UNIVERSITEIT VAN AMSTERDAM

Journée Snakemake

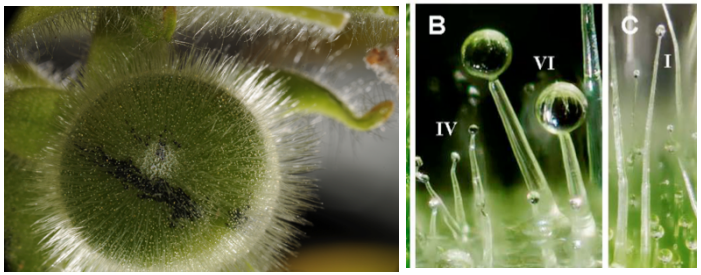
RNA-Seq pipelines with Snakemake (small RNAs, mRNA)

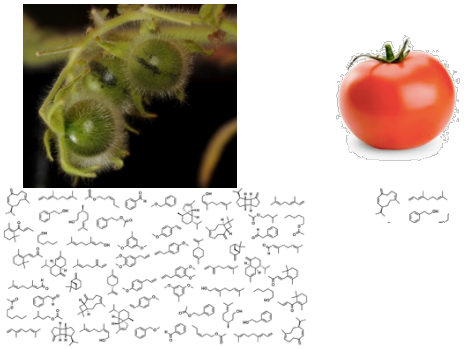
Marc Galland

Paris, December 12, 2016

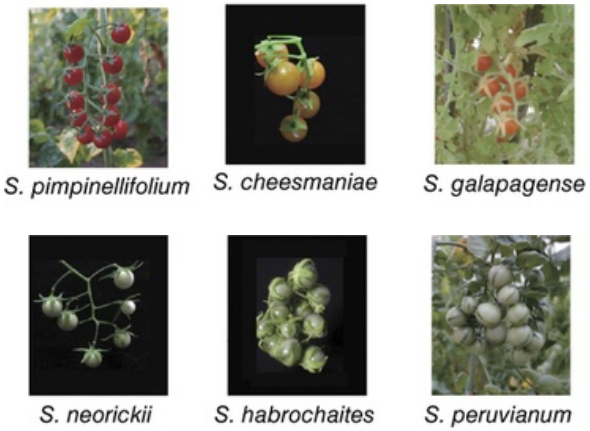
Wild tomatoes produce natural insecticides

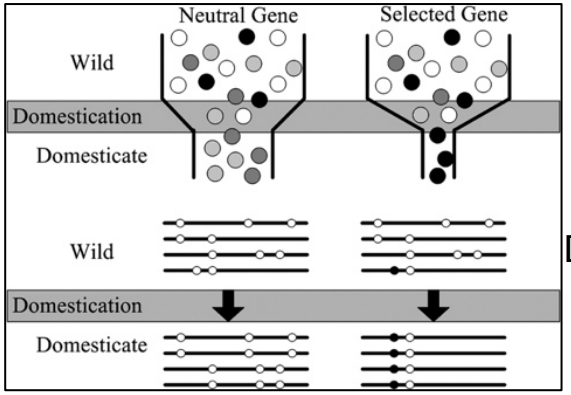
Whitefly outbreaks

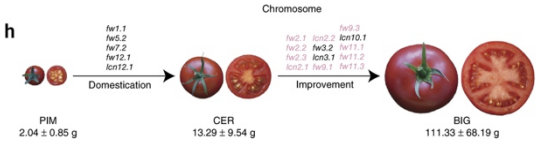




Wild tomato are more resistant to insects due to the synthesis of insect-toxic compounds in their trichomes







