BMC Genomics

DATABASE

Open Access

RGeasy: a reference gene analysis tool for gene expression studies via RT-qPCR

Micaele Rodrigues de Souza[1], Ivo Pontes Araújo[2], Wosley da Costa Arruda[2], André Almeida Lima[3], Solange Aparecida Ságio[1], Antonio Chalfun-Junior[3] and Horllys Gomes Barreto[1*]

Abstract

Gene expression through RT-qPCR can be performed by the relative quantification method, which requires the expression normalization through reference genes. Therefore, it is essential to validate, experimentally, the candidate reference genes. Thus, although there are several studies that are performed to identify the most stable reference genes, most them validate genes for very specific conditions, not exploring the whole potential of the research since not all possible combinations of treatments and/or conditions of the study are explored. For this reason, new experiments must be conducted by researchers that have interest in analyzing gene expression of treatments and/or conditions present, but not explored, in these studies. Here, we present the *RGeasy* tool, which aims to facilitate the selection of reference genes, allowing the user to choose genes for a greater number of combinations of treatments/conditions, compared to the ones present in the original articles, through just a few clicks. *RGeasy* was validated with RT-qPCR data from gene expression studies performed in two coffee species, *Coffea arabica* and *Coffea canephora*, and it can be used for any animal, plant or microorganism species. In addition to displaying a rank of the most stable reference genes for each condition or treatment, the user also has access to the primer pairs for the selected reference genes.

Keywords Endogenous genes, Relative expression, Normalization

Background

The Reverse Transcription-quantitative Polymerase Chain Reaction (RT-qPCR) has been considered the gold standard technique for gene expression analysis for the last ten years or more, mainly due to its sensitivity and precision [1, 2]. Quantification of gene expression through this technique can be performed through the absolute or relative methods [3]. Absolute quantification is carried out through a standard curve, usually generated with plasmid DNA or in vitro transcribed RNA, which enables to determine the exact quantity of target-DNA molecules present in the samples [4–6]. On the other hand, in the relative quantification method, gene expression is determined by order of magnitude, obtained through the comparison of the given sample to a reference one. This method of quantification is preferred to expression analysis of samples submitted to different types of treatments, since it permits to control different sources of variations through the use of reference genes [6].

Different factors can affect the reliability of the data generated from relative expression assays, such as RNA integrity and purity, cDNA quality, primer efficiency, and

*Correspondence:
Horllys Gomes Barreto
horllys@uft.edu.br
[1]Laboratory of Molecular Analysis (LAM), Department of Life Sciences, Federal University of Tocantins, UFT, University Campus of Palmas, Palmas, TO 402-970, Brazil
[2]Computer Science Course, Federal University of Tocantins, University Campus of Palmas, Palmas, TO, Brazil
[3]Laboratory of Plant Molecular Physiology, Department of Biology, Federal University of Lavras, Lavras, MG, Brazil

the proper selection of the reference genes, since they are used in the normalization process [2]. Reference genes enable data normalization due to their relative constant expression levels, thus acting as internal controls [5–7]. However, it is essential that the chosen reference genes are experimentally validated, in order to prove their stability and thus avoid imprecise data normalization of RT-qPCR studies [8].

RT-qPCR data normalization is performed by using the Cycle of Quantification (Cq) values, which is defined as the cycle where the fluorescence level reaches a threshold that can be manually or automatically established [6]. Currently, the different tools that aid in choosing the most stable genes use Cq values to calculate the relative expression of the genes, such as the RefFinder tool, which classifies reference genes through the integration of different algorithms: GeNorm [9], NormFinder [10], BestKeeper [11], and delta-Ct [12]. Thus, considering the classification from each algorithm, an appropriated weight is assigned for each individual gene and the geometric mean of their weights is calculated for the final overall classification, resulting in a ranking [13, 14]. However, a drawback comes from this process considering that this type of information is not always available in scientific papers. Therefore, here we present the tool *REFERENCE GENE EASY (RGeasy)* (http://rgeasy.com.br), a free tool which targets two different audiences. The first group is composed by researchers that have developed or are developing reference gene validation studies. Researchers are able to deposit their data from published studies on the *RGeasy* database, thus providing the required information to enable the classification of reference genes for all possible combinations of treatments and/or conditions of the study (Fig. 1A). The second group is composed by researchers planning to develop gene expression studies, with *RGeasy* enabling them to skip the reference gene validation step (Fig. 1B).

Results and discussion

Graphic interface

The initial graphic interface of *RGeasy* provides information about its developers, as well as the features it has to offer.

When using RGeasy, users have access to the species registered on the tool by clicking on "Species" in the navigation bar located in the upper part of the initial interface. Species deposited on *RGeasy* are separated into three categories: Animals, Plants, and Microorganisms. Currently, RGeasy's database has five, five, and three registered species of animals, plants, and microorganisms, respectively (Fig. 2). By clicking on the species of interest, it is automatically shown all the reference gene validation studies registered on *RGeasy* for that species (Fig. 3). In this interface, users have access to each study by clicking on its title. Under the title from each study, it is displayed the types of samples analyzed on them.

In order to define the desired combination of treatments or conditions, users must select the samples of interest by clicking on the icon beside them (Fig. 3). The result is instantly shown by clicking on "Run RefFinder". Since *RGeasy* uses the RefFinder tool [13, 14] to analyze the stability of the reference genes, a table is generated on the results page with a gene ranking according to the following algorithms: RefFinder, Delta CT, Bestkeeper, Normfinder and Genorm.

In addition, on the results page, *RGeasy* provides a table with some additional information for each reference gene, according to the stability ranking from RefFinder [13, 14]. For each reference gene, the primer pair, the correlation coefficient (R^2), the amplification efficiency, the accession number, and the database from which the sequence was obtained, are made available to the user. All this information is provided during species registration by the researchers that conducted the study of reference gene validation.

Tests (data sets)

RGeasy validation was performed using gene expression data from two coffee species. Searches for coffee reference gene validation studies on the *Web of Science* and *Scopus* databases resulted in nine studies, similarly to what was found by Fernandes-Brum et al. [15]. A thorough analysis of these papers allowed the observation that five of them did not include all the possible combinations of treatments or conditions, being indicated on the "Treatment combination" column from Table 1 as "not analyzed".

The absence of Cqs values in most RT-qPCR papers (supplementary material) makes it impossible to define the reference genes for the combination not explored in the original articles. This scenario implies that new experiments are necessary for analyzing combinations of treatments or conditions not included by the researchers in their original work, resulting in a greater demand of time and other resources that, with RGeasy, this can now be avoided. In general, it could be found that in the studies that did not include all possible combinations, an average of 10 new combinations of treatments and conditions could be found (Supplementary material). For instance, Barsalobres-Cavallari et al. [16] have analyzed the effects of biotic stress in different coffee (*Coffea arabica*) tissues (roots, stem, leaves, flowers, fruits, and all of these tissues together) and identified the best reference genes for each tissue. However, this study has not evaluated paired combinations among the tissues, such as roots and leaves, or leaves and fruits and so on (Table 1).

