

NEWSLETTER DU HUB



EN NOVEMBRE AU HUB NOVEMBER AT THE HUB

Journées départementales du DBC	1
<i>DBC Department days</i>	
Publications	4
<i>Publications</i>	
Dernières nouvelles	4
<i>Latest news</i>	
Défi du mois	5
<i>Monthly challenge</i>	
Nouveaux-elles membres	6
<i>New members</i>	
Ce mois-ci	6
<i>This month</i>	

JOURNÉES DÉPARTEMENTALES DU DBC

DBC department days

Date : Du 29 Novembre au 01 Décembre 2021

Lieux : Center Parcs - Les bois francs



Scientific Program

Monday November 29, afternoon

Session chair: R. Chikhi

- 14:15 C. Zimmer: Welcome and introduction
- 14:30 Imaging and Modeling Unit
 - C. Zimmer: Imaging and modeling high resolution chromatin structure
 - J. Bai: Enhancing super-resolution microscopy with deep learning
- 15:00 Bioinformatics & Biostatistics Hub
 - **C. Malabat: Overview of the Hub**
 - **G. Millot: A New Route For Integron Cassette Dissemination Among Bacterial Genomes**
- 15:20 Microbial Evolutionary Genomics Unit
 - J. Moura de Sousa: To catch a hijacker: abundance, evolution and genetic diversity of P4-like bacteriophage satellites
- 15:40 InBio
 - G. Batt: Experimental and Computational Methods for Modeling Cellular Processes
 - J. Ruess: From Single Cells To Populations And Back: Optogenetic Control Of Microbial Communities

Session chair: David Di Gregorio

- 17:00 Microbial Paleogenomics Unit
 - N. Rascovan: The computational challenges of Microbial paleogenomics
- 17:20 Systems Biology Group
 - B. Schwikowski: The Systems Biology Group
 - E. Zakiev: Search for gene signatures in ovarian cancer
- 17:50 Bioinformatics & Biostatistics Hub
 - **E. Permal: Overview of Hub contribution through Covid pandemic crisis**
- 18:10 Direction des Applications de la Recherche et des Relations Industrielles
 - A. Borderia & E. Brient-Litzler: How algorithms become products: key challenges and examples
- 18:15 Poster session 1

Tuesday November 30, morning

Session chair: Thomas Gregor

- 9:30 Sequence Bioinformatics Group
 - R. Chikhi: The Sequence Bioinformatics group
 - L. Denti: Human structural variant detection with accurate long reads
- 10:00 Structural Bioinformatics Unit
 - O. Sperandio: In silico contributions to the design of therapeutic compounds against protein-protein interactions

Session chair: Nicolas Rascovan

- 10:50 Spatial Organisation of Genomes Unit
 - A. Cournac: Spatial Organization of Genomes
- 11:10 Institut Necker Enfant Malades
 - M. Pontoglio: HNF1B, a bookmarking transcription factor at the heart of cell quiescence/differentiation
- 11:30 Institut Necker Enfant Malades
 - G. Panasyuk : Metabolomics sheds light on surprising function of the ancient signaling pathway in maintaining metabolic homeostasis
- 11:50 Physics of Biological Function Unit
 - T. Gregor: Linking DNA polymer and transcription dynamics

Tuesday November 30, afternoon

Session chair: Ganna Panasyuk

- 14:15 Keynote 1:
 - J-P. Vert: Deep language models for DNA and proteins
- 15:00 Synapse and Circuit Dynamics Unit
 - D. DiGregorio: Temporal predictions using biological neural networks
- 18:15 Poster session 2

Wednesday December 1, morning

Session chair: Benno Schwikowski

- 9:30 Keynote 2:
 - I. Bloch: Hybrid AI for biomedical image understanding – Towards explainability
- 10:55 Decision and Bayesian Computation Group
 - **F. Laurent: Single-molecule localization microscopy**
- 11:15 Bioinformatics & Biostatistics Hub
 - **O. Mirabeau: Spatial transcriptomics in Zebrafish**
- 11:35 Bioinformatics & Biostatistics Hub
 - **R. Torchet + H. Julienne: Gender Speaking Differences in Academia**
- 11:55 Statistical Genetics Group
 - H. Aschard: Genetic Epidemiology in the Big Data era
 - A. Frouin: Use of new kernel matrix for bacterial GWAS
 - C Boetto: Dissecting The Predictors Of Microbiome Variability

