Weihua Pan

University of California, Riverside (951)323-3581
Computer Science and Engineering wpan005@ucr.edu

Education

- **University of California, Riverside (Current)** 2014.9 - 2019.6 (expected)
  Ph.D, Computer Science *Riverside, CA*
  - GPA: 3.92
- **University of California, Riverside (Current)** 2016.9-2018.6 (expected)
  M.S, Statistics *Riverside, CA*
  - GPA: 3.80 (until Summer 2017)
- **University of Science and Technology of China** 2011.9-2014.6
  M.E, Computer Software and Theory *Hefei, China*
  - GPA: 3.68 (rank: 8/90)
- **Nanjing Normal University** 2007.9-2011.6
  B.E, Computer Science and Technology *Nanjing, China*
  - GPA: 3.58 (rank: 2/100) This rank is a comprehensive evaluation for admission to graduate school.

Skills

Programming Languages: C/C++, Python, R, MatLab, SQL, Java

Programming Tools: MPI, OpenMP, TensorFlow(learning), Hadoop, MySQL

Operating Systems: Unix/Linux, MacOS X, Windows

Bioinformatics Tools: BLAST, BWA, Bast2Go, Samtools, etc.

Biology Knowledge: undergraduate level molecular biology class and genetics class took, and basic knowledge about epigenetics, cell biology and biochemistry

Research Interests

- Computational Biology
  - Contiguity Improvement for Genome Assembly
  - Nucleosome Movement and its relationship with Gene Expression and Function
  - Large-scale Haplotype Inference
  - Metagenomics Sequence Analysis and Functional Analysis
- Combinatorial Optimization Approximate Algorithm
- Statistics Modeling / Machine Learning / Deep Learning / Data Mining

Research Projects

- **Nucleosome Movement and its relationship with Gene Expression and Function (in progress)**
  - We already propose a method *ThIEF* for tracking nucleosomes across multiple time points. Now I'm trying to characterize nucleosome movement using tracks generated by *ThIEF*, and study its relationship with gene expression and gene function. If strong covariance can be found, I plan to apply deep learning method to predict gene expression level and gene function by nucleosome movement information.
- **Chimeric Contigs Correction for *de novo* Genome Assemblies via Low-quality Optical Maps (almost finished)**
  - I propose a novel chimeric contigs correction method called *Novo&Chimeric* that can correct chimeric contigs and chimeric optical maps by each other. The problem is modeled into weighted vertex cover problem for deciding whether to correct contigs or optical maps when they conflict. Experiments show that our tool can outperform manual work done by experienced researcher of genome assembly problem using much less time.
- ***De novo* Genome Assemblies Reconciliation via Optical Maps (finished)**
  - I propose a novel assembly reconciliation method called *Novo&Stitch* that can take advantage of optical maps to accurately carry out assembly reconciliation. Combinatorial optimization models and technologies such as graph model, dynamic programming, weighted vertex cover model on hypergraph, greedy strategy, linear programming are used for solving some subproblems like data reduction, error correction and post-processing. Extensive experimental results demonstrate that our tool can significantly improve the contiguity of de novo genome assemblies without introducing misassembles or reducing completeness.
- **Large-scale Haplotype Inference for Population Data (finished)**
  - I improve a haplotype inference algorithm *WinHAP* to version 2 in two aspects: (1) a divide-and-conquer strategy is utilized to solve the challenge of huge computer memory (2) the *OpenMP* parallel computing mode is implemented to utilize all the computing power in a multi-core computer cluster. *WinHAP2* algorithm phases 500 genotypes with 1,000,000 *SNP*s using just 12.8*MB* in memory and 2.5 hours on a personal computer.
- **Metagenomics Functional Analysis Pipeline (finished)**
  - We improve a genomics functional analysis tool *Parallel-MATA* to version 2. *Parallel-MATA2.0* enhances the taxonomical analysis based on multiple databases, improved parallel computation's efficiency, and enabled interactive visualization of results. Furthermore, it includes functional analysis for metagenomic samples, which is based on dual computational engines: one based on *SEED* database and another based on *GO* (Gene Ontology) hierarchical structure.
- **Metagenome *NGS* Reads Binning (finished)**
  - I propose a supervised metagenome reads binning method MetaObtainer by combining both similarity-based and composite-based binning methods. It can deal with very short *NGS* reads because similarity-based method is used for pre-grouping reads before characterization, and it's much faster than most similarity-based methods because it's alignment-free.

Publications

- **Pan, W.**, Lonardi, S., Novo&Stitch: Accurate Reconciliation of Multiple de novo Genome Assemblies via Optical Maps. Submitted to *RECOMB' 18* conference.
- Polishko, A., Hasan, M., **Pan, W.**, Bunnik, E.M., Le Roch, K. and Lonardi, S., 2017. ThIEF: Finding Genome-wide Trajectories of Epigenetics Marks. *WABI' 17*, 19:1-19:16.
- **Pan, W.**, Chen, B. and Xu, Y., 2015. MetaObtainer: A Tool for Obtaining Specified Species from Metagenomic Reads of Next-generation Sequencing. *Interdisciplinary Sciences: Computational Life Sciences*,7(4), pp.405-413.
- **Pan, W.**, Zhao, Y., Xu, Y. and Zhou, F., 2014. WinHAP2: an extremely fast haplotype phasing program for long genotype sequences. *BMC bioinformatics*,15(1), p.164.
- Su, X.$, **Pan, W.$**, Song, B., Xu, J. and Ning, K., 2014. Parallel-META 2.0: enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. *PLoS One*,9(3), p.e89323. ($ represents co-first author)

Tools

- **Novo&Chimeric**: Optical Map based Chimeric Contigs Correction Tool
  `https://github.com/ucrbioinfo`
- **Novo&Stitch**: Optical Map based de novo Genome Assemblies Reconciliation Tool
  `https://github.com/ucrbioinfo`
- **WinHAP2**: Large-scale Haplotype Inference Tool
  `http://staff.ustc.edu.cn/~xuyun/winhap`
- **Parallel-META 2.0**: Metagenomics Functional Analysis Pipeline
  `http://bioinfo.single-cell.cn/parallel-meta.html`
- **MetaObtainer**: Supervised Metagenome NGS Reads Binning Tool
  `http://www.cs.ucr.edu/~wpan005/metaobtainer`

Working Experience

- **Research Assistant / Teaching Assistant** — 2014.9 - present
  Department of Computer Science and Engineering, UCR — *Riverside, CA*
- **Summer Internship** — Summer 2012
  Qingdao Institute of Bioenergy and Bioprocess Technology, CAS — *Qingdao, China*
- **Research Assistant / Teaching Assistant** — 2011.9 - 2014.6
  School of Computer Science and Technology, USTC — *Hefei, China*