GDC Genetic Diversity Centre Zurich

Genetic Diversity: Analysis

16 June - 27 June 2025

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

GDA ▷ Welcome

Genetic Diversity Centre Zurich

Aria Silvia Tamara Jean-Claude Nik

At the Genetic Diversity Centre (**GDC**), we like our DNA diverse and our organisms anything but standard. As a knowledge and technology platform within the USYS department at ETH Zurich, we support research into the genetic and genomic diversity of all kinds of life — especially the weird, wild, and wonderful non-model organisms that don't usually get the spotlight. Whether you're tracking trout genes or decoding desert microbes, we're here with the tools and expertise to help.

Genetic Diversity Centre Zurich

"I cannot teach anybody anything, I can only make them think."

—Socrates

"Science is a way of thinking much more than it is a body of knowledge."

—Carl Sagan

Genetic Diversity Centre Zurich

What is your thinking mode?

Source: A framework popularised by behavioural scientist Jonathan Haidt, which describes how people often switch between different 'modes of thinking' - especially when they are debating, persuading or making decisions.

Humans are surprisingly flexible thinkers - but not always in the way you'd expect. Depending on the situation, we often slip into different mental 'modes' that shape the way we argue, reason and react:

Scientist mode - Curious, evidence-driven and open to changing our minds. This is when we genuinely seek the truth, test hypotheses and follow the data - even when it challenges our beliefs. Unfortunately, this mode is rarer than we'd like to think.

Preacher mode - We are defending sacred beliefs or values and want to inspire others to share them. In this mode, we don't ask questions - we give answers.

Prosecutor Mode - We focus on finding flaws in others' arguments. We don't test our own ideas - we attack someone else's, often to win a debate or score points.

Politician Mode - We're more concerned with how things look than what's true. We shape our words to appeal to our audience, seek approval and build coalitions - not necessarily to discover facts.

While each mode has its place, it's the scientist mode that's most valuable for honest inquiry - and often the hardest to maintain. The trick is to know which hat you're wearing... and when to take it off.

GDC Genetic Diversity Centre Zurich

Power-Disa-Point-Ability

Power-Disa-Point-Ability: When a presenter confuses “more slides” with “more insight” and hides behind animations, charts, and bullet points like a kid behind the sofa during a horror movie.

