

Scientific Programming

Practical 1

Introduction

Luca Bianco - Academic Year 2019-20
luca.bianco@fmach.it

Outline

- ❖ Personal introduction
- ❖ Introduction to the practical
- ❖ Hands-on practical

About me

Computer Science

Ph.D. at the University of Verona, Italy, with thesis on Simulation of Biological Systems

Research Fellow at Cranfield University - UK

Three years at Cranfield University working at proteomics projects (GAPP, MRMAid, X-Tracker...)

Module manager and lecturer in several courses of the MSc in Bioinformatics

Bioinformatician at IASMA – FEM

Currently bioinformatician in the Computational Biology Group at Istituto Agrario di San Michele all'Adige – Fondazione Edmund Mach, Trento, Italy

Collaborator uniTN - CiBio

I ran the Scientific Programming Lab for QCB for the last couple of years

Fondazione Edmund Mach

FEM – San Michele, Trento - Italy



Agricultural Institute

Research and Innovation Centre

Genomics, metabolomics wet labs on
fruits (apple, grape, small fruits,...)

Bioinformatics and computational biology

Bioinformatics @FEM (UBC)

❖ Genomics

- Assembly and annotation of complex genomes (plants, insects, etc.)
- Development of SNP Chips for genetic screening
- Resequencing of genomes / Variant discovery

❖ Metagenomics

- Targeted metagenomic data
- *Feature selection* algorithms
- Algorithms for strain-level identification from un-targeted metagenomics

❖ Transcriptomics

- RNA-seq data analysis, gene and pathway enrichment
- Data integration and compilation of expression atlases

❖ Metabolomics

- Data analysis pipelines for targeted and untargeted data
- Methods for MS imaging

❖ Statistical data analysis

- Integration of -omic data and analysis of correlation networks



Bix @FEM - Examples

Genome assembly

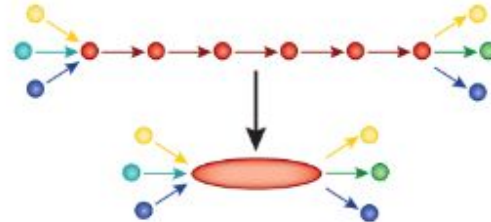
1. Fragment DNA and sequence



2. Find overlaps between reads

...AGCCTAGACCTACA**GGATGCGCGACACGT**
GGATGCGCGACACGTCGCATATCCGGT

3. Assemble overlaps into contigs



4. Assemble contigs into scaffolds



[from M. Baker, Nature Methods, 2014]



Bix @FEM - Examples

Genome assembly of DH of Pear and Apple

Input data:

Illumina: ~60x – 100x PE information + (mate pairs for Apple)

Pacific Biosciences ~ 30x + 30x (35x only for Apple)

Bionano optical maps: ~ 600x (for both)

Hi-C: pear only

Genetic maps: integrated genetic map from 21 mapping populations (Apple only)

Output result (example for Apple):

Chromosome scale assembly

Contigs: 2150 for a total of **625Mb**

N50 Contigs (hybrid dbg2olc): ~ **620Kbps**

280 **Scaffolds**, for an N50 **5,6Mb**

17 chromosomes + 1g0 unanchored sequences

[Daccord et al, Nature Genetics, 49, 2017; Linsmith et al., GigaScience, 2020]





Bix @FEM - Examples

SNP-Chips development for GWAS

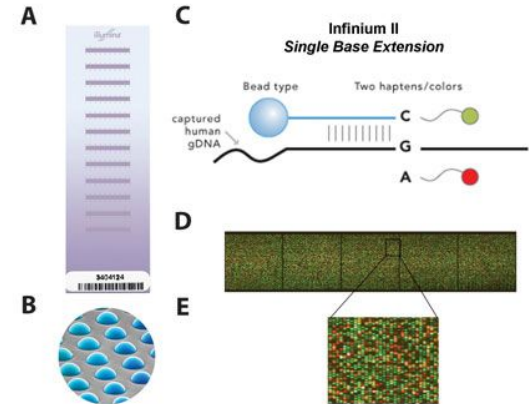
20K SNP Illumina Infinium II Array (reseq of 16 Apple cultivars, Illumina 30x)

487K SNP Affymetrix Axiom Array (reseq of 63 Apple cultivars, Illumina 20-30x)

1. Reads alignment and filtering
2. SNP calling
3. Identification of most reliable SNPs
4. Selection of (20K) 487K target SNPs

Several Terabytes of data produced!!!!

Peach, pear and walnut done too!



