritos para estas enzimas en otras especies relacionadas, confirmando la asignación recibida en la anotación. De este modo, el

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

análisis bioinformático realizado puede ser útil para la descripción detallada del potencial biosintético de esta planta a través de la caracterización funcional de los candidatos seleccionados.
Palabras clave: *Argemone mexicana*; alcaloides bencilisoquinolínicos; citocromo P450; dihidrobenzofenantridina oxidasa; sanguinarina reductasa; tetrahydroprotoberberina oxidasa; transcriptómica.

Introduction

Argemone mexicana L., commonly known as Mexican prickly poppy, cardosanto or chicalote, belongs to the Papaveraceae family. It is widely spread through tropical and subtropical ecosystems and often used in traditional medicine. Ancient Mesoamerican cultures used it against parasitic and microbial infections [1–3]. However, it is also a poisonous plant, particularly for the presence of alkaloids in its seeds [4]. Although it accumulates more than 20 alkaloids, berberine, a protoberberine, and sanguinarine, a benzophenanthridine, represent the most prominent ones [5,6]. Sanguinarine has potent antiviral and cytotoxic effects [7,8], whereas berberine displays important antidiabetic effects on hyperglycemic rats [9]. Moreover, recent studies have demonstrated that berberine ameliorates the inflammatory response in severely affected COVID-19 patients [10]. Hence, a growing interest exists in obtaining both these alkaloids. Unfortunately, only low quantities of them could be recovered directly from plant tissues, so other viable alternatives are under analysis. Although their chemical synthesis or semi-synthesis has been explored, the presence of multiple chiral centers makes large-scale production economically unfeasible [11,12]. Metabolic engineering represents an interesting approach for the up-yield production of plant specialized metabolites (PSM), especially in non-model species. However, this approach requires an extensive knowledge of the genes involved in the metabolic pathways of interest, not only for the biosynthetic process, but also for its regulation and intermediary transport [13]. In this sense, high throughput technologies ("omic" sciences) constitute a valuable tool for the exploration, analysis and application of the vast gene pool (genomics, transcriptomics) and molecular diversity (metabolomics) related to PSM [14,15]. Next generation sequencing technologies, along with current methods for functional annotation, offer comprehensive tools to predict and characterize proteins, within reasonably closeness to the actual *in vivo* functions, offering a wide range of resources to be applied in alkaloid engineering research [16]. In this study, a high sensitivity bioinformatic method was used to identify a set of unique protein-coding sequences involved in BIA biosynthesis, within an *A. mexicana* transcriptomic dataset in the absence of a reference genome.

Experimental

Plant material

In vitro plantlets were obtained from seeds that were collected and disinfested as previously described [17]. Developing seedlings were harvested after the emergence of the first pair of true leaves (non-cotyledonary) and formation of secondary roots (ca. 20 days after germination). Both sanguinarine and berberine are actively synthetized at these developmental phases [17]. Plantlets were sectioned into shoots (hypocotyl and leaflets) and radicles (main and lateral roots). Fresh tissues were quickly processed in a ribonuclease-free environment and immediately subjected to the extraction process to avoid important changes at the transcriptional level.

RNA isolation and sequencing

Total RNA was extracted using the Spectrum Plant Total RNA kit (Sigma-Aldrich, St Louis MO), followed by DNA decontamination with Turbo DNA-free kit (Invitrogen, Waltham MA), according to the manufacturer's protocols. Samples containing 5 µg of total RNA were preserved in RNAstable matrix (Biomatrica, San Diego CA) until analysis. RNA NGS was performed using an Ion torrent semiconductor sequencing platform by an external service provider. Prior to library preparation, RNA quality was verified by capillary electrophoresis on the RNA 6000 pico gel matrix (Agilent Technologies, Santa Clara CA), loaded on the Agilent 2100 Bioanalyzer. At this step, only two samples with an RNA Integrity Number (RIN) above 8

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

were selected for further processing. These QC-passed samples (one from root and other from aerial tissues) were purified and polyadenylated mRNA enriched with Dynabeads mRNA DIRECT Micro Purification kit (Ambion Austin TX). cDNA synthesis was conducted with Total RNA-Seq Kit v2 (Thermo Fisher Scientific, Watham MA), using strand-compatible Ion Torrent sequencing adapters. RNA sequencing template was prepared with the Ion PI Hi-Q OT2 200 Kit to deliver into the Ion Proton™ System (Thermo Fisher Scientific), with a sequencing depth about 30 million reads and a length up to 200 bp fragment reads.

Transcriptome processing and assembling

Raw reads were examined and filtered with the FastQC tool [18]. Low complexity sequences and those with a length below 20 bp and/or with a Phred score <23 were discarded. Next, all read sets were trimmed and filtered with FASTX [19] to eliminate adapters and artifacts. In order to obtain a robust reference transcriptome in terms of tissue representability, a de novo assembly of pooled read sets was conducted with Oases RNA-seq assembler [20] configured to a k-mer of 27, after assaying different values (19, 21, 27 and 29). The assembly's completeness was assessed with the Benchmarking Universal Single-Copy Orthologue (BUSCO) tool v5.7.1 [21], using the eudicots odb10 dataset in transcriptome mode. Additionally, N50 and N90 statistics were calculated employing a custom script.

Functional annotation

All the analyzed sequences were retrieved from the above described XT1 *A. mexicana* assembled transcriptome. The XT1 transcriptome was deposited as a Transcriptome Shotgun Assembly (TSA) project, publicly available at NCBI Sequence Read Archive (SRA) under the run accessions SRR18335407 and SRR18335408. Bioproject: PRJNA814261. Biosample: SAMN26542188.

