AskoR, A R Package for Easy RNASeq Data Analysis

Susete Alves Carvalho, Kévin Gazengel, Anthony Bretaudeau, Stéphanie Robin, Stéphanie Daval, Fabrice Legeai

proceedings

MDPI

Proceedings

AskoR, a R Package for Easy RNA-Seq Data Analysis †

Susete Alves-Carvalho [1,2,‡], Kévin Gazengel [1,‡], Anthony Bretaudeau [1,2], Stéphanie Robin [1,2], Stéphanie Daval [1] and Fabrice Legeai [1,2,*]

[1] IGEPP, INRAE, Institut Agro, Univ Rennes, 35000, Rennes, France
[2] Univ Rennes, Inria, CNRS, IRISA, 35000, Rennes, France
* Correspondence: fabrice.legeai@inrae.fr
† Presented at the 1st International Electronic Conference on Entomology (IECE 2021), 1–15 July 2021; Available online: https://iece.sciforum.net/.
‡ These authors contributed equally to this work.

Abstract: For facilitating the process of transcriptomics data, and to guarantee the reproducibility of our analyses, we developed AskoR, which is a R library for performing a suite of statistical analysis and graphical output from gene expression data obtained by sequencing (RNA-Seq). From raw counts, it makes it possible to filter and normalize the data, to check the consistency of the samples, and to carry out differential expression tests, GO terms enrichments, and clusters of co-expression, with a large number of figures in the output. AskoR can be downloaded and used in your favorite R environment or directly accessible through a Galaxy portal like the one which is hosted by the BioInformatics Platform for the Agroecosystems Arthropods (BIPAA).

Keywords: BioInformatics Platform for Agroecosystems Arthropods; RNA-Seq; differential expression; clustering; Gene Ontology enrichment

Citation: Alves-Carvalho, S.; Gazengel, K.; Masanelli, S.; Bretaudeau, A.; Robin, S.; Daval, S.; Legeai, F. AskoR, a R Package for Easy RNA-Seq Data Analysis. *Proceedings* **2021**, *1*, 0. https://doi.org/10.3390/IECE-10646

Published: 9 July 2021

1. Introduction

The BioInformatics Platform for the Agroecosystems Arthopods (BIPAA) is a bioinformatics platform from the French Research Institute for Agriculture, Food and Environment (INRAE). It is dedicated to support genomics and post-genomics programs developed on insects associated with agroecosystems, and assists cooperation and coordination of multiple communities working on arthopod genomics. The Information System has been created more than 10 years ago to support the International Aphid Genomics Consortium (IAGC), in order to annotate and curate the pea aphid genome [1], and has been continuously improved and extended until the recent achievement of the genomes of phylloxera (*Daktulosphaira vitifoliae*) [2], various parasitoid wasps (*Hyposoter didymator*, *Campoletis sonorensis* [3], *Cotesia congregata* [4], *Aphidius ervi* and *Lysiphlebus fabarum* [5]) or *Spodoptera frugiperda* [6]. Consequently, BIPAA is the home of several public reference databases including AphidBase, LepidoDB and ParWaspDB, each hosting multiple insect genomes. Altogether, 38 genomes are currently available online, and its infrastructure has evolved to support the load of numerous new genomes and to facilitate browsing and navigating. For each species, a collection of web applications allows users to explore reference genomes or transcriptome assemblies and annotations (e.g. genome browser, gene reports), to compare genomics regions (synteny viewer), to analyze these data with multiple tools (e.g. alignment of various sequences, annotation, SNP prediction etc.) through a dedicated Galaxy server [7] or specific web applications (e.g. a blast form), or to correct or add information by curating the genome annotations within Apollo [8]).

RNA-Seq studies are now affordable and widely used in many laboratories. In insect science, this method is still currently use for studying phenomena, such as the molecular responses of whole insects, organs, and tissues to different biotic or abiotic stresses, including exposure to insecticides, infection with microbes, or feeding on different hosts, and improving our knowledge of gene expression changes associated with immunity,

detoxification, chemoreception, or reproduction [9–15]. Many statistical methods have been developed, so far, for analyzing RNA-Seq data from raw counts and to identify differentially expressed genes (DEGs) [16–18], search for enrichment of functions in gene-lists [19,20], create Venn diagrams [21], or for searching for clusters of co-expressed genes [22,23]. Many of these analysis tools require programming skills in R or paid software subscriptions, such as OmicsBox[24] or CLC Workbenches [25], making it difficult for many researchers to implement reproducible workflows. On the other hand, some reproducible workflows have been designed [26] for managing RNA-Seq data from the mapping on a reference annotation (genome or transcriptome) but these workflows do not include a full set of supplementary analysis for statistically and graphically exploring data by classification or enrichment.

