DNA quality requirements for Single-Molecule sequencing

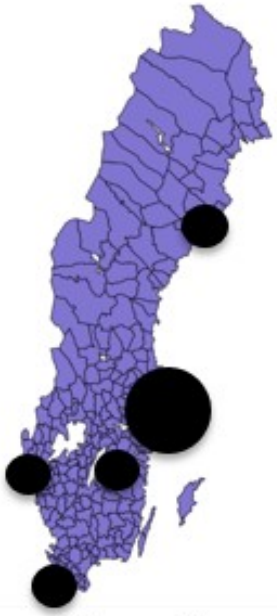
Olga Vinnere Pettersson, PhD

Project coordinator

NGI-Sweden / SciLifeLab (UU)

SWEDEN

Stockholm

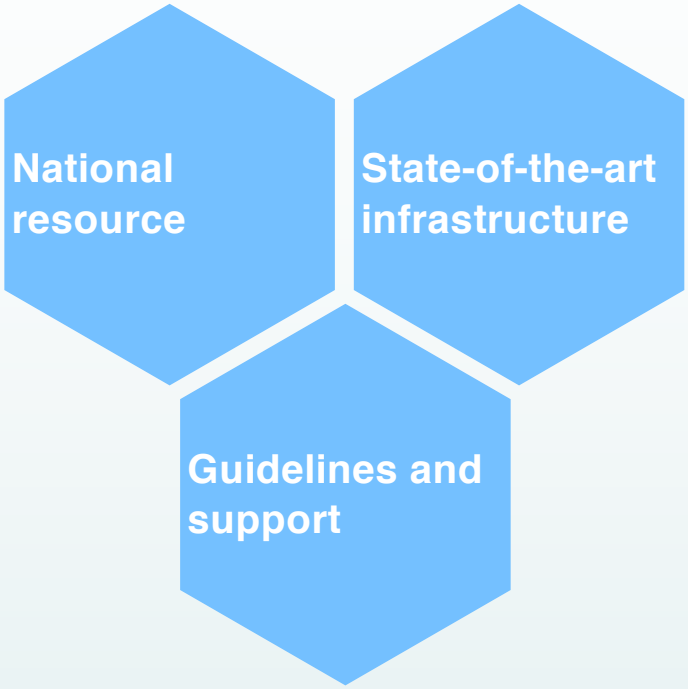


Umeå
Linköping
Gothenburg
Lund

Uppsala

NGI: Mission statement

Since Jan 1, 2013, National Genomics Infrastructure (NGI)
is a **national resource** for next generation sequencing (NGS)



Our mission

- To make a **state-of-the-art infrastructure** for massively parallel NGS and SNP genotyping available to researchers all over Sweden enabling internationally competitive research in genomics.

- To provide **guidelines and support** for study design, sample collection, protocol selection and bioinformatic analysis.

NATIONAL GENOMICS INFRASTRUCTURE

SciLifeLab

NATIONAL GENOMICS INFRASTRUCTURE

"This research infrastructure is world class and a jewel in the crown of Swedish bioscience." (Swedish Research Council)