In order to predict proteins encoded by the XT1 reference transcriptome, open reading frame (ORF) calling was performed with the TransDecoder module of Trinity suite [22], using default parameters. Translated transcriptomes (longest_orfs.pep files) were header-formatted and indexed with SAMtools [23], in conjunction with BEDtools [24] that was used to manipulate and extract sequence subsets for further analysis. The predicted coding sequences (CDS) were analyzed using HMMER3 [25] versus the Pfam-A database [26] by hmmscan searches for domain composition.

Best-hit results (cutoff e-value < 0.001) were screened for Pfam domains related to BIAs biosynthesis. Protein families retrieved included pyridoxal-dependent decarboxylases, pathogenesis-related Bet v 1, mycolic acid cyclopropane synthetase, berberine and berberine like, O-methyltransferase, Cytochrome P450, NAD(P)H-binding domain and FAD binding domain proteins, corresponding to PF00282, PF00407, PF02353, PF08031, PF00891, PF00067, PF13460 and PF01565 records, respectively [27–30].

Function of the Pfam selected CDS was further confirmed by local alignment against the NCBI RefSeq non-redundant (nr) protein database (February 12, 2022 release) using DIAMOND v2.0.14 [31]. In addition, sequences were re-annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) Ortholog family (KOfam) assignment, based on Hidden Markov Model (HMM) profiles [32]. Metabolic assignments of CDS were reconstructed and visualized, employing KEGG mapper [32]. Finally, Gene Ontology (GO) terms of annotated proteins were obtained at the PANNZER2 web server [34]. Visualization of GO terms were conducted with REVIGO [35].

Selection and analysis of protein-coding sequences

Annotated sequences related to BIA biosynthesis were manually filtered to remove duplicates and incompletes. DBOX [EC. 1.3.1.12] and SanR [EC. 1.3.1.107], corresponding to last step in sanguinarine biosynthesis and dihydrosanguinarine interconversion, were selected for subsequent analysis, due to the lack of reports in *A. mexicana*. Multiple Sequence Alignments (MSA) were obtained with ClustalW, followed by a phylogenetic analysis of maximum likelihood and JTT matrix-based evolutionary model [36]. All these processes were conducted using MEGA X software [37]. For the above, the query sequences of previously characterized enzymes in related species were obtained from public databases. *Papaver somniferum* ADOX5, ADOX7 and ADOX8 (NCBI accessions: AGL44334.1, AGL44335.1 and AGL44336.1), respectively, were included for DBOX, as well as a FAD linked oxidase from Macleaya cordata (OVA00265.1) and the (S)-tetrahydroprotoberberine oxidase from *A. mexicana* (ADY15027.1). In the case of SanR, the UNIPROT record A0A6J0ZSP7, two sequences from *Papaver somniferum* (XP_026412126.1 and XP_026397103.1) and one

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

from Eschscholzia californica (ADE41047.1) were included. The selected A. mexicana SanR candidates (AmSanR1 and AmSanR2) were submitted to in silico 3D structural analysis using PHYRE2 (http://www.sbg.bio.ic.ac.uk/phyre2/) [38] and the 3D viewer FirstGlance (http://firstglance.jmol.org).

Results and discussion

Transcriptome overview

A total of 75,378 CDS, corresponding to at least 100 amino acid length putative proteins (PP), were extracted from 97,592 *de novo* assembled transcripts from the *A. mexicana* seedling XT-1 transcriptome. Comparison to Pfam and nr data bases assigned functional groups to over 70 % of these PP. in contrast, KEGG and GO only assigned functional groups to 32 and 49 % of them, respectively (Table 1; Fig 1 and 2). KEGG identified 24,364 *Argemone* PP, and 1,063 of them were assigned catalytic functions. Those PP were distributed among 479 metabolic pathways; it is noteworthy to point out that they were involved in amino acid biosynthesis and cofactors, as well as carbon metabolism (131, 100 and 104 hits, respectively; Fig. 1). Interestingly, up to 30 PP were related to isoquinoline alkaloid biosynthesis (map: 00950; Fig. 1). KEGG comparison also revealed *Argemone* PP involved in the processing of environmental and genetic information, cellular processes, and organismal systems (Supplementary file 1; Fig. S1).