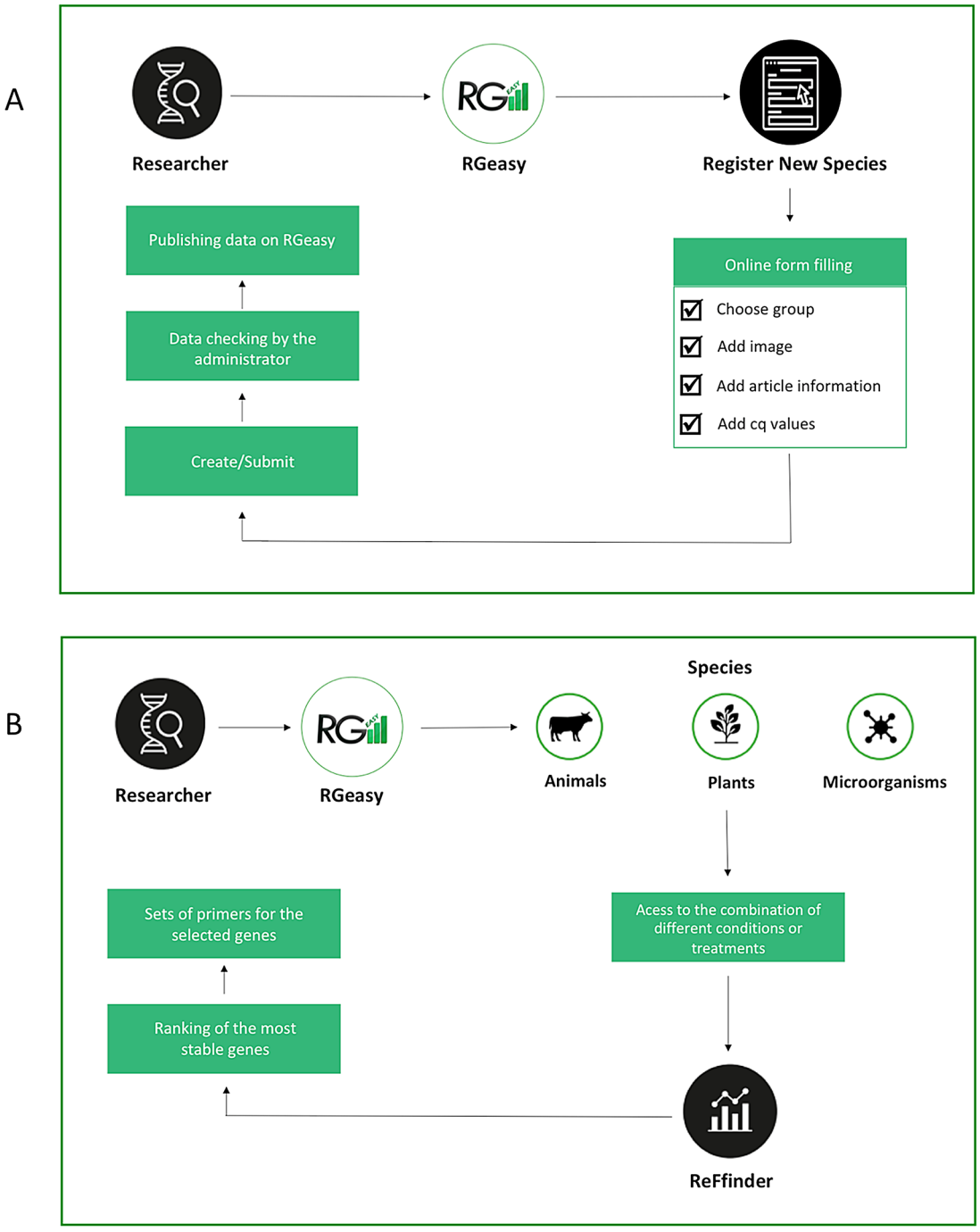

Fig. 1 Registration process of a new dataset on Rgeasy. Steps of data deposition (**A**). Steps of RGeasy use flow (**B**)

Case study

In order to confirm if RGeasy was properly working, after adding the data of the Freitas et al. [23] and Fernandes-Brum et al. [15] studies, it was verified if the results generated by the tool were in accordance with results from the combinations of treatments and conditions previously described in the two articles. This analysis allowed us to prove *RGeasy*'s efficiency, since the same results were obtained, as demonstrated in Figs. 4 and 5 for the data from *C. canephora*.

When compared to the combinations of treatments and conditions explored in the original study conducted by Freitas et al. [23], 16 new combinations were identified by *RGeasy*. The two most stable reference genes were identified to each new combination (Tables 2, 3 and 4), considering that normalization against a single reference gene is not acceptable, unless investigators present

