COURSE CONTENTS

10. A simple application
    - FASTQ format
    - processing DNA reads in FASTQ files
    - decoding quality values in FASTQ files
    - counting quality values
    - a full application to show the quality distribution in a FASTQ file

11. More about strings
    - Python's formatting mini-language
    - strings, bytes, and encodings

12. Memory-efficient containers
    - bytes
    - typed arrays

13. Iteration (itertools)
    - generators
    - enumerate
    - combinatorics
    - list comprehensions and generator expression
    - map function

14. Collections
    - defaultdict
    - counter

15. Random numbers
    - random
    - numpy.random

16. Duplicate estimation in an NGS dataset
    - criteria for PCR duplicates
    - hashing DNA reads
    - algorithm for duplicate identification
    - designing the command-line interface
    - putting it all together
    - beautifying the code

17. Documentation
    - docstrings
    - Pythonic code

GENERAL INFORMATION

VENUE

DECHEMA-Haus
Theodor-Heuss-Allee 25
60486 Frankfurt am Main, Germany

LANGUAGE

The course will be held in English.

REGISTRATION

Please complete and return the enclosed form or contact:

DECHEMA-Forschungsinstitut
Training department
P.O. Box 17 03 52
D-60077 Frankfurt am Main

Phone: +49 69 7564 253
Fax: +49 69 7564 414
Internet: http://dechema-dfi.de/kurse
E-mail: gruss@dechema.de

REGISTRATION FEE

765,- €

750.- € (personal DECHEMA members)

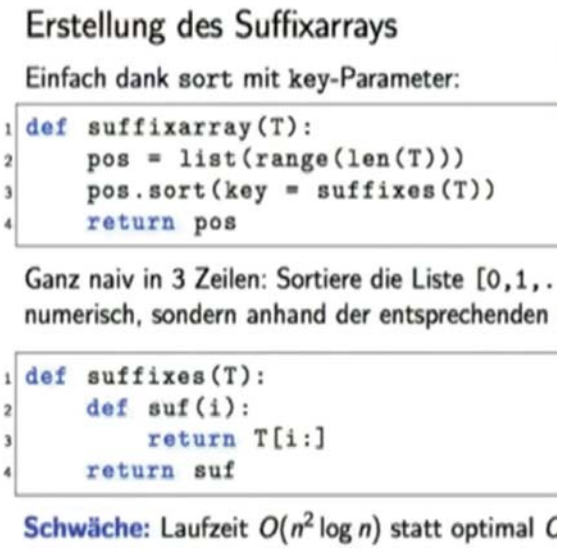
(incl. course materials, certificate of attendance, lunch, coffee breaks)

DECHEMA

TRAINING COURSE

7 - 8 July 2014
Frankfurt am Main/Germany

Introduction to Python for the Biosciences



SUMMARY

This two-day course teaches the basics of the Python programming language. Python is an open-source programming language that runs on each major operating system and offers high readability and programming productivity. No previous programming experience is required. However, we assume that participants come with a working Python installation on their notebooks. Python's language elements will be taught by examining example tools from high throughput DNA sequence analysis with next generation sequencing (NGS) data.

Python is based on the concept of objects defined by classes and operations associated with them. For example, a DNA sequence is a Python object, for which we can implement the reverse complement operation. We thus introduce the terminology of object-oriented programming. In parallel, we discuss Python's statements, which are similar to those in many other programming languages (assignments, loops, conditionals, context managers, error handling by exceptions). We next discuss Python's data types: basic types (booleans, numbers, strings), sequence and container types (lists, bytes, arrays, sets), and dictionaries. To interact with the outside world, we discuss how to write command line programs and work with files. As an example, we compute and output the base quality distribution in a FASTQ file using only elementary programming techniques.

We then develop a more complex application: the rapid estimation of the rate of (PCR) duplicates in a sequencing run and its visualization. For this, we will explore several of Python's advanced features (iterators, generators, comprehensions), standard library modules and extensions (e.g. collections, itertools, numpy, matplotlib). We also discuss how to write good (readable and "Pythonic") code and write documentation.

GOALS

After the course, the participants will be able to write their own simple Python scripts and applications, especially in the context of NGS data. They will find it easier to extend their Python knowledge on their own based on solid foundations and worked examples taught in the course.

TARGET AUDIENCE

Biologists, Chemists, Biotechnologists without substantial programming experience who need to write small data analysis or format conversion scripts in their day-to-day jobs; scientists interested in analysis of sequencing data with Python

COURSE DURATION

Two days, six to seven 45-min lectures on each day.

There will be sufficient time for general discussion and questions, and there will be tutors during exercises for individual questions and problems.

INSTRUCTORS

Prof. Dr. Sven Rahmann, Professor for Genome Informatics at the Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen and at the Faculty of Computer Science, TU Dortmund. Sven Rahmann has been using Python for teaching and in research projects for many years. He has also been a speaker at PyConDE.

Members of the research group, all active Python programmers, will act as tutors.

COURSE CONTENTS

1. A first look - the interactive Python interpreter
    - interpreted vs. compiled languages
    - duck typing
2. Names and values
3. Basic data types
    - Booleans (true, false)
    - numeric types (int, float)
    - strings

COURSE CONTENTS

4. Python statements
    - assignments
    - loops
    - conditionals
    - conditional assignments
    - function definitions
    - exceptions
    - context managers
5. Object-oriented programming (OOP)
    - classes
    - attributes
    - methods
    - constructor __init__
    - inheritance
    - encapsulation
6. Container types
    - lists and tuples
    - sets
    - dictionaries
7. Modules
    - module namespaces
    - import
8. Command-line arguments
    - designing a command-line interface
    - argparse module
    - writing a command-line application
9. Working with files
    - opening a file
    - reading from files
    - writing to files
    - standard input, standard output, standard error
    - redirection

I

Reply form
(Fax-No.: +49 69 7564-414)

DECHEMA-Forschungsinstitut
Training department
P.O. Box 17 03 52
D-60077 Frankfurt am Main

Registration to the DECHEMA training course 7172 Python
"Introduction to Python for the Biosciences" Frankfurt am Main, 7 - 8 July 2014
Deadline for registration: 16 June 2014

Participant

Ms ☐ Mr ☐ Academic degree _____

Name _____ Surname _____

Company _____

Department _____

Street/POB _____

Code/Place _____

Phone/Fax _____ E-mail _____

I am a personal DECHEMA-member ☐ yes ☐ no

Invoice address (if different)

Company _____

Department _____

Street/POB _____

Code/Place _____

Method of payment

☐ bank transfer after receipt of invoice

☐ by credit card:

☐ Mastercard ☐ Visa

Card number _____ Expiration date _____/_____

The course fee amounts to 765.- € / 750.- € (personal DECHEMA members). If we receive a notice of withdrawal at least two weeks prior to the beginning of the course, the participation fee less 10% for administration expenses will be reimbursed. Thereafter, a reimbursement will not be possible.

_____ _____

Place, date signature + company stamp