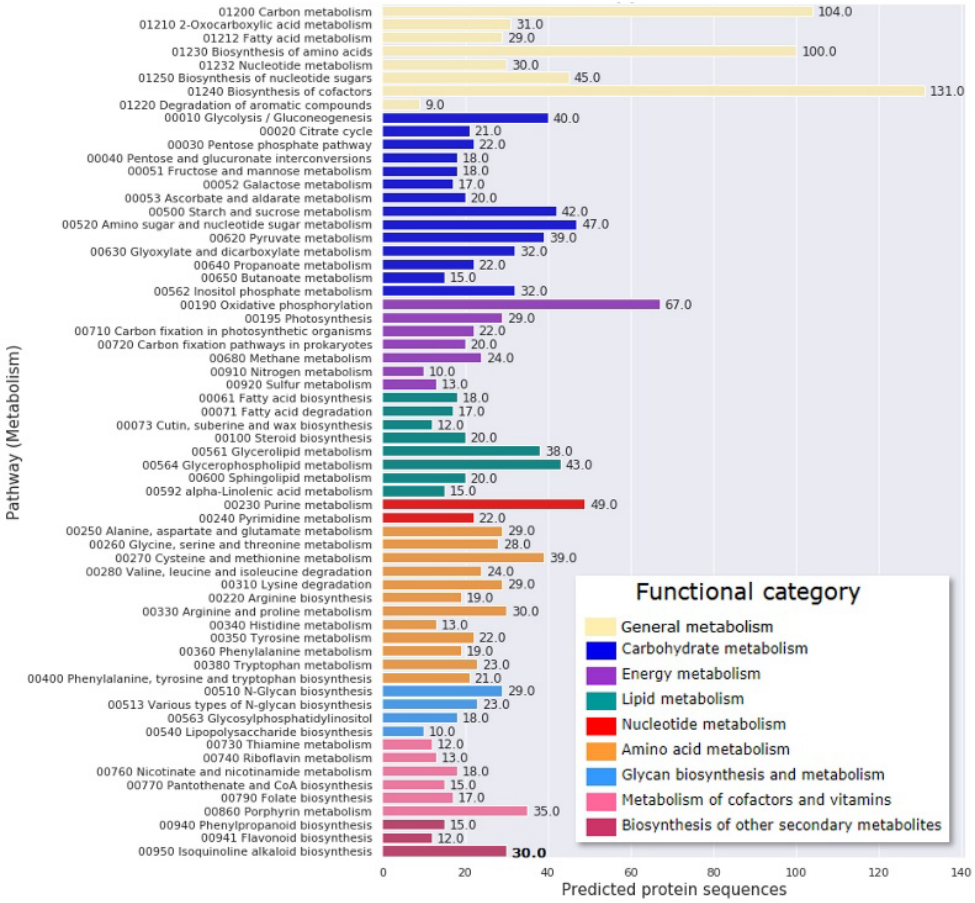

Fig. 1. Reconstruction of the metabolic pathways represented in the XT1 *A. mexicana* transcriptome by KEGG mapper. Number of putative proteins identified of the different categories are indicated at the end of each bar.

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

Table 1. Summary of XT1 sequencing, assembly statistics, and functional annotation features.

Aerial body post QC reads	17 011 810
Root system post QC reads	17 757 669
Percentage of assembled reads	63.31%
Assembled contigs/transcripts	97 592
Largest contig/transcript size	1 0951 nt
N50 contig size	1 409 nt
N90 contig size	375 nt
Total BUSCO groups searched	2 326
Complete BUSCOs	1 853 (80%)
Putative proteins from unigenes*	75 378
Pfam annotated	54 867 (72.79%)
nr annotated	61 969 (82.21%)
KEGG annotated (KOfam)	24 364 (32.32%)
GO annotated	37 241 (49.4%)

*Number of predicted protein-coding sequences using TransDecoder.LongOrfs algorithm that identifies ORFs that are at least 100 amino acids long.

GO annotation assigned 37,241 PP into three major categories (Fig. 2**(A)**). Most of GO terms related to berberine and sanguinarine biosynthesis, such as tyrosine decarboxylase (GO:0004837), (S)-coclaurine-N-methyltransferase (GO:0030794) and (S)-tetrahydroprotoberberine N-methyltransferase (GO:0030782) activities, were highly represented within the recognized molecular functions, and this was consistent with assignments made by Pfam and KEGG. Likewise, coincidences were also detected in terms associated to plant growth regulators, such as receptors for ethylene (GO:0038199) and auxins (GO:0038198) (Fig. 2**(B)**).

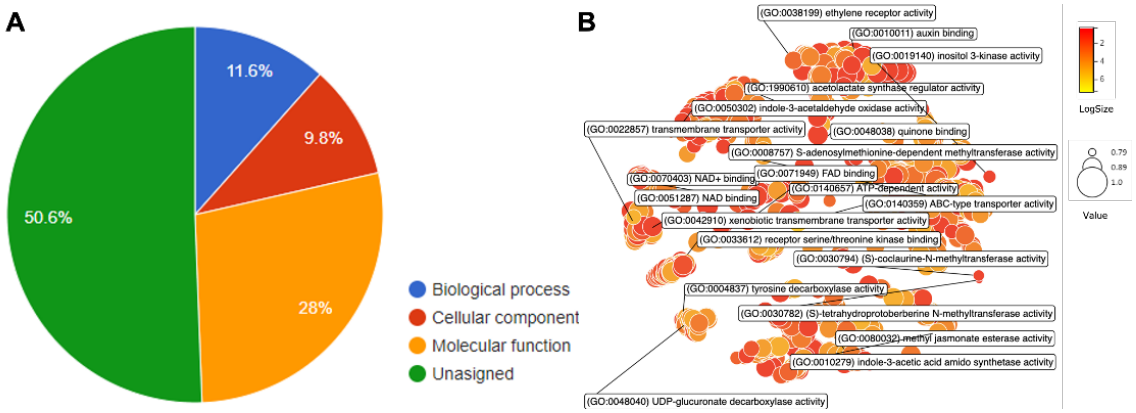

Fig. 2. GO annotation of XT1 transcriptome. **(A)** GO major category classification of protein-coding sequences. **(B)** Revigo visualization of the molecular function terms. Labeled bubbles in B include terms representative for specialized metabolism, amino acid biosynthesis and growth regulators.

