EMBnet.journal
Bioinformatics in Action

Ds-Seq: an integrated pipeline for in silico small RNA sequence analysis for host-pathogen interaction studies

Fingerprinting Breast Milk; insights into Milk Exosomics

Exosomal Epigenetics

AND MORE...

29
2024

EMBnet.journal
Bioinformatics in Action

Contents

Editorial 2

Reports

EpiCass and CassavaNet4Dev Advanced Bioinformatics Workshop
Andreas Gisel1,2, Livia Stavolone3, Temitayo Olagunju1, Michael Landi4, Renaud van Damme4, Adnan Niazi4, Laurent Falquet5, Trushar Shah1, Erik Bongcam-Rudloff 3

Reviews

The effect of the planned behaviour theory and the transtheoretical behaviour model on physical activity. A systematic review.
Anastasia Dermatis, Flora Bacopoulou, Ioulia Kokka, Dimitrios Vlachakis, Georgios Lyrakos, Despina Menti, Christina Darviri 9

Research Papers

Ds-Seq: an integrated pipeline for in silico small RNA sequence analysis for host-pathogen interaction studies
Temitayo Adebanji Olagunju, Angela Uche Makolo, Andreas Gisel 15

Fingerprinting Breast Milk; insights into Milk Exosomics
Eleni Papakonstantinou, Konstantina Dragoumani, Antonia Mataragka, Flora Bacopoulou, Christos Yapijakis, Nikolaos AA Balatsos, Katerina Pissaridi, Dimitris Ladikos, Aspasia Efthymiadou, George Katsaros, Evangelos Gikas, Pantelis Hatzis, Martina Samiotaki, Michalis Aivaliotis, Vasileios Megalooikonomou, Antonis Giannakakis, Costas Iliopoulos, Erik Bongcam-Rudloff, Sofia Kossida, Elias Eliopoulos, George P Chrousos, Dimitrios Vlachakis 25

Exosomal Epigenetics
Eleni Papakonstantinou, Konstantina Dragoumani, George P Chrousos, Dimitrios Vlachakis 31

Milk exosomes and a new way of communication between mother and child
Eleni Papakonstantinou, Konstantina Dragoumani, Thanasis Mitsis, George P Chrousos, Dimitrios Vlachakis 36

Protein Spotlight

Protein Spotlight 254........ 41

Protein Spotlight 258........ 43

Protein Spotlight 262........ 45

Protein Spotlight 264........ 47

Editorial

Issue 29 of EMBnet.journal showcases the dynamic and impactful work being done in the field of bioinformatics and computational biology. This edition highlight the importance of international collaboration, cutting-edge research, and the application of advanced technologies to address pressing global health challenges.

One of the key highlights is the "EpiCass and CassavaNet4Dev Advanced Bioinformatics Workshop," article which emphasizes the role of bioinformatics education in agricultural development and food security. The workshop brought together experts from various regions to share knowledge and enhance skills in next-generation sequencing (NGS) and data analysis, essential for improving crop resilience and combating diseases.

Another significant contribution is the research on "Ds-Seq: an integrated pipeline for in silico small RNA sequence analysis for host-pathogen interaction studies in plants." This innovative tool represents a major step forward in understanding the molecular interactions between hosts and pathogens, which is crucial for developing new breeding strategies for various crops essential to food production.

The review article on the application of the planned behaviour theory and the transtheoretical model in promoting physical activity provides a comprehensive analysis of behavioural interventions. This research is particularly relevant in the context of public health, as it offers insights into designing effective programs to encourage healthier lifestyles.

The articles in Issue 29 collectively underscore the critical role of bioinformatics in addressing a wide range of global challenges, from enhancing public health and agricultural productivity to advancing education. They also demonstrate the power of collaboration across borders, bringing together expertise and resources to tackle complex problems.

As we continue to face global health crises, such as the rise of multidrug-resistant bacteria and the impacts of climate change on disease dynamics, the contributions of bioinformatics become increasingly indispensable. This issue of EMBnet.journal not only celebrates scientific achievements but also calls for continued cooperation and innovation in the field.

By leveraging the collective knowledge and technological advancements presented in this issue, we can make significant strides in improving global health outcomes and building a more resilient future. We also encourage more researchers to submit their work for the upcoming Issue 30 of the journal. With an impact factor of 3 over the last four years, we aim to achieve an even higher score in future issues.

Erik Bongcam-Rudloff
Editor-in-Chief
erik.bongcam@slu.se

http://dx.doi.org/10.14806/ej.29.0.1058

Cover picture: Vinbärsfuks, Currant furge (*Polygonia c-album*), Engsö slott, Västmanland, Sweden, 2022 © Erik Bongcam-Rudloff

EMBnet.journal
Bioinformatics in Action

Reports

EpiCass and CassavaNet4Dev Advanced Bioinformatics Workshop

Andreas Gisel[1,2], Livia Stavolone[3], Temitayo Olagunju[1], Michael Landi[4], Renaud van Damme[4], Adnan Niazi[4], Laurent Falquet[5], Trushar Shah[1], Erik Bongcam-Rudloff[4]✉

[1] International Institute of Tropical Agriculture, Ibadan, Nigeria
[2] Institute for Biomedical Technologies, National Research Council, Bari, Italy
[3] Institute for Sustainable Plant Protection, National Research Council, Bari, Italy
[4] Swedish University of Agricultural Sciences, Uppsala, Sweden
[5] University of Fribourg, Fribourg, Switzerland

Competing interests: AG none; LS none; TO none; ML none; RvD none; AN none; LF none; TS none; EBR none

Abstract

EpiCass and CassavaNet4Dev are collaborative projects funded by the Swedish Research Council between the Swedish University of Agriculture (SLU) and the International Institute of Tropical Agriculture (IITA). The projects aim to investigate the influence of epigenetic changes on agricultural traits such as yield and virus resistance while also providing African students and researchers with advanced bioinformatics training and opportunities to participate in big data analysis events. The first advanced bioinformatics training workshop took place from May 16th to May 18th, 2022, followed by an online mini-symposium titled "Epigenetics and crop improvement" on May 19th. The symposium featured international speakers covering a wide range of topics related to plant epigenetics, cassava viral diseases, and cassava breeding strategies. A new online and on-site teaching model was developed for the three-day workshop to ensure maximum student participation across Western, Eastern, and Southern Africa. Initially planned in Nigeria, Kenya, Ethiopia, Tanzania, and Zambia, the workshop ultimately focused on Nigeria, Kenya, and Ethiopia due to a lack of qualified candidates in the other countries. Each classroom hosted 20 to 25 students, with at least one bioinformatician present for support. The classrooms were connected via video conferencing, whereas teachers located in different places in Africa and Europe joined the video stream to conduct teaching sessions. The workshop was divided into theoretical classes and hands-on sessions, where participants could run data analysis with support from online teachers and local bioinformaticians. To enable participants to run guided, CPU and RAM-intensive data analysis workflows and overcome local computing and internet access restrictions, a system of virtual machines (VMs) hosted in the cloud was developed. The teaching platform provided teaching and exercise materials to support the use of the VMs. Some students could not run heavy data analysis workflows due to unforeseen restrictions in the cloud. Currently, these issues have been solved and in the future all participants will have the opportunity to run the analysis steps independently in the cloud using the protocols hosted on the teaching platform.

Introduction

EpiCass and CassavaNet4Dev are two projects that aim to investigate the epigenetics of cassava (Manihot esculenta) and its influence on agronomic traits. They also aim to disseminate the data and information generated and train African researchers in high-throughput data analysis. EpiCass focuses primarily on research, while CassavaNet4Dev focuses on dissemination, training, and networking within the African plant science and breeding community.

EpiCass is divided into two research lines. The first line examines the DNA methylation profiles of four cassava genotypes and correlates these profiles with the highly variable storage root yield. The second line investigates the DNA methylation profile of two genetically similar cassava genotypes with different resistance against the African Cassava Mosaic Virus. In both cases, the goal is to extract and analyse sequencing data to produce knowledge that can improve these traits or better monitor them during breeding processes. EpiCass also proposes a workshop on next-generation sequencing (NGS) data analysis for African researchers

Article history
Received: 12 April 2023
Published: 08 June 2023

© 2023 Gisel *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at http://journal.embnet.org.

to gain a basic understanding of sequencing technologies and introduce them to epigenetics and its potential in crop improvement.

On the other hand, CassavaNet4Dev aims to develop a network of African scientists and students interested in learning NGS data analysis and joining a network of plant scientists and breeders. This network plans to include selected participants to collaborate on the data analysis of the EpiCass project, which intends to generate raw data of about 36 GB long-read re-sequencing data, 2.7 TB DNA methylation profiling data, and 130 GB small non-coding RNA data.

The two projects collaborated to organise the first bioinformatics training workshop, which took place from 16-18 May 2022, with a mini-symposium on 19 May providing an overview of epigenetics in plants and possible applications for crop improvement. In the following paragraphs, we describe in more detail the organisation, execution, problems, and reviews of this workshop.

Workshop set-up

With our extensive experience in physical and online bioinformatics training, we have developed a hybrid workshop model combining the benefits of physical interaction with the ability to involve participants and teachers over a larger geographical area, thereby reducing travel and accommodation costs for both participants and organisers. The idea of such a workshop model was developed based on our observation that it is challenging to follow participants with different experience levels in online courses, especially during hands-on sessions. We also wanted to allow scientists and students from various African regions to participate in our workshop without incurring the high costs of long-distance travel.

The EpiCass project covers Nigeria and Kenya in Eastern and Western Africa, respectively. Additionally, the International Institute of Tropical Agriculture (IITA) has essential stations in Tanzania and Zambia in Southern Africa and significant collaborations in Ethiopia, such as the Bio & Emerging Technology Institute (BETin) in Addis Ababa. We planned physical classes of up to 20 participants at IITA stations to ensure a certain infrastructure level, including a good internet connection and a bioinformatician in each classroom to follow the group. However, due to limited funds, only local students or those with their own finances could participate. Fortunately, BETin offered to host the classroom and sponsor the event in Addis Ababa and allowed Ethiopian students from other universities outside Addis Ababa to travel to BETin.

The workshop classrooms were connected through a video conference system (in our case, Zoom) to allow for comments and questions in each classroom, with additional teachers from Africa and Europe joining the live stream to teach and follow the practical hands-on sessions. It's worth mentioning that the three classrooms were in two different time zones, with Ibadan in WAT (UTC +1), Nairobi and Addis Ababa in EAT (UTC+3), and the teacher from Europe in CEST (UTC+2). When the classes started at 8 am in Ibadan, it was 9 am in Europe and 10 am in Nairobi and Addis Ababa. This posed a logistical challenge, especially during the breaks. Another important element of the workshop was the development of documented classes on a teaching platform. In our case, we used the open-access platform Moodle, where each teacher hosted presentations and detailed exercises for the hands-on sessions. This platform enabled participants to preview future classes, follow presentations on their laptops, run hands-on sessions at their own pace, and run hands-on sessions as many times as needed to understand the different data analysis steps. Each participant was registered in Moodle, hosted at SLU in Uppsala, and had full access to the teaching material and communication channels.

The programme for the 3-day workshop introduced participants to NGS data analysis, with one full day dedicated to learning and using the operation system (OS) LINUX. This was necessary because many African scientists know little or nothing about LINUX. The morning was dedicated to introducing LINUX and explaining the most important commands to enable participants to start with NGS data analysis. In the afternoon, each participant was connected to an external LINUX server to run all the exercises and become more familiar with the OS.

The second day introduced the concept of NGS and the first steps in the NGS data analysis workflow, including quality control of raw reads, cleaning raw reads, and mapping them against a reference sequence. In the afternoon, participants were exposed to data analysis running tools for quality control, filtering, and mapping.

On the third day, we gave participants an introduction to epigenetics, the main topic of the EpiCass project, followed by classes on NGS assembly and hands-on sessions in the afternoon.

Selection of participants

To ensure a successful workshop, selecting motivated participants who are engaged in the research work and at a similar level is crucial. As the workshop focuses on NGS data analysis, we set basic bioinformatic knowledge

Figure 1. Percentage of LINUX knowledge among the workshop's applicants.

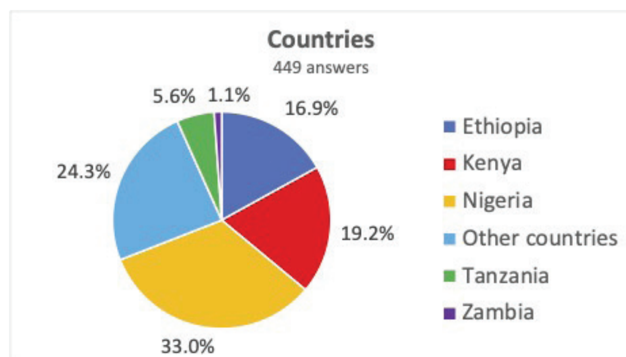


Figure 2. Distribution of the applicants by country

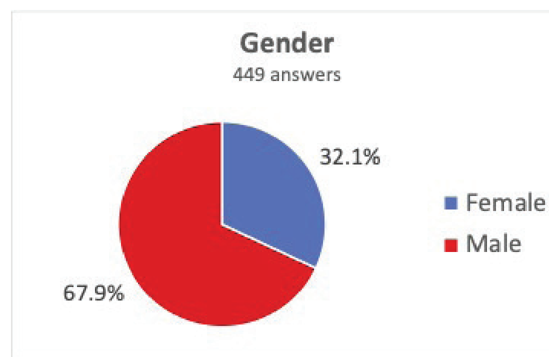


Figure 3. Distribution of gender and level of education of the applicants.

as a selection criterion for access to the workshop. We developed a Google Form registration module that collects personal information and includes questions about bioinformatics and LINUX to measure the level of the applicants.

The registration form was distributed within the IITA network in Africa and to collaborating universities in Nigeria, Ethiopia, Kenya, Tanzania, and Zambia. We received 449 registrations, mainly from Nigeria, Kenya, and Ethiopia (Figure 2). Interestingly, 25% of the registrations came from countries other than the selected ones, including Egypt, Sudan, Uganda, Tunisia, and countries outside Africa such as India and China. Most of the applicants were male researchers (almost 70%), and half of the applicants were master's students (Figure 3).

As mentioned earlier, we were looking for applicants with basic bioinformatics knowledge. To this aim, we included five questions about basic bioinformatics and five questions to evaluate LINUX experience in the registration form. Each question was worth one point, and a maximum of ten points were possible. Figure 4 shows that most applicants had zero points, indicating no experience in either bioinformatics or LINUX. Only 25% (114) of the applicants had at least one point, which indicates an urgent need for bioinformatics education in Africa. Many applicants reported being involved in analysing biological data without proper training, highlighting the need for bioinformatics training in Africa.

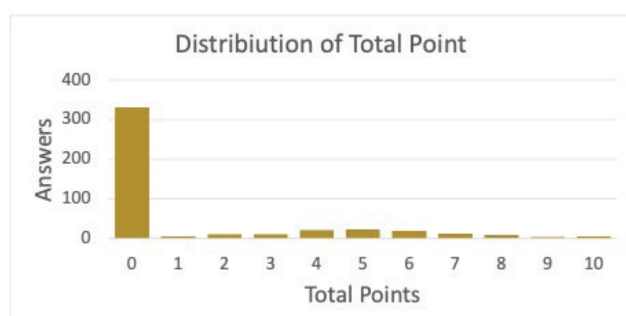


Figure 4. Distribution of the total points of all applicants for the quiz with five questions about basic bioinformatics and five questions about LINUX.

Unfortunately, we did not receive enough applications from Zambia, and the applications from Tanzania were not qualified enough. As a result, we focused our workshop on Nigeria (Ibadan - IITA), Kenya (Nairobi - IITA), and Ethiopia (Addis Ababa – BETin). Each local team selected the final participants based on their bioinformatics and LINUX experience and whether they had a running project requiring data analysis. We invited 21 participants from Ethiopia, 22 from Nigeria, and 24 from Kenya and hosted 29 (4 female and 25 male), 19 (10 female and nine male), and 18 (nine female and nine male), respectively.

As previously mentioned, we had hands-on exercises during the afternoons, and each student had access to a LINUX server to execute the data analysis. In the following paragraph, we will describe how we tackled the challenge of hosting about 60 students for high RAM and CPU data analysis.

Set-up of the computing environment

NGS data analysis, especially NGS assembly, is a computationally intensive task that requires high RAM and CPU usage. Laptops and personal computers are inadequate for such applications, and it has become increasingly difficult to host external participants for such workshops on institutional servers due to security measures. To address these issues, we developed a computing strategy using cloud computing. We have extensive experience with the Ubuntu 18.04 LINUX operating system and have set up an Ubuntu 18.04 virtual machine (VM) with all the necessary software (see Table 1) for the 3-day workshop. We use Amazon Web Services (AWS) via its graphical web-based interface and Command Line Interface (CLI) to host the VM. We support various VM file formats, such as Open Virtualization Archive (OVA), Virtual Machine Disk (VMDK), Virtual Hard Disk (VHD/VHDX), and use the free VM software VirtualBox to save the VM in one of those formats.

When you register for an AWS account, you automatically set up an account with complete access to all AWS services and resources in the account. This account is called the AWS account root user, and it

EMBnet.journal
Bioinformatics in Action

Gisel et al. (2023) EMBnet.journal 29, e1045
http://dx.doi.org/10.14806/ej.29.0.1045

Figure 5. Classrooms at the three locations.

is accessed by signing in with the email address and password that you used to create the account. AWS strongly recommends not using the root account for your everyday tasks. Instead, it is best to create an Identity and Access Management (IAM) user, to which you grant specific permissions to administer and use services in the AWS account. Therefore, you should never expose the password or access key to collaborators. IAM accesses the AWS services with an account ID (created by AWS), a username, and a password, all created by the root user. It is possible to set up different IAMs for different tasks. For our project, we created an IAM that can create storage buckets, upload data from external resources into these buckets, create Amazon Machine Images (AMI), and start and stop instances. After defining the credentials and permissions for the IAM, you can upload the above-mentioned VM into the Amazon storage bucket with a single CLI command from your terminal. While creating the storage bucket, you must decide in which geographical region to create it. You can choose from US, Canada, Europe, Africa, the Middle East, South America, and Asia. The location of your storage is important since you will have to use the cloud computing in the area where you have your storage. Depending on the location, computing prices and available instance types may vary. After you have uploaded the VM into the storage bucket, you need to import it into the computing environment EC2 (Amazon Elastic Computing Cloud), where it will be converted and stored as an Amazon Machine Image (AMI) from where it can be launched at any time, creating an instance. During the launch, you can select your AMI from "My AMIs," describe the instance, choose the instance type and needed storage size, and define who will have access to the running instance. If you want a public IP accessible to anybody, you can choose "Anywhere" under the "network settings" "Allow SSH traffic from." Otherwise, you can select specific IP addresses to limit access and increase security.

We chose to use the EC2 environment because we needed computing power quickly and wanted the freedom to select specific instance types (*i.e.* number of CPUs and amount of RAM) based on the applications we planned to run. For the first hands-on exercise on the afternoon of the first day, during which we used various LINUX commands, we required only a small amount of computing power. Therefore, we started an instance with 2 CPUs and 4 GiB RAM (using Amazon instance type t2.medium) for each classroom. When an instance is started, it is assigned a temporary public IP address which participants use to connect to the virtual machine using the Secure Shell Protocol (SSH). Each time an instance is started, AWS assigns a new public IP address, which will most likely be different from the previous one. If you need a fixed IP address each time you start the same instance, you will need to set up an elastic IP address, which comes at a cost but might be worthwhile in certain situations. All 66 participants were able to connect to one of the three virtual machines and successfully complete the LINUX exercise.

For the hands-on exercise on the second day, in which we planned to perform quality control, filtering, and mapping of a real NGS dataset, we intended to provide each participant with their own virtual machine with 4 CPUs and 16 GiB RAM to obtain results in a reasonable amount of time. Unfortunately, we encountered a limit that we did not experiment with during our tests and studies of AWS documentation. The available CPUs, without any additional request,

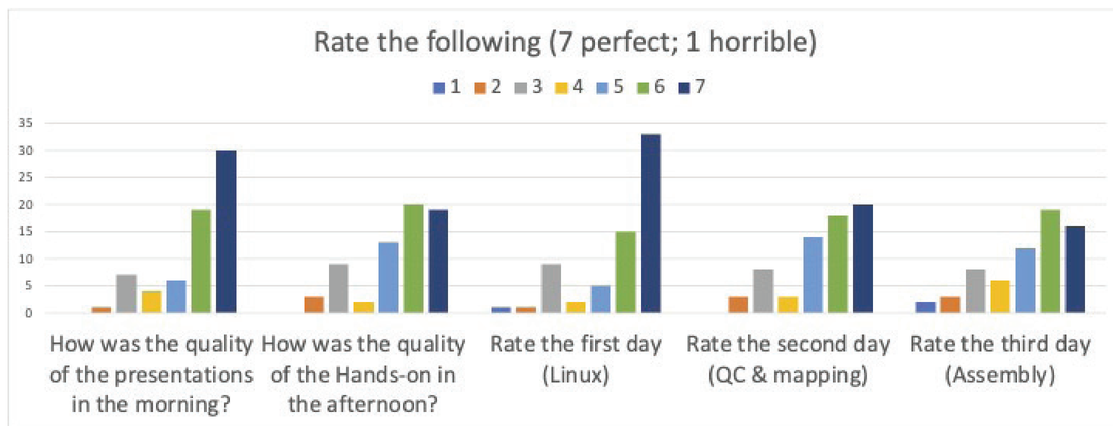


Figure 6. Evaluation of sections of the workshop.

are limited to 32, whereas we planned to use 264 CPUs. Only in the EC2 dashboard, under "Limits", can you find a CPU limits calculator that shows how much CPU you have available and how much you would need for your task. If you require more CPUs than are assigned, you can request an increase in the actual limit, which we did. However, it took three days for our request to increase the limit from 32 to 264 CPUs, and even then, we only received an increase of 192 CPUs, which was too late for the workshop and not enough for what we had planned. The take-home message is that even when using the EC2 environment and paying a higher price for on-demand computing resources, there are very low limits, and requests for increases can take a long time. Therefore, it is essential to plan well in advance for the needed resources and expect you may not receive the number of requested resources.

With 32 CPUs, unfortunately, we could not offer a reasonable solution for participants to run their own NGS data analysis workflows. We launched three VMs, each with 8 CPUs and 32GiB RAM (Amazon instance type t2.2xlarge), but participants were unable to run their analyses. As a result, local teachers ran the analysis, and participants followed the steps on the screen and had the opportunity to ask questions or request a repetition of some steps. The same scenario occurred during the hands-on exercises on day 3, where we demonstrated how to set up NGS data for an NGS assembly and run the assembly, including some basic quality tests of the assembled sequences.

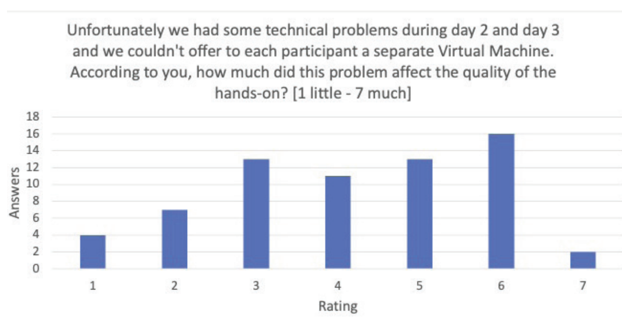


Figure 7. Rating of the influence on the quality of training due to the lack of access to computing.

At the end of the workshop, we requested that participants evaluate the event. Specifically, we wanted to understand their thoughts about the impact of being unable to access their server and follow the data analysis exercises. As shown in Figure 6, the difference between the first day and the following two days reflects the drop in the quality of the hands-on experience due to the inability to use their server. Surprisingly, with the help of local teachers demonstrating the data analysis, we were able to maintain a high level of quality, and participants did not rate this lack as a total failure.

Mini-symposium

After the three-day training, we closed the workshop with a one-day mini-symposium covering topics related to the EpiCass project, "Epigenetics and Crop Improvement." Once again, participants were present in their classrooms, and speakers connected from Africa, Germany, Switzerland, and the USA. Since we advertised the event as open and free, we also had up to 123 external attendees, mainly from Africa and Europe (Figure 8). We invited speakers to cover topics such as epigenetics in plants, especially in trees, viral diseases in cassava, molecular breeding in cassava, and epigenetics for crop improvement.

Ueli Grossniklaus from the University of Zurich in Switzerland was the first speaker and gave an interesting introduction to plant epigenetics, including an overview of results from his laboratory. He then developed the concept of population epigenetics and performance. Next on the program was Frank Johannes from the Technical University of Munich in Germany, who presented results of epigenetic analysis in trees and demonstrated that trees have different DNA methylation profiles in each branch originating from a different bud, as somatic epimutations accumulate continuously as a function of tree age. Claude Becker from the Ludwig Maximilians University in Munich, Germany, presented his group's work on epigenetics in abiotic and biotic stress, developing the concept that epigenetic variation can be associated with defence capacity.

The next two talks dealt with ongoing activities in cassava research. Stefan Winter from DSMZ (German

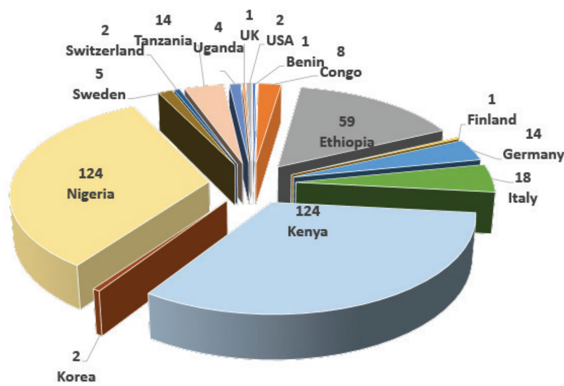


Figure 8. Distribution of countries of external attendees of the mini-symposium.