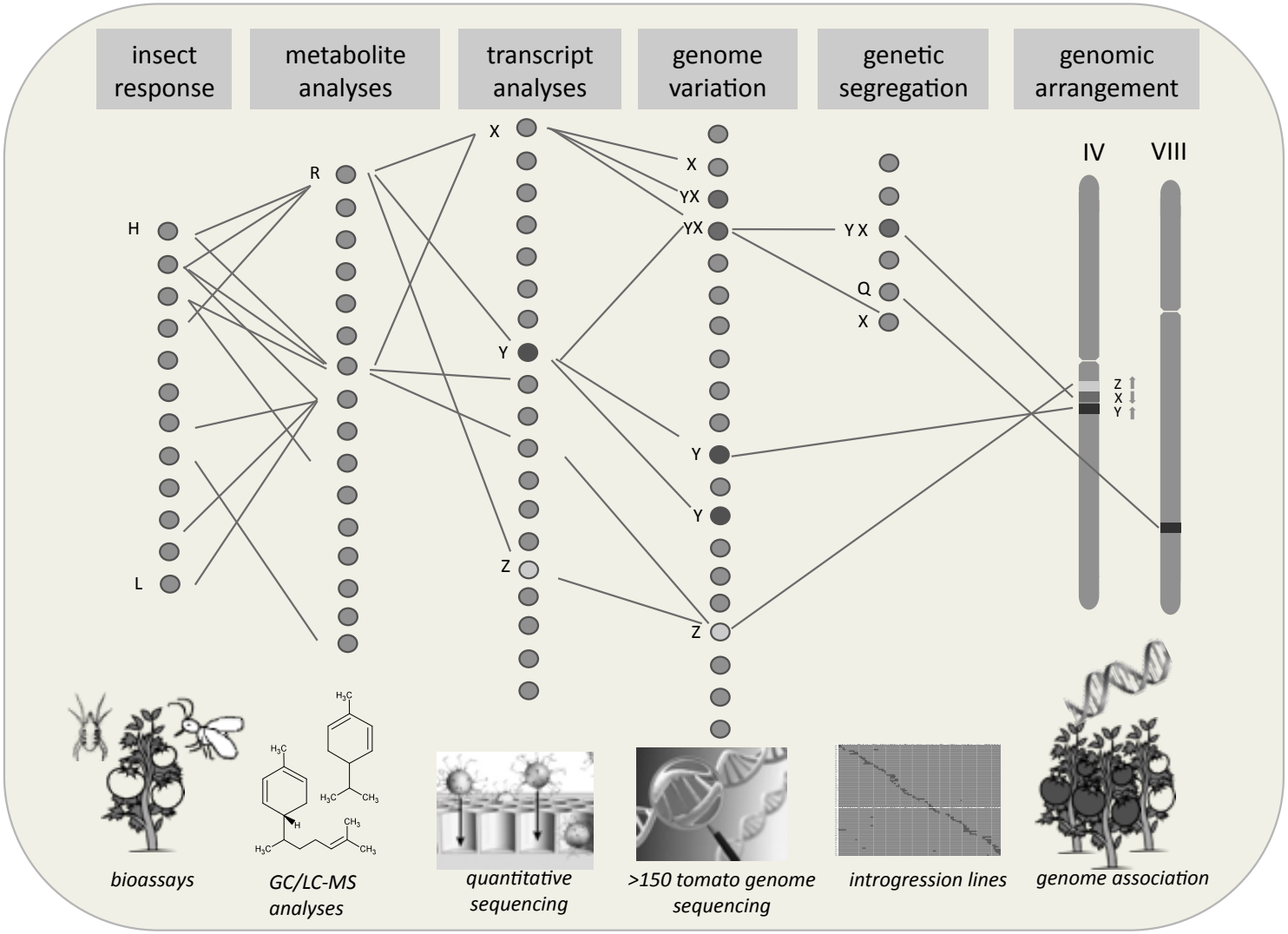
Drastic reduction of genetic diversity

Domestication population ~300 plants

Ross-Ibarra *et al.,* (2007) *PNAS* 104:8641-8648; Lin *et al.*, (2014) *Nat.Gen.* 46:1220-1226; Aflitos *et al.,* (2014) *Plant J* 80(1):136-148

Project overview

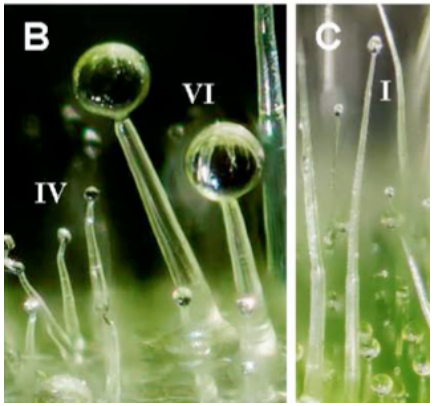
Main goal: Explore natural variation in wild tomatoes to identify metabolic pathways and their transcriptional regulators to improve resistance in commercial tomato