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

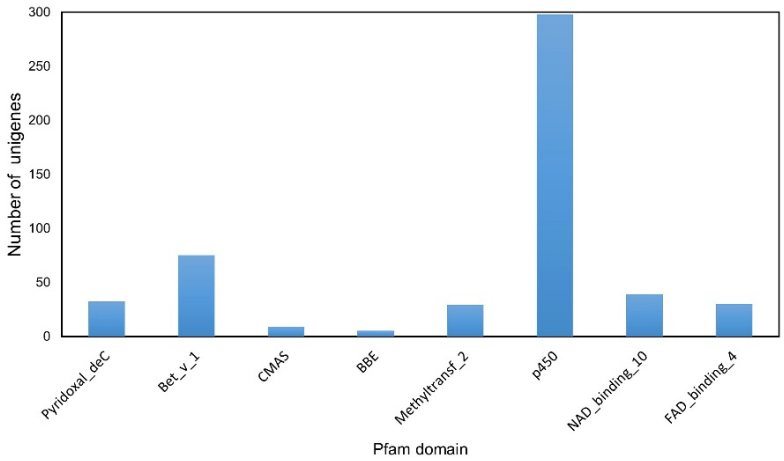

Fig. 3. Pfam classification of proteins involved in BIAs biosynthesis identified in the XT-1 *A. mexicana* transcriptome. **Pyridoxal_deC**, Pyridoxal-dependent aminoacid decarboxylase; **Bet_v_1**, Pathogenesis-related protein Bet v 1 Family; **CMAS**, Mycolic acid cyclopropane synthetase; **BBE**, Berberine and berberine like; Methyltransf_2, O-methyltransferase; **p450**, Cytochrome P450; **NAD_binding_10**, NAD(P)H-binding; **FAD_binding_4**, FAD binding domain.

BIA related proteins in XT1 transcriptome

A total of 514 unigenes encoding putative BIA biosynthetic enzymes were identified within the Pfam screened transcriptome (Fig. 3). This set, together with sequences related to the secondary metabolism identified by KEGG and GO functional annotations, were analyzed by BLAST to further confirm their identity (Supplementary file 2; Dataset S1). The larger group corresponded to Cytochrome P450 proteins (CYPs) followed by Pathogenesis-related protein 10/Bet v 1, from the major allergen protein family (PR10/Bet v1-like), which shares homology with norcoclaurine synthase (NCS).

Interestingly, candidates for the complete biosynthetic routes for both berberine and sanguinarine were reconstructed from the assembled transcriptome by phylogenetic analysis (Supplementary file 1; Fig. S2). The related Pfam records, identified as described in Materials and Methods (see Functional Annotation), are displayed in Fig. 4 and Table 2, which also lists the enzyme acronyms and EC numbers. The first set of reactions involves the transformation of two tyrosine units into the three-hydroxylated intermediary norcoclaurine, with the participation of TyDC and NCS (Table 2; Fig. 4). The next track of reactions consists in the transformation of norcoclaurine into reticuline. This requires the addition of three methyl groups: two on hydroxyl substitutions and one on the heterocyclic N of the structure. Further closing of the reticuline methylene bridge, performed by BBE, produces scoulerine, the last common intermediary for sanguinarine and berberine synthesis (Table 2; Fig. 4). Sanguinarine synthesis from scoulerine proceeds via a dehydro- intermediary and requires the formation of two methylendioxy bridges, followed by oxidation and hydroxylation, as depicted in Fig. 4. This requires the involvement of CheSyn (CYP719A14), StySyn (CYP719A13), TNMT, MSH and PH6. Final steps consist in the interconversion of dihydrosanguinarine and sanguinarine with the participation of SanR and DBOX (Fig. 4). On the other hand, berberine synthesis from scoulerine initiates with a methylation, followed by the formation of the corresponding methylendioxy bridge and further oxidation (Fig. 4). These reactions are performed by SOMT, CDS (CYP719A13) and STOX (Table 2) [6]. In this way, developing *A. mexicana* seedlings possess the complete set of enzymes involved in the synthesis of BIAs from both, the benzophenanthridine and protoberberine, groups. This suggests that genes involved in coordinating the operation of both branches are also present in the XT-1 transcriptome. RPKM values (see Supplementary) indicated a lower BIA gene expression in aerial parts (maximum 200 RPKM) than roots (Fig. S3**(a)**), which is consisted with previous observations in is *Argemone* developing seedlings [5,17]. However, general transcriptional activity in aerial tissue seems considerable, as suggested by a selected photosynthetic marker (ribulose bisphosphate carboxylase oxidase small subunit; RuBisCO; rbcs; Fig. S3**(a)**). Higher BIA gene expression in roots was confirmed by RT-qPCR analysis of a group of selected genes (Fig. S3**(b)**) which showed a similar trend.

