Scientific Programming Practical 1 (QCB)

--> Data science in A103

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 Scienitific 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,…)

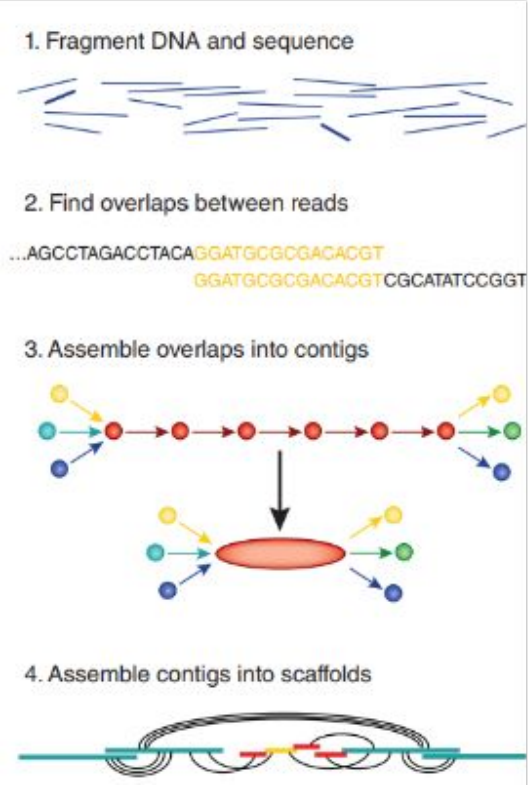
Bioinformatics and computational biology

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



In a nutshell… (Tunis' version…)

Reads

Assembled genome

[Virgil and the Muses, Bardo Museum, Tunis]

[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 + lg0 unanchored sequences

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

Bix @FEM - Examples

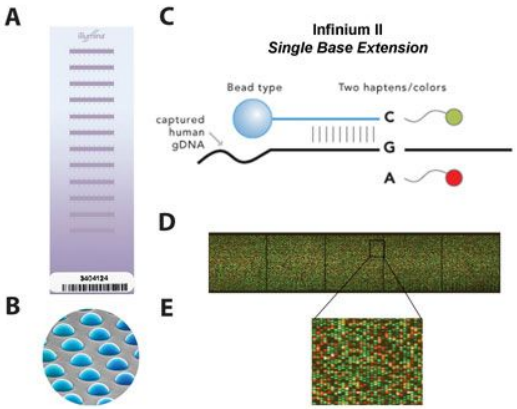
SNP-Chips development for GWAS

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)
600K SNP Affymetrix Axiom Array Walnut (reseq. 18 cultivars, Illumina 80x)

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!

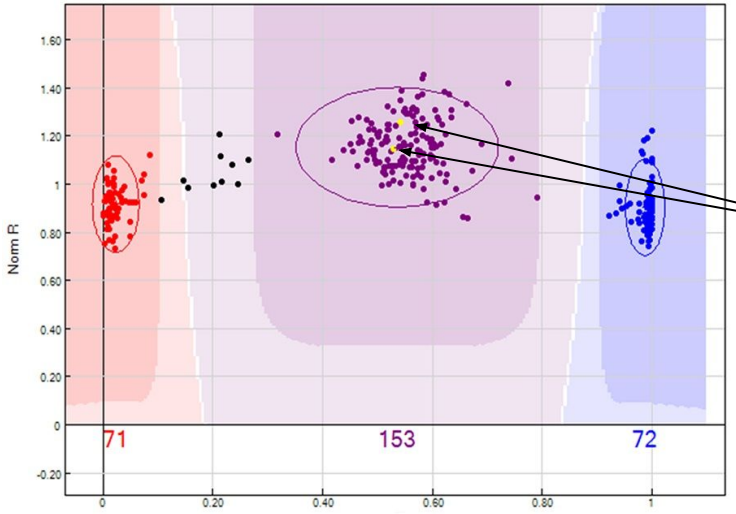


[Bianco et al., PloS One, 2014; Bianco et al., the Plant Journal 2016; Marrano et al., the Plant Journal 2018]

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)
600K SNP Affymetrix Axiom Array Walnut (reseq. 18 cultivars, Illumina 80x)

