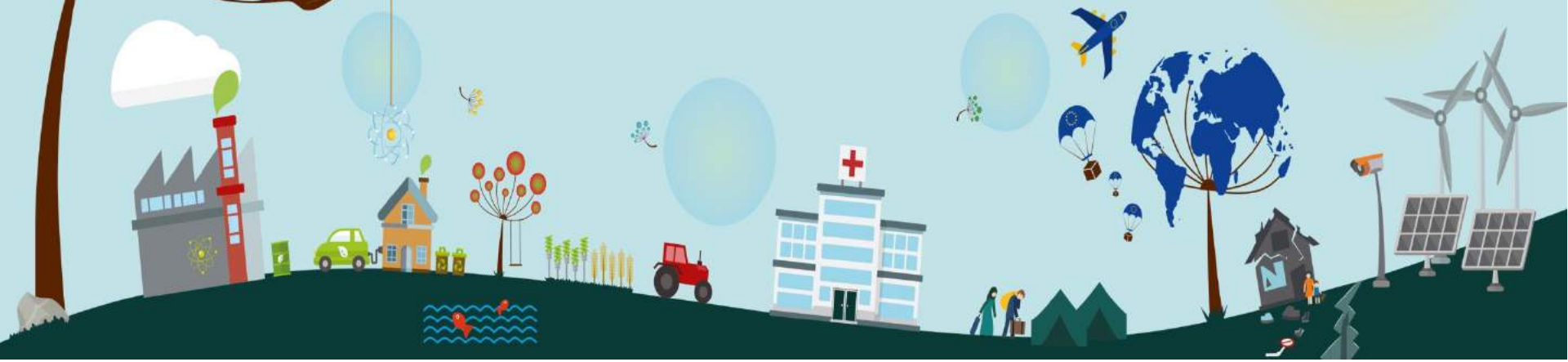


The European Commission's science and knowledge service

Joint Research Centre

Nextflow for supporting the European Reference
Laboratory for GM Food and Feed...and not only



JRC's Mission

***“As the science and knowledge service
of the Commission our mission is to support
EU policies with independent evidence
throughout the whole policy cycle*”**



Antonio Puertas Gallardo & Mauro Petrillo
JRC.F.7 Knowledge for Health and Consumer Safety Unit
CRG Barcelona

JRC sites

Headquarters in Brussels
and research facilities located
in **5 Member States:**

- Belgium (Geel)
- Germany (Karlsruhe)
- Italy (Ispra)
- The Netherlands (Petten)
- Spain (Seville)

