

---

# BINF\*6210 Software Tools

Fall 2025 - Karl Cottenie

---



COTTENIELAB.org

---

---

# Introductions

---

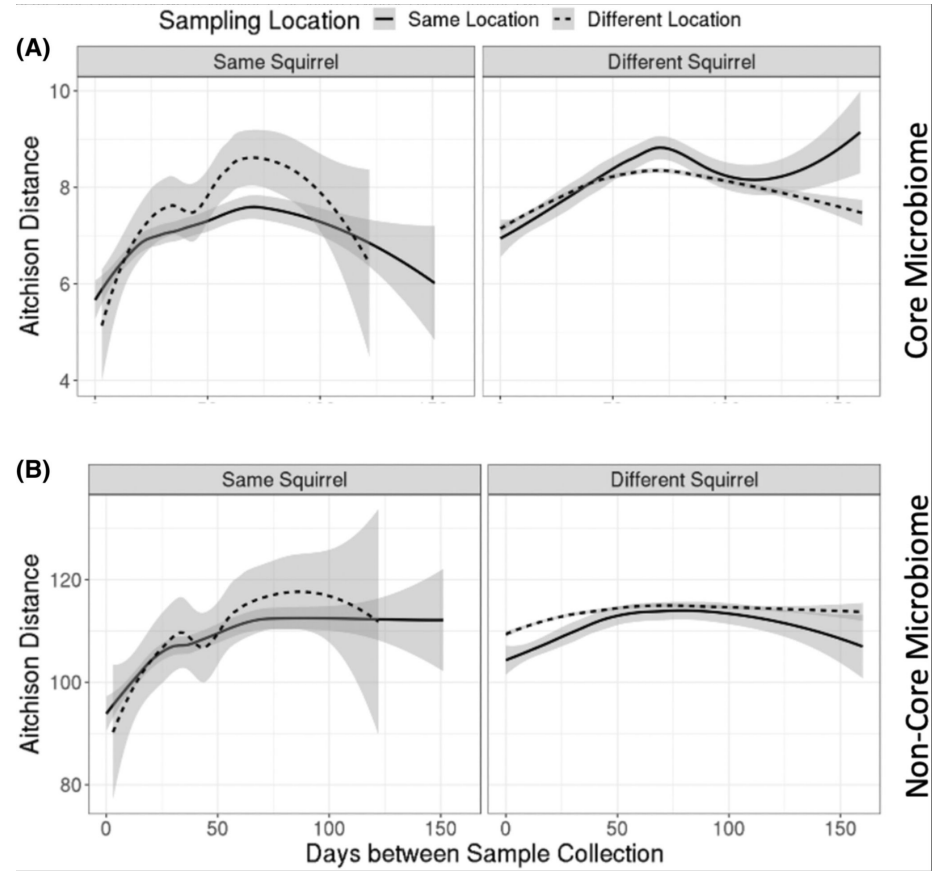
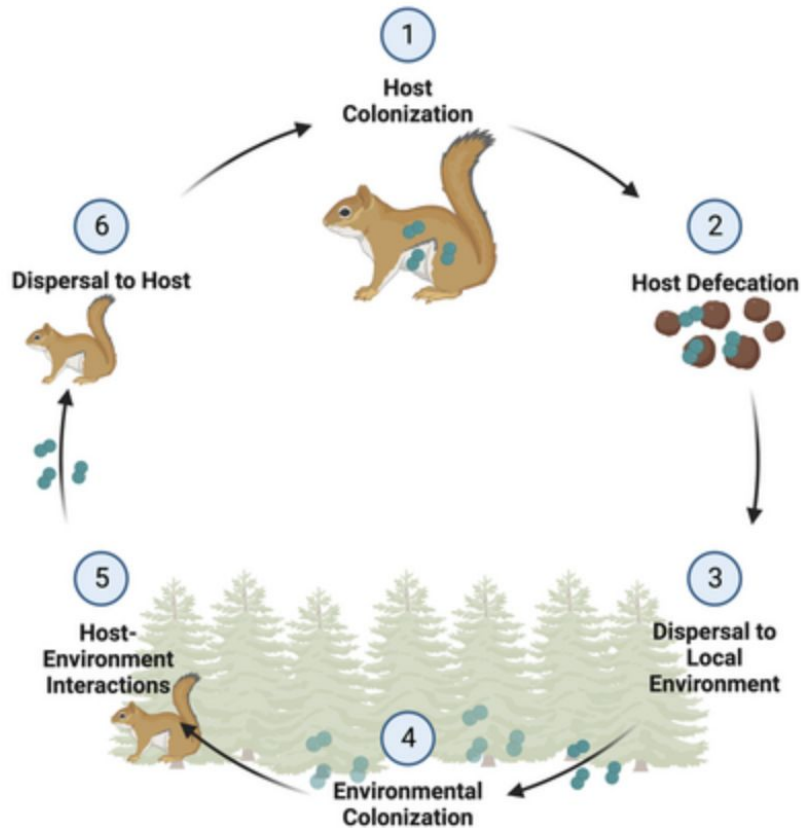
---