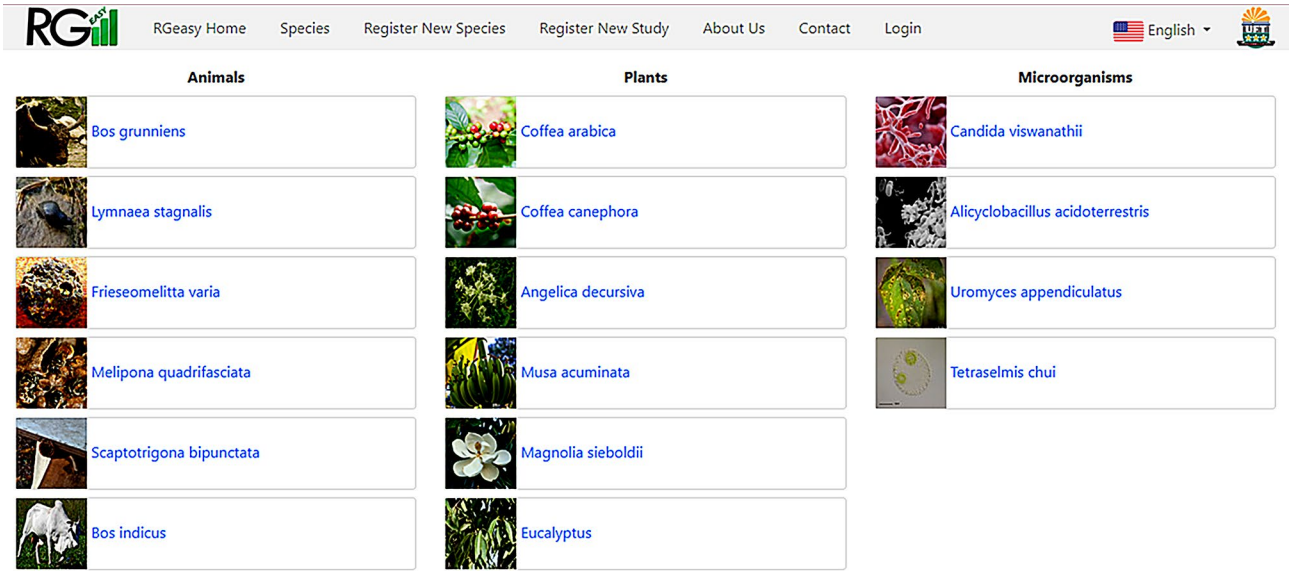

Fig. 2 Species categories on *RGeasy*

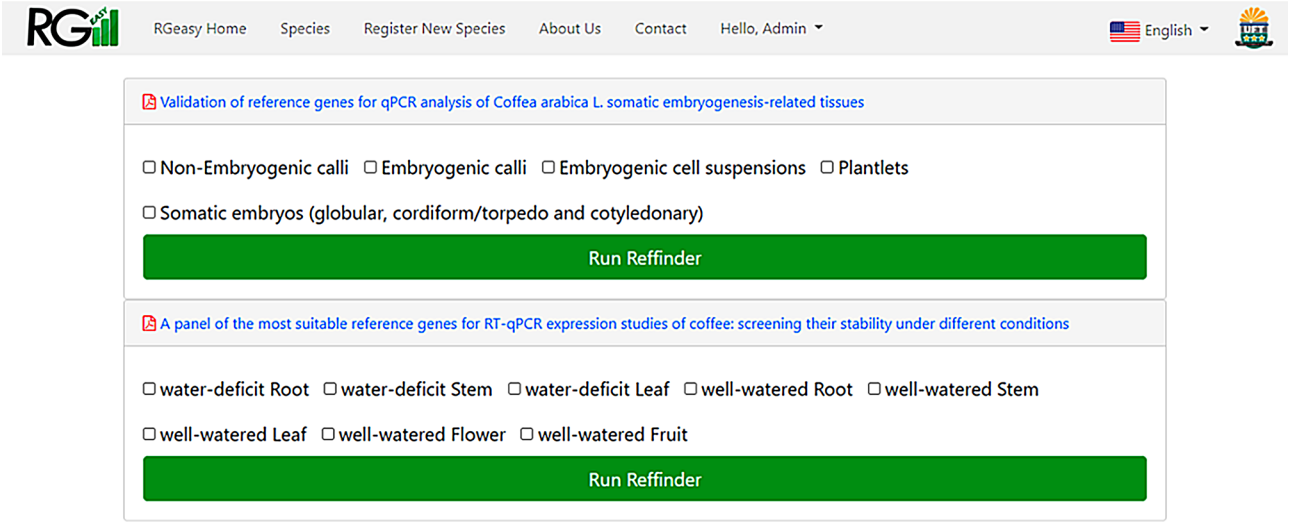

Fig. 3 Reference gene validation studies registered on *RGeasy* for coffee species and the samples analyzed in each study

clear evidence that the gene is stably expressed under the experimental conditions of the study [24].

From the 12 tested genes by Freitas et al. [23], eight were among the most stable genes for the new combinations, with *24 S* and *PP2A* being the most stable ones for combinations including somatic embryos, and *UBQ* being the most stable gene in only one of the new combinations (Table 2). The other combinations displayed specific pairs of reference genes (Table 2), not previously observed in the original research, indicating the importance of validating the candidate reference genes for each treatment/condition.

In relation to the study conducted by Fernandes-Brum et al. [15], 27 and 21 new treatment and condition combinations could be evaluated for *C. arabica* and *C. canephora*, respectively (Tables 3 and 4). For *C. arabica*, considering the two conditions analyzed (water-deficit / well-watered), the two new combinations of tissues showed that *AP47* and *RPL39* were the most stable genes, similar to previous combinations of the original study. Under water-deficit conditions, three new combinations could be analyzed, in which four of the 12 reference genes evaluated were shown to be the most stable genes (Table 3). Among these genes, three of them (*AP47/ UBQ/ RPL39*) had been identified as stable reference genes in previous combinations of the original article [15] but for the new combination "Roots and Stem", the gene *24 S*, one of the most stable genes in this case, had not been indicated as a stable gene for any combination of the study. Under well-watered conditions, 22 new

Table 1 Description of the coffee reference gene validation studies found on the web of Science and Scopus databases and their status towards the possible combinations of treatments or conditions analyzed in the study. When all possible combinations were evaluated in the study, it was categorized as "Analyzed", otherwise it was categorized as "not analyzed"

Article	Species	Organ/Tissue	Experimental condition	*N°* of tested genes	Recommended genes	Treatment combinations
[16]	*C. arabica*	Roots, stem, leaves, flowers, fruits, and their combination	Non-inoculated (control) and inoculated with *Hemileia vastatrix*.	8	*GAPDH*, *14-3-3*, and *RPL7*	Not analyzed
[17]	*C. arabica*	Roots and leaves	Control versus drought-stressed leaves and control versus drought-stressed roots	8	*AP47*, *24 S*, *UBI9*, *GAPDH*, and *UBQ10*	Not analyzed
		Leaves	Different cultivars		*AP47* and *GAPDH*	Analyzed
		Leaves, stem, roots, and cherry fruits	Tissue combination		*UBQ10*, *24 S*, and *UBI9*	Not analyzed
[18]	*C. arabica*	Leaves	In vitro samples and in planta with *Hemileia vastatrix*/ Control samples in ungerminated *Hemileia vastatrix*	7	*40S_Rib*, *GADPH*, and *Hv00099*	Analyzed
[19]	*C. arabica*	Leaves	Cold stress	10	*UBQ10*, *GAPDH*, *ACT*, and *EF1α*	Analyzed
			Drought stress		*GAPDH*, *ACT*, *EF1α*, and *Apt*	
			Multiple stresses		*UBQ10*, *GAPDH*, *ACT*, and *elf-4 α*	
			Different cultivars/Control (not subjected to stress)		*GAPDH*, *UBQ10*, *AP47*, and *EF1α*	
			Stress and cultivar combination		*GAPHD*, *Cycl*, and *UBQ10*	
[20]	*C. arabica*	Hypocotyls	Biotic stress (*Colletotrichum kahawae*) in different genotypes (susceptible and resistant)/Control (non-inoculated)	10	*β-Tub9* and *IDE*	Analyzed
[21]	*C. arabica*	Leaves and roots	Nitrogen deficiency/Control (non-stressed)	10	*MDH*, *EF1*, *GAPDH*, and *EF1α*	Not analyzed
		Leaves	Salt stress/Control (non-stressed)		*EF1*, *EF1α* and *UBQ10*	
			Temperature stress/Control (non-stressed)		*MDH*, *GAPDH*, and *EF1α*	
[22]	*C. arabica*	Leaves	Genotype, [CO_2], temperature, multiple stress interaction and total stress interaction	10	*MDH*, *ACT*, and *S15*	Analyzed
[23]	*C. arabica*	Somatic embryos, suspension cells, embryogenic and non-embryogenic calli, and plantlets	Different tissues of somatic embryogenesis-related/ Control (non-embryogenic calli)	12	*24 S* and *PP2A*	Not analyzed
[15]	*C. arabica*	Roots, stem, leaves, flowers and fruits	Well-watered and water-deficit	12	*AP47*, *UBQ*, *RPL39*, and *EF1α*	Not analyzed
	C. canephora		Well-watered	8	*ADH2*, *ACT*, *GAPDH*, and *UBQ*	