Additionally, DEGs can often be integrated with other data types to improve the robustness of studies, such as QTL analysis, orthology studies, epigenomic landscapes, and proteomics or metabolomic studies. Thereupon, the semantic web technology (consisting in data modeling in structured format such as the Ressource Description Framework (RDF) allowing its querying with the SPARQL language), gives the opportunity to link various information from various sources into graphs of data. Moreover, we are developing AskOmics [27,28] which provides a web interface to upload and integrate heterogeneous data files (GFF, BED, and tabulated formats) into RDF and a visual SPARQL query builder software to allow the experts to compose and execute expressive and semantically-rich queries.

Here, we report AskoR, a R pipeline for the analysis of gene expression data with a simple and reproducible script. It includes several steps (data filtering, normalization, sample validation, differential expression analysis, Gene Ontology enrichment and co-expression) and produces numerous files directly uploadable into an AskOmics instance in order to be linked with other data.

2. Materials and Methods

2.1. Installation and deployment

AskoR can be directly installed from the git repository (https://github.com/askomics/askoR). An AskoR Docker image, including a preconfigured RStudio instance, is maintained on GitHub (https://github.com/genouest/docker-galaxy-rstudio-askor). This image can be used in a Galaxy server as an interactive tool, allowing to use the preconfigured RStudio-AskoR environment directly from the Galaxy web interface. Instructions on configuring a Galaxy server to make use of this image are available on the GitHub repository.

2.2. AskoR implementation, usage and dependencies

AskoR is a R package, it consists in a R library, which has to be loaded in a script. A user guide is available at the web site. As a template or example, we are providing a R script running sequentially each function which can be adapted to the needs of the user. AskoR has many parameters and flags (for example method of normalization, p-values or CPM thresholds) with default values which can be modified before or during the process. A complete list of parameters is available on the wiki. Because AskoR takes advantages of various R packages, it has many dependencies which have to be installed locally on the user R environment, or already available in the docker environment. The main dependencies are *edgeR*, *limma*, *ggfortify*, *ggplot2*, *topGO*, *UpSetR* and *coseq*.

2.3. Differential expression analysis

After a filtering step where the genes with low counts were excluded from further analyses (i.e. keeping genes with CPM values higher than a threshold for a minimal number of samples). AskoR uses the *calcNormFactors* function of the edgeR library [16], for scaling the data among all libraries and removing the effects of outliers. The default method of this normalization procedure uses a trimmed mean of M-values (TMM) between

each pair of samples, but can be changed by the users to Relative Log expression (RLE) or upperquartile. Next, AskoR applies the Cox-Reid profile-adjusted likelihood method of edgeR for estimating the dispersion then a generalized linear model (GLM) for testing the differential expression, the latter is based on a "quasi likelihood F-test" *qlf*, but can be changed to the likelihood Ratio Test *LRT* with the *glm* parameter.

2.4. Venn and Upset diagrams

For the automatic production of Venn and Upset diagrams with VennDiagram and UpSetR R packages, we use two parameters indicating which gene-list has to be involved in the graphs. The first parameter *compaVD* or *upset_list* (for Venn or Upset diagrams respectively) allows to choose which contrasts will be included in the comparisons (multiple graphs can be produced). The second parameter *VD* or *upset_type* allows to select for each contrast the subset of DEGs to compare (up, down or both).

2.5. GO-term Enrichment

By default, for each gene list, the Gene-Ontology enrichment is evaluated using a Fisher test adjusted by the Benjamini-Hochberg (BH) method with the weight01 algorithm against the complete list of GO assigned genes given by the user. All statistical tests (fisher, ks, t, globaltest, sum or ks.ties) and/or algorithms (classic, elim, weight, weight01, lea or parentchild) supported by TopGO [29] can be selected with the parameters *GO_stats* and *GO_algo* respectively.

2.6. Clustering

The clustering of expression profiles is handled with the coseq R package [22]. However, the methods can be chosen with the *coseq_model* parameter as well as the transformation with *coseq_transformation* and the range of cluster numbers to be evaluated for identifying the best K value *coseq_ClustersNb*.

2.7. Graphics

The graphs allowing the representation of clusters and enrichments are produced with the R package *ggplot2*, *ComplexHeatmap*, and *circlize* allowing the visualization of the intersections of DEG lists between contrasts produced using the *VennDiagram* and *UpSetR* R packages.

2.8. Production of tabular files for AskOmics

In a so-called "AskoTables" directory, AskoR produces files which could be imported directly into AskOmics. Some are directly derived from the input file, describing 3 entities : 1) *Condition* : a group of samples with corresponding characteristics (tissue, stress, development stage, etc...); 2) *Context* : a group of conditions which are compared within a contrast; 3) *Contrast* : the comparison of 2 contexts.