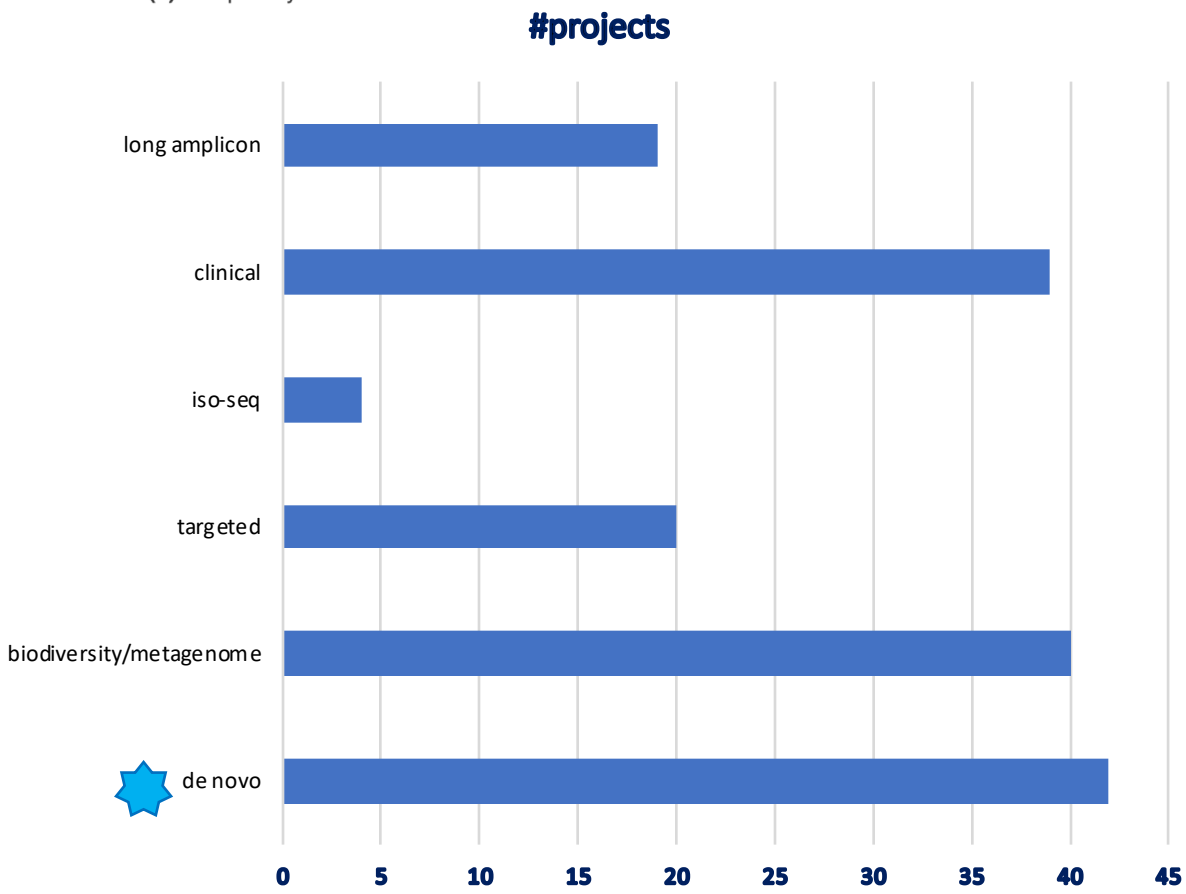
SciLifeLab

SMRT smörgåsbord at NGI: every project is unique

Uppsala, Sweden

National Genomics Infrastructure hosted @ SciLifeLab

Phone: +46-70-4250134
Email: olga.pettersson@igp.uu.se
Instrument(s): Sequel System



154 projects in 2018

Protists

Bacteria

Fungi

Insects

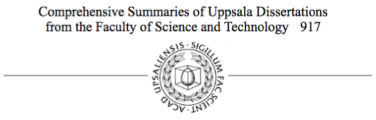
Plants

Mammals

Arachnids

Who am I to judge…

1997 – 2004 Mycologist by training (plant pathogenic fungi)

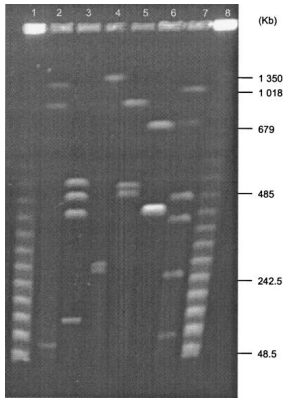


Approaches to Species Delineation in Anamorphic (mitosporic) Fungi: A Study on Two Extreme Cases

BY

OLGA VINNERE

2004 – 2007 Genomic architecture of Bartonellae
(HMW-DNA & PFGE)



2007 – 2012 Fungal genomics (extremophiles)

Research article | Open Access

Genome and physiology of the ascomycete filamentous fungus *Xeromyces bisporus*, the most xerophilic organism isolated to date

Su-lin L. Leong, Henrik Lantz, Olga V. Pettersson, Jens C. Frisvad, Ulf Thrane, Hermann J. Heipieper, Jan Dijksterhuis, Manfred Grabherr, Mats Pettersson, Christian Tellgren-Roth, Johan Schnürer

First published: 20 August 2014 | https://doi.org/10.1111/1462-2920.12596 | Cited by: 14

2012 – current time: Project Coordinator at NGI
over 600 projects
circa 200 *de novo* projects on PacBio and ONT

RESEARCH ARTICLE Open Access

A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds

Andreas Wallberg, Ignas Bunikis, Olga Vinnere Pettersson, Mai-Britt Mosbech, Anna K. Childers, Jay D. Evans, Alexander S. Mikheyev, Hugh M. Robertson, Gene E. Robinson and Matthew T. Webster

Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications

Matthias H. Weissensteiner, Andy W.C. Pang, Ignas Bunikis, Ida Höijer, Olga Vinnere-Petterson, Alexander Suh and Jochen B.W. Wolf

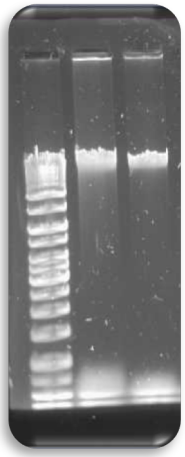
TECHNICAL NOTE Open Access

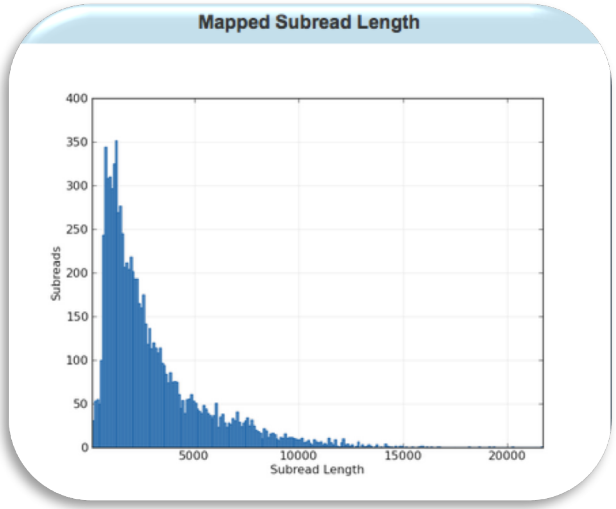
De novo assembly of *Dekkera bruxellensis*: a multi technology approach using short and long-read sequencing and optical mapping