# Karl Cottenie

---

- [Google Scholar profile](#)
- Contact me
  - For course-related questions: discussion board on CourseLink
  - For personal questions: via email - [cottenie@uoguelph.ca](mailto:cottenie@uoguelph.ca)
- Computational ecology
- I love R
- Most of my publications: multivariate statistics in R
- Last 10 years
  - Barcoding
  - Microbiome studies
- Fun fact: [I love salsa dancing](#)

# Red squirrel microbiome



# TA: next week

---



# Thank you: Dr. Sally Adamowicz

---

- Previous course instructor
- Provided me with all her material
- I could not have done this without her hard work



# You as a person

---

- In break-out group;
- Show your partner(s) the last photo on your phone (that you are comfortable sharing - of course) and why you took it;
- When back in class, describe your partner's photo.

# Peoples and places

---

With great respect, we acknowledge that the University of Guelph resides on the ancestral lands of the Attawandaron people and the treaty lands and territory of the Mississaugas of the Credit. We uphold the significance of the Dish with One Spoon Covenant and the continuing relationship our Indigenous neighbours have with this land. We recognize that today this gathering place is home to many First Nations, Métis and Inuit peoples and acknowledging them reminds us of our relationships to this land where we learn and work.



# Equity, Diversity and Inclusion (EDI) Statement



Our classroom is a space where everyone is valued, respected, and supported. You are welcome here regardless of your colour, race, background, identity, or experiences.

This semester, we encourage you to ask questions and engage with one another in an open and respectful way. If at any point you face barriers, feel excluded, or need support, please remember that you are not alone and we are here to listen, to learn, and to help.

Together, let's create a community where everyone feels they belong and can thrive. We are excited to learn and grow with you, and we can't wait to see what we accomplish together this year.

# I want to get to know you

---

- Active participation for today
  - Picture
  - How you want me to call you in class

---

---

# Learning outcomes

---

---

# Learning outcomes for today

---

By the end of this class, you should:

- Be able to answer the question: Do I want to take this class?
- Understand the main learning outcomes for the course

# Learning outcomes for the course

---

By the end of the course, you should be able to:

1. Obtain data from key databases relevant for bioinformatics and to understand the sources and limitations of these data
  - Database systems (data + database management system)
  - Genetic/genomic data (e.g., [NCBI](#), [BOLD](#), [Silva](#))
  - Gene functional annotation data (e.g., [GO](#), [KEGG](#))
  - Other biological data (e.g., phenotype, [PanTHERIA](#), [FishBase](#))
  - Environmental data (e.g., [WorldClim](#))
  - Some web tools, but emphasis on scraping data from within your script

# Learning outcomes for the course

---

By the end of the course, you should be able to:

2. Filter, manipulate, analyze, and visualize bioinformatics data (with non-exclusive emphasis on the R programming language and software resources available through Bioconductor)
  - Reformatting, quality control, filtering, subsetting
  - Fundamental data science skills using base R and tidyverse suite of packages
  - Analyses (e.g., for population genetics, functional analysis, biome research, phylogenetics)
  - Visualization
  - Emphasis on [Bioconductor tools](#)

# You as a person

---

- Arrange yourself according to your R background

# Learning outcomes for the course

---

By the end of the course, you should be able to:

3. Conduct reproducible analyses and use software tools for version control and collaboration
  - Organization of files
  - Commenting of code
  - Formatting of code
  - Version control
  - Collaborative coding
  - [Github](#)
  - Benchmarking and coding for computational efficiency
  - Use of IDE (Integrated Development Environment, e.g., [RStudio](#))
  - **LLM?**



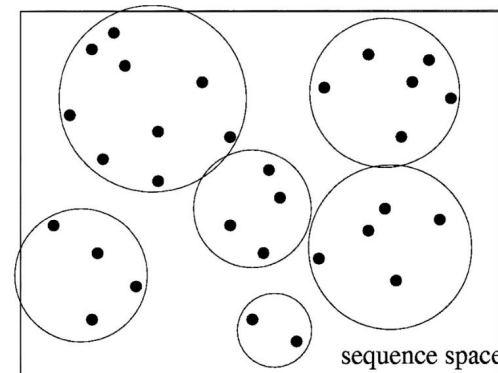
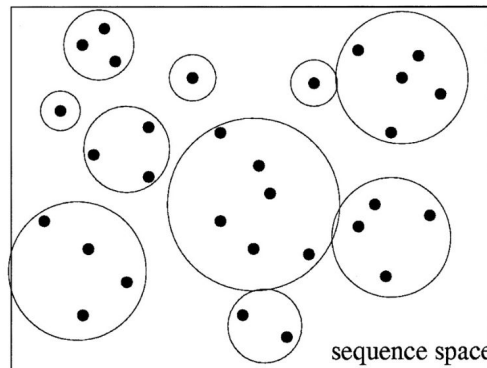
# Learning outcomes for the course

---

By the end of the course, you should be able to:

## 4. Understand and apply selected algorithms and metrics commonly used in bioinformatics

- Sequence alignment
- Clustering
- Phylogeny
- Fst
- Genetic diversity
- Phylogenetic diversity



# Learning outcomes for the course

---

By the end of the course, you should be able to:

5. Adapt the above skills to learn new tools and conduct new analyses not explicitly covered in class
  - How to use help resources
  - How to use software documentation
  - Online resources
  - Data analysis cycle
  - Problem-solving skills
  - Conduct your own projects with both technical proficiency and creativity
  - (generative AI tools?)

