Tool deployment and sustainability in bioinformatics



Bérénice Batut, Jonas Weber, Rolf Backofen, Björn Grüning

University of Freiburg

Bioinformatics for Human Health and Disease Heidelberg - November 2016

Biomedical research

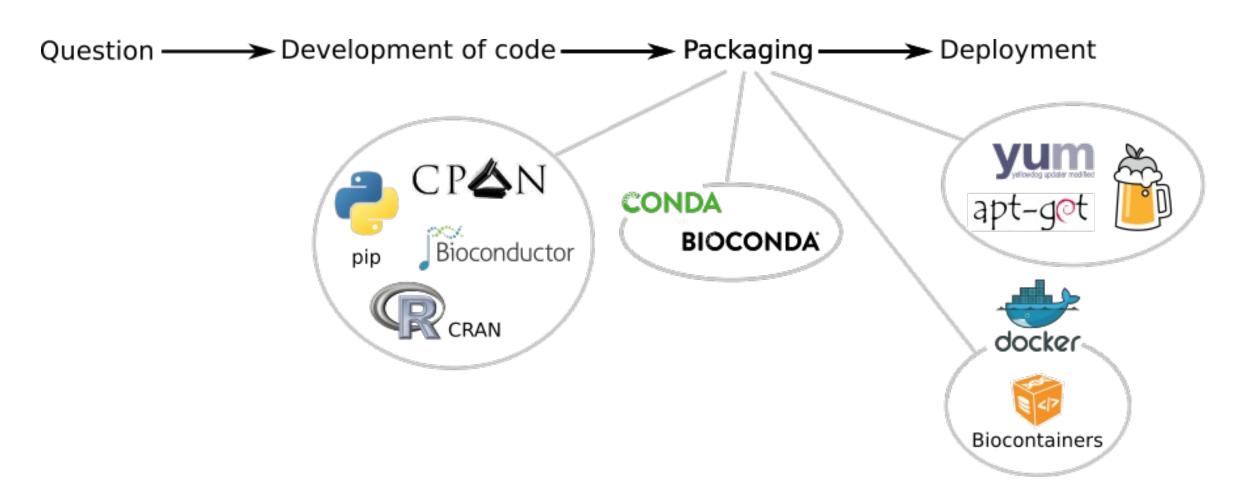
Generation of massive amounts of diverse data

→ Need of bioinformatic solutions and tools

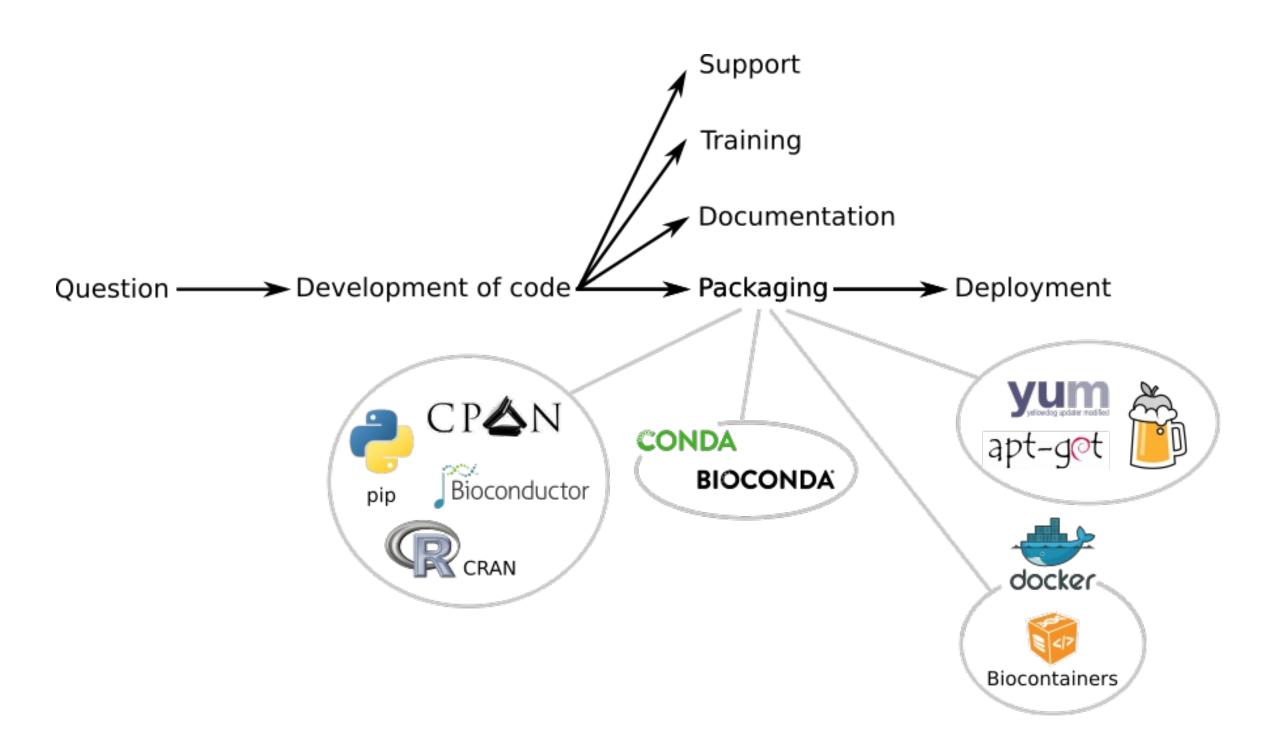
How to develop a tool?