Wednesday December 1, afternoon

Session chair: Gaël Millot

- 14:15 Centre de Ressources en information scientifique (CeRiS)
 - A-C Delétoille: Policy on the management and sharing of research data and software code
- 14:30 Direction des Systèmes Informatiques (DSI)
 - Y. Ghorbal: HPC resources @Pasteur (past, current and future)
- 14:45 Metagenomic signatures
 - S. Kennedy: Perinatal mother-infant microbiota
- 15:05 Evolutionary Bioinformatics Unit
 - **A. Zhukova: Cuban history of CRF19 recombinant subtype of HIV-1**
- 15:25 G. Batt: Closing words

La retraite départementale du DBC a eu lieu au Center Park du lundi 29 Novembre au mercredi 1er décembre. Une dizaine de membre du Hub ont participé soit pour des poster ou des présentations. Cela a permis des rencontres avec d'autres personnes du département et des échanges fructueux. Elise Jacquemet a gagné le prix du meilleur poster pour le groupe Stats et toutes les présentations du Hub ont été bien accueillies. Un grand merci à Vincent Guillemot pour son implication dans l'organisation de cet évènement.

CBD retreat was located at Center Park resort, Monday 29th November until Wednesday 1st December. Around ten Hub members were present to either do talks or poster presentations. It helps meeting other people from the CBD and gave fruitful exchanges. Elise Jacquemet from Stats group won the best poster contest, and all Hub presentations were well appreciated. Special thanks to Vincent Guillemot who actively participate to the organization.

Toutes les informations de la retraite sont disponibles à cette adresse : <https://research.pasteur.fr/en/event/2021-computational-biology-department-days/>

All information on this retreat is available on the following website: <https://research.pasteur.fr/en/event/2021-computational-biology-department-days/>

The Four Horsemen of neglecting experimental design

Elise JACQUEMET¹, Emeline PERTHAME¹, Stevven VOLANT¹, Thomas OBADIA¹, Hugo VARET¹, François LAURENT¹, Pascal CAMPAGNE¹

¹ Institut Pasteur, Université de Paris, Hub de Bioinformatique et Biostatistique, F-75015 Paris, France

Introduction

Reproducibility of experimental results has become a central concern in biological sciences. Increasing attention is brought on how to carefully design experiments and choose appropriate statistical tools. Controlled experiments may be hard to conduct as they are subjected to many technical constraints, especially in animal experiments. Regular issues are such that (i) experimental outcomes may not be properly designed, statistically speaking; (ii) basic statistical tools may not be suited to properly analyse data that are partly shaped by technical constraints. Both issues are not rare in studies made at Pasteur and their impact on results are seldom considered. A downside of neglecting experimental design is an uncontrolled inflation of false-positives that may occur well above the typical 5% control threshold. Furthermore, a lack of power insidiously leads to overestimating the magnitude of experimental effects, all the more when they are coupled with confounding factors, due to ill-defined designs. Here we present four problems we often encounter in our daily life of data analysts. We illustrate why & how they may drastically alter the validity of some scientific findings.

Carrying out underpowered studies

Intrepid biologist speaking: I work on a contagious disease that require mice to be kept in an isolator. This will be a very expensive experiment, I have no other choice than working with reduced sample sizes. This should be fine since I expect large effects...

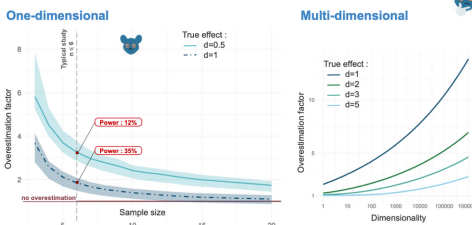


Fig 1 - Over-estimation of experimental effects in reduced sample size due to a lack of power. Data were simulated to compare mean values obtained in two experimental groups with a t-test (using fixed values of d , i.e. the standardised difference of means among groups, and various sample sizes). Standardised differences among groups were recorded only when the p -value ≤ 0.05 (10,000 simulated comparisons for each parameter combination). The curves and envelopes represent the average and interquartile ranges of the Overestimation factor (i.e., the ratio between estimated and true effect).

Sequentially recruiting samples

Intrepid biologist speaking: I want to investigate the experimental effect of two treatments in zebrafish, but I have no guess about the strength of this effect. To minimize the number of fish used in my study, I will make a first experiment with a small sample size and I will repeat this experiment until I find a significant difference.