In addition, AskoR provides for each contrast a tabular file, with all the tested genes in lines with the AskOmics compliant columns : Test_id (a unique id for a test), measured_in@Contrast (the name of the contrast), is@gene (the name of the gene), FC (fold-change) and logFC, PValue and FDR (p-value and adjusted p-value of the test), Expression and Significance. Furthermore AskoR supplies a summarized table including all the genes (and annotation) and their significativity at each contrast.

3. Results

AskoR requires only a few tabular files including standard raw counts of reads for each gene, descriptions of samples, including biological and technical replicates, and a list of contrasts between the different treatments and conditions. The structure of the table describing the contrasts is rather simple as it includes a line for each condition and each contrast reported to a column containing "+" or "-" for the condition to be compared and 0 for the others. Additionally, the users can provide a gene ontology assignation file for

performing the GO enrichment step, and a list of complementary annotations which will be transferred to the final tabular files.

To expedite analysis and improve false discovery rate, genes with low CPM values should be removed prior to performing the differential expression analysis, which would be unlikely to be detected as differentially expressed anyhow. The user can adjust several parameters, such as *threshold_cpm* and *replicate_cpm*, and visualize the results in density graphs to determine whether the low expressed genes have been successfully removed.

From the matrix of CPM, AskoR produces Multi-dimensional Scaling (MDS) plot, displaying the coordinates of the samples on 3 axes, a heatmap of the correlation between the samples (with dendograms) and with their respective conditions encoded by a color, and a correlogram (Fig. 1A). This result provides an overview of the samples and allows for the identification of outliers or inconsistencies in biological or technical replicates.

The normalization and differential expression analyses are performed with the popular edgeR package [16], including the functions *calcNormFactors* for the normalization, *estimateGLMCommonDisp*, *estimateGLMTrendedDisp*, *estimateGLMTrendedDisp* or *estimateDisp* for the estimation of the dispersion and *glmFit* and *glmLRT* or *glmQLFit*, and *glmQLFTest* for the DE tests with the Generalized Linear Models (GLM). With a set of default parameters which can be changed by the users, such as the normalization and dispersion methods (*normal_method*, *glm_disp*), GLM (*glm*) or multi-test correction (*p_adj_method*), the AskoR pipeline performs the adequate functions.

For each contrast, a file is created to report the results of the tests in a format readable by AskOmics to facilitate the integration of the results. It also compiles and summarizes all the results of the tests for a batch of contrasts into a single tabular file. It generates as well numerous graphical outputs such as volcano plots, mean-difference plots and heatmaps of top list of the DEGs.

Venn and Upset diagrams allow users to identify common DEGs shared between multiple contrasts. While Venn diagrams allow to compare up to 4 lists, Upset allows users to make comparisons among more constrats (Fig. 1B). However, when making many contrasts, a complete graph including all gene-lists may be unreadable or meaningless, then with only two parameters (*VD* and *compaVD* or *upset_list* and *upset_type*) the user can select precisely the lists of genes to be displayed in the graphs.

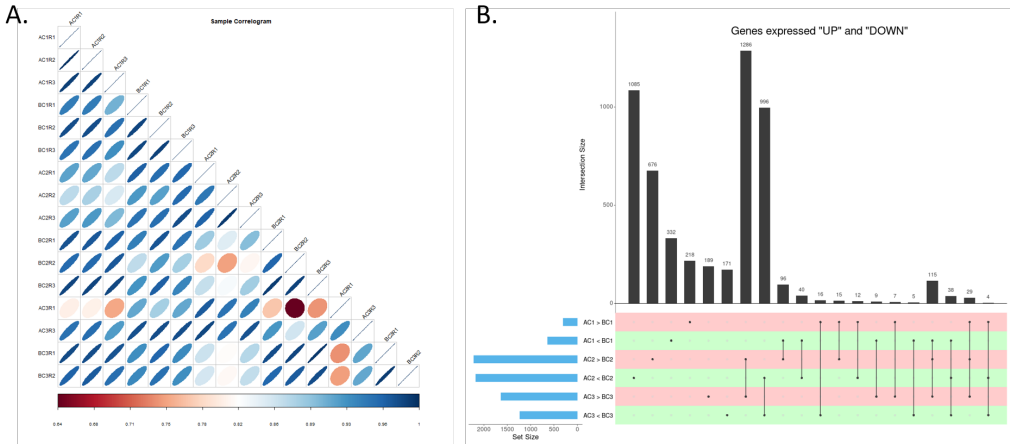

Figure 1. **A**. Correlogram plot generated by AskoR to describe sample correlations. Each cell represents a pairwise comparison and each correlation coefficient is represented by an ellipse whose 'diameter', direction, and color depict the accordance for that pair of samples. Highly correlated samples are depicted as thin blue ellipses, while poorly correlated samples are depicted as red ellipses with wide diameters. **B**. Intersections of DEGs lists between contrasts. DEGs lists in each contrast and their size are shown in the horizontal histograms. The lines connected by dots represent the intersection between these lists. The vertical histograms represent the number of DEGs in each intersection. Colors (green/red) represents UP- and DOWN- regulated DEGs.