C. canephora	WW	All tissues grouped	*ADH2/ACT/GAPDH/UBQ*
C. canephora	WW	Root	*ACT/ADH2*
C. canephora	WW	Stem	*ADH2/UBQ*
C. canephora	WW	Leaf	*ADH2/RPL7*
C. canephora	WW	Flower	*RPL7/PSAB*
C. canephora	WW	Fruit	*UBQ/PSAB*

Fig. 4 Most stable reference genes obtained from the study conducted by Fernandes-Brum et al. [15] for all tissues grouped, roots, stem, leaves, flowers and fruits from *C. canephora* plants

combinations could be analyzed through *RGeasy* and nine, out of the 12 reference genes analyzed, were identified in at least one of these combinations (Table 3). Interestingly, the gene *14.3.3*, previously shown to be among the most stable genes only for leaves, was not present among the most appropriate genes for the new combinations involving leaves, with the same occurring for *CYCL* in the case for new combinations comprising flowers.

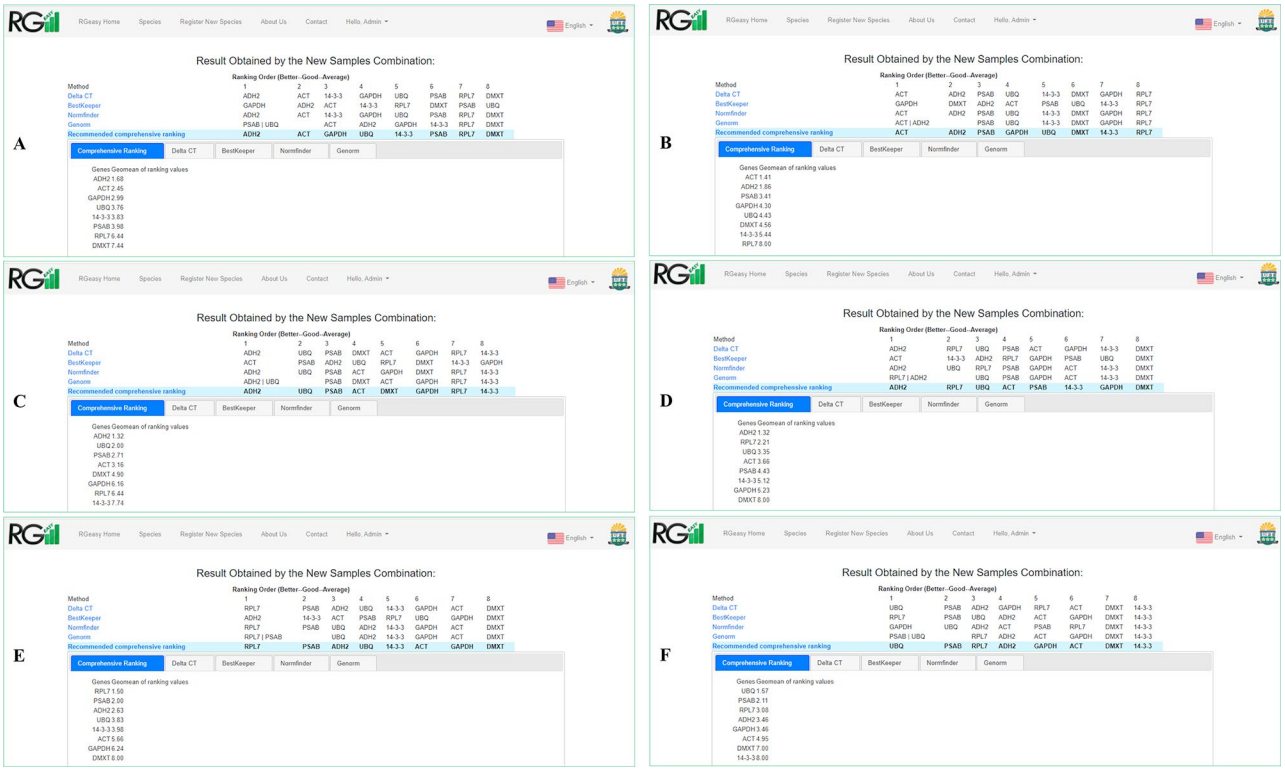

Fig. 5 Reference gene ranking generated by *RGeasy* for all tissues grouped (**A**), roots (**B**), stem (**C**), leaves (**D**), flowers (**E**) and fruits (**F**) from *C. canephora* plants analyzed by Fernandes-Brum et al. [15]

These two genes were also found to be among the most stable reference genes for combinations involving other tissues, such as roots, stem, and fruits, as observed for the "Root and Flower" and "Stem and Fruit" combinations. In addition, *RGeasy* showed that *APT1*, previously considered as a stable gene only for *C. arabica* fruits [15] was among the most stable genes for combinations comprising different tissues, some of them not including fruits (Table 3).

According to Fernandes-Brum et al. [15], from the eight tested reference genes in *C. canephora* tissues, five (*ADH2*, *ACT*, *UBQ*, *RPL7*and *PSAB*) were identified as stably expressed genes when tissues were individually analyzed. However, for the new combinations analyzed by *RGeasy*, *RPL7* was not reported among the stable reference genes. In addition, the analysis revealed that *ADH2* is not present as one of the two most stable reference genes in only four of the new 21 combinations (Table 4).

As previously mentioned, in addition to the classification of the reference genes, *RGeasy* also provides the primer pairs for each reference gene analyzed (Fig. 6), thus minimizing one more factor that can potentially affect the reliability of RT-qPCR data.

Conclusions

In conclusion, *RGeasy*'s database allows the selection of reference genes for a greater number of treatment and condition combinations that is usually present in the original published articles with just a few clicks, revolutionizing the research related to reference gene selection for gene expression studies through RT-qPCR. The tool can thus prevent new experiments from being carried out by exploring RT-qPCR data on its total, thus reducing cost and time that would otherwise be spent to analyze combinations that are of interest for other researchers but were not evaluated in the original articles. In addition, *RGeasy* provides a greater dissemination of published articles, since the selection of reference genes through the tool requires that the researchers cite the original paper used to generate the new treatment combinations, as described at the bottom part of RGeasy's results page.

Methods

System architecture

The *RGeasy* website is implemented in HTML, CSS, and Javascript for the user interface and Laravel Framework with the Mysql database for the backend infrastructure code. The Web Server architecture of RGeasy follows the Model-View-Controller (MVC) pattern.

