

Masterthesis (d/f/m) - in Tübingen Parameter Optimization for Pan-Genome Graph Construction

Computomics is a growing Bioinformatics AgTech company with a vision of contributing to a sustainable future through innovative AI solutions for plant breeding, food safety and food security.

As for now we are looking for a motivated Master Student for the project.

The Project

Pan-genomics is revolutionizing the way we understand genomic diversity across species. As the number of available genome assemblies grows, we can surpass the limitations of linear reference genomes by constructing genome graphs, shared reference structures that combine multiple assemblies. This approach provides a comprehensive representation of sequence space while maintaining synteny and evolutionary relationships.

One way to create genome graphs is through multiple whole-genome alignments, transformed into graph structures. However, this process is inherently complex, requiring careful parameter tuning to balance sequence compression and synteny preservation. While aligning smaller sequences or pairs of genomes is relatively well-established, the alignment of multiple whole genomes remains a significant challenge.

The Challenge

Genomes, even within the same species, exhibit patterns of structural complexity due to transposable elements, structural variants, and other evolutionary events that disrupt synteny. These patterns and their extent differ from species to species, presenting unique alignment challenges, requiring tailored parameter optimization for graph construction. Despite advances in pan-genomics, our understanding of how alignment parameters influence genome graph complexity, linearity, and correctness is still incomplete.

Your primary goals and tasks are:

- **Metric Development:** Develop novel metrics to evaluate graph linearity and correctness, providing insights into the quality of the constructed genome graphs.
- **Parameter Tuning:** Explore and optimize parameters for pan-genome graph construction across multiple species (model organisms and agricultural species).

Your Qualifications:

- Background in bioinformatics, biology, genetics, or a related field.
- Familiarity with genome assembly, sequence alignment, or programming (e.g., Python, nextflow, etc.) is highly desirable
- Strong written and verbal communication
- > Capable of working autonomously as well as a being proactive team player

Our Offer:

- Working in an innovative, dynamic and international company
- Comfortable working environment with an openminded and supportive team
- Onboarding plan and mentoring
- Chance to grow with us
- Family-friendly workplace
- Fresh fruit, sweets, coffee, tea and water free of charge
- Company events (BBQ, company outing etc.)
- Ergonomic office furniture and equipment
- Inspiring and beautiful office view from charming Tübingen city

For more information or to apply, please contact: careers@computomics.com