Collection of Microorganisms and Cell Cultures), who leads the Plant Virus Department in Germany, gave an overview of the viral threats in cassava and the latest research results on Cassava Brown Streak Virus and possible resistance against this virus. This virus is considered the biggest threat to food security in coastal Eastern Africa and around the eastern lakes and is slowly expanding towards Western Africa. Secondly, to provide an overview of advances in cassava breeding, Adenike Ige, a collaborator of Ismail Rabbi from IITA in Ibadan, Nigeria, presented the latest developments in marker-assisted selection and genetic gain they worked on over the last ten years.

The final presentation was given by Chad Niederhuth from Michigan State University in East Lansing, USA. Chad provided an overview of research in the last few years about epigenetics and possible crop improvement applications, mixed with his group's research results. He summarised that epigenetics could have wide-ranging applications in crop improvement, but we need much more research to uncover clear correlations.

The whole workshop, bioinformatics training and mini-symposium was organized by the EpiCass team, including Erik Bongcam-Rudloff, SLU, Uppsala, Sweden; Renaud van Damme SLU, Uppsala, Sweden; Livia Stavolone IPSP-CNR, Bari, Italy; Laurent Falquet, UniFr, Switzerland, Adnan Niazi, SLU, Uppsala, Sweden; Trushar Shah, IITA, Nairobi, Kenya; Michael Landi, IITA, Nairobi, Kenya, Temitayo Olagunju, IITA, Ibadan, Nigeria; and Andreas Gisel, ITB-CNR, Bari, Italy & IITA, Ibadan, Nigeria.

Acknowledgements

This work was supported by the Swedish Research Council (VR) funded projects, EpiCass grant number 2020-04457 and Cassavanet4Dev grant number 2021-05105.

Reviews

The effect of the planned behaviour theory and the transtheoretical behaviour model on physical activity. A systematic review.

Anastasia Dermatis[1], Flora Bacopoulou[2,3]✉, Ioulia Kokka[2,4], Dimitrios Vlachakis[3,5], Georgios Lyrakos[6], Despina Menti[6], Christina Darviri[2]

[1]Psychology Laboratory, CityU Research Center, Athens, Greece

[2]Postgraduate Course on the Science of Stress and Health Promotion, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece

[3]Center for Adolescent Medicine and UNESCO Chair on Adolescent Health Care, First Department of Pediatrics, School of Medicine, National and Kapodistrian University of Athens, Aghia Sophia Children's Hospital, Athens, Greece

[4]University Medical School, Outpatient Specialty Clinic for Sexual Health, First Department of Psychiatry, Eginition Hospital, Athens

[5]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece

[6]Department of Psychology, Cardiff Metropolitan University at City Unity College, Athens, Greece

Competing interests: AD none; FB none; IK none; DV none; GL none; DM none; CD none

Abstract

Systematic physical activity (PA) is crucial in preventing illnesses that can become life-threatening, such as colon and breast cancer, heart disease and ischemic stroke, cardio-respiratory disease, type II diabetes, and depression. Many theory–based interventions have been applied to achieve positive outcomes in an individual's behavioural change and the ability to engage in systematic PA. This systematic review investigates the influence of the Transtheoretical model of behaviour (TTM) and the theory of planned behaviour (TPB) on PA. A substantial search in Science Direct, Wiley Online Library databases and PubMed was performed to obtain articles about the topic. Data exportation was possible after the reviewers applied exclusion–inclusion criteria to estimate evidence quality. Empirical evidence was assessed with the CONSORT checklist to appraise the risk of bias. The primary search identified 195 studies. Of those, ten original studies were comprised. All studies indicated a positive influence of TPB and TTM on physical activity in non–health and healthy populations. In particular, it was found to have an impact on energy expenditure, balance and body strength. Theory-based interventions are notably effective in promoting physical activity behaviour. Researchers and health professionals must select and utilise interventions based on the above mentioned theories and aim to enhance PA behavioural change on individual and interpersonal factors. Although the positive outcomes of theory-based interventions on PA behaviour, it is necessary for further research to be conducted.

Introduction

The definition of physical activity (PA) states that it is "an activity of the body generated by the skeletal muscles which influence the enhancement of metabolic rate over resting energy expenditure" (Neufer *et al.*, 2015). PA is a complex behaviour classified into low, moderate, and high intensity. Regulate PA is linked to considerable health benefits, such as reduced type II diabetes, breast and colon cancer, depression, ischemic stroke, and cardiovascular disease (Welch *et al.*, 2019). Medical professionals recommend PA to relieve chronic pain significantly. It is estimated that 33% of the world's population is affected by chronic pain (Dureja *et al.*, 2014). Moreover, research has indicated that exercise and PA can amplify the operation of the central nervous system, reduce cognitive operational erosion associated with ageing, and decrease the possibility of dementia (Foster *et al.*, 2018; Kennedy *et al.*, 2017). Furthermore, PA induces neurotransmitters such as norepinephrine and serotonin, which subscribe to stress reduction (Chauvet – Gelinier and Bonin, 2017).

To reduce non – communicable diseases, the World Health Organization, under the Global Action Plan

Article history
Received: 01 May 2023
Accepted: 23 May 2023
Published: 21 September 2023

© 2023 Dermatis *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at http://journal.embnet.org.

Dermatis et al. (2023) EMBnet.journal 29, e1046
http://dx.doi.org/10.14806/ej.29.0.1046

(WHO, 2017), proposed a 25% decrease in the jeopardy of premature mortality from cancer, diabetes, and chronic respiratory disease, by 2025. Numerous reports underline that a significant percentage of people worldwide cannot become involved long–term with PA (Kohl III *et al.*, 2012). Thus, differentiating the characteristics of effective PA promotion plans has become a significant concern for public health authorities. Theory-based projects appear to be most efficacious when considering behavioural change.

Such projects are developed according to valid theories of behavioural change, which verify the assumption that behaviour moderators need to be modified to establish a behaviour change (Glanz and Bishop, 2010). A vast amount of theories have been applied to attain positive behavioural outcomes. It has been acknowledged that theory-based interventions allow scientists to collect data, test hypotheses, distinguish variables that influence behaviour, and propose mechanisms which should be included during behavioural interventions (Davis *et al.*, 2015; Michie *et al.*, 2014). A theoretical approach with a significant effect on behavioural change towards PA is the theory of planned behaviour (TPB). The TPB focuses on theoretical apprehensions concerning an individual's motivational features as determinants in establishing specific behaviour (Ajzen, 2015). Furthermore, a theory-based model is highly recognised for its ability to interpret when and how individuals are willing to change their behaviour, also known as the transtheoretical model of change (TTM). The theory mainly centers on how individuals transition from one stage to the next, decisional balance, and self–efficacy. Also, TTM explores an individual's readiness to conform to new behaviour and incorporates five behavioural change stages (Prochaska and DiClemente, 1983).

This review investigates the influence of both theories-based models on physical activity.

Methods

Study selection

Wiley Online Library databases, Science Direct, and PubMed were used to identify studies. Keywords used in the search equation were: Physical Activity (PA), Transtheoritical Model (TTM), Theory of Planned Behavior (TPB), and Exercise and Physical Activity.

Inclusive criteria: 1) studies that evaluated the effect of TPB or TTM on PA, 2) regardless of age and gender of participants, 3) randomised control trial design studies, 4) studies including non–clinical and clinical population, 5) studies including interventions, regardless of a specific structure (sequence of events, duration, location).

Exclusive criteria: 1) studies subject matter unrelated to the topic, 2) studies published in languages other than English, 3) other systematic reviews and meta-analyses, and 4) scientific protocols published providing inadequate outcomes.

Study evaluation

Specific criteria were used to assess all studies of this systematic review from the Consolidating Standards of Reporting Trial Checklist 2010 (CONSORT) (Schulz *et al.*, 2010). Criteria required for this article were: structured abstract, eligibility of participants, sufficient sample size, number of participants allocated to groups, flow diagram, percentage of drop–outs for each group, demographics and clinical information, and limitations.

Results

The primary search submitted 195 articles. Due to the absence of theory-based interventions, eight articles were excluded. Furthermore, thirty articles were excluded due to the absence of full-text availability. The non-reference to PA excluded forty-five articles. Lack of randomised control trial design excluded thirty articles. Finally, in this systematic review, ten articles were evaluated. The flow chart of article selection is presented in Figure 1.

According to the CONSORT 2010 criteria checklist, a structured abstract was not provided for three articles (Chatzisarantis *et al.*, 2015; Darker *et al.*, 2010; Shirazi *et al.*, 2007). All articles comprised eligibility criteria and pre–specified primary and secondary estimations. The sample size was determined in only half of the articles (Darker *et al.*, 2010; Marshall *et al.*, 2003; Jennings *et al.*, 2014; Shirazi *et al.*, 2007). Six articles indicated a control trial design and generation (Darabi *et al.*, 2017; Chatzisarantis *et al.*, 2015; Darker *et al.*, 2010; Jennings *et al.*, 2014; Marshall *et al.*, 2003; Shafieinia *et al.*, 2016). All articles identified statistical methods, while only two did not present a flow chart (Darker *et al.*, 2010; Mostafavi *et al.*, 2015). Participants' withdrawal rates and losses were described in all ten articles. Two articles did not provide baseline demographic tables (Darker *et al.*, 2010; Mostafavi *et al.*, 2015). Limitations were indicated in all ten articles. Overall, the studies being evaluated

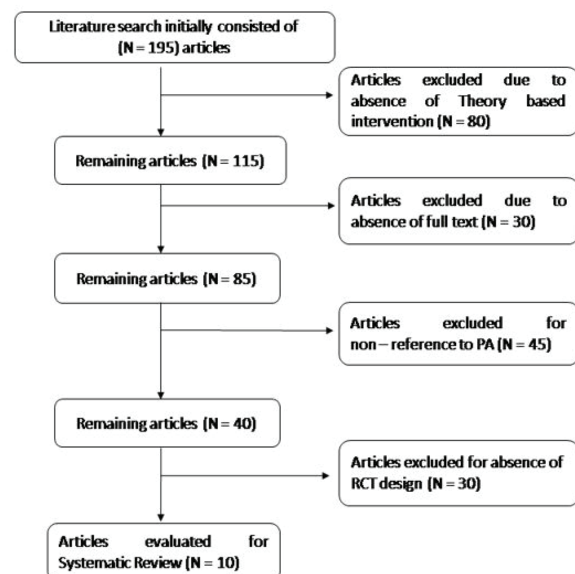


Figure 1.

Dermatis et al. (2023) EMBnet.journal 29, e1046
http://dx.doi.org/10.14806/ej.29.0.1046

Table 1. Studies included in the systematic review.

STUDY	PARTICIPANTS	CONTROL GROUP	INTERVENTION	RESULTS
Author, Year, Origin	**sample, age, gender, characteristics**	**sample, mean age, gender**		**Main Outcomes**
Shirazi *et al.*, 2007, Iran	n=61, 40-65 yrs, females, general population osteoporosis prevention	n=55, females, 40-65 yrs	TTM 12-week intervention included an exercise program costumed for each participant on weight training and walking and an educational program based on individual (SoC).	The active group increased lower body strength and balance. Psychological (SoC) also showed positive progress. From being sedentary (SoC) to active.
Proper *et al.*,2003, Germany	n=299,males and females, over 18	n=168	Seven 20 min TTM counselling sessions structured on participants' SoC for nine months based on PACE program measures. In addition, healthy lifestyle factors were provided to increase PA behaviour.	Increased cardiovascular capacity and energy expenditure indicated TTM interventions positively affected PA.
Pinto *et al.*,2005, Nertherlands	n=43, females, 18 and older, breast cancer patients	n=43, females, 18 and over, breast cancer patients	PA counseling 12 week TTM intervention based on weekly exercise info and (SoC) to enhance PA behavior.	There was a rise in moderate intensity PA levels and high energy expenditure per week after increased PA behavioral promotion.
Marshall *et al.*, 2003, Australia	n=227, males and females, 40 – 60 yrs., general population	n=235, males and females, 40 – 60 yrs., general population.	TTM intervention included 1 time mailing of 4 brochures on PA structured on participants SoC to promote PA behavior.	Short – term PA behavior was increased after TTM intervention.
Mostafavi *et al.*, 2015, Iran	n=71, females, mean age 52 yrs. Metabolic Syndrome	n=71, females, mean age 52 yrs. Metabolic Syndrome	TTM Intervention consisted of five 1 hour educational sessions and exercise instructions to promote PA behavior.	TTM intervention had a positive increase in PA behavior. Lowered triglyceride, and HDL cholesterol levels.
Shafieinia *et al.*, 2016, Iran	n=54, females, age between 18 – 60 yrs. general population	n=54, females, age between 18 – 60 yrs. general population	TTM intervention four 90' educational sessions and e mail reminding correspondence once every 2 weeks to increase PA behavior.	TPB had major increase on PA behavior. Increase of objective to walk longer, inspired participants to indeed walk longer.
Chatzisarantis *et al.*, 2015. Australia	n=1028, male and female adolescents, age 12 – 18 yrs. secondary students	n=344, male and female adolescents, age 12 – 18 yrs. secondary students	TPB intervention to enhance PA included reading students persuasive messages and vigorous PA four days a week for eight weeks during leisure time.	Leisure physical activity was increased and showed medium – to large change following an effective TPB intervention program.
Darker *et al.*, 2010, Ireland, UK, Aberdeen.	n=66, male and female, age 16 – 65 yrs,general population	n=64, male and female, age 16 – 65yrs, general population	During a 77 days TPB intervention to promote PA participants committed to increase walking time 10 – 20' each week. Intervention was based on coping planning, goal setting, and action planning.	TPB intervention significantly influenced walking behavior. Participants increased walking behavior during the week.
Jenning *et al.*, 2014, Australia, Canada.	n=184, males and females, age <18yrs,type II diabetes	n=202, males and females, age <18yrs,type II diabetes	12-week web - based TPB intervention to increase PA.	TPB intervention changed PA behavior. A web – based programme indicated a short term change in physical activity. All PA measures, including steps per day, moderate and vigorous physical activity ,were increased.
Darabi *et al.*, 2017, Iran.	n=289, female, age 12 – 16yrs.	n=289, female, age 12 – 16yrs.	6 months TPB intervention with 90 min educational sessions and complementary media training to increase PA behavior.	Adolescent girls improved overall PA after completing a TBA educational intervention .

were suitable to be incorporated in this review and are outlined in Table 1.

Five (50%) of the studies involved TPB interventions and five (50%) of the studies involved TTM interventions. 40% of the research focused on female participants. Eight of the studies investigated adult participants, whereas two were secondary school students. Four studies investigated TPB and TTM intervention results on clinical populations diagnosed with osteoporosis, metabolic syndrome, breast cancer, and type II diabetes. The seven remaining articles focused on non–clinical population. The mean sample size was n=232.2. The smallest size was n=43, and the largest size was n=1.028.

Discussion

This systematic review aims to investigate the influence of the TPB and the TTM on behavioural change concerning PA. Outcomes showed that both models were efficient in increasing PA. TTM- and TPB-based interventions positively influenced PA by incrementing fitness levels in clinical and non–clinical populations. A TTM 12 – week home–based intervention was conducted to decrease osteoporosis in female participants. The intervention included an educational program on weight training and walking. Secondary results indicated an increase in PA by showing an improvement in participants' lower body strength and balance. Moreover, a significant benefit in psychological SoC was apparent in the experimental group but not in the control group (Shirazi *et al.*, 2007). A different study evaluating the effect of a TTM intervention on PA indicated a significant improvement in PA by measuring the total increase in energy expenditure. Intervention periods were short and tailored for each individual on SoC. The positive increase in PA was evident in outcome measures.

According to SoC, results of the TTM 12-week intervention based on PA weekly exercise information and consultations showed a substantial improvement in PA behaviour. Differentiation among intervention and control groups showed an increase in overall time of PA at a modest – intensity per week. Moreover, the experimental group indicated a significant rise in high and very high energy expenditure over the control group (Pinto *et al.*, 2005). The outcomes of an RCT study structured as a TTM intervention on adults showed an increase in short–term PA behaviour. The intervention included a one–time mailing of four pamphlets and a letter to a randomly distributed experimental group. The brochures were evaluated carefully to represent participants' current SoC. At two months' baseline the experimental group showed a significant PA increase of 78 minutes per week, whereas at six months, PA behaviour had sustained higher for the experimental group than the control group but had been reduced insignificantly compared to two months post–baseline calculations (Marshall *et al.*, 2003). Empirical evidence in a study investigating the impact of a TTM intervention to increase PA for females with metabolic syndrome showed an increase in PA levels for the intervention group, with significant advancement in all SoC. Participants in the control group did not indicate any shift in PA behaviour. All shifts in TTM constructs were notable in the experimental group. Both exercise instructions and educational sessions positively increased PA behaviour (Mostafavi *et al.*, 2015). The efficient outcomes of both theories are highlighted in two more studies. Scientists designed an RCT intervention based on TPB to increase PA behaviour among women. The intervention was structured on a combination of emails sent once every two weeks for three months, sent once every two weeks for three months, and four educational sessions with a duration of 90'. Results indicated an enhancement in all TPB variables, with the exception of subjective norms, and an increase in PA behaviour in the intervention group compared to the control group (Shafieinia *et al.*, 2016). In addition, a TPB school–based intervention was designed to alter leisure–time PA behaviour. The intervention included a high-intensity 40' PA session four days a week for eight weeks. The study outcomes revealed a shift in PA behaviour by showing a medium to high increase in PA leisure– time (Kawabata *et al.*, 2018). A one-week TPB intervention was implemented to increase PA among the general public. The intervention aimed to promote walking based on objective measures. Results showed that brisk walking was increased from 20' to 32' over one week, thus indicating a significant enhancement in PA behavioural change (Darker *et al.*, 2010).

The positive effect of the TPB on the increase of PA was also perceivable in a 12 – week web–based intervention for adults with type II diabetes. A self–management strategy was used to increase PA behavioural change. Study's outcome by the end of the 12 weeks indicated an overall increase in PA behaviour (Jennings *et al.*, 2014). Results were consistent with those of Baker and Mutrie, (2005). Researchers designed a four–week TTM intervention to increase PA using a pedometer. Data analysis showed a positive change in PA behaviour.

Moreover, results were consistent with those of Dinger *et al.*, (2007), who indicated a significant PA behavioural change using a TTM web–based intervention. Also, research outcomes were consistent with those of Kawabata *et al.*, (2018), showing an increase in leisure–time PA after completing a four–week PA intervention among secondary school students based on the TPB. In addition, Vandelanotte *et al.*, (2007), empirically indicated an important increase in PA behaviour change utilising a web–based TPB structured intervention.

The current systematic review does not lack limitations. It is necessary to mention that 40% of the articles being evaluated had high rates of sample dropouts (Mostafavi *et al.*, 2015; Shirazi *et al.*, 2007; Darker *et al.*, 2010; Shafieinia *et al.*, 2016). Furthermore, for only a few studies, fair homogeneity was evident among participants when examining socioeconomic

status and race. This factor could limit generalisation of the outcomes (Shirazi *et al.*, 2007; Pinto *et al.*, 2003; Darabi *et al.*, 2017; Proper *et al.*, 2003). Also, a significant limitation is using self–report questionnaires in all of the articles being reviewed. Worth mentioning is that the study conducted by Marshall *et al.*, (2003) held interviews via phone. This may have as a result social desirability bias, a tendency of the respondent to answer in ways favourable to the scientist (Akbulut, 2017). Moreover, a limitation in one study designed by Darker *et al.*, (2010) is that volunteer participants were recruited to participate in the study's TPB intervention. Such action could develop an intervention group willing to increase walking behaviour than a group of non – volunteers who would be less accepting.

Conclusions

Physical activity is a significant factor in maintaining optimum health status and preventing life–threatening, non – communicable diseases. A systematical exercise program has been proven to positively affect physical and mental health. The TPB and TTM theories addressing behavioural change have been applied to modify individuals' positions towards PA. The outcomes of the current review indicated that both theoretical models positively enhance PA for non–healthy and healthy populations. However, all studies showed promising results, the fact that the sample sizes were small dictates that scientific outcomes cannot be generalised. Future studies should aim for a greater number of participants, which would increase the opportunity to generalise the findings.

Key Points

- Regular physical activity can prevent mental and physical health issues.
- Behavioural theory-based models can affect an individual's behaviour.
- The theory of planned behaviour and the transtheoretical model of behaviour can alter behaviour and increase physical activity.
- Larger sample sizes are required for research.

Funding

This systematic review did not receive any specific grant from funding agencies in the commercial, public, or not-for-profit sectors.

References

Ajzen I (2015) The theory of planned behaviour is alive and well, and not ready to retire: a commentary on Sniehotta, Presseau, and Araújo-Soares. Health psychology review **9**(2), 131-137. http://dx.doi.org/10.1080/17437199.2014.883474

Akbulut Y, Dönmez O, Dursun ÖÖ (2017) Cyberloafing and social desirability bias among students and employees. Computers in Human Behavior **7**(2), 87-95. http://dx.doi.org/10.1016/j.chb.2017.02.043.

Baker G, Gray SR, Wright A, Fitzsimons C, Nimmo M *et al.* Scottish Physical Activity Research Collaboration (SPARColl) (2008) The effect of a pedometer-based community walking intervention" Walking for Wellbeing in the West" on physical activity levels and health outcomes: a 12-week randomized controlled trial. International Journal of Behavioral Nutrition and Physical Activity **5**, 1-15. http://www.ijbnpa.org/content/5/1/44

Neufer P D, Bamman MM, Muoio DM, Bouchard C, Cooper *et al.* (2015) Understanding the cellular and molecular mechanisms of physical activity-induced health benefits. Cell metabolism **22**(1), 4-11. http://dx.doi.org/10.1016/j.cmet.2015.05.011

Chatzisarantis NLD, Kamarova S, Kawabata M, Wang, JCK (2015) Developing and evaluating utility of school-based intervention programs in promoting leisure-time physical activity : An application of the theory of planned behaviour. International Journal of Sport Psychology **46**(2), 95-116. http://hdl.handle.net/10497/17251

Chauvet-gelinier J, Bonin B (2017) Science Direct Stress , anxiety and depression in heart disease patients : A major challenge for cardiac rehabilitation. Annals of Physical and Rehabilitation Medicine **60**(1), 6–12. http://dx.doi.org/10.1016/j.rehab.2016.09.002.

Darabi F, Kaveh MH, Majlessi F, Farahani FKA, Yaseri M *et al.* (2017) Effect of theory-based intervention to promote physical activity among adolescent girls: a randomized control trial. Electronic physician **9**(4), 4238. http://dx.doi.org/10.19082/4238

Darker CD, French DP, Eves FF (2010) An intervention to promote walking amongst the general population based on an "extended" theory of planned behaviour: a waiting list randomised controlled trial. Psychology of Health **25**(1), 71–88. http://dx.doi.org/doi:10.1080/08870440902893716.

Davis R, Campbell R, Hildon Z, Hobbs L, Michie S (2015) Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health psychology review **9**(3), 323-344. http://dx.doi.org/10.1080/17437199.2014.941722

Dinger MK, Heesch KC, Cipriani G (2007) Comparison of two email-delivered, pedometer-based interventions to promote walking among insufficiently active women. Science of Medicine and Sports **10**(5), 297–302. http://dx.doi.org/doi:10.1016/j.jsams.2006.07.011.

Dureja GP, Jain PN, Shetty N, Mandal SP, Prabhoo R *et al.* (2014) Prevalence of chronic pain, impact on daily life, and treatment practices in India. Pain Practice **14**(2), E51-E62. http://dx.doi.org/10.1111/papr.12132

Foster C, Shilton T, Westerman L, Bull F (2018) World Health Organisation to develop global action plan to promote physical activity: time for action. British Journal of Sports Medicine **52**, 484–485. http://dx.doi.org/10.1136/bjsports-2017-098070

Glanz K, and Bishop DB (2010) The role of behavioral science theory in development and implementation of public health interventions. Annual review of public health **31**, 399-418. http://dx.doi.org/10.1146/annurev.publhealth.012809.103604

Jennings CA, Vandelanotte C, Caperchione CM (2014) Effectiveness of a web-based physical activity intervention for adults with Type 2 diabetes-a randomised controlled trial. Preventive Medicine **60**,33–4. http://dx.doi.org/doi:10.1016/j.ypmed.2013.12.011.

Kawabata M, Chua K, Chatzisarantis NL (2018) A school-based intervention program in promoting leisure-time physical activity: trial protocol. BMC public health **18**(1), 1-7. http://dx.doi.org/10.1186/s12889-018-5320-1

Kennedy G, Hardman RJ, Macpherson H, Scholey AB (2017) How Does Exercise Reduce the Rate of Age-Associated Cognitive Decline ? A Review of Potential Mechanisms. Journal of Alzheimer's disease **55**,1–18. http://dx.doi.org/10.3233/JAD-160665.

Kohl HW, Craig, CL, Lambert EV, Inoue S, Alkandari JR *et al.* (2012) The pandemic of physical inactivity: global action for public health. The lancet 380(9838), 294-305. http://dx.doi.org/10.5167/uzh-69552.

Marshall AL, Bauman AE, Owen N, Booth ML, Crawford DMB (2003) Population-based randomised controlled trial of a stage-targeted physical activity intervention. Annals of Behavioural Medicine **25**(3), 194–202. http://dx.doi.org/doi:10.1207/S15324796ABM2503_05.

Michie S, West R, Campbell R, Brown J, Gainforth (2014) ABC of behaviour change theories (ABC of behavior change): An essential resource for researchers, policy makers and practitioners. Silverback Publishing (Silverback IS).

Mostafavi F, Ghofranipourm F, Feizi A (2015). Improving physical activity and metabolic syndrome indicators in women: a transtheoretical model-based intervention. International Journal of Preventive Medicine **6**(28). http://dx.doi.org/doi:10.4103/2008-7802.154382.

Pinto BM, Frierson GM, Rabin C, Trunzo JJ (2005) Home-based physical activity intervention for breast cancer patients. Journal of Clinical Oncology **23**(15), 3577–3587. http://dx.doi.org/doi:10.1200/JCO.2005.03.080.

Prochaska JO, DiClemente CC (1983) Stages and processes of self-change of smoking: toward an integrative model of change. Consulting Clinical Psychology **51**(3), 390–395. https://psycnet.apa.org/doi/10.1037/0022-006X.51.3.390

Proper KI, Hildebrandt VH, Beek AJ, Van Der, Twisk JWR (2003) Effect of Individual Counseling on Physical Activity Fitness and Health A Randomized Controlled Trial in a Workplace Setting. American journal of preventive medicine **24**(3), 218–226. http://dx.doi.org/10.1016/S0749-3797(02)00645-1.