Table 2 Comparison between the treatment or treatment combinations and their best reference genes, ranked by the RefFinder tool [13, 14], identified by RGeasy (first two columns) and the original study conducted by Freitas et al. [23] (two last columns)

Combinations of treatments and conditions not analyzed in the original study	Genes	Combinations of treatments and conditions analyzed in the original study	Genes
Somatic embryo and embryogenic calli	*PP2A / 24 S*	All samples	*24 S / PP2A*
Somatic embryo and non-embryogenic calli	*24 S / PP2A*	embryogenic cell suspensions,	*APRT / EF1a*
Somatic embryo and plantlets	*PP2A / EF1a*	non-embryogenic calli	*UBQ / ACT*
Somatic embryo, embryogenic calli, and non-embryogenic calli	*PP2A / 24 S*	embryogenic call	*ACT / 24 S*
Somatic embryo, embryogenic cell suspensions, embryogenic calli, and non-embryogenic calli	*24 S / PP2A*	combined embryogenic and non-embryogenic calli samples	*RPL39 / 24 S*
Somatic embryo, embryogenic cell suspensions, embryogenic calli, and plantlets	*24 S / PP2A*	somatic embryos	*PP2A / RPL39*
Somatic embryo, embryogenic calli, non-embryogenic calli, and plantlets	*24 S / PP2A*	plantlets	*PP2A / AP47*
Embryogenic cell suspensions e embryogenic calli	*ACT / APRT*	-	-
Embryogenic cell suspensions and non-embryogenic calli	*APRT / EF1a*	-	-
Embryogenic cell suspensions and plantlets	*TUB / ACT*	-	-
Embryogenic cell suspensions, embryogenic calli, and non-embryogenic calli	*EF1a / APRT*	-	-
Embryogenic cell suspensions, embryogenic calli, and plantlets	*ACT / TUB*	-	-
Embryogenic cell suspensions, embryogenic calli, non-embryogenic calli, and plantlets	*TUB / EF1a*	-	-
Embryogenic calli and plantlets	*RPL39 / 24 S*	-	-
Non-embryogenic calli and plantlets	*UBQ / PP2A*	-	-
Embryogenic calli, non-embryogenic calli, and plantlets	*RPL39 / 24 S*	-	-

The tool works from the input of a dataset, which it is structured and stored on its database [25]. The relational database contains tables of several pieces of information, such as species, stability data and information about primer composition (Correlation coefficient - R^2, accession, gene name, forward and reverse primers).

In order to generate specific's user treatment combinations, RGeasy uses Laravels' Object Relational Mapping (ORM) to create a dynamic table with the samples of biological repetitions from all genes of a study Then, it runs the Reffinder application to return the results, and finally shows them alongside each primer's data, ordered from the best to the worst candidate reference gene.

Use case diagram

In addition to the layered architecture used in the basic infrastructure of *RGeasy*, the Unified Modeling Language (UML) was used for the creation of use case and class diagrams. UML is a graphical representation of software modeling that assists in visualizing and documenting the system through various types of diagrams [26].

The use case diagram (Fig. 7) is a document where it is specified the requirements, allowing to observe how the end user interacts with the system in a determined context. This context can be represented by a text or a list of tasks that define the flow of operations [26].

On the other hand, the class diagram displays the static structure of a system, based on its classes (Fig. 8).

Based on the previously mentioned diagrams, Entity-Relationship diagrams were designed (Fig. 9). These diagrams are responsible for representing the general logical structure of the database [27]. The Entity-Relationship diagram of the *RGeasy* is divided into three groups: framework information tables, information tables related to the registered species and the system permission tables.

The *framework* provides a code infrastructure and applies conventions to reduce the code and learning curve of the tool [28]. For RGeasy, the *Laravel framework* was used.

Management of access control and data collections were performed through the MySQL Database Management System [27], which employs the Structured Query Language (SQL).

Tests (Data sets).

RGeasy validation was performed with a dataset from two coffee species, *Coffea arabica* and *Coffea canephora*, originated from two papers that analyzed the best reference genes in different plant developmental stages and conditions: (1) Validation of reference genes for RT-qPCR analysis of *Coffea arabica* L. somatic embryogenesis-related tissues [23]; (2) A panel of the most suitable reference genes for RT-qPCR expression studies of coffee: screening their stability under different conditions [15]. Data from each article were separately inserted on RGeasy, according to the species under study (*C. arabica* and *C. canephora*). In the case of the study conducted by Fernandes-Brum et al. [15] a different set of reference genes were tested for each species, the study was thus registered twice.

Table 3 Comparison between the treatment or treatment combinations and their best reference genes, ranked by the RefFinder tool [13, 14], identified by RGeasy (first two columns) and the original study conducted by Fernandes-Brum et al. [15] (two last columns) for *C. Arabica*

Combinations of treatments and conditions not analyzed in the original study	Genes	Combinations of treatments and conditions analyzed in the original study	Genes
Roots and stem (water-deficit / well-watered)	*AP47 / RPL39*	All tissues grouped (water-deficit / well-watered)	*AP47/UBQ/RPL39/ EF1α*
Leaves and stem (water-deficit /well-watered)	*AP47 / RPL39*	Root (water-deficit / well-watered)	*APT1/AP47/RPL39/EF1α*
Roots and stem (water-deficit)	*AP47 / 24 S*	Stem (water-deficit / well-watered)	*ACT/AP47/PP2A*
Roots and leaves (water-deficit)	*UBQ / RPL39*	Leaf (water-deficit / well-watered)	*AP47/RPL39*
Stem and leaves (water-deficit)	*RPL39 / AP47*	Root/Leaf (water-deficit / well-watered)	*AP47/RPL39*
Roots and stem (well-watered)	*RPL39 / ACT*	All tissues grouped (water-deficit)	*AP47/UBQ*
Roots and leaves (well-watered)	*AP47 / RPL39*	Root (water-deficit)	*AP47/UBQ*
Roots and flowers (well-watered)	*14.3.3 / RPL39*	Stem (water-deficit)	*ACT/RPL39*
Roots and fruits (well-watered)	*APT1 / RPL39*	Leaf (water-deficit)	*PP2A/14.3.3*
Roots, stem, and leaves (well-watered)	*RPL39 / AP47*	All tissues grouped (well-watered)	*AP47/RPL39*
Roots, Stem, and flowers (well-watered)	*APT1 / ACT*	Root (well-watered)	*24 S/PP2A*
Roots, stem, and fruits (well-watered)	*APT1 / AP47*	Stem (well-watered)	*ACT/UBQ*
Roots, leaves and flowers (well-watered)	*24 S / RPL39*	Leaf (well-watered)	*RPL39/14.3.3*
Roots, leaves, and fruits (well-watered)	*RPL39 / AP47*	Flower (well-watered)	*RPL39/CYCL*
Roots, flowers, and fruits (well-watered)	*APT1 / RPL39*	Fruit (well-watered)	*APT1/TUB-b/CYCL*
Roots, stem, leaves, and flowers (well-watered)	*AP47 / UBQ*	-	-
Roots, stem, leaves, and fruits (well-watered)	*AP47 / RPL39*	-	-
Roots, flowers, and fruits (well-watered)	*APT1 / AP47*	-	-
Roots, leaves, flowers, and fruits (well-watered)	*AP47 / RPL39*	-	-
Stem and leaves (well-watered)	*RPL39 / PP2A*	-	-
Stem and flowers (well-watered)	*APT1 / AP47*	-	-
Stem and fruits (well-watered)	*APT1 / CYCL*	-	-
Stem, flowers, and fruits (well-watered)	*APT1 / AP47*	-	-
Leaves and flowers (well-watered)	*AP47 / PP2A*	-	-
Leaves and fruits (well-watered)	*AP47 / RPL39*	-	-
Leaves, flowers, and fruits (well-watered)	*AP47 / UBQ*	-	-
Flowers and fruits (well-watered)	*AP47 / APT1*	-	-