Remi-Andre Olsen, Ignas Bunikis, Ievgeniia Tiukova, Kicki Holmberg, Britta Lötstedt, Olga Vinnere Pettersson, Volkmar Passoth, Max Käller and Francesco Vezzi

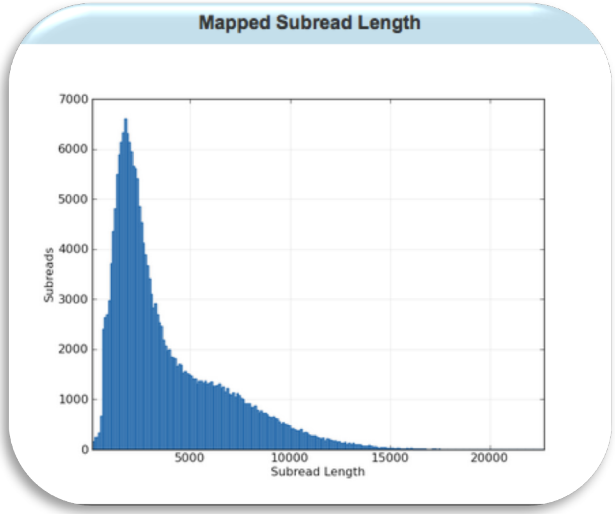
2018 – current time: member of the VGP sample-prep group

2013: a wake-up call





Polished Contigs	223	Max Contig Length	36,298
N50 Contig Length	2,932	Sum of Contig Lengths	480,087



Polished Contigs	9	Max Contig Length	1,508,929
N50 Contig Length	1,353,702	Sum of Contig Lengths	7,813,244

For Long Reads one needs to have *long and pure* DNA

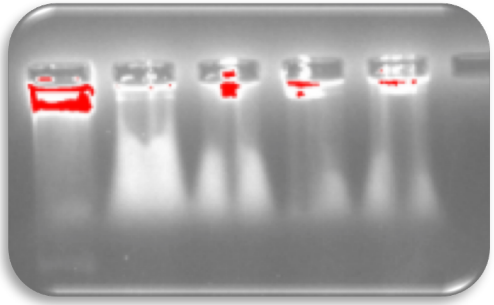
SciLifeLab

”Easy” example: *de novo, Corvus sp.*

Genome: 1.2 Gb (2n)
Application: *de novo*
Sequencing: 20 kb insert size
60x (30x per strand)
6 hour movies
Sequel I

Sample requirements:
HMW DNA
10 ug minimum
260/280 = 1.8 – 2.0
260/230 = 2.0 – 2.2

May 15, 2015

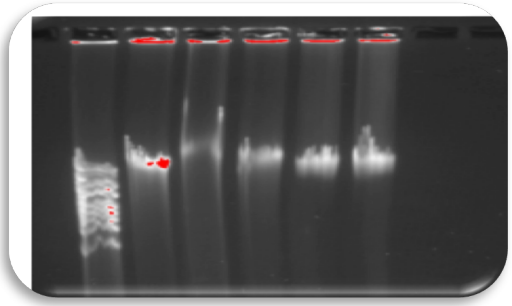


260/280 = 0.78
260/230 = 1.64

74 e-mails

Intermediate QC

Jun 16, 2015



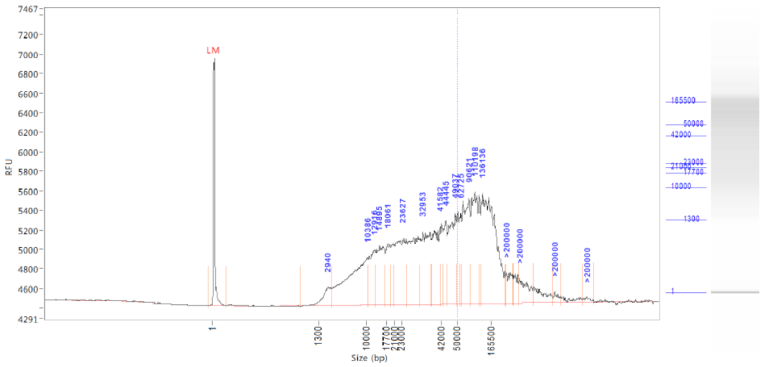
260/280 = 1.86
260/230 = 2.24

Real life examples

Corvus sp.

