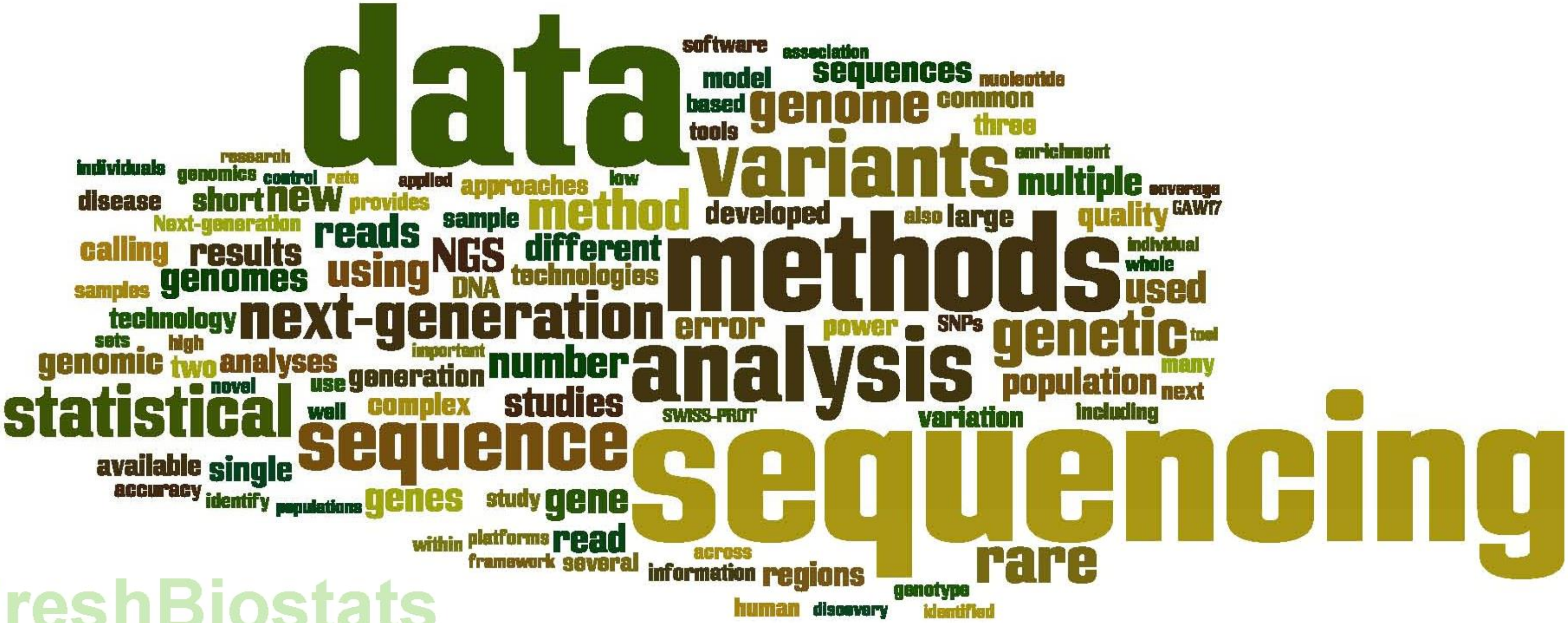
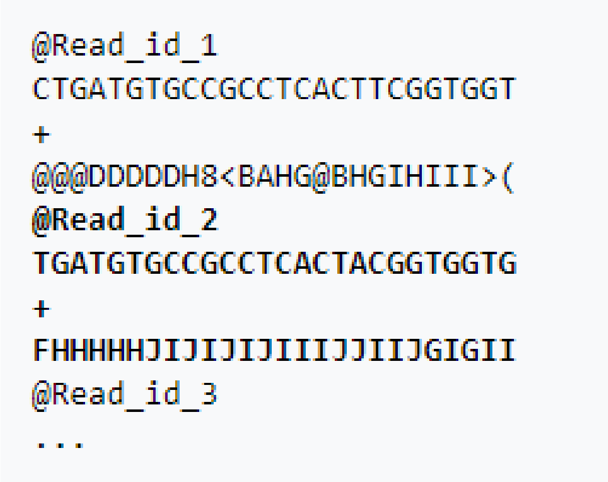
CAB

KASETSART UNIVERSITY