Genetic Diversity Centre Zurich

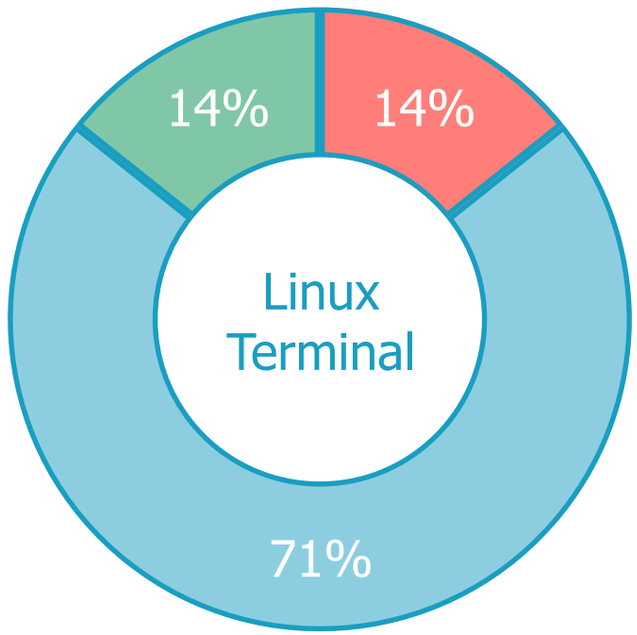
Request #1: Survey

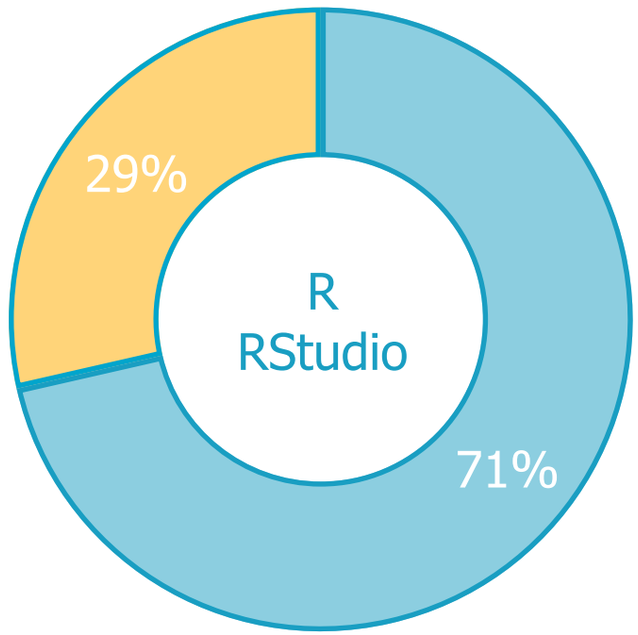
Request #2: R-Code

Request #3: Requirements

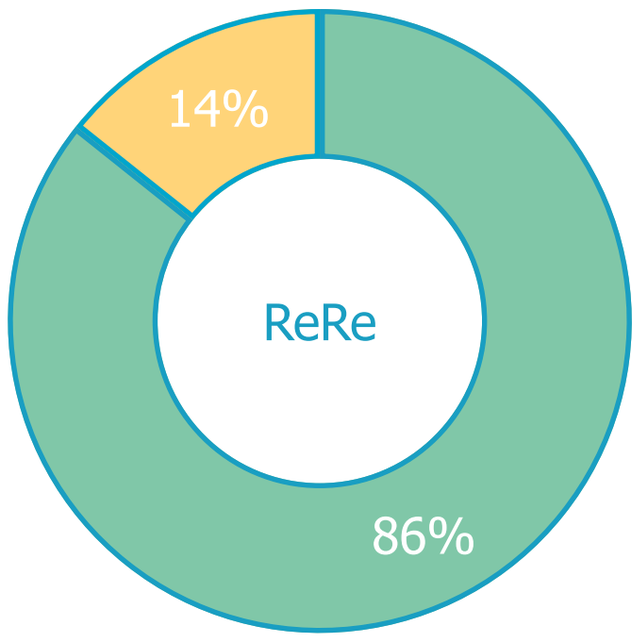
Genetic Diversity Centre Zurich

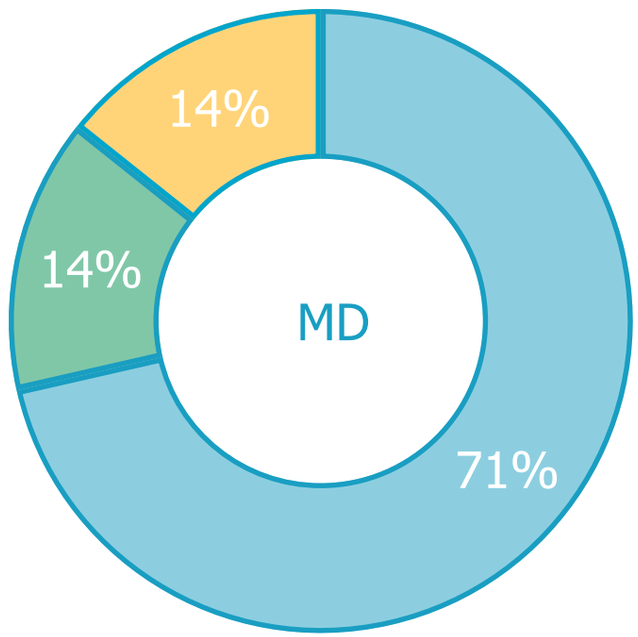
Request #1: Survey





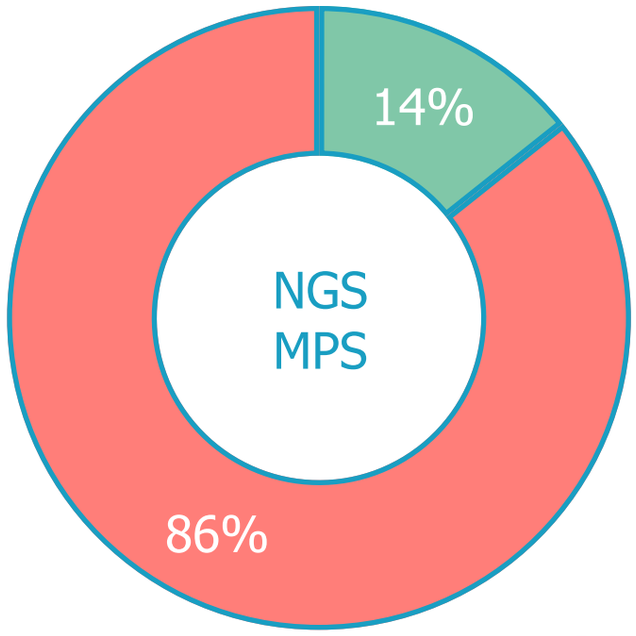