CONTIGS	Primary	Associated	Unzipped Primary	Unzipped haplotigs
# contigs	2996	2136	1481	6349
# >50Kb	457	586	354	3193
Largest	52,5Mb	0,18Mb	52,6Mb	3,11Mb
N50	11,4Mb	0,05Mb	12,1Mb	0,42Mb
Total	1,12Gb	97,8Mb	1,07Gb	1,01Gb

20 kb library – perfect DNA
60x

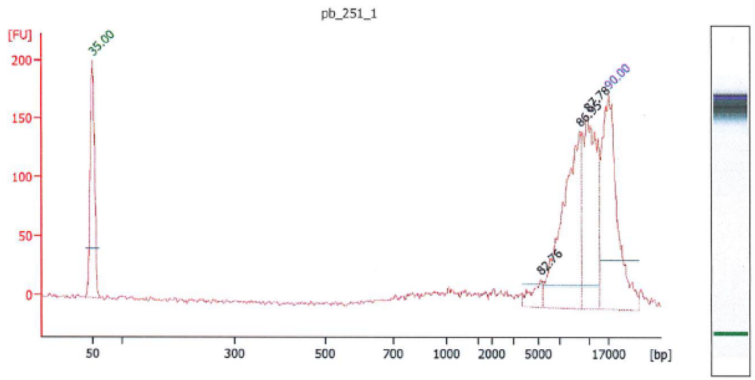


Bird 2

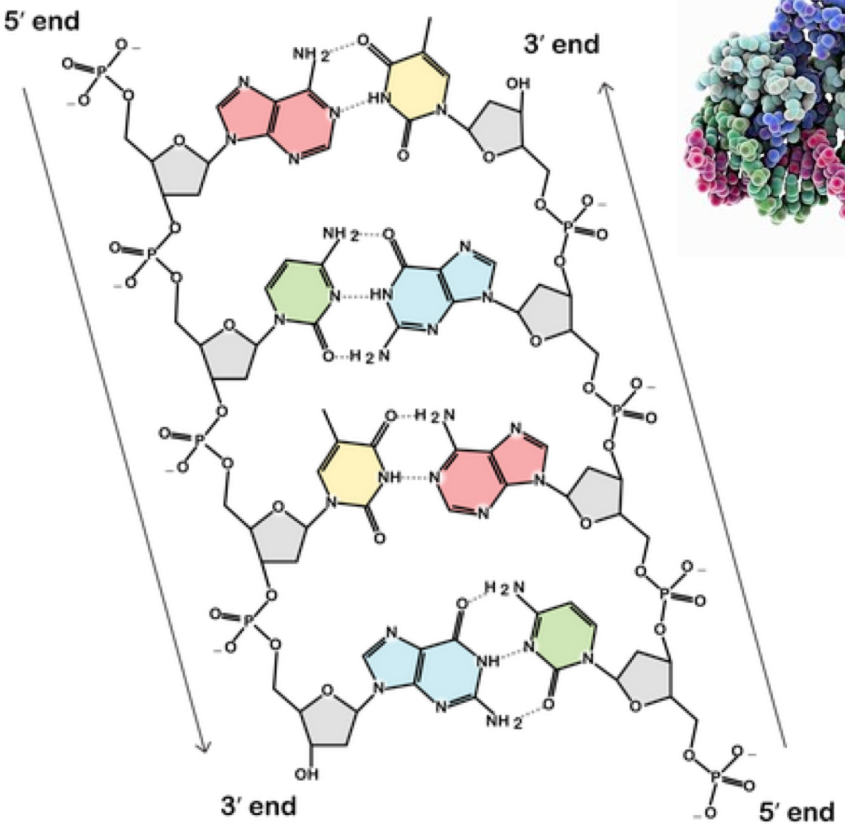
SCAFFOLDS	Unphased
# scaffolds	38231
# >50Kb	201
Largest	2,01 Mb
N50	0,21 Mb
Total	1,21Gb

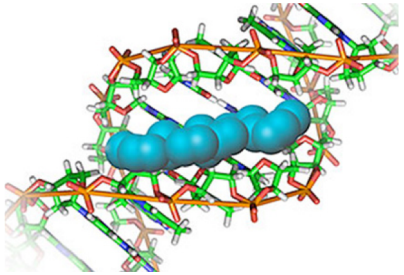
User insisted to proceed despite warning

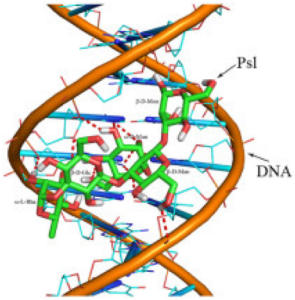
10 kb library – short, dirty & little DNA
36X



