Latest sequencing technologies to analyze population structures and dynamics

Eveline Verhulst

Laboratory of Entomology

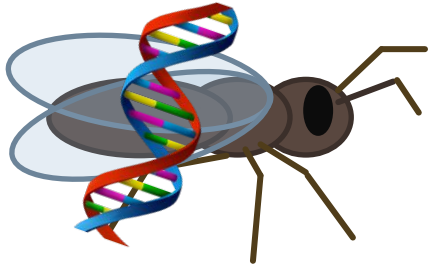
BINGO Workshop, 23 Jan 18

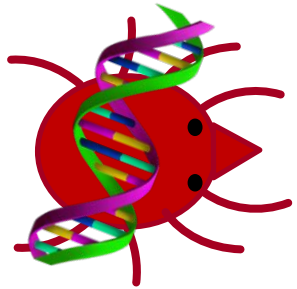
www.evelineverhulst.nl

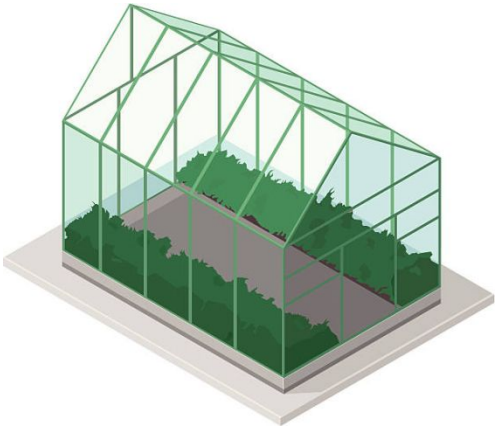
@ECVerhulst

WAGENINGEN
UNIVERSITY & RESEARCH

Biocontrol and evolution







What is my (starting) population?

Species composition

- Host plant and pest
- Origin of biocontrol population
  - Cryptic species
  - Hybridizations
    - Hybrid breakdown
- Unintended release of exotic species

Infections

- (Endo)Symbionts
- Microorganisms
- Diseases (fungi, bacteria, nematodes)

WHAT IS OUT THERE?

- and how much?
- and what are the dynamics?

Population genetics

- Study of genetic variation within and among populations
- Explained by evolutionary factors
- Based on the Hardy–Weinberg law
  - population size is large
  - mating is at random
  - mutation, selection and migration are negligible
- Else allele frequencies and genotype frequencies may change from one generation to the next.
  - allowing us to infer evolutionary patterns

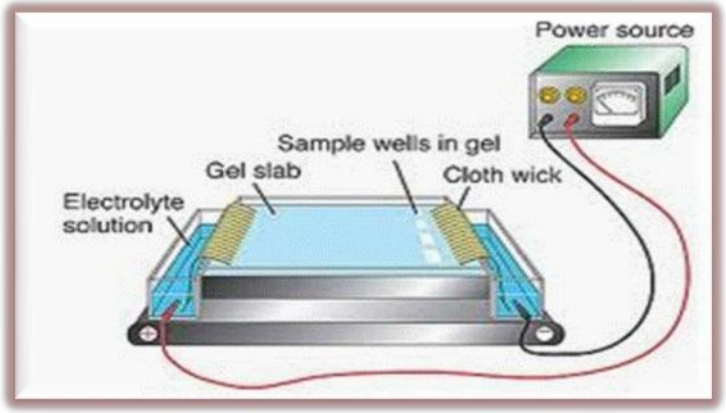
Measuring genetic variation

A bit of history

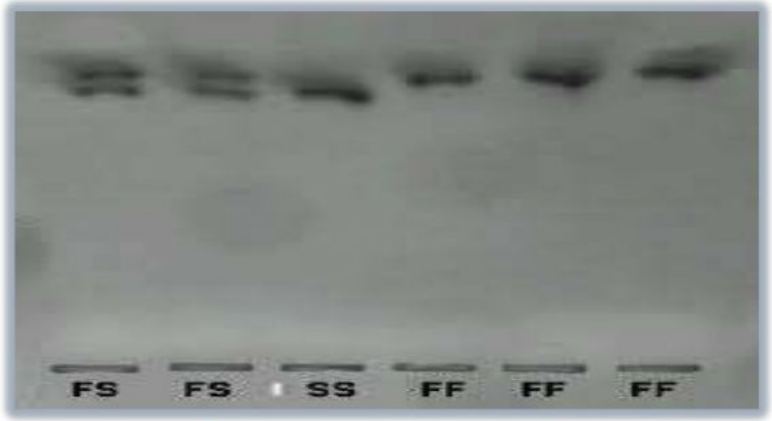
Types of population analysis

- Allozymes
- Microsatellites
- SNPs and Indels
  - Large genomic scale
    - RADSeq
    - RAPD
    - AFLP
    - RFLP
  - Small genomic scale
    - Sequence fragments of ITS2, COI, 16S

Allozymes



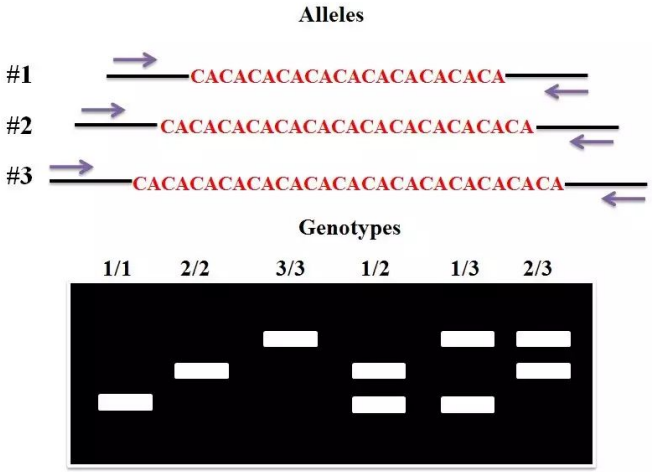
- low abundance
- low level of polymorphism