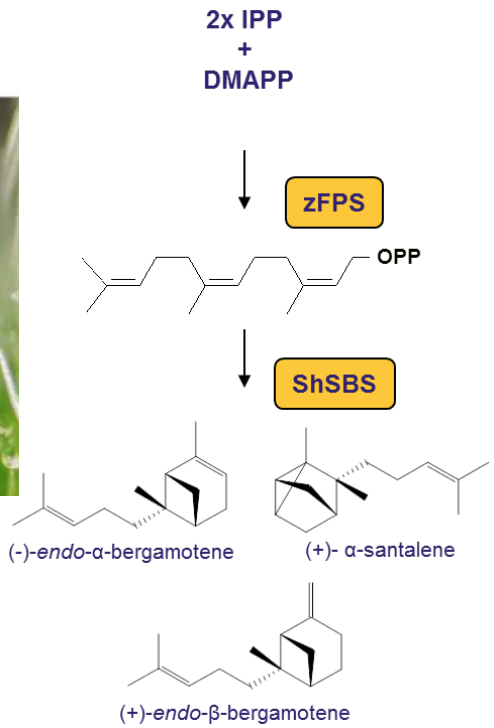


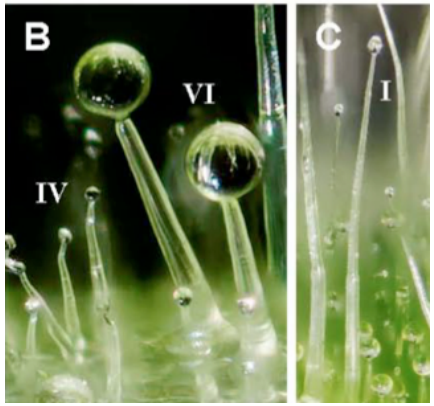
My main two questions

What are the genes related to production of insecticides in glandular trichomes?

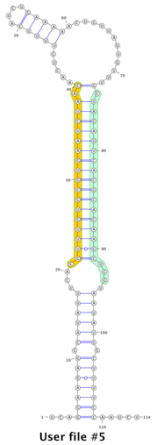
How is the synthesis of specialized metabolites regulated in glandular trichomes ?





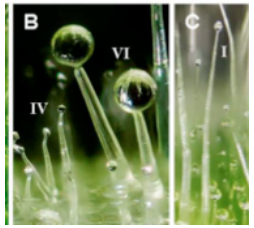


A role for small RNAs ?

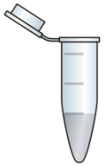


Trichome transcriptomics
RNA sequencing

80 mRNA-Seq samples: 20 accessions x 4 biological replicates

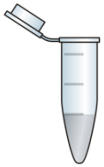


RNA isolation
from trichomes



mRNAs

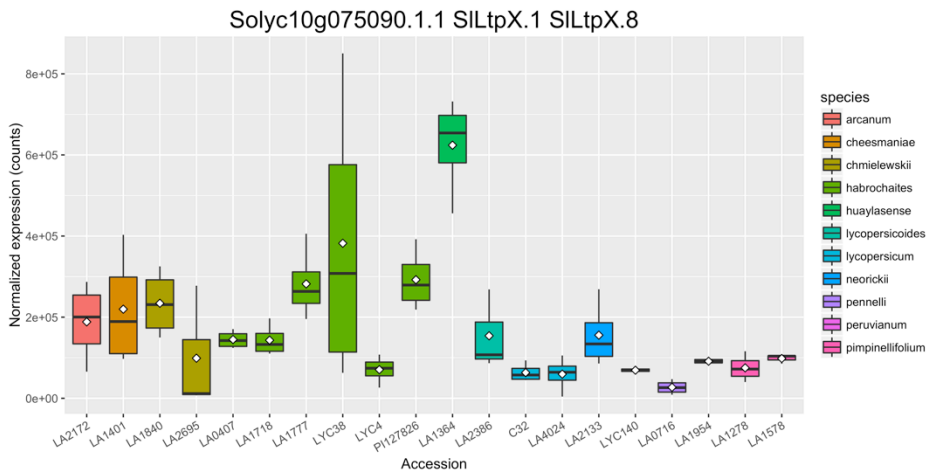
mRNA-Seq (proton ion sequencer)