Focus on chemical purity: sorting out the culprits







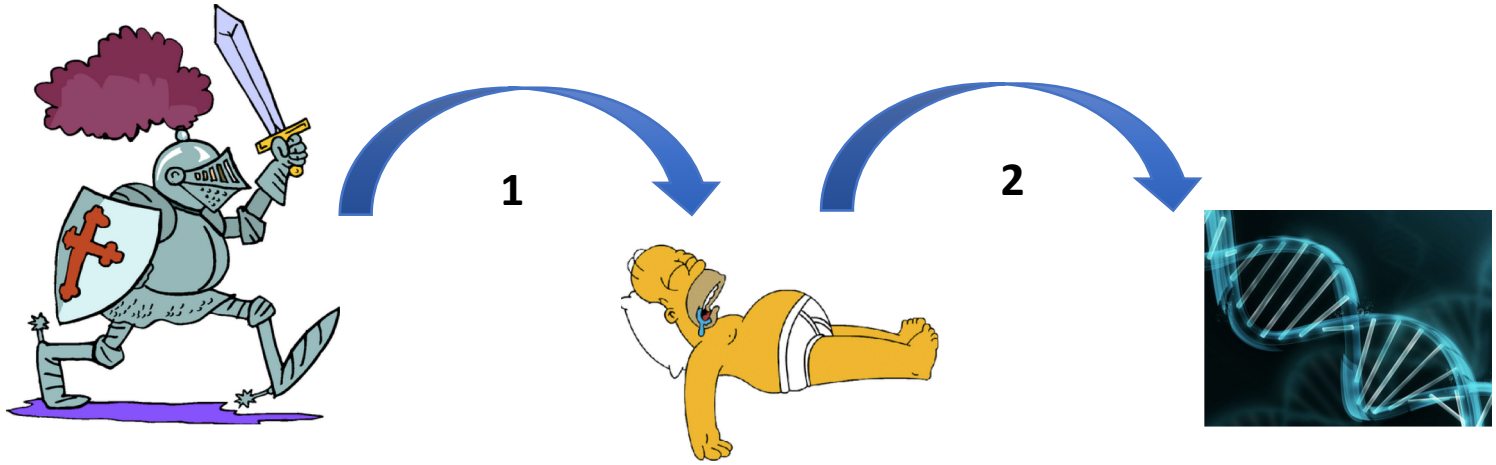
1. **Carry-over from host cells**
   - you name it…

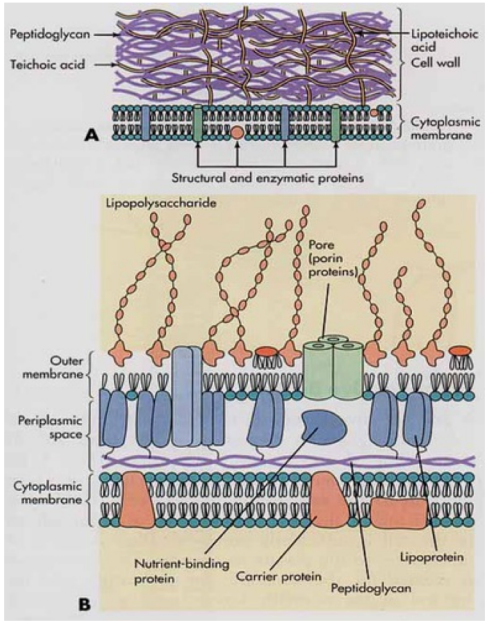
2. **Carry-over of extraction chemicals**
   - Phenol
   - Guanidinium
   - Ethanol
   - EDTA

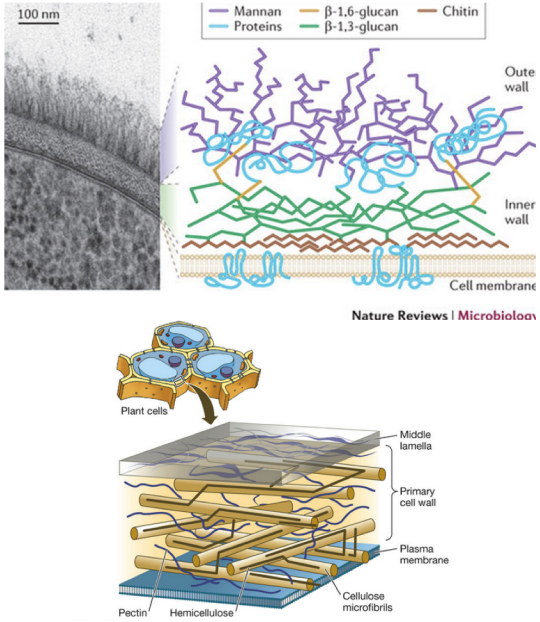
Most have dual action:
- Enzyme inhibition
- DNA-binding

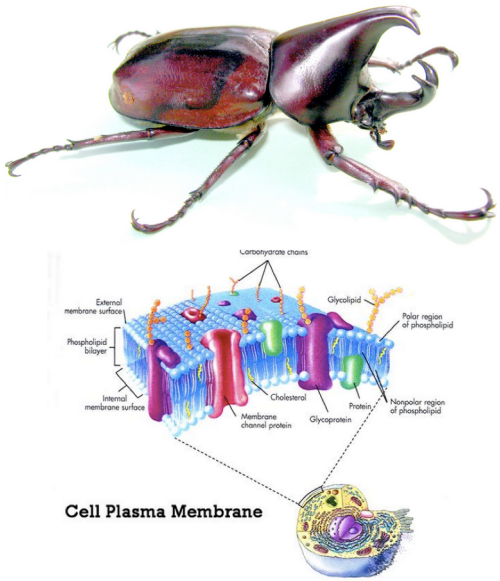
"DNA chemical structure," by Madeleine Price Ball