N=7

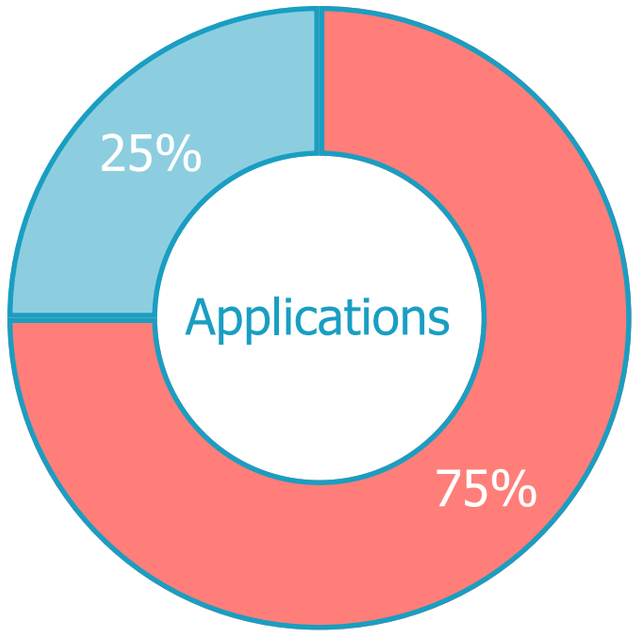




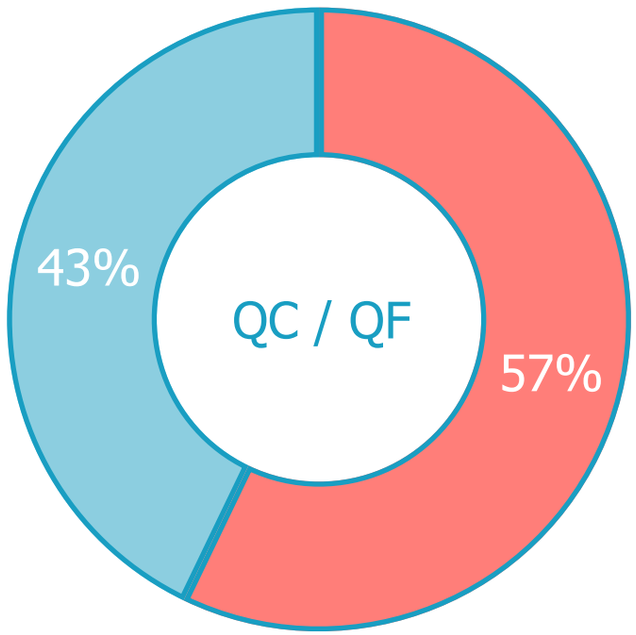
GDC Genetic Diversity Centre Zurich

Request #1: Survey



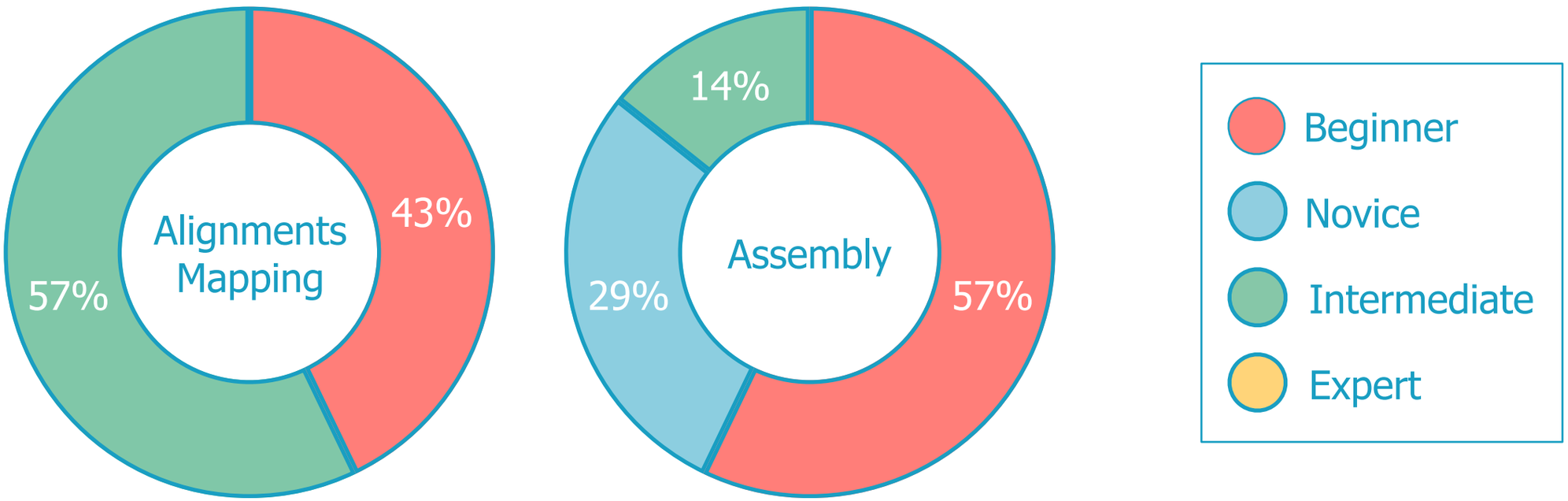


N=7

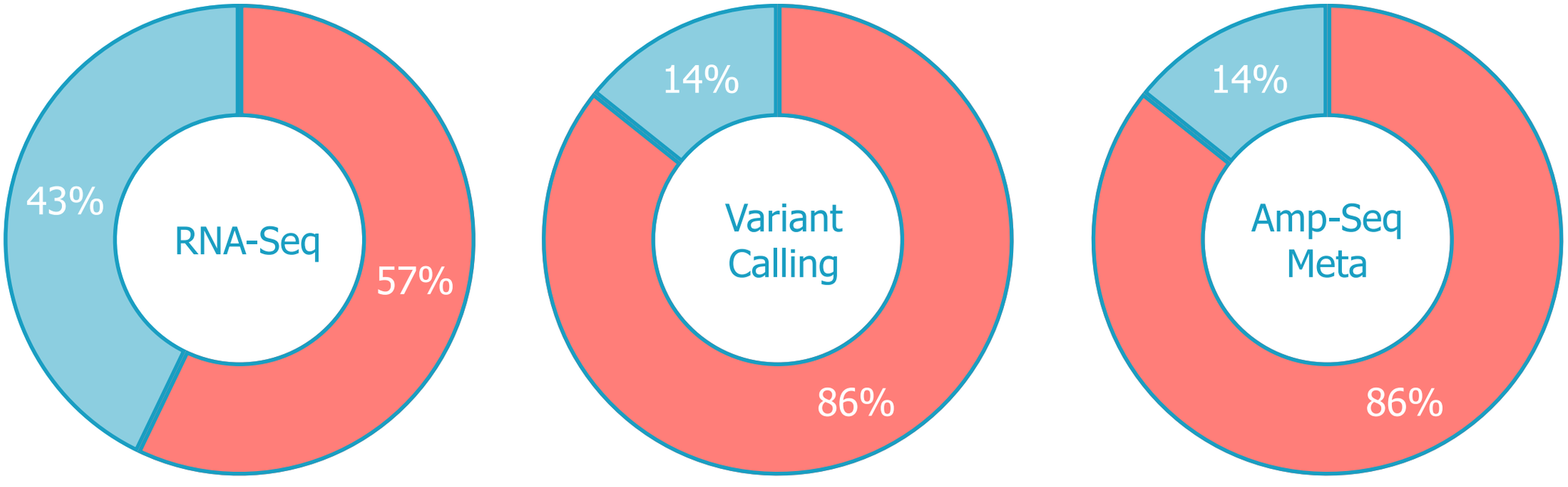


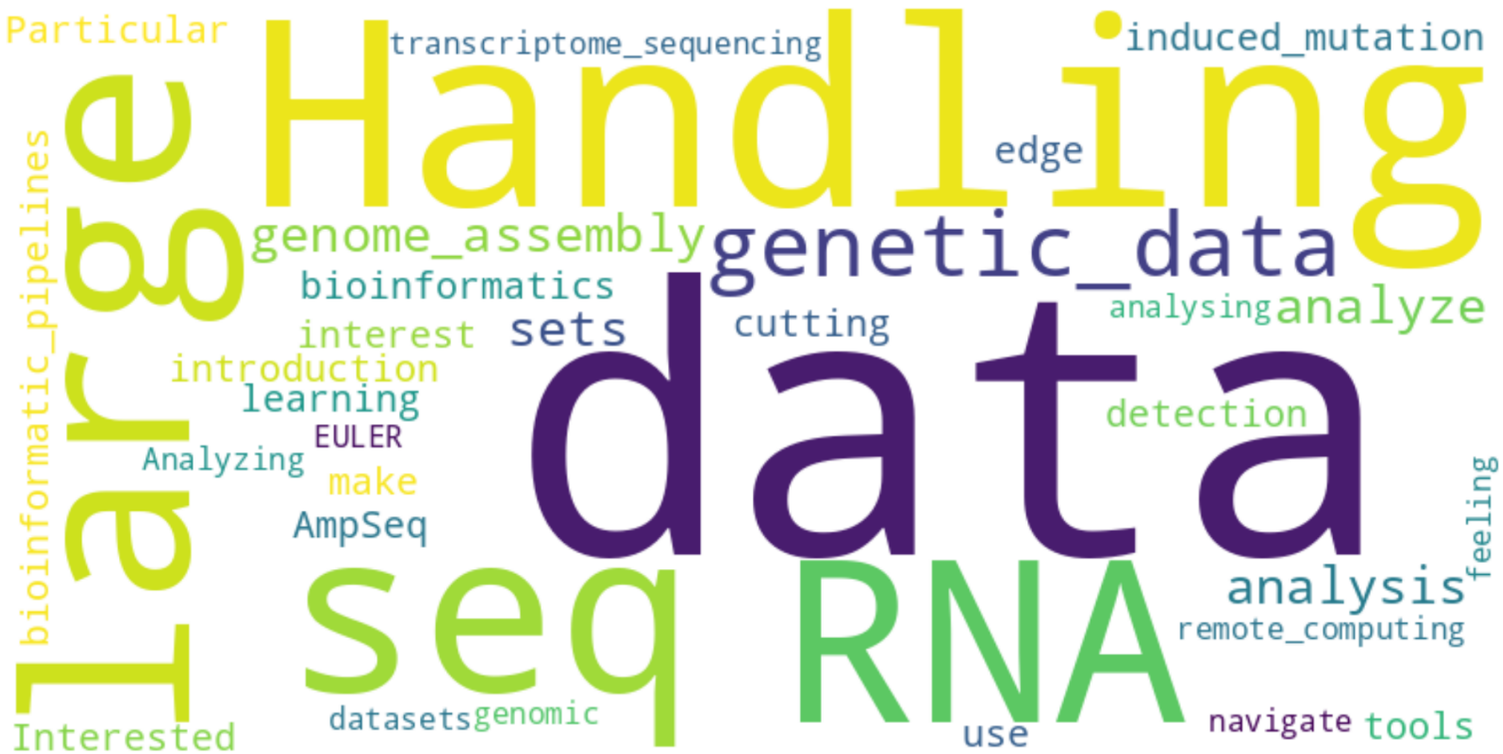
Genetic Diversity Centre Zurich

Request #1: Survey



N=7





Summary of Students' learning interests: The most prominent themes include **RNA-seq analysis**, **genome assembly**, **handling large genetic datasets**, and **bioinformatic pipelines - ** all centred on practical, hands-on genomics and computational biology skills. Let me know if you'd like this image exported or used in a slide.

