

2023 美國半年研究實習研究生甄選(NAL) Half-year internship in the U.S. (NAL)

2023 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徵選碩博士生參與！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。

2023 internship at National Agricultural Library (NAL), United States Department Of Agriculture. More info about NAL at www.nal.usda.gov.

申請資格 Applicant Qualification	臺大碩、博士班研究生 2 Master and Ph.D students in NTU
實習期間 Internship Period	半年 (約自 2023 年 2 月到 2023 年 8 月) Half year (approximately starting from Feb. 2023 to Aug. 2023.)
申請文件 Application Documents	<ul style="list-style-type: none"> • 2023 NAL 申請書 2023 NAL Application form • 成績單 Undergraduate and postgraduate official transcripts (GPA) • 申請動機 Statement of Interest • 推薦信 2 封 2 reference letters • 程式編寫相關著作 (論文、期刊等) Papers (journal or conference) or sample examples of programming skills • 英文能力檢定 English Language Certificate <p><u>*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。</u></p> <p><u>* 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.</u></p>
審查時程 Application Timeline	<ul style="list-style-type: none"> • 申請截止日期：2022 年 10 月 12 日 Application Deadline：October 12, 2022 • 最終錄取結果公告：預定 2022 年 12 月 15 日前 (由 NAL 決定) NAL final Announce acceptance：Before December 15, 2022 (Depends on NAL) • 確認接受截止日期：預定 2022 年 12 月 15 日前 (由 NAL 決定) Acceptance deadline for students：Before December 15, 2022 (Depends on NAL)

	<ul style="list-style-type: none"> 簽證申請 (NAL 參訪用) : NAL 另行通知 Visa application (for a site visit at NAL) : Until further notice by NAL 出發前往美國參訪 (預定參訪 5 日內) : NAL 另行通知 Departure to US for a site visit (within 5 days) : Until further notice by NAL <p><u>*申請簽證相關時程為預估，請注意個人學業生涯規劃</u> <u>*The appropriate time for visa application is estimated. Please be aware of your academic career plan.</u></p>
薪資 Salary	<p>月薪約 1,500 美金內 (依據國立臺灣大學之標準給薪) Maximum \$1,500 (USD) salary/per month</p> <p><u>*The amount of salary will be paid according to the policy of NTU.</u></p>
福利 Benefits	<ul style="list-style-type: none"> 臺美往返機票，補助上限為每位學生 2,000 美金 Flight tickets between Taiwan and U.S. under maximum \$2,000 (USD) per student 參訪津貼 (食宿等) : 每位學生 1,706 美金 Visiting allowance : \$1,706 (USD)
實習報告 Internship Report	<ul style="list-style-type: none"> 500 字實習報告 Evaluation of the Internship experience with 500 words 實習照片 5 張 (電子檔) 5 photos (in electronic form) of the internship period
<p>申請資料下載 Download information : https://www.space.ntu.edu.tw/navigate/s/732DCF67876A4A868E9FED3BD94BEC60QQY</p>	

本計畫收件截止日期為 **2022 年 10 月 12 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 October 12, 2022**.
The college of EECS will review all applicants and make the nomination to NAL.

For Internship Announcement

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://doi.org/10.1093/nar/gku983); <https://github.com/NAL-i5K>). Projects focus on application/tool development and biocuration services for our research community.

The i5k Workspace is a web resource providing dissemination, visualization, and curation tools for insect or arthropod genome projects. This is an actively growing project, and there are many possible development opportunities. The i5k Workspace actively supports several large-scale genome sequencing projects, including the i5k initiative (<http://i5k.github.io>); the Ag100Pest initiative (<http://i5k.github.io/ag100pest>); and the Beenome 100 Project (<https://www.ars.usda.gov/news-events/news/research-news/2022/collecting-a-library-of-bee-genomes/>). These initiatives are part of the Earth BioGenome project (<https://www.earthbiogenome.org/>), an ambitious initiative to provide high-quality genome assemblies of all extant eukaryotes.

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)
- **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)

- Content management (Drupal, Tripal)
- Version control (Git, Github/GitLab)
- Continuous integration and testing (Travis, Jenkins CI, Codacy, etc)

Potential Opportunities

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

- Test new data visualization tools, such as Jbrowse 2. The i5k Workspace currently supports over 90 different organisms. One of our key principles is to provide useful ways for users to view and interact with data. Jbrowse 2 now supports new ways of viewing and combining data sets and data sources, and an intern interested in this project would be able to test the new functionality, with a specific focus on cross-species synteny viewing.
- Rescuing microbial DNA sequence in genome assemblies. Many genome assemblies have some level of microbial DNA that is included, and is often not identified, or if it is identified, it is usually removed and discarded. This project idea could include identifying, removing and saving microbial sequence in our genome assemblies, and researching and providing resources to make this data usable by our end users.
- 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.
- Container development/ cloud optimization. We have developed containers for some of our applications and are working to complete this for all of our resources. We have recently migrated our servers to the cloud and are interested in optimizing our software for cloud based environments.
- Improve Standard Operating Procedure documentation to reflect best practices.

Examples of past and current projects:

- Containerization of the genomics-workspace application. 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
- Python programs for updating GFF3 coordinates to new assembly versions: (<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>)
- Development of a novel BLAST user-interface (<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>)
- Program to compare two GFF3 files (<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/>)
- Implementation of HMMER and ClustalW web services (<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

Publications of past interns

- Chen, M.-J. M., Lin, H., Chiang, L., Childers, C.P., and Poelchau, M.F. Methods in Molecular Biology. 2019. The GFF3toolkit – QC and Merge Pipeline for Genome Annotation. Pp. 75-87; doi:10.1007/978-1-4939-8775-7_7. Interns Mei-Ju May Chen, Yu-Yu Lin and Li-Mei Chiang are co-authors.
 - Panfilio, K.A., et al. 2019. Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biology* 20(64); doi: 10.1186/s13059-019-1660-0. Interns Mei-Ju May Chen and Chien-Yueh Lee are co-authors.
 - Armisen, David, et al. 2018. The genome of the water strider *Gerris buenoi* reveals expansions of gene repertoires associated with adaptations to life on the water. *BMC genomics* 19.1: 832. Intern Li-Mei Chiang is a co-author.
 - Poynton, H.C., et al. 2018. The Toxicogenome of *Hyalella azteca*: A Model for Sediment Ecotoxicology and Evolutionary Toxicology. *Environmental Science & Technology* 52:10; doi: 10.1021/acs.est.8b00837. Interns Mei-Ju May Chen and Yu-Yu Lin are co-authors.
 - Schoville, S., et al. 2018. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Scientific Reports*, 8; doi: 10.1038/s41598-018-20154-1. Intern Mei-Ju May Chen is a co-author.
 - Poelchau, M.F., Chen, M.-J. M., Lin, Y.-Y., and Childers, C.P. 2018. Methods in Molecular Biology. Navigating the i5k Workspace@NAL – a resource for arthropod genomes. Pp. 557-577; doi:10.1007/978-1-4939-7737-6_18. Interns Mei-Ju May Chen and Yu-Yu Lin are co-authors.
 - Saha, S., et al. (2017). Improved annotation of the insect vector of Citrus greening disease: Biocuration by a diverse genomics community. *Database*, bax032, <https://doi.org/10.1093/database/bax032>. Intern Mei-Ju May Chen is a co-author.
 - Benoit, J.B., et al. (2016). Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. *Nat Commun* 7. Interns Chien-Yueh Lee and Han Lin are co-authors.
 - 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
 - 博理館 202 辦公室