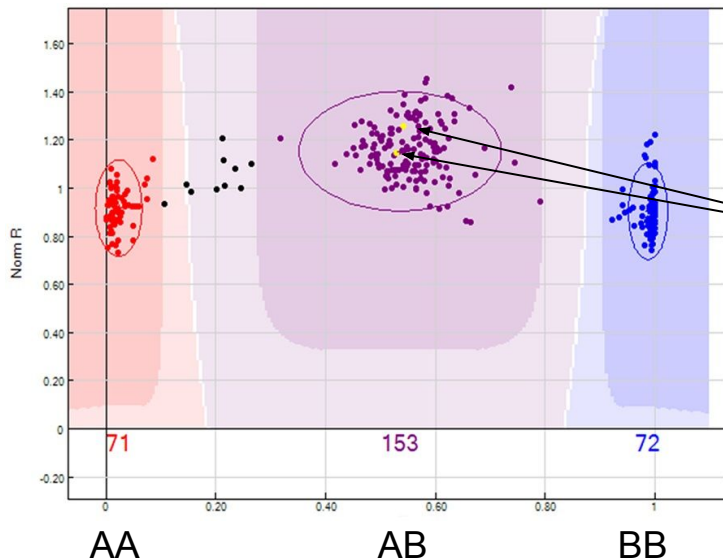


Bix @FEM - Examples

SNP-Chips development for GWAS

20K SNP Illumina Infinium II Array (reseq of 16 Apple cultivars, Illumina 30x)

487K SNP Affymetrix Axiom Array (reseq of 63 Apple cultivars, Illumina 20-30x)



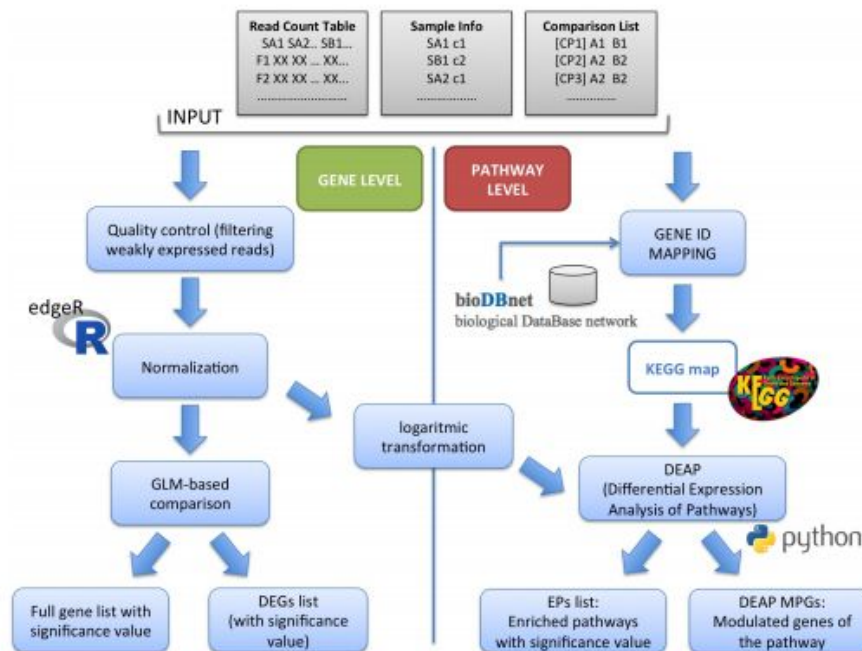
Task:

Analyze 500,000 of these...
(1 x SNP)

parents

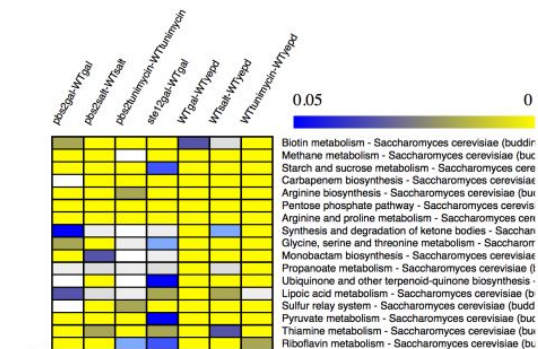
Bix @FEM - Examples

RNAseq data analysis with Pathway Inspector



Bix @FEM - Examples

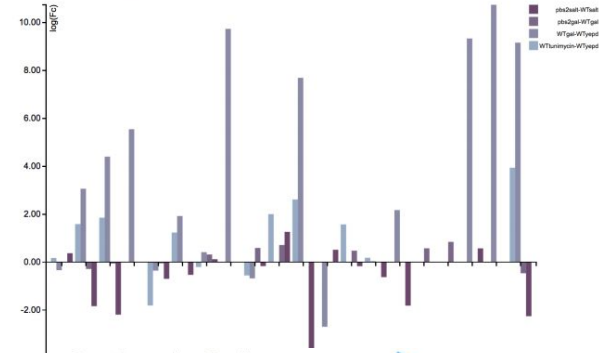
RNAseq data analysis with Pathway Inspector