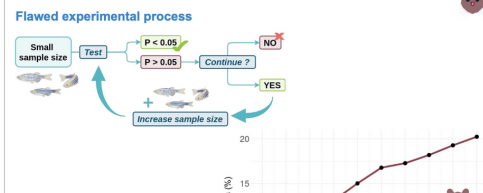


Fig 3 - Inflation of false-positives when repeating an experiment, due to sequential recruiting. Data were simulated to compare mean values obtained in two experimental groups with a t-test, under H_0 . A repetition consists in making the same experiment again with a constant sample size (adding 3 individuals per group). The figure reads as follow: e.g. after recruiting samples with 5 sequential repetitions, the rate of false positives reaches 15% instead of remaining at 5% as mistakenly assumed.

Confounding experimental effects

Intrepid biologist speaking: For convenience reasons I will prepare libraries and proceed to sequencing samples as they come, by experimental group. I should be able to catch the effects of both Age and Sex anyhow.

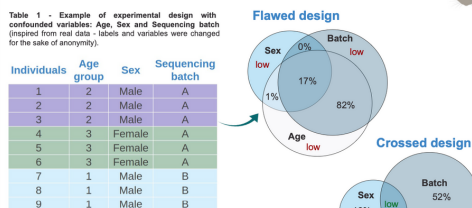


Fig 4 - Partition of variance. Venn diagrams representing the amount of variation explained by 3 variables: age, sex and sequencing batch. Data were simulated by attributing identical additive effects of variables across two experimental designs: (i) a flawed design, as presented in Table 1 and (ii) a proper blocked design where all experimental and technical factors are crossed. Partition of variance was obtained by using linear models (such as ANOVA).

Overlooking nestedness in experiments

Intrepid biologist speaking: I work on an infectious disease using a mouse model. I made an experiment with 2 groups of 18 mice. To avoid contamination among individuals, I can't mix the two groups in the same cages and I will proceed with batches (i.e., in each group: 3 cages of 6 mice). Then I will analyse data with usual statistical tests (e.g., Student's test, etc.).

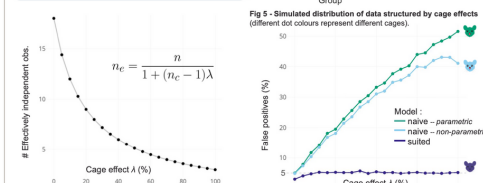
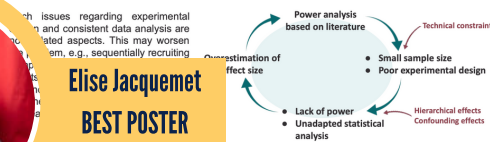


Fig 5 - Hierarchical (cage) effects λ decrease the effective number of independent observations (n_e). The cage effect λ is the fraction of variance bound to consistent differences among experimental units (cages). Consequence is an overestimated test-statistic value, leading to an uncontrolled inflation of false positives (see Fig. 7). Parameters: $n=18$, $n_c=6$ animals per cage.

A non-reproducibility loop



Support & tools provided by the Hub





PUBLICATIONS

Publications

Devosia equisanguinis sp. nov., isolated from horse blood. Kämpfer P, Busse H-J, Clermont D, **Criscuolo A**, Glaeser SP. *International Journal of Systematic and Evolutionary Microbiology* 2021 Nov. doi: 10.1099/ijsem.0.005090

OCT4-mediated inflammation induces cell reprogramming at the origin of cardiac valve development and calcification. Farrar EJ, Hiriart E, Mahmut A, Jagla B, Peal DS, Milan DJ, Butcher JT, Puceat M. *Sci. Adv.* 7 (7). 2021 Nov. doi: 10.1126/sciadv.abf7910

DERNIÈRES NOUVELLES

Latest news

wgetENAHTS v4.0

wgetENAHTS is a command line program written in Bash to download gzipped FASTQ files from the European Nucleotide Archive (ENA) ftp repository. Every download is performed using the standard tool wget.

Update : Significant running time reduction by automatically selecting the best downloading protocol and by reimplementing the file integrity assessment step

Infos : <https://gitlab.pasteur.fr/GIPhy/wgsENAHTS>

DNA2ORF v1.0

DNA2ORF is a command line program written in Java to find and translate open reading frames (ORF) inside nucleotide sequence(s), an ORF being unequivocally defined here as a nucleotide segment of length divisible by 3 that is free of in-frame STOP codons (therefore not necessarily containing a START codon, e.g. Woodcroft et al. 2016, Sieber et al. 2018).