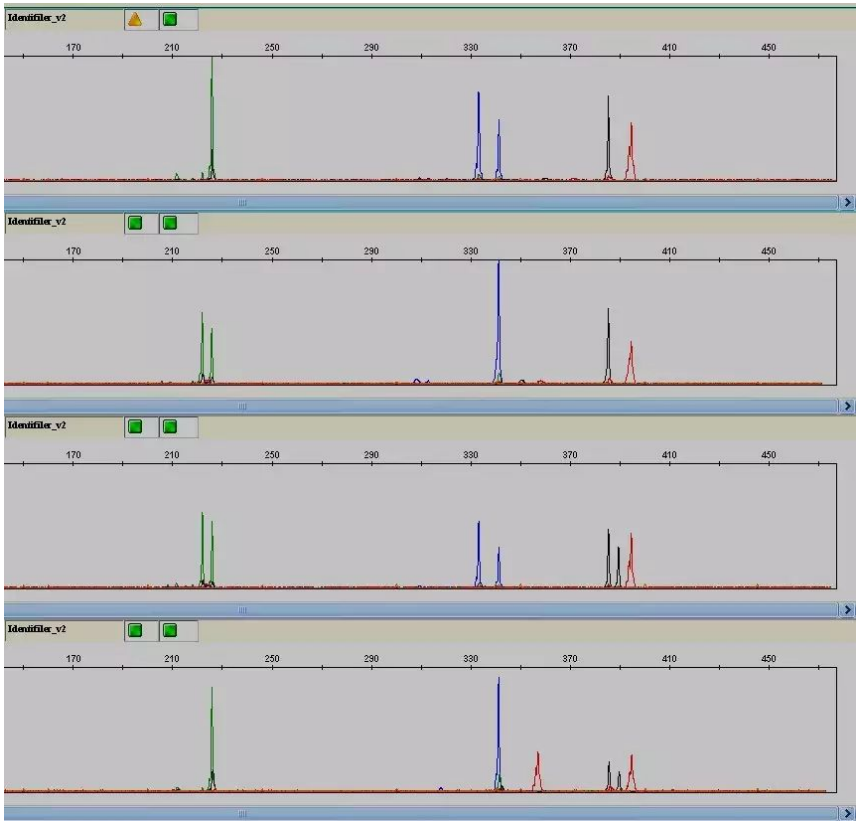


Types of population analysis

- Allozymes
- Microsatellites
- SNPs and Indels
  - Large genomic scale
    - RADSeq
    - RAPD
    - AFLP
    - RFLP
  - Small genomic scale
    - Sequence fragments of ITS2, COI, 16S