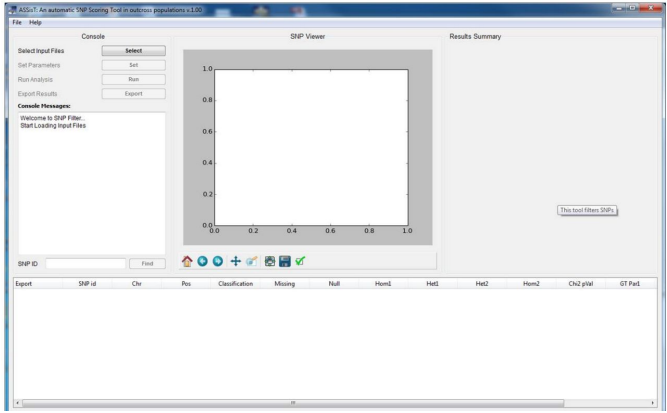


AA AB BB genotypes

Task:
Analyze 500,000 of these…
(1 x SNP)

parents

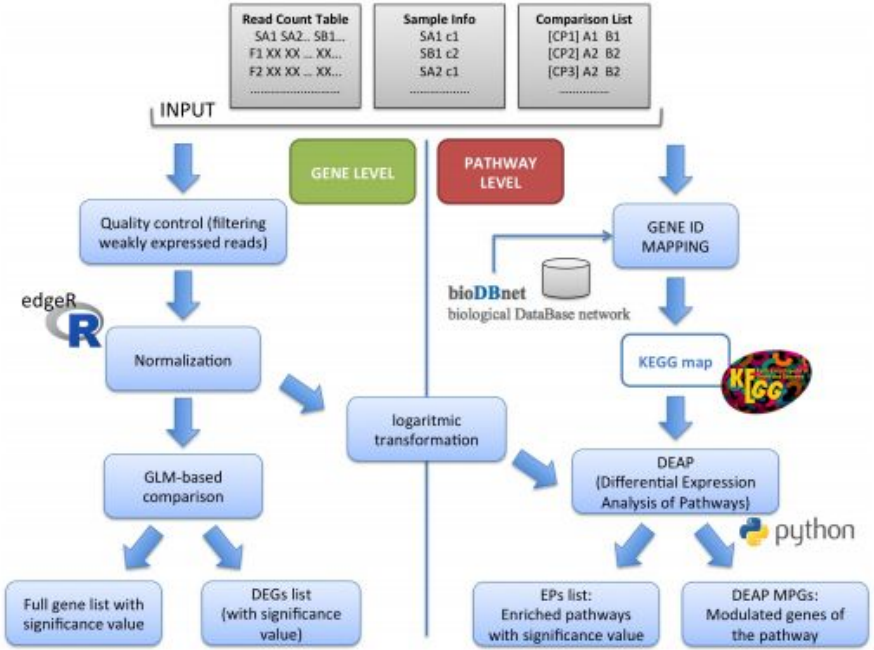
ASSIsT



[Di Guardo et al., Bioinformatics, 2015]

Bix @FEM - Examples

RNAseq data analysis with Pathway Inspector

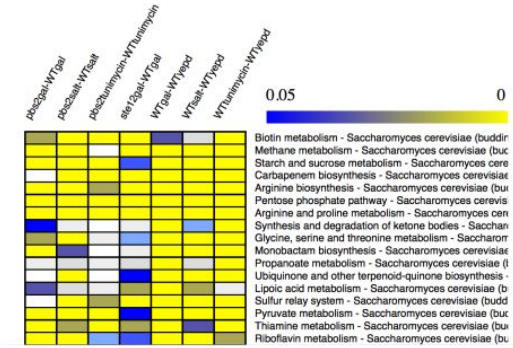


https://pathwayinspector.fmach.it

[Bianco et al., Bioinformatics, 2017]