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

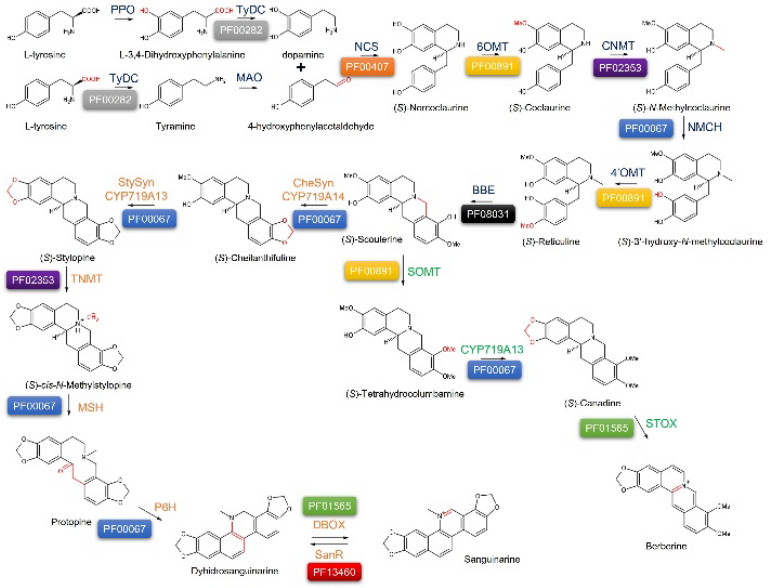

Fig. 4. Schematic representation of sanguinarine and berberine biosynthetic pathways in *A. mexicana*. Colored boxes indicate the Pfam record associated to the functional domain of the different enzymes. **PPO**, Polyphenol oxidase; **TyDC**, tyrosine/DOPA decarboxylase; **MAO**, Monoamine oxidase; **NCS**, norcoclaurine synthase; **6OMT**, norcoclaurine *6-O*-methyltransferase; **CNMT**, coclaurine *N*-methyltransferase; **NMCH**, *N*-methylcoclaurine 3'-hydroxylase; **4'OMT**, *3'*-hydroxy-*N*-methylcoclaurine *4'-O*-methyltransferase; **BBE**, berberine bridge enzyme; **CheSyn/CYP719A14**, cheilanthifoline synthase; **CYP719A13**, trifunctional (S)-stylopine synthase/(S)-nandinine synthase/(S)-canadine synthase; **TNMT**, tetrahydroprotoberberine *N*-methyltransferase; **MSH**, *N*-methyl-stylopine 14-hydroxylase; P6H, protopine 6-hydroxylase; **DBOX**, dihydrobenzophenanthridine oxidase; **SanR**, sanguinarine reductase; **SOMT**, scoulerine 9-*O*-methyltransferase; **STOX**, tetrahydroprotoberberine oxidase.

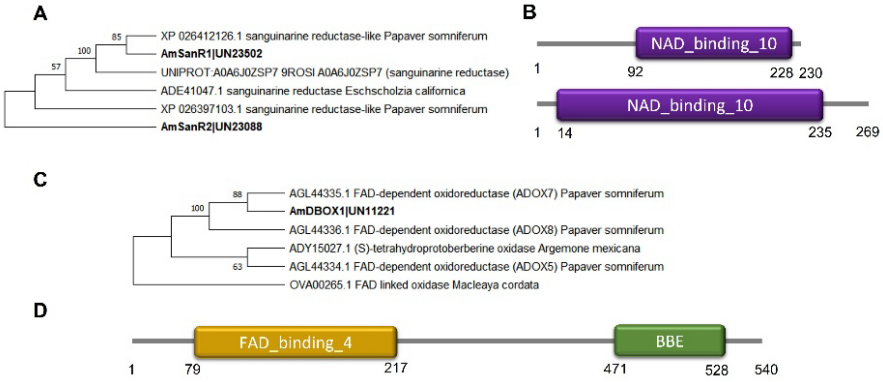

Fig. 5. *In silico* analysis of the putative SanR (AmSanR) and DBOX (AmDBOX) retrieved from the XT1 *A. mexicana* transcriptome. **(A)** Maximum likelihood phylogenetic tree of sanguinarine reductase related sequences. **(B)** Domain architecture scheme of UN23502 and UN23088 transcripts encoding a putative SanR isoforms (AmSanR1 and AmSanR2). **(C)** Maximum likelihood phylogenetic tree of dihydrobenzophenantridine oxidase related sequences. **(D)** Domain architecture scheme of UN11221 transcript encoding a putative DBOX (AmDBOX1). Bootstrap frequencies for each clade are percentages of 1000 iterations. Species and associated GenBank accession numbers for phylogenetic tree construction are indicated in each taxon name. Numbers at the bottom of the predicted architectures indicate the domain position in amino acids.

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

Table 2. BIA biosynthetic enzymes identified in XT1 transcriptome.

Enzyme (Abbreviation)	**Pathway**	**Transcript ID (Length of predicted coding product in amino acids)**	**Best hit accession and/or reference**
Tyrosine/DOPA decarboxylase (TyDC)	Common	UN06517 (532) UN06518 (532) UN07029 (537)	ACJ76782.1
Norcoclaurine synthase (NCS)	Common	UN18055 (351) UN06380 (274)	ACO90257.2 [51] ACJ76785.1
Norcoclaurine *6-O*-methyltransferase (6OMT)	Common	UN45871 (137)	XP_026447518.1
Coclaurine *N*-methyltransferase (CNMT)	Common	UN22975 (365) UN19321 (322)	XP_017224815.1 ANY58190.1
N-methylcoclaurine 3'-hydroxylase (NMCH)	Common	UN16831 (488) UN13116 (513) UN10440 (368)	XP_010253990.1
3'-hydroxy-*N*-methylcoclaurine *4'-O*-methyltransferase (4'OMT)	Common	UN18945 (356)	XP_026440860.1
Berberine bridge enzyme (BBE)	Common	UN14424 (548)	ACJ76783.1
Cheilanthifoline synthase (CheSyn/CYP719A14)	Sanguinarine	UN16884 (405)	B1NF20.1
Trifunctional (S)-stylopine synthase/(S)-nandinine synthase/(S)-canadine synthase (CYP719A13)	Sanguinarine and/or berberine	UN11761 (504)	B1NF19.1 [45]
Tetrahydroprotoberberine *N*-methyltransferase (TNMT)	Sanguinarine	UN14356 (362) UN15660 (387)	XP_026421878.1
N-methyl-stylopine 14-hydroxylase (MSH)	Sanguinarine	UN08536 (465)	OVA14716.1
Protopine 6-hydroxylase (P6H)	Sanguinarine	UN04162 (511)	XP_026456227.1
Dihydrobenzophenanthridine oxidase (DBOX)	Sanguinarine	UN11221 (540)	AGL44334.1[44]
Sanguinarine reductase (SanR)	Sanguinarine	UN23502 (230) UN23088 (269)	ADE41047 [41] XP_026397103.1
Scoulerine 9-*O*-methyltransferase (SOMT)	Berberine	UN12378 (371)	ALY11061.1
Tetrahydroprotoberberine oxidase (STOX)	Berberine	UN10862 (454)	ADY15027.1 [52]

