**2021 美國一年研究實習研究生甄選 (NAL)**
1-year internship in the U.S. (NAL)

2021 年 United States Department Of Agriculture 美國農業部國家農業研究局 (USDA-ARS) 與台大電資學院合作之研究實習計畫甄選碩博士生參與！實習單位為美國農業部國家農業圖書館 (National Agricultural Library, NAL) 網頁連結請見：www.nal.usda.gov。


<table>
<thead>
<tr>
<th>申請資格</th>
<th>Applicant Qualification</th>
</tr>
</thead>
<tbody>
<tr>
<td>臺大碩、博士班研究生</td>
<td>2 Master and Ph.D students in NTU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>實習期間</th>
<th>Internship Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>一年（約自2021年8月中起，前半年在台灣，後半年在美國）</td>
<td>1 year (approximately starting from mid-Aug. 2021; the first 6 months in Taiwan; the last 6 months at NAL)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>申請文件</th>
<th>Application Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 2021 NAL 申請書</td>
<td>2021 NAL Application form</td>
</tr>
<tr>
<td>• 成績單</td>
<td>Undergraduate and postgraduate official transcripts (GPA)</td>
</tr>
<tr>
<td>• 申請動機</td>
<td>Statement of Interest</td>
</tr>
<tr>
<td>• 推薦信 2 封</td>
<td>2 reference letters</td>
</tr>
<tr>
<td>• 程式編寫相關著作（論文、期刊等）</td>
<td>Papers (journal or conference) or sample examples of programming skills</td>
</tr>
<tr>
<td>• 英文能力檢定</td>
<td>English Language Certificate</td>
</tr>
</tbody>
</table>

*若無英語能力檢定成績，依然可提申請；若獲 NAL 錄取，需於辦理簽證前補交檢定證明。若英語能力經 NAL 檢定後認定須加強，學生可獲 NAL 補助，並於出國前完成要求之英文課程。

*English Language Certificate can be provided after NAL acceptance before visa application. If NAL considers the further improvement in English proficiency necessary, the NAL-accepted student needs to complete required English courses before departure to US.*

<table>
<thead>
<tr>
<th>審查時程</th>
<th>Application Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 申請截止日期：2021年3月26日</td>
<td>Application Deadline: March 26, 2021</td>
</tr>
<tr>
<td>• 最終錄取結果公告：2021年4月20日</td>
<td>NAL final Announce acceptance: April 20, 2021</td>
</tr>
<tr>
<td>• 確認接受截止日期：2021年4月30日</td>
<td>Acceptance deadline for students: April 30, 2021</td>
</tr>
<tr>
<td>• 簽證申請：2021年8月起</td>
<td>Visa application: Since August 2021</td>
</tr>
<tr>
<td><strong>Salary</strong></td>
<td>Departure to US: mid-February in 2022. <em>The appropriate time for visa application is estimated. Please be aware of your academic career plan.</em></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>Maximum 25000 USD annual salary (the first 6 months paid by NTU; the last 6 months paid by NAL). <em>The salary will be different according to the policy of NTU and NAL.</em></td>
</tr>
</tbody>
</table>

申請資料下載 Download information: [https://www.space.ntu.edu.tw/navigate/s/93F52061C09A475484201FBE09C57DF9QQY](https://www.space.ntu.edu.tw/navigate/s/93F52061C09A475484201FBE09C57DF9QQY)

本計畫收件截止日期為 2021年3月26日中午12:00前，有意願之申請者請於收件截止日前，將申請書連同所有資料以電子檔及紙本方式交至博理館202室陳貞伶小姐(cathychen@ntu.edu.tw)，由電資學院統一審查後再將決定之推薦名單送至NAL。Please submit all application documents in **electronic and hard copy form** to Ms. Cathy Chen (cathychen@ntu.edu.tw) in room 202 in Barry Lam Hall before 12 pm on March 26, 2021. The college of EECS will review all applicants and make the nomination to NAL.
The National Agricultural Library of the United States Department of Agriculture (NAL; https://www.nal.usda.gov/) is now recruiting new students to an internship program on computational bioscience and scientific big data management. Master’s and PhD-level students with experience in bioinformatics and/or software development are invited to participate on projects related to the i5k Workspace@NAL, a USDA database for arthropod genomics (https://i5k.nal.usda.gov; doi: 10.1093/nar/gku983; https://github.com/NAL-i5K). The i5k Workspace is a web resource providing dissemination, visualization, and curation tools for ‘orphaned’ insect or arthropod genome projects. This is an actively growing project, and there are many possible development opportunities. Projects focus on application/tool development and biocuration services for our research community.

Due to the COVID-19 pandemic, the first 6 months of this internship will take place in Taiwan (August 2021 – February 2022). While in Taiwan, interns will work with supervisors at the NAL via videoconference. The last 6 months on the internship will take place at the National Agricultural Library in Beltsville, MD, USA.

Desired Skills and Related Experience

Exact skills, knowledge and experience will vary based on the projects selected. The list below contains desired, but not required skills.

