

## JOB DESCRIPTION & PERSON SPECIFICATION

**1. Job title:** Machine learning and genomics informatician      **Centre:** CCR 30      **Group:** Specialist  
**Location:** HQ

**Reports to (Job title):** Richard Harrison      **Line Manages (Job titles):**

### 2. Main purpose of role:

NIAB is the UK's fastest growing crop science organisation, with rapidly expanding research capabilities in plant genetics, agronomy, farming systems and data science, the largest national field trials capability, and strong research links with industry, Government and academia. With headquarters in Cambridge, and regional offices across the country, employing around 400 people across the UK, NIAB provides scientific research, technical services and practical advice to improve the yield, efficiency and resilience of crop production across the arable, forage and horticulture sectors. NIAB (Cambridge, UK) is an Independent Research Organisation eligible for competitive BBSRC funding calls. Research departments at NIAB's sites in Cambridge and East Malling have more than 40 current BBSRC grants.

Pathogens are a constant threat to society and in the past have led to profound changes to our civilisation. Despite many advances in disease control, plant pathogens remain of great concern to the security of our food production, but the predication of the host range and pathogenic 'potential' of microbes remains elusive.

Working with the data sciences and pathology department, NIAB is seeking a talented and motivated postdoctoral informatician to investigate machine-learning approaches to predict host range from genome sequence using the bacterial plant pathogen *Pseudomonas* as a model, building on previous work- e.g. Hulin et al.

(<https://doi.org/10.1111/ppa.13189>, <https://doi.org/10.1111/nph.15182>, <https://doi.org/10.1111/ppa.12834>).

Candidates should also have a PhD in a relevant subject and a publication track record in a relevant discipline with a strong background in informatics and statistics, ideally having used machine learning approaches previously. Although desirable, prior experience in biology or the analysis of high-throughput sequencing datasets and comparative genomics is not required. However the candidate should have excellent skills in scientific programming/scripting and a strong desire to learn. You will be guided by senior data scientists in the Data Sciences department at NIAB. NIAB will also provide unique computational infrastructures designed for big-data analytics, deep learning and algorithmic development, through which your knowledge and skills in AI and data sciences will be strengthened.

Using supervised machine learning and deep learning (so-called AI) approaches the post holder will deploy a variety of methods to identify signatures of host adaptation present in the genome of *Pseudomonas* which will be experimentally verified.

Prior experience in the use of various classification and regression techniques eg:

- Support vector machines
- Random forests
- Convolutional Neural Networks

as well as experience with high level languages and environments Python, Tensorflow, R etc.

They will work collaboratively with wet-lab scientists producing data and participate in the research activities of the institution as a whole. They will be expected to contribute their own innovative ideas, resulting in potential independent projects. The postholder will be expected to write scientific research papers and present results at conferences. They must be comfortable working at the interface of disciplines in a constructive and collaborative

manner.

NIAB is committed to equal opportunity and gender balance and is also recognised by the commitment to work with international partners as evidenced by the number of Newton and GCRF current projects on global staple crops.

**3. Financial authority/responsibility** (e.g. delegated budget, authorisation level, approx value of contracts etc):

-

**4. Key relationships** (external and internal): Postdoctoral researchers, industry stakeholders, group leaders

Tasks/responsibilities (in order of priority)	Approx % of time
Developing pipelines for machine learning/ deep learning analysis	40
Developing methods for comparative genomics / data processing	40
Writing papers	20

**6. Working conditions :**

Role will include working in the office and lab. Flexible, agile working is required. Travelling where necessary to represent NIAB nationally and internationally is of great importance as is the requirement to work to short deadlines and with variable workloads.

**7. PERSON SPECIFICATION****Education/Qualifications:**

Essential:	Desirable:
<p>A PhD in bioinformatics or other scientific discipline which has included activities in bioinformatics, data analysis or computer sciences</p> <p>A degree in a Biological, Mathematical and/or Computational science</p>	

**Experience:**

Essential:	Desirable:
<p>At least 2 years of working experience in developing software in informatics or Computational Biology</p> <p>Strong background in machine learning/ deep learning or requisite mathematical skills to learn</p>	<ul style="list-style-type: none"> <li>• Demonstrable experience of leading the development of novel machine learning tools and methodologies</li> <li>• Extensive experience in artificial intelligence/machine learning field</li> </ul>

**Specialist Training:**

Essential:	Desirable:
<p>Strong programming skills in Python or other appropriate language</p> <p>Familiarity with Unix command line</p>	<p>Detailed knowledge of machine learning languages, for example Tensorflow, PyTorch</p> <p>Familiarity with the use of job scheduling systems for the use of High performance computing</p> <p>Use of version control (github)</p>

**Personal Qualities (skills, behaviours and competencies)**

Essential:	Desirable:
<p>Attention to detail and ability to solve scientific problems</p> <p>Work independently or with minimal supervision</p> <p>Excellent communication and interpersonal skills</p>	<p>Ability to work effectively in a collaborative environment</p> <p>Enthusiastic to learn and help the scientific community</p>

**Date of description:** July 2020**Compiled by:** Richard Harrison