Schulz KF, Altman DG, Moher D (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Journal of Pharmacology and pharmacotherapeutics **1**(2), 100-107. http://dx.doi.org/10.4103/0976-500X.72352

Shafieinia M, Hidarnia A, Kazemnejad A, Rajabi R (2016) Effects of a Theory Based Intervention on Physical Activity Among Female Employees : A Quasi-Experimental Study. Asian journal of sports medicine **7**(2) http://dx.doi.org/10.5812/asjsm.31534

Sharabi A, Margalit M (2011) The mediating role of internet connection, virtual friends, and mood in predicting loneliness among students with and without learning disabilities in different educational environments. Journal of Learning Disabilities **44**(3), 215–227. http://dx.doi.org/10.1177/0022219409357080.

Shirazi K, Wallace LM, Niknami S, Hidarnia A, Torkaman *et al.* (2007) A home-based , transtheoretical change model designed strength training intervention to increase exercise to prevent osteoporosis in Iranian women aged 40 – 65 years : a randomised controlled trial. Health Education Research **22**(3), 305–317. http://dx.doi.org/10.1093/her/cyl067.

Vandelanotte C, Spathonis KM, Eakin EG, Owen N (2007) Website-delivered physical activity interventions: A review of the literature. American journal of preventive medicine **33**(1), 54-64. http://dx.doi.org/doi:10.1016/j.amepre.2007.02.041.

Welch WA, Strath SJ, Brondino M, Walker R, Swartz AM (2019) Duration–Response of Light-Intensity Physical Activity and Glucose Dynamics in Older Adults. Journal of Physical Activity and Health **16**(1), 37–42. http://dx.doi.org/10.1123/jpah.2017-0557.

World Health Organization. (2017). Preventing noncommunicable diseases (NCDs) by reducing environmental risk factors (No. WHO/FWC/EPE/17.01). World Health Organization. https://tumourclassification.iarc.who.int/

EMBnet.journal
Bioinformatics in Action

Olagunju et al. (2024) EMBnet.journal 29, e1037
http://dx.doi.org/10.14806/ej.29.0.1037

Research Papers

Ds-Seq: An Integrated Pipeline for In Silico Small RNA Sequence Analysis for Host-pathogen Interaction Studies

Temitayo Adebanji Olagunju[1,2], Angela Uche Makolo[1], Andreas Gisel[2,3]✉

[1]University of Ibadan, Ibadan, Nigeria
[2]Bioscience Centre, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria
[3]Institute for Biomedical Technologies, National Research Council of Italy (CNR), Bari, Italy
Competing interests: TAO none; AUM none; AG is a member of EMBnet.journal Editorial Board

Abstract

Plant-pathogen interactions activate molecular activities wherein the host defends the pathogen while the pathogen tries to suppress the plant response. Small RNAs (sRNAs) mediate major mechanisms, including post-transcriptional gene silencing, histone modification and DNA methylation by which plants respond to the presence of pathogens. Genome-wide profiling of host and pathogen sRNAs is therefore pivotal to uncovering the mechanisms underlying the host-pathogen interaction and mechanisms for host resistance. sRNA high throughput sequencing (HTS) data analysis often involves multiple stages/tools. Most necessary tools are accessible only through the command line, making it challenging for those without a high level of Unix/Linux skills. Furthermore, installation of some of these tools may become difficult due to dependencies and software version compatibility. We have developed an integrated open-source pipeline, Ds-Seq, for end-to-end in silico analysis of sRNA HTS data with improved reproducibility. The pipeline combines in-house scripts and public tools in a shell script, which can be invoked with a single command. The pipeline's usefulness has been demonstrated with testing on publicly available and published data from independent sRNA-seq datasets of host-pathogen interaction studies. Ds-Seq is available on GitHub, while a Docker image can be obtained from the Docker hub.

Introduction

Small RNAs (sRNAs), generally between 20-30 nt long, have regulatory functions which are critical to many biological processes (Won *et al.*, 2014). The 21-24 nt subset, also referred to as micro RNAs (miRNAs), are evolutionarily conserved between species (Zhang *et al.*, 2006). In host-pathogen interaction, sRNAs play a significant role in the defense response of the host to pathogen invasion through the RNA interference (RNAi) machinery (Pumplin and Voinnet, 2013; Carbonell *et al.*, 2019). sRNAs can also regulate gene expression through DNA methylation and histone modification (Wang *et al.*, 2018; Tamiru *et al.*, 2018; Diezma-Navas *et al.*, 2019). These multiple pathways of sRNA-mediated gene expression regulation underscore the importance of genome-wide characterization of sRNAs as an essential first step to further investigation of molecular mechanisms orchestrated by sRNAs in response to different environmental cues.

High throughput sequencing (HTS) technologies such as small RNA sequencing (sRNA-seq) and computational analysis tools have been successfully applied to carry out genome-wide profiling of the sRNA landscape of many organisms to investigate roles played by sRNAs in different conditions (Li *et al.*, 2018; Hu *et al.*, 2020; Wenlei *et al.*, 2020).

Knowledge discovery using computational tools for in silico analysis of HTS data often involves multiple stages, determined mainly by the specific research questions to be answered. Each step usually entails selecting a computational tool from a buffet of available tools. Many of these tools are accessible only through a command-line interface (CLI) (Seemann, 2013), making it a challenge, especially for those not familiar with a Unix/Linux operating system environment (Xu *et al.*, 2014; Morais *et al.*, 2018; Joppich and Zimmer, 2019). The command-line design, however, favours a batch processing approach for computational pipelines where the output file of a stage, after manipulation, is parsed to the tool at the next step. Furthermore, installation of some of these required tools for analysis may become difficult due to dependencies and software version

Article history
Received: 23 September 2023
Accepted: 19 November 2023
Published: 25 January 2024

© 2024 Olagunju *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at http://journal.embnet.org.

compatibility with user operating systems (List *et al.*, 2017; Mangul *et al.*, 2018). More efficient and effective research can only be conducted when wet biologists are left to focus on knowledge discovery from computational analysis and not waste time setting up tools, troubleshooting or learning to use them (Smith, 2013; Smith, 2014). It is important to note that the problem of reproducibility of results (Stodden *et al.*, 2018; Beaulieu-Jones, 2017) also becomes pronounced against the backdrop of the challenges of bioinformatics tools installation and management.

Some pipelines for the analysis of sRNAs are already in existence but are not tailored for host-pathogen interaction experiments. The Java-based UEA sRNA Workbench (Beckers *et al.*, 2017) was developed to explore the sRNA landscape based mainly on the differential expression analysis of sRNA libraries. This tool is thus limited in its scope. sRNAPipe was developed by (Pogorelcnik, 2018) as a GUI-based pipeline on the Galaxy platform (Afgan *et al.*, 2018) for the exploratory analysis of sRNA libraries. The pipeline maps the reads to only the host genome assembly, after which they are classified into four main sub-groups for further exploration and does not include a module for differential expression analysis of the sRNAs in the libraries. sRNAbench and sRNAtoolbox 2019 (Aparico-puerta *et al.*, 2019), which is an update of the sRNAtoolbox (Rueda *et al.*, 2015), provides a more comprehensive suite of tools for the exploration of the sRNA landscape through a GUI, including differential expression analysis, visualisation of genome-mapped reads on a genome browser, and sRNA target prediction. Although the sRNAtoolbox provides a differential expression analysis module, unlike the UEA sRNA Workbench and the sRNAPipe, it does not offer segregation of the sRNAs by the library to aid quick identification of unique and common sRNAs in the libraries.

In this paper, we present Dockerized sRNA-Seq tool, Ds-Seq, an open-source in-silico analysis pipeline for investigating host-pathogen interaction in plants with available genome assemblies through profiling the sRNA landscape. With parameters specified by the user through a configuration file, it takes the fastq files of the sRNA-Seq libraries as input, maps the reads to the user-supplied host and pathogen genome assemblies, produces a differential expression profile of the sRNAs in the libraries, segregates the host-mapped sRNAs by the library, identifies known miRNAs, and predicts novel sRNAs. Only a single command is required to initiate the analysis with this pipeline, making it relatively easy to use in a CLI environment without requiring high technical skills in UNIX/Linux environment. We also introduced the use of a Docker container to provide a consistent environment for reproducibility of results (Baker and Penny, 2016) and eliminate the challenges occasioned by software versioning and compatibility issues. This pipeline has been tested with publicly available sRNA-Seq data from host-pathogen interaction studies with accompanying published results. The outcome of the testing generally showed agreement between results obtained by Ds-Seq and published data. Ds-Seq is freely available from the GitHub repository[1] and Docker hub with ID cephas/ds-seq[2].

Materials, Methodologies and Techniques

Ds-Seq (with schematic shown in Figure 1) was developed with Perl, Python and R scripts, all wrapped in a single shell script that can be run with a single command. It was designed with a modular structure such that each module corresponds to a stage of the analysis that could be carried out separately with other bioinformatics tools. The user can select specific modules of interest or all the modules through a configuration file where all the parameters for the analysis are defined. It is important to note that Ds-Seq does not contain a database, so all the files required for analysis, including reference genome sequences, have to be supplied by the user. Parameters for the independent tools are also to be defined by the user through an accompanying configuration file. Due to the wide variation of the sRNA family, based on the sequence length, the class of sRNA of interest to a user can be defined and controlled with the length parameters in the configuration file. Further exclusion of other types of non-coding RNAs such as tRNA, piwiRNAs, snoRNAs etc., can be made from the analysis through a multi-fasta file containing the sequences to be excluded. Sequences from Rfam (Kalvari *et al.*, 2020) are examples of sequences that could be excluded from any analysis carried out with Ds-Seq.

Design

Ds-Seq was designed to analyze sRNA-seq data easily and with reproducibility, especially for wet biologists with minimal experience with Unix/Linux operating system command line interface. As such, the invocation of the pipeline is achieved with a single line of command after parameters have been defined in the accompanying configuration file and user files have been correctly placed. Each module of Ds-Seq has specific file requirements such that the modules selected by a user would determine the required input files and data. File requirements are shown in Table 1 for the modules (i) NGS sequences filtering, (ii) expression profiling, (iii) novel sRNA prediction, (iv) genome-wide host and pathogen sRNA mapping, and (v) conserved sRNA mapping. Reads mapping is carried out generally using Bowtie (Langmead, 2009), being a memory-efficient and ultra-fast short DNA sequence alignment program, using default parameters for mapping in addition to reporting the best alignment. At the same time, the user defines the number of accepted nucleotide mismatches between sequences through the configuration file.

[1]https://github.com/CEPHAS-01/small-RNASeq.ngs
[2]https://hub.docker.com/r/cephas/ds-seq

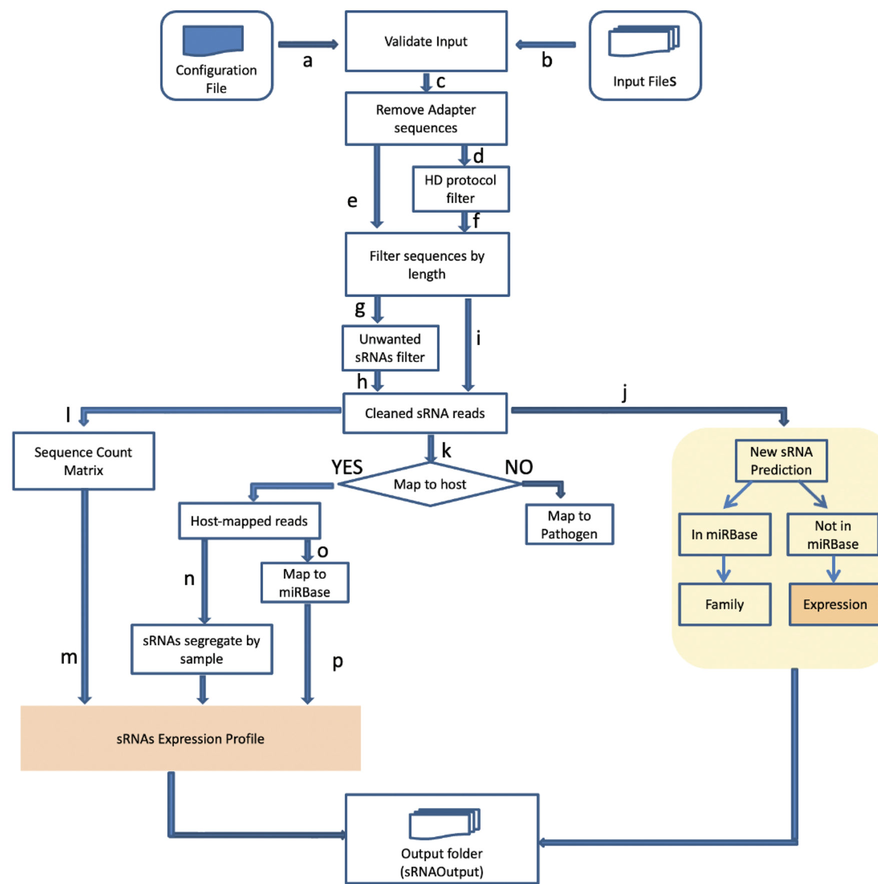


Figure 1. Schematic of sRNA sequencing analysis pipeline.
(a) Parameters and file paths are defined in a configuration file. **(b)** Raw sequence reads of the samples in .fastq or .gz file format, and other data are supplied to the validation module. **(c)** With successful validation stage, reads are passed on for removal of adapter sequences where raw reads without adapters are discarded while adapter sequences are removed from the other reads. **(d)** Adapter-cleaved reads are further processed if the HD protocol was used in the library preparation. **(e)** Adapter-cleaved reads are filtered for the length specified in the configuration file to produce clean reads. **(f)** Reads processed for HD protocol are filtered with length to produce clean reads. **(g)** Length-filtered reads are optionally processed to remove unwanted non-coding RNAs supplied by the user **(h)** Clean reads are produced for further analysis **(i)** cleaned reads are produced for further analysis **(j)** Cleaned reads are parsed for novel small RNA prediction. **(k)** Cleaned reads are mapped to the host genome and separated into host-mapped and unmapped sRNAs **(l)** Cleaned reads are parsed for expression matrix generation, and **(m)** expression matrices are parsed for differential expression analysis **(n)** Host-mapped sRNAs are segregated by sample **(o)** Host-mapped sRNAs are mapped to known miRNAs repository to identify known miRNAs **(p)** Known miRNAs expression profile in samples are produced.

Input raw sequence files can be left in compressed zip (.gz) format without the need to first decompress as the pipeline can discriminate the file type and handle it appropriately. The raw sequence files of the sRNA-Seq libraries must be saved in replicates in a folder named by the sample and located in the 'data' directory in a manner representative of the experimental design. A unique analysis ID is generated for each non-redundant sequence and is consistent across all the libraries. The checkpoint feature is implemented at strategic stages of the pipeline to avoid repeating steps previously executed, especially tasks requiring intensive compute time, such as reference sequence indexing. The configuration file can be used for processing other analyses while retaining the same parameters to enable the user to compare different libraries to identify similarities and differences. The analysis modules in Ds-Seq are further explained in the following sections.

NGS reads filtering

The NGS data filtering is carried out using in-house Perl scripts to remove adapter sequences and sequences without an adapter from the input raw reads library. An optional further clipping of 4 nt from the 5' and 3' ends of the cleaned reads is done if the High Definition protocol (Sorefan *et al.*, 2012) was used during the library preparation. For each input raw reads file, each sequence is aggregated into one, and its abundance in the input file is appended to the sequence header information to produce a multi-fasta file. Non-redundant reads with a length that meets the user-defined range and reads abundance not less than the defined threshold in the configuration file are retained for use downstream of the analysis pipeline as clean reads. An optional step

Research Papers

Table 1. Files requirement for each modular stage of the analysis pipeline.

Analysis module	Other input parameters (through configuration file)	Reference genome assembly	Annotation file	Chromosome length file
Filter NGS data	Yes	No	No	No
Host mapping	Yes	Yes	No	No
Conserved sRNA	Yes	Yes	No	No
Pathogen mapping	Yes	Yes	No	No
Novel sRNA prediction	Yes	Yes	Yes	Yes
Expression profile	Yes	No	No	No

also removes any other unwanted non-coding RNA sequences from the reads, such as sequences from Rfam (Kalvari *et al.*, 2020) or different user-defined sequences supplied through a multi-fasta file. A bar plot of the length distribution of the cleaned reads from all replicates of the samples is also produced to give an insight into the profile of the sRNAs within the libraries and as a further check of the quality of the sRNA-Seq libraries. This information helps determine the quality of the library. It could also provide information on the specific molecular pathway origin of the sRNAs based on the activities of Dicer (Mueth *et al.*, 2015).

Differential Expression Profiling

This module in the pipeline is used to reveal the differential expression profile of all sRNAs across the libraries and experimental conditions of interest. Count matrices of the cleaned reads across pairwise combinations of all libraries are produced, using the sequence as a unique identifier to achieve sequence identity uniformity across all the libraries. These matrices are parsed as input into the edgeR (Robinson *et al.*, 2009; McCarthy *et al.*, 2012) package for the differential expression analysis. Reads that meet the fold change and p-value criteria defined in the configuration file are returned as differentially expressed.

Conserved/known miRNAs identification

Micro RNAs (miRNAs) belong to a sub-class of sRNAs with sequence lengths between 21-24 nt. Some miRNAs are evolutionarily conserved across species (Zhang *et al.*, 2006) and are identified through a homology search of the known mature sequences in the miRNA repository miRBase (Kozomara *et al.*, 2019) (www.mirbase.org). In this pipeline module, all known mature sequences hosted in miRBase are supplied as a multi-fasta sequence file indexed with Bowtie (Langmead *et al.*, 2009) before mapping. The user can alternatively provide a list of conserved sequences of interest in a multi-fasta format as a reference in place of the mature sequences from miRBase. All sequences with the best hit within the defined acceptable nucleotide mismatch parameter defined in the configuration file mapping to this repository are reported as known sRNAs. The sequences specific to the host are reported as known host sRNAs.

Prediction of novel sRNAs

Plants produce stress-responsive sRNAs in response to biotic and abiotic stressors. Due to the specificity of production of these sRNAs (Sunkar *et al.*, 2012), novel prediction is required to uncover those present in the sRNA-Seq libraries. To this aim we use miRDeepP (Yang and Li, 2011), an open-source software designed specifically for predicting sRNAs in plants. mirDeepP comprises nine different Perl scripts. The critical parameters required at some of the stages of this analysis have been included in the parameters supplied by the user through the configuration file. The reader is referred to (Yang and Li, 2011) for further information on the distinct stages, and the parameters required. The predicted sRNA sequences are further filtered by mapping to miRBase to discard known miRNAs, while unmapped sequences are retained as potential novel sRNAs.

Genome-wide Host and Pathogen sRNA Profiling

To differentiate between host and pathogen(s) sRNAs, the raw reads are mapped against the corresponding reference genomes supplied by the user. The indexing of the genomes and the subsequent mapping is done by the mapping software and does not require any further action by the user. The sequences are mapped first to the host reference genome, and only reads that fail to map to the reference genome are mapped to the pathogen reference genome. This step provides the mapping profiles of each sRNA-Seq library, reporting the number of reads mapped to the reference genomes (host and pathogen), the distribution of the sequence lengths of interest defined by the user in the configuration file across all the libraries, and the loci of each sRNA on the reference genome. A repository of pathogens genome assemblies can be supplied as a reference in multi-fasta file format for a comprehensive pathogen sRNA profiling.

Input

To use Ds-Seq, it is expected that the NGS sequences have passed quality control and are deemed suitable for further downstream analysis. The input files to the pipeline consist mainly of (i) a configuration text file and (ii) data, including the raw single-ended sRNA sequence reads in fastq or zipped fastq formats (.gz),

Table 2. List of third-party tools used in the pipeline and their sources.

Usage	Tool	Source
Differential Expression	EdgeR	https://bioconductor.org/packages/release/bioc/html/edgeR.html
Sequence Alignment	Bowtie	http://bowtie-bio.sourceforge.net/index.shtml
Novel miRNA Prediction	miRDeep-P	http://sourceforge.net/projects/mirdp/
MFE: Vienna RNA Package (RNA fold)	RNAfold	https://www.tbi.univie.ac.at/RNA/
Plots	GGPlot2	https://cran.r-project.org/web/packages/ggplot2/index.html

annotation file corresponding to the genome assembly build, chromosome length file and reference sequences (Table 2). The pipeline can parse unzipped input fastq.gz files without requiring any pre-processing. A validation module checks and validates all the input parameters from the configuration and input files. If an error is flagged, the analysis does not commence, and a report is written to a log file.

Output

The results of the different stages of the analysis are organised into subfolders based on the modules of the pipeline, and all the sub-folders are presented in a single output folder. A schematic showing the organisation of the pipeline output is shown in Supplementary Figure S1[3], while a more detailed description is presented in Supplementary Table 13[4]. The output files are formatted as tab-delimited flat files to make it easy to port the results to other tools for further downstream analysis. Results produced include the profile of sRNA in the samples highlighting the host and pathogen sequences, the differential expression profile between libraries, conserved sRNAs, and novel predicted sRNAs. Publication-ready plots of the sRNA length distribution across libraries and frequency of the nucleotides at the 5' positions of the reads are produced. For differential expression analysis, volcano plots, a multidimensional scaling plot (MDS) and a biological coefficient of variation (BCV) (McCarthy *et al.*, 2012) plot are also produced.

Validation

Ds-Seq was validated using two publicly available datasets obtained from host-pathogen interaction studies with published results.

(i) Study A: sRNA-seq data obtained from the work of (Yang *et al.*, 2018) where the regulatory roles of Rice Stripe Virus (RSV)-derived small interfering RNAs were investigated in rice. Although this work entailed comparative sRNA data analysis from rice and the insect vector *Laodelphax striatellus*, we focused on only the RSV interaction with rice. Two samples of mock and RSV-infected rice in two replicates each were used. The data was obtained from the National Center for Biotechnology Information (NCBI), accession number GSE113555.

(ii) Study B: NGS data obtained in a study of sRNAs from the interaction between Turnip Mosaic Virus (TuMV) and the Tanto and Drakkar cultivars of oilseed rape *Brassica napus* (Pitzalis *et al.*, 2020). Only the data from the Drakkar cultivar was used for testing Ds-Seq. The sRNA-seq data in this study were retrieved from the NCBI Sequence Read Archive (SRA) using accession code PRJNA508739.

Pipeline Availability

Ds-Seq can be downloaded as scripts from the GitHub repository[5] and run on a UNIX/Linux operating system or obtained as a docker image from Docker hub with the ID cephas/ds-seq[6]. Further details of how to download and use the pipeline are contained in the README.md file on the GitHub repo. A user manual describing how to use the pipeline is available for download at the Github repository.

Technical Information

The versions of the software used in the Docker container for this pipeline and under which it has been tested are listed as follows:

1. Ubuntu - 18.04
2. R-base - 3.6.1
3. Perl - 5.26.1
4. Python - 2.7.17
5. RNAfold 2.4.11

Information on third-party tools used in the pipeline is presented in Table 2. We recommend running the pipeline using the Docker container since all the software are already bundled with the image.

However, the pipeline can still be run as scripts on a UNIX/Linux machine if an environment with these versions of tools is provided. This is not to say that other versions of these software will not work, but that the pipeline has performed correctly in an environment with these specific software versions.

Supporting data

The two published datasets used for testing the pipeline were obtained from the National Center for Biotechnology Information (NCBI) Gene Expression

[3]http://journal.embnet.org/index.php/embnetjournal/article/downloadSuppFile/1037/1037_supp_fig

[4]http://journal.embnet.org/index.php/embnetjournal/article/downloadSuppFile/1037/1037_supp_tab

[5]https://github.com/CEPHAS-01/small-RNASeq.ngs

[6]https://hub.docker.com/r/cephas/ds-seq

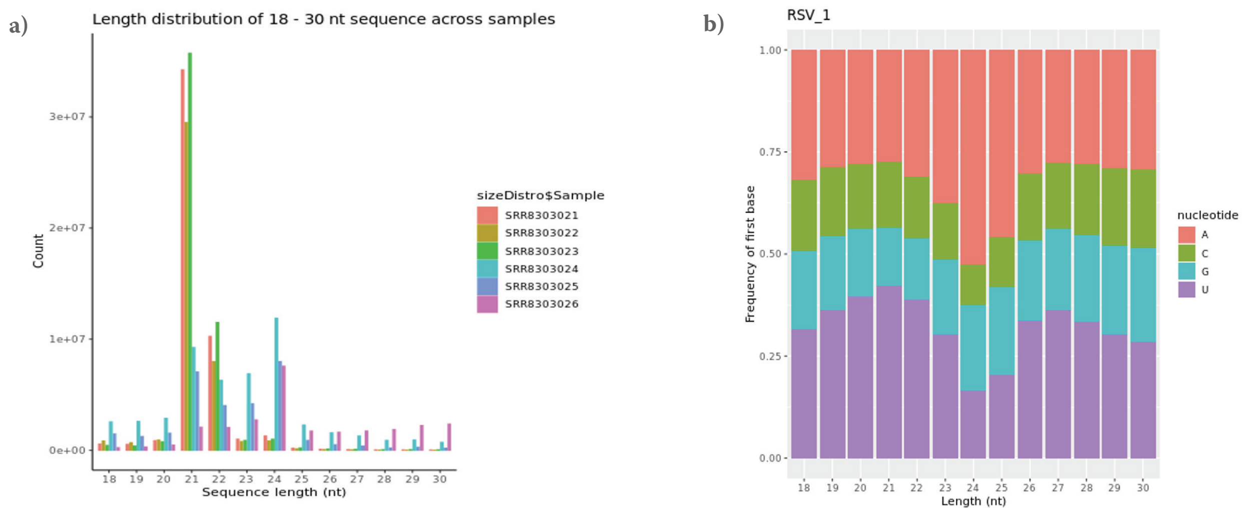


Figure 2. Examples of plots generated by Ds-Seq showing **(a)** the length distribution of the Mock and TuMV-infected Brassica napus sRNA reads across all libraries showing higher accumulation of 21-24 nt se-quences over other lengths in Study A. **(b)** Nucleotide frequency distribution for the sRNA sequence lengths in the two Rice RSV-infected samples showing dominance of nucleotide U at 21 nt and A at 24 nt positions for Study B.