The DNA extraction process









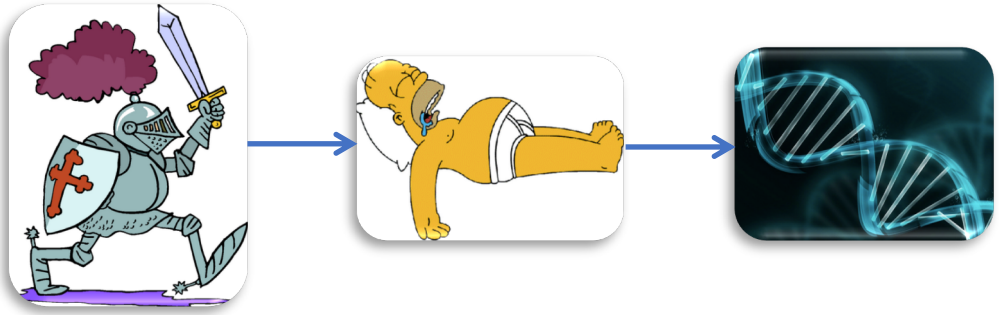
Sources of DNA contamination

Carry-over
from DNA extraction:

- Native cell wall components
- RNA & proteins
- Secondary metabolites

- Phenol
- Salts
- Ethanol

goo.gl/u8OiGb



Bacteria: LPS, secondary metabolites

Fungi: chitin, secondary metabolites,
proteins, pigments, polysaccharides

Plants: polyphenols and other aromatics,
polysaccharides,
secondary metabolites, pigments

Insects: proteins, chitin, pigments

Animals: tissue specific

What do absorption ratios tell us?

Pure DNA 260/280: 1.8 – 2.0

< 1.8:

Too little DNA compared to other components of the solution; presence of organic contaminants: proteins and phenol; glycogen - **absorb at 280 nm**.

> 2.0:

High share of RNA.

Pure DNA 260/230: 2.0 – 2.2

<2.0:

Salt contamination, humic acids, peptides, aromatic compounds, polyphenols, urea, guanidine, thiocyanates (latter three are common kit components) – **absorb at 230 nm**.

>2.2:

High share of RNA, very high share of phenol, **high turbidity,** dirty instrument, wrong blank.

Focus on chemical purity: why so important?

SciLifeLab

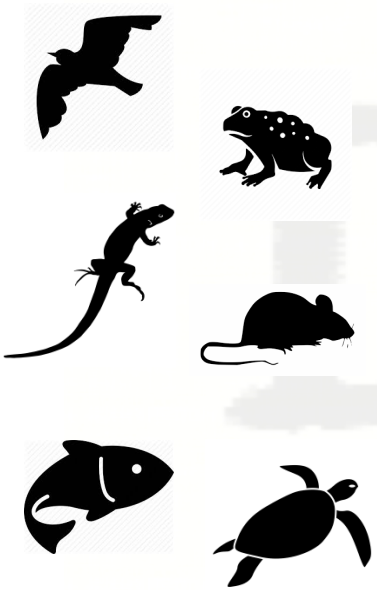
Lower chemical purity = worse loading and shorter reads,

***ergo* higher sequencing and analysis costs.**

Organism type, *de novo* application, 60x	2 Gb / SMRT	4 Gb / SMRT	7 Gb / SMRT
Bacterium (3.2 Mb)	2-4 strains	6-8 strains	10-12 strains
Insect (300 Mb)	N cells: 9 Price: 10 kEUR	N cells: 5 Price: 6 kEUR	N cells: 3 Price: 4 kEUR
Bird (1.2 Gb)	N cells: 43 Price: 47 kEUR	N cells: 22 Price: 25 kEUR	N cells: 12 Price: 14 kEUR
Mammal (3.2 Gb)	N cells: 96 Price: 105 kEUR	N cells: 48 Price: 53 kEUR	N cells: 27 Price: 30 kEUR

Numbers are old (Sequel I, V2 chem), but the problem remains!

Earth Biogenome Project: VGP sample prep group



Tissues
Collection method
Preservation
Extraction
QC
Storage

VERTEBRATE GENOMES PROJECT
A PROJECT OF THE G10K CONSORTIUM

3.4.2.QV40 PHASED METRICS

Insects and DNA extraction

If possible: use pupae
If dissectable imago: use thorax, leg muscles, male genitalia
Avoid gut at all cost

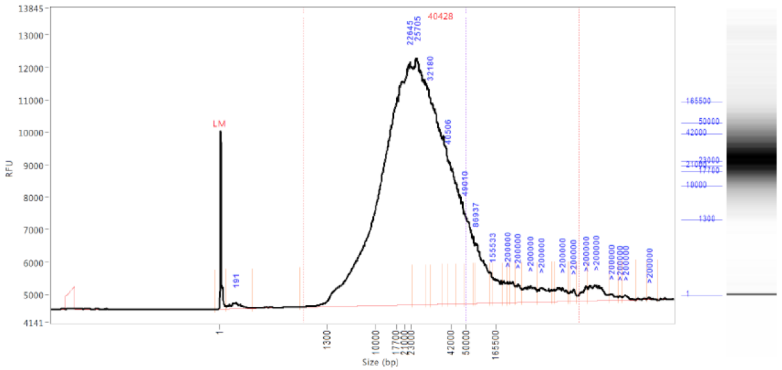
Use the whole body *only if there is no other choice*
shorturl.at/rsuvR
Atuta.slack.com (Mara Lawniczak)