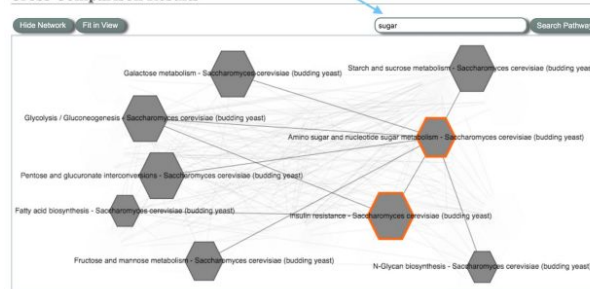
Intersection pbs2salt-WTsalt WTtunimycin-WTyeptd pbs2gal-WTgal WTgal-WTyeptd

| Gene | Comparison | P-value | Fold Change | FDR |
|---------|---------------------|----------------------|--------------------|----------------------|
| YOR04W | pbs2gal-WTsalt | 0.028480382028946 | 0.2420311772254 | 0.0481361748189205 |
| | WTtunimycin-WTyeptd | 2.38460349315057e-07 | 0.524172138345184 | 4.474873728693e-07 |
| | pbs2gal-WTgal | 0.00087890548504029 | -0.301105531814261 | 0.0207830810500887 |
| | WTgal-WTyeptd | 2.86788386434103e-15 | 0.83218983220708 | 7.1912488256014e-15 |
| YIL140W | pbs2gal-WTsalt | 3.85247940741733e-08 | 0.833589168058476 | 1.88548837803052e-07 |
| | WTtunimycin-WTyeptd | 1.58817870171733e-24 | 1.001609102890 | 5.8239103817754e-24 |
| | pbs2gal-WTgal | 1.87826251152733e-09 | 0.30764051192865 | 1.49862020124624e-08 |
| | WTgal-WTyeptd | 0.3034912942367044 | 0.276548277440217 | 0.0548210627660931 |
| YLR130C | pbs2gal-WTsalt | 0.301481010188815 | 0.31882150138888 | 0.0308287774872549 |
| | WTtunimycin-WTyeptd | 3.0140342077802e-06 | -1.58765788614284 | 3.8332186378679e-05 |
| | pbs2gal-WTgal | 2.7670331688813e-09 | 0.488214138134541 | 2.3895374291845e-08 |
| | WTgal-WTyeptd | 8.80372033688032e-18 | -0.8312188730247 | 1.8224549730366e-17 |
| YOR011W | pbs2gal-WTsalt | 1.00574728111917e-16 | 0.72183031688741 | 8.2817487473501e-16 |
| | WTtunimycin-WTyeptd | 0.0108623879303186 | 0.21641810221331 | 0.0147157873255527 |

Comparative analysis for Amino sugar and nucleotide sugar metabolism - Saccharomyces cerevisiae (budding yeast)



Cross-Comparison Results



Bix @FEM - Examples

Pedigree-based haplotype visualization

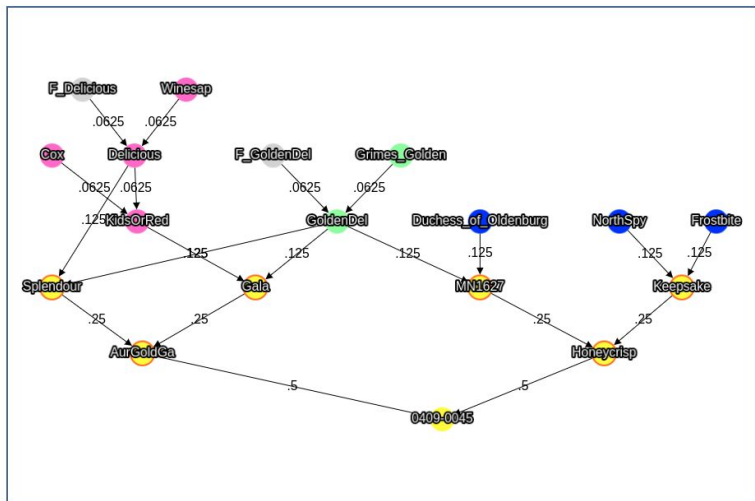
Family of sample: 0409-0045

0409-0045

Fit in View

View global stats

Pedigree browse



Legend
Sample Mother Father Both No data

Choose one ancestor or other accessions

- Generation 1
 - Honeycrisp
 - AurGoldGa
 - Generation 2
 - Keepsake
 - MN1627
 - Gala
 - Splendour
 - Generation 3
 - Generation 4
 - Generation 5
- Relatedness 0-10% [2]
 - Relatedness 10-20% [2]
 - Relatedness 20-30% [17]
 - Relatedness 30-40% [182]
 - Relatedness 40-50% [183]
 - Relatedness 50-60% [9]
 - Relatedness 60-70% [0]
 - Relatedness 70-80% [0]
 - Relatedness 80-90% [0]
 - Relatedness 90-100% [0]