Omnibus (GEO) database for Study A using accession GSE113555, and Sequence Read Archive (SRA) for Study B using accession PRJNA508739. All the data produced by the pipeline have been reported in this paper and the supplementary files. A test data for the pipeline will be made available for review on request pending the time that it would be hosted permanently in a public repository.

Results

Ds-Seq was tested with two different publicly available datasets, as highlighted in the validation section. For each test, the versions of the public dataset and the parameters used for the analysis by the authors in their study were replicated as much as possible in the pipeline.

For Study A (Yang *et al.*, 2018), the average cleaned reads produced with Ds-Seq for the Mock and RSV-treated samples were 13,664,509 and 15,206,600, respectively, compared to 13,876,277 and 15,400,564 reads reported in the study for the same data (Supplementary Table 1[7]). Furthermore, the average percentage of host-mapped reads obtained with Ds-Seq was 69.92% for Mock samples and 66.73% for RSV-treated samples while in contrast, 64.90% and 59.92% were reported respectively for Mock and RSV samples in Study A (Supplementary Table 2[7]). Negligible reads were reportedly mapped to the RSV genome for Mock samples (0.00%) by Ds-Seq and in Study A, while for RSV-treated samples 0.54% and 0.29% reads were reported mapped to RSV genome by Ds-Seq and in Study A respectively (Supplementary Table 2[7]).

The sRNA reads length distribution produced by Ds-Seq showed an accumulation of 21-24 nt sRNAs over other reads length (Supplementary Figure S3[8]), in agreement with the report in Study A. The distribution of the nucleotides at each position of the reads in the libraries produced by the pipeline revealed that at the 5' end, most of the 21nt sRNAs had Uracil (U) while most of the 24 nt sRNAs had Adenosine (A) (Figure 1B and Supplementary Figure S2[8]), consistent with the report in Study A. Differential expression profiles upon virus infection as reported in the study showed that 450 and 1,558 sRNAs were upregulated and downregulated respectively (Yang *et al.*, 2018), compared with 3789 and 1924 sRNAs identified by Ds-Seq at $|logFC| \geq 2.0$ and adjusted p-value < 0.05 (Supplementary Tables 7[7] and 8[7]). Although not reported in Study A, Ds-Seq identified 628 (167), 687 (172), 925 (178) and 749 (176) conserved miRNAs (Rice-specific conserved miRNAs) in the Mock1, Mock2, RSV1 and RSV2 samples respectively (Supplementary Table 1[7]). Ds-Seq also reported 81 and 115 sRNAs from the Mock and RSV samples, respectively, as likely novel sRNAs predicted by miRDeep-P (Supplementary Tables 9[7] and 10[7]), which were also not mentioned in Study A.

In Study B (Pitzalis *et al.*, 2020), the percentage of cleaned reads produced by Ds-Seq was between 48.2% to 74.8% for the Mock samples and 83.5% and 95.4% for the infected samples (Supplementary Table 5[7]). In comparison, 48.3% to 74.5% (Supplementary Table 5[7]) and 83.8% to 96.0% were reported in Study B for the Mock and infected samples respectively. For clean reads mapped to the Drakkar transcriptome, Ds-Seq reported 48.4% to 51.9% in the Mock samples, and 19.7% to 19.9% in the infected samples, as well as 0.01% to 0.03% and 30.7% to 33.6% of the reads, mapped to the virus genome in the Mock and virus-infected samples respectively

[7]http://journal.embnet.org/index.php/embnetjournal/article/downloadSuppFile/1037/1037_supp_tab

[8]http://journal.embnet.org/index.php/embnetjournal/article/downloadSuppFile/1037/1037_supp_fig

Table 3. Comparison of bioinformatics tools and parameters used at specific stages of the analysis by Ds-Seq and the independent investigation of the published datasets used to test Ds-Seq.

Analysis	Adapter removal	Reads mapping (parameters)	Differential expression (parameters)	Conserved miRNAs reference
Ds-Seq	In-house Perl script	Bowtie (Langmead *et al.*, 2009) (-a --best)*	edgeR (Robinson *et al.*, 2009; McCarthy *et al.*, 2012) (p-value < 0.05)	mirBase (Kozomara *et al.*, 2019)
Study A (Yang *et al.*, 2018)	Cutadapt (Martin, 2011)	Bowtie (Langmead *et al.*, 2009))	edgeR (Robinson *et al.*, 2009; McCarthy *et al.*, 2012) (adjusted p-value < 0.05)	
Study B (Pitzalis *et al.*, 2020)	Cutadapt (Martin, 2011)	Bowtie2 (Langmead and Salzberg, 2012)) (-t -N 0–end- to-end–very-sensitive–score-min C,0,0)	DESeq2 (Love *et al.*, 2014) (>= 150 mean reads and p-value < 0.05)	mirBase (Kozomara *et al.*, 2019) and other sources described in the publication.

(Supplementary Table 6[9]). This is compared with Study B, where 41.9% to 48.6% of the Mock samples mapped to the Drakkar (host) transcriptome and 13.4% to 15.3% mapped to the virus genome (Supplementary Table 6[9]). Similarly, 0.01% to 0.06% of the cleaned reads mapped to the virus genome in the Mock sample, while in the infected samples, 64.56% to 68.23% of the reads mapped to the virus (Supplementary Table 6[9]). The sRNA profile generated by Ds-Seq showed a general accumulation of the 21-24 nt reads more than other lengths. Still, more 24 nt reads were recorded in the mock-treated samples while 21-22 nt reads were the primary sizes in the infected plants (Figure 2A), as was reported in Study B.

The number of differentially expressed sRNA reads was reported from Ds-Seq as 49,426 and 18,255 upregulated and downregulated, respectively, at $|logFC| \geq 1.5$ and pValue < 0.05 (Supplementary Tables 11[9] and 12[9]). In Study B, differential expression was reported for only conserved miRNAs present in the samples with at least ten mean normalized reads showing that about 150 miRNAs were upregulated and 90 were downregulated (Pitzalis *et al.*, 2020). For the conserved miRNAs from miRBase Ds-Seq reported between 575 and 1302 conserved miRNAs compared to 1047 identified in Study B, while those specific to B. napus ranged from 64 to 73 miRNAs across both mock and infected samples (Supplementary Table 5[9]). Some plots from Ds-Seq from the analysis of data from Study A and Study B are shown in Figure 2.

Discussion

sRNA profiling has emerged as an essential step in studies regarding the control of gene expression in plants due to the implication of sRNAs in various molecular mechanisms by which the growth and develop-ment of plant species are regulated. These multiple mechanisms underscore the importance of a pipeline for sRNA profiling as a prime step toward discovering the molecular underpinnings of gene regulation involving sRNAs. In host-pathogen interaction, sRNAs play a significant role in the defense response of the host to pathogen invasion. The profiling of the sRNA landscape in host-pathogen interaction using NGS data opens a path toward the identification of sRNAs involved in the host defense and to further elucidate the molecular mechanism of the host defense response or pathogen action (Yang *et al.*, 2018; Pitzalis *et al.*, 2020). sRNAs of interest can then be extracted for further downstream analysis, such as identification of the target genes, co-regulation or DNA methylation.

In this paper, we introduce Ds-Seq, an integrated in silico pipeline for sRNA studies in plant host-pathogen interaction. The pipeline performance and utility were tested on two publicly available NGS data sets from which independent genome-wide sRNA analysis results have been published. The results of testing the automated pipeline generally showed agreement with the reported results of the independent analysis of the two published datasets (Yang *et al.*, 2018; Pitzalis *et al.*, 2020).

Ds-Seq identified a lower number of host-specific conserved miRNAs than the reported figures in the independent analysis, which could be attributed to the slight differences in tools with the runtime options applied at specific stages of the analyses.

In fact, two factors were noted regarding the difference with the results of the analysis of the conserved miRNA in Study B. Firstly, reads mapping was done using Bowtie2 (Langmead and Salzberg, 2012) with further restrictive parameters to filter the output, compared to the use of Bowtie (Langmead *et al.*, 2009) with all alignments reported in the pipeline (Table 3). Secondly, in Ds-Seq, only *B. napus* was used as the conserved miRNAs of interest from miRbase. In contrast, in the analysis of (Pitzalis *et al.*, 2020), the pool of conserved miRNAs was drawn from *B. napus*, *B. rapa*, *B. oleracea*, *A. lyrate* and *A. thaliana* (Table 3). The configuration file used for Ds-Seq can take only one plant name as an argument for the conserved miRNA of interest and could not capture the multiple plant names

[9]http://journal.embnet.org/index.php/embnetjournal/article/downloadSuppFile/1037/1037_supp_tab

Olagunju et al. (2024) EMBnet.journal 29, e1037
http://dx.doi.org/10.14806/ej.29.0.1037

specified in Study B . This would be addressed in future updates to the pipeline to permit specifying multiple names of plants of interest.

A difference was also recorded between the number of differentially expressed sRNAs in study A and the results from the pipeline. A total of 384 and 181 sRNAs were reported to be upregulated and downreg-ulated respectively by the pipeline using the parameter p-value < 0.05. This was different from the reported values in the independent study, where differentially expressed sRNAs were defined using an adjusted p-value < 0.05. A further step was taken on the results from the pipeline to extract differentially expressed sRNAs using adjusted p-value as was used in the independent study of Yang *et al.* (2018) to have a common basis for comparison. The result still showed a disparity in the pipeline-reported values and the independent analysis results, which could be due to omission in reporting some steps in the analysis or some parameters used in the separate study published by the authors.

Furthermore, a high number of differentially expressed sRNAs were reported by Ds-Seq with the data from study B. EdgeR (Robinson, 2009; McCarthy *et al.*, 2012) was used in the pipeline and differentially expressed reads were defined by the user through the configuration file using a fold change cut-off and p-value. In Pitzalis *et al.* (2020), however, DESeq2 (Love *et al.*, 2014) was used with extra parameters based on the number of reads used to filter the results. Therefore, using different tools (Maza, 2016; Costa-Silva *et al.*, 2017) could be responsible for the observed disparity in reported differentially expressed sRNAs.

The slight differences in the observed results will be addressed in the future release of Ds-Seq, especially those due to the use of different tools. We have plans to include more tools at various stages of the analysis to present users with options to choose from.

For differential expression analysis, the pipeline has been tested successfully on samples with two and three biological replicates each but can support samples with more than three biological replicates. The next iteration of Ds-Seq will accommodate other alignment software for the reads mapping module and integration of a module to predict the gene targets of the sRNAs.

Importantly, to ensure the reproducibility of the results (Baker and Penny, 2016), the pipeline can be containerized from its Docker image obtainable from the Docker hub repository. Application containerization with Docker[10] is employed to provide a consistent and self-contained software environment for running applications. Only the required dependencies of an application are installed before deployment to a host operating system. Docker containers have been demonstrated to promote reproducibility of genomic data analysis with minimal adverse effects on the outcome (Di Tommaso *et al.*, 2015). Aside from ensuring a consistent environment for applications, this approach also eliminates software version compatibility issues in application deployment.

[10]https://www.docker.com

Conclusions

In this paper, we have presented Ds-Seq, an automated pipeline for in silico analysis of sRNAs which can be run from a UNIX/Linux machine command line interface or as a Docker container. The pipeline combines several open-source sRNA analysis tools, in-house Perl, Python and R scripts, into a single shell script that runs end-to-end sRNA analysis from sRNA-Seq libraries with a single command. Although Ds-Seq was designed and applied to the study of plant host-pathogen interaction, it can also be adapted to other studies that seek to investigate the influence of abiotic factors on any plant with an available genome assembly. We have demonstrated the pipeline's capabilities using two publicly available datasets, and the results obtained generally indicated agreement with the published results from the same datasets, although with some differences attributable to the use of different tools and analysis parameters. The observed differences between the published dataset and the results of this pipeline underscore the challenge of reproducibility of analysis results which can be eliminated with the use of a predefined analysis pipeline such as Ds-Seq in a Docker environment.

Key Points

- An in-silico automated end-to-end analysis pipeline, Ds-Seq, is presented for small RNA profiling from NGS data in a host-pathogen interaction scenario.
- The pipeline has been tested with publicly available published datasets from host-pathogen interaction studies and the results showed general agreement with those obtained from the independent analysis of the datasets used.
- Ds-Seq containerization with Docker image promotes reproducibility of results through a consistent software environment.
- The pipeline features a modular structure that enables a user to choose modules of interest through a configuration file.
- Legible reports and results are presented as tab-delimited flat files for portability to other tools for further downstream analysis and publication-ready plots.

Funding

This work was supported by the Bill and Melinda Gates Foundation through the grant INV-008053 "Metabolic Engineering of Carbon Pathways to Enhance Yield of Root and Tuber Crops" and Roots and Tubers Pro-gramme (RTB) of the Consultative Group on International Agricultural Research (CGIAR).

Data Availability Statement: The two published datasets used for testing the pipeline were obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database for Study A using accession GSE113555, and Sequence Read Archive (SRA) for Study B using accession PRJNA508739. All the data produced by the pipeline have been reported in this paper and the supplementary files.

A test data for the pipeline will be made available for review on request pending the time that it would be hosted permanently in a public repository.

Acknowledgments

The authors are grateful to Yinka Fakunle and his ICT team for their help in maintaining the bioinformatics infrastructure.

References

Afgan,E., Baker,D., Batut,B., van den Beek \,M., Bouvier,D., Cech,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. (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. **46**(W1):W537-W544. http://dx.doi.org/10.1093/nar/gky379. PMID: 29790989; PMCID: PMC6030816

Aparicio-Puerta,E., Lebrón,R., Rueda,A., Gómez-Martín,C., Giannoukakos,S., Jaspez,D., Medina,J.M., Zubkovic,A., Jurak,I., Fromm,B., Marchal,J.A., Oliver,J., Hackenberg,M. (2019) sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. Nucleic Acids Res. **47**(W1):W530-W535. http::/dx.doi.org/10.1093/nar/gkz415. PMID: 31114926; PMCID: PMC6602500.

Baker,M. and Penny,D. (2016) Is there a reproducibility crisis? Nature, **533**, 452–454. http://dx.doi.org/10.1038/d41586-019-00067-3

Beaulieu-Jones,B.K. and Greene,C.S. (2017) Reproducibility of computational workflows is automated using continuous analysis. Nat. Biotechnol., **35**, 342–346. http://dx.doi.org/10.1038/nbt.3780

Beckers,M., Mohorianu,I., Stocks,M., Applegate,C., Dalmay,T., Moulton,V. (2017) Comprehensive processing of high-throughput small RNA sequencing data including quality checking, normalization, and differential expression analysis using the UEA sRNA Workbench. RNA. 2017 Jun;**23**(6):823-835. http://dx.doi.org/10.1261/rna.059360.116. Epub 2017 Mar 13. PMID: 28289155; PMCID: PMC5435855.

Carbonell,A., Lisón,P. and Daròs,J.A. (2019) Multi-targeting of viral RNAs with synthetic trans-acting small interfering RNAs enhances plant antiviral resistance. Plant J., **100**, 720–737 http://dx.doi.org/10.1111/tpj.14466

Costa-Silva,J., Domingues,D. and Lopes,F.M. (2017) RNA-Seq differential expression analysis: An ex-tended review and a software tool. PLoS One, **12**, 1–18. http://dx.doi.org/10.1371/journal.pone.0190152

Diezma-Navas,L., Pérez-González,A., Artaza,H., Alonso,L., Caro,E., Llave,C. and Ruiz-Ferrer,V. (2019) Crosstalk between epigenetic silencing and infection by tobacco rattle virus in Arabidopsis. Mol. Plant Pathol., **20**, 1439–1452. http://dx.doi.org/10.1111/mpp.12850

Di Tommaso,P., Palumbo,E., Chatzou,M., Prieto,P., Heuer,M.L. and Notredame,C. (2015) The impact of Docker containers on the performance of genomic pipelines. PeerJ, 2015, 1–10. http://dx.doi.org/ 10.7717/peerj.1273

Hu,T., Wang,T., Li,H., Wassie,M., Xu,H. and Chen,L. (2020) Genome-wide small RNA profiling reveals tiller development in tall fescue (Festuca arundinacea Schreb). BMC Genomics, **21**, 1–13. http://dx.doi.org/10.1186/s12864-020-07103-x

Joppich,M. and Zimmer,R. (2019) From command-line bioinformatics to bioGUI. PeerJ, 2019. http://dx.doi.org/10.7717/peerj.8111

Kalvari,I., Nawrocki,E.P., Ontiveros-Palacios,N., Argasinska,J., Lamkiewicz,K., Marz,M. (2021) Rfam 14: expanded coverage of metagenomics, viral and microRNA families. Nucleic Acids Res. **49**, D192–D200. http://dx.doi.org/10.1093/nar/gkaa1047

Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) MiRBase: From microRNA sequences to function. Nucleic Acids Res., **47**, D155–D162 http://dx.doi.org/10.1093/nar/gky1141

Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol., **10**. http://dx.doi.org/10.1186/gb-2009-10-3-r25

Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods, **9**, 357–359. http://dx.doi.org/10.1038/nmeth.1923

Li,A., Li,G., Zhao,Y., Meng,Z., Zhao,M., Li,C., Zhang,Y., Li,P., Ma,C. Le, Xia,H., *et al.* (2018) Combined small RNA and gene expression analysis revealed roles of miRNAs in maize response to rice black-streaked dwarf virus infection. Sci. Rep., **8**, 1–14. http://dx.doi.org/10.1038/s41598-018-31919-z

List,M., Ebert,P. and Albrecht,F. (2017) Ten Simple Rules for Developing Usable Software in Computa-tional Biology. PLoS Comput. Biol., **13**, 1–5. http://dx.doi.org/10.1371/journal.pcbi.1005265

Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol., **15**, 550. http://dx.doi.org/10.1186/s13059-014-0550-8

Mangul,S., Mosqueiro,T., Abdill,R.J., Duong,D., Mitchell,K., Sarwal,V., Hill,B., Brito,J., Littman,R.J., Statz,B., *et al.* (2018) Challenges and recommendations to improve installability and archival stability of omics computational tools. bioRxiv. http://dx.doi.org/10.1101/452532.

Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMB-net.journal, **17**, 10–12. http://dx.doi.org/10.14806/ej.17.1.200

Maza,E. (2016) In papyro comparison of TMM (edgeR), RLE (DESeq2), and MRN normalization methods for a simple two-conditions-without-replicates RNA-seq experimental design. Front. Genet., **7**, 1–8. http://dx.doi.org/10.3389/fgene.2016.00164

Morais,D., Roesch,L.F.W., Redmile-Gordon,M., Santos,F.G., Baldrian,P., Andreote,F.D. and Pylro,V.S. (2018) BTW-Bioinformatics Through Windows: An easy-to-install package to analyze marker gene data. PeerJ, 2018, 10–16. http://dx.doi.org/10.7717/peerj.5299

McCarthy,D.J., Chen,Y. and Smyth,G.K. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res., **40**, 4288–4297. http://dx.doi.org/10.1093/nar/gks042

Mueth,N.A., Ramachandran,S.R. and Hulbert,S.H. (2015) Small RNAs from the wheat stripe rust fungus (Puccinia striiformis f.sp. tritici). BMC Genomics, **16**, 1–16. http://dx.doi.org/10.1186/s12864-015-1895-4

Pitzalis,N., Amari,K., Graindorge,S., Pflieger,D., Donaire,L., Wassenegger,M., Llave,C. and Heinlein,M. (2020) Turnip mosaic virus in oilseed rape activates networks of sRNA-mediated interactions between viral and host genomes. Commun. Biol., **3**, 1–16. http://dx.doi.org/10.1038/s42003-020-01425-y

Pogorelcnik,R., Vaury,C., Pouchin,P., Jensen,S. & Brasset,E. (2018) sRNAPipe: A Galaxy-based Pipeline for Bioinformatic in-depth Exploration of Small RNAseq Data. Mobile DNA, Vol. 9, No. 25. http://dx.doi.org/10.1186/s13100-018-0130-7.

Pumplin,N. and Voinnet,O. (2013) RNA silencing suppression by plant pathogens: Defence, coun-ter-defence and counter-counter-defence. Nat. Rev. Microbiol., **11**, 745–760. http://dx.doi.org/10.1038/nrmicro3120

Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, **26**, 139–140. http://dx.doi.org/ 10.1093/bioinformatics/btp616

Rueda,A., Barturen,G., Lebrón,R., Gómez-Martín,C., Alganza,Á., Oliver,J.L., & Hackenberg,M. (2015) sRNAtoolbox: an integrated collection of small RNA research tools. Nucleic acids research, **43**(W1), W467–W473. http://dx.doi.org/10.1093/nar/gkv555

Seemann,T. (2013) Ten recommendations for creating usable bioinformatics command line software. Gi-gascience, **2**, 2–4. http://dx.doi.org/0.1186/2047-217X-2-15

Smith,D.R. (2013) The battle for user-friendly bioinformatics. Front. Genet., **4**, 1–2. http://dx.doi.org/10.3389/fgene.2013.00187

Smith,D.R. (2014) Buying in to bioinformatics: An introduction to commercial sequence analysis software. Brief. Bioinform., **16**, 700–709. http://dx.doi.org/10.1093/bib/bbu030

Sorefan,K., Pais,H., Hall,A.E., Kozomara,A., Griffiths-Jones,S., Moulton,V. and Dalmay,T. (2012) Re-ducing ligation bias of small RNAs in libraries for next generation sequencing. Silence, **3**, 1–11. http://dx.doi.org/10.1186/1758-907X-3-4

Stodden,V., Seiler,J. and Ma,Z. (2018) An empirical analysis of journal policy effectiveness for compu-tational reproducibility. Proc. Natl. Acad. Sci. U. S. A., **115**, 2584–2589. http://dx.doi.org/ 10.1073/pnas.1708290115.

Sunkar,R., Li,Y.F. and Jagadeeswaran,G. (2012) Functions of microRNAs in plant stress responses. Trends Plant Sci., **17**, 196–203. http://dx.doi.org/10.1016/j.tplants.2012.01.010

Yang,X. and Li,L. (2011) miRDeep-P: A computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics, **27**, 2614–2615. http://dx.doi.org/10.1093/bioinformatics/btr430

Tamiru,M., Hardcastle,T.J. and Lewsey,M.G. (2018) Regulation of genome-wide DNA methylation by mobile small RNAs. New Phytol., **217**, 540–546. http://dx.doi.org/10.1111/nph.14874

Wang,C., Wang,C., Xu,W., Zou,J., Qiu,Y., Kong,J., Yang,Y., Zhang,B. and Zhu,S. (2018) Epigenetic changes in the regulation of Nicotiana tabacum response to cucumber mosaic virus infection and symptom recovery through single-base resolution methylomes. Viruses, **29**;10(8):402. http://dx.doi.org/ 10.3390/v10080402

Wenlei,C., Xinxin,C., Jianhua,Z., Zhaoyang,Z., Zhiming,F., Shouqiang,O. and Shimin,Z. (2020) Com-prehensive Characteristics of MicroRNA Expression Profile Conferring to Rhizoctonia solani in Rice. Rice Sci., **27**, 101–112. http://dx.doi.org/10.1016/j.rsci.2019.04.007

Won,S.Y., Yumul,R.E. and Chen,X. (2014) Small RNAs in Plants: In Molecular Biology, Pp. 95-127. http://dx.doi.org/10.1007/978-1-4614-7570-5_3

Xu,G., Strong,M.J., Lacey,M.R., Baribault,C., Flemington,E.K. and Taylor,C.M. (2014) RNA CoMPASS: A dual approach for pathogen and host. PLoS One, **9**. http://dx.doi.org/10.1371/journal.pone.0089445

Yang,M., Xu,Z., Zhao,W., Liu,Q., Li,Q., Lu,L., Liu,R., Zhang,X. and Cui,F. (2018) Rice stripe vi-rus-derived siRNAs play different regulatory roles in rice and in the insect vector Laodelphax striatellus. BMC Plant Biol., **18**(1):219. http://dx.doi.org/10.1186/s12870-018-1438-7

Zhang,B., Pan,X., Cannon,C.H., Cobb,G.P. and Anderson,T.A. (2006) Conservation and divergence of plant microRNA genes. Plant J., **46**(2):243-59. http://dx.doi.org/10.1111/j.1365-313X.2006.02697.x

EMBnet.journal
Bioinformatics in Action

Papakonstantinou et al. (2024) EMBnet.journal 29, e1048
http://dx.doi.org/10.14806/ej.29.0.1048

Research Papers

Fingerprinting Breast Milk; insights into Milk Exosomics

Eleni Papakonstantinou[1,2], Konstantina Dragoumani[1], Antonia Mataragka[1], Flora Bacopoulou[2], Christos Yapijakis[2], Nikolaos AA Balatsos[3], Katerina Pissaridi[4], Dimitris Ladikos[4], Aspasia Eftymiadou[5], George Katsaros[5], Evangelos Gikas[6], Pantelis Hatzis[7], Martina Samiotaki[8], Michalis Aivaliotis[9], Vasileios Megalooikonomou[10], Antonis Giannakakis[11], Costas Iliopoulos[12], Erik Bongcam-Rudloff[13], Sofia Kossida[14], Elias Eliopoulos[1], George P Chrousos[2]✉, Dimitrios Vlachakis[1,2,12]✉

[1]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece

[2]University Research Institute of Maternal and Child Health & Precision Medicine, and UNESCO Chair on Adolescent Health Care, National and Kapodistrian University of Athens, "Aghia Sophia" Children's Hospital, Athens, Greece

[3]Department of Biochemistry and Biotechnology, University of Thessaly, Biopolis, Larissa, Greece

[4]YIOTIS S.A., Athens, Greece

[5]Institute of Technology of Agricultural Products, Hellenic Agricultural Organization "DEMETER", Lykovrisi, Greece

[6]Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Athens, Greece

[7]Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, Athens, Greece

[8]Institute for Bioinnovation, Biomedical Sciences Research Center "Alexander Fleming", Vari, Greece

[9]Basic and Translational Research Unit, Special Unit for Biomedical Research and Education, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

[10]Multidimensional Data Analysis and Knowledge Management Laboratory, Computer Engineering and Informatics Department, School of Engineering, University of Patras, Patras, Greece

[11]Laboratory of Gene Expression, Molecular Diagnostics and Modern Therapeutics, Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece

[12]School of Informatics, Faculty of Natural & Mathematical Sciences, King's College London, London, U.K.