If gene ontology terms are provided, AskoR can perform enrichment analysis in lists of DEGs annotated and produce a table (Table 1) with a statistical test from TopGO R package [29] and dedicated figure (Fig. 2).

It is also often useful to group genes that show a similar expression in several conditions (expression profile), to identify co-regulated genes and to characterize genes with no function having similar expression to candidate genes. AskoR performs a clustering with the Coseq R package [22] that uses two classification algorithms (kmeans, and gaussian mixture) with respective transformations, tests for the best number of clusters (K), and produces statistics and graphics helping to check the quality and robustness of the chosen model. Additionally, AskoR outputs several graphs (Fig. 3) to display the expression profiles for each contrast, and search for GO enrichment in each cluster.

Table 1. Gene-Ontology terms for a contrast, each GO term from the 3 categories (molecular function, biological process or cellular component) is reported in the table. The *Annotated* column indicates the number of genes assigned to the term in the complete gene set. The *Significant* column shows the number of genes assigned to the term in the tested list while the *Expected* column gives the expected value, the *Ratio* value corresponds to the *Significant* column divided by *Expected* column, the *statisticTest* displays the adjusted p-value of the test, and *GO_cat* is the GO category of the term ("CC" for cellular component, "BP" for biological process, and "MF" for molecular function).

GO.ID	Term	Annotated	Significant	Expected	statisticTest	Ratio	GO_cat
GO:0022627	cytosolic small ribosomal subunit	21	13	4.75	0.00012	2.73684210526316	CC
GO:0004812	aminoacyl-tRNA ligase activity	51	24	11.8	0.00014	2.03389830508475	MF
GO:0005840	ribosome	193	79	43.63	0.00016	1.81068072427229	CC
GO:0042273	ribosomal large subunit biogenesis	28	21	6.29	0.00017	3.33863275039746	BP
GO:0030687	preribosome, large subunit precursor	10	8	2.26	0.00019	3.53982300884956	CC
GO:0000460	maturation of 5.8S rRNA	10	8	2.25	0.00019	3.55555555555556	BP
GO:0005929	cilium	89	36	20.12	0.00024	1.78926441351889	CC

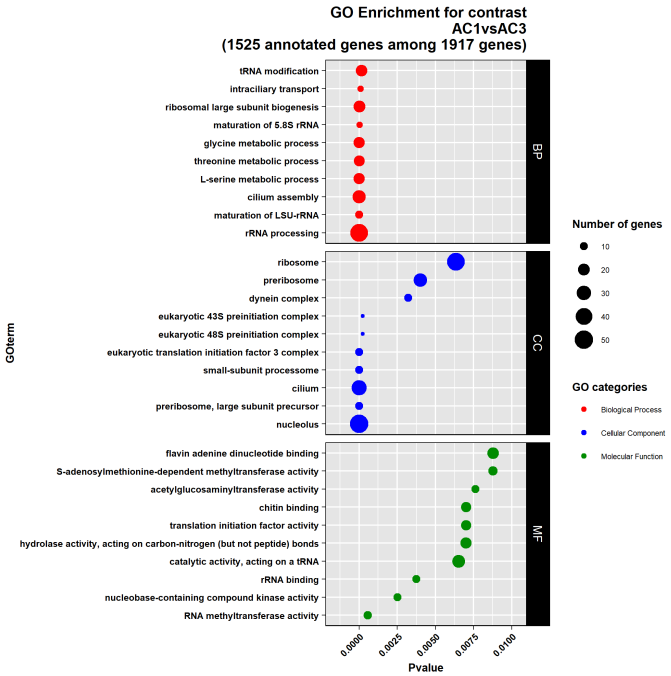

Figure 2. Plot of GO enrichment for one contrast. Each line displays a significant term grouped by GO category, the position of the dot refers to the significance of the test and the size of the circle corresponds to the number of observed genes in the tested list which are assigned to that terms.