Microsatellites



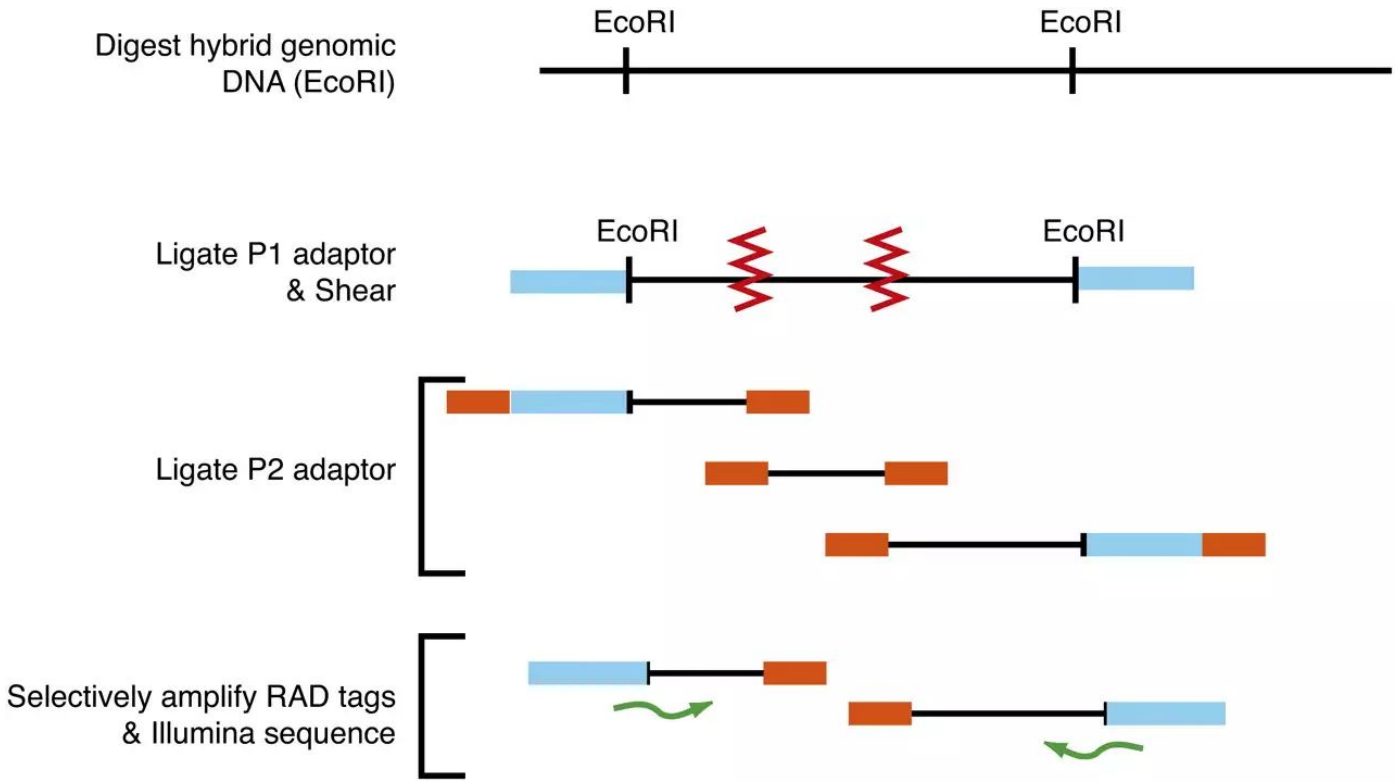


- Very robust
- Lower resolution

Types of population analysis

- Allozymes
- Microsatellites
- SNPs and Indels
  - Large genomic scale
    - RADSeq
    - RAPD
    - AFLP
    - RFLP
  - Small genomic scale
    - Sequence fragments of ITS2, COI, 16S

RAD-seq


Digest hybrid genomic DNA (EcoRI)
EcoRI
EcoRI
Ligate P1 adaptor & Shear
EcoRI
EcoRI
Ligate P2 adaptor
Selectively amplify RAD tags & Illumina sequence

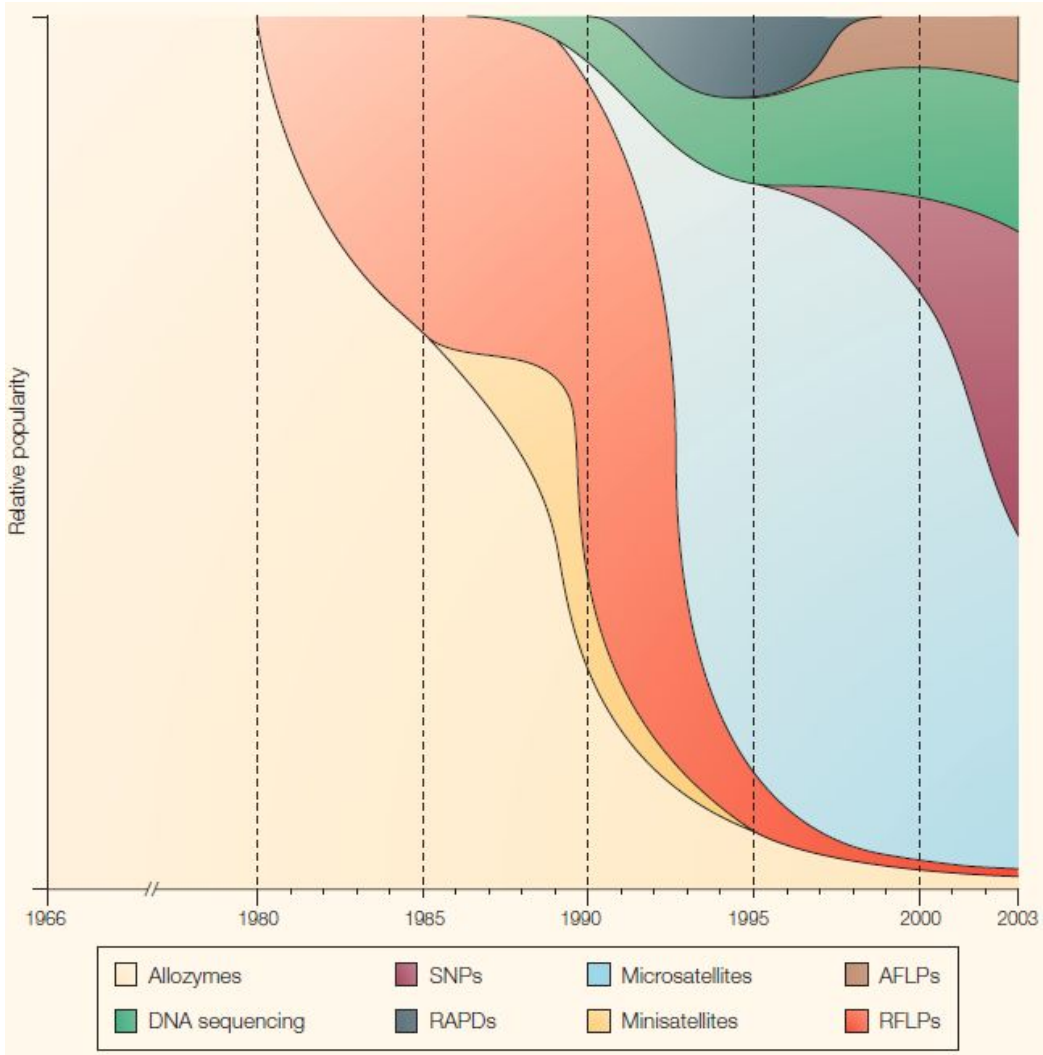
Types of analysis

- Allozymes
- Microsatellites
- SNPs and Indels
  - Large genomic scale
    - RADSeq
    - RAPD
    - AFLP
    - RFLP

    }
    - Not robust
    - Higher resolution
  - Small genomic scale
    - Sequence fragments of ITS2, COI, 16S, etc…

ITS2, COI, 16S

- Resolution often issue
  - Sanger sequencing cannot detect different types of polymorphisms unless cloned
- Sequence length
  - Sanger sequencing up to 500 bp