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

Phylogenetic analysis and protein structure modelling of DBOX and SanR

Interconversion of sanguinarine to dihydrosanguinarine is carried out by the pair SanR/DBOX and plays a critical role in protecting plant cells from the possible toxic effects caused by an excessive accumulation of sanguinarine [39]. Although these enzymes have been described in other species, such as *E. californica* [39], these ones in *A. mexicana* have been not reported. Candidates for AmSanR and AmDBOX were identified in the XT-1 transcriptome. For AmSanR, two proteins; AmSanR-1 and -2 were selected due to their phylogenetic proximity to PsSanR1 (XP_026397103.1) and EcSanR1 (ADE41047) (Fig. 5**(A)**). Both PP presented the required domains for binding of $NAD(P)^+$ and for reductase activity (Fig. 5**(B)**). Interestingly, despite being only 230 and 269 residues long, respectively, both AmSanR-1 and -2 corresponded to complete proteins, as revealed by the pairwise comparison to other SanR and for the presence of the untranslatable regions (UTRs) at both ends of the original transcripts (UN23502 and UN23088). These results suggest that AmSanR1 and -2 represent possible isoforms, as it could be deduced from multiple alignment and protein *in silico* modeling (Fig. 6). On the other hand, the complete AmDBOX1 coding sequence, belonging to the UN11221 transcript, was 540 amino acids long and displayed 50 and 68 % identity to the *P. somniferum* DBOX (PsADOX5) and PsADOX7, respectively (Fig. 5**(C)**). The predicted architecture of AmDBOX1 included the typical structural DBOX features, such as the FAD-binding pocket and the catalytic BBE-like domain (Fig. 5**(D)**).

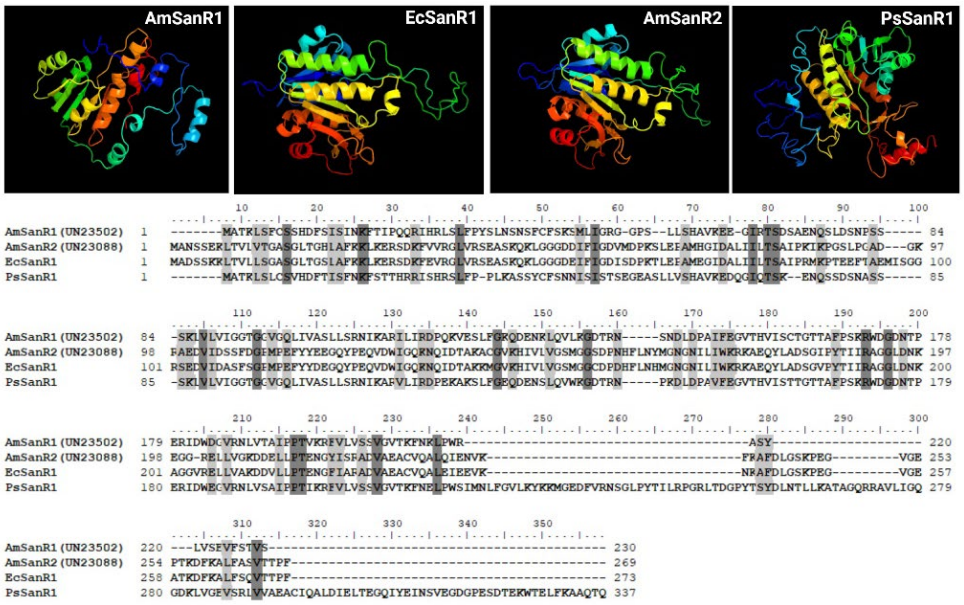

Fig. 6. AmSanR1 and AmSanR2 alignment and 3D structure. The 3D structure can be visualized at the top panel. The alignment of both sequences with that of the PsSanR1 and EcSanR1 can be visualized at the lower panel which indicates the divergent and conserved residues between them. The 3D structure and alignment were realized through PHYRE2 and BioEdit respectively. The caption was created with BioRender.com.

Conclusions

In exploring an *A. mexicana* trancriptomic data set by homology of protein domains using the hidden Markov models, Pfam classification and predictive structural modeling allowed the identification of strong candidates of coding sequences involved in the complete BIA biosynthetic pathway in *A. mexicana*, a non-model plant species used in traditional medicine. Hence, this approach may be applied in similar studies, such as the identification of regulatory proteins or those involved in metabolite transportation.

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

Acknowledgments

We thank Ing. L.A. Puc-Canul for providing support for the use of the supercomputing infrastructure (HOBON server; CICY). In addition, the authors thank Dr. M.L. Miranda-Ham for her critical review and Dr. A. Muñoz-Sánchez and MSc. M. Monforte-González for technical support during different stages of this work.

This work was supported by CONAHCYT (National Council for Humanities, Science and Technology, Mexico), grants CB-2016-0285887 and CBF 2023-2024-1879. JGS-G is awarded a research assistance scholarship from SNI/CONACYT. LLM was recipient of a postdoctoral scholarship from CONAHCYT (514907/289293).