[13]Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

[14]IMGT®, the international ImMunoGenetics information system®, Laboratoire d'ImmunoGénétique Moléculaire LIGM, Institut de Génétique Humaine, (IGH), Centre National de la Recherche Scientifique (CNRS), Université de Montpellier (UM), Montpellier, France

Competing interests: EP none; KD none; AM none; FB none; CY none; NAB none; KP none; DL none; AE none; GK none; EG none; PH none; MS none; MA none; VM none; AG none; CI none; EBR is a member of EMBnet.journal Editorial Board; SK none; EE none; GPC none; DV is a member of EMBnet.journal Editorial Board

Abstract

Breast milk, often referred to as "liquid gold," is a complex biofluid that provides essential nutrients, immune factors, and developmental cues for newborns. Recent advancements in the field of exosome research have shed light on the critical role of exosomes in breast milk. Exosomes are nanosized vesicles that carry bioactive molecules, including proteins, lipids, nucleic acids, and miRNAs. These tiny messengers play a vital role in intercellular communication and are now being recognized as key players in infant health and development. This paper explores the emerging field of milk exosomics, emphasizing the potential of exosome fingerprinting to uncover valuable insights into the composition and function of breast milk. By deciphering the exosomal cargo, we can gain a deeper understanding of how breast milk influences neonatal health and may even pave the way for personalized nutrition strategies.

Introduction

Breastfeeding has long been recognized as the "gold standard" of infant nutrition, providing not only essential nutrients but also numerous bioactive molecules crucial for a newborn's growth and development. While the composition of breast milk has been extensively studied, recent research has unveiled a new layer of complexity

Article history
Received: 26 September 2023
Accepted: 03 October 2023
Published: 22 May 2024

© 2024 Papakonstantinou *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at http://journal.embnet.org.

- milk exosomes. Exosomes, small extracellular vesicles (EVs) measuring around 50-150 nanometers, are now known to be present in breast milk and are believed to play a pivotal role in infant health (Melnik *et al.*, 2021). Herein, we will delve into the field of milk exosomics, exploring what exosomes are, their potential significance in breast milk, and how fingerprinting these exosomes can provide valuable insights into milk composition and its effects on infant health.

Exosomes: Tiny Messengers with Big Potential

Exosomes constitute a subset of EVs secreted by various cell types, including epithelial cells, immune cells, and even mammary epithelial cells. These vesicles are enclosed by a lipid bilayer membrane and carry a cargo of proteins, lipids, nucleic acids, and microRNAs (miRNAs) from their parent cells (Mosquera-Heredia *et al.*, 2021). Exosomes serve as messengers between cells, facilitating intercellular communication and the transfer of information (Kalluri and LeBleu, 2020).

Exosome composition is highly dynamic and dependent on the originating cell type (Mathieu *et al.*, 2019). Common components of exosomes include proteins, lipids, nucleic acids, and other biomolecules, for example, exosomes contain a diverse array of proteins, including those involved in membrane transport and fusion (*e.g.*, tetraspanins), heat shock proteins, cytoskeletal proteins, and enzymes (Zhang *et al.*, 2019). Exosomal membranes are rich in cholesterol, sphingomyelin, and phosphatidylserine, contributing to exosome stability and uptake by target cells (Mosquera-Heredia *et al.*, 2021; Zhang *et al.*, 2019). Exosomes carry various forms of nucleic acids, including DNA, mRNA, and miRNA, which can be functional when transferred to recipient cells, influencing gene expression. Exosomes have diverse functions in physiological and pathological processes, modulating immune responses, promoting tissue repair, and contributing to cell signaling and homeostasis (Zhang *et al.*, 2019). Importantly, exosomes are implicated in various diseases, including cancer, neurodegenerative disorders, and inflammatory conditions (He *et al.*, 2023; Kogure *et al.*, 2013; Song *et al.*, 2015).

Exosomes in Breast Milk

Breast milk stands as a quintessential source of nutrition for infants, offering a tailored blend of nutrients, growth factors, immunoglobulins, and cells essential for their growth and immune development(Lessen and Kavanagh, 2015). Recent studies have unveiled the presence of a significant quantity of exosomes in breast milk, introducing a novel dimension to our understanding of its role in infant health (Galley and Besner, 2020).

These exosomes, characterized by their lipid bilayer membrane, encapsulate a cargo of bioactive molecules, including immune-related proteins and microRNAs (miRNAs), which hold promise in modulating various physiological processes in infants. The cargo of exosomes in breast milk is of particular interest. Exosomes in breast milk carry immune-related proteins and miRNAs, potentially aiding in the development of the infant's immune system.

Proteomic analysis of breast milk-derived exosomes has unveiled a diverse repertoire of proteins intricately involved in various cellular components, molecular functions, and biological processes, including immune-related processes such as defense response, phagocytosis, complement activation, and regulation of cytokine responses. Notably, these exosomes carry growth factors crucial for tissue development, such as the gastrointestinal tract (Hock *et al.*, 2017; Pisano *et al.*, 2020). Studies have illuminated the multifaceted role of breast milk EVs in modulating both innate and adaptive immune responses. They stimulate specific Toll-like receptors (TLRs) while concurrently tempering the response of endosomal TLRs, suggesting a regulatory role in immune activation (Kim *et al.*, 2023). Additionally, milk EVs have been shown to inhibit the activation and differentiation of CD4+ T cells, implying their involvement in regulating adaptive immune responses and facilitating immune maturation in infants (Zonneveld *et al.*, 2021).

Proteomic profiling of milk EVs across different mammal species and lactation stages has unraveled a diverse array of proteins with significant nutritional, enzymatic, hormonal, and immunomodulatory functions, and these proteins underpin the various properties of milk EVs, including their immunomodulatory effects and potential as biomarkers for assessing mammary gland health. Notably, the protein cargo of EVs exhibits dynamic changes during lactation stages, with colostrum EVs enriched in immune-related proteins and mature milk EVs harboring proteins involved in ribosome regulation and cell growth. Understanding the protein content of milk EVs provides valuable insights into milk's nutritional value and its potential applications in health assessment (Buratta *et al.*, 2023).

The lipid composition of milk EVs constitutes a pivotal determinant of their properties and biological effects. Enriched in sphingolipids and glycerophospholipids containing saturated fatty acids, EV membranes exhibit rigidity and stability in biological fluids (Tenchov *et al.*, 2022). These lipids also serve as precursors for bioactive molecules involved in immune signaling and inflammation. Despite abundant research on the nucleic acid cargo of milk EVs, information regarding their lipid content and biological roles remains limited. Initial investigations have revealed a shared lipid composition enriched in phosphatidylserine (PS) and sphingomyelin (SM) among milk EVs from bovine and human sources compared to milk fat globules (MFG). These findings underscore the resemblance of milk EVs to those isolated from other body fluids and cell culture media. Recent studies have delved into characterizing the lipid profiles of milk EVs, unveiling high levels of phosphatidylcholine (PC) and phosphatidylethanolamine (PE) in bovine milk

EVs, with PS enrichment compared to milk. Similarly, human milk EVs exhibit diverse lipid species, with discernible differences between preterm and term samples. Notably, these lipid variances may contribute to the prevention of conditions like necrotizing enterocolitis (NEC) through the modulation of signaling pathways (Buratta *et al.*, 2023).

Metabolomic analysis of human milk (HM) involves the identification and annotation of metabolites utilizing various databases and spectral libraries. Techniques such as NMR, LC-MS, CE-MS, and GC-MS facilitate the detection of a comprehensive array of metabolites in HM, including carbohydrates, fatty acids, amino acids, and organic acids. Common metabolites detected across multiple analytical platforms include glucose, lactose, amino acids like tyrosine, fatty acids such as capric acid/caprate and caprylic acid/caprylate, and organic acids like citric acid/citrate and pyruvic acid/pyruvate. These metabolites represent vital sources of energy, essential for infant growth and development, emphasizing the nutritional richness of human milk (Ramos-Garcia *et al.*, 2023). The lipidomic and metabolomic profile of milk EVs have the potential to unveil their diverse biological functions and potential therapeutic applications. Understanding these components provides valuable insights into the nutritional value of milk and its implications for infant health and development.

Breast milk exosomes also carry small RNAs, such as mRNAs, lncRNAs, siRNAs and more, which can influence gene expression and regulation in the infant (Melnik *et al.*, 2021). MicroRNAs (miRNAs) found in breast milk play a crucial role in regulating various physiological processes in infants. The selective packaging of miRNAs into exosomes suggests a regulated process rather than random selection, highlighting the potential significance of exosomes in early neonatal development. The composition of miRNAs in human breast milk varies depending on factors such as gestational age at delivery and lactation stage, with preterm milk exhibiting distinct miRNA profiles compared to term milk. Changes in miRNA expression patterns in preterm milk may reflect adaptations to promote neurodevelopment and mitigate the challenges faced by preterm infants, contributing to the maturation of the preterm infant's brain (Slyk-Gulewska *et al.*, 2023).

Furthermore, specific miRNAs detected in human breast milk target genes involved in synaptic components, neurogenesis, and neurodevelopmental processes, suggesting their potential roles in regulating neuronal function and connectivity. Dysregulation of these miRNAs has been implicated in various neurological disorders, highlighting their importance in maintaining proper brain function and development. Future research directions may include longitudinal studies to further investigate the evolution of miRNA cargo in human breast milk over lactation stages and its potential impact on infant neurodevelopment. Additionally, miRNA sequencing of the whole milk sample could provide insights into differences in miRNA composition between exosomal and non-exosomal fractions (Freiria-Martinez *et al.*, 2023).

Studies have examined the influence of different factors on the miRNA profile of breast milk and its implications for infant health. Research indicates that while the total concentration of miRNAs remains relatively stable during the first 6 months of lactation, significant changes occur in miRNA composition, particularly at 4 months compared to 2 months postpartum, suggesting an adaptation of miRNA profiles to meet the evolving needs of the infant. Additionally, mothers who deliver prematurely exhibit altered hormonal profiles and miRNA expression in their breast milk, potentially benefiting premature infants by influencing glucose homeostasis, adipogenesis, and immune function (Slyk-Gulewska *et al.*, 2023).

The delivery method also impacts miRNA expression in breast milk, with cesarean delivery disrupting the balance of specific miRNAs and potentially increasing the risk of type 2 diabetes mellitus (T2DM) later in life (Slyk-Gulewska *et al.*, 2023). Furthermore, correlations between maternal stress levels and changes in extracellular vesicle-derived miRNAs suggest potential impacts on fatty acid metabolism, steroid biosynthesis, and organ growth regulation in infants. Breast milk from overweight or obese mothers contains altered levels of miRNAs associated with infant growth and neurodevelopment, possibly contributing to an increased risk of childhood obesity due to lower levels of certain miRNAs. Regarding diabetes, breast milk from mothers with gestational diabetes mellitus (GDM) exhibits abnormal miRNA levels associated with metabolic outcomes in their infants. Similarly, mothers with type 1 diabetes mellitus (T1DM) demonstrate differences in miRNA composition in breast milk exosomes, potentially impacting immune response processes. Breastfeeding protects against T2DM development, partly due to the absence of certain miRNAs present in cow's milk, which may contribute to β-cell dysfunction and insulin resistance. Breastfeeding also facilitates epigenetic programming through exosome miRNAs, whereas infant formulas lack these bioactive components, potentially modulating gene expression related to obesity predisposition and reducing the risk of obesity and related diseases. Understanding the diverse roles of miRNAs in breast milk and their modulation by various factors holds promise for the prevention and treatment of various diseases in infants, including necrotizing enterocolitis (NEC), atopic diseases, diabetes, obesity, and cancer. Further research is imperative to elucidate the specific mechanisms and therapeutic potential of miRNAs in breast milk (Slyk-Gulewska *et al.*, 2023).

Human milk exosomes also contain various miRNAs associated with adipogenesis, the process of fat cell development, while concerns have arisen regarding the potential transfer of adipogenic miRNAs across species through milk consumption, particularly as humans consume milk from other animals throughout their lives (Abbas *et al.*, 2023). However, evidence

linking dairy milk consumption to increased obesity rates remains inconclusive. Effective biological action of milk exosomal miRNAs relies on sufficient quantities reaching target cells. Studies have found that the miRNA profiles in the milk of obese women differ, with lower levels of miRNA-148a and miRNA-30b observed. This variation may influence fat accumulation in breastfeeding infants. Certain milk exosomal miRNAs, such as miRNA-29, miRNA-148, miRNA-30b, and miRNA-125b, are believed to have epigenetic effects on recipients, potentially impacting gene expression and cellular function. Furthermore, milk exosomes possess the ability to traverse the gastrointestinal barrier, making them promising candidates for oral drug delivery. They can also be engineered and loaded with specific miRNAs involved in adipocyte differentiation, conversion, or browning, offering potential therapeutic benefits. By modifying the miRNA cargo of exosomes, it may be possible to enhance their health-promoting effects and provide an alternative to traditional pharmaceutical interventions (Abbas *et al.*, 2023).

Additional research has investigated the therapeutic potential of human milk exosomes in treating necrotizing enterocolitis (NEC) using a rat model, where both exosomes from full-term and preterm human breast milk were found to alleviate NEC severity, with preterm-derived exosomes showing better promotion of intestinal epithelial cell proliferation (Yan *et al.*, 2022). High-throughput sequencing revealed differential expression of long non-coding RNAs (lncRNAs) and messenger RNAs (mRNAs) between the exosome types, with associated pathways implicated in cell proliferation, such as such as the JAK-STAT signaling pathway, bile secretion, and the AMPK signaling pathway, highlighting the protective role of human milk exosomes against NEC and giving insight into potential therapeutic mechanisms (Yan *et al.*, 2022).

While the exact functions of breast milk exosomes are still being elucidated, several potential roles have been proposed. Exosomes may help educate the infant's immune system by delivering immune-related molecules and miRNAs. Growth factors carried by exosomes may contribute to the development and maturation of the infant's gastrointestinal tract and other tissues. Exosomes may play a role in shaping the infant's gut microbiota, which has a significant impact on overall health (Slyk-Gulewska *et al.*, 2023).

Fingerprinting Breast Milk Exosomes

Fingerprinting breast milk exosomes involves the comprehensive analysis of their cargo to understand their composition and potential effects on infant health. This process combines techniques from various scientific disciplines, including proteomics, genomics, metabolomics, lipidomics, and miRNA profiling (Shenker *et al.*, 2020). Proteomic analysis of breast milk exosomes involves the identification and quantification of the proteins they carry. Mass spectrometry-based approaches are commonly used to determine the protein composition, providing insights into the functional roles of exosomes in breast milk (Beck *et al.*, 2015). Genomic and miRNA analysis involves sequencing the nucleic acids carried by breast milk exosomes. This can reveal the specific miRNAs and genes involved in immune modulation, tissue development, and other processes crucial for infant health (Hunt *et al.*, 2011). Lipidomic and metabolomic analysis focuses on the lipid content and metabolite cargo of exosomes, respectively. This can uncover lipid signatures associated with breast milk exosomes, shedding light on their stability and potential interactions with recipient cells (Garwolinska *et al.*, 2017).

Fingerprinting breast milk exosomes has the potential to have significant implications for infant health and nutrition. By deciphering the exosomal cargo in breast milk, it may be possible to tailor breastfeeding advice and strategies to meet the specific needs of individual infants (Shenker *et al.*, 2020). Personalized nutrition based on exosomal profiles could optimize the health and development of newborns. Breast milk exosomes may serve as biomarkers for various maternal and infant health conditions (Mourtzi *et al.*, 2021). Changes in exosomal cargo could indicate infections, metabolic disorders, or other health issues in either the mother or the infant. Understanding the cargo of breast milk exosomes could lead to the development of novel therapeutic approaches. Exosomes could be isolated and used as targeted drug delivery vehicles, or their cargo could inspire the development of new therapies for infant health issues.

While the field of milk exosomics holds great promise, it also faces several challenges and areas for further research. Standardisation of exosome isolation and analysis methods is essential for reproducibility and comparability of research findings. Understanding the functional significance of breast milk exosomes and their cargo remains a priority (Mitsis *et al.*, 2020). In vitro and in vivo studies are needed to validate their roles in infant health, as well as longitudinal studies required to assess the long-term effects of breast milk exosomes on infant health and development. A holistic omics profiling of human milk exosomes, whilst further bioinformatics and AI analysis is in hand, will be able to advance our knowledge on their specific role and the personalized benefits for the afterbirth biological interplay between the mother and the infant (Papakonstantinou *et al.*, 2023). As exosome research advances, ethical considerations regarding the collection and use of breast milk samples for research purposes need to be addressed.

Conclusion

Breast milk exosomes represent a fascinating frontier in the field of neonatal nutrition and health. These tiny vesicles, carrying a cargo of proteins, lipids, and nucleic acids, have the potential to shape the development and

well-being of newborns. Fingerprinting breast milk exosomes through proteomic, genomic, and lipidomic analysis can provide invaluable insights into milk composition and its impact on infant health. As research in milk exosomics continues to advance, we may unlock the secrets of personalized nutrition for infants and discover new therapeutic avenues for addressing neonatal health challenges. While many questions remain unanswered, the future of milk exosomics holds great promise for improving the health and well-being of the youngest members of our society. Further research in this domain promises to uncover novel strategies for enhancing infant nutrition and mitigating health risks.

Key Points

- Exosomes in breast milk exhibit diverse cargo including proteins, metabolites, and nucleic acids, crucial for intercellular communication and immune modulation. Understanding their dynamic composition and roles in physiological processes provides insights into their significance in infant health and development.
- Breast milk exosomes carry immune-related proteins and microRNAs, contributing to the infant's immune system development. They stimulate specific immune receptors while regulating immune responses, suggesting a regulatory role in immune activation and maturation, thus highlighting their potential in enhancing infant immunity.
- The composition of microRNAs in breast milk exosomes varies with factors like gestational age and delivery method, potentially influencing neurodevelopment in infants, whilst dysregulation of these microRNAs has been linked to neurological disorders, underscoring the importance of breast milk exosomes in promoting proper brain function and development.
- Milk exosomes show promise in treating conditions like necrotizing enterocolitis (NEC) and as potential biomarkers for maternal and infant health conditions. They could be harnessed for targeted drug delivery or inspire the development of novel therapies, indicating their therapeutic potential in neonatal health.
- While milk exosome research offers exciting possibilities, standardisation of isolation methods and understanding their functional significance remain critical. Longitudinal studies are needed to assess their long-term effects on infant health, and ethical considerations regarding sample collection and use must be addressed as the field advances.

Funding

The authors would like to acknowledge funding from 'MilkSafe: A novel pipeline to enrich formula milk using omics technologies', a research co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T2EDK-02222).

References

Abbas MA, Al-Saigh NN and Saqallah FG (2023) Regulation of adipogenesis by exosomal milk miRNA. Rev Endocr Metab Disord **24**(2), 297-316. http://dx.doi.org/10.1007/s11154-023-09788-3

Beck KL, Weber D, Phinney BS, Smilowitz JT, Hinde K *et al.* (2015) Comparative Proteomics of Human and Macaque Milk Reveals Species-Specific Nutrition during Postnatal Development. J Proteome Res **14**(5), 2143-2157. http://dx.doi.org/10.1021/pr501243m

Buratta S, Urbanelli L, Tognoloni A, Latella R, Cerrotti G *et al.* (2023) Protein and Lipid Content of Milk Extracellular Vesicles: A Comparative Overview. Life (Basel) **13**(2). http://dx.doi.org/10.3390/life13020401

Freiria-Martinez L, Iglesias-Martinez-Almeida M, Rodriguez-Jamardo C, Rivera-Baltanas T, Comis-Tuche M *et al.* (2023) Human Breast Milk microRNAs, Potential Players in the Regulation of Nervous System. Nutrients **15**(14). http://dx.doi.org/10.3390/nu15143284

Galley JD and Besner GE (2020) The Therapeutic Potential of Breast Milk-Derived Extracellular Vesicles. Nutrients **12**(3). http://dx.doi.org/10.3390/nu12030745

Garwolinska D, Hewelt-Belka W, Namiesnik J and Kot-Wasik A (2017) Rapid Characterization of the Human Breast Milk Lipidome Using a Solid-Phase Microextraction and Liquid Chromatography-Mass Spectrometry-Based Approach. J Proteome Res **16**(9), 3200-3208. http://dx.doi.org/10.1021/acs.jproteome.7b00195

He A, Wang M, Li X, Chen H, Lim K *et al.* (2023) Role of Exosomes in the Pathogenesis and Theranostic of Alzheimer's Disease and Parkinson's Disease. Int J Mol Sci **24**(13). http://dx.doi.org/10.3390/ijms241311054

Hock A, Miyake H, Li B, Lee C, Ermini L *et al.* (2017) Breast milk-derived exosomes promote intestinal epithelial cell growth. J Pediatr Surg **52**(5), 755-759. http://dx.doi.org/10.1016/j.jpedsurg.2017.01.032

Hunt KM, Foster JA, Forney LJ, Schutte UM, Beck DL *et al.* (2011) Characterization of the diversity and temporal stability of bacterial communities in human milk. PLoS One **6**(6), e21313. http://dx.doi.org/10.1371/journal.pone.0021313

Kalluri R and LeBleu VS (2020) The biology, function, and biomedical applications of exosomes. Science **367**(6478). http://dx.doi.org/10.1126/science.aau6977

Kim KU, Han K, Kim J, Kwon DH, Ji YW *et al.* (2023) The Protective Role of Exosome-Derived MicroRNAs and Proteins from Human Breast Milk against Infectious Agents. Metabolites **13**(5). http://dx.doi.org/10.3390/metabo13050635

Kogure T, Yan IK, Lin WL and Patel T (2013) Extracellular Vesicle-Mediated Transfer of a Novel Long Noncoding RNA TUC339: A Mechanism of Intercellular Signaling in Human Hepatocellular Cancer. Genes Cancer **4**(7-8), 261-272. http://dx.doi.org/10.1177/1947601913499020

Lessen R and Kavanagh K (2015) Position of the academy of nutrition and dietetics: promoting and supporting breastfeeding. J Acad Nutr Diet **115**(3), 444-449. http://dx.doi.org/10.1016/j.jand.2014.12.014

Mathieu M, Martin-Jaular L, Lavieu G and Thery C (2019) Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. Nat Cell Biol **21**(1), 9-17. http://dx.doi.org/10.1038/s41556-018-0250-9

Melnik BC, Stremmel W, Weiskirchen R, John SM and Schmitz G (2021) Exosome-Derived MicroRNAs of Human Milk and Their Effects on Infant Health and Development. Biomolecules **11**(6). http://dx.doi.org/10.3390/biom11060851

Mitsis T, Pierouli K, Diakou KL, Papakonstantinou E, Bacopoulou F *et al.* (2020) Exosomics. EMBnet J **26**. http://dx.doi.org/10.14806/ej.26.0.934

Mosquera-Heredia MI, Morales LC, Vidal OM, Barcelo E, Silvera-Redondo C *et al.* (2021) Exosomes: Potential Disease Biomarkers and New Therapeutic Targets. Biomedicines **9**(8). http://dx.doi.org/10.3390/biomedicines9081061

Mourtzi N, Siahanidou T, Tsifintaris M, Karamichali E, Tasiopoulou A *et al.* (2021) lncRNA NORAD is consistently detected in breastmilk exosomes and its expression is downregulated in mothers of preterm infants. Int J Mol Med **48**(6). http://dx.doi.org/10.3892/ijmm.2021.5049

Papakonstantinou E, Vlamos P, Bacopoulou F, Chrousos GP, Eliopoulos E *et al.* (2023) Human Breast Milk Omics and Neurodevelopment. Adv Exp Med Biol **1423**, 235-236. http://dx.doi.org/10.1007/978-3-031-31978-5_22

Pisano C, Galley J, Elbahrawy M, Wang Y, Farrell A *et al.* (2020) Human Breast Milk-Derived Extracellular Vesicles in the Protection Against Experimental Necrotizing Enterocolitis. J Pediatr Surg **55**(1), 54-58. http://dx.doi.org/10.1016/j.jpedsurg.2019.09.052

Ramos-Garcia V, Ten-Domenech I, Albiach-Delgado A, Gomez-Ferrer M, Sepulveda P *et al.* (2023) Isolation and Lipidomic Screening of Human Milk Extracellular Vesicles. Methods Mol Biol **2571**, 177-188. http://dx.doi.org/10.1007/978-1-0716-2699-3_18

Shenker NS, Perdones-Montero A, Burke A, Stickland S, McDonald JAK *et al.* (2020) Metabolomic and Metataxonomic Fingerprinting of Human Milk Suggests Compositional Stability over a Natural Term of Breastfeeding to 24 Months. Nutrients **12**(11). http://dx.doi.org/10.3390/nu12113450

Slyk-Gulewska P, Kondracka A and Kwasniewska A (2023) MicroRNA as a new bioactive component in breast milk. Noncoding RNA Res **8**(4), 520-526. http://dx.doi.org/10.1016/j.ncrna.2023.06.003

Song J, Kim D, Han J, Kim Y, Lee M *et al.* (2015) PBMC and exosome-derived Hotair is a critical regulator and potent marker for rheumatoid arthritis. Clin Exp Med **15**(1), 121-126. http://dx.doi.org/10.1007/s10238-013-0271-4

Tenchov R, Sasso JM, Wang X, Liaw WS, Chen CA *et al.* (2022) Exosomes horizontal line Nature's Lipid Nanoparticles, a Rising Star in Drug Delivery and Diagnostics. ACS Nano **16**(11), 17802-17846. http://dx.doi.org/10.1021/acsnano.2c08774

Yan X, Liu L, Yao S, Chen Y, Yu Q *et al.* (2022) LncRNA and mRNA profiles of human milk-derived exosomes and their possible roles in protecting against necrotizing enterocolitis. Food Funct **13**(24), 12953-12965. http://dx.doi.org/10.1039/d2fo01866g

Zhang Y, Liu Y, Liu H and Tang WH (2019) Exosomes: biogenesis, biologic function and clinical potential. Cell Biosci **9**, 19. http://dx.doi.org/10.1186/s13578-019-0282-2

Zonneveld MI, van Herwijnen MJC, Fernandez-Gutierrez MM, Giovanazzi A, de Groot AM *et al.* (2021) Human milk extracellular vesicles target nodes in interconnected signalling pathways that enhance oral epithelial barrier function and dampen immune responses. J Extracell Vesicles **10**(5), e12071. http://dx.doi.org/10.1002/jev2.12071

EMBnet.journal
Bioinformatics in Action

Papakonstantinou et al. (2024) EMBnet.journal 29, e1049
http://dx.doi.org/10.14806/ej.29.0.1049

Research Papers

Exosomal Epigenetics

Eleni Papakonstantinou[1,2], Konstantina Dragoumani[1], George P Chrousos[2]✉, Dimitrios Vlachakis[1,2,3]✉

[1]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece

[2]University Research Institute of Maternal and Child Health & Precision Medicine, and UNESCO Chair on Adolescent Health Care, National and Kapodistrian University of Athens, “Aghia Sophia” Children’s Hospital, Athens, Greece

[3]School of Informatics, Faculty of Natural & Mathematical Sciences, King’s College London, London, U.K.