Question — Development of code — Deployment

How to develop a tool?



How to develop a tool? The Golden Path



But...

Who has never dealed with missing tool dependencies? Or an older version of a tool that could not be installed?

Major issues in deployment and sustainability

Deployment issue

Wanted

A package manager

- Programming language agnostic
- OS independent
- No root privileges needed
- Management of multiple version
- HPC and Cloud compatible

A great solution CONDA

- Open Source package manager
- Independent of any programming language and OS
- Fast, robust and easy package installation

\$ conda install deeptools

Installation and management of multiple versions

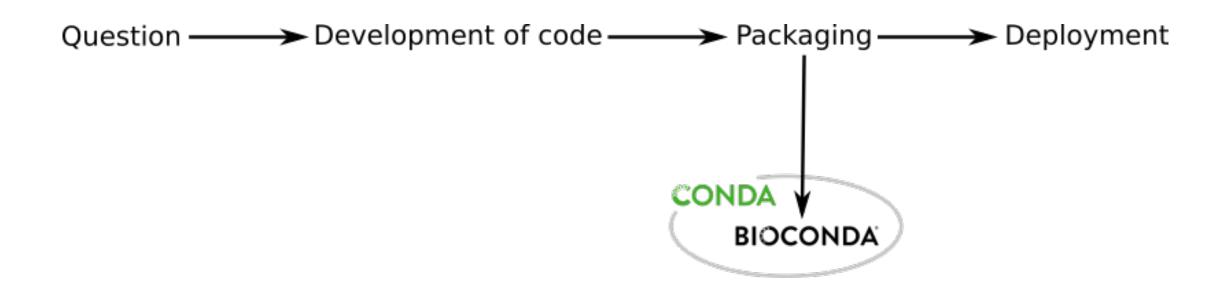
A Conda package

```
name: deeptools
 vension: '2.4.8'
source:
fn: deepTools-2.4.8.tar.gz
 url: https://pypi.python.org/packages/bd/73/56827615F73941843338F4922c56c1a94745845b75db17d3b673d3657dae/deepTools-2.4.0.tar.gz
build:
number: 0
requirements:
build:
  - python
  - numpy >=1.9.0
  - scipy >=0.17.0
  - matplotlib >=1.4.0
  - pysan >=0.8.2
  - py2bit >=8.2.8
  - gcc
 run:
  - pybigsig >=0.2.3
  - numpy >=1.9.0
                                                                                   #1/bin/bash
  - scipy >=0.17.0
  - matplotlib >=1.4.8
  - pysan >=8.8.2
                                                                                   $PYTHON setup.py install
  - py201t >-0.2.0
test:
                                                                                            build.sh
imports:
commands:
  - bamCompare --version
about:
home: https://github.com/fidelram/deepTools
 summary: A set of user-friendly tools for normalization and visualization of deep-sequencing data
                                              meta.yaml
```

→ Easy to write!