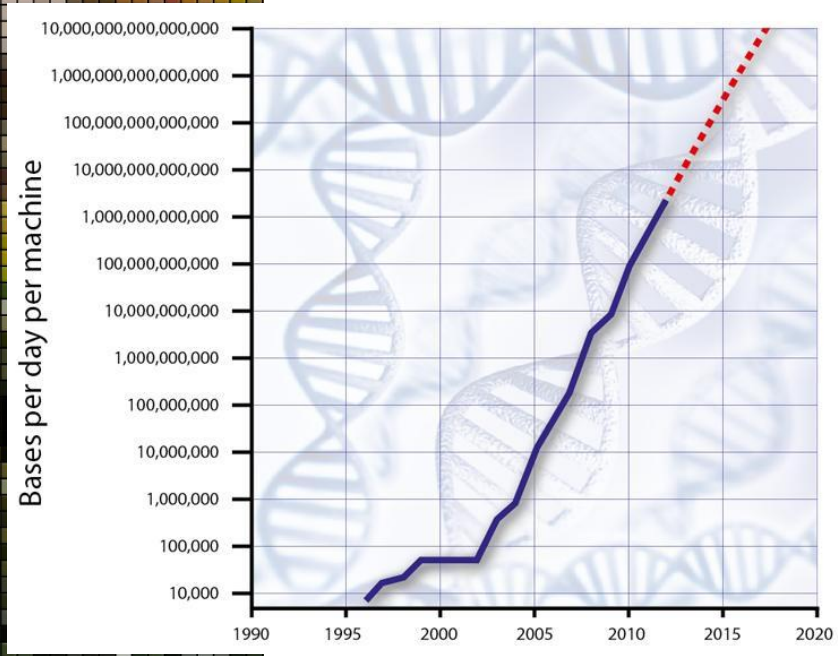
Schlötterer
2004, Nature

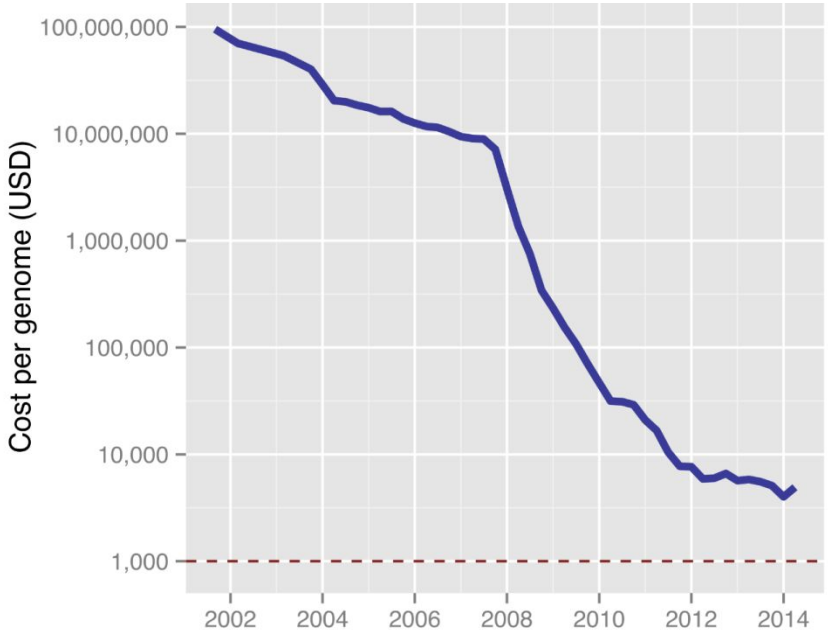
Marker	Advantages	Disadvantages
SNPs	• Low mutation rate • High abundance • Easy to type • New analytical approaches are being developed at present • Cross-study comparisons are easy; data repositories already exist	• Substantial rate heterogeneity among sites • Expensive to isolate • Ascertainment bias • Low information content of a single SNP
Microsatellites	• Highly informative (large number of alleles, high heterozygosity) • Low ascertainment bias • Easy to isolate	• High mutation rate • Complex mutation behaviour • Not abundant enough • Difficult to automate • Cross-study comparisons require special preparation
Allozymes	• Cheap • Universal protocols	• Requirement for fresh or frozen material • Some loci show protein instability • Limited number of available markers • Potentially direct target of selection
RAPDs and derivatives	• Cheap • Produces a large number of bands, which can then be further characterized individually (for example, converted into single locus markers)	• Low reproducibility • Mainly dominant • Difficult to analyse • Difficult to automate • Cross-study comparisons are difficult
DNA sequencing	• Highest level of resolution possible • Not biased • Cross-study comparisons are easy; data repositories already exist	• Still significantly more expensive than the other techniques

RAPD, randomly amplified polymorphc DNA; SNP, single nucleotide polymorphism.