Bix @FEM - Examples

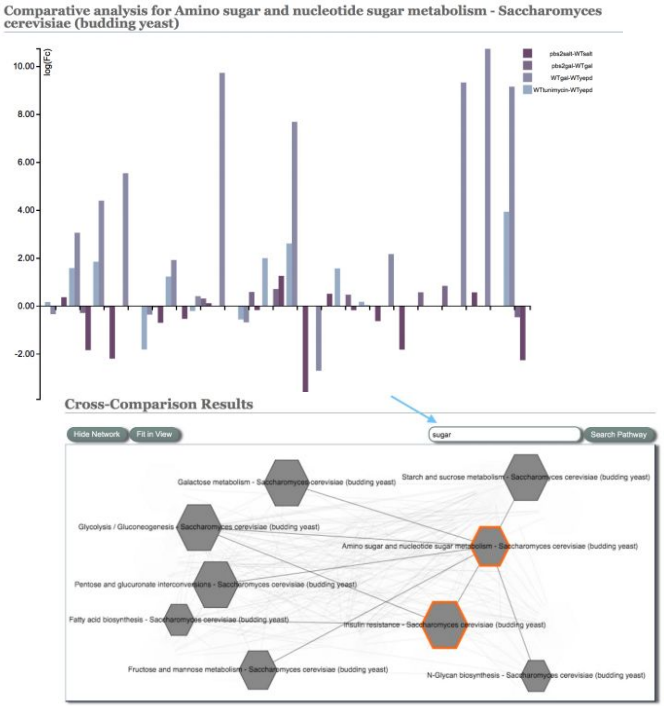
RNAseq data analysis with Pathway Inspector



Intersection pbs2salt-WTsalt WTtunimycin-WTyepd pbs2gal-WTgal WTgal-WTyepd

Export Table

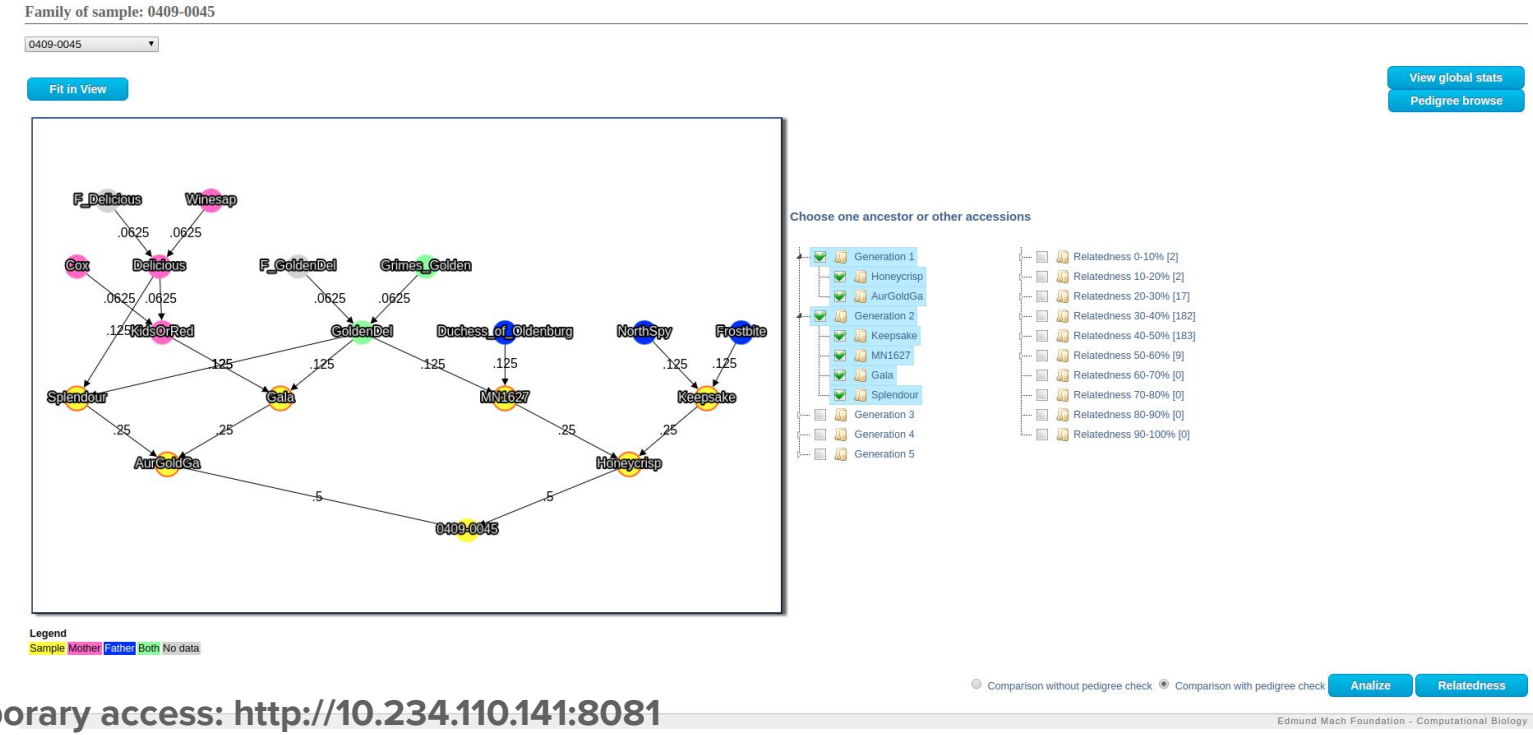
Gene	Comparison	P-value	Fold Change	FDR
YGR046W	pbs2salt-WTsalt	0.02648088292984 6	0.24203131773235 4	0.0481391746189205
	WTtunimycin-WTyepd	2.38460949315057e-07	0.52417213634018 4	4.4746733128663e-07
	pbs2gal-WTgal	0.00086789048530400 9	-0.30110553181420 1	0.00297830816509087
	WTgal-WTyepd	2.86768386434163e-15	0.83218963322068	7.19124882258014e-15
YIL140W	pbs2salt-WTsalt	3.86247940741723e-08	0.63358919069476	1.68546937903529e-07
	WTtunimycin-WTyepd	1.58617879717633e-24	1.00164091529 93	5.92391639177941e-24
	pbs2gal-WTgal	1.67826251752703e-09	0.39764045118286 5	1.49962602124929e-08
	WTgal-WTyepd	0.00349129423675144	0.27654927744021 7	0.00482106276605931
YLR130C	pbs2salt-WTsalt	0.00148100161555513	0.31582150138085 8	0.00396287774872549
	WTtunimycin-WTyepd	3.0140342977802e-86	-1.96765789614264	3.83321563378879e-85
	pbs2gal-WTgal	2.76753316898143e-09	0.46821413913454 1	2.39953774291843e-08
	WTgal-WTyepd	5.89372053985592e-18	-0.83112196773047	1.62245497936366e-17
YOR011W	pbs2salt-WTsalt	1.00574726111917e-16	0.72183053160874 1	8.28174673473501e-16
	WTtunimycin-WTyepd	0.01096236759355166	0.21641810023153 1	0.0147157873255527



