

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

- **Honours** and **PhD** in Plant 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**
- **Honorary Senior Fellow**, Faculty of Science, University of Melbourne
- Certifications in **Data Science, Machine Learning, Data Analytics, Website Design** and **Chatbot development**
- Currently expanding cloud-computing capability through AWS Educate training, including **cloud fundamentals, storage, compute, and networking**.
- 69 **publications** (37 peer-reviewed articles, 5 book chapters, 4 editorials, 18 conference papers, 2 patents, 2 preprints, 1 obituary)
- 11 **datasets** in public repositories (MassIVE, Pride, (Giga)DB) + 36 GitHub **repositories** + 7 Tableau **vizzes**
- h-index 22, i10-index 29, cited 3,492 times
- Active **GitHub** profile documenting reproducible analytical pipelines across omics, machine learning, NLP and data visualisation, SQL and web designs
- Expertise in **experimental design, high-dimensional biological data analysis, statistical modelling, ML/NLP, multi-omics integration, visualisation, scientific reporting and publication**
- **Management** and **project leadership** 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 proteogenomics for gene model validation, 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

Honorary Senior Fellow

- 2026-2028: [Faculty of Science](#), University of Melbourne, Parkville VIC, Australia
 - Contributing scientific expertise in proteomics, bioinformatics, omics analytics, and research translation through university-based collaborations
 - Supporting interdisciplinary research activities involving data analysis, scientific interpretation, publication development, grant application and student/research mentoring

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](#).

TECHNICAL SKILLS & DIGITAL CAPABILITIES

Programming, Analytics & Workflow Development

Advanced experience developing reproducible analytical workflows in Python/JupyterLab, R/RStudio, SQL, Bash, Git/GitHub and VS Code. Python experience includes data wrangling, statistical modelling, machine learning, NLP, automation, web scraping, API-oriented workflows, image/media processing and scientific computing using libraries such as pandas, NumPy, SciPy, scikit-learn, statsmodels, TensorFlow, PyTorch, XGBoost, LightGBM, Keras, matplotlib, seaborn, BeautifulSoup, Selenium, FastAPI, Pydantic and Hugging Face Transformers. R experience includes tidyverse, ggplot2, Bioconductor and statistical/data-visualisation workflows.

Data Science, Machine Learning & Natural Language Processing

Experienced in data cleaning, feature engineering, exploratory data analysis, statistical modelling, classification, clustering, regression, dimensionality reduction and predictive modelling. Methods include ANOVA, linear models, correlation analysis, PCA, PLS, LDA, k-means, hierarchical clustering, self-organising maps, Random Forest, Support Vector Machines, ensemble models, neural networks and deep learning. NLP experience includes tokenisation, text preprocessing, TF-IDF, word embeddings, topic modelling, semantic analysis, transformer-based workflows and AI-assisted analytical pipelines.

Generative AI, Agentic AI & Confidential LLM Workflows

Practical experience using generative AI and agentic AI tools to support coding, debugging, data analysis, research synthesis, technical writing, documentation, website development and workflow optimisation. Tools used include ChatGPT, Gemini, GitHub Copilot and LM Studio. Experience includes prompt engineering, AI-assisted code development, validation of AI-generated outputs, iterative refinement of analytical workflows, and use of local LLM environments for confidential projects where data privacy and offline processing are required.

Bioinformatics, Omics Analytics & Scientific Data Mining

Specialist expertise in proteomics, metabolomics, transcriptomics and multi-omics data analysis, with strong experience in functional annotation, pathway analysis, genome/protein mapping, biological interpretation and scientific reporting. Tools and resources include BLAST+, Cytoscape, UniProtKB, Gene Ontology, STRING, KEGG, Reactome, PlantReactome, BioCyc, BreadWheatCyc, InterPro, AlphaFold, ColabFold, Galaxy, Apollo/JBrowse, Integrated Genome Browser, ProteoWizard, Trans-Proteomic Pipeline, PeptideAtlas, ProteinAtlas, MetaboAnalyst, REVIGO and ShinyGO.

Data Visualisation, Reporting & Research Communication

Experienced in transforming complex analytical outputs into clear, interpretable visual narratives for scientific, technical and non-technical audiences. Visualisation experience includes dashboards, schematic workflows, publication-ready figures, exploratory plots and interactive reporting using Python, R, Power BI and Tableau. Plot types include heatmaps, volcano plots, box plots, violin plots, scatterplots, bar charts, circular/Circos plots, radar plots, treemaps, word clouds, Gantt charts and multi-panel scientific figures. Strong record of data storytelling through peer-reviewed publications, Jupyter notebooks, dashboards and research websites.

Web Design, Front-End Development & Hosting Administration

Designed, deployed and maintained scientific websites and research communication resources using HTML, CSS, JavaScript, PHP, GitHub Pages, cPanel and WebCentral hosting workflows. Experience includes front-end development, content structuring, media optimisation, secure email handling, hosting administration, domain/hosting troubleshooting, website maintenance and practical deployment support for [Data Biome](#) and other scientific communication projects.

Cloud, Collaboration & Workplace Technologies

Fully proficient in Microsoft Windows and Microsoft 365 workplace technologies, including Word, Excel, PowerPoint, Outlook, Teams, SharePoint, OneNote and Power BI. Currently expanding cloud-computing capability through AWS Educate training, with completed badges in all the “Getting Started” series and becoming familiar with AWS products such as S3, EC2, VPC, RDS, IAM, and Lambda.

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

AWS Educate / Amazon Web Services

- 2026: AWS Educate / “Getting Started” series training badges: Introduction to the AWS Management Console, Introduction to Cloud Computing 101, Getting Started with Storage, Getting Started with Compute, Getting Started with Networking, Getting Started with Databases, Getting Started with Security, Getting Started with Cloud Ops, and Getting Started with Serverless.
- Continuing structured self-directed training through AWS Educate, completing the available courses sequentially across the Getting Started, AI and Machine Learning, and Core Concepts series.
- Using Amazon SageMaker Studio Lab and an AWS Free Tier account as hands-on practice environments to consolidate AWS cloud, AI/ML, compute, storage, and deployment concepts.

CodeCademy courses

- 2024: Career Path: Data Scientist – Analytics Specialist. [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: Introduction to Prompt engineering + Introduction to Large Language Model (LLM). [Certificate](#)
- 2025: Skill Path: Build chatbots with Python. [Certificate](#)
- 2026: Certification Path: AIF-C01: AWS Certified AI Practitioner (in progress)

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.

Scientific committees

- 2019-2023: EU COST Action “Protein Production in Photosynthetic Organisms”
- 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 \(1947-2022\)](#)