

Student project SnoekLab 2026/2027/2028

Student projects are available for microbiome research as well as plant genetics and multi-omics. We also offer collaborative projects within the UU and external at NIOO-KNAW and WUR.

Contact: Basten Snoek | l.b.snoek@uu.nl

Soil microbiome

1) Microbial function

Microbes (bacteria, archaea, and fungi) are important for life on earth because they are the foundation of food webs and play essential roles in nutrient cycling (including carbon and nitrogen), human health, and food production. Recent advances in sequencing have enabled efficient measurement of microbiomes and their composition, distribution, and impact on their environment have been extensively studied. However, it is still not well-known which species contributes to specific microbial functions such as carbon respiration, nitrogen fixation, and many others. To answer such questions, we need a reliable and relatively fast way of estimating bacterial functions. For this, we can use the gene content of bacterial genomes. Existing databases hold millions of bacterial genomes that can be used as a resource. From these genomes, we'll construct a species by gene function matrix to be used in functional estimation of environments and species. Apart from predicting which Bacteria are important for which functions, we are also interested in studying which functions are always found together, which are mutually exclusive, and which are linked to specific parts of the phylogenetic tree. Furthermore, we plan to use this to enrich bacterial co-response networks (<https://doi.org/10.1093/ismeco/ycaf036>) and learn what functions are enriched in different environments and microbial communities.

2) Link between microbiome and environment

Microbes are everywhere, yet the environment determines which microbes can live where. Although the major drivers of microbiome assembly are known, such as pH in soil environments or salinity and depth marine environments, it is not known in detail which environmental variables affect which species. In this project, large-scale microbiome data will be collected and combined with environmental metadata to study and predict which environmental variables matter for which species.

3) Predictive modeling

Want to model the microbiome and predict the possible outcomes when it is changing? What effects play a roll, what outcomes can be predicted? We have a set of ~1000 microbiomes of agricultural soils of which we want to know how they affect soil health and if we can use them for prediction of the effects of soil management

4) Co-op NIOO

Want to do an internship at the top ecological institute NIOO in Wageningen. There are several projects available such as: Hyperspectral data of the long-term ecological fields, the Clue-fields or on the microbiome as response to drought and plant type.

5) Expand SpeSpeNet-webtool (microbial network visualisations)

We recently designed and published a web-tool called SpeSpeNet (<https://doi.org/10.1093/ismeco/ycaf036>), that makes networks of microbiome data and visualizes them. In these networks, nodes represent microbial taxa and edges represent correlations between them. The nodes can be overlayed with environmental associations, such that the networks show which environmental variables drive microbial community assembly.

SpeSpeNet was written in R-Shiny and is publicly available (https://utrecht-university.shinyapps.io/SpeSpeNet_v1/). Due to the approach and packages used, the quality of the network visualizations is not quite up to publication standard. Hence, if the user wants to make SpeSpeNet figures for their own publication they currently need to manually adjust the visualization using R scripts available on the github (<https://github.com/SnoekLab/SpeSpeNet>). Ideally, the visualizations in the webtool would be publication ready without any adjustments. In this project we want a student who is proficient in software development to improve SpeSpeNet. Preferably, this would be done in R-shiny such that the current webtool can be updated, but an entirely new implementation is also acceptable provided that the student takes care that the same settings produce exactly the same network as in the current implementation. Once the visualizations are improved to publication quality, the student can add additional functionality to SpeSpeNet such as plotting functional annotations (such as from funguild) on top of the nodes in the network. The student can also suggest their own ideas regarding the direction to take.

Plant genetics and multi-omics

1) Phenomics and GWAS

We imaged populations of ~200 genotypes of Lettuce and ~200 genotypes of its wild ancestor *Lactuca serriola* under different conditions (control, salt stress, far-red light). Using the images, we want to investigate what genes affect phenotypes under standard conditions and if and how these change under salt and far-red light stress. We have sequenced all plants in the two populations, and we want to link the phenotypes to the plant genotypes using GWAS. First, the phenotypes have to be extracted from the obtained images. In this project, we want a student to quantify different phenotypes from the imaging data and use these as traits for GWAS. We can investigate several different phenotypes such as leaf shape, leaf color, or hyperspectral reflectance. In this project, the student will learn how to extract meaningful phenotypes from images, do GWAS, and interpret the results, while gaining experience in general research and writing skills.