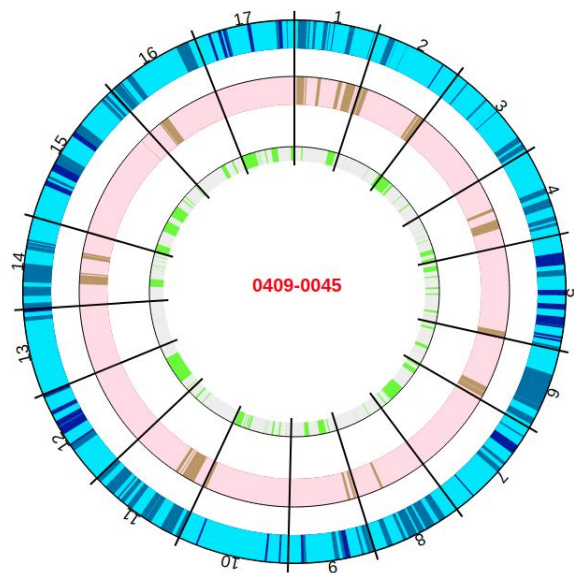
Comparison without pedigree check Comparison with pedigree check

Analyze

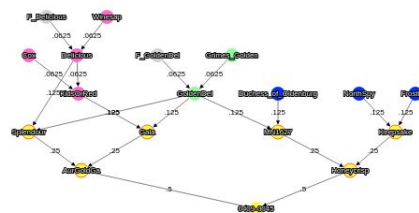
Relatedness

Bix @FEM - Examples

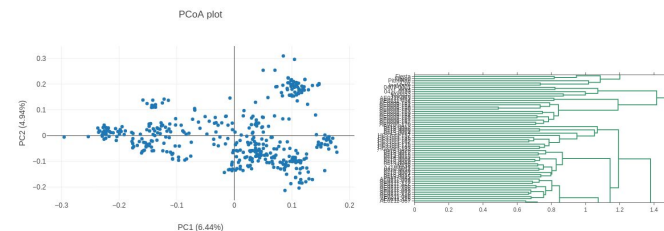
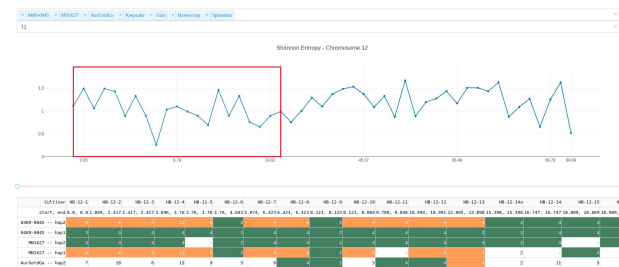
Pedigree-based haplotype visualization



Pedigree of 0409-0045



| Cultivars | Relatedness | Color | |
|------------|-------------|--------|---|
| MN1627 | 0.39 | 1CE6FF | ✓ |
| AurGoldGa | 0.55 | FF4A46 | ✓ |
| Keepsake | 0.42 | 008FA6 | ✓ |
| Gala | 0.48 | FFDBE5 | ✓ |
| Honeycrisp | 0.52 | 0000A6 | ✓ |
| Splendour | 0.48 | B79762 | ✓ |



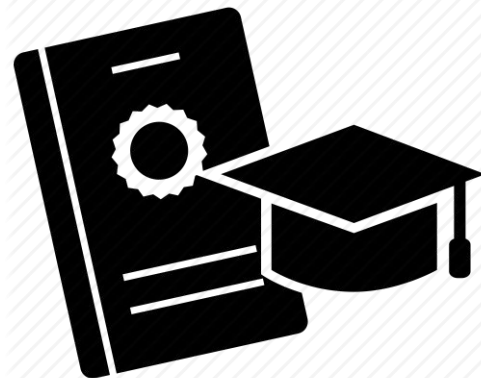
Opportunities @FEM

MSc External thesis

Are you interested in a bioinformatics project in NGS data analysis, RNA Seq, data integration?

Come and talk to me or email me at:

luca.bianco@fmach.it



Scientific Programming Practical

Back to business now!