Request #2: R-Code

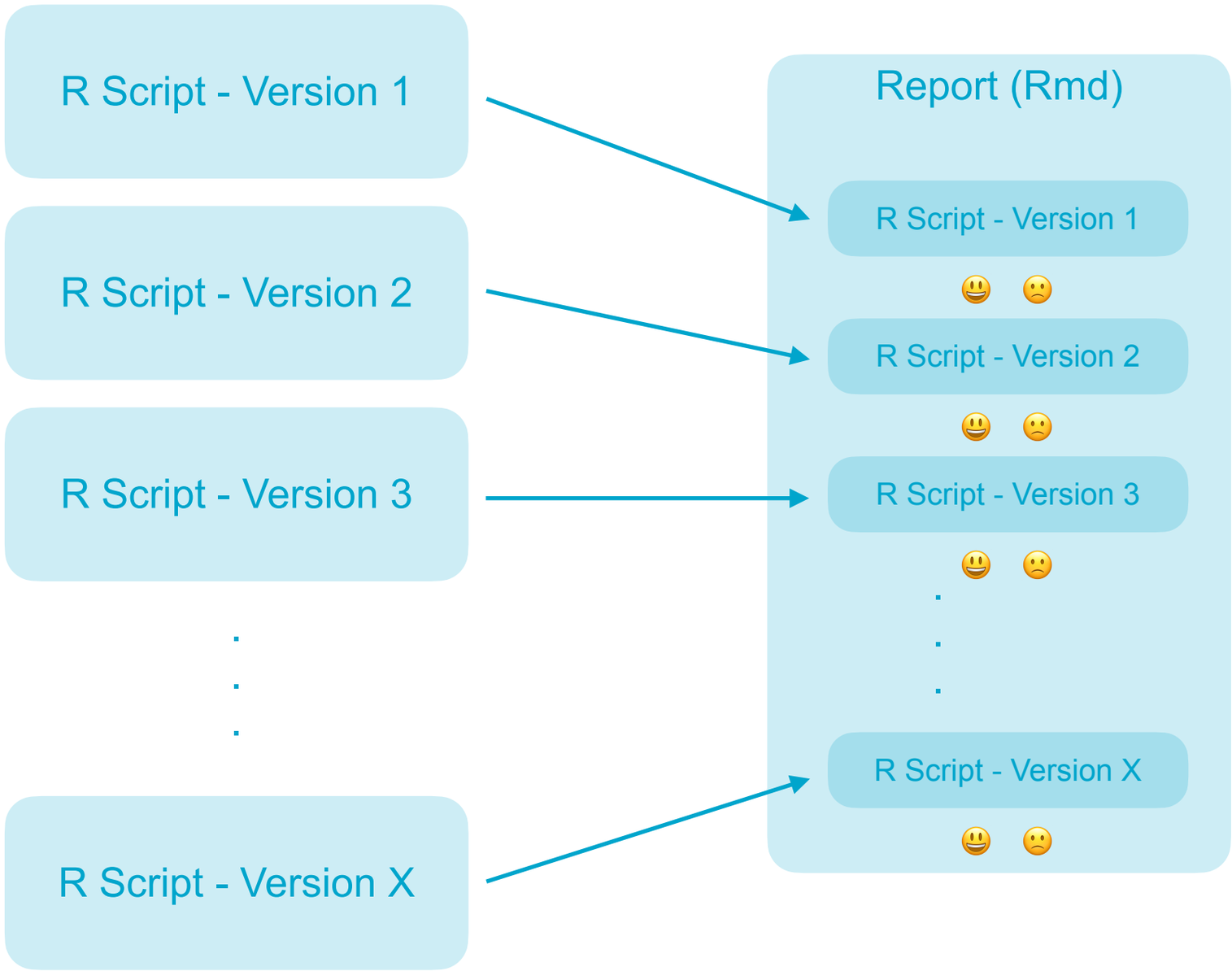
Please complete a simple task in R: Write a script that creates a text file with the results of **10 rolls** of a **pair of dice**.

When reviewing the R scripts, we noticed a few areas that could be improved:

- **Script header**: Does the script include a clear and informative header with the author, date, and purpose?
- **Coding style**: Are formatting, indentation, and naming conventions consistent and tidy?
- **Consistency**: Are similar tasks handled in a consistent way throughout the script?
- **Typos**: Are variable names and function calls free of spelling mistakes?
- **Functionality**: Does the script run without errors and produce the expected results?
- **Reproducibility**: Can someone else run the script and achieve the same results?
- **Readability**: Are variable names meaningful, and are helpful comments included?
- **Simplicity**: Is the code as simple as possible, without unnecessary complexity?
- **Correctness**: Are the tasks implemented correctly and aligned with the assignment's goals?

```
## Turner Helix
## Rolling Dices (D6)
## Version: 250517

# Function to simulate rolling a pair of dice
roll_dice <- function() {
  return(sample(1:6, 2, replace = TRUE) %>% sum())
}

# Simulate rolling a pair of dice 10 times
results <- sapply(1:10, function(x) roll_dice())

# Create a text file and write the results
file_path <- "10_dice_rolls.txt"
writeLines(c("The results of 10 rolls of a pair of dice
are:", paste(results, collapse = ", ")), file_path)
```

```
## Turner Helix
## Rolling Dices (D6)
## Version: 250519

write.table(TeachingDemos::dice(10,2), "dice2x10.txt")
```




```{r First Attempt, file = "dice_v1.r"}
```

```{r V2 Faster, file = "dice_v2.r"}
```

```{r V3 Flexible, file = "dice_v3.r"}
```

```{r VX Final, file = "dice_vx.r"}
```

GDA: Rolling Dice Example

Jean-Claude Walser

22/06/23

Intro

Request #2 - An important part of this course is to apply the knowledge you have learned. For this reason, we would like you to solve the following assignment using R:

“Write an R script (R code) that creates a text file containing the results of 10 rolls of a pair of dice.”

