RESUME

Dr. Delphine VINCENT

Data Science, Machine Learning, NLP & Bioinformatics

Melbourne area. Dual Citizenship: Australian and French.

Website | Google Scholar | ORCID | LinkedIn | GitHub | Tableau Public | email

SNAPSHOT

- Certifications in **Data Science Machine Learning, Data Analytics, Website Design** and **Chatbot development**
- **Honours** and **PhD** in Plants Science, 4 post-docs in biology, 10 years as a Senior Research Scientist, 3 years as a Data Scientist
- 27 years expertise in **biology, data analysis** and **bioinformatics** in academia, and government
- 3 years as an **Omics analytics, data science, machine learning, NLP and bioinformatics consultant t**
- 68 **publications** (37 peer-reviewed articles, 5 book chapters, 4 editorials, 18 conference papers, 2 patents, 1 preprint, 1 obituary) + 11 **datasets** in public repositories (MassIVE, Pride, (Giga)^{n}DB) + 35 GitHub **repositories** + 7 Tableau **vizzes**
- h-index 21, i10-index 29, cited 3,413 times
- Active **GitHub** profile with 35 repositories documenting reproducible analytical pipelines across omics, machine learning, NLP and data visualisation
- In-depth expertise in **experimental design, high-dimensional biological data analysis, statistical modelling, machine learning, natural language processing, multi-omics integration, visualisation, scientific reporting, manuscript preparation, and publication**
- Managerial and **project leading** experiences
- **Editor** and **reviewer** for multiple scientific journals

EMPLOYMENT HISTORY

Omics Analytics, Data Science, Machine Learning, NLP & Bioinformatics

- 2023-present: Independent Research Consultant, Melbourne area VIC 3000, Australia
  - Current projects include multi-omics biomarker discovery, metabolomics modelling, and scientific machine learning for university-based collaborators
  - Delivered end-to-end omics and machine learning analyses for academic and industry collaborators across proteomics, metabolomics, transcriptomics, and NLP-driven biological datasets
  - Developed reproducible Python and R workflows for statistical modelling, clustering, predictive analytics, and multi-omics interpretation
  - Designed and deployed scientific data communication tools including GitHub repositories, dashboards, and research websites

Senior Research Scientist

- 2013-2023: Agriculture Victoria Research (AVR), Victorian Government, AgriBio Bundoora VIC 2083, Australia

Post-doctorates

- 2009-2011: Australian National University (ANU), Canberra ACT 0200, Australia
- 2007-2009: UMR BIOGECO INRA 69 route d'Arcachon, 33612 Cestas, France
- 2005-2006: UMR Genetique Vegetale INRA, Ferme du Moulon, 91190 Gif-sur-Yvette, France
- 2003-2005: University of Nevada, Reno (UNR), 1664 North Virginia Street, Reno NV 89557 USA

EDUCATIONAL QUALIFICATION

PhD: 1999-2003: "Proteomic analysis of response to water stress in maize based on cell differentiation". PhD doctorate granted with Highest Distinctions from Universite Paris-Saclay, France.

DATA SCIENCE & WEB DESIGN SKILLS

Computer OS

Fully skilled with MS Windows software (Word, Excel, Powerpoint, Sharepoint, Teams, OneNote, Outlook, Power BI etc…).

Programming (code languages / IDE)

R/RStudio (tidyverse, readr, tidyr, dplyr, ggplot2, Bioconductor), **Python/JupyterLab** (pandas, NumPy, scikit-learn, SciPy, statsmodels.api, MiniSom, TensorFlow, UpSet, wordcloud, spaCY, Hugging Face Trasnformers, squarify, RegEx, NLTK, UMAP, gradio, iPython, matplotlib, seaborn, Regex, Pickle, Gensim, pyTorch, XGBoost, LightGBM, Keras, plotpy, BeautifulSoup, Selenium, tdqm, skimage, pillow, imageio, librosa, OpenCV, mido,Pydub, gTTS, MoviePy, Subprocess, FastAPI, Pydantic, RapidFuzz, JSON), Bash Shell, Git, **GitHub**, Structured Query Language (SQL), **Hypertext Markup Language** (HTML), javascript (JS), Hypertext PreProcessor (PHP), **Cascading Style Sheet** (CSS), Visual Studio (VS) Code, **Anaconda navigator**.

Data Science, Machine Learning (ML) and Natural Language Processing (NLP)

Data wrangling, feature engineering, descriptive (means, SDs, quartiles, medians, box plots, histograms, etc..), **univariate** (linear models, ANOVA, and regression), **bivariate** (correlation) and **multivariate** (principal component analysis (PCA), partial least squares (PLS), linear discriminant analysis (LDA), k-means clustering, self-organising maps, hierarchical clustering, and supervised machine

learning) analyses; ML algorithms (Random Forest, Ensemble, Support Vector Machine, Stacked Model, Multi-Layer Perceptron, Deep Learning, Neural Networks), NLP algorithms (tokenisation, stopword removal, stemming, lemmatisation, BoW, TF-IDF, word embedding, POS tagging, topic modelling).