Competing interests: EP none; KD none; GPC none; DV is a member of EMBnet.journal Editorial Board

Abstract

Epigenetics is the study of heritable changes in gene expression that occur without changes to the underlying DNA sequence. Epigenetic modifications can include DNA methylation, histone modifications, and non-coding RNAs, among others. These modifications can influence the expression of genes by altering the way DNA is packaged and accessed by transcriptional machinery, thereby affecting cellular function and behavior. Epigenetic modifications can be influenced by a variety of factors, including environmental exposures, lifestyle factors, and aging, whilst abnormal epigenetic modifications have been implicated in a range of diseases, including cancer, neurodegenerative disorders, and cardiovascular disease. The study of epigenetics has the potential to provide new insights into the mechanisms of disease and could lead to the development of new diagnostic and therapeutic strategies. Exosomes can transfer epigenetic information to recipient cells, thereby influencing various physiological and pathological processes, and the identification of specific epigenetic modifications that are associated with a particular disease could lead to the development of targeted therapies that restore normal gene expression patterns. In recent years, the emerging role of exosomal epigenetics in human breast milk, highlighting its significance in infant nutrition and immune development. Milk exosomes are shown to carry epigenetic regulators, including miRNAs and long non-coding RNAs, which can modulate gene expression in recipient cells. These epigenetic modifications mediated by milk exosomal RNAs have implications for the development of the gastrointestinal tract, immune system, and metabolic processes in infants.

Introduction

Exosomal epigenetics refers to the study of how epigenetic modifications, which are chemical changes to DNA and associated proteins that regulate gene expression, can be transferred between cells via exosomes. Exosomes are small extracellular vesicles that are released by many different types of cells and contain a variety of bioactive molecules, including DNA, RNA, and proteins (Foo *et al.*, 2021). Exosomes can play a role in epigenetic regulation by carrying epigenetic information between cells (Qian *et al.*, 2015). For example, exosomes released by cancer cells have been found to contain DNA methylation and histone modifications that can be taken up by recipient cells and alter gene expression patterns (Behbahani *et al.*, 2016). Similarly, exosomes released by stem cells have been shown to contain miRNAs that can regulate gene expression in recipient cells (Foo *et al.*, 2021). Understanding the role of exosomal epigenetics in health and disease has the potential to provide new insights into the mechanisms underlying cellular communication and could lead to the development of new therapeutic approaches.

Exosomal epigenetic modifications refer to the changes in the epigenetic state of cells that are mediated by exosomes. Exosomes can carry various epigenetic modifications such as DNA methylation, histone modifications, and non-coding RNAs that can regulate gene expression in recipient cells (Zhang *et al.*, 2019). For example, exosomes released by cancer cells have been found to contain DNA methylation and histone modifications that can be taken up by recipient cells and alter gene expression patterns, leading to tumor growth and progression (Behbahani *et al.*, 2016).

Exosomal epigenetic modifications

Exosomal epigenetic modifications refer to the chemical changes to DNA and associated proteins that regulate gene expression, which can be transferred between cells via exosomes (Qian *et al.*, 2015). There are several types of epigenetic modifications that can be

Article history
Received: 26 September 2023
Accepted: 03 October 2023
Published: 22 May 2024

© 2024 Papakonstantinou *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at http://journal.embnet.org.

transferred via exosomes, including DNA methylation, histone modifications, and non-coding RNAs such as miRNAs, that can influence gene expression, and as a result, they have the potential to contribute to a range of physiological and pathological processes (Zhao *et al.*, 2023). When exosomes are taken up by recipient cells, the miRNAs and mRNAs they contain can influence gene expression patterns in the recipient cell, leading to changes in cellular behavior and function (Valadi *et al.*, 2007). For example, exosomes released by stem cells have been shown to contain miRNAs that can promote cell proliferation and inhibit apoptosis in recipient cells, leading to tissue repair and regeneration (Foo *et al.*, 2021). Similarly, exosomes released by cancer cells can contain miRNAs and mRNAs that promote tumor growth and invasion by regulating the expression of genes involved in cell proliferation, migration, and invasion. In addition to cancer, exosomal epigenetic modifications have been implicated in a range of other diseases, including neurological disorders, cardiovascular disease, and autoimmune diseases; for instance, exosomes derived from mesenchymal stem cells have been shown to promote neuronal differentiation and neurite outgrowth by transferring miRNAs to recipient cells (Zhao *et al.*, 2023).

Exosomal DNA methylation has been shown to play a role in cancer progression, with cancer cells transferring hypermethylated DNA fragments to neighboring cells via exosomes, leading to the silencing of tumor suppressor genes in recipient cells (Behbahani *et al.*, 2016). Additionally, exosomal histone modifications have been implicated in regulating gene expression during embryonic development, and abnormal levels of histone modifications in exosomes have been associated with various diseases, including cancer and neurodegenerative disorder (Volker-Albert *et al.*, 2020). Getting insights in the exosomal epigenetic modifications has the potential to provide new insights into the mechanisms of cellular communication.

Exosomal gene regulation

Exosomal gene regulation refers to the process by which genetic material carried by exosomes, such as miRNAs, mRNAs, lncRNAs, and other non-coding RNAs, can influence gene expression in recipient cells. Exosomes released by the parent cells can contain various biomolecules, including genetic material (Kalluri and LeBleu, 2020). When exosomes are taken up by recipient cells, the genetic material they contain can regulate the expression of genes in the recipient cell by targeting specific mRNA transcripts and influencing the stability and translation of these transcripts (Lloret-Llinares *et al.*, 2018). For example, exosomal miRNAs have been shown to regulate a variety of cellular processes, including cell proliferation, apoptosis, and differentiation, by targeting specific mRNAs and suppressing their expression (Hu *et al.*, 2012). Similarly, exosomal mRNAs have been shown to be translated in recipient cells, leading to the expression of proteins that can influence cell behavior and function (Hu *et al.*, 2012). The role of exosomal gene regulation in health and disease is an area of active research, with potential implications for the development of new therapeutic strategies. For example, exosomes carrying specific miRNAs could be used to deliver targeted therapies for diseases such as cancer, while exosomes carrying mRNAs could be used to promote tissue repair and regeneration (Fang *et al.*, 2022).

Exosomal Biomarkers

The use of exosomal biomarkers for disease diagnosis and monitoring has several advantages over traditional biomarkers, such as their stability in body fluids and their ability to be isolated from a variety of sources, including blood, urine, and saliva, and they have been studied as potential indicators of cancer, neurodegenerative disorders, and cardiovascular disease (Huda *et al.*, 2021). In cancer, exosomes have been shown to carry specific proteins, such as CD63 and CD81, that are often upregulated in cancer cells, as well as specific miRNAs and mRNAs that can be used to monitor disease progression and treatment response (Mathew *et al.*, 2021). Exosomal epigenetic information can be transferred between cells and influence gene expression in recipient cells. For example, exosomes have been shown to carry DNA fragments that are hypermethylated at specific sites, leading to the silencing of tumor suppressor genes in recipient cells (Aslan *et al.*, 2019). Similarly, exosomal miRNAs can regulate gene expression in recipient cells by binding to specific mRNA transcripts and regulating their stability and translation (Foo *et al.*, 2021).

In addition to DNA and miRNAs, exosomes can also carry other epigenetic factors, such as histone modifications and other non-coding RNAs. These factors can influence gene expression in recipient cells by altering the accessibility of DNA to transcriptional machinery or by regulating the stability and translation of mRNA transcripts (Zhang *et al.*, 2019). In addition to their diagnostic and prognostic potential, exosomal biomarkers also have the potential to be used as therapeutic targets. For example, exosomes carrying specific miRNAs or proteins could be targeted to inhibit disease progression or promote tissue repair (Aslan *et al.*, 2019; Mathew *et al.*, 2021).

Epigenetic Effects of Human Breast Milk Exosomes

Human breast milk is an intricate fluid teeming with a multitude of compounds essential for infant nutrition and the development of their immune systems. Among its constituents are secretory immunoglobulins (IgA), leucocytes, lysozyme, and lactoferrin, all of which play crucial roles in conferring passive immunity to infants and regulating the development of their immune systems (Kim and Yi, 2020). Beyond its nutritional value, breast milk contains a rich array of exosomes, which play a crucial role in intercellular communication and the transfer of bioactive molecules between maternal mammary

epithelial cells and infant cells (O'Reilly *et al.*, 2021). Emerging evidence suggests that exosomes present in human breast milk carry epigenetic information through the delivery of miRNAs, DNA fragments, and histones to recipient cells, that can influence gene expression and developmental programming in the infant (Leroux *et al.*, 2021).

The transfer of epigenetic information via exosomal cargo is thought to play a critical role in infant health and development. MiRNAs encapsulated within exosomes have been shown to regulate gene expression by targeting specific mRNA transcripts in recipient cells. These miRNAs can influence various cellular processes, including immune function, metabolism, and neuronal development, thereby shaping the developmental trajectory of the infant (Abeysinghe *et al.*, 2020; Zhou *et al.*, 2012). Additionally, exosomal DNA fragments and histones can be transferred to infant cells, where they may contribute to epigenetic modifications and gene regulation. DNA methylation patterns carried by exosomes may influence the establishment of DNA methylation profiles in the infant's genome, potentially impacting gene expression and long-term health outcomes (Takahashi *et al.*, 2017).

Early-life exposures to maternal exosomal miRNAs and epigenetic regulators could shape the developmental trajectory of key physiological systems, potentially influencing susceptibility to chronic diseases, such as obesity, diabetes, and cardiovascular disorders, in adulthood (Rashidi *et al.*, 2022). The epigenetic effects of milk exosomal RNAs, particularly miRNAs, play a crucial role in promoting intestinal health and immune regulation in infants (Alsaweed *et al.*, 2015). Studies have shown that milk exosomes and their RNA cargoes can enhance intestinal epithelial cell growth and protect against intestinal injury and inflammation (Zeng *et al.*, 2021). For instance, miRNAs such as miR-200a-3p, miR-4334, miR-219, and miR-338 have been found to mitigate intestinal inflammation and damage by targeting proinflammatory genes and pathways (Sun *et al.*, 2013; Xie *et al.*, 2019). Moreover, milk exosomal miRNAs are implicated in immune modulation, with reports suggesting their potential role in regulatory T-cell induction and immune protection (Admyre *et al.*, 2007). Various immune-related miRNAs abundant in milk exosomes have been shown to regulate processes such as B-cell tolerance, plasma cell differentiation, and cytokine expression, thereby influencing immune responses in infants (Chen *et al.*, 2014; Mourtada-Maarabouni *et al.*, 2008; Quan *et al.*, 2020).

Milk exosomal RNAs, including miRNAs and lncRNAs, contribute to epigenetic regulation by targeting genes involved in DNA methylation and histone modification. For example, miRNAs such as miR-148a, miR-152, and miR-29b target DNA methyltransferases, potentially affecting genomic DNA methylation patterns and gene expression (Melnik and Kakulas, 2017). These epigenetic modifications mediated by milk exosomal RNAs have implications for the development of the gastrointestinal tract, immune system, and metabolic processes in infants. While milk exosomal RNAs offer potential benefits for intestinal health and immune function, their implications in metabolic diseases have also raised concerns. Some miRNAs abundant in milk exosomes, such as miR-148a, miR-29b, and miR-21, have been associated with adipogenesis, insulin resistance, and osteoporosis, raising questions about their impact on metabolic health in recipients (Bian *et al.*, 2015; Guglielmi *et al.*, 2017; Monda *et al.*, 2013).

Breast milk contains miRNAs (miRNAs) that are pivotal in orchestrating gene expression in infants, with recent research shedding light on a subset of miRNAs termed xeno-miRNAs (xenomiRs) (Zhang *et al.*, 2012). XenomiRs originate from non-human sources, primarily maternal diet, and are present in human circulation, exerting regulatory effects on gene expression and potentially influencing the immune system. The composition of breast milk miRNAs, including xenomiRs, is intricately linked to maternal dietary intake, highlighting the profound impact of maternal nutrition on infant health outcomes (Stephen *et al.*, 2020). Through vertical transmission via breast milk, xenomiRs derived from dietary sources may modulate gene expression in infants, offering a fascinating glimpse into the intricate interplay between maternal diet and infant health. The presence of xenomiRs in breast milk underscores the importance of considering dietary factors in shaping infant immunity and underscores the complex dynamics of cross-species communication (Liao *et al.*, 2017).

This emerging field of research on milk exosomics not only deepens our understanding of the role of breast milk in infant nutrition and immune development but also opens new avenues for optimizing infant health through targeted nutritional interventions.

Conclusion

Recent research has focused on exploring the role and applications of exosomes, particularly in human breast milk, to elucidate their epigenetic effects and mechanisms underlying exosomal epigenetic regulation in health and disease. Promoting breastfeeding initiation and duration can yield far-reaching benefits for infant health and development due to the diverse array of bioactive components, including exosomes, present in human breast milk. Additionally, efforts to optimise maternal nutrition and lifestyle factors during lactation may enhance the composition and functionality of exosomes in breast milk, further augmenting their epigenetic effects on infant health.

While substantial progress has been made in elucidating the epigenetic effects of exosomes in human breast milk, several questions remain unanswered. Future research endeavors should focus on characterizing the epigenetic effects of exosomes in breast milk, delineating their mechanisms of action, and discerning their long-term effects on infant health and disease susceptibility. By unraveling the mechanisms underlying the transfer

and impact of exosomal cargo in breast milk, the scientific community aims to harness this knowledge to devise innovative strategies for promoting infant health and disease prevention.

Key Points

- Elucidating the mechanisms by which exosomes facilitate the transfer of epigenetic cues between cells, modulating gene expression patterns and cellular responses.
- The influence of exosomal epigenetic cargo present in human breast milk on neonatal health and developmental outcomes, with a focus on immune modulation and metabolic regulation, is investigated.
- Exosomal biomarkers have a diagnostic, prognostic, and therapeutic potential for various diseases, including their utility in targeted therapeutic interventions.
- Evaluating the impact of maternal dietary microRNA content in breast milk on neonatal gene expression and immune function is key modulator in shaping neonatal health.

Funding

The authors would like to acknowledge funding from 'MilkSafe: A novel pipeline to enrich formula milk using omics technologies', a research co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T2EDK-02222).

References

Abeysinghe P, Turner N, Morean Garcia I, Mosaad E, Peiris HN *et al.* (2020) The Role of Exosomal Epigenetic Modifiers in Cell Communication and Fertility of Dairy Cows. Int J Mol Sci **21**(23). http://dx.doi.org/10.3390/ijms21239106

Admyre C, Johansson SM, Qazi KR, Filen JJ, Lahesmaa R *et al.* (2007) Exosomes with immune modulatory features are present in human breast milk. J Immunol **179**(3), 1969-1978. http://dx.doi.org/10.4049/jimmunol.179.3.1969

Alsaweed M, Hartmann PE, Geddes DT and Kakulas F (2015) MicroRNAs in Breastmilk and the Lactating Breast: Potential Immunoprotectors and Developmental Regulators for the Infant and the Mother. Int J Environ Res Public Health **12**(11), 13981-14020. http://dx.doi.org/10.3390/ijerph121113981

Aslan C, Maralbashi S, Salari F, Kahroba H, Sigaroodi F *et al.* (2019) Tumor-derived exosomes: Implication in angiogenesis and antiangiogenesis cancer therapy. J Cell Physiol **234**(10), 16885-16903. http://dx.doi.org/10.1002/jcp.28374

Behbahani GD, Khani S, Hosseini HM, Abbaszadeh-Goudarzi K and Nazeri S (2016) The role of exosomes contents on genetic and epigenetic alterations of recipient cancer cells. Iran J Basic Med Sci **19**(10), 1031-1039.

Bian Y, Lei Y, Wang C, Wang J, Wang L *et al.* (2015) Epigenetic Regulation of miR-29s Affects the Lactation Activity of Dairy Cow Mammary Epithelial Cells. J Cell Physiol **230**(9), 2152-2163. http://dx.doi.org/10.1002/jcp.24944

Chen T, Xi QY, Ye RS, Cheng X, Qi QE *et al.* (2014) Exploration of microRNAs in porcine milk exosomes. BMC Genomics **15**(1), 100. http://dx.doi.org/10.1186/1471-2164-15-100

Fang Z, Zhang X, Huang H and Wu J (2022) Exosome based miRNA delivery strategy for disease treatment. Chinese Chemical Letters **33**(4), 1693-1704. http://dx.doi.org/https://doi.org/10.1016/j.cclet.2021.11.050

Foo JB, Looi QH, How CW, Lee SH, Al-Masawa ME *et al.* (2021) Mesenchymal Stem Cell-Derived Exosomes and MicroRNAs in Cartilage Regeneration: Biogenesis, Efficacy, miRNA Enrichment and Delivery. **14**. http://dx.doi.org/10.3390/ph14111093

Guglielmi V, D'Adamo M, Menghini R, Cardellini M, Gentileschi P *et al.* (2017) MicroRNA 21 is up-regulated in adipose tissue of obese diabetic subjects. Nutr Healthy Aging **4**(2), 141-145. http://dx.doi.org/10.3233/NHA-160020

Hu G, Drescher KM and Chen XM (2012) Exosomal miRNAs: Biological Properties and Therapeutic Potential. Front Genet **3**, 56. http://dx.doi.org/10.3389/fgene.2012.00056

Huda MN, Nafiujjaman M, Deaguero IG, Okonkwo J, Hill ML *et al.* (2021) Potential Use of Exosomes as Diagnostic Biomarkers and in Targeted Drug Delivery: Progress in Clinical and Preclinical Applications. ACS Biomater Sci Eng **7**(6), 2106-2149. http://dx.doi.org/10.1021/acsbiomaterials.1c00217

Kalluri R and LeBleu VS (2020) The biology, function, and biomedical applications of exosomes. Science **367**(6478). http://dx.doi.org/10.1126/science.aau6977

Kim SY and Yi DY (2020) Components of human breast milk: from macronutrient to microbiome and microRNA. Clin Exp Pediatr **63**(8), 301-309. http://dx.doi.org/10.3345/cep.2020.00059

Leroux C, Chervet ML and German JB (2021) Perspective: Milk microRNAs as Important Players in Infant Physiology and Development. Adv Nutr **12**(5), 1625-1635. http://dx.doi.org/10.1093/advances/nmab059

Liao Y, Du X, Li J and Lonnerdal B (2017) Human milk exosomes and their microRNAs survive digestion in vitro and are taken up by human intestinal cells. Mol Nutr Food Res **61**(11). http://dx.doi.org/10.1002/mnfr.201700082

Lloret-Llinares M, Karadoulama E, Chen Y, Wojenski LA, Villafano GJ *et al.* (2018) The RNA exosome contributes to gene expression regulation during stem cell differentiation. Nucleic Acids Res **46**(21), 11502-11513. http://dx.doi.org/10.1093/nar/gky817

Mathew B, Mansuri MS, Williams KR and Nairn AC (2021) Exosomes as Emerging Biomarker Tools in Neurodegenerative and Neuropsychiatric Disorders-A Proteomics Perspective. Brain Sci **11**(2). http://dx.doi.org/10.3390/brainsci11020258

Melnik BC and Kakulas F (2017) Milk Exosomes and microRNAs: Potential Epigenetic Regulators.In: Patel V and Preedy V (Eds.) Handbook of Nutrition, Diet, and Epigenetics. Springer International Publishing, Cham, pp. 1-28.

Monda KL, Chen GK, Taylor KC, Palmer C, Edwards TL *et al.* (2013) A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat Genet **45**(6), 690-696. http://dx.doi.org/10.1038/ng.2608

Mourtada-Maarabouni M, Hedge VL, Kirkham L, Farzaneh F and Williams GT (2008) Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). J Cell Sci **121**(Pt 7), 939-946. http://dx.doi.org/10.1242/jcs.024646

O'Reilly D, Dorodnykh D, Avdeenko NV, Nekliudov NA, Garssen J *et al.* (2021) Perspective: The Role of Human Breast-Milk Extracellular Vesicles in Child Health and Disease. Adv Nutr **12**(1), 59-70. http://dx.doi.org/10.1093/advances/nmaa094

Qian Z, Shen Q, Yang X, Qiu Y and Zhang W (2015) The Role of Extracellular Vesicles: An Epigenetic View of the Cancer Microenvironment. Biomed Res Int 2015, 649161. http://dx.doi.org/10.1155/2015/649161

Quan S, Nan X, Wang K, Jiang L, Yao J *et al.* (2020) Characterization of Sheep Milk Extracellular Vesicle-miRNA by Sequencing and Comparison with Cow Milk. Animals (Basel) **10**(2). http://dx.doi.org/10.3390/ani10020331

Rashidi M, Bijari S, Khazaei AH, Shojaei-Ghahrizjani F and Rezakhani L (2022) The role of milk-derived exosomes in the treatment of diseases. Front Genet **13**, 1009338. http://dx.doi.org/10.3389/fgene.2022.1009338

Stephen BJ, Pareek N, Saeed M, Kausar MA, Rahman S *et al.* (2020) Xeno-miRNA in Maternal-Infant Immune Crosstalk: An Aid to Disease Alleviation. Front Immunol **11**, 404. http://dx.doi.org/10.3389/fimmu.2020.00404

Sun Q, Chen X, Yu J, Zen K, Zhang CY *et al.* (2013) Immune modulatory function of abundant immune-related microRNAs in microvesicles from bovine colostrum. Protein Cell **4**(3), 197-210. http://dx.doi.org/10.1007/s13238-013-2119-9

Takahashi A, Okada R, Nagao K, Kawamata Y, Hanyu A *et al.* (2017) Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. Nat Commun **8**, 15287. http://dx.doi.org/10.1038/ncomms15287

Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ *et al.* (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol **9**(6), 654-659. http://dx.doi.org/10.1038/ncb1596

Volker-Albert M, Bronkhorst A, Holdenrieder S and Imhof A (2020) Histone Modifications in Stem Cell Development and Their Clinical Implications. Stem Cell Reports **15**(6), 1196-1205. http://dx.doi.org/10.1016/j.stemcr.2020.11.002

Xie MY, Hou LJ, Sun JJ, Zeng B, Xi QY *et al.* (2019) Porcine Milk Exosome MiRNAs Attenuate LPS-Induced Apoptosis through Inhibiting TLR4/NF-kappaB and p53 Pathways in Intestinal Epithelial Cells. J Agric Food Chem **67**(34), 9477-9491. http://dx.doi.org/10.1021/acs.jafc.9b02925

Zeng B, Chen T, Luo JY, Zhang L, Xi QY *et al.* (2021) Biological Characteristics and Roles of Noncoding RNAs in Milk-Derived Extracellular Vesicles. Adv Nutr **12**(3), 1006-1019. http://dx.doi.org/10.1093/advances/nmaa124

Zhang L, Hou D, Chen X, Li D, Zhu L *et al.* (2012) Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. Cell Res **22**(1), 107-126. http://dx.doi.org/10.1038/cr.2011.158

Zhang Y, Liu Y, Liu H and Tang WH (2019) Exosomes: biogenesis, biologic function and clinical potential. Cell Biosci **9**, 19. http://dx.doi.org/10.1186/s13578-019-0282-2

Zhao Z, Zhang L, Ocansey DKW, Wang B and Mao F (2023) The role of mesenchymal stem cell-derived exosome in epigenetic modifications in inflammatory diseases. Front Immunol **14**, 1166536. http://dx.doi.org/10.3389/fimmu.2023.1166536

Zhou Q, Li M, Wang X, Li Q, Wang T *et al.* (2012) Immune-related microRNAs are abundant in breast milk exosomes. Int J Biol Sci **8**(1), 118-123. http://dx.doi.org/10.7150/ijbs.8.118

Papakonstantinou et al. (2024) EMBnet.journal 29, e1050
http://dx.doi.org/10.14806/ej.29.0.1050

Milk exosomes and a new way of communication between mother and child

Eleni Papakonstantinou[1,2], Konstantina Dragoumani[1], Thanasis Mitsis[1], George P Chrousos[2]✉, Dimitrios Vlachakis[1,2,3]✉

[1]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece

[2]University Research Institute of Maternal and Child Health & Precision Medicine, and UNESCO Chair on Adolescent Health Care, National and Kapodistrian University of Athens, "Aghia Sophia" Children's Hospital, Athens, Greece

[3]School of Informatics, Faculty of Natural & Mathematical Sciences, King's College London, London, U.K.