Scientific Programming Practical

In this practical you will

1. Install Python 3.x (and pip)
2. Install Visual Studio Code
3. Get familiar with the Python console
4. Start using Visual Studio Code and advanced features (like debugging)
5. End the session with some exercises

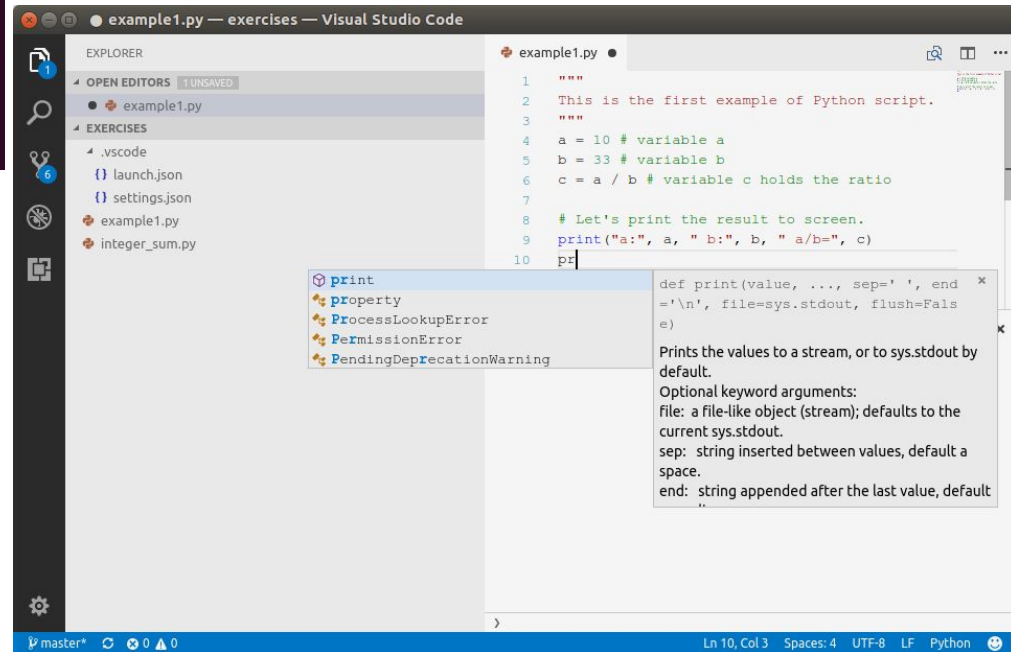


Scientific Programming Practical

Console VS. Integrated Development Environment (IDE)

```
biancol@bluhp:~$ python3
Python 3.5.2 (default, Aug 18 2017, 17:48:00)
[GCC 5.4.0 20160609] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

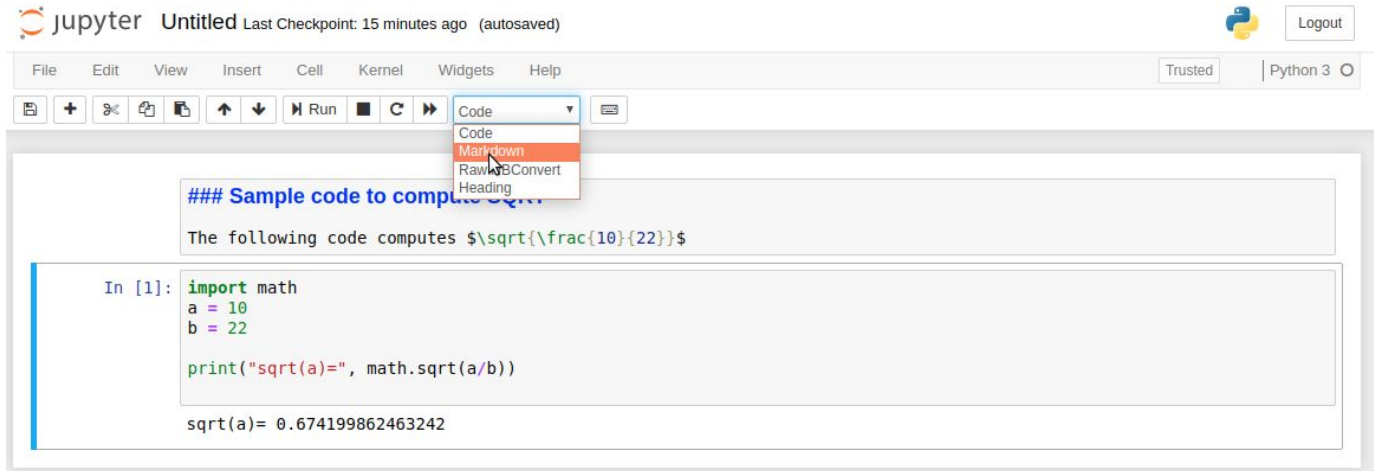
```
>>> print("Hi there")
Hi there
>>> print("{} + {} = {}".format(10,20, 10+20))
10 + 20 = 30
>>>
```



Notebooks and Jupyter

Notebooks contain both the **code**, some **text describing the code** and the **output of the code execution**,

Jupyter is becoming the de-facto standard for writing technical documentation.