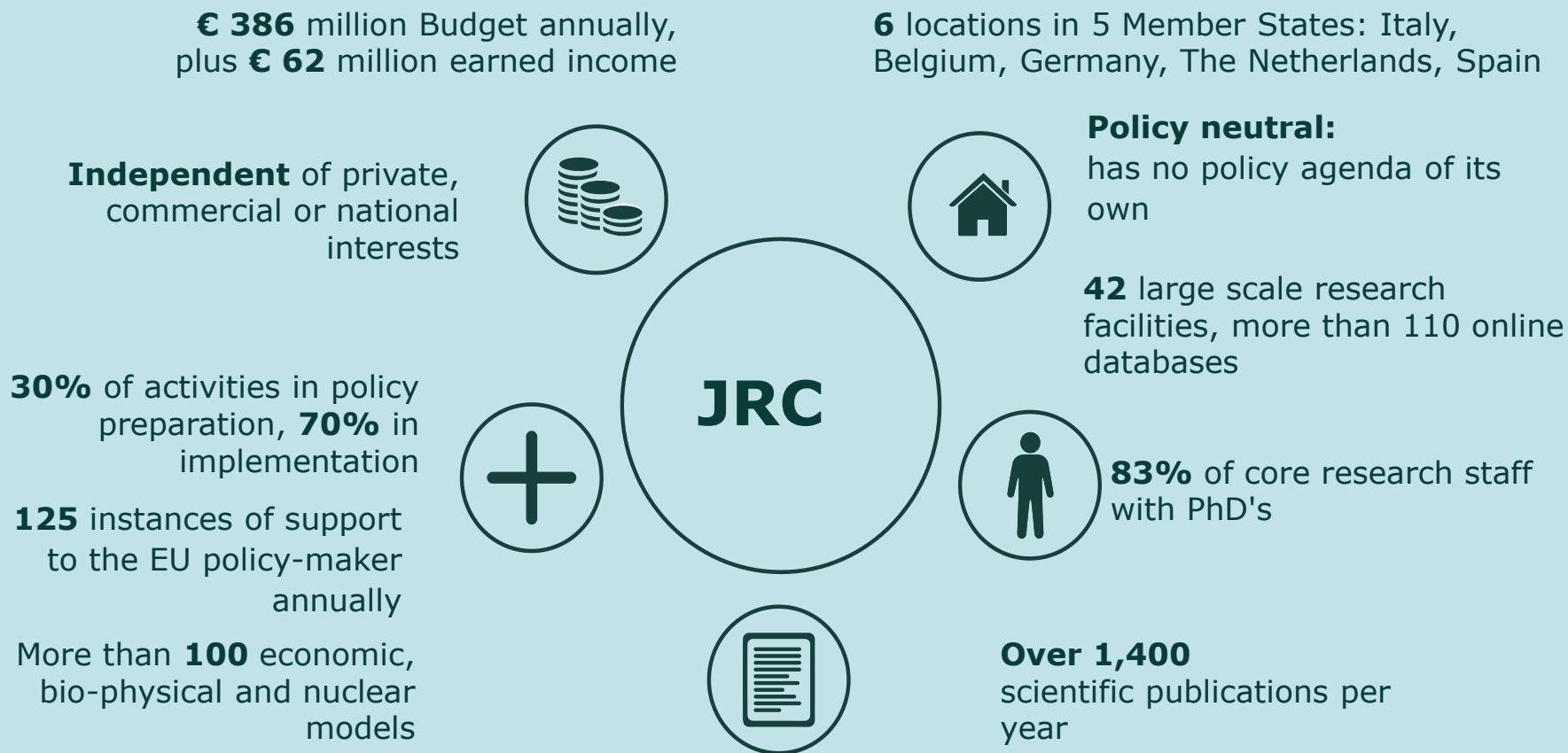


Keeping the pace with the development of research

- Fully **policy-relevant** and world class **knowledge production**
- Priorities driven **knowledge and competence management**
- One JRC – **anticipating** emerging issues, **understanding** complexities and **bridging** silos
- **Addressing challenges of research** (information deluge, multidisciplinary, integrity, reproducibility)



JRC Role: facts & figures



JRC role in support to EU GMO legal frame

EURL GMFF (mandate according to Reg.
EC No 1829/2003)

Support to the GMO authorisation process

"Validation of methods for detection,
Identification and quantification of
GM events as mandatory requirement
for the authorisation process"



MetScan

Is a Bioinformatics Pipeline to support EURL GMFF for in-silico analysis of GMO detection methods

Submission

Processing^{1,2}

Report



JOINT RESEARCH CENTRE
European Union Reference Laboratory for GM Food and Feed

MetScanUI Application Requests Log out panel

New MetScan Request

* Name:

Comments:

* Password:



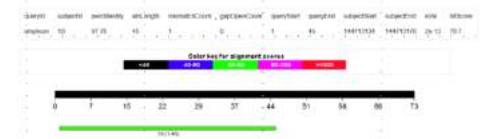
Dossier (EURL-VL-04-17) provided by Syngenta (2 x primer2)

Analyses results

Plant genomes BLAST

hits found:

Zea_mays.AGPv3.27.dna.toplevel



1 Towards Plant Species Identification in Complex Samples: A Bioinformatics Pipeline for the Identification of Novel Nuclear Barcode Candidate
<https://doi.org/10.1371/journal.pone.0147692>

2 Novel nuclear barcode regions for the identification of flatfish species
<https://doi.org/10.1016/j.foodcont.2017.04.009>

MetScan: How does it work ?



