



"Looking for that perfect fit: Overcoming roadblocks to an optimal sequencing strategy with biocontrol agent *Nesidiocoris tenuis* as a case study" by Kim Ferguson is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

# Looking for that perfect fit:

Overcoming roadblocks to an optimal sequencing strategy with biocontrol agent *Nesidiocoris tenuis* as a case study

Kim Ferguson

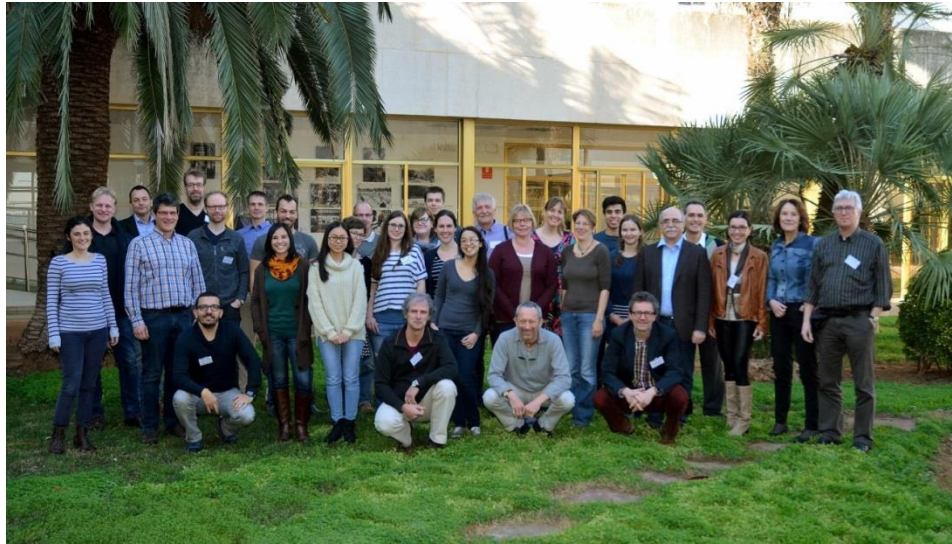
PhD Candidate

Wageningen University

 @kfergy

Given at the Royal Entomological Society's 2<sup>nd</sup>  
Meeting of the Insect Genomics Special Interest  
Group, 16<sup>th</sup> May 2017

# **BINGO: Breeding Invertebrates for Next Generation BioControl**



- Key idea: Improvement of biocontrol agents via natural genetic variation
- Network of universities, research institutes, and companies
  - 13 PhDs
  - 24 senior researchers
  - 12 partners from Netherlands, Germany, France, Spain, Czech Republic, Austria, Switzerland, Greece and Portugal

# My project:

Assemble the genome of three important biocontrol agents and compare the genetic variation between commercial strains and native populations

- genetic bottlenecks
- lab adaptation
- coding and non-coding regions
- aimed at making genome research attractive to biocontrol

*Trichogramma brassicae*



Sophie Chattington, Uni Bremen

*Amblyseius swirskii*



Angeliki Paspatis, IVIA

*Nesidiocoris tenuis*



Milena Chinchilla-Ramirez, IVIA

# Meet Nesi



- *Nesidiocoris tenuis*, mirid
- World-wide distribution
- Used throughout Spain against tomato pests
  - *Tuta absoluta*, whitefly, spider mites, thrips

- Zoophytophagous

## Aspects that complicate sequencing

- Diploid
- Generation time: ~ 1 month
- No reference genome
- Unknown :
  - Inbreeding ability
  - Standing genetic variation



# Goldilocks & the 3 Sequencing Strategies



Arthur Mee and Holland Thompson, eds. *The Book of Knowledge* (New York, NY: The Grolier Society, 1912)

# 1<sup>st</sup> Strategy: PacBio/Illumina hybrid *de novo*

THEORETICAL SIZE

ILLUMINA



PACBIO



## Benefits:

- High coverage
- Less scaffolds, and other assembly concerns
  - Tandem repeat areas covered by long reads
  - Errors from PB corrected with short reads
- By far, the most recommended approach