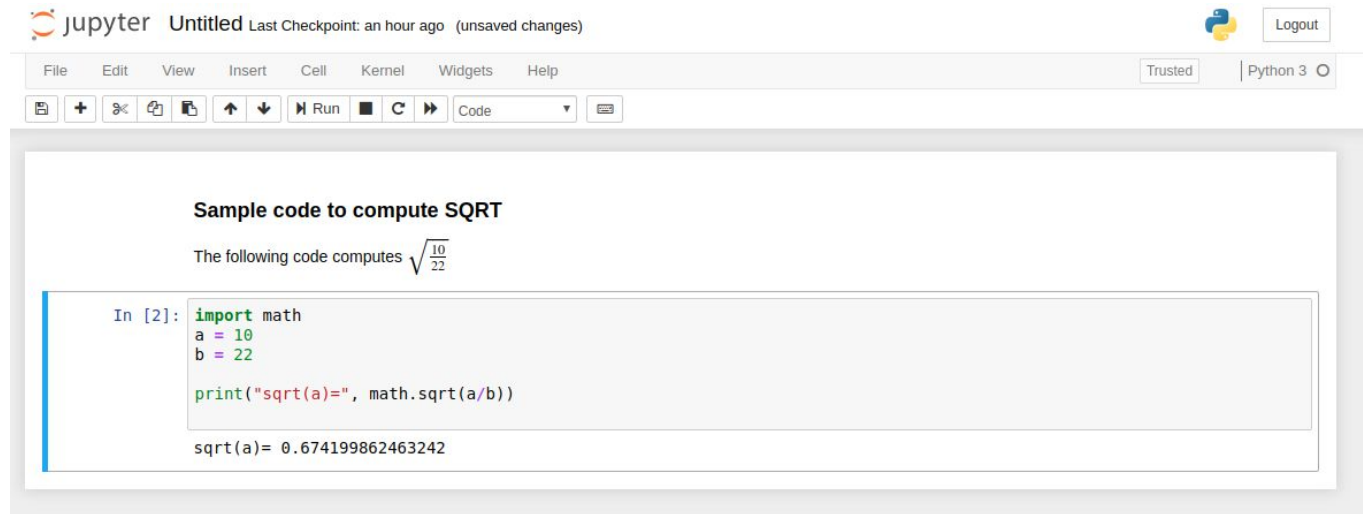


Notebooks and Jupyter

Notebooks contain both the **code**, some **text describing the code** and the **output of the code execution**,

Jupyter is becoming the de-facto standard for writing technical documentation.

A cell can be executed by clicking on **Run**



The screenshot displays the Jupyter Notebook web interface. At the top, the title bar shows 'jupyter' with its logo, followed by 'Untitled' and 'Last Checkpoint: an hour ago (unsaved changes)'. On the right, there is a 'Logout' button and a 'Python 3' selector. Below the title bar is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. A toolbar contains icons for saving, adding a new cell, deleting a cell, undo, redo, and running the code. The main area shows a code cell with the following content:

Sample code to compute SQRT

The following code computes $\sqrt{\frac{10}{22}}$