Insect (large body, large genome) example

QC stats	Insect 1	Insect 2
260/280 260/230	2.0 1.8	3.5 1.83
Char. fragment length, FEMTO	25 kb	14 kb
Insert size	25 kb	9.5 kb
Size-selection	17 kb	8 kb
Loading, per SMRT	7 Gb	6.5 Gb
Read N50	17 kb	8 kb
#contigs	279	21 808
Longest contig	17.4 Mb	2.4 Mb
Contig N50	7.9 Mb	281 kb

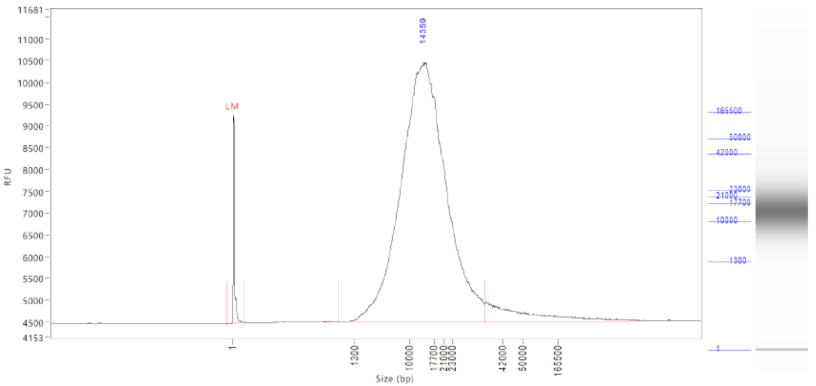
Insect 1:

Pupae, flash-frozen,
agarose plug extraction



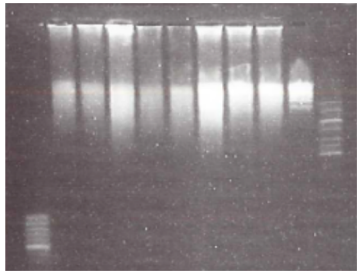
Insect 2:

Imago, frozen in 95% EtOH,
leg muscles, MagAttract kit



A painful journey of seed beetles

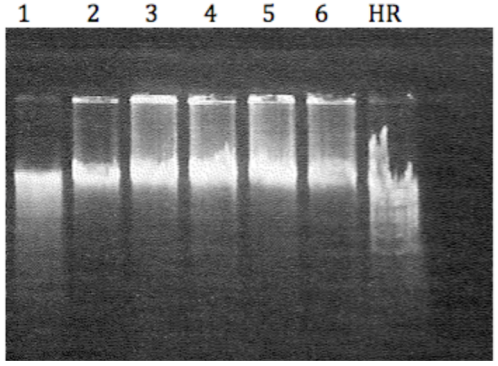
1: High-salt / ethanol protocol, whole body -> black DNA -> **FAIL**
2. MagAttract, entire body ->
260/280 = 2.1 – 2.2
260/230 = 0.32 - 0.45 -> **FAIL**
3. MagAttract, muscle -> A260 within range



-> **FAIL**

Callosobruchus maculatus

4. MagAttract -> Zymo DNA purification -> A260 low -> **FAIL**
5. High-salt / ethanol -> Zymo DNA purification -> A260 low -> **FAIL**
6. GenomicTip 20G stand-alone, muscle -> A260 in range -> **FINALLY!!!**



The genomic footprint of sexual conflict

Ahmed Sayadi, Alvaro Martinez Barrio, Elina Immonen, Jacques Dainat, David Berger, Christian Tellgren-Roth, Björn Nystedt & Göran Arnqvist

Nature Ecology & Evolution **3**, 1725–1730(2019) | Cite this article

2467 Accesses | **75** Altmetric | Metrics

The Evolution of Dark Matter in the Mitogenome of Seed Beetles

Ahmed Sayadi, Elina Immonen, Christian Tellgren-Roth, Göran Arnqvist Author Notes

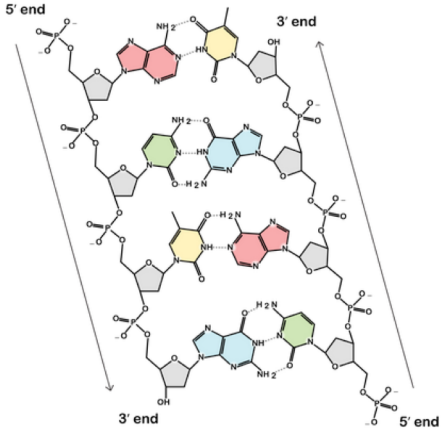
Genome Biology and Evolution, Volume 9, Issue 10, October 2017, Pages 2697–2706, https://doi.org/10.1093/gbe/evx205

Published: 27 September 2017 **Article history**

A journey to HMW-DNA world

SciLifeLab

HMW-DNA behaves very differently from LMW-DNA



High viscosity:

1. **Bad hydration** -> topological issues
-> getting HMW-DNA in solution without any concentration gradient is an issue

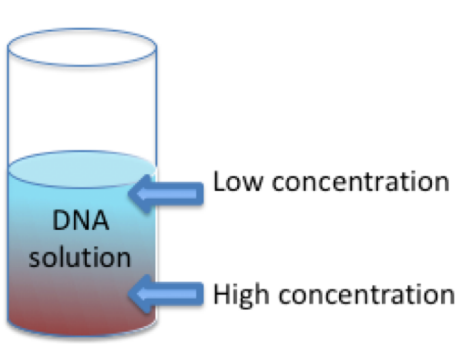
2. **Presence of contaminants**

Getting ”correct” readings of HMW-DNA

SciLifeLab

DNA-spa: let the molecules relax!