Schlötterer 2004, Nature

Drastic changes in sequencing technology

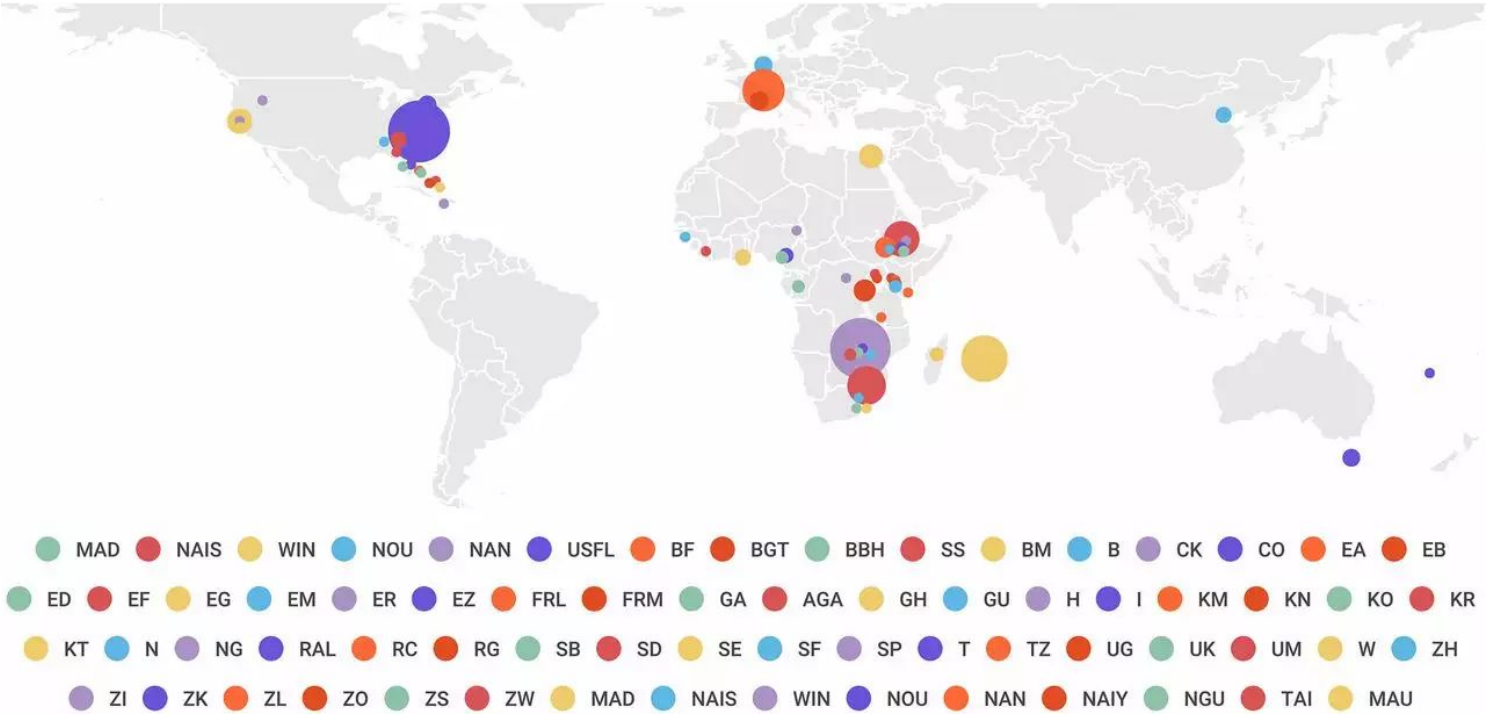




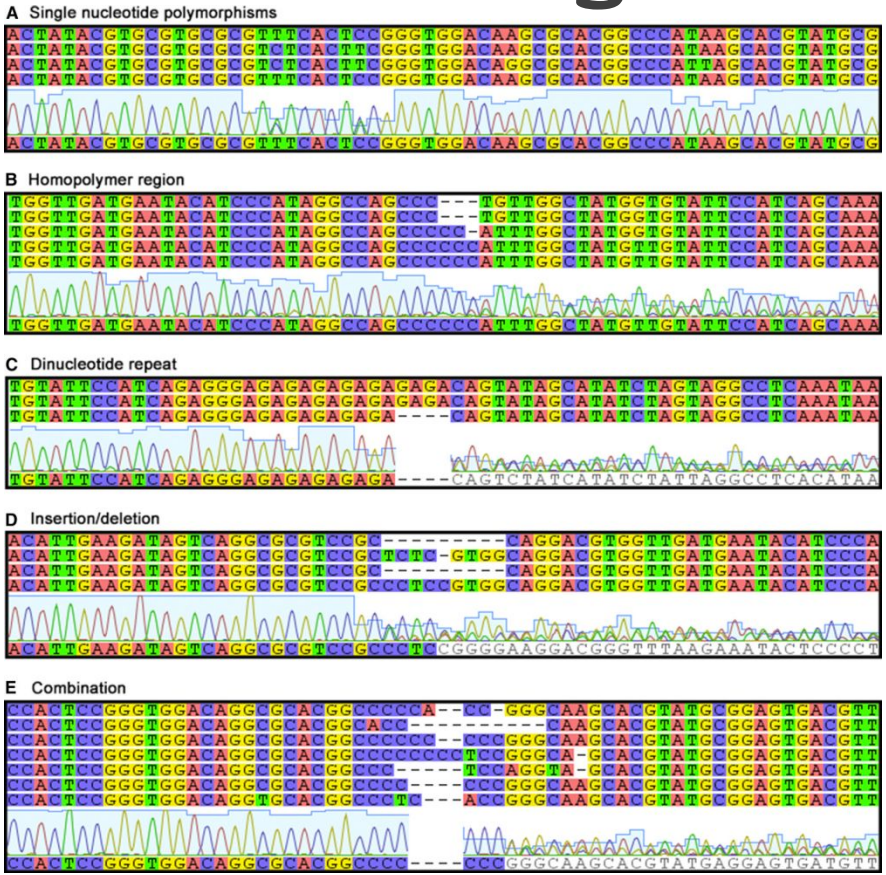
Population genomics

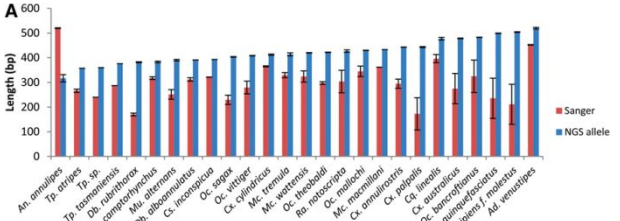
- Whole individual genome comparisons would give the highest resolution
- In *D. melanogaster*, >6,000,000 natural variants (SNPs and indels) have been described (Huang et al. 2014).