## Risks:

- Requires relatively large amounts of DNA (~10-20 µg)
- Inbreeding necessary
- “Recommended” ≠ affordable

**But why is inbreeding necessary?**

# Inbreeding for homozygosity

- Why? As little polymorphism as possible
  - $2n \rightarrow n$
  - Error rate of PacBio and ease of assembly
  - Decrease complexity  $\rightarrow$  genomes are a generalization



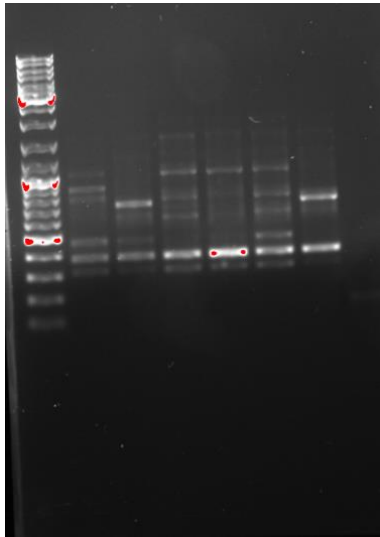
3 attempts at lab rearing over 1 year  
None survived to the 2<sup>nd</sup> generation

- 1 month/generation  $\rightarrow$  10 inbred generations = 10 months
  - not sustainable to keep trying past 1 year
- Inbreeding may be possible, but impractical for this project

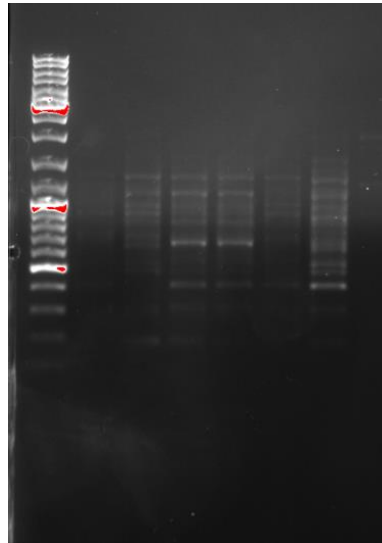
**Maybe it's already homozygous? Or close?**

# Well, how variable is the population?

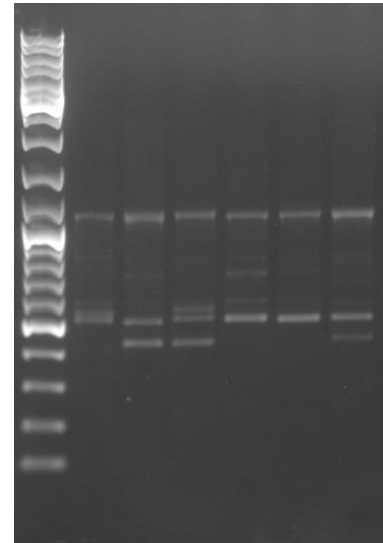
- Single-source population for more than a decade
- Maybe there's little variation i.e. inbreeding not req'd
- Random amplified polymorphic DNA (RAPD) markers
  - 6 individuals, 10 RAPD markers (OPA series 1-10)



OPA1



OPA2



OPA4

Ferguson, KB (unpublished data)



# 1<sup>st</sup> Strategy: PacBio/Illumina hybrid *de novo*

THEORETICAL SIZE

ILLUMINA



PACBIO



## Benefits:

- Increased coverage
- Less scaffolds, and other assembly concerns
  - Tandem repeat areas covered by long reads
  - Errors from PB corrected with short reads
- By far, the most recommended approach

## Risks:

- Requires relatively large amounts of DNA (~10-20 µg)
- Inbreeding necessary
- “Recommended” ≠ affordable