Bioinformatics tools

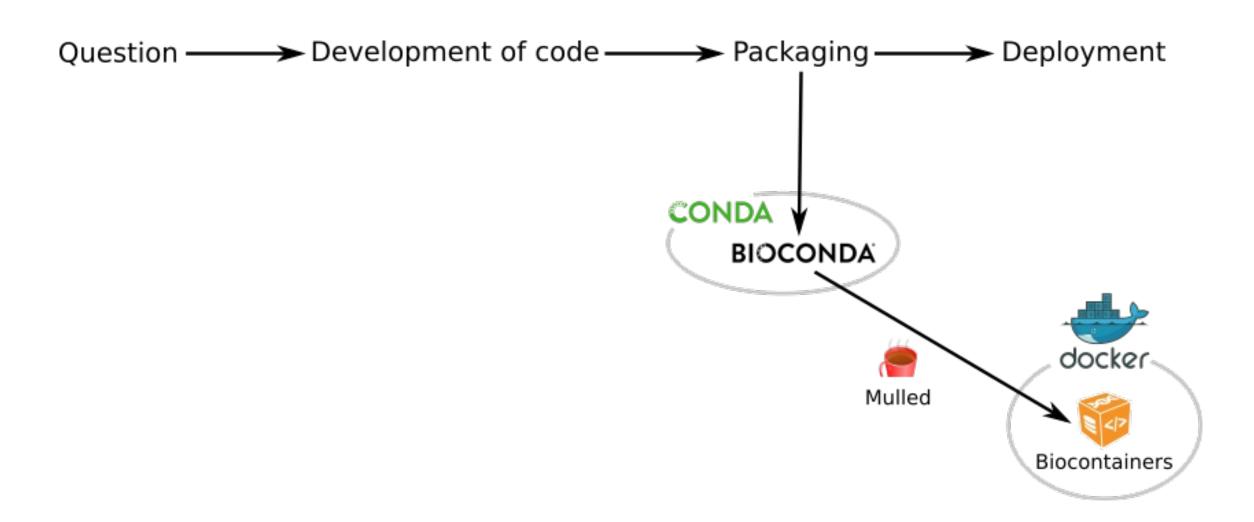
BIOCONDA



- More than 1,700 bioinformatic packages
- Big, fast-growing and newcomer-friendly community

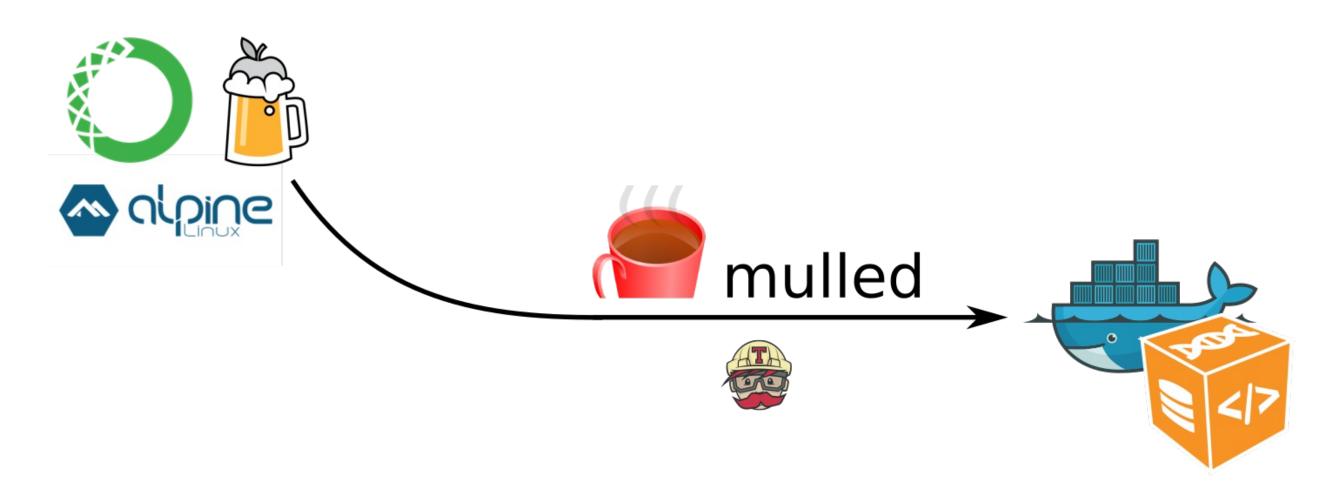
Want to have a Docker image?