>1000 fruit fly genomes



Sanger vs NGS

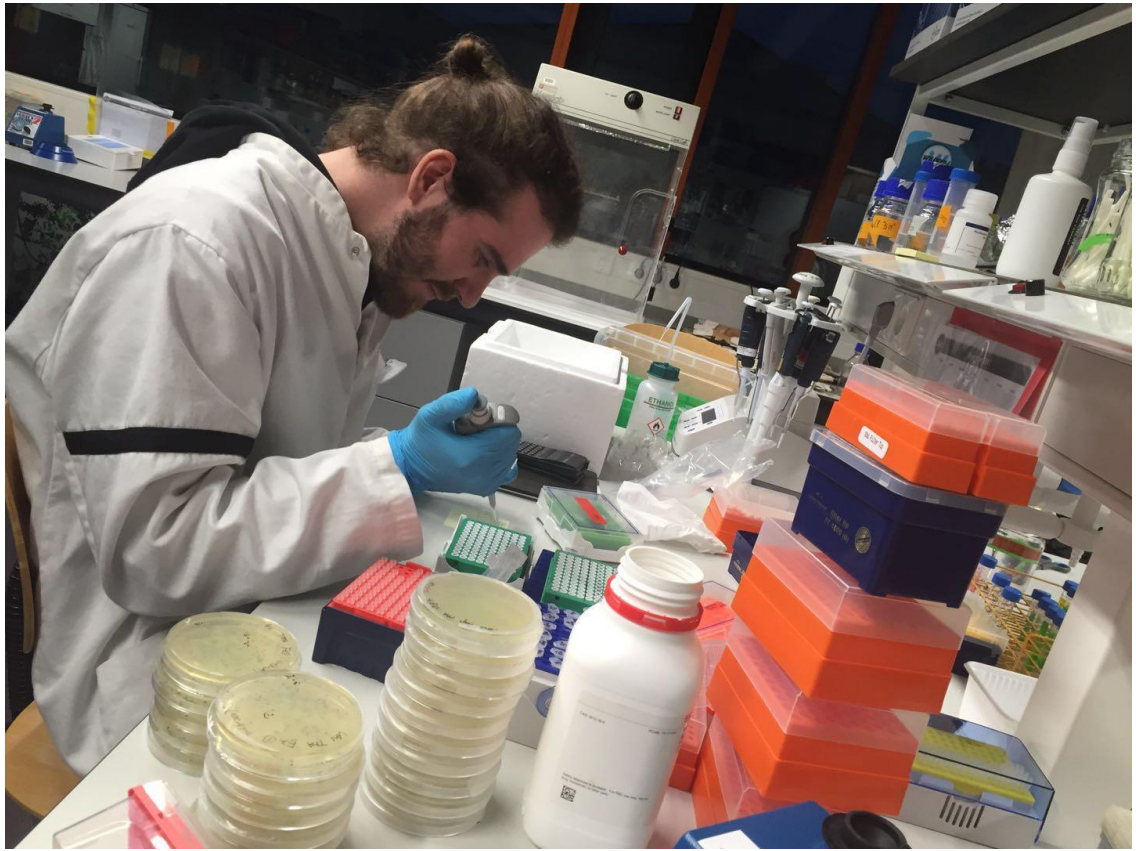




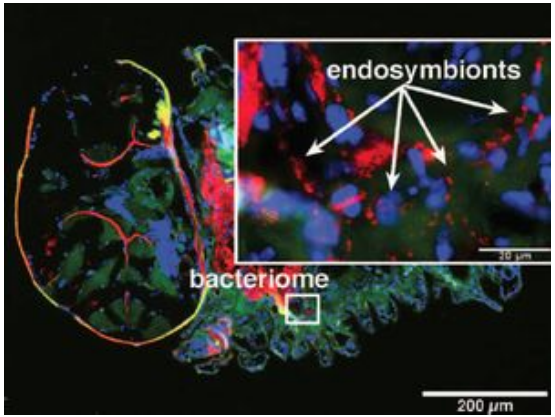
Batovska et al. 2017, G3

Newest sequencing technologies

Need lab, equipment



Population dynamics in greenhouse



What about fieldwork?

Barcode sequencing

Sequence in the field

ONT MinION

Oxford Nanopore Technology:
“Our goal is to enable the analysis of any living thing, by any person, in any environment.”