Competing interests: EP none; KD none; GPC none; DV is a member of EMBnet.journal Editorial Board

Abstract

Extracellular vesicles are a heterogeneous group of lipid-bound vesicles released by cells into the extracellular space. EVs are an important mediator of intercellular communications and carry a wide variety of molecules that exert a biological function, such as lipids, nucleic acids, proteins, ions, and adenosine triphosphate (ATP). Extracellular vesicles are classified into microvesicles, exosomes, and apoptotic bodies depending on their biogenesis and size. Exosomes are spherical lipid-bilayer vesicles with a diameter of about 40 to 100 nm. Exosomes originate from intracellular endosomal compartments, while microvesicles originated directly from a cell's plasma membrane and apoptotic bodies originate from cells undergoing apoptosis and are released via outward blebbing and fragmentation of the plasma membrane. Specifically, exosomes have garnered great attention since they display great potential as both biomarkers and carriers of therapeutic molecules.

Extracellular vesicles

Extracellular vesicles (EVs) are a heterogeneous group of lipid-bound vesicles released by cells into the extracellular space (Abels and Breakefield, 2016; Doyle and Wang, 2019). EVs are an important mediator of intercellular communications and carry a wide variety of molecules that exert a biological function, such as lipids, nucleic acids, proteins, ions, and adenosine triphosphate (ATP) (Shetty and Upadhya, 2021; Zeng *et al.*, 2022). EVs are classified into microvesicles, exosomes, and apoptotic bodies depending on their biogenesis and size (Tang *et al.*, 2019). Exosomes are spherical lipid-bilayer vesicles with a diameter of about 40 to 100 nm (Zhang *et al.*, 2020). Exosomes originate from intracellular endosomal compartments, while microvesicles originated directly from a cell's plasma membrane and apoptotic bodies originate from cells undergoing apoptosis and are released via outward blebbing and fragmentation of the plasma membrane (Willms *et al.*, 2018). Specifically, exosomes have garnered great attention since they display great potential as both biomarkers and carriers of therapeutic molecules (Zhang *et al.*, 2019).

Exosome composition

Exosome composition depends on its cell of origin, though several structural components remain constant among different populations. A cohort of distinct proteins is scattered among them. Most exosomes carry tetraspanins, such as CD9, CD37, CD63, CD81, and CD82, endosomal sorting complex required for transport (ESCRT) proteins such as TSG101 and Alix, cell adhesion molecules such as CD31 and CD44, heat shock proteins such as HSP27, HSP60, HSP70, HSP90, and Rab GTPases such as Rab11 and Rab27 (Burtenshaw *et al.*, 2022). Depending on the donor cell type, some exosomes may also display class 1 and class 2 major histocompatibility complex molecules (MHC class I and MHC class II). The exosome bilayer membrane is quite rigid and consists of lipids such as cholesterol, sphingomyelin, and ceramides (Gurung *et al.*, 2021).

An exosome's unique cargo depends on the originating cell. Different cell types secrete different exosomes with different functions, while alterations in their status due to inflammation, viral infection, or other pathological conditions like cancer and neurodegenerative disorders also play a role in exosomal cargo (Chen *et al.*, 2021). For example, astrocyte-derived

Article history
Received: 23 September 2023
Accepted: 03 October 2023
Published: 22 May 2024

© 2024 Papakonstantinou *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at http://journal.embnet.org.

exosomes play an important role in neuroplasticity and neuronal function due to their unique cargo of bioactive compounds like neuroglobin, glutaminase, and prostaglandin D2 synthase (Xia *et al.*, 2022). On the other hand, in pathological conditions like Parkinson's Disease (PD) neuron-derived exosomes appear to carry α-syn oligomers whose aggregation is a main characteristic of disease progression (Yu *et al.*, 2020).

Although the biological origin of exosomes is known, the exact specifics of their formation and cargo sorting are still under research (Zhang *et al.*, 2020). Specifically, early endosomes are created via the invagination of the plasma membrane and the early accumulation of bioactive compounds (Zhang *et al.*, 2020). These early endosomes grow into late endosomes through acidification, protein content alterations, and an increase in their ability to fuse with other membranes. Ultimately late endosomes form multivesicular bodies (MVBs) via reverse budding during which the endosomal membrane invaginates to create intraluminal vesicles (ILVs). Some MVBs later fuse with the cell membrane and release ILVs into the extracellular space as exosomes (Vlachakis *et al.*, 2021; Zhang *et al.*, 2020). The exact formation and cargo sorting of exosomes may depend on the ESCRT machinery or other components like tetraspanins and lipid rafts (Zhang *et al.*, 2020).

Exosome-Mediated Communication

Exosomes facilitate intercellular communication through various mechanisms. Upon release, exosomes may interact directly with receptors on recipient cell surfaces, triggering signaling pathways without cargo delivery, or alternatively, they can be internalised by recipient cells, leading to the release of their cargo into the cytosol. The uptake and processing of exosomes by recipient cells involve complex molecular interactions and endocytic processes, although the specifics remain under investigation. Even though exosomes exhibit a high degree of selectivity in target cell recognition, often favoring cells of similar origin, a small fraction of exosomes may be taken up by non-similar cells, suggesting potential roles in intercellular crosstalk beyond cell-type specificity.

Released exosomes may manage to transmit information between cells in multiple ways by acting directly on the receptor's cell surface without cargo delivery. Exosomes that carry MHCI and MHCII act as antigen-presenting agents and activate T-cell receptors (TCRs) on T lymphocytes (Anel *et al.*, 2019; Mathieu *et al.*, 2019). In other instances, exosomes are internalised through and led to the lysosome for degradation or recycled and re-secreted. Most target cells receive information by exosome uptake where the vesicle's cargo is released in the cytosol. The specifics of exosome uptake and cargo delivery are also still under research. Exosomes could dock via distinct molecular interactions that make use of membrane-exposed proteins, lipids (inset), or sugars, or via non-specific endocytosis like macropinocytosis or micropinocytosis. The exosomes entering the recipient cell may target the endosomes. Once internalised, exosomes either release their cargo through the indirect route of endosomal escape, are recycled and re-secreted, or are targeted for degradation at the lysosome. Endosomal escape is possibly mediated by mechanisms resembling those used by viruses that harbor fusogenic proteins similar to the glycoprotein G of the vesicular stomatitis virus (VSVG). Alternatively, exosomes may directly release their cargo into the cytosol after vesicle fusion with the plasma membrane (Mathieu *et al.*, 2019).

Exosomes' selective trafficking and communication mechanisms are quite complex. Exosomes are vastly uptaken by cells similar to the cell of origin. It appears that exosomes released from a distinct cell type carry a conserved signature that is used as a recognition moiety for the same type of cell. Therefore, exosome targeting is highly selective highlighting an essential role in intercellular communication between similar cell lines. Nevertheless, a very small percentage of exosomes are uptaken by proximal cells regardless of the cell of origin (Sancho-Albero *et al.*, 2019).

There is a possibility that the small number of exosomes uptaken by non-similar to the cell of origin adjacent cells potentially play another role in intercellular communication. Particularly, extended exosome recycling between non-similar cells may underlie complex biological mechanisms. For instance, breast cancer animal models have established that breast cancer cells (BCCs) receive CD81+ exosomes from fibroblasts and later re-secrete them. In this case, BBC-produced Wnt11 is incorporated in fibroblast-derived exosomes during their localisation in the recipient cell. The re-secreted wnt11-associated CD81+ exosomes then promote breast cancer motility via Wnt-related signaling (Luga *et al.*, 2012). We speculate that this type of recycling mechanism may govern physiological mechanisms.

This theory supposes that a small number of exosomes incorporate elements from different cell types under physiological conditions. For example, an exosome released from a type A cell will carry type A elements, which after being uptaken and recycled by a type B cell will carry both type A and type B elements. This exosome in turn will be uptaken and recycled by a type C cell and thus will carry type A, type B, and type C elements. If this process continues, the end result would be a hybrid cell that contains multiple moieties from vastly different sources (Figure 1). These exosomes could mediate the transport of cellular information that highlights the status of entire systems and not single cell types across distant cell groups. Given that EVs have the ability to cross the blood-brain-barrier (Banks *et al.*, 2020), if such exosomes exist, they could -via their cargo-inform the CNS of the health status of every cell type in the body. This kind of ability could uncover an entirely new system of health monitoring by the brain.

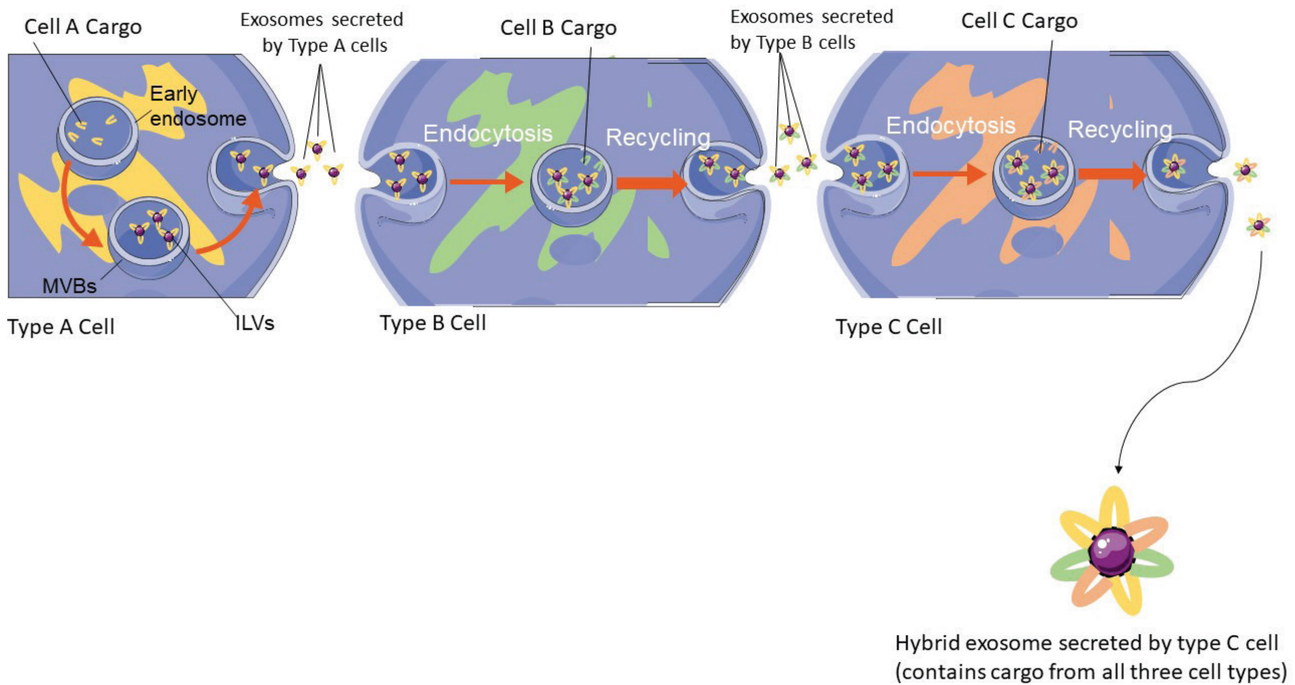


Figure 1. A possible mechanism for repeated exosome recycling. Early endosomes are formed in the type A cell, during which bioactive compounds are accumulated there to later serve as exosomal cargo. The early endosomes mature to form MVBs which themselves create ILVs where the bioactive compounds will be incorporated. These ILVs are later released by the type A cell as exosomes and are uptaken by the type B cell. These exosomes are internalised through endocytosis by the type B cell where instead of releasing their cargo or led to the endosome, they undergo a process called recycling. In exosome recycling the type B cell incorporates its own cargo on the vesicles and later re-releases them. The exosomes released by the type B cell later undergo a similar process by being uptaken and recycled by a type C cell. The final exosome released by the type C cell will be hybrid in nature since it contains cargo from all three cell types. This process could be repeated among multiple cell types.

Several biological applications could also be developed based on the aforementioned hypothesis. Specifically, the existence of hybrid exosomes could pave the way for new methods of diagnosis. Since exosomal cargo can hint at the existence of pathological conditions in the cell of origin, a hybrid exosome could work as previously mentioned a health marker of entire systems and not a specific cell line. Thus, isolation of exosomes and identification of their contents could provide a holistic view of a patient's health and help diagnose complex pathologies.

We should reinstate that only a small number of exosomes are uptaken by cells non-similar to their cell of origin, and exosome recycling is still a mechanism under research. Therefore, further studies are needed to verify if such hybrid exosomes exist. Nevertheless, hybrid exosomes could greatly expand our current knowledge of intercellular communication.

Maternal-Infant Communication via Milk Exosomes

Milk exosomes represent a fascinating avenue for maternal-infant communication, offering a conduit for the transfer of bioactive molecules from mother to child. During lactation, mammary epithelial cells secrete exosomes into breast milk, encapsulating a diverse cargo reflective of maternal physiological status and environmental exposures. These milk exosomes are ingested by the infant during breastfeeding, facilitating the transfer of essential nutrients, immune factors, and signaling molecules. Notably, the survivability of human milk exosomal miRNAs upon simulated digestion has been confirmed, showing the great potential effect of milk exosomes (Liao *et al.*, 2017). Emerging evidence also suggests that milk exosomes play a crucial role in immune system maturation, gastrointestinal development, and neurodevelopment in the breastfeeding infant. miRNAs encapsulated within exosomes, facilitating their stability and functionality in the gastrointestinal tract are implicated in epigenetic regulations promoting intestinal health in infants, and show a protective role against inflammation and injury (Alsaweed *et al.*, 2015; Zeng *et al.*, 2021). Studies also suggest that miRNAs present in breast milk may also contribute to infant immune regulation, highlighting the complexity of communication via milk-derived exosomal miRNAs (Admyre *et al.*, 2007). Furthermore, the selective packaging of specific molecules into milk exosomes may enable tailored communication between mother and child, optimizing infant health and development.

Conclusions

To further enhance our understanding of milk exosomes and their role in maternal-infant communication, future research should focus on elucidating the specific

mechanisms by which milk exosomes exert their effects on infant health and development. This includes investigating the regulatory pathways involved in the transfer of exosomal cargo from mother to child, as well as exploring how environmental factors and maternal health status influence the composition and function of milk exosomes. Additionally, longitudinal studies tracking the long-term health outcomes of infants exposed to different profiles of milk exosomal cargo could provide valuable insights into the potential implications for lifelong health trajectories. Moreover, efforts to develop non-invasive methods for isolating and analyzing milk exosomes could facilitate clinical applications, such as the development of diagnostic tools or therapeutic interventions aimed at optimizing maternal and infant health. Overall, continued investigation into milk exosomes holds promise for uncovering novel strategies to support infant health and development.

Key Points

- Extracellular Vesicles encompass a diverse array of lipid-bound vesicles released by cells, acting as crucial mediators of intercellular communication, and containing various bioactive molecules. Exosomes, in particular, drawing specific attention due to their potential as biomarkers and therapeutic carriers.
- Exosomes exhibit a unique composition influenced by their cell of origin, containing a range of proteins, lipids, nucleic acids, and other molecules. Their cargo reflects the physiological state of the parent cell and can vary under different pathological conditions.
- Exosomes facilitate bidirectional communication between cells through various mechanisms, including direct interaction with cell surface receptors and cargo delivery upon internalisation.
- While exosomes primarily target cells of similar origin, a small fraction may interact with non-similar cells, suggesting broader implications for intercellular crosstalk. Investigating the complex dynamics of exosome trafficking and cargo exchange is essential for comprehensively understanding their role in cellular communication networks.
- Milk exosomes represent a fascinating avenue for maternal-infant communication, facilitating the transfer of essential nutrients, immune factors, and signaling molecules from mother to child during breastfeeding. These exosomes play a crucial role in immune system maturation, gastrointestinal development, and neurodevelopment in the infant, with implications for long-term health outcomes.

Funding

The authors would like to acknowledge funding from 'MilkSafe: A novel pipeline to enrich formula milk using omics technologies', a research co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T2EDK-02222).

References

Abels ER and Breakefield XO (2016) Introduction to Extracellular Vesicles: Biogenesis, RNA Cargo Selection, Content, Release, and Uptake. Cell Mol Neurobiol **36**(3), 301-312. http://dx.doi.org/10.1007/s10571-016-0366-z

Admyre C, Johansson SM, Qazi KR, Filen JJ, Lahesmaa R *et al.* (2007) Exosomes with immune modulatory features are present in human breast milk. J Immunol **179**(3), 1969-1978. http://dx.doi.org/10.4049/jimmunol.179.3.1969

Alsaweed M, Hartmann PE, Geddes DT and Kakulas F (2015) MicroRNAs in Breastmilk and the Lactating Breast: Potential Immunoprotectors and Developmental Regulators for the Infant and the Mother. Int J Environ Res Public Health **12**(11), 13981-14020. http://dx.doi.org/10.3390/ijerph121113981

Anel A, Gallego-Lleyda A, de Miguel D, Naval J and Martínez-Lostao L (2019) Role of Exosomes in the Regulation of T-cell Mediated Immune Responses and in Autoimmune Disease. Cells **8**(2). http://dx.doi.org/10.3390/cells8020154

Banks WA, Sharma P, Bullock KM, Hansen KM, Ludwig N *et al.* (2020) Transport of Extracellular Vesicles across the Blood-Brain Barrier: Brain Pharmacokinetics and Effects of Inflammation. Int J Mol Sci **21**(12). http://dx.doi.org/10.3390/ijms21124407

Burtenshaw D, Regan B, Owen K, Collins D, McEneaney D *et al.* (2022) Exosomal Composition, Biogenesis and Profiling Using Point-of-Care Diagnostics-Implications for Cardiovascular Disease. Front Cell Dev Biol **10**, 853451. http://dx.doi.org/10.3389/fcell.2022.853451

Chen H, Wang L, Zeng X, Schwarz H, Nanda HS *et al.* (2021) Exosomes, a New Star for Targeted Delivery. Front Cell Dev Biol **9**, 751079. http://dx.doi.org/10.3389/fcell.2021.751079

Doyle LM and Wang MZ (2019) Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. Cells **8**(7). http://dx.doi.org/10.3390/cells8070727

Gurung S, Perocheau D, Touramanidou L and Baruteau J (2021) The exosome journey: from biogenesis to uptake and intracellular signalling. Cell Communication and Signaling **19**(1), 47. http://dx.doi.org/10.1186/s12964-021-00730-1

Liao Y, Du X, Li J and Lonnerdal B (2017) Human milk exosomes and their microRNAs survive digestion in vitro and are taken up by human intestinal cells. Mol Nutr Food Res **61**(11). http://dx.doi.org/10.1002/mnfr.201700082

Luga V, Zhang L, Viloria-Petit AM, Ogunjimi AA, Inanlou MR *et al.* (2012) Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. Cell **151**(7), 1542-1556. http://dx.doi.org/10.1016/j.cell.2012.11.024

Mathieu M, Martin-Jaular L, Lavieu G and Théry C (2019) Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. Nature Cell Biology **21**(1), 9-17. http://dx.doi.org/10.1038/s41556-018-0250-9

Sancho-Albero M, Navascués N, Mendoza G, Sebastián V, Arruebo M *et al.* (2019) Exosome origin determines cell targeting and the transfer of therapeutic nanoparticles towards target cells. J Nanobiotechnology **17**(1), 16. http://dx.doi.org/10.1186/s12951-018-0437-z

Shetty AK and Upadhya R (2021) Extracellular Vesicles in Health and Disease. Aging Dis **12**(6), 1358-1362. http://dx.doi.org/10.14336/ad.2021.0827

Tang TT, Lv LL, Lan HY and Liu BC (2019) Extracellular Vesicles: Opportunities and Challenges for the Treatment of Renal Diseases. Front Physiol **10**, 226. http://dx.doi.org/10.3389/fphys.2019.00226

Vlachakis D, Mitsis T, Nicolaides N, Efthimiadou A, Giannakakis A *et al.* (2021) Functions, pathophysiology and current insights of exosomal endocrinology (Review). Mol Med Rep **23**(1), 26. http://dx.doi.org/10.3892/mmr.2020.11664

Willms E, Cabañas C, Mäger I, Wood MJA and Vader P (2018) Extracellular Vesicle Heterogeneity: Subpopulations, Isolation Techniques, and Diverse Functions in Cancer Progression. Front Immunol **9**, 738. http://dx.doi.org/10.3389/fimmu.2018.00738

Xia X, Wang Y, Qin Y, Zhao S and Zheng JC (2022) Exosome: A novel neurotransmission modulator or non-canonical neurotransmitter?

Ageing Research Reviews **74**, 101558. http://dx.doi.org/https://doi.org/10.1016/j.arr.2021.101558

Yu H, Sun T, An J, Wen L, Liu F *et al.* (2020) Potential Roles of Exosomes in Parkinson's Disease: From Pathogenesis, Diagnosis, and Treatment to Prognosis. Front Cell Dev Biol **8**, 86. http://dx.doi.org/10.3389/fcell.2020.00086

Zeng B, Chen T, Luo JY, Zhang L, Xi QY *et al.* (2021) Biological Characteristics and Roles of Noncoding RNAs in Milk-Derived Extracellular Vesicles. Adv Nutr **12**(3), 1006-1019. http://dx.doi.org/10.1093/advances/nmaa124

Zeng Y, Qiu Y, Jiang W, Shen J, Yao X *et al.* (2022) Biological Features of Extracellular Vesicles and Challenges. Front Cell Dev Biol **10**, 816698. http://dx.doi.org/10.3389/fcell.2022.816698

Zhang Y, Bi J, Huang J, Tang Y, Du S *et al.* (2020) Exosome: A Review of Its Classification, Isolation Techniques, Storage, Diagnostic and Targeted Therapy Applications. Int J Nanomedicine **15**, 6917-6934. http://dx.doi.org/10.2147/ijn.s264498

Zhang Y, Liu Y, Liu H and Tang WH (2019) Exosomes: biogenesis, biologic function and clinical potential. Cell Biosci **9**, 19. http://dx.doi.org/10.1186/s13578-019-0282-2

EMBnet.journal
Bioinformatics in Action

on a tightrope

Vivienne Baillie Gerritsen

Too much of anything is never good. Excess alcohol, and our faculties are impaired. Excess heat, and drought spreads. Excess cold, and vineyards die. Too much, too, of what is paradoxically essential to life frequently turns out to be toxic. Consider oxygen, iron, zinc or vitamins to name but four. Though we may be acquainted with the symptoms of what 'too much' entails, these are merely the superficial echo of cells under stress. Over the aeons, and throughout the living kingdom, organisms have had to deal with periodical over-abundances of many things. While selecting systems to use them in small doses, they promoted others to keep them in check. As an illustration, iron is vital for ferrying oxygen in organisms, and it is crucial in DNA synthesis, DNA repair and other fundamental cellular processes. Yet, too much iron will kill a cell – a process known as ferroptosis. Though this may be an ideal way of ridding a tissue of unhealthy cells, alternative processes have also evolved to stabilise things and prevent ferroptosis. One of these processes involves a protein known as ferroptosis suppressor protein 1, or FSP1.

Danseuse, by Helen Phillips

The Annex Galleries, Santa Rosa, California

Ferroptosis is defined as the death of a cell brought about by an overwhelming presence of iron. Though iron is vital for all species – as it is required to transport and deliver oxygen to organs, to ferry electrons in mitochondria or as a cofactor for instance – too much of it can be toxic as it can hinder a cell's antioxidant capacity. This means that noxious 'lipid reactive oxygen species' begin to accumulate, ultimately leading to oxidative cell death where, in the case of ferroptosis, plasma membranes typically rupture and mitochondria shrink while the cells, swell. This is unlike other programmed cell deaths, such as apoptosis for instance where cells typically bleb and diminish in size, which is why scientists regard ferroptosis as a category apart.

What exactly is meant by 'oxidative cell death'? In the case of ferroptosis, this implies lipid peroxidation that is driven by an iron-containing enzyme lipoxygenase. Lipid peroxidation involves the degradation of lipids caused by the addition of molecular oxygen which attacks their carbon-carbon double bonds. As the main constituent of plasma membranes are phospholipids, these present an ideal target for peroxidation modification which interferes not only with the assembly of plasma membranes but also with its structure and dynamics. As a result, if nothing is done to counter peroxidation, plasma membranes are damaged and the cell eventually dies.

Molecules known as antioxidants regulate the level of oxidation in a cell. CoQ_{10} is one. When CoQ_{10} was first discovered, it was called 'vitamin Q_{10}', where 'Q' stands for 'quinone' and '10' refers to the number of isoprenyl chemical subunits in the molecule's tail. In time, vitamin Q_{10} was renamed 'ubiquinone' because of its ubiquitous presence in lipid membranes. CoQ_{10} is a lipid-soluble antioxidant meaning that it can slip into the plasma membrane of

cell organelles, such as the mitochondrion, the endoplasmic reticulum or the Golgi apparatus. Here, ubiquinone must remain in a reduced state – $CoQ_{10}H_2$ also known as ubiquinol – so as to halt the propagation of lipid peroxides. How is ubiquinol kept in a reduced state in animals? Thanks to FSP1.

Ferroptosis suppressor protein 1 (FSP1) is an enzyme or, more specifically, a CoQ_{10} plasma membrane oxidoreductase. This is where the 'Co' springs from in CoQ_{10} since the molecule is also referred to as 'coenzyme Q_{10}' as it is required by FSP1 for its catalytic activity. Like its coenzyme, FSP1 is also present in plasma membranes and probably targeted there thanks to a post-translational modification known as myristoylation which is the addition of a long fatty acid at the enzyme's N-terminus. Myristoylation not only directs FSP1 towards plasma membranes but also helps it squeeze between the phospholipids. So now we have CoQ_{10} and FSP1 in the plasma membrane – and all we need to suppress ferroptosis is a source of hydrogen which is supplied by the omnipresent and universal cofactor NADPH. Thanks to NADPH, FSP1 can produce the reduced form of ubiquinone, $CoQ_{10}H_2$ or ubiquinol, which is then armed to fight lipid peroxidation.

Having sorted out this rather complex biochemistry, another molecule popped up: vitamin K. Vitamin K, like CoQ_{10}, belongs to the family of quinones and is prescribed in substantial doses to patients taking warfarin, a popular blood thinner. Why? Because vitamin K counters warfarin poisoning – but no one really understood why. It turns out that vitamin K is also involved in suppressing ferroptosis. Indeed, vitamin K and CoQ_{10} happen to share similar structural properties. Consequently, vitamin K can take the place of CoQ_{10} in FSP1, where it is reduced. The resulting reduced vitamin K (VKH_2), like $CoQ_{10}H_2$ is also a potent inhibitor of lipid peroxidation. Scientists then realised that the reduction of vitamin K led by FSP1 is also responsible for the effect vitamin K has against warfarin poisoning.