https://pathwayinspector.fmach.it

Bix @FEM - Examples

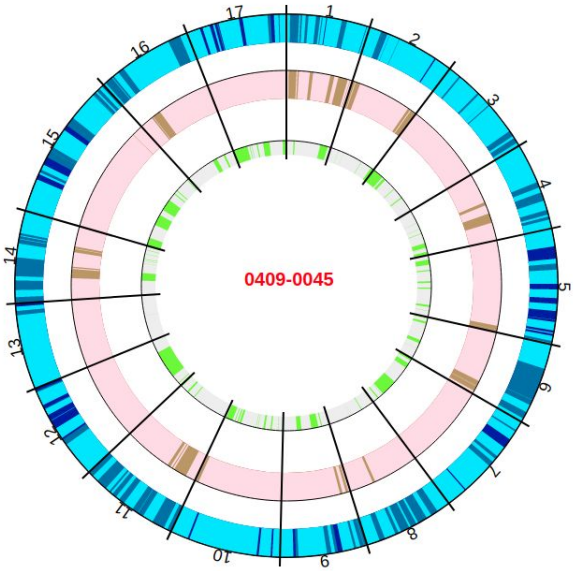
Pedigree-based haplotype visualization

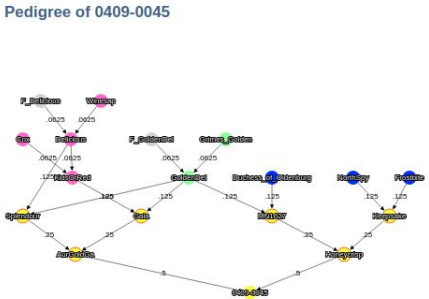


Temporary access: http://10.234.110.141:8081

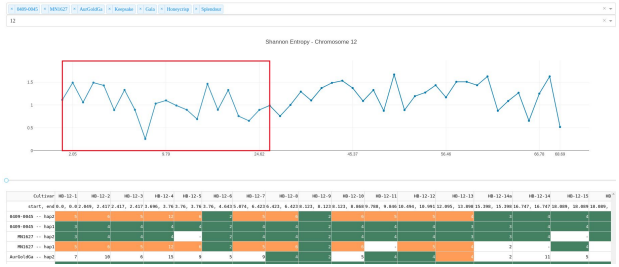
Bix @FEM - Examples

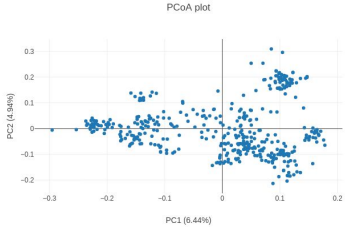
Pedigree-based haplotype visualization

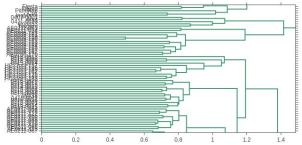




Cultivars	Relatedness	Color
MN1627	0.39	1CE6FF
AurGoldGa	0.55	FF4A46
Keepsake	0.42	006FA6
Gala	0.48	FFDBE5
Honeycrisp	0.52	0000A6
Splendour	0.48	B79762







Temporary access: http://10.234.110.141:8081

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

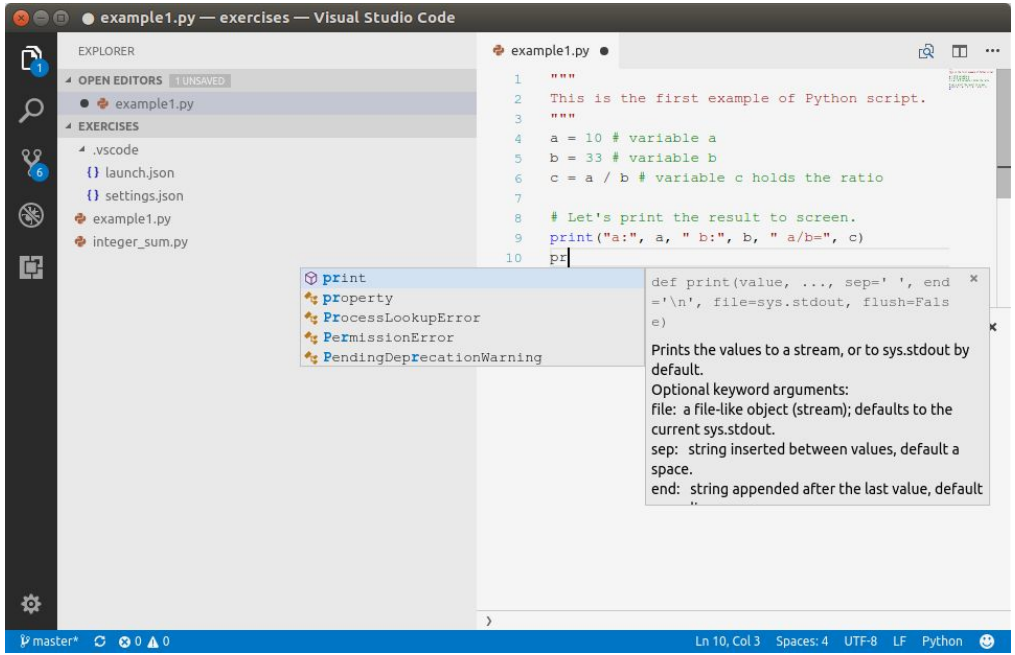
1. Install Python 3.x (and pip)
2. Install Visual Studio Code
3. Get familiar with the Python console
4. Start using Visual Studio Code and advanced features (like debugging)
5. End the session with some exercises