```
In [2]: import math
a = 10
b = 22

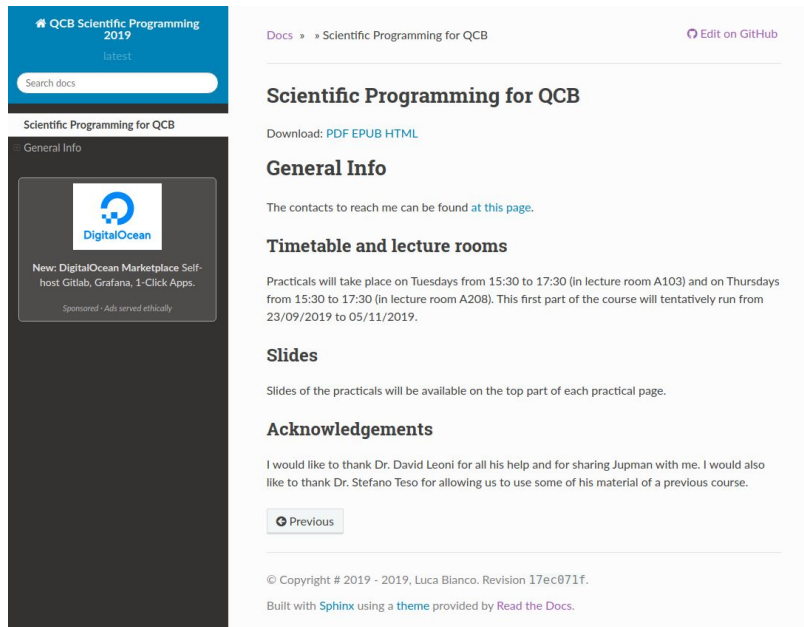
print("sqrt(a)=", math.sqrt(a/b))
```

sqrt(a)= 0.674199862463242

Resources

All material regarding practicals will be found here:

<http://qcbsciprolab2019.readthedocs.io>




The screenshot shows the documentation page for 'QCB Scientific Programming 2019' on Read the Docs. The page has a dark blue header with the title and a search bar. Below the header, there's a sidebar with 'General Info' and a DigitalOcean logo. The main content area has a light blue background and contains sections for 'Scientific Programming for QCB', 'General Info', 'Timetable and lecture rooms', 'Slides', and 'Acknowledgements'. The 'Acknowledgements' section mentions Dr. David Leoni and Dr. Stefano Teso. At the bottom, there's a 'Previous' button and a copyright notice.

QCB Scientific Programming 2019
latest

Search docs

Scientific Programming for QCB

General Info


New: DigitalOcean Marketplace Self-host Gitlab, Grafana, 1-Click Apps.
Sponsored - Ads served ethically

Docs » Scientific Programming for QCB [Edit on GitHub](#)

Scientific Programming for QCB

Download: [PDF](#) [EPUB](#) [HTML](#)

General Info

The contacts to reach me can be found [at this page](#).

Timetable and lecture rooms

Practicals will take place on Tuesdays from 15:30 to 17:30 (in lecture room A103) and on Thursdays from 15:30 to 17:30 (in lecture room A208). This first part of the course will tentatively run from 23/09/2019 to 05/11/2019.

Slides

Slides of the practicals will be available on the top part of each practical page.

Acknowledgements

I would like to thank Dr. David Leoni for all his help and for sharing Jupman with me. I would also like to thank Dr. Stefano Teso for allowing us to use some of his material of a previous course.

[Previous](#)

© Copyright # 2019 - 2019, Luca Bianco. Revision 17ec071f.
Built with [Sphinx](#) using a [theme](#) provided by [Read the Docs](#).



luca.bianco@fmach.it

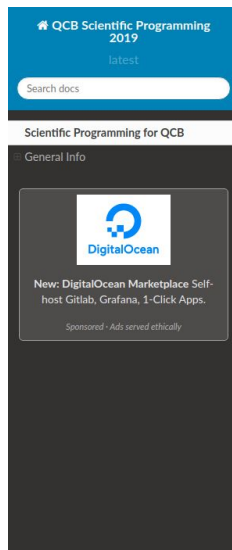
Timetable

Tuesdays:

A103: 15,30 - 17,30

Thursdays:

A208: 15,30 - 17,30



Docs » » Scientific Programming for QCB

[Edit on GitHub](#)

Scientific Programming for QCB

Download: [PDF](#) [EPUB](#) [HTML](#)

General Info

The contacts to reach me can be found [at this page](#).

Timetable and lecture rooms

Practicals will take place on Tuesdays from 15:30 to 17:30 (in lecture room A103) and on Thursdays from 15:30 to 17:30 (in lecture room A208). This first part of the course will tentatively run from 23/09/2019 to 05/11/2019.

Slides

Slides of the practicals will be available on the top part of each practical page.

Acknowledgements

I would like to thank Dr. David Leoni for all his help and for sharing Jupman with me. I would also like to thank Dr. Stefano Teso for allowing us to use some of his material of a previous course.

[Previous](#)