The R script should be easy to use (user friendly) and flexible for extensions (e.g. more rolls, more dices, or flexible output). Think about ways to implement reproducibility, if possible? We don’t need the perfect solution, partial solutions are fine. If the task is beyond your R skills, a step-by-step description of a solution would also be fine. Please send your solution (zipped) via e-mail directly to me using subject: GDA23: Request #2.

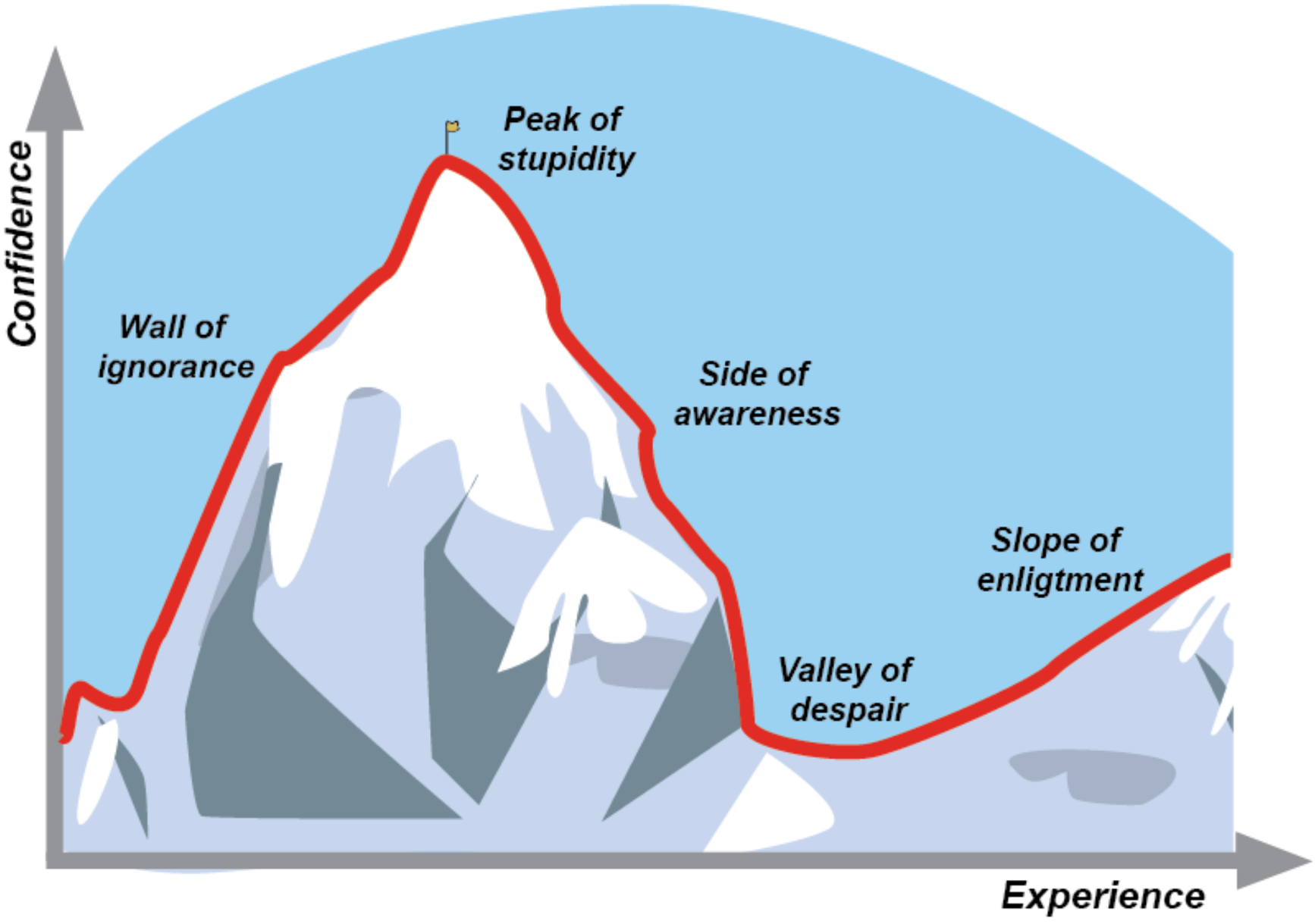
Differnt Versions of the Same

My First Attempt | My Second Attempt | My Third Attempt | **My Forth Attempt**

There is a package `TeachingDemo` and it contains a function dice.

```
# GDA: Rolling Dice Script
# Author: Jean-Claude Walser
# Date: 19.06.23

require(TeachingDemos)
```