**Without homozygosity,  
too risky to pursue**

## 2<sup>nd</sup> Strategy: i5K approach



- Specifically, their protocol for difficult to inbreed species
- Pooled individuals for a larger sized library , and single individual for smaller insert sizes

### Benefits:

- Homozygosity not required
- Established assembly pipeline

### Risks:

- More time-consuming assembly method
- Enough material from individual

## 2<sup>nd</sup> Strategy: i5K approach



- Specifically, their protocol for difficult to inbreed species:
- Pooled individuals for a larger sized library , and single individual for smaller insert sizes

### Benefits:

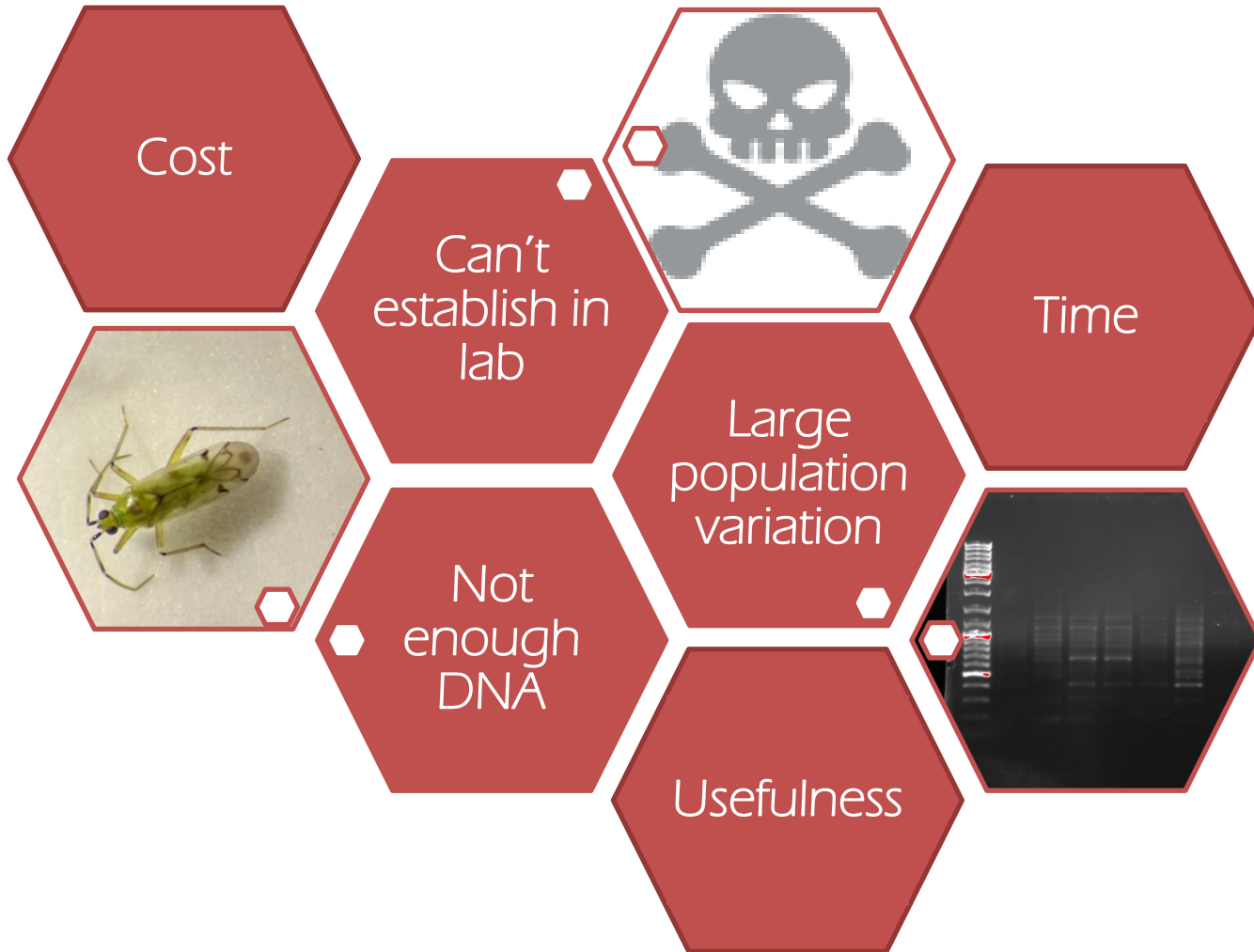
- Homozygosity not required
- Established assembly pipeline