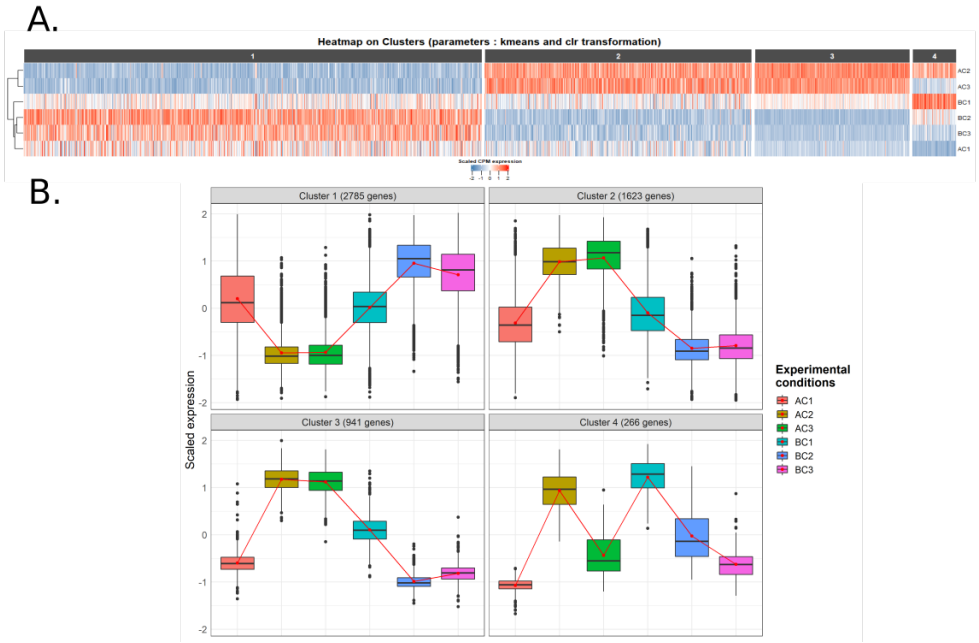

Figure 3. Gene clustering based on expression profiles. **A**. Heatmap of scaled expression. Each line represents a condition and each column a gene. The color indicates the scaled expression. The genes are grouped in clusters. The condition are ordered by the user or grouped by correlation (with a dendogram). **B**. Boxplots of scaled expression by clusters

4. Discussion

AskoR is highly customizable, and includes many parameters with default values modifiable according to the needs of the analysis. Thus, it is straightforward to run an analysis with a single file and a set of data. The results of the analyses are then stored in multiple files under a named directory. Then, any set of results obtained with a particular set of parameters can be stored, compared and reproduced.

AskoR is working as a standalone library and no web interface has been developed yet (with R Shiny for instance). Nevertheless, it can be run as a single script and easily modified directly on any laptop or server or integrate directly in workflows for an improved reproducibility. Furthermore, the docker image of AskoR can be integrated directly as an interactive environment in a Galaxy instance. As a result, with the help of the online documentation, it is straighforward to explore transcriptomics data and produce publishable figures. Consequently, AskoR has already been used for various analysis [12,30–32].

Additionally, AskoR prepares output files following the AskOmics requirements. Within this tool, complex and expressive queries can be generated via a graphical interface in order to extract the most interesting genes by combining several lists of genes by a logical operator (e.g. DEGs from one or more particular contrast or cluster). For instance, it is possible to extract genes over-expressed in a condition at some selected time-points but or under-expressed (or not differentially expressed) at other time points. Furthermore, DEGs can be supplemented with other data such as genome locations, assignation to metabolic networks, miRNA targets, or epigenomics landscapes to promote selection of candidate genes.

AskoR is open-source available online at a git repository, where anyone can contribute, push an issue or ask for a specific request. For example, we are adding functions to create graphical representation combining the expression levels and the of a set of genes list defined by a user.

Author Contributions: Conceptualization, S.D. and F.L.; software development S.A-C, K.G., S.M. and F.L.; application, tests and validation, S.A-C, K.G. and S.R., deployment, A.B.

Funding: The internship of Sylvin Masanelli has been granted by the Santé des Plantes et Environnement (SPE) department of INRAE

Data Availability Statement: AskoR is available at https://github.com/askomics/askoR the docker image for being deploying Askomics in a Galaxy instance is available at https://github.com/genouest/docker-galaxy-rstudio-askor

Acknowledgments: We are thankful to the GenOuest platform for the Galaxy environment, to Emmanuelle Becker and Fanny Casse for clustering and enrichment graphs, to Andréa Rau for the help while implanting the coseq package and to Régine Delourme, Jean-Philippe Vernadet, Jérémy Gauthier, Florence Prunier, Ambre-Aurore Josselin, Mélanie Huguet for testing and helping us to design the tool, and to Olivier Dameron and Anne Siegel for the discussion around Askomics