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
>>>
```

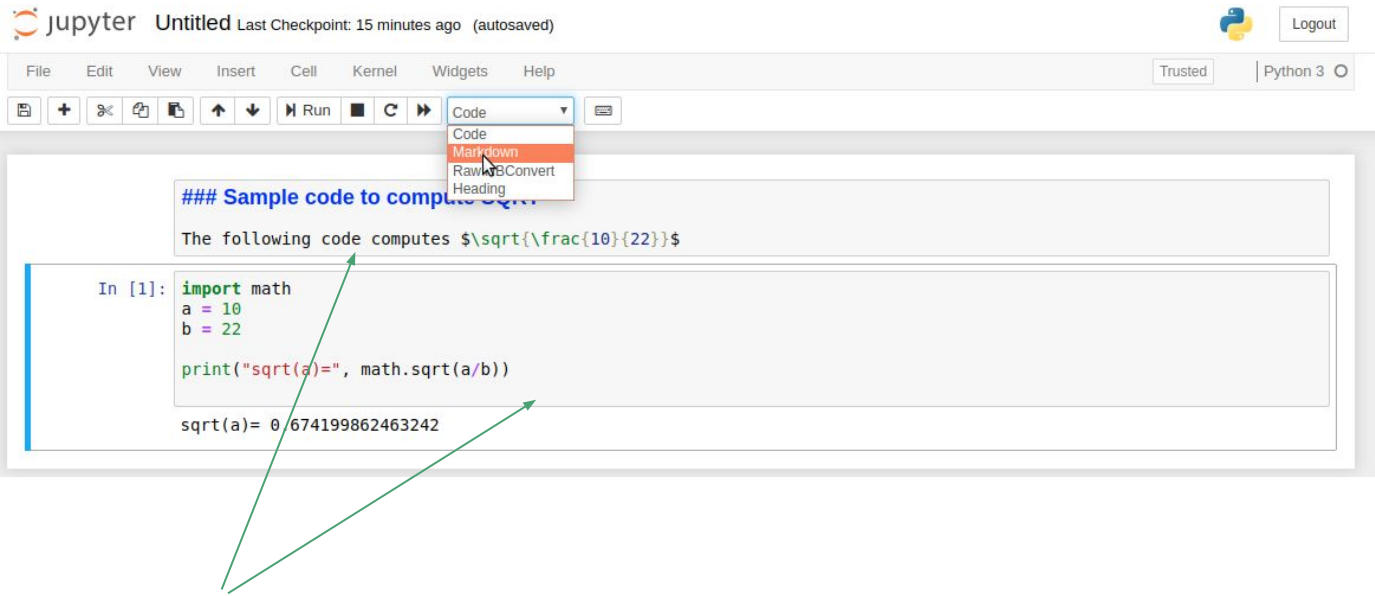


Notebooks and Jupyter

"Jupyter is a web-based interactive development environment for python/R.. notebooks, code, and data."

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.



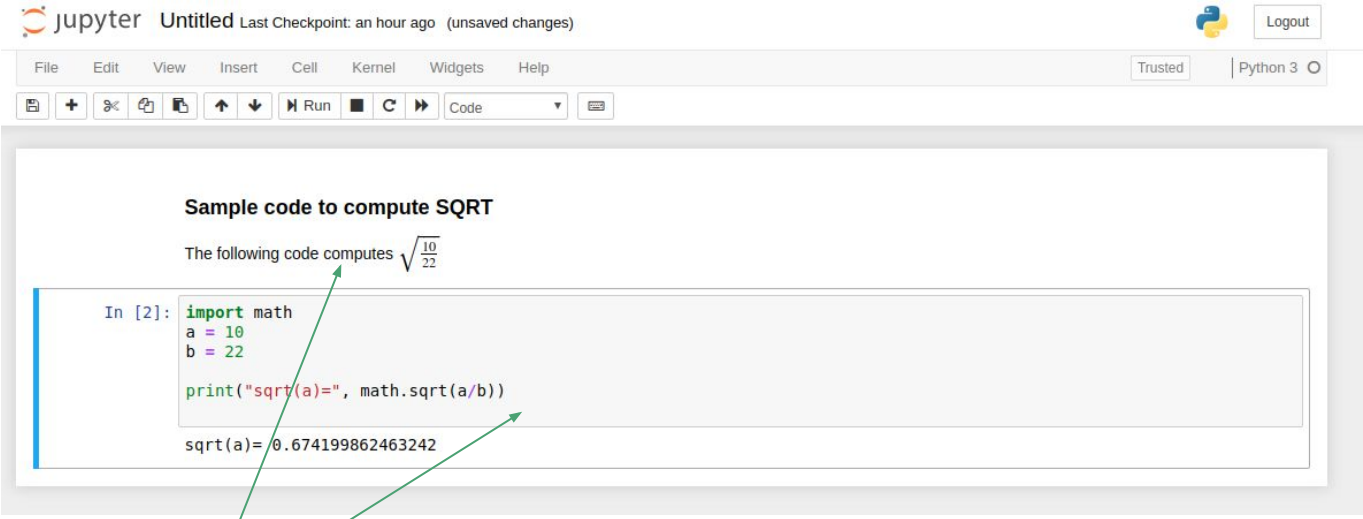
Cells

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**

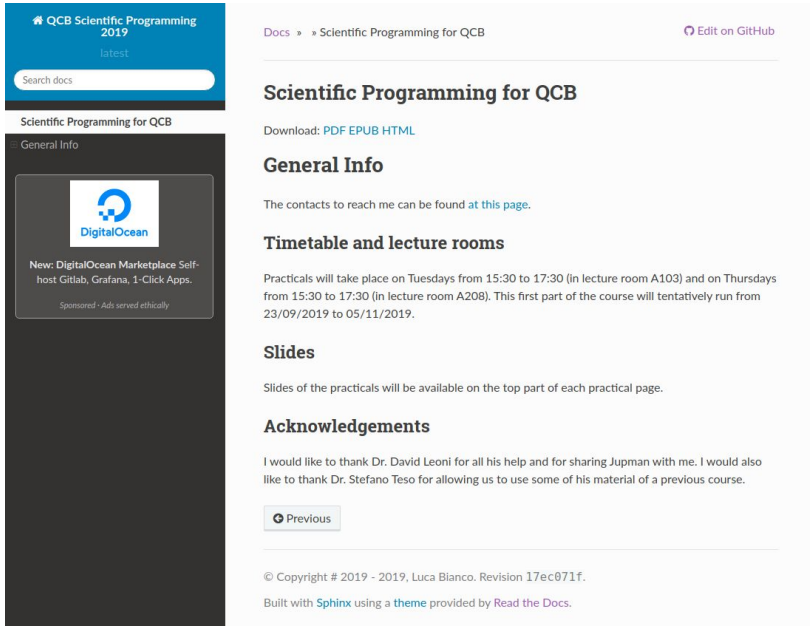