### Risks:

- More time-consuming assembly method
- Enough material from individual

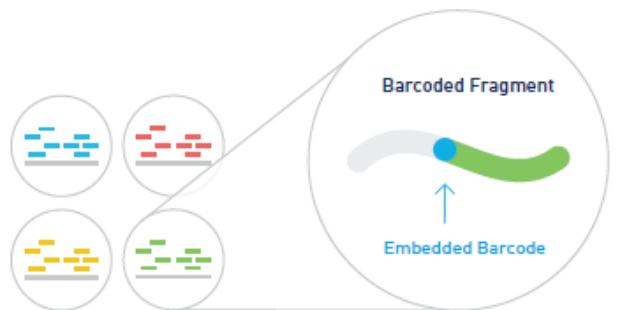
**Not enough DNA from an individual Nesi**

# Roadblocks

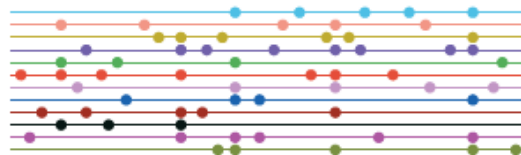


# 3<sup>rd</sup> Strategy: 10x Genomics

- The 10x Chromium library prep uses barcodes on short sequences
- Illumina for the sequencing



Genome is replicated and barcoded via a low-level enzymatic replication.



Lines represent Linked-Reads. Dots represent reads.  
Color indicates barcode. Reads with same barcode originated from molecules encapsulated in the same partition.