Dunning-Kruger Effect Curve

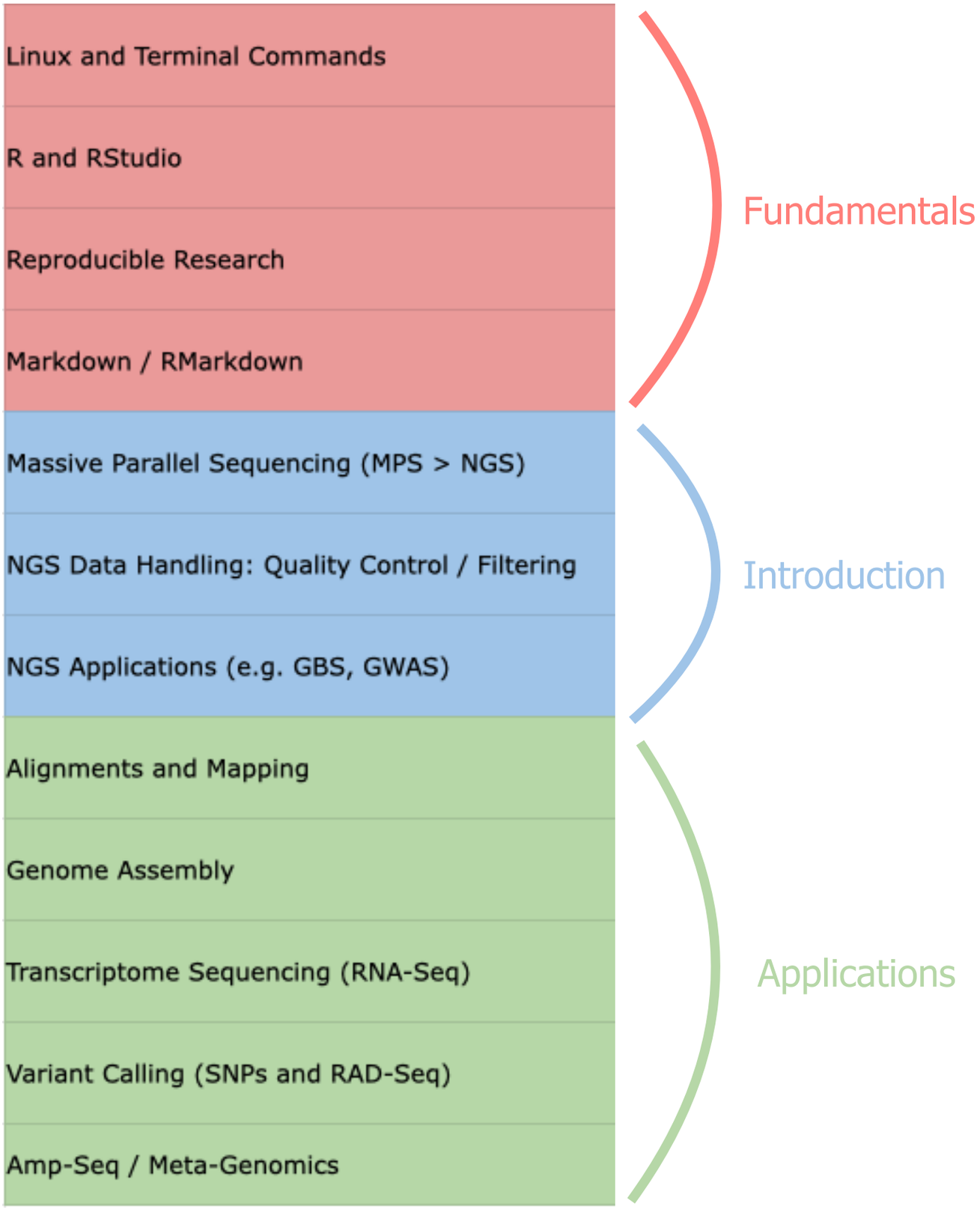


Request #3: Requirements

- ✅ Internet access
- ✅ Linux terminal
- ✅ Code (text) editor
- ✅ MarkDown editor
- ✅ R and RStudio

Genetic Diversity Centre Zurich

Course Structure


Linux and Terminal Commands
R and RStudio
Reproducible Research
Markdown / RMarkdown
Massive Parallel Sequencing (MPS > NGS)
NGS Data Handling: Quality Control / Filtering
NGS Applications (e.g. GBS, GWAS)
Alignments and Mapping
Genome Assembly
Transcriptome Sequencing (RNA-Seq)
Variant Calling (SNPs and RAD-Seq)
Amp-Seq / Meta-Genomics
Fundamentals
Introduction
Applications

GDC Genetic Diversity Centre Zurich

Genetic Diversity: Analysis - 701-1425-00L

GDA25
Week 25+26

Welcome	Time	Topic	Lecturer
Monday, 16.06.25	09:00-10:00	First Things First	JCW / NZ

NGS	Time	Topic	Lecturer
Monday, 16.06.25	10:00-12:00	Ctrl+Alt+Genome: Starting Out in Bioinformatics	JCW
	13:00-15:00	NGS Demystified: A Beginner's Guide to Sequencing	JCW
	15:00-16:00	Discussion & Takeaways	JCW