Cells
(after Run)

Resources

All material regarding practicals will be found here:

http://qcbsciprolab2019.readthedocs.io



@
luca.bianco@fmach.it

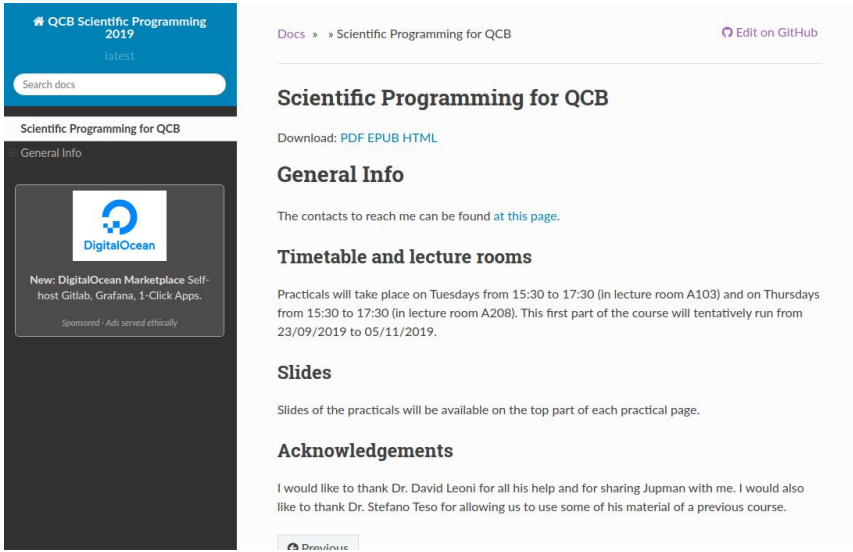
Timetable

Tuesdays:

A107: 15,30 - 17,30

Thursdays:

A107: 15,30 - 17,30



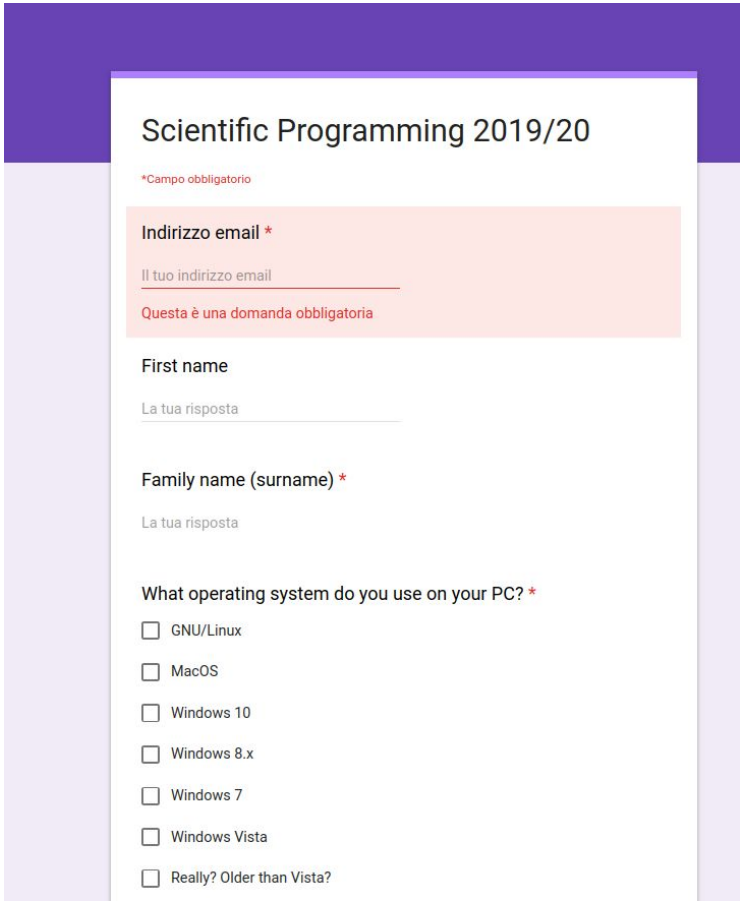
http://qcbsciprolab2019.readthedocs.io

luca.bianco@fmach.it

Please, fill the form at

https://tinyurl.com/y6nlnx7l