10x Genomics

## Benefits:

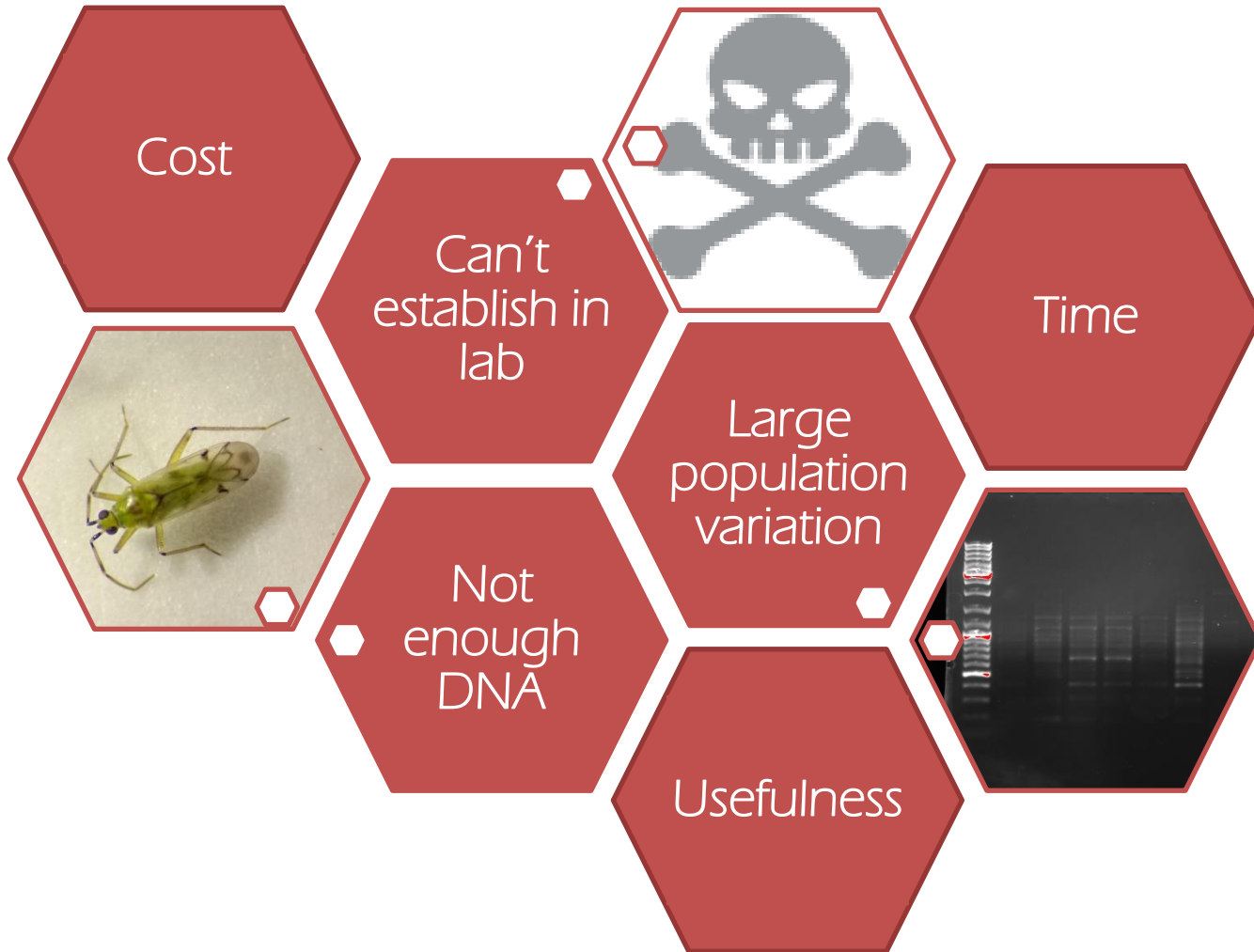
- High coverage
- Less scaffolds
- Small amount of DNA
- Phased genomes with structural variants