Bioinformatics and Data Mining

Multi-omics **data mining**, **functional annotation**, **pathway analysis** and **biological interpretation** using BLAST+, Jexpress, Cluster-Treeview, Blast2GO, Goblet, SignalP, TargetP, SecretomeP, ChloroP,NetCGly, GPI-SOM, Goanna PowerBI, PeptideAtlas, ProteinAtlas, cytoscape, UniprotKB, Gene Ontology (GO), STRING, SwissModel, Apollo Jbrowser, Proteowizard, TransProteomics Pipeline, Pretzel, AgriGO, REVIGO, ShinyGO, Reactome, PlantReactome, BioCyc, Pathway Tools, KEGG, AraCyc, BreadwheatCyc, Circos, MetaboAnalyst, DeconTools, EMBL, EBI InterPro, PDBeKB, RCSB PDB Archive, AlphaFold, ColabFold, Cn3D, Bio edit, Mega XI, Expasy, Gene Ontology, OpenAI, PaintOmics, MetaboAnalyst, NCBI Analyse, Mercator, Galaxy platforms, JBrowser, Integrated Genome Browser.

Data Visualisation

Circos plots, radar plots, violin plots, word clouds, treemaps, treemap bars, histograms, line plots, bar plots, pie charts, scatterplots, donut charts, heat maps, box plots, volcano plots, Gantt chart, LM plots, density plots, join plots, pair plots, pair grid plots, heatmaps, etc.. using **PowerBI**, **Tableau**, **R** and **Python**.

Data interpretation and Storytelling

Attested by my published articles in which I summarise the process in a simplified schematic diagram, present all results with appropriate charts to facilitate interpretation, as well as by my Jupyter Notebooks and Tableau / PowerBI dashboards.

Web design and front-end development

Designed and deployed scientific web resources to support research visibility and data communication with secure email handling, optimised media performance, and a WebCentral-compliant deployment pipeline for Data Biome.

PUBLICATIONS AND SCRIPTS

Full article list here and on Google Scholar.
Active GitHub repositories containing Python, R, SQL, and web-development code for reproducible workflows.
Tableau dashboards on Public Tableau.
HTML and CSS code on my professional website.

DATASETS ON PUBLIC REPOSITORIES

Mass Spectrometry Interactive Virtual Environment (MassIVE): 2016-2022: MSV000090572 DOI:10.25345/C53N20J8S, MSV000088253 DOI:10.25345/C5585Q, MSV000085379 DOI:10.25345/C5QQ6J, MSV000084216 DOI:10.25345/C5C95X, MSV000083970 DOI:10.25345/C56S7Z, MSV000083191 DOI:10.25345/C5HG7N, MSV000082070, MSV000081618, MSV000080036

Protein identification database (PRIDE): 2016: PXD002529

(Giga)^{n} DB: 2023: Supporting data for Vincent et al, 2023

Apollo Jbrowse: 2024: Mapping of bread wheat genome, transcriptome and proteome (proteogenomics)

CERTIFICATIONS & PROFESSIONAL DEVELOPMENT

CODECADEMY ONLINE COURSE

- 2024: Career Path: Data Scientist – Analytics Specialist (SQL, Python 3, Tableau, Excel, GitHub). Certificate
- 2024: Skill Path: Build a website with HTML, CSS and GitHub pages. Certificate
- 2025: Career Path: Data Scientist – Machine Learning Specialist (Python 3). Certificate
- 2025: Courses: Introduction to Prompt engineering + Introduction to Large Language Model (LLM). Certificate
- 2025: Skill Path: Build chatbots with Python. Certificate

TEACHING AND STAFF MANAGEMENT

- 2013-2023: Manager of Science A staff at AVR. Mentoring and supervising students.
- 2013-2020: Demonstrator and lecturer in the Systems Biology Workshop held yearly at AgriBio.
- 2010: Demonstrator in sessions of the Advances in Molecular Plant Science course for undergraduate students at ANU.
- 2007-2008: Guest lecturer in the Biochemistry Master at Bordeaux University (France).

OCCUPATIONAL HEALTH AND SAFETY (OHS), FIRST AID, WELLBEING, MENTORING

- 2013-2016: OHS Representative (HSR), designated first aider (DFA), wellbeing champion for AVR.
- 2022-2023: mentor for the national Young Indigenous Women's STEM Academy (YIWSA) program.

AUSTRALIAN AND INTERNATIONAL COMMITTEES

- 2019-2023: European Cooperation in Science and Technology Action "Protein Production in Photosynthetic Organisms (R3PO)"
- 2021-2023: Proteomics and Metabolomics Victoria (PMV) committee member

EDITORIAL BOARDS

- 2021-2023: BMC Plant Biology editor
- 2015-2023: Editor of the following Research Topics/Special Issues:
  How can secretomics help unravel the secrets of plant-microbe interactions?
  Secretomics: More Secrets to Unravel on Plant-Fungus Interactions. Vol 1
  Secretomics: More Secrets to Unravel on Plant-Fungus Interactions. Vol 2
  Proteomics: Technologies and Their Applications
  Plant Adaptation to Their Biotic and Abiotic Environment Through the Lens of Secretomics
  State-of-the-Art Molecular Plant Sciences in Australia
  Sowing the seed to ensure the future of plant proteomics: Commemorative Issue in Honor of Dr. Dominique Job (1947-2022)