Conflicts of Interest: The authors declare no conflict of interest

References

1. Legeai, F.; Shigenobu, S.; Gauthier, J.P.; Colbourne, J.; Rispe, C.; Collin, O.; Richards, S.; Wilson, A.C.C.; Murphy, T.; Tagu, D. AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect Molecular Biology* **2010**, *19*, 5–12. doi:10.1111/j.1365-2583.2009.00930.x.
2. Rispe, C.; Legeai, F.; Nabity, P.D.; Fernández, R.; Arora, A.K.; Baa-Puyoulet, P.; Banfill, C.R.; Bao, L.; Barberà, M.; Bouallègue, M.; Bretaudeau, A.; Brisson, J.A.; Calevro, F.; Capy, P.; Catrice, O.; Chertemps, T.; Couture, C.; Delière, L.; Douglas, A.E.; Dufault-Thompson, K.; Escuer, P.; Feng, H.; Forneck, A.; Gabaldón, T.; Guigó, R.; Hilliou, F.; Hinojosa-Alvarez, S.; min Hsiao, Y.; Hudaverdian, S.; Jacquin-Joly, E.; James, E.B.; Johnston, S.; Joubard, B.; Goff, G.L.; Trionnaire, G.L.; Librado, P.; Liu, S.; Lombaert, E.; ling Lu, H.; Maïbèche, M.; Makni, M.; Marcet-Houben, M.; Martínez-Torres, D.; Meslin, C.; Montagné, N.; Moran, N.A.; Papura, D.; Parisot, N.; Rahbé, Y.; Lopes, M.R.; Ripoll-Cladellas, A.; Robin, S.; Roques, C.; Roux, P.; Rozas, J.; Sánchez-Gracia, A.; Sánchez-Herrero, J.F.; Santesmasses, D.; Scatoni, I.; Serre, R.F.; Tang, M.; Tian, W.; Umina, P.A.; van Munster, M.; Vincent-Monégat, C.; Wemmer, J.; Wilson, A.C.C.; Zhang, Y.; Zhao, C.; Zhao, J.; Zhao, S.; Zhou, X.; Delmotte, F.; Tagu, D. The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biology* **2020**, *18*. doi:10.1186/s12915-020-00820-5.
3. Legeai, F.; Santos, B.F.; Robin, S.; Bretaudeau, A.; Dikow, R.B.; Lemaitre, C.; Jouan, V.; Ravallec, M.; Drezen, J.M.; Tagu, D.; Baudat, F.; Gyapay, G.; Zhou, X.; Liu, S.; Webb, B.A.; Brady, S.G.; Volkoff, A.N. Genomic architecture of endogenous ichnoviruses reveals distinct evolutionary pathways leading to virus domestication in parasitic wasps. *BMC Biology* **2020**, *18*. doi:10.1186/s12915-020-00822-3.
4. Gauthier, J.; Boulain, H.; van Vugt, J.J.F.A.; Baudry, L.; Persyn, E.; Aury, J.M.; Noel, B.; Bretaudeau, A.; Legeai, F.; Warris, S.; Chebbi, M.A.; Dubreuil, G.; Duvic, B.; Kremer, N.; Gayral, P.; Musset, K.; Josse, T.; Bigot, D.; Bressac, C.; Moreau, S.; Periquet, G.; Harry, M.; Montagné, N.; Boulogne, I.; Sabeti-Azad, M.; Maïbèche, M.; Chertemps, T.; Hilliou, F.; Siaussat, D.; Amselem, J.; Luyten, I.; Capdevielle-Dulac, C.; Labadie, K.; Merlin, B.L.; Barbe, V.; de Boer, J.G.; Marbouty, M.; Cônsoli, F.L.; Dupas, S.; Hua-Van, A.; Goff, G.L.; Bézier, A.; Jacquin-Joly, E.; Whitfield, J.B.; Vet, L.E.M.; Smid, H.M.; Kaiser, L.; Koszul, R.; Huguet, E.; Herniou, E.A.; Drezen, J.M. Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. *Communications Biology* **2021**, *4*. doi:10.1038/s42003-020-01623-8.
5. Dennis, A.B.; Ballesteros, G.I.; Robin, S.; Schrader, L.; Bast, J.; Berghöfer, J.; Beukeboom, L.W.; Belghazi, M.; Bretaudeau, A.; Buellesbach, J.; Cash, E.; Colinet, D.; Dumas, Z.; Errbii, M.; Falabella, P.; Gatti, J.L.; Geuverink, E.; Gibson, J.D.; Hertaeg, C.; Hartmann, S.; Jacquin-Joly, E.; Lammers, M.; Lavandero, B.I.; Lindenbaum, I.; Massardier-Galata, L.; Meslin, C.; Montagné, N.; Pak, N.; Poirié, M.; Salvia, R.; Smith, C.R.; Tagu, D.; Tares, S.; Vogel, H.; Schwander, T.; Simon, J.C.; Figueroa, C.C.; Vorburger, C.; Legeai, F.; Gadau, J. Functional insights from the GC-poor genomes of two aphid parasitoids, Aphidius ervi and Lysiphlebus fabarum. *BMC Genomics* **2020**, *21*. doi:10.1186/s12864-020-6764-0.
6. Gimenez, S.; Abdelgaffar, H.; Goff, G.L.; Hilliou, F.; Blanco, C.A.; Hänniger, S.; Bretaudeau, A.; Legeai, F.; Nègre, N.; Jurat-Fuentes, J.L.; d'Alençon, E.; Nam, K. Adaptation by copy number variation increases insecticide resistance in the fall armyworm. *Communications Biology* **2020**, *3*. doi:10.1038/s42003-020-01382-6.
7. Afgan, E.; Baker, D.; Batut, B.; van den Beek, M.; Bouvier, D.; Čech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grüning, B.A.; Guerler, A.; Hillman-Jackson, J.; Hiltemann, S.; Jalili, V.; Rasche, H.; Soranzo, N.; Goecks, J.; Taylor, J.; Nekrutenko, A.; Blankenberg, D. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* **2018**, *46*, W537–W544. doi:10.1093/nar/gky379.
8. Dunn, N.A.; Unni, D.R.; Diesh, C.; Munoz-Torres, M.; Harris, N.L.; Yao, E.; Rasche, H.; Holmes, I.H.; Elsik, C.G.; Lewis, S.E. Apollo: Democratizing genome annotation. *PLOS Computational Biology* **2019**, *15*, e1006790. doi:10.1371/journal.pcbi.1006790.
9. Boulain, H.; Legeai, F.; Jaquiéry, J.; Guy, E.; Morlière, S.; Simon, J.C.; Sugio, A. Differential Expression of Candidate Salivary Effector Genes in Pea Aphid Biotypes With Distinct Host Plant Specificity. *Frontiers in Plant Science* **2019**, *10*. doi:10.3389/fpls.2019.01301.