- **Bioinformatics**
  - Basic knowledge of genome assemblies and gene prediction
  - Gene prediction and functional annotation
  - Biocuration
  - Unix command line (including usage of high-performance computing systems)
  - Natural Language Processing

- **Data visualization**
  - Google APIs and SDKs (Analytics, Charts, Maps)
  - Other JavaScript libraries (such as jQuery, Flot/jQuery, D3.js, Processing.js, etc.)
  - User experience/User interface

- **Programming languages**
  - Object oriented programming (Java, Python)
  - Dynamic scripting languages (Perl, PHP)
  - Statistical programming (R)

- **Database design and programming**
  - Entity-relation modeling and normalization
  - Performance tuning
  - Data warehouses, business intelligence, and data mining

- **Open source software**
  - LAMP (Linux/Apache/MySQL/PHP or Python, etc.) software stack
  - Web framework (Django, Google Web Toolkit)
  - Middleware (JBoss, Tomcat)
Potential Opportunities

Interns also have the opportunity to develop their own project ideas related to the i5k Workspace.

- **Develop workflows for data processing.** The i5k Workspace accepts data from many genomes, and will increase the number of genomes hosted in the future. In this internship project, the intern will work with the i5k Workspace team to develop and improve a pipeline in a workflow language (e.g. CWL). This workflow will expedite adding new content to the i5k Workspace’s applications. In the process, the intern will add new content to our database, and/or improve existing content.

- **Bioinformatics application development.** The i5k Workspace is exploring new services for our users. Interns will develop or refine workflows to automatically generate new data types for our genome browsers, such as mapped RNA-Seq reads, methylation data, or lateral gene transfers.

- **Text mining.** Interns with experience in python and Natural Language Processing (NLP) will have the opportunity to develop text mining workflows of insect genetics abstracts.

- **Develop tools and integrate data to support comparative analyses of arthropod genomic data.** Homology data are available for many i5k species. This information can be added to gene pages, visualized in browsers, and represented in many other ways to support comparative analysis.

- **Biocuration.** The i5k Workspace enables manual curation of gene models by the i5k community. We are seeking an intern interested in manual curation and biocuration to: identify workflows for new annotators; create manual curation tutorials; and interact with the i5k manual curation community to identify curation needs.

- **Tripal development.** Interns with experience in PHP can contribute towards developing modules for the Tripal software ([http://tripal.info/](http://tripal.info/)).

- **Improve i5k Workspace systems, tools and approaches.** Interns will review computational approaches in our existing software and develop improved algorithms. For example, interns can review our coordinate conversion workflow for genome assemblies and research improved algorithms to optimize for memory usage, compute efficiency or storage usage.

- **Create tests to improve i5k Workspace system functionality and efficiency.** Interns will develop or improve uptime, build, coverage, and functional tests for the genomics workspace ([https://github.com/NAL-i5K/genomics-workspace](https://github.com/NAL-i5K/genomics-workspace)) with the potential to extend to other applications.

Examples of past and current projects:

- Development of an RNA-Seq mapping pipeline: [https://github.com/NAL-i5K/NAL_RNA_seq_annotation_pipeline](https://github.com/NAL-i5K/NAL_RNA_seq_annotation_pipeline)
- Text mining of insect genetics abstracts
- Containerization of the genomics-workspace application. [https://github.com/NAL-i5K/genomics-workspace_docker](https://github.com/NAL-i5K/genomics-workspace_docker)
• Workflow development for genomics content processing using the Common Workflow Language. [https://github.com/NAL-i5K/Organism_Onboarding](https://github.com/NAL-i5K/Organism_Onboarding)

• Python programs for updating GFF3 coordinates to new assembly versions: ([https://github.com/NAL-i5K/remap-gff3](https://github.com/NAL-i5K/remap-gff3))

• Reduce unnecessary/redundant code, implement build tests, incorporate coverage, unit and functional tests ([https://github.com/NAL-i5K/genomics-workspace](https://github.com/NAL-i5K/genomics-workspace))

• Development of a novel BLAST user-interface ([https://github.com/hotdogee/django-blast](https://github.com/hotdogee/django-blast))

• Development of a single-sign on system to the i5k Workspace using Django

• Programs to check the quality of the GFF3 format ([https://github.com/hotdogee/gff3-py](https://github.com/hotdogee/gff3-py))

• Program to compare two GFF3 files ([https://github.com/chienyuehlee/gff-cmp-cat](https://github.com/chienyuehlee/gff-cmp-cat))

• A ‘toolkit’ for the GFF3 annotation format, including programs to QC and merge two gff3 files for Official Gene Set generation ([https://github.com/NAL-i5K/GFF3toolkit/](https://github.com/NAL-i5K/GFF3toolkit/))

• Implementation of HMMER and ClustalW web services ([https://github.com/NAL-i5K/genomics-workspace](https://github.com/NAL-i5K/genomics-workspace))

• Application stress testing (our internal tests were incorporated into the Apollo codebase)

• A program to convert several file types between assembly coordinate systems: [https://github.com/NAL-i5K/coordinates_conversion](https://github.com/NAL-i5K/coordinates_conversion)

Publications of past interns


- McKenna, D.D., et al. (2016). Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *Genome Biol.* 17, 227. Interns *Chien-Yueh Lee* and *Han Lin* are co-authors.

- Poelchau, M., et. al. (2014). The i5k Workspace@NAL--enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Res.* (43), D714-D719. Interns *Chien-Yueh Lee*, *Han Lin* and *Jun-Wei Lin* are co-authors.

- 計畫承辦人： 陳貞伶 02-3366-3697
- [cathychen@ntu.edu.tw](mailto:cathychen@ntu.edu.tw)
- 博理館 202 辦公室