## Risks:

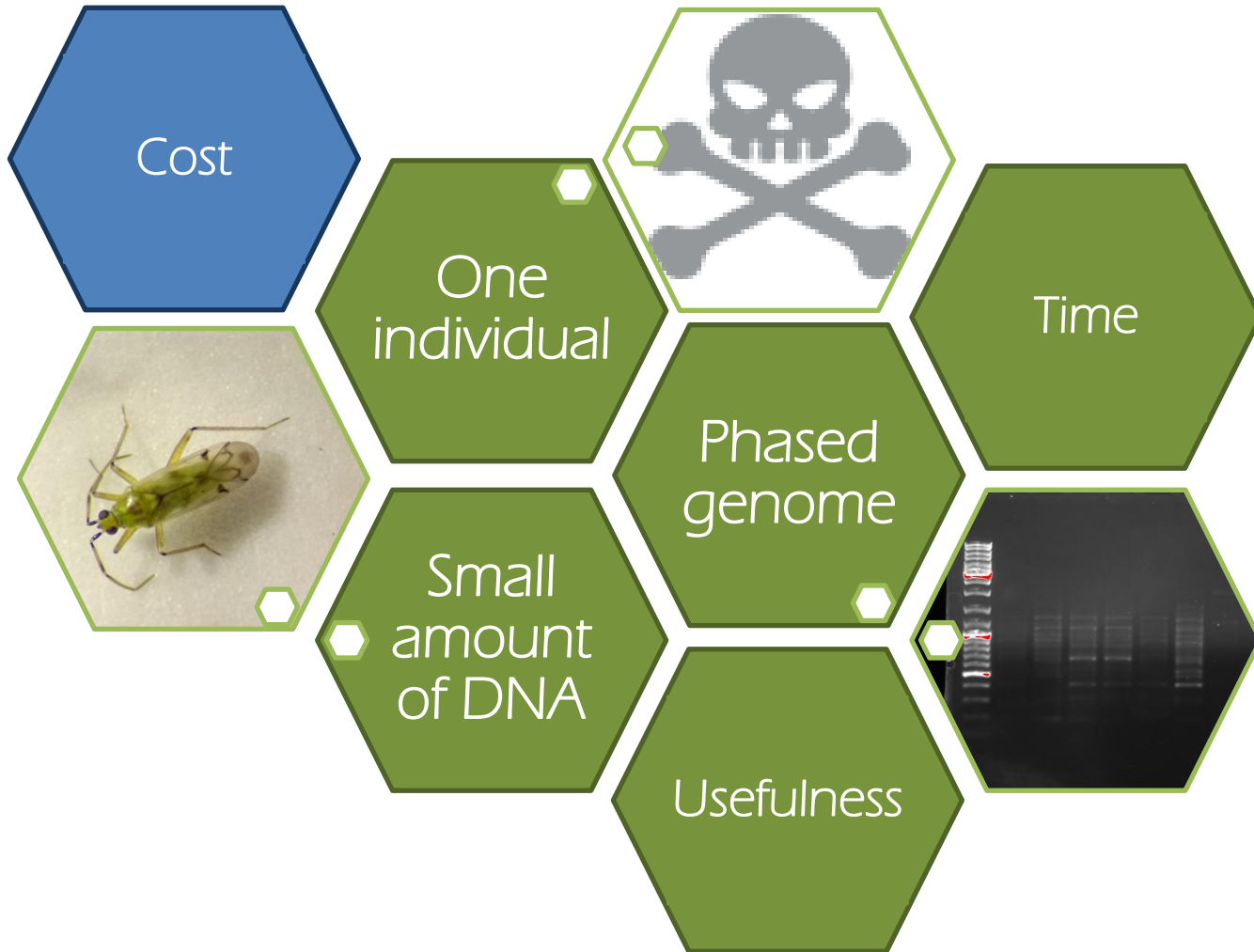
- Proprietary library prep, assembly algorithm, and visualizations
- Cost prohibitive
- Chromium reaction could fail



# Roadblocks



# Opportunities



# The End?

So far:

- BioScience at Wageningen University
- Confirmed 10x barcoding reaction
- Currently being sequenced
- Bioinformatics to be performed soon
- Goal: Collaborative, open annotation strategy



Jessie Willcox Smith - *Goldilocks and the Three Bears*, 1916

# Acknowledgements

- My supervisors:
  - Dr. Bart Pannebaker, Dr. Eveline Verhulst, and Dr. Bas Zwaan
- Laboratory of Genetics, Wageningen University
- Dr. Elio Schijlen at BioScience (Wageningen University and Research)
- Koppert Biocontrol Systems
- Milena Chinchilla-Ramirez and IVIA



[www.bingo-itn.eu](http://www.bingo-itn.eu)





# Questions?

Kim Ferguson

 @kfergy



# Photo Credits

- Slide 2: BINGO-ITN ([www.bingo-itn.eu](http://www.bingo-itn.eu))
- Slide 3:
  - *Trichogramma brassicae*: ©BugsInThePicture
  - *Amblyseius swirskii*: ©Koppert BV
  - *Nesidiocoris tenuis*: Kim Ferguson, 2017 CC by 4.0
- Slide 4:
  - Tomatoes: Kim Ferguson, 2017 CC by 4.0
  - Tomato plant: Instituto Valenciano de Investigaciones Agrarias (IVIA)
- Slide 5: Arthur Mee and Holland Thompson, 1912
- Slide 10: i5K logo, <https://www.hgsc.bcm.edu/arthropods/i5k>
- Slide 13: 10x Genomics promotional image, [10genomics.com](http://10genomics.com)
- Slide 16: Jessie Willcox Smith, 1916
- Emojis are provided by EmojiOne (2.3)