Amplicon generation bash scripts running on HPC cluster nodes



Blast ncbi-blast-2.2.31+ installed
And running on HPC cluster nodes

Problems:

- Dependency on the Hardware
- Software versioning (very fast pace)
- Sys Admin burden (increase with the size of the HPC cluster)
- Prone to human errors
- Long time to test new versions
- ...



e-PCR cmdline tool version 2.3.12
running on HPC cluster nodes

MetScan

nextflow

Advantages:

- Independency of hardware
- No Software versioning
- No Sys Admin burden
- No Prone to human errors
- Less hands-on intervention

```
ATCATCATGCTGCTAGGTAACCCGATTAACCTTACAGGATCGGATTCATTGGAATCGGA
```



docker prova_amplicon

```
tcaggagtgaagctgcagacctttgcggtgaggtgcacagctcatac
tcaggagtgaagctgcagacctttgcggtgaggtgcacagctcatac
```

```
ccaagagtgaagcaatgcaagatttattgcaagagtgaagaact
ccaagagtgaagcaatgcaagatttattgcaaga-----act
```

```
atggaagggaacccattgggtgacctgctgctcaggcagtcta
gtggaagggaacccagtggtgacctgctgctgggtgggcagcctt
```

```
atctgctccaccacacatctgctgattaggtcc
atctgctccaccacacatctgctgattaggtcc
```



docker prova_blast



docker prova_epcr

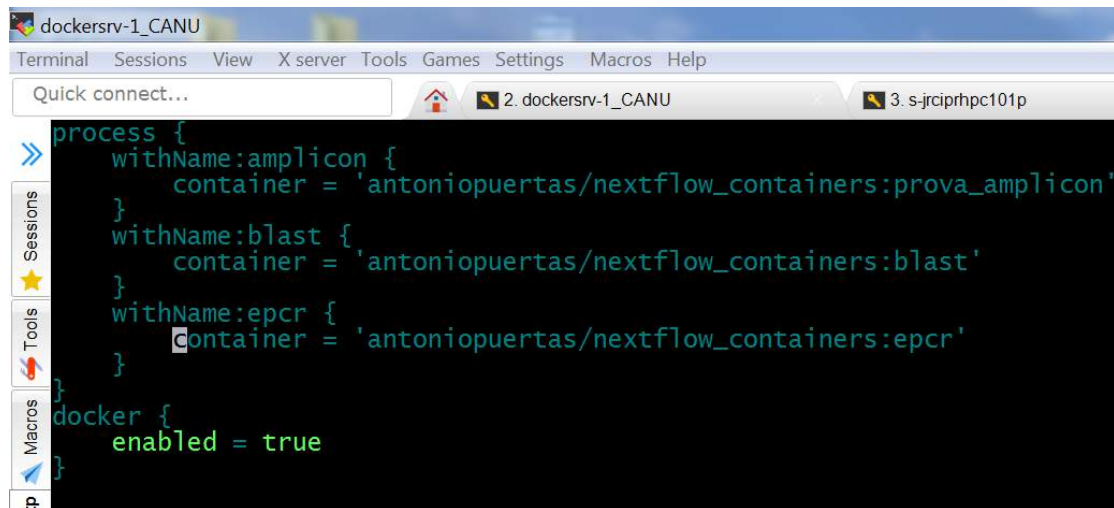
Nextflow: Docker & Nextflow config file

Built 3 dockers: amplicon, blast and e-pcr with the same versions of MetScan software

```
[root@dockersrv-1 scripts]# docker images
```

REPOSITORY	TAG	IMAGE ID	CREATED	SIZE
antoniopuertas/nextflow_containers	prova_amplicon	49ec99a7b5cf	7 days ago	200 MB
prova_amplicon	latest	49ec99a7b5cf	7 days ago	200 MB
antoniopuertas/nextflow_containers	blast	df850fa24a80	7 weeks ago	774 MB
prova_blast	latest	df850fa24a80	7 weeks ago	774 MB
antoniopuertas/nextflow_containers	e-pcr	bfb17cd0cd4d	7 weeks ago	208 MB
prova_epcr	latest	bfb17cd0cd4d	7 weeks ago	208 MB
docker.io/centos	latest	5182e96772bf	2 months ago	200 MB
docker.io/nextflow/examples	latest	318c7d0648b8	24 months ago	979 MB