References

1. Gbesso, G. H. F.; Gbesso, F. K.; Adoukonou, R. C. F.; Akabassi, G. C.; Padonou, E. A.; Tente, A. B. H. *Ethnobot. Res. Appl.* **2021**, *21,* 1–11. DOI: https://doi.org/10.32859/era.21.20.1-11
2. Priya, C. L.; Rao, Kokati V. B. *Int. J. Pharm. Sci. Res.* **2012**, *36,* 2143–48. DOI: http://dx.doi.org/10.13040/IJPSR.0975-8232.3(7).2143-48
3. Rubio-Pina, J.; Vazquez-Flota, F. *Curr. Top. Med. Chem.* **2013**, *13,* 2200–2207. DOI: https://doi.org/10.2174/15680266113139990152
4. Babu, C. K.; Khanna, S. K.; Das, M. *Antioxid. Redox Signaling.* **2007**, *9,* 515–525. DOI: https://doi.org/10.1089/ars.2006.1492
5. Vázquez-Flota, F; Rubio-Piña, J.; Xool-Tamayo, J.; Vergara-Olivares, M.; Tamayo-Ordoñez, Y.; Monforte-González, M.; Guízar-González, C.; Mirón-López, G. *Rev. Fitotec. Mex.* **2018**, *41,* 13–21. DOI: https://doi.org/10.35196/rfm.2018.1.13-21
6. Laines-Hidalgo, J. I.; Muñoz-Sánchez, J. A.; Loza-Müller, L.; Vázquez-Flota, F. *Molecules,* **2022**, *27, 1378.* DOI: https://doi.org/10.3390/molecules27041378
7. Chang, Y. C.; Chang, F. R.; Khalil, A. T.; Hsieh, P. W.; Wu, Y. C. *Zeitschrift für Naturforschung C.* **2003**, *58,* 521–526. DOI: https://doi.org/10.1515/znc-2003-7-813
8. Gali, K.; Ramakrishnan, G.; Kothai, R.; Jaykar, B. *Int. J. Pharmtech. Res.* **2011**, *3,*1329–1333.
9. Nayak, P.; Kar, D. M.; Maharana, L. *Pharmacologyonline*. **2011**, *1, 889–903.*
10. Zhang, B. Y.; Chen, M.; Chen, X. C.; Cao, K.; You, Y.; Qian, Y. J.; Yu, W. K. *Br. J. Surg.* **2021**, *108,* e9–e11. DOI: https://doi.org/10.1093/bjs/znaa021
11. Namdeo, A. G.; Jadhav, T. A.; Rai, P. K.; Gavali, S.; Mahadik, K. *Phcog. Rev*. **2007**, *1, 227–231.*
12. Isah, T.; Umar, S.; Mujib, A.; Sharma, M. P.; Rajasekharan, P. E.; Zafar, N.; Frukh, A. *Plant Cell Tiss. Organ Cult.* **2018**, *132,* 239–265. DOI: https://doi.org/10.1007/s11240-017-1332-2
13. Yadav, A. N.; Kour, D.; Rana, K. L.; Yadav, N.; Singh, B.; Chauhan, V. S.; Rastegari, A. A.; Hesham, A. E.; Gupta, V. K., in: *New and Future Developments in Microbial Biotechnology and Bioengineering: Microbial Secondary Metabolites Biochemistry and Applications*. Chapter 20, Vijai Kumar Gupta and Anita Pandey, Elsevier, **2019**, 279-320. DOI: https://doi.org/10.1016/B978-0-444-63504-4.00020-7
14. Dasgupta, A.; Chowdhury, N.; De, R. K. *Comput. Methods Programs Biomed*. **2020**, *192,* 105436. DOI: https://doi.org/10.1016/j.cmpb.2020.105436
15. Marchev, A. S.; Yordanova, Z. P.; Georgiev, M. I. *Crit. Rev. Biotechnol*. **2020**, *40,* 443–458. DOI: https://doi.org/10.1080/07388551.2020.1731414
16. Yamada, Y.; Sato, F. *Biomolecules.* **2021***, 11,* 1719. DOI: https://doi.org/10.3390/biom11111719
17. Xool-Tamayo, J.; Serrano-Gamboa, G.; Monforte-González, M.; Mirón-López, G.; Vázquez-Flota, F. *Biotechnol. Lett.* **2017**, *39,* 323–330. DOI: https://doi.org/10.1007/s10529-016-2250-9
18. http://www.bioinformatics.babraham.ac.uk/projects/fastqc, accessed in October 2023.
19. http://hannonlab.cshl.edu/fastx_toolkit, accessed in January 2024