Cells are able to cope with certain levels of lipid peroxidation; it just must never reach levels that become detrimental to the cells. It all comes down to equilibrium. This said, FSP1 is not the only enzyme involved in taming ferroptosis; there are other pathways too, each of which seem to supplement the other. In recent years, medical scientists have taken a growing interest in ferroptosis as it seems to be at the heart of neurodegenerative diseases such as Alzheimer's and Parkinson's but also diseases like cancer. Promoting ferroptosis could kill cancer cells for example while, in neurodegenerative diseases, checking ferroptosis could keep neurons alive. This would make FSP1 an ideal therapeutic target besides proving, perhaps, to be a good biomarker.

Life certainly tiptoes along a narrow tightrope. From an evolutionary point of view, the involvement of components as biologically vital as iron and vitamin K in ferroptosis is intriguing. Iron has an essential role in life that dates back billions of years. When oxygen started to accumulate in the Earth's atmosphere and organisms began to use oxygen to drive many of their vital processes which were already dependent on iron, they were concomitantly building one of Nature's conundrums: she was going to have to find a way of keeping toxicity at bay. In this light and long before the existence of FSP1, vitamin K could actually be the most ancient member of anti-ferroptotic quinones since ferroptosis is a cell-death mechanism that has been conserved from prokaryotes to plants and mammals. Fascinating.

Cross-references to UniProt

Ferroptosis suppressor protein 1, *Homo sapiens* (Human): Q9BRQ8

References

1. Mishima E., Ito J., Wu Z., et al.
   A non-canonical vitamin K cycle is a potent ferroptosis suppressor
   Nature 608: 778-783(2022)
   PMID: 35922516

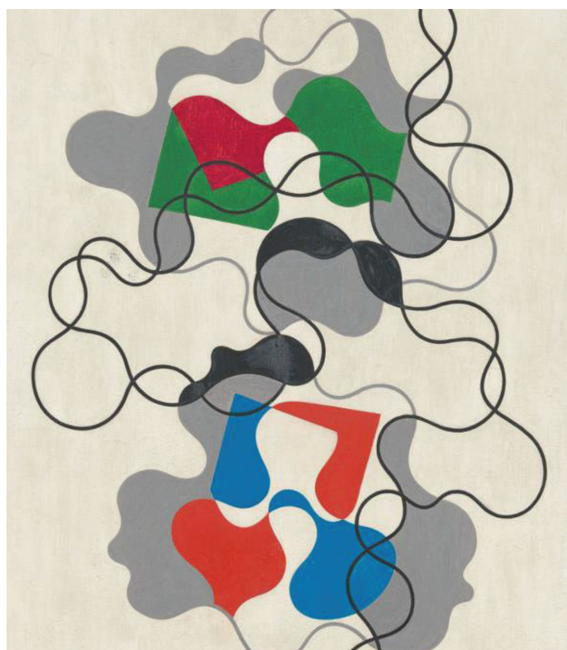
2. Doll S., Porto Freitas F., Shah R., et al.
   FPS1 is a glutathione-independent ferroptosis suppressor
   Nature 575: 693-698(2015)
   PMID: 31634899

proteinspotlight
Swiss Institute of Bioinformatics
Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

a shrewd tweak

Vivienne Baillie Gerritsen

The chairs were rickety. So I rummaged around the kitchen drawer, extracted an old knife and used its tip to drive a few screws back into the wood. The knife kept on losing grip and I kept on swearing. The fastest and least infuriating way to have done the job would have been to go down to the cellar and find a screwdriver. Both utensils can be used to drive in screws, but one has been intentionally manufactured to perform just that, simultaneously reducing the time and energy involved. Nature, too, has its screwdrivers. Given time and chance, it will always take the opportunity to select a commodity which will make things, if not easier, at least more in tune with what is needed. One example: ribosomes are huge molecular complexes whose role is to synthesize proteins in cells. Until recently, it was thought that all ribosomes were alike. A bit like kitchen knives. However, it turns out that some ribosomes differ slightly in their makeup and are found only in certain kinds of cell – presumably because they synthesize proteins particular to these cells. One such ribosome has been discovered in sperm cells, along with a protein known as large ribosomal subunit protein eL39-like*, or RPL39L.



Lines of Summer (detail), 1942

by Sophie Taeuber-Arp

Ribosomes are among the most fundamental molecular complexes to be found in organisms. And with good reason: they are where cells synthesize proteins. Ribosomes are themselves an assemblage of various proteins mingled with RNA, known as ribosomal RNA (rRNA). And yes, ribosomes need ribosomes to synthesize the proteins that are part of their own constitution. From bacteria to fungi, plants, insects and mammals, ribosomes are all built according to the same architectural plan: one large subunit, one small subunit, and a handful of rRNA. Prokaryotic ribosomes are composed of one large subunit, itself a complex of about 30 proteins and two rRNAs, and a small subunit of about 20 proteins and one rRNA. In eukaryotes, the large subunit is characteristically composed of 47 proteins and three rRNAs, and the small subunit of 33 proteins and one rRNA.

Uniting as many as 47 proteins and three rRNAs into one large ribosomal subunit – which does not fall apart and performs the multiple tasks involved in protein synthesis – demands careful assembly. Making a ribosome is like building a factory while also hiring employees to carry out different tasks. In eukaryotes, everything begins in the nucleolus, a region located within the cell nucleus that is dedicated to the first steps of ribosome formation. Here the rRNAs are prepared and added to pre-ribosomal proteins. Both the small and the large subunits then begin, independently, to migrate towards the cell cytoplasm, where they will finally bind to one another, ready to do their job. During their migration, the subunits slowly mature as the parts which make them up are folded, processed, rotated, checked and finally channelled through the nuclear membrane.

In ribosomes, each protein has a different role – as do the rRNAs. Certain proteins are needed to stabilize the overall architecture while others help the large

and small subunits assemble, and of course you have the proteins that are involved in exporting the ribosome from the nucleolus to the cytoplasm in the first place. Then you need the proteins that take an active part in protein synthesis *per se* – a process so incredibly delicate and intricate, it takes any student days, if not weeks, to grasp. Briefly – so much so, it may seem criminal – to synthesize one protein sequence, ribosomes read and translate (into amino acids) genes from their mRNA. The required amino acids are taken from the cell cytoplasm by tRNAs and added, in the correct order, to a growing protein chain which protrudes from a spot on the large subunit of the ribosome. A spot known as the nascent polypeptide exit tunnel, or NPET. When the sequence is complete, the protein slips out of the NPET and is sent to where it is needed in the cell, or outside the cell. What do rRNAs do? Like proteins, certain RNAs also have roles. In particular, rRNAs help ribosomes assemble the amino acids in the correct order – which is of utmost importance. With all the molecules and the many steps involved in protein synthesis, needless to say, it is one of the most costly activities of a cell, lapping up over 70% of its energy!

Around the NPET, ribosomes do an extra bit of quality checking, to make sure that the genes have been correctly read and their sequence properly assembled. If they have not, the faulty nascent chain is directed towards another part of the cell where it is disposed of. Recently, researchers discovered that the very end of NPET in mammalian sperm cells differs from its counterparts and seems to have become specialised. As in other cells, protein sequences are double-checked in sperm NPETs, but sperm NPETs also seem to keep an extra good look out for proteins that are essential for sperm function. This would imply that sperm NPETs do not serve the exact same purpose as NPETs in other ribosomes, which turns out to be the case. A protein known as RPL39 is usually located in NPETs where it forms part of the wall. In mammalian sperm, RPL39 has been replaced by a paralog, termed RPL39L.

RPG39L illustrates well the advantage of heterogenous ribosomes. Though RPL39L and its paralog RPL39 differ by only three amino acids, it is enough to impart to RPL39L roles that are advantageous to sperm. What is more, RPL39L and RPL39 are not interchangeable as is often the case with paralogs, further suggesting that RPL39L is cell-specific. RPL39L seems to have been tweaked to monitor the well-being of sperm in particular. In which way? Sperm motility requires a lot of energy. Few cells need to travel so far – what is more by their own means – to reach their destination. Such a trip requires energy, or ATP, which is produced by a cell's mitochondria. It so happens that RPL39L has a role in mitochondrion formation. In its absence, the organelles are malformed and sperm motility is defective. This could be explained by a role for RPL39L in double-checking the correct formation of mitochondrial proteins in the NPET. Moreover, RPL39L also seems to be involved in the faithful assembly of the large ribosomal subunit which, besides quality checking, is paramount to protein synthesis.

At the very heart of what could be defined as life, ribosomes represent the elemental passage from DNA to protein. The need for protein is unending within any living organism as each of our activities – visible or invisible – requires it. Cell homeostasis, on which life depends, relies on protein homeostasis. So, too, does sperm motility, and hence fertility. If sperm are unable to wriggle and swim because their flagella are not beating properly, they will never reach their destination. Or if they do, they may not have the wherewithal to fight and forage their way through the egg's coat and fertilize it. RPL39L could therefore constitute a therapeutical target to help counter infertility – at least the type of infertility caused by sperm that lack stamina.

** this is the protein's new name which will be used from July 2023; otherwise, it is known as: 60S ribosomal protein L39-like*

Cross-references to UniProt

60S ribosomal protein L39-like / large ribosomal subunit protein eL39-like, *Homo sapiens* (Human) : Q96EH5
60S ribosomal protein L39-like / large ribosomal subunit protein eL39-like, *Mus musculus* (Mouse) : Q9CQD0

References

1. Li h., Huo Y., He X. *et al.*
   A male germ-cell-specific ribosome controls male fertility
   Nature 612:725-731(2022)
   PMID: 36517592

2. Zou Q., Yang L., Shi R. *et al.*
   Proteostasis regulated by testis-specific ribosomal protein RPL39L maintains mouse spermatogenesis
   iScience 2021 Oct 30;24(12):103396
   PMID: 34825148

Swiss Institute of Bioinformatics
proteinspotlight
Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

self-reliant

Vivienne Baillie Gerritsen

In times of perplexity, every now and then it is better to deal with things yourself. It can be a time-saver, and sometimes, too, an energy-saver. If faced, say, by an oncoming downpour, it is wiser to run for shelter rather than wait for someone to bring you an umbrella. Such decisions also exist at the cellular level. Take our immune system for instance. When attacked by a virus, our body begins by rapidly firing off a first round of artillery as it awaits further and more complex lines of defence that involve myriads of other factors. It turns out that the use of fast lanes such as these also occur at a far smaller scale. In this light, scientists recently discovered quite an extraordinary protein, known as ophMA, that belongs to the fungus *Omphalotus orealius,* which methylates its own C-terminal tail instead of depending on another transferase to do the job. The tail is then cleaved and folds into a cyclic peptide, an omphalotin, that has anti-nematode properties. Besides its talent for independence, ophMA also adopts a rare catenane arrangement, very similar to two rings that have been interlocked.

The Wrestlers

by Henri Gaudier Brzeska (1891-1915)

Omphatolins are produced by *Omphalotus olearius* – a mushroom also known as Jack O'Lantern, because of its deep orange colour akin to that of pumpkins and its soft green bioluminescent glow at night. Omphatolins are cyclic highly-methylated peptides and have been known, since the turn of the 21st century, to specifically target the nematode *Meloidogyne incognita*, a common plant pathogen. Since the peptide's structure resembles that of another renowned fungal cyclic peptide, cyclosporine, scientists assumed that omphatolins were synthesized in the same way. That is to say, their synthesis does not rely on the help of ribosomes but rather on huge multi-modular enzymes called non-ribosomal peptide synthetases (NRPs). Such enzymes are able to build peptide sequences – i.e. to create amide bonds – without resorting to the ribosome's complex machinery.

No one, however, was able to pin down what was making omphalotin. Until, much to everyone's surprise, its sequence was spotted on *Omphalotus olearius*' genome. This could imply that omphalotin is synthesized by a ribosome – something no one had expected, let alone considered, given its similarity to cyclosporine. As their name implies, RiPPs (for **Ri**bosomally synthesized and **P**osttranslationally modified **P**eptides) are synthesized via the regal ribosomal machinery with its tango of mRNA, rRNA, amide bonding, post-translational modifications and proteolytic cleavage. This turned out to be partly true for omphalotin: it is a RiPP, but of a different kind. Omphalotin actually forms the C-terminal end of a far longer sequence whose N-terminal end is a methyltransferase – the very methyltransferase that methylates omphalotin! Since this discovery, many other peptides of the same nature continue to surface and are now collectively known as borosins, i.e. macrocyclic peptides whose leader sequence (*gross modo* the methyltransferase) is very long.

OphMA designates both the methyltransferase (M) and omphalotin (A). About 400 amino acids long, ophMA is divided, roughly, into three main

domains: a large N-terminal methyltransferase domain, followed by what has been called a clasp domain and then a C-terminal core peptide that will give rise to the cyclic peptide omphalotin – itself barely twelve amino acids long. When omphalotin is required, the methyltransferase methylates nine out of twelve of the peptide's amide bonds. Methylation is believed to occur, amide bond by amide bond and in the same direction, either as the peptide is being cyclised or once it has been cyclised.

The process sounds straightforward but is actually quite complex as it involves a very rare chemical arrangement known as a catenane. A catenane is a molecular architecture where two – or more – macrocycles are interlocked in the manner of intertwined rings. Remember the Christmas decorations you made at school? Where you took a strip of coloured paper that you glued at the ends to form a circle? Then you took a second strip, inserted one end through the paper circle you had just made, and glued its ends? That is a two-ringed catenane – which can only be disrupted if the covalent bonds holding one ring are broken, or if you tear one of the strips of coloured paper.

OphMA acts as a homodimer where each monomer locks into the other – just like our Christmas decorations – to adopt a catenane arrangement. How does it happen? First, the methyltransferase and the clasp domains of one monomer move towards each other to form a ring-like structure. The clasp domain of one monomer then wraps around the methyltransferase domain of the other, and the core domain (i.e. the future omphalotin) of one monomer inserts itself into the active site of the other monomer. The result is a very rare catenane arrangement of two enzymes – most probably providing stability to ophMA.

The active site becomes an extended hydrophobic tunnel in which sits the peptide substrate, ready to be methylated. Conformational changes abound as methylation occurs – with the help of a cofactor, S-adenosyl methionine or SAM – on the amide bonds, one transfer after another, as the substrate performs 180 degree flips to bring the next residue into the active site. Consequently, each methyltransferase in the ophMA dimer methylates the substrate peptide that belongs to its monomer as opposed to its own.

It is a wonderful strategy. A sort of "you scratch my back, I'll scratch yours". What is more, with the methyltransferase at the substrate's immediate disposal, there is no need to find an external transferase, so to speak, which would only involve additional energy- and time-consuming biological processes. The methylated peptide is cleaved by another enzyme (called ophP) and cyclised to yield the finished cyclic peptide with nine methylated amide bonds conferring both stability and cell permeability. At this point, one would imagine that the catenane arrangement has collapsed – or is on the point of collapsing – as the methyltransferase is discarded and omphalotin is sent to its target.

Despite knowing in great detail how omphalotin is made, how it is toxic to *Meloidogyne incognita* is still not understood. Perhaps, like cyclosporine, it is able to glide easily through the cell membranes and block important enzymes in the cytoplasm. Although no one yet knows which enzymes are blocked… What is particularly interesting for researchers is that omphalotin is only specific to *Meloidogyne incognita*, a known and wide-spread plant pathogen, yet it is not toxic to bacteria or other fungi. So, as a pesticide against a given crop, it should do little harm to the environment. What is more, ophMA does not seem to be very watchful and joyfully methylates residues that have been artificially replaced by others – which could help scientists suggest fine-tuned versions of the peptide for a given therapeutical or agricultural use.

Cross-references to UniProt

Methyltransferase/ribosomally synthesized cyclic peptide omphalotin A precursor ophMA, *Omphalotus olearius* (Jack o'lantern): A0A2R2JFI5

References

1. Ramm S., Krawczyk B., Mülenweg A., *et al.*
   A self-sacrificing *N*-methyltransferase is the precursor of the fungal natural product omphalotin
   Angew. Chem. Int. Ed. 56:9994-9997(2017)
   PMID: 28715095

2. Song H., Naismith J.H.
   Enzymatic methylation of the amide bond
   Current Opinion in Structural Biology 65:79-88(2020)
   PMID: 32653730

proteinspotlight
Swiss Institute of Bioinformatics

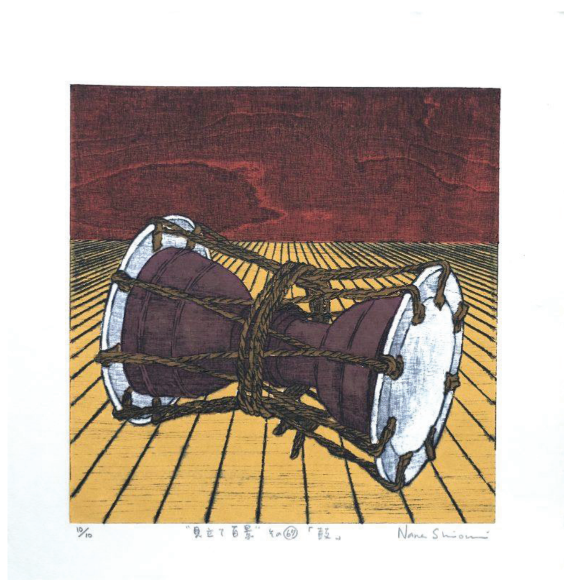
Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

EMBnet.journal
Bioinformatics in Action

sound and silence

Vivienne Baillie Gerritsen

Like smells and tastes, sounds can whizz you back to forgotten places. The shriek of a seagull. The wash of waves. The crack of lightning. A motor's rumble. A Christmas carol. A childhood tune. More often than not, these castaway memories emerge wrapped in a delicate veil of magic. It is a wonderful feeling, of something you would like to know again but cannot, although it is there hidden deep inside you. A feeling we would be unable to remember were it not for our ears. Sound is sensed by way of vibrations that hit parts of our inner ear, an intricate part of mammalian anatomy. Here, vibrations are amplified, causing messages to be relayed to our brain that translates them into noise – which is what the act of hearing is. All sorts of proteins are involved in the perception of sound* but also in its unfortunate contrary: hearing impairment. One particular protein, connexin 26, forms intercellular channels in the cochlea of the inner ear so that molecules can transit from one cell to another. Connexin26 also happens to be involved in many forms of hearing impairment because, when dysfunctional, molecules are no longer able to pass.



One Hundred Views of Mitate No.67 Drum
woodcut print, 2004

by Nana Shiomi

Connexin 26, or Cx26, is one of several proteins that form channels which gather at specific locations – called gap junctions – on cell membranes. Each channel spans the plasma membrane of one cell, crosses a thin extracellular region before plunging into the plasma membrane of a neighbouring cell – thus allowing the passage of molecules from one cell to another. Gap junctions are a means for cells to communicate with one another, while keeping things levelled and regulated, a little like orchestrating a roundtable without letting anything trouble the theme under discussion. As can be expected, gap junctions are expressed in a wide variety of cells, such as neural cells where they support neural differentiation and proliferation, cardiac cells where they dictate contraction, in the lens where mutations can bring about cataracts and in the inner ear to help us hear. In this respect, several mutations in Cx26 are responsible for certain forms of hearing impairment (HI) – an ailment which affects 1 to 3 children out of 1000 at birth or during early childhood.

Gap junctions are dynamic structures, with channels being replaced and recycled on a regular basis. The channels themselves are formed in several steps. To begin with, half-channels are produced within each cell and whipped off to the plasma membrane along microtubular tracks, probably as part of a secretory pathway. In this way, one cell produces one half of a future channel, while the next one over produces the second half. Once in the plasma membrane, each half-channel then travels to a docking region in a gap junction, probably with the help of cytoskeletal actin. Once the two halves have reached their destination, they join in the extracellular region. Thus docked and locked, the two former half-channels now form an entire intercellular channel permeable to specific molecules.

Several proteins form gap junction channels in the mammalian inner ear, which is made up of three main parts: the cochlea, the labyrinth and the vestibule, all three bathed in various fluids. The cochlea is what we use for hearing, while the vestibule and the labyrinth support our sense of balance. Cx26 is widely expressed in the cochlea, where it is involved in the maintenance of ionic and metabolic homeostasis as well as in intercellular signalling. Potassium ions, K^+, are known to be at the heart of sensory transduction. Cx26 could be involved in maintaining cochlear homeostasis by sustaining the dynamics – removal and recycling – of K^+ within the organ. However, there is a great chance that Cx26 is involved in the flux of other molecules too, such as Ca^{2+} ions and inositol phosphates.

The half-channels discussed above are also known as connexons. Each connexon is a hexamer of monomers, or connexins. A fully-fledged gap junction channel is therefore a dodecamer, i.e. the union of two connexons. There are 21 different connexins (one of which is Cx26) in the human proteome, which combine to form homo- or heteromeric connexons. Connexons will then go on to form homo- or heterotypic channels. It is not hard to understand that such combinations create an astounding diversity of gap junction channel composition and function.

Cx26, in particular, has been extensively studied because of its role in sound transduction and hearing impairment. Each connexin has three distinctive structural elements essential for the channel's overall function: four transmembrane alpha helices, an N-terminal helix which protrudes into the lumen of the channel, and two extracellular loops that are essential for connexon docking. Imagine six sets of four alpha helices, i.e. a connexon, that joins to another six sets of four alpha helices to form a full channel. The overall ribbon representation looks like a wonderful firework of party streamers that is reflected on water. More prosaically, channels such as these have been compared to the Japanese hand drum, the tsuzumi, which is narrow in its centre while both ends widen out.

How do molecules travel from one cell to another? Are the channels always open? Or do they close? An elegant 'plug gating' model has been proposed. Each connexin's N-terminal helix (NTH) protrudes into the channel's lumen. When there is no difference between the cells' membrane voltages, the NTHs are held against the inside of the channel thus leaving room for molecules to pass, similar to you pressing against a wall to let something wide pass. When there is a difference in membrane voltage between the two cells, the NTHs are released from the sides and join in the channel's centre to form a plug, through which nothing can pass.

It is a wonderfully refined model. What is more, it seems that Cx26 frequently combines with another connexin (connexin 30), which could explain why waves of Ca^{2+} spread far faster through cells with a combined channel as opposed to one composed only of Cx26. It is easy to understand that if Cx26 is dysfunctional, the transduction of sound will be affected. So far, about 90 mutations have been characterised in the Cx26 gene, several of which directly influence hearing either because connexons are malformed, or because they are mistargeted or fail to dock for example. Today, roughly 6% of the world's population suffers from HI due to genetic and environmental factors. It is an affliction which spreads silently across the world. Understanding the nooks and crannies of our cochlea may help to bring sound back to those who have lost it.

**Protein Spotlight issue 22*

Cross-references to UniProt

Gap junction beta-2 protein, *Homo sapiens* (Human): P29033

References

1. Defourny J., Thiry M.
   Recent insights into gap junction biogenesis in the cochlea
   Developmental Dynamics 252:239-246(2022)
   PMID: 36106826

2. Maeda S., Tsukihara T.
   Structure of the gap junction channel and its implications for its biological functions
   Cell. Mol. Life Sci. 68:1115-1129(2011)
   PMID: 20960023

proteinspotlight
Swiss Institute of Bioinformatics
Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

EMBnet.journal
Bioinformatics in Action

ISSN 1023-4144

DEAR READER,

If you have any comments or suggestions regarding this journal we would be very glad to hear from you. If you have a tip you feel we can publish then please let us know. Before submitting your contribution read the "Authors guidelines"[1] and send your manuscript and supplementary files using our on-line submission system[2].

Past issues are available as PDF files from the web archive[3].

Visit EMBnet website for more information: www.journal.embnet.org

EMBNET.JOURNAL EXECUTIVE EDITORIAL BOARD

Editor-in-Chief

Erik Bongcam-Rudloff
Department of Animal Breeding and Genetics, SLU, SE
erik.bongcam@slu.se

Deputy Editor-in-Chief

Dimitrios Vlachakis
Assistant Professor, Genetics Laboratory, Department of Biotechnology
Agricultural University of Athens, GR
dimvl@aua.gr

Editorial Board Secretary

Laurent Falquet
University of Fribourg & Swiss Institute of Bioinformatics
Fribourg, CH
laurent.falquet@unifr.ch

Executive Editorial Board Members

Emiliano Barreto-Hernandez
Biotechnology Institute, National University of Colombia
Bogota, CO
ebarretoh@unal.edu.co

Domenica D'Elia
Institute for Biomedical Technologies, CNR, Bari, IT
domenica.delia@ba.itb.cnr.it

Sissy Efthimiadou
Agricultural Research Institute ELGO Dimitra, GR
sissyefthimiadou@gmail.com

Elias Eliopoulos
Genetics Lab, Biotechnology Department, Agricultural University of Athens, GR
eliop@aua.gr

Andreas Gisel
CNR, Institute for Biomedical Technologies, Bari, IT
andreas.gisel@ba.itb.cnr.it
International Institute of Tropical Agriculture, Ibadan, NG
a.gisel@cgiar.org

Lubos Klucar
Institute of Molecular Biology, SAS Bratislava, SK
klucar@EMBnet.sk

Assistant Editors

Eleni Papakonstantinou
Agricultural University of Athens, GR
eleni.ppk@gmail.com

Katerina Pierouli
Agricultural University of Athens, GR
pierouli.katerina@gmail.com

Gianvito Pio
Department of Computer Science, University of Bari Aldo Moro, IT
gianvito.pio@uniba.it

PUBLISHER

EMBnet Stichting p/a
CMBI Radboud University
Nijmegen Medical Centre
6581 GB Nijmegen
Th e Netherlands

Email: erik.bongcam@slu.se
Tel: +46-18-67 21 21

[1]http://journal.embnet.org/index.php/embnetjournal/about/submissions#authorGuidelines
[2]http://journal.embnet.org/index.php/embnetjournal/author/submit
[3]http://journal.embnet.org/index.php/embnetjournal/issue/archive