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

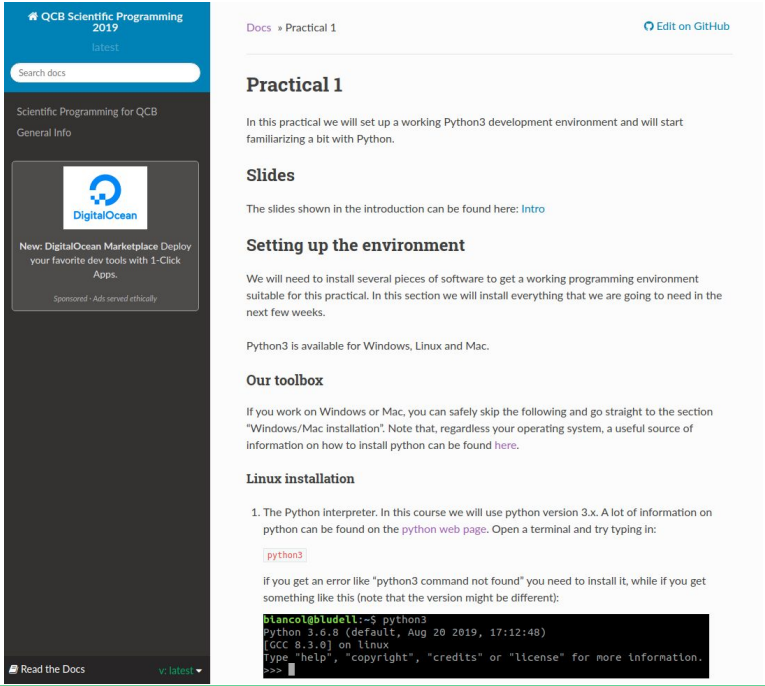
What operating system do you use on your PC? *

- 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



QCB Scientific Programming 2019

latest

Search docs

Scientific Programming for QCB

General Info

Read the Docs v: latest

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):

```
biancol@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