Infos : <https://gitlab.pasteur.fr/GIPhy/DNA2ORF>

Ecole de bioinformatique AVIESAN-IFB

21 to 26 November 2021

L'EBAI21 s'est déroulé du 21 au 26 novembre à la station biologique de Roscoff. Elle a réuni 37 élèves et 30 bioinformaticiens de toute la France.

Coordination scientifique et organisation : Rachel Legendre

debutR course

15/16/18/19 & 26 November 2021

Cours basique de R, organisé dans le cadre de la formation continue, destiné aux débutants complet en R

Infos : <https://vguillemot.github.io/debuter/>



DÉFI DU MOIS

Monthly challenge

Les fêtes de fin d'année approchent et les petites traditions avec elles. Traditionnellement à cette époque de l'année nous organisons le repas de Noël du Hub avec victuailles et petits cadeaux. Malheureusement cette année ce ne sera pas possible. Pour palier à ce manque cruel nous vous proposons malgré tout d'organiser deux événements festifs:

L'échange de cadeaux (aka. Secret Santa)

Pour les personnes intéressées merci de remplir rapidement ce petit formulaire

<https://framaforms.org/festivites-de-fin-dannee-1638267245>

Tirage au sort **Vendredi 10 Décembre.**

Cette année étant particulière les règles changent un peu histoire de faciliter l'organisation, vous connaîtrez la personne à qui vous offrez le cadeau ainsi que son adresse personnelle, merci de préserver la magie et de garder ces informations pour vous. L'idée est de mettre cela en place rapidement pour permettre un envoi de cadeau courant de la semaine prochaine.

La valeur du cadeau à offrir devra être comprise entre 5 et 10€. Dans l'idéal pensez à un cadeau utile et/ou sympathique

Un goûter virtuel

Pour les personnes intéressées merci de répondre via le formulaire <https://framaforms.org/festivites-de-fin-dannee-1638267245>

Le Goûter de Noël aura lieu le **Judi 16 Décembre à 16h** Afin de pallier aux soucis d'affichage de Teams de certains OS nous vous proposons de passer exceptionnellement par Gatherthown, un lien vous sera envoyé ultérieurement. Nous espérons que ce petit programme vous permettra de finir l'année sur une note agréable et de se voir "toustes ensemble" une dernière fois avant la nouvelle année qui on l'espère sera plus douce.

The end of the year is approaching and the some traditions come with it. Traditionally at this time of year we organize the Hub's Christmas dinner with food and small gifts.

Unfortunately this year again it will not be possible. To overcome this cruel lack, we are thought it could be nice to set two festive events:

Gift exchange (aka. Secret Santa)

For those interested, please quickly fill out this small form <https://framaforms.org/festivites-de-fin-dannee-1638267245>
Random drawing on **Friday December 10th.**

This year being special the rules are changing a bit to facilitate the organization, you will know the person to whom you are giving the gift as well as the person personal address, thank you for preserving the magic and keeping this information to yourself. The idea is to get this in place quickly so that we can send gifts before holidays. The value of the gift to offer must be between 5 and 10 €. Ideally, think of a useful and / or sympathetic gift.

Online goûter

For those interested, please quickly fill out this small form <https://framaforms.org/festivites-de-fin-dannee-1638267245>
The Online goûter will be on **Thursday December 16th at 4PM**

In order to alleviate the Teams display concerns we will use Gatherthown for this event, a link will be sent to you soon. We hope that this little program will allow you to end the year on a pleasant note and to see each other "all together" one last time before the new year which we hope will be better.



NOUVEAUX-ELLES MEMBRES

New members



Hippolyte KENNGNI

Ingénieur de recherche
CDD 1 an
IFB

CE MOIS CI

This month

13-17 Dec. 2021

Introduction à la Phylogénie Moléculaire : Concepts, méthodes et interprétation

Formation continue + RIIP + ENS

Equipe pédagogique : A. Criscuolo, C. Dauga, J. Guglielmini, F. Lemoine, G. Perrière, N. Taib, A. Zhukova

Infos : <https://research.pasteur.fr/en/b/nc3>

13-15 Dec. 2021

Journées MAMOV I

Infos :

<http://www.cmap.polytechnique.fr/~bansaye/JourneesMAMOVI>

15 Dec. 2021

Présentation des résultats de stage de Jérémy au CETEA