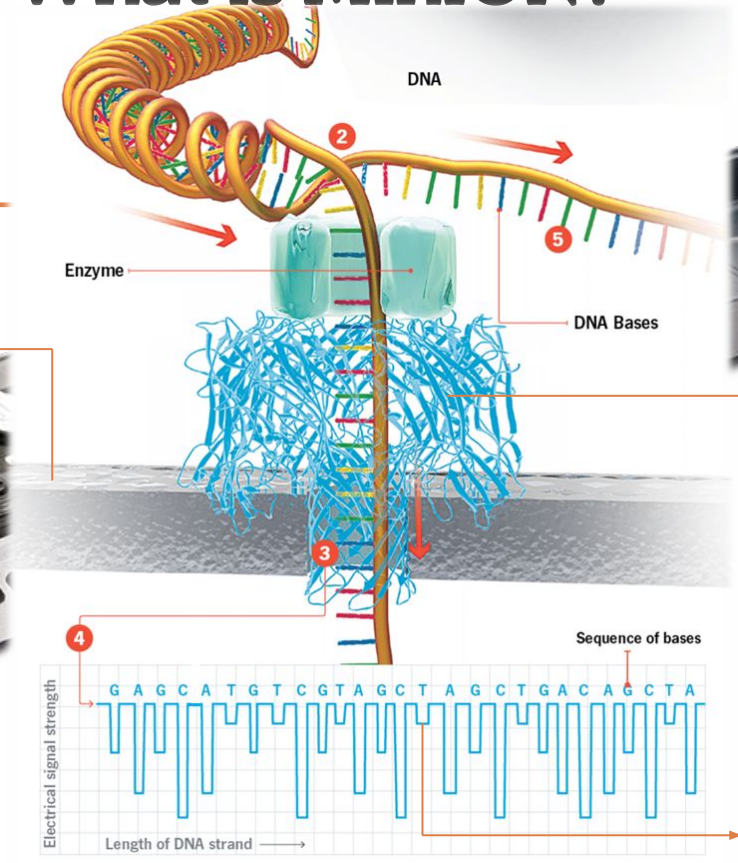
In the rainforest

Menegon et al. 2017, Plos One

What is MinION?

Flowcell

Membrane



Nanopore

Electrical signal

Capabilities

- flow cell generates 10–20 Gb

Long reads	Real time	Easy, rapid prep	On demand	Accessible
Choose your read length: 5kb? >200kb? Longer?	Immediate access to data	1D library prep: <10mins, 2 pots	Run different experiments in sequence on one flow cell	No capital cost
Easier assembly, phasing	Rapid time to result – move on	Low cost of materials	Barcode for even more samples	Easy install
Covering repetitive regions	Rapid insight into whether status of sample	De-skill prep	Run many experiments on one device	No additional lab infrastructure requirements

MinION uses

Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick, Nicholas J. Loman ✉ […] Miles W. Carroll

Nature **530**, 228–232 (11 February 2016)
doi:10.1038/nature16996

Received: 18 November 2015
Accepted: 15 January 2016

Profiling bacterial communities by MinION sequencing of ribosomal operons

Lee J. Kerkhof, Kevin P. Dillon, Max M. Häggblom and Lora R. McGuinness

Microbiome 2017 **5**:116

https://doi.org/10.1186/s40168-017-0336-9 © The Author(s). 2017

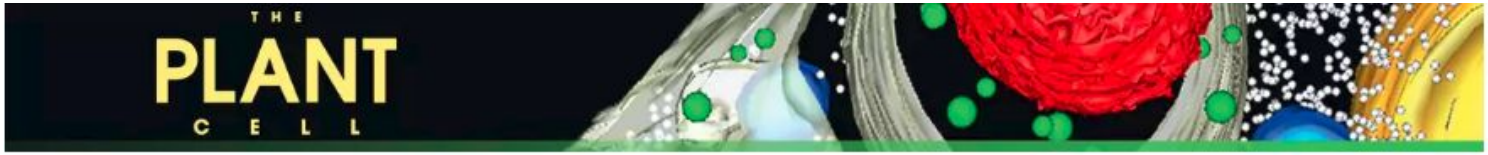
Received: 25 April 2017 | Accepted: 30 August 2017 | Published: 15 September 2017

Real-time DNA barcoding in a remote rainforest using nanopore sequencing

Aaron Pomerantz, Nicolas Penafiel, Alejandro Arteaga, Lucas Bustamante, Frank Pichardo, Luis A Coloma, Cesar L Barrio-Amoros, David Salazar-Valenzuela, Stefan Prost

doi: https://doi.org/10.1101/189159

This article is a preprint and has not been peer-reviewed [what does this mean?].



De Novo Assembly of a New *Solanum pennellii* Accession Using Nanopore Sequencing

Maximilian H.-W. Schmidt, Alexander Vogel, Alisandra K. Denton, Benjamin Istace, Alexandra Wormit, Henri van de Geest, Marie E. Bolger, Saleh Alseekh, Janina Maß, Christian Pfaff, Ulrich Schurr, Roger Chetelat, Florian Maumus, Jean-Marc Aury, Sergey Koren, Alisdair R. Fernie, Dani Zamir, Anthony M. Bolger, Björn Usadel

Published October 2017. DOI: https://doi.org/10.1105/tpc.17.00521 Altmetric 120

- Obtained an error rate of <0.02%
- Reached a gene completeness of 96.53%
- Complemented with Illumina sequences for polishing
- 31 flowcells yielded 134.8 Gb of data in total,
- 110.96 Gb (representing ~100-fold coverage) after filtering
- Total yield per flowcell varied 0.96 and 6.02 Gb after filtering.