The study conducted by Freitas et al. [23] comprised five different coffee tissues (embryogenic and non-embryogenic calli, two cell lines of embryogenic cell suspensions with different culture times and six culture times, somatic embryos, and plantlets), and 12 candidate reference genes, resulting in total of 1,728 samples (including the biological and technical replicates). On the other hand, Fernandes-Brum et al. [15] analyzed five different tissues (roots, stem, leaves, flowers, and fruits) for two coffee species (*C. arabica* and *C. canephora*) under well-watered conditions, with roots, stem, and leaves from *C. arabica* being also analyzed under water stress conditions. The stability of 12 candidate reference genes were analyzed on *C. arabica* tissues, resulting in a total of 864 samples. For *C. canephora*, eight candidate reference genes were analyzed, resulting in a total of 360 samples (including the biological and technical replicates).

In addition to these two studies, it was performed a search on the *Web of Science* and *Scopus* databases, using the keywords "reference gene", "housekeeping gene", "endogenous gene", "RT-qPCR", and "Coffea", for studies of reference gene validation for gene expression analysis in coffee species.

Table 4 Comparison between the treatment or treatment combinations and their best reference genes, ranked by the RefFinder tool [13, 14], identified by RGeasy (first two columns) and the original study conducted by Fernandes-Brum et al. [15] (two last columns) for *C. Canephora*

Combinations of treatments and conditions not analyzed in the original study	Genes	Combinations of treatments and conditions analyzed in the original study	Genes
Roots and Stem	*ADH2 / ACT*	All tissues grouped	*ADH2/ACT/GAPDH/UBQ*
Roots and leaves	*ACT / ADH2*	Root	*ACT/ADH2*
Roots and flowers	*ADH2 / ACT*	Stem	*ADH2/UBQ*
Roots and fruits	*ACT / ADH2*	Leaf	*ADH2/RPL7*
Roots, stem, and leaves	*ADH2 / ACT*	Flower	*RPL7/PSAB*
Roots, stem, and flowers	*ADH2 / ACT*	Fruit	*UBQ/PSAB*
Roots, stem, and fruits	*ACT / ADH2*	-	-
Roots, leaves, and flowers	*ADH2 / ACT*	-	-
Roots, leaves and fruits	*ACT / ADH2*	-	-
Roots, stem, leaves and flowers	*ADH2 / ACT*	-	-
Roots, stem, leaves, and fruits	*ADH2 / ACT*	-	-
Stem and leaves	*PSAB / ADH2*	-	-
Stem and flowers	*PSAB / ADH2*	-	-
Stem and fruits	*UBQ / ACT*	-	-
Stem, leaves, and flowers	*PSAB / ADH2*	-	-
Stem, leaves, and fruits	*UBQ / ACT*	-	-
Stem, leaves, flowers, and fruits	*UBQ / ADH2*	-	-
Leaves and flowers	*ADH2 / PSAB*	-	-
Leaves and fruits	*UBQ / ADH2*	-	-
Leaves, flowers, fruits	*UBQ / PSAB*	-	-
Flowers and fruits	*UBQ / PSAB*	-	-

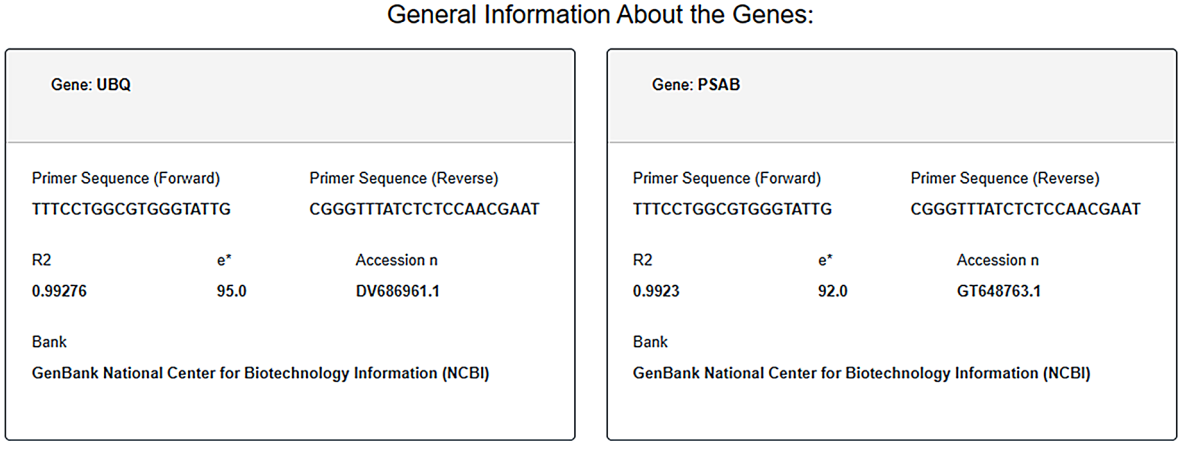

Fig. 6 Additional information, including primer sequence, correlation coefficient (R^2), amplification efficiency, accession number, and database where the sequence is deposited, provided by *RGeasy* for the reference genes deposited on its database