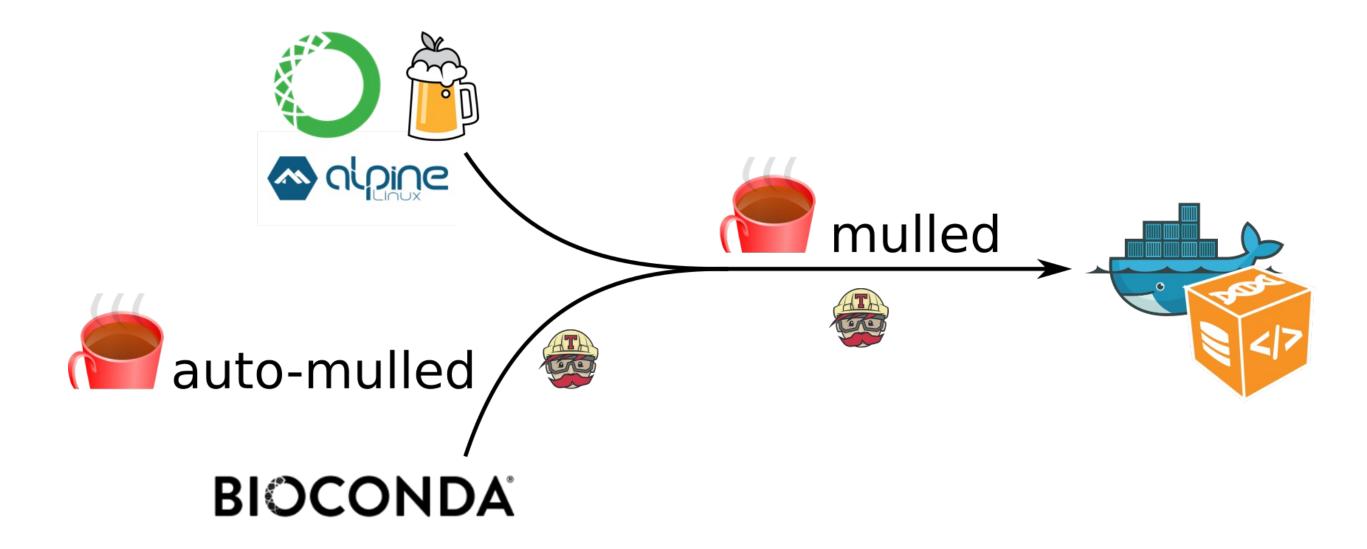


Mulled

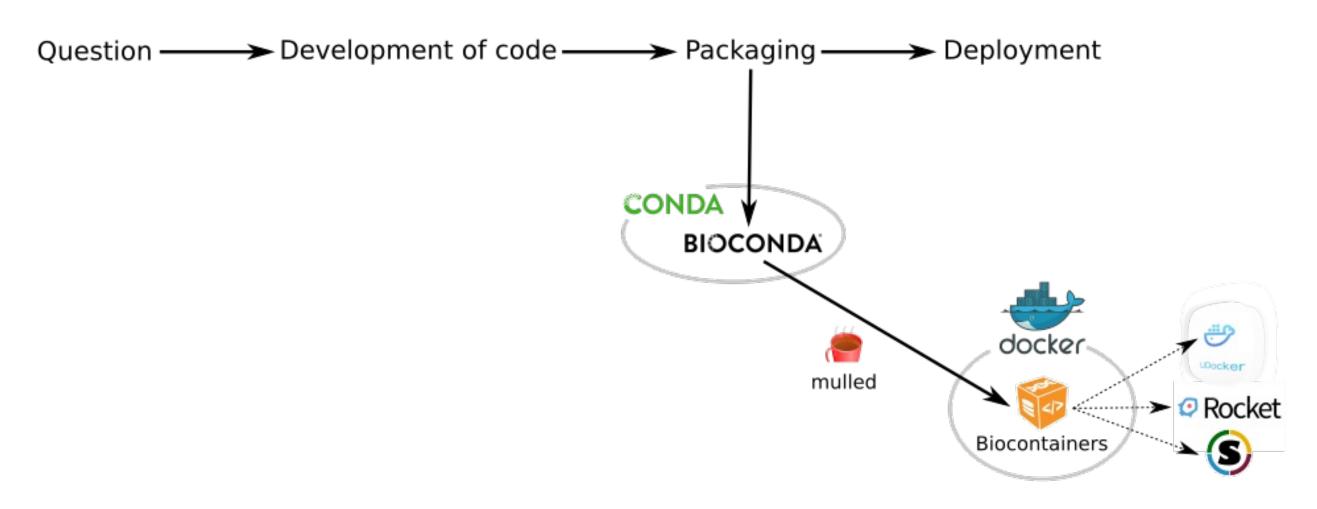
A layer donning approach to build Docker containers (without Dockerfile)



Mulled



Deployment issue → ✓ Fixed!