BioComputing	Time	Topic	Lecturer
Tuesday, 17.06.25	09:00-11:00	Navigating the Terminal: A Beginner's Guide	JCW
	11:00-12:00	Biocomputing with R	NZ
	13:00-15:00	Think & Tinker Time	Self Study
	15:00-16:00	Discussion & Takeaways	JCW / NZ

Help / Advanced Discussion	Time	Topic	Lecturer
Wednesay, 18.06.25	09:00-10:00	Prompt Support (Terminal + R)	JCW / NZ
	10:00-15:00	Think & Tinker Time	Self Study
	15:00-16:00	Advanced Discussion & Takeaways	JCW / NZ

Reproducible Science	Time	Topic	Lecturer
Thursday, 19.06.25	09:00-11:00	Introduction / Group work	NZ
	11:00-12:00	Markdown, RMarkdown & RegEx: Your Toolkit for Clean Code & Reports	NZ
	13:00-15:00	Think & Tinker Time	Self Study
	15:00-16:00	Advanced Discussion & Takeaways	NZ

Genetic Diversity Centre Zurich

9:00 - 12:00 Lecture

Self-study with Challenges

Questions
Problems

15:00 - 16:00 Discussion

Genetic Diversity Centre Zurich

Genetic Diversity Centre (GDC) - Course Webpage

Search

Learning Objectives

Main
◇ Understand the basic principles and differences between massive-parallel sequencing (MPS) platforms.
◇ Understand sample indexing and sequence (read) types.
Minor
◇ Understand the importance of a read data archive and be able to access such data.
◇ Understand the advantages, limitations and applications of MPS in e.g. genomics, transcriptomics or metabarcoding.

Lecture Notes

⬇ MPS (NGS)

Massively Parallel Sequencing

Massive parallel sequencing or massively parallel sequencing is any of several high-throughput approaches to DNA sequencing using the concept of massively parallel processing.

MPS ~ NGS

I use the term massively parallel sequencing (MPS) because it is a more general term and it also includes newer sequencing technologies. Next-generation sequencing (NGS) refers to PCR-based sequencing technologies (e.g. Roche 454, Illumina), while single-molecule sequencing (e.g. PacBio or ONT) is often referred to as next-next or third-generation sequencing. I keep it simple and use MPS.

Genetic Diversity Centre Zurich

Topic	Webside
MPS (NGS)	MPS (NGS)
Terminal	Linux 1 - Local Terminal
	Linux 2 - Remote Terminal
	Biocomputing

Genetic Diversity Centre Zurich

https://www.gdc-docs.ethz.ch/GeneticDiversityAnalysis/GDA/site/

- ➜ Infos
- ➜ Challenges
- ➜ Links
- ➜ Handouts

GDC Genetic Diversity Centre Zurich

Good to know!

E-Mail	jean-claude.walser(at)env.ethz.ch niklaus.zemp(at)env.ethz.ch

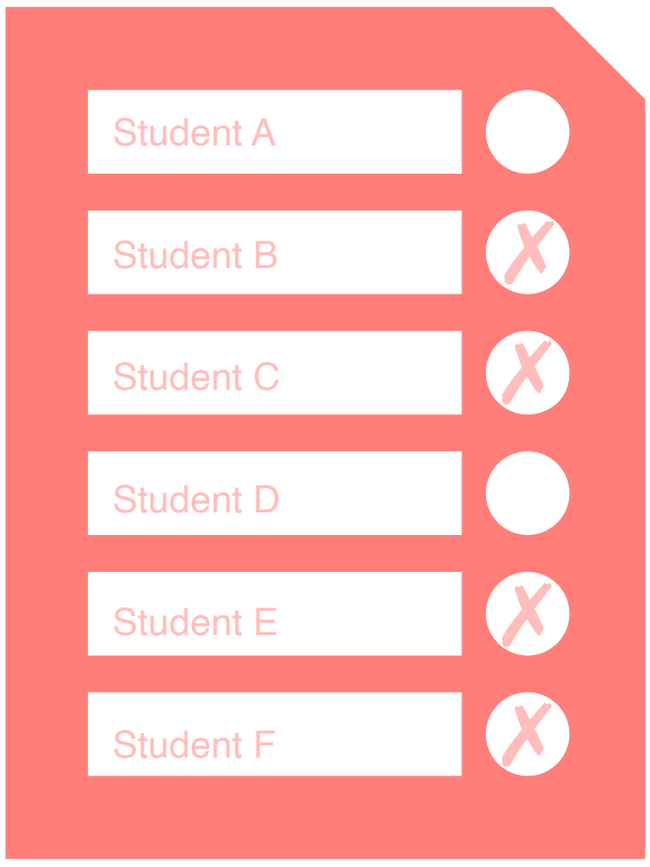
Subject **GDA25: Question**

WWW	https://www.gdc-docs.ethz.ch/GeneticDiversityAnalysis/GDA/site/

SSH	ssh guest01@gdc-vserver.ethz.ch

Genetic Diversity Centre Zurich

GDC Genetic Diversity Centre Zurich



- it is your course
- it is your choice
- it is your decision

- pick a project
- hand in a report (md/html)
- hand it in before deadline
- feedback if you wish

Feedback

helpful information or criticism that is given to someone to say what can be done to improve a performance.