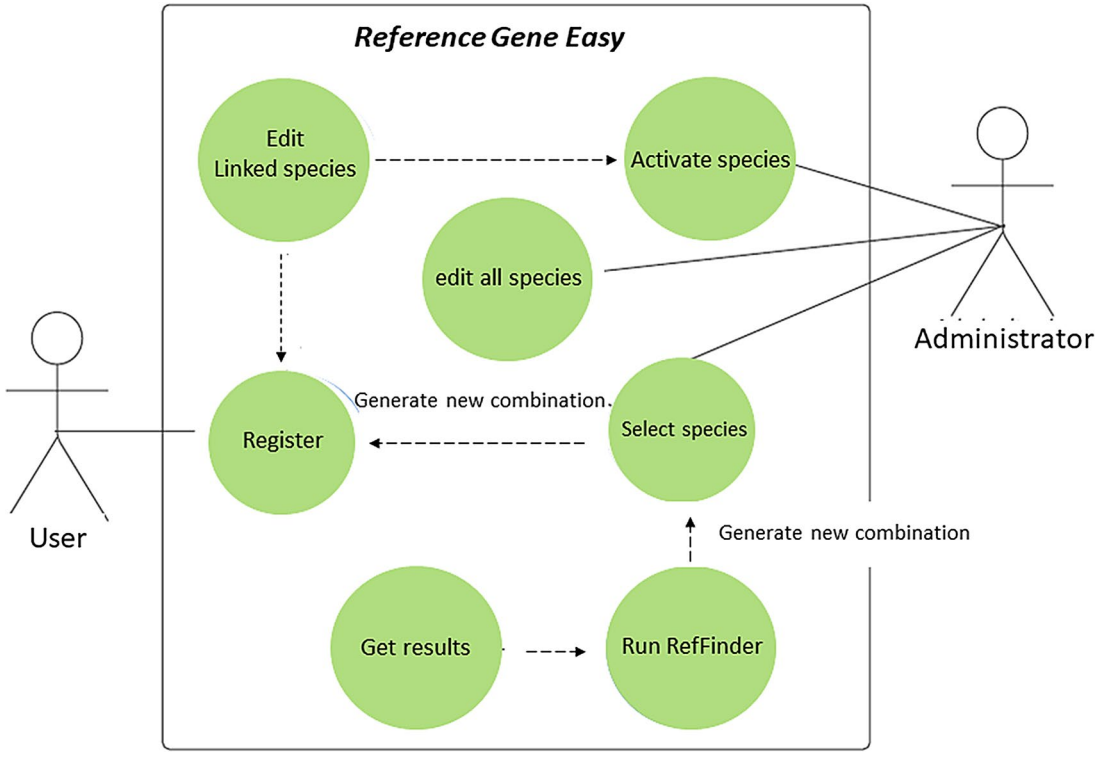

Fig. 7 Use case diagram illustrating the registration process and the management of species

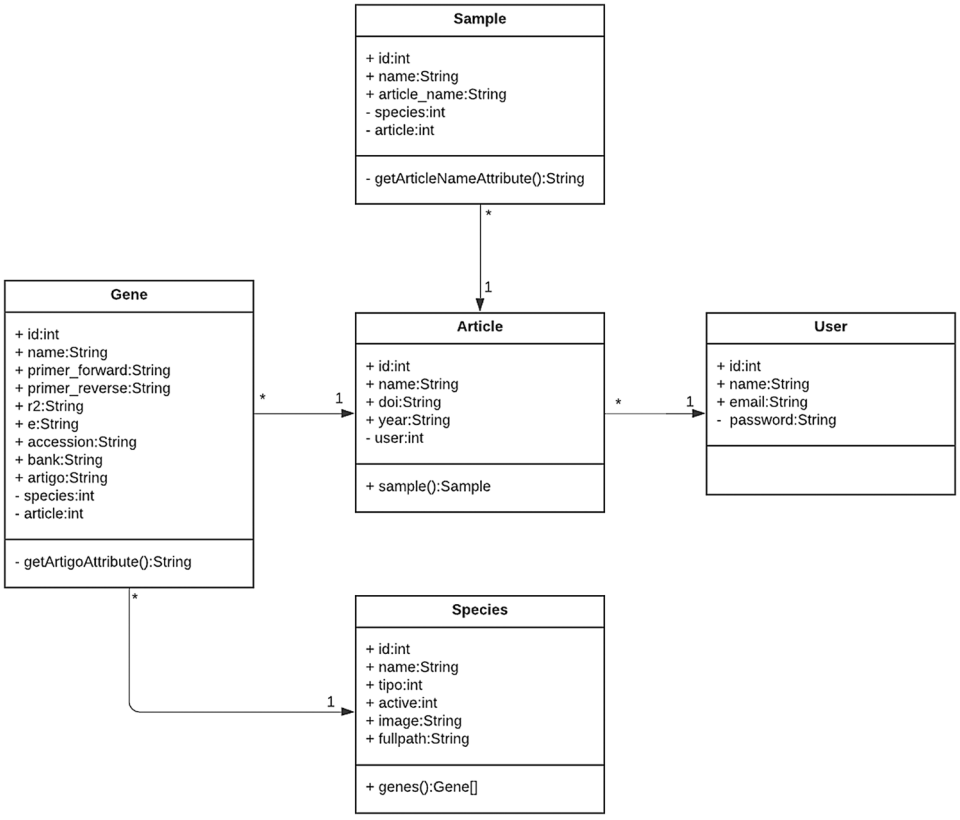

Fig. 8 Class diagram showing the list of the most important classes (Sample, Article, Gene, and Species) of the RGeasy system. For each article that is registered on RGeasy's database there is a sample group and a specific group of genes analyzed for a given species

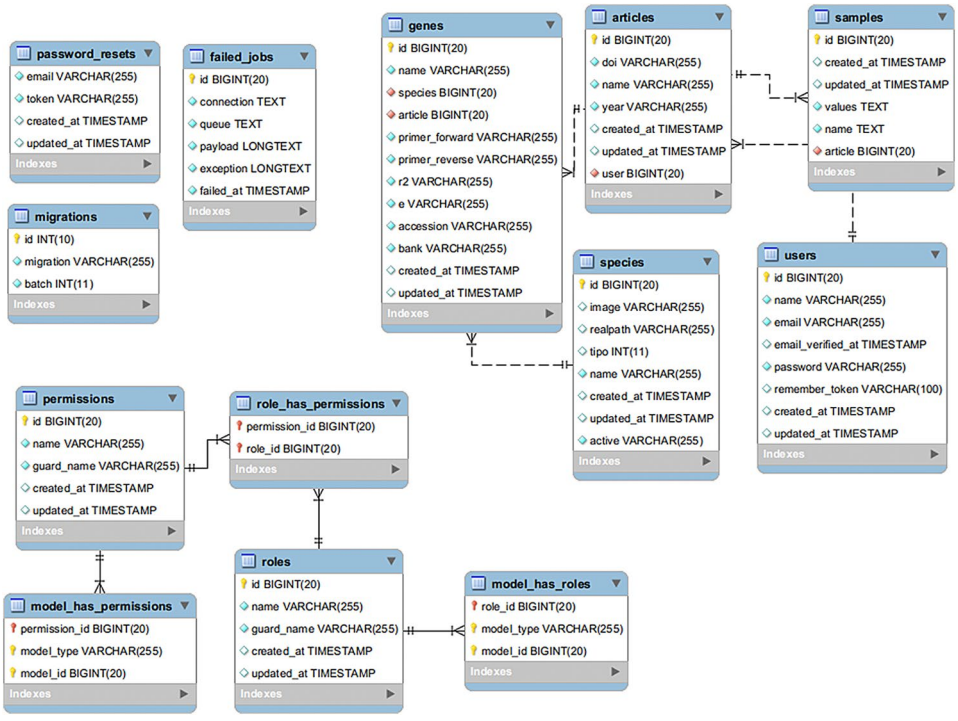

Fig. 9 Entity-Relationship diagrams of RGeasy's database showing its tables. Each table from this diagram displays the specific information of each attribute, such as s maximum word length, metadata and relationships

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10808-y.