10. Moné, Y.; Nhim, S.; Gimenez, S.; Legeai, F.; Seninet, I.; Parrinello, H.; Nègre, N.; d'Alençon, E. Characterization and expression profiling of microRNAs in response to plant feeding in two host-plant strains of the lepidopteran pest Spodoptera frugiperda. *BMC Genomics* **2018**, *19*. doi:10.1186/s12864-018-5119-6.
11. Tetreau, G.; Dhinaut, J.; Galinier, R.; Audant-Lacour, P.; Voisin, S.N.; Arafah, K.; Chogne, M.; Hilliou, F.; Bordes, A.; Sabarly, C.; Chan, P.; Walet-Balieu, M.L.; Vaudry, D.; Duval, D.; Bulet, P.; Coustau, C.; Moret, Y.; Gourbal, B. Deciphering the molecular mechanisms of mother-to-egg immune protection in the mealworm beetle Tenebrio molitor. *PLOS Pathogens* **2020**, *16*, e1008935. doi:10.1371/journal.ppat.1008935.
12. Lorenzi, A.; Ravallec, M.; Eychenne, M.; Jouan, V.; Robin, S.; Darboux, I.; Legeai, F.; Gosselin-Grenet, A.S.; Sicard, M.; Stoltz, D.; Volkoff, A.N. RNA interference identifies domesticated viral genes involved in assembly and trafficking of virus-derived particles in ichneumonid wasps. *PLOS Pathogens* **2019**, *15*, e1008210. doi:10.1371/journal.ppat.1008210.
13. Rondoni, G.; Roman, A.; Meslin, C.; Montagné, N.; Conti, E.; Jacquin-Joly, E. Antennal Transcriptome Analysis and Identification of Candidate Chemosensory Genes of the Harlequin Ladybird Beetle, Harmonia axyridis (Pallas) (Coleoptera: Coccinellidae). *Insects* **2021**, *12*, 209. doi:10.3390/insects12030209.
14. Meslin, C.; Bozzolan, F.; Braman, V.; Chardonnet, S.; Pionneau, C.; François, M.C.; Severac, D.; Gadenne, C.; Anton, S.; Maibèche, M.; Jacquin-Joly, E.; Siaussat, D. Sublethal Exposure Effects of the Neonicotinoid Clothianidin Strongly Modify the Brain Transcriptome and Proteome in the Male Moth Agrotis ipsilon. *Insects* **2021**, *12*, 152. doi:10.3390/insects12020152.
15. Gonzalez, F.; Johny, J.; Walker, W.B.; Guan, Q.; Mfarrej, S.; Jakše, J.; Montagné, N.; Jacquin-Joly, E.; Alqarni, A.A.; Al-Saleh, M.A.; Pain, A.; Antony, B. Antennal transcriptome sequencing and identification of candidate chemoreceptor proteins from an invasive pest, the American palm weevil, Rhynchophorus palmarum. *Scientific Reports* **2021**, *11*. doi:10.1038/s41598-021-87348-y.
16. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2009**, *26*, 139–140. doi:10.1093/bioinformatics/btp616.
17. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **2014**, *15*. doi:10.1186/s13059-014-0550-8.
18. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szcześniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.; Mortazavi, A. A survey of best practices for RNA-seq data analysis. *Genome Biology* **2016**, *17*. doi:10.1186/s13059-016-0881-8.