2) eGWAS and gene regulatory networks

We measured gene expression of populations of ~200 genotypes of Lettuce and ~200 genotypes of its wild ancestor *Lactuca serriola* under different conditions (control, salt stress, far-red light). Since these genotypes have all been sequenced, we can now link expression variation to genetic variation using eGWAS. We will also investigate how the genetic regulation of gene expression changes between the standard condition versus when the plants encounter salt or far-red light stress. Furthermore, this analysis can be extended by constructing gene regulatory networks (GRNs), elucidating the different pathways affected by genetic variation. In this project you can learn how to work with RNA-seq data, how to do GWAS, how to construct GRNs and interpret the results, while gaining experience in general research and writing skills.

3) Data integration and exploration tool(s)

Results from QTL mapping and GWAS give insight into the effect of genetic variation on phenotypic and gene expression variation. However, these results are complex and not easy to navigate for someone without coding experience. Additionally, integrating the results with that of previous literature can yield valuable context but the effort required is prohibitive for most researchers. For some model organisms, tools exist to facilitate researchers in exploring and integrating QTL data. For example, the AraQTL platform ([eQTL visualization and investigation](#)) enables easy exploration, discovery, and integration of QTL studies in Arabidopsis. We would like a similar setup for *L. sativa* (lettuce) and expand the tool in such a way that construction of gene regulatory networks becomes feasible.

4) AraNL and phenomics

We recently collected a population of *Arabidopsis* throughout the Netherlands. Several hundred plants from this population have been imaged in the NPEC facility, by RGB, fluorescence, and hyperspectral cameras. We want to use this data to investigate how phenotypic variation is linked to geography and/or environmental condition of their place of origin. In this project, the student will do data extractions from the images output by NPEC and calculate phenotypes such as leaf color, leaf shape, and others. Then, the student can explore the link between phenotypic variation and geography/environment of their place of origin. In this project you can learn how to extract meaningful phenotypes from images, how to obtain environmental data from e.g. satellites and relate these to phenotypic variation, and interpret the results, while gaining experience in general research and writing skills.

5) Conservation of gene function: Between species comparison of gene expression

In biology, we often work with model systems, assuming that lessons learned in one species can be transferred to another. In plants, the most well researched model system is *Arabidopsis thaliana*, followed by tomato, wheat, and rice. However, to what extent can we really transfer results between plant species? In this student project, we will investigate whether and how many different aspects of plant genetics and physiology are conserved between species. We will investigate how often orthologous genes still have the same function between plant species. We'd also like to know how well and until what taxonomic rank gene expression levels are conserved. This conservation (or lack thereof) can then be linked to gene functions and pathways, so a prediction can be made on how well certain processes in one species translate to another. In this project you will learn how to work with RNA-seq data, comparative genomics and between species comparisons of gene functions, while gaining experience in general research and writing skills.

6) GWAS of hypocotyl length

In this project, the student will study the genetic regulation of the response to heat stress in *Arabidopsis*. Previously, we obtained the hypocotyl lengths of ~1000 different genotypes from imaging data. These 1000 genotypes have all been sequenced and can be used in GWAS to uncover loci involved in heat induced hypocotyl elongation. In this project you will learn how to do GWAS, how to search for candidate genes and interpret the GWAS results, while gaining experience in general research and writing skills.

7) Subset GWAS of the *Arabidopsis* 1001 population

GWAS results can depend on the population used. In *Arabidopsis*, several studies have measured phenotypes such as flowering time for up to 1000 different genotypes. We want to investigate how using a smaller set of genotypes for GWAS impacts the GWAS results compared to taking the full set of 1000 genomes. This is highly relevant for the field of genetics, because most non-human studies do not phenotype 1000s of genotypes, instead using only 100-200 genotypes. To investigate this, the student will do in-silico sub-setting of GWAS populations and study the different outcomes. These differences can highlight more complex genetic interactions or population specific effects, possibly masked by the larger population. The study also has broad implications for genetics studies in regard to how large the population needs to be to consistently pick up the regulators (which are known for flowering time phenotypes) with high significance. In this project you can learn how to do GWAS, search for candidate genes and interpret the results, while gaining experience in general research and writing skills.