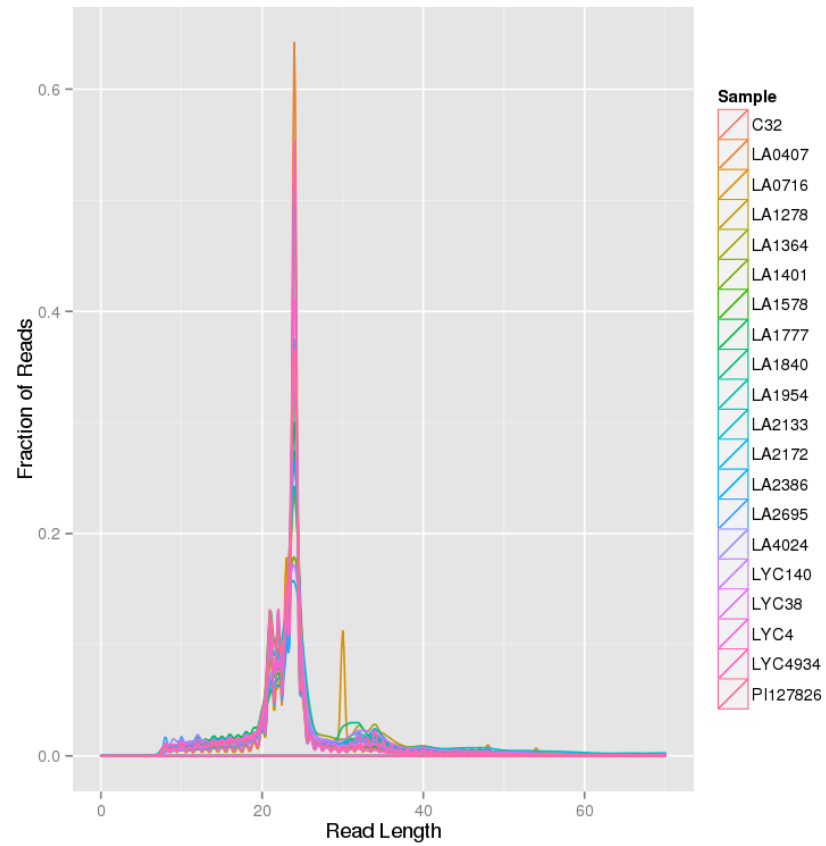
small RNAs (< 200 nts)
≈ 7% of total

Small RNA-Seq (proton ion sequencer)

What do I use Snakemake for?



Gene expression estimates from mRNA-Seq



Small RNA analyses

Snakemake for mRNA-Seq

NATURE BIOTECHNOLOGY | RESEARCH | BRIEF COMMUNICATIONS

日本語要約

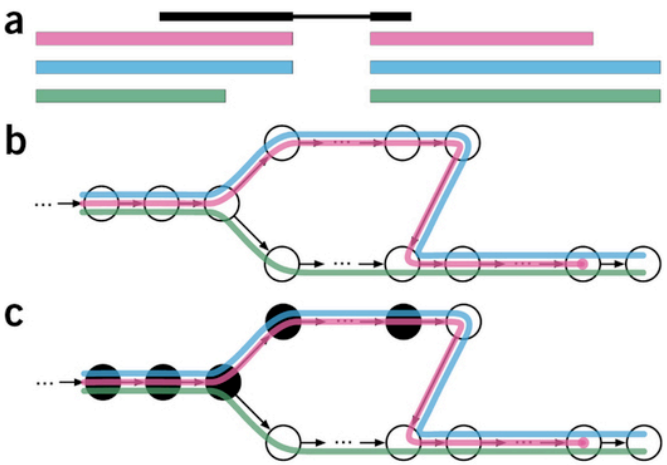
Near-optimal probabilistic RNA-seq quantification