Sustainability issue

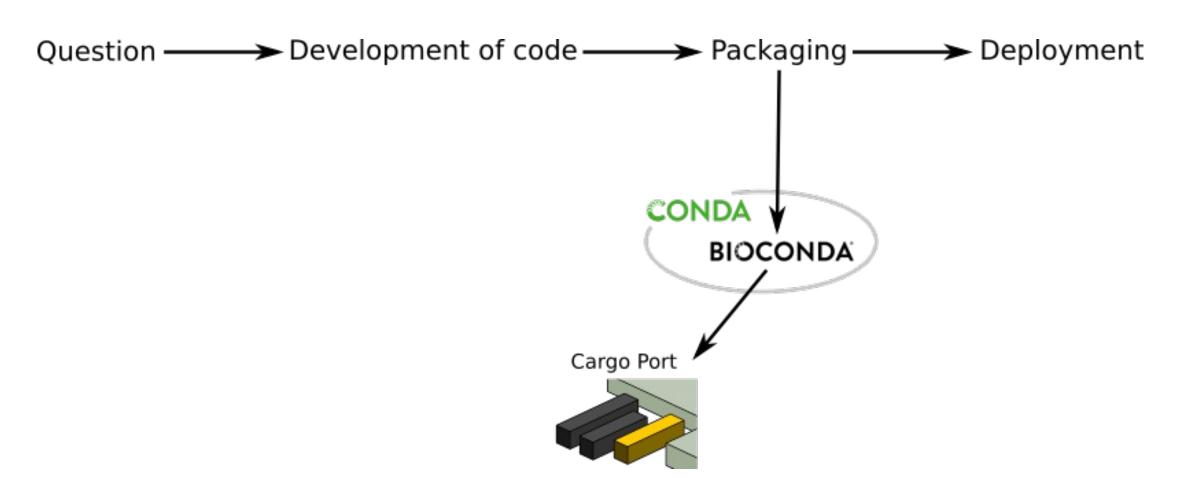
Wanted

- Accessibility of all packages
- Cache of all versions of the packages
- Permanent access

A solution Cargo Port

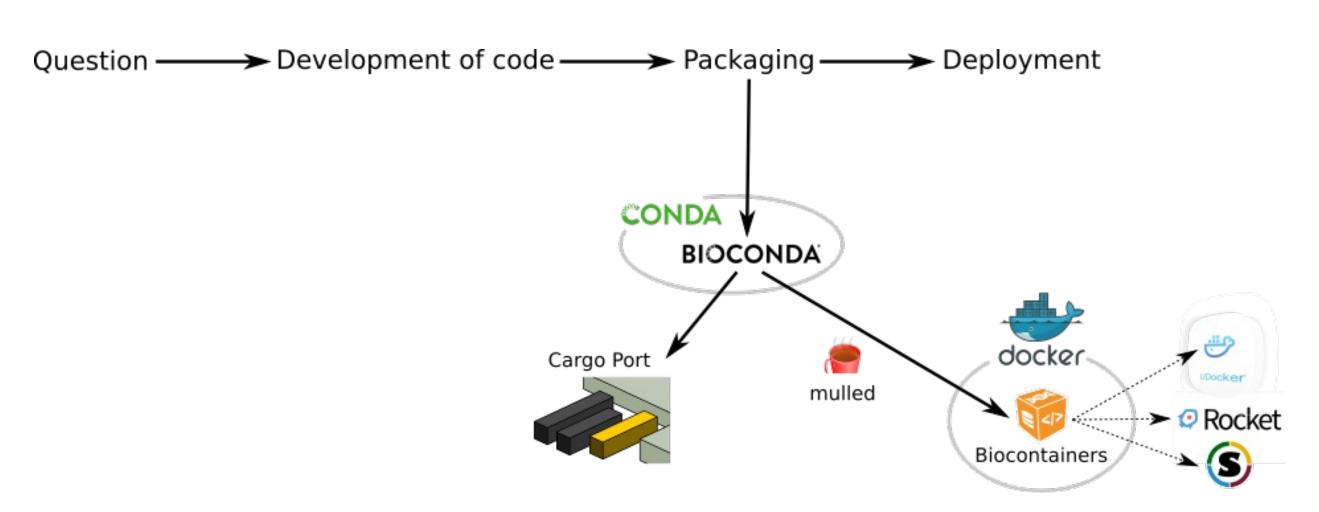
- A public repository, kindly hosted by the Galaxy Project
- Stable and long term storage
- Daily archives of the BioConda packages

Sustainability issue → ✓ Fixed!



Tool deployment and sustainability in bioinformatics





Thank You! Questions?