---

---

Fit into the MBINF program?

---

---

# 3 Semesters

---

- Fall
  - Topics (BINF\*6890) - variety of concepts, reading papers, writing and presenting skills
  - Software Tools (BINF\*6210) - R, databases, biodiversity, foundational algorithms
  - Programming (BINF\*6410) - Unix operating system, Bash, Python programming (elective, but highly recommended for most students)
- Winter
  - Statistical Bioinformatics (BINF\*6970) - data analysis for high-dimensional data
  - Genomic Methods (BINF\*6110) - large-scale genomics and high-performance computing
  - Elective course
- Summer
  - Research Project (BINF\*6999)

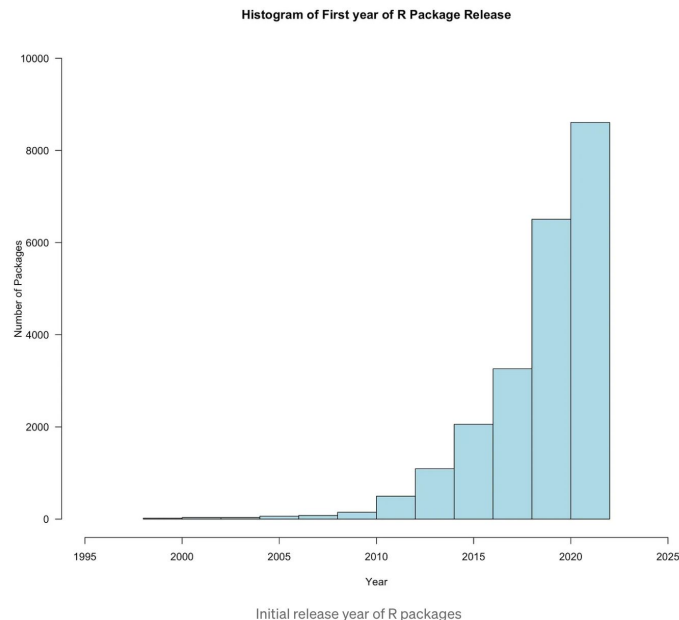
# Why these languages?

---

- Recent literature
- Consulted recent alumni
- Online sources about trends
- Practitioners and job postings
- Considered availability of software tools
- Learning curve and timeline
  - You will be able to do lots of things quickly!
  - You will need to keep learning over time!

# Why these languages?

- Unix/Linux
  - Often used for computational heavy lifting: genome assembly, mapping reads to reference genome
- Python/R
  - High-level language
  - flexible for data manipulation, especially R
- Python
  - Libraries for machine learning
  - Numerous contributed tools
- R
  - Unparalleled array of user-contributed packages
  - Excels for statistics and visualization



# Some benefits of R

---

- Huge global community of authors/developers
- Open-source language
- Origins in statistics
- Excels in data wrangling, visualization, as well as statistical computing
- Many resources for bioinformatics
- High-level language
  - Hides complexity, increases efficiency for the programmer
  - Functions are human readable
  - [Core R is written in C, Fortran, and R](#)

---

---

# Organization of the course

---

---



# Course organization

---

- Required: read the course outline after this class (important information on due dates, academic integrity, etc.)
- Available through CourseLink
- Use the Discussion board through CourseLink to ask questions about logistics and assignments
- Class structure
  - Some background information in advance of class
  - Class time: primarily active learning components
  - Be sure to go through example scripts
  - Required and optional readings

# Coding strategy

---

- Tailored to students with a life sciences background
- From less to more complex bioinformatics tasks
- Learn programming while doing biological analyses
- A series of mini-modules, consisting of data and analytical tools
- Biological examples and concepts
  - DNA barcoding
  - Biodiversity and molecular ecology
  - Phylogenetics
  - Microbiome
  - Differential gene expression and functional analysis

# You as a person

---

- Arrange yourself according to your background/interest in bioinformatics

# Let's get started :-)

---

- Getting R installed and running
- A few basics about R
- Arithmetic in R

# Tips about homework

---

- Get started on your homework right away
- 3 courses/semester is considered a full-time course load at the graduate level
- Working on the material outside of class time is critical to your progression, you can only learn bioinformatics/coding/R by doing it yourself
- 10-12 hours per week per 0.5 credit course (or 2-3 hours of time outside class for every hour of class time)
- You may work in groups to increase your understanding

---

---

# Learning outcomes - revisited

---

---

# Learning outcomes for today

---

By now, you should:

- Be able to answer the question: Do I want to take this class?
- Understand the main learning outcomes for the course