Explore extreme conditions

Nanopore DNA Sequencing and Genome Assembly on the International Space Station

Deep Sequen
Off-Grid Nan

Arwyn Edwards, A

Extren
Svalba

Arwyn E

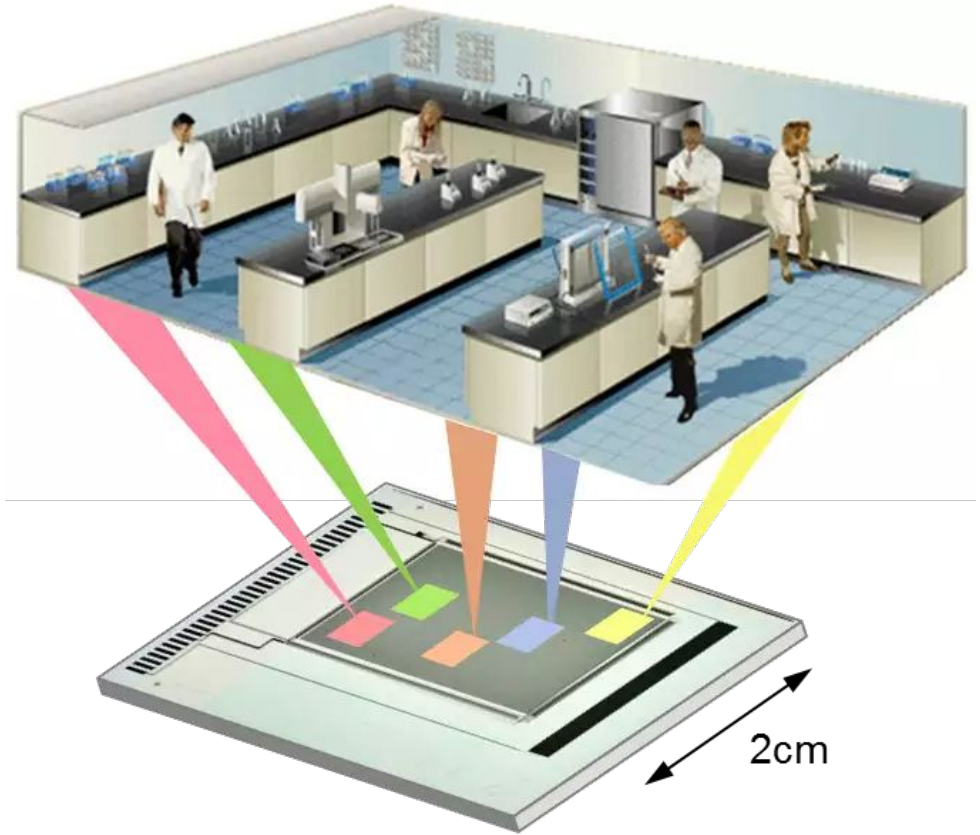
Real-Time DNA
Nanopore Seque

Sarah S. Johnson,[1]

Comparison of sequencing techniques

	Error	Read length	Amplification	Cost/bp	Portability	Flexibility
Illumina	<1%	<800bp	Yes	+	-	-
Pacbio	≈10%	Average 10,000bp Rare 60,000bp	Yes	-	-	-
MinION	≈10%	Unlimited, longest so far 950Kb	No	--	++	++

Lab-on-a-chip



VolTRAX: library prep on a chip

VOITRAX
VOITRAX

SmidgION

Issue 15, 2012

From the journal:
Lab on a Chip

Integrated rapid-diagnostic-test reader platform on a cellphone

Onur Mudanyali,[a] Stoyan Dimitrov,[a] Uzair Sikora,[a] Swati Padmanabhan,[a] Isa Navruz[a] and Aydogan Ozcan*[abc]

Biosensors and Bioelectronics 26 (2011) 4070–4075

Contents lists available at ScienceDirect

Biosensors and Bioelectronics

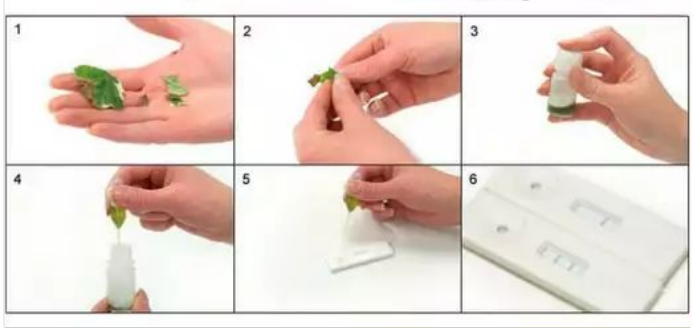
journal homepage: www.elsevier.com/locate/bios

Development of a lab-on-a-chip device for diagnosis of plant pathogens

Sandra Julich[a], Marko Riedel[b,1], Mark Kielpinski[a], Matthias Urban[a], Robert Kretschmer[c,d], Stefan Wagner[b], Wolfgang Fritzsche[a], Thomas Henkel[a], Robert Möller[c,e,*], Sabine Werres[b,**]

POCKET® DIAGNOSTIC

Plant disease test kits

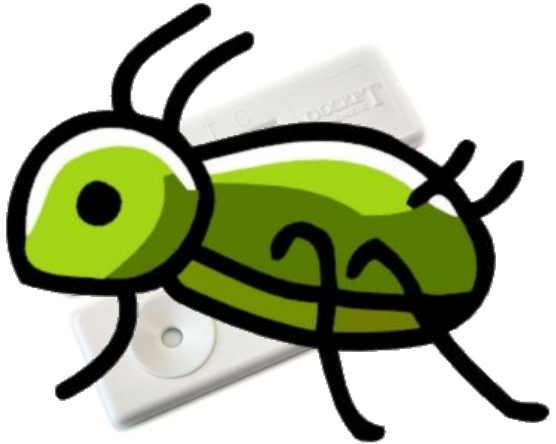


Rapid Positive/Negative results in the field or laboratory

"Pregnancy test" for

- Erwinia amylovora
- Potato virus Y
- Ralstonia solanacearum
- Phytophthora
- Mycotoxins

- Also used for PCR fragment analysis



Future: Biocontrol tricorder

Questions?