Supplementary Material 1

Acknowledgements

We thank the "Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq, the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)", the "Rede de Biodiversidade e Biotecnologia da Amazônia Legal (Bionorte)", the "Programa de Mestrado em Agroenergia da UFT (Agroenergia-UFT)", the "Universidade Federal do Tocantins (PROPESQ-UFT)" and the "Fundação de Amparo à Pesquisa do Estado do Tocantins (FAPT)" for the financial support. We would like to thank all the researchers involved in the development and publication of the RefFinder tool for their excellent contribution to the analysis of real-time PCR data.

Author contributions

Conceived and designed the experiments H.G.B, A.C.J & S.A.S; Experiment, data collection & Software analysis, M.R.S, I.P.A, W.C.A, A.A.L & H.G.B; Supervision, H.G.B & A.C.J; Writing-original draf, M.R.S & I.P.A; Made critical revisions of the content of the paper, H.G.B, A.C.J, W.C.A, S.A.S & A.A.L.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Data availability

The source code is available on GitHub (https://github.com/rgeasy/refencegeneseasy).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 December 2023 / Accepted: 17 September 2024
Published online: 30 September 2024

References

1. Udvardi MK, Czechowski T, Scheible W-R. Eleven golden rules of quantitative RT-PCR. Plant Cell. 2008;20:1736–7.
2. Derveaux S, Vandesompele J, Hellemans J. How to do successful gene expression analysis using real-time PCR. Methods. 2010;50:227–30.
3. Arabkari V, Clancy E, Dwyer RM, Kerin MJ, Kalinina O, Holian E, et al. Relative and absolute expression analysis of micrornas associated with luminal a breast cancer–a comparison. Pathol Oncol Res. 2020;26:833–44.
4. QIAGEN, QuantiFast SYBR, Green PCR. Handbook. For fast, quantitative, real-time PCR and two-step RT-PCR using SYBR Green. 2014.
5. Marcelino-Guimarães F, Darben L, Carvalho K, Marin S. Aplicações da qPCR. In: Introdução às técnicas de PCR convencionm tempo real e digital. 1st edition. Ribeirão Preto: Sociedade Brasileira de Genética; 2018. pp. 129–40.
6. Magalhães D, Arena G, Barbosa E, Litholdo-Junior C. Princípios da qPCR. In: Introdução às técnicas de PCR convencionm tempo real e digital. 1st edition. Ribeirão Preto: Sociedade Brasileira de Genética; 2018. pp. 113–28.
7. Dos Santos CP, da Cruz Saraiva KD, Batista MC, Germano TA, Costa JH. Identification and evaluation of reference genes for reliable normalization of real-time quantitative PCR data in acerola fruit, leaf, and flower. Mol Biol Rep. 2020;47:953–65.
8. Foquet B, Song H. There is no magic bullet: the importance of testing reference gene stability in RT-qPCR experiments across multiple closely related species. PeerJ. 2020;8:e9618.
9. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3:research00341.

10. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse Transcription-PCR data: a model-based Variance Estimation Approach to identify genes suited for normalization, Applied to bladder and Colon Cancer Data sets. Cancer Res. 2004;64:5245–50.
11. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. Biotechnol Lett. 2004;26:509–15.
12. Silver N, Best S, Jiang J, Thein SL. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol Biol. 2006;7:33.
13. Xie F, Xiao P, Chen D, Xu L, Zhang B, miRDeepFinder. A miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol Biol. 2012;80:75–84.
14. Xie F, Wang J, Zhang B. RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes. Funct Integr Genomics. 2023;23.
15. Fernandes-Brum CN, de Oliveira Garcia B, Moreira RO, Ságio SA, Barreto HG, Lima AA, et al. A panel of the most suitable reference genes for RT-qPCR expression studies of coffee: screening their stability under different conditions. Tree Genet Genomes. 2017;13:1–13.
16. Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG. Identification of suitable internal control genes for expression studies in Coffea arabica under different experimental conditions. BMC Mol Biol. 2009;10:1–11.
17. Cruz F, Kalaoun S, Nobile P, Colombo C, Almeida J, Barros LMG, et al. Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. Mol Breeding. 2009;23:607–16.
18. Vieira A, Talhinhas P, Loureiro A, Duplessis S, Fernandez D, do Céu Silva M, et al. Validation of RT-qPCR reference genes for in planta expression studies in Hemileia Vastatrix, the causal agent of coffee leaf rust. Fungal Biol. 2011;115:891–901.
19. Goulao LF, Fortunato AS, Ramalho JC. Selection of reference genes for normalizing quantitative real-time PCR gene expression data with multiple variables in Coffea spp. Plant Mol Biol Rep. 2012;30:741–59.
20. Figueiredo A, Loureiro A, Batista D, Monteiro F, Várzea V, Pais MS, et al. Validation of reference genes for normalization of qPCR gene expression data from Coffea spp. hypocotyls inoculated with Colletotrichum kahawae. BMC Res Notes. 2013;6:1–10.
21. de Carvalho K, Bespalhok Filho JC, Dos Santos TB, de Souza SGH, Vieira LGE, Pereira LFP, et al. Nitrogen starvation, salt and heat stress in coffee (Coffea arabica L.): identification and validation of new genes for qPCR normalization. Mol Biotechnol. 2013;53:315–25.
22. Martins MQ, Fortunato AS, Rodrigues WP, Partelli FL, Campostrini E, Lidon FC, et al. Selection and validation of reference genes for accurate RT-qPCR data normalization in Coffea spp. under a climate changes context of interacting elevated [CO2] and temperature. Front Plant Sci. 2017;8:307.
23. Freitas NC, Barreto HG, Fernandes-Brum CN, Moreira RO, Chalfun-Junior A, Paiva LV. Validation of reference genes for qPCR analysis of Coffea arabica L. somatic embryogenesis-related tissues. Plant Cell Tissue Organ Cult (PCTOC). 2017;128:663–78.
24. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M et al. The MIQE Guidelines: M inimum I nformation for Publication of Q uantitative Real-Time PCR E xperiments. 2009.
25. Pressman RS, Maxim BR. Software Engineering.
26. Valente MT. Engenharia de Software Moderna: Princípios e Práticas para Desenvolvimento de Software com Produtividade. 2020.
27. Coronel C, Morris S, Crockett K, Blewett C. Data Base Principles- fundamentals of design, implementation, and management. 3rd edition. Marinda Louw; 2020.
28. Stauffer M. Laravel Up & Running A Framework for Building Modern PHP Apps.

Publisher's note