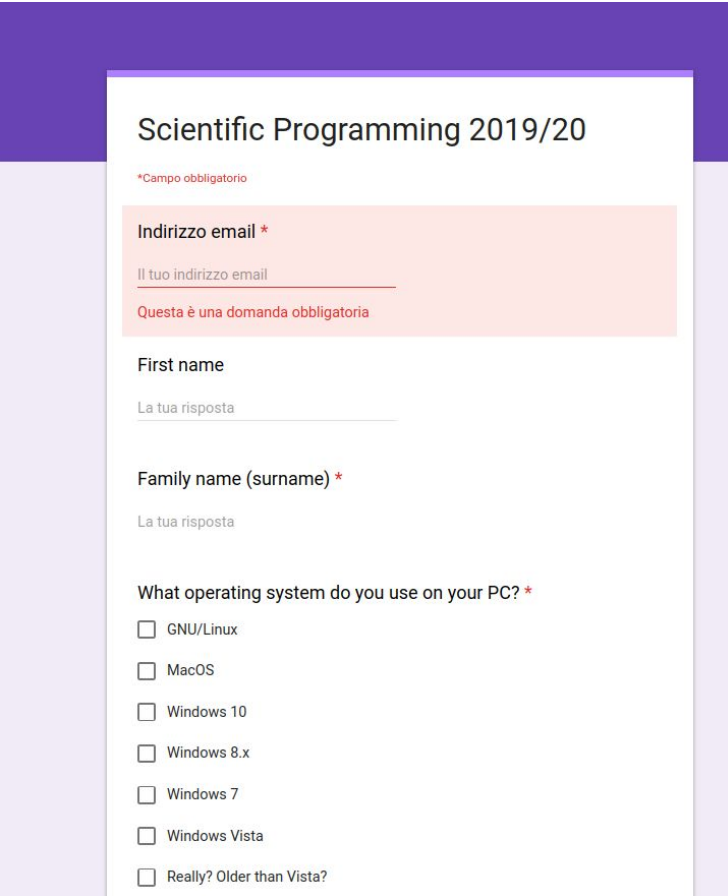


luca.bianco@fmach.it

Please, fill the form at

<https://tinyurl.com/y4qszzv8>

Deadline Sunday, September 29th



Scientific Programming 2019/20

*Campo obbligatorio

Indirizzo email *

Il tuo indirizzo email

Questa è una domanda obbligatoria

First name

La tua risposta

Family name (surname) *

La tua risposta

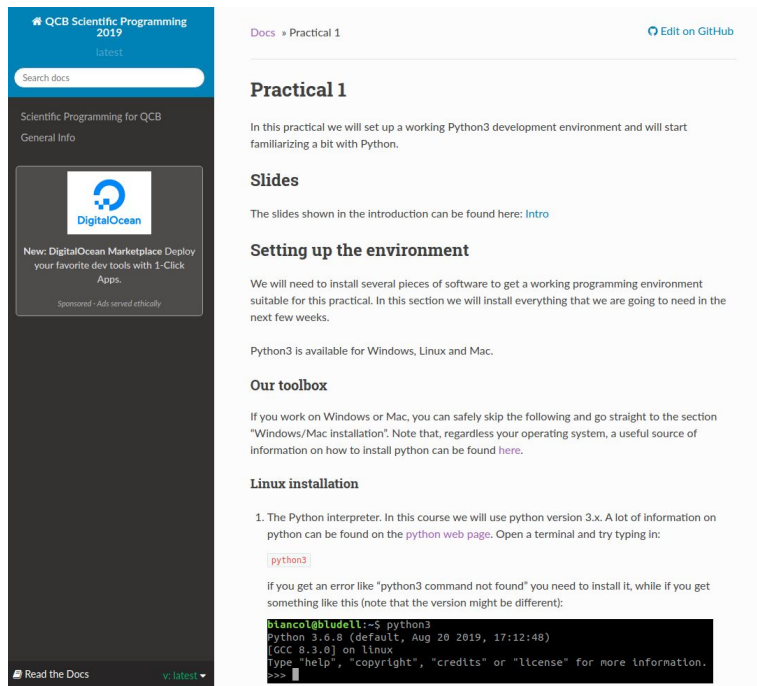
What operating system do you use on your PC? *

- ☐ GNU/Linux
- ☐ MacOS
- ☐ Windows 10
- ☐ Windows 8.x
- ☐ Windows 7
- ☐ Windows Vista
- ☐ Really? Older than Vista?

Any questions?

If not, please go to:

<https://qcbsciprolab2019.readthedocs.io/en/latest/introduction.html>



The screenshot shows the documentation page for 'QCB Scientific Programming 2019' on the Read the Docs platform. The left sidebar contains a search bar, navigation links for 'Scientific Programming for QCB' and 'General Info', a DigitalOcean advertisement, and a 'Read the Docs' footer. The main content area is titled 'Practical 1' and includes sections for 'Slides', 'Setting up the environment', 'Our toolbox', and 'Linux installation'. The 'Linux installation' section lists the Python interpreter as the first step and includes a terminal snippet showing the command to run 'python3' and its output.

QCB Scientific Programming 2019
latest

Search docs

Scientific Programming for QCB
General Info

New: DigitalOcean Marketplace Deploy your favorite dev tools with 1-Click Apps.
Sponsored - Ads served ethically

Docs » Practical 1 [Edit on GitHub](#)

Practical 1

In this practical we will set up a working Python3 development environment and will start familiarizing a bit with Python.

Slides

The slides shown in the introduction can be found here: [Intro](#)

Setting up the environment

We will need to install several pieces of software to get a working programming environment suitable for this practical. In this section we will install everything that we are going to need in the next few weeks.

Python3 is available for Windows, Linux and Mac.

Our toolbox

If you work on Windows or Mac, you can safely skip the following and go straight to the section "Windows/Mac installation". Note that, regardless your operating system, a useful source of information on how to install python can be found [here](#).

Linux installation

1. The Python interpreter. In this course we will use python version 3.x. A lot of information on python can be found on the [python web page](#). Open a terminal and try typing in:

```
python3
```

If you get an error like "python3 command not found" you need to install it, while if you get something like this (note that the version might be different):

```
blanco@bludell:~$ python3
Python 3.6.8 (default, Aug 20 2019, 17:12:48)
[GCC 8.3.0] on linux
Type "help", "copyright", "credits" or "license()" for more information.
>>>
```



luca.bianco@fmach.it