Prepare the following configuration file, which call to the docker repository



The screenshot shows a terminal window titled 'dockersrv-1_CANU' with a menu bar (Terminal, Sessions, View, X server, Tools, Games, Settings, Macros, Help) and a tab bar (Quick connect..., 2. dockersrv-1_CANU, 3. sjrciprhc101p). The terminal displays a Nextflow configuration file with the following content:

```
process {
  withName:amplicon {
    container = 'antoniopuertas/nextflow_containers:prova_amplicon'
  }
  withName:blast {
    container = 'antoniopuertas/nextflow_containers:blast'
  }
  withName:epcr {
    container = 'antoniopuertas/nextflow_containers:epcr'
  }
}
docker {
  enabled = true
}
```

Nextflow: Launch script

Launch nextflow -C running the following dir_proc.nf file and calling the dockers in the repository

```
process amplicon {
  container 'prova_amplicon:latest'
  containerOptions '--volume /nfs/data/:/data'
  ... /bin/bash /amplycon/amplycon.sh /data/source_sequences
}
process blast {
  container 'prova_blast:latest'
  containerOptions '--volume /nfs/data/:/dbs'
  output: file 'output.txt'

  blastn -query /dbs/myseq -db /dbs/_EXCHANGE/gmobiongs/from/metscandbs/ensembl_plants/last/Ostreococcus_lucimarinus.GCA_000092065.1.27.dna.toplevel.fa > output.txt && cat output.txt > /dbs/results/blast/ostreococcus_lucimarinus.GCA_000092065.1.27.dna.toplevel.fa.blast.out
  blastn -query /dbs/myseq -db /dbs/_EXCHANGE/gmobiongs/from/metscandbs/phytozome_plants/last/Csativus_122_v1.fa > output.txt && cat output.txt > /dbs/results/blast/Csativus_122_v1.fa.blast.out
  blastn -query /dbs/myseq -db /dbs/_EXCHANGE/gmobiongs/from/metscandbs/phytozome_plants/last/Athaliana_167_TAIR9.fa > output.txt && cat output.txt > /dbs/results/blast/Athaliana_167_TAIR9.fa.blast.out
}
process epcr {
  container 'prova_epcr:latest'
  containerOptions '--volume /nfs/data/:/dbs'

  e-PCR -n 2 -g 2 -f 3 -t 4 -m 1000 -d 20-1000 -o /dbs/results/e-pcr/ostreococcus_lucimarinus.GCA_000092065.1.27.dna.toplevel.fa.pcr.out /dbs/myamplicon /dbs/_EXCHANGE/gmobiongs/from/metscandbs/ensembl_plants/last/Ostreococcus_lucimarinus.GCA_000092065.1.27.dna.toplevel.fa
  e-PCR -n 2 -g 2 -f 3 -t 4 -m 1000 -d 20-1000 -o /dbs/results/e-pcr/Cyanidioschyzon_merolae.ASM9120v1.27.dna.toplevel.fa.pcr.out /dbs/myamplicon /dbs/_EXCHANGE/gmobiongs/from/metscandbs/ensembl_plants/last/Cyanidioschyzon_merolae.ASM9120v1.27.dna.toplevel.fa
  e-PCR -n 2 -g 2 -f 3 -t 4 -m 1000 -d 20-1000 -o /dbs/results/e-pcr/Athaliana_167_TAIR9.fa.pcr.out /dbs/myamplicon /dbs/_EXCHANGE/gmobiongs/from/metscandbs/phytozome_plants/last/Athaliana_167_TAIR9.fa
}
```

Nextflow: Next steps

We believed Nextflow could give us support in answering to the following questions

AMR workshop held at the JRC in 2018¹

- How to generate, distribute and update pipelines ?
- How to evaluate and demonstrate pipelines accuracy ?
- Who is the institution entitled to do that ?

Regulatory bioinformatics²

- *in NGS*: "One of the key issues in applying NGS/bioinformatics in regulatory decision making is ensuring the accuracy and quality of information"
- *In computational Toxicology*:
- ... "Validation of different software packages, however, remains an issue"..
- standardization and application of different software and computational methods must be resolved.

¹ The challenges of designing a benchmark strategy for bioinformatics pipelines in the identification of antimicrobial resistance determinants using next generation sequencing technologies
[doi:10.12688/f1000research.14509](https://doi.org/10.12688/f1000research.14509)

² Regulatory bioinformatics for food and drug safety <https://doi.org/10.1016/j.vrtph.2016.05.021>



Questions?

You can find me at: antonio.puertas-gallardo@ec.europa.eu