- 3-7 days at RT or +4°C
- Gentle agitation
- Playing with ionic strength

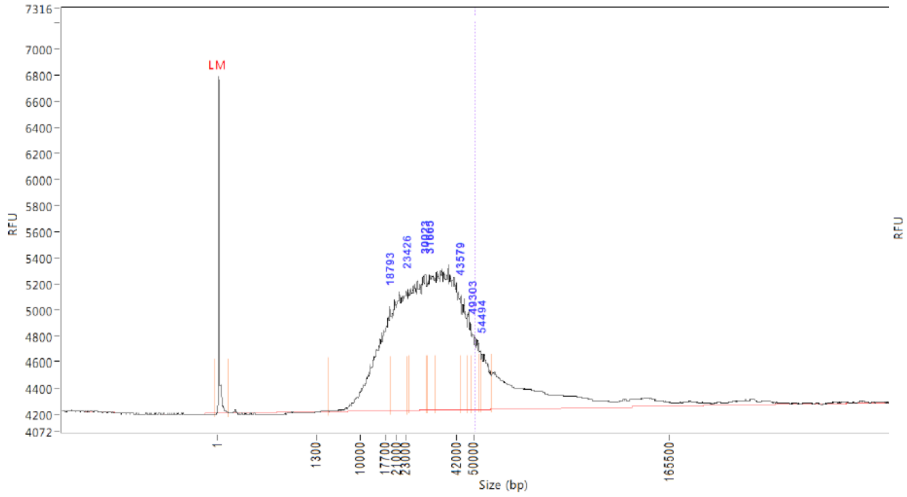


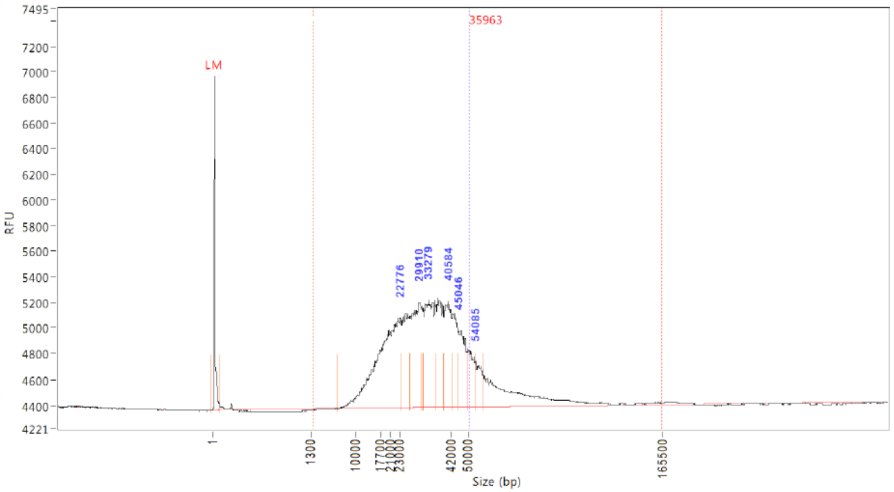
LMW-DNA

HMW-DNA

HMW-DNA	Read N50, kb
E.coli in situ - fresh	13.3
E.coli in situ - relaxed	22.3

FemtoPulse measurements: several replicates (and dilution series)





Causes of DNA degradation

Mechanical damage during tissue homogenization.

Wrong pH and ionic strength of extraction buffer (-> hydrolysis).

Incomplete removal / contamination with **nucleases**.

Phenol: too old, or inappropriately buffered (**pH 7.8 – 8.0**); incomplete removal.

Wrong pH of the **DNA solvent** (acidic water).
Recommended: Low TE for short-term storage, 1xTE for long-term storage.

Vigorous pipetting (wide-bore pipet tips).

Vortexing of DNA in high concentrations.

Too many **freeze-thaw** cycles (*we tested 5, still Ok*).

Sequence-dependency

General recommendations

Treat DNA as a crystal vase: it is fragile when in solution

As soon as DNA is released from the cells – use **wide-bore** tips

Limit pipetting to minimum

Never vortex!

Do not heat above 65°C

Reduce amount of freeze-thaw cycles

Store at maximum -70°C, in TE-buffer

NEVER store DNA in water

ALWAYS were gloves when handling tubes

If shipping: solid frozen on dry ice

Acknowledgements

Mai-Britt Mossbech
Tomas Klingström

Jamshid Fatehi

Ulf Gyllensten