19. Hedegaard, J.; Arce, C.; Bicciato, S.; Bonnet, A.; Buitenhuis, B.; Collado-Romero, M.; Conley, L.N.; SanCristobal, M.; Ferrari, F.; Garrido, J.J.; Groenen, M.A.; Hornshøj, H.; Hulsegge, I.; Jiang, L.; Jiménez-Marín, Á.; Kommadath, A.; Lagarrigue, S.; Leunissen, J.A.; Liaubet, L.; Neerincx, P.B.; Nie, H.; van der Poel, J.; Prickett, D.; Ramirez-Boo, M.; Rebel, J.M.; Robert-Granié, C.; Skarman, A.; Smits, M.A.; Sørensen, P.; Tosser-Klopp, G.; Watson, M. Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. *BMC Proceedings* **2009**, *3*. doi:10.1186/1753-6561-3-s4-s5.
20. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; Mesirov, J.P. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **2005**, *102*, 15545–15550, [https://www.pnas.org/content/102/43/15545.full.pdf]. doi:10.1073/pnas.0506580102.
21. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33*, 2938–2940. doi:10.1093/bioinformatics/btx364.
22. Godichon-Baggioni, A.; Maugis-Rabusseau, C.; Rau, A. Clustering transformed compositional data using K-means, with applications in gene expression and bicycle sharing system data. *Journal of Applied Statistics* **2018**, *46*, 47–65. doi:10.1080/02664763.2018.1454894.
23. Vidman, L.; Källberg, D.; Rydén, P. Cluster analysis on high dimensional RNA-seq data with applications to cancer research - An evaluation study. *PLOS ONE* **2019**, *14*, e0219102. doi:10.1371/journal.pone.0219102.
24. Bioinformatics, B. https://www.biobam.com/omicsboxOmicsBox – Bioinformatics Made Easy, 2019.
25. QIAGEN. https://digitalinsights.qiagen.com/QIAGEN CLC Genomics Workbench 20.0.
26. Ewels, P.A.; Peltzer, A.; Fillinger, S.; Patel, H.; Alneberg, J.; Wilm, A.; Garcia, M.U.; Tommaso, P.D.; Nahnsen, S. The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* **2020**, *38*, 276–278. doi:10.1038/s41587-020-0439-x.
27. AskOmics. https://askomics.org/.
28. Garnier, X.; Bretaudeau, A.; Legeai, F.; Siegel, A.; Dameron, O. a user-friendly interface to Semantic Web technologies for integrating local datasets with reference resources. *Proceedings of JOBIM 2019* **2019**.
29. Alexa, A.; Rahnenfuhrer, J. *topGO: Enrichment Analysis for Gene Ontology*, 2020. R package version 2.40.0.
30. Gazengel, K.; Aigu, Y.; Lariagon, C.; Humeau, M.; Gravot, A.; Manzanares-Dauleux, M.J.; Daval, S. Nitrogen supply and host-plant genotype modulate the transcriptomic profile of *Plasmodiophora brassicae*. *Frontiers in Microbiology (in press)*.
31. Prunier, F.; Persyn, E.; Legeai, F.; McClure, M.; Meslin, C.; Robin, S.; Alves-Carvalho, S.; Mohammad, A.; Blugeon, C.; Jacquin-Joly, E.; Montagné, N.; Elias, M.; Gauthier, J. Comparative transcriptome analysis at the onset of speciation in a mimetic butterfly, the Ithomiini Melinaea marsaeus. *Journal of Evolutionary Biology (submitted)*.
32. Daval, S.; Gazengel, K.; Belcour, A.; Linglin, J.; Guillerm-Erckelboudt, A.; Sarniguet, A.; Manzanares-Dauleux, M.J.; Lebreton, L.; Mougel, C. Soil microbiota influences clubroot disease by modulating *Plasmodiophora brassicae* and *Brassica napus* transcriptomes. *Microbial Biotechnology* **2020**, *13*, 1648–1672. doi:10.1111/1751-7915.13634.