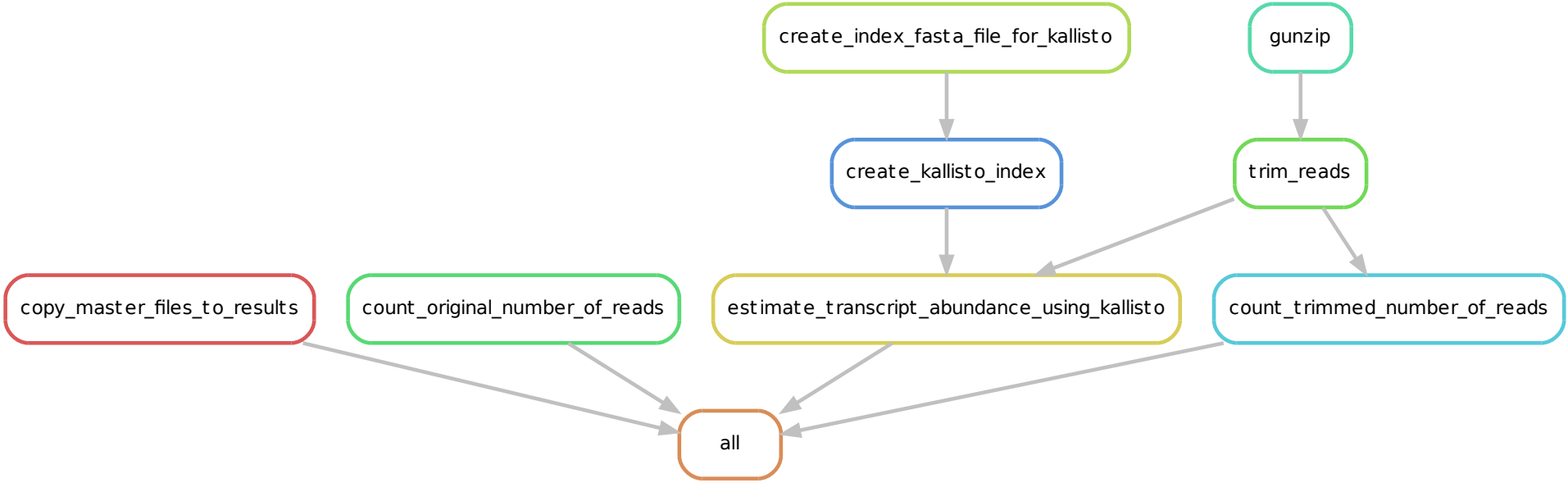
Nicolas L Bray, Harold Pimentel, Páll Melsted & Lior Pachter

Affiliations | Contributions | Corresponding author

Nature Biotechnology **34**, 525–527 (2016) | doi:10.1038/nbt.3519

Received 15 October 2015 | Accepted 25 February 2016 | Published online 04 April 2016





Snakemake for small RNA-Seq

Snakemake tips and tricks

- Great system. Saved my life as a computational biologist!
- Main issue at start: wildcards
  - From output files
  - Using a function to infer wildcards
- Problem with R util (Snakemake version 3.3)
- Did not use sub-workflows. Is it a good idea to get modular? Better to have one Snakefile per input/output(s)?
- Snakemale & computer grids. Maybe a problem not related to snakemake!

Acknowledgments

Swammerdam Institute of Life Sciences

Greenhouses

Ludek Tikovsky & Harold Lemereis

"RNA Biology & Applied Bioinformatics"

Han Rauwerda
Wim de Leeuw
Rob Dekkers
Selina van Leeuwen
Wim Ensink
Marina van Olst
Timo Breit

"The trichome team" (T^3)

Petra Bleeker (PI)
Ruy Kortbeek (PhD student)
Aleksandra Muras (Technician)

"Biosystems and Data Analysis" group

Johan Westerhuis
Yipeng Song

Funding

NWO-STW VIDI "Tomatoes arm themselves against insects"
"DURETO" - Marie Curie Individual Fellowship 2014

Thank you for your attention