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

20. Schulz, M. H.; Zerbino, D. R.; Vingron, M.; Birney, E. *Bioinformatics*. **2012**, *28,* 1086–1092. DOI: https://doi.org/10.1093/bioinformatics/bts094
21. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E. V.; Zdobnov, E.M. *Bioinformatics*, **2015**, *31*, 3210-3212. DOI: https://doi.org/10.1093/bioinformatics/btv351
22. Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A. *Nat. Biotechnol.* **2011**, *29,* 644–652. DOI: https://doi.org/10.1038/nbt.1883
23. Danecek, P.; Bonfield, J. K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M. O.; Whitwham, A.; Keane, T.; McCarthy, S. A.; Davies, R. M.; Li, H. *GigaScience*. **2021**, *10,* giab008. DOI: https://doi.org/10.1093/gigascience/giab008
24. Quinlan, A. R.; Hall, I. M. *Bioinformatics*. **2010**, *26,* 841–842. DOI: https://doi.org/10.1093/bioinformatics/btq033
25. Eddy, S. R. *PLoS Computational Biology*. **2011**, *7,* e1002195. DOI: https://doi.org/10.1371/journal.pcbi.1002195
26. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G. A.; Sonnhammer, E. L. L.; Tosatto, S. C. E.; Paladin, L.; Raj, S.; Richardson, L. J.; Finn, R. D.; Bateman, A. *Nucleic Acids Res.* **2021**, *49,* D412–D419. DOI: https://doi.org/10.1093/nar/gkaa913
27. Dastmalchi, M.; Park, M. R.; Morris, J. S.; Facchini, P. *Phytochem. Rev*. **2018**, *17,* 249–277. DOI: https://doi.org/10.1007/s11101-017-9519-z
28. Gracz-bernaciak, J.; Mazur, O.; Nawrot, R. *Int. J. Mol. Sci.* **2021**, *22,* 12427. DOI: https://doi.org/10.3390/ijms222212427
29. Facchini, P. J.; Morris, J. S. *Front. Plant Sci*. **2019**, *10,* 1058. DOI: https://doi.org/10.3389/fpls.2019.01058
30. Zhong, F.; Huang, L.; Qi, L.; Ma, Y.; Yan, Z. *Plant Mol. Biol.* **2020**, *102,* 477–499. DOI: https://doi.org/10.1007/s11103-019-00959-y
31. Buchfink, B.; Reuter, K.; Drost, H.-G. *Nat. Methods*. **2021**, *18,* 366–368. DOI: https://doi.org/10.1038/s41592-021-01101-x
32. Aramaki, T.; Blanc-Mathieu, R.; Endo, H.; Ohkubo, K.; Kanehisa, M.; Goto, S.; Ogata, H. *Bioinformatics*. **2020**, *36, 2251–2252*. DOI: https://doi.org/10.1093/bioinformatics/btz859
33. Kanehisa, M.; Sato, Y. *Protein Sci*. **2020**, *29, 28–35*. DOI: https://doi.org/10.1002/pro.3711
34. Törönen, P.; Medlar, A.; Holm, L. *Nucleic Acids Res*. **2018**, *46,* W84–W88. DOI: https://doi.org/10.1093/nar/gky350
35. Jones D.T.; Taylor W.R. *FEBS Lett* **1994**, *339,* 269–275. DOI:10.1016/0014-5793(94)80429-X.
36. Supek, F.; Bošnjak, M.; Škunca, N.; Šmuc, T. *PLoS ONE*. **2011**, *6,* e21800. DOI: https://doi.org/10.1371/journal.pone.0021800
37. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. *Mol. Biol. Evol.* **2018**, *35,* 1547–1549. DOI: https://doi.org/10.1093/molbev/msy096
38. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. *Nat. Protoc.* **2015**, *10,* 845–858. DOI: https://doi.org/10.1038/nprot.2015.053
39. Vogel, M.; Lawson, M.; Sippl, W.; Conrad, U.; Roos, W. *J. Biol. Chem.* **2010**, *285,* 18397–18406. DOI: https://doi.org/10.1074/jbc.M109.088989
40. Deng, X.; Zhao, L.; Fang, T.; Xiong, Y.; Ogutu, C.; Yang, D.; Vimolmangkang, S.; Liu, Y.; Han, Y. *Hortic. Res.* **2018**, *5,* 29. DOI: https://doi.org/10.1038/s41438-018-0035-0
41. Hagel, J. M.; Morris, J. S.; Lee, E.-J.; Desgagné-Penix, I.; Bross, C. D.; Chang, L.; Chen, X.; Farrow, S. C.; Zhang, Y.; Soh, J.; Sensen, C. W.; Facchini, P. J. *BMC Plant Biol.* **2015**, *15,* 227. DOI: https://doi.org/10.1186/s12870-015-0596-0
42. Pei, L.; Wang, B.; Ye, J.; Hu, X.; Fu, L.; Li, K.; Ni, Z.; Wang, Z.; Wei, Y.; Shi, L.; Zhang, Y.; Bai, X.; Jiang, M.; Wang, S.; Ma, C.; Li, S.; Liu, K.; Li, W.; Cong, B. *Hortic. Res.* **2021**, *8,* 5. DOI: https://doi.org/10.1038/s41438-020-00435-5

J. Mex. Chem. Soc. **2025**, *69*(4)
Regular Issue

ISSN-e 2594-0317

43. Morris, J. S.; Caldo, K. M. P.; Liang, S.; Facchini, P. J. *ChemBioChem*. **2021**, *22,* 264–287. DOI: https://doi.org/10.1002/cbic.202000354
44. Hagel, J. M.; Beaudoin, G. A. W.; Fossati, E.; Ekins, A.; Martin, V. J. J.; Facchini, P. J. *J. Biol. Chem.* **2012**, *287, 42972–42983*. DOI: https://doi.org/10.1074/jbc.M112.420414
45. Díaz Chávez, M. L.; Rolf, M.; Gesell, A.; Kutchan, T. M. *Arch. Biochem. Biophys.* **2011**, *507,* 186–193. DOI: https://doi.org/10.1016/j.abb.2010.11.016
46. Leong, B. J.; Last, R. L. *Curr. Opin. Struct. Biol.* **2017**, *47, 105–112.* DOI: https://doi.org/10.1016/j.sbi.2017.07.005
47. Waki, T.; Takahashi, S.; Nakayama, T. *BioEssays*. **2021**, *43,* 2000164. DOI: https://doi.org/10.1002/bies.202000164
48. Loza-Muller, L.; Shitan, N.; Yamada, Y.; Vázquez-Flota, F. *Planta*. **2021**, *254,* 122. DOI: https://doi.org/10.1007/s00425-021-03780-4
49. Kato, N.; Dubouzet, E.; Kokabu, Y.; Yoshida, S.; Taniguchi, Y.; Dubouzet, J. G.; Yazaki, K.; Sato, F. *Plant Cell Physiol*. **2007**, *48,* 8–18. DOI: https://doi.org/10.1093/pcp/pcl041
50. Yamada, Y.; Kokabu, Y.; Chaki, K.; Yoshimoto, T.; Ohgaki, M.; Yoshida, S.; Kato, N.; Koyama, T.; Sato, F. *Plant Cell Physiol*. **2011**, *52,* 1131–1141. DOI: https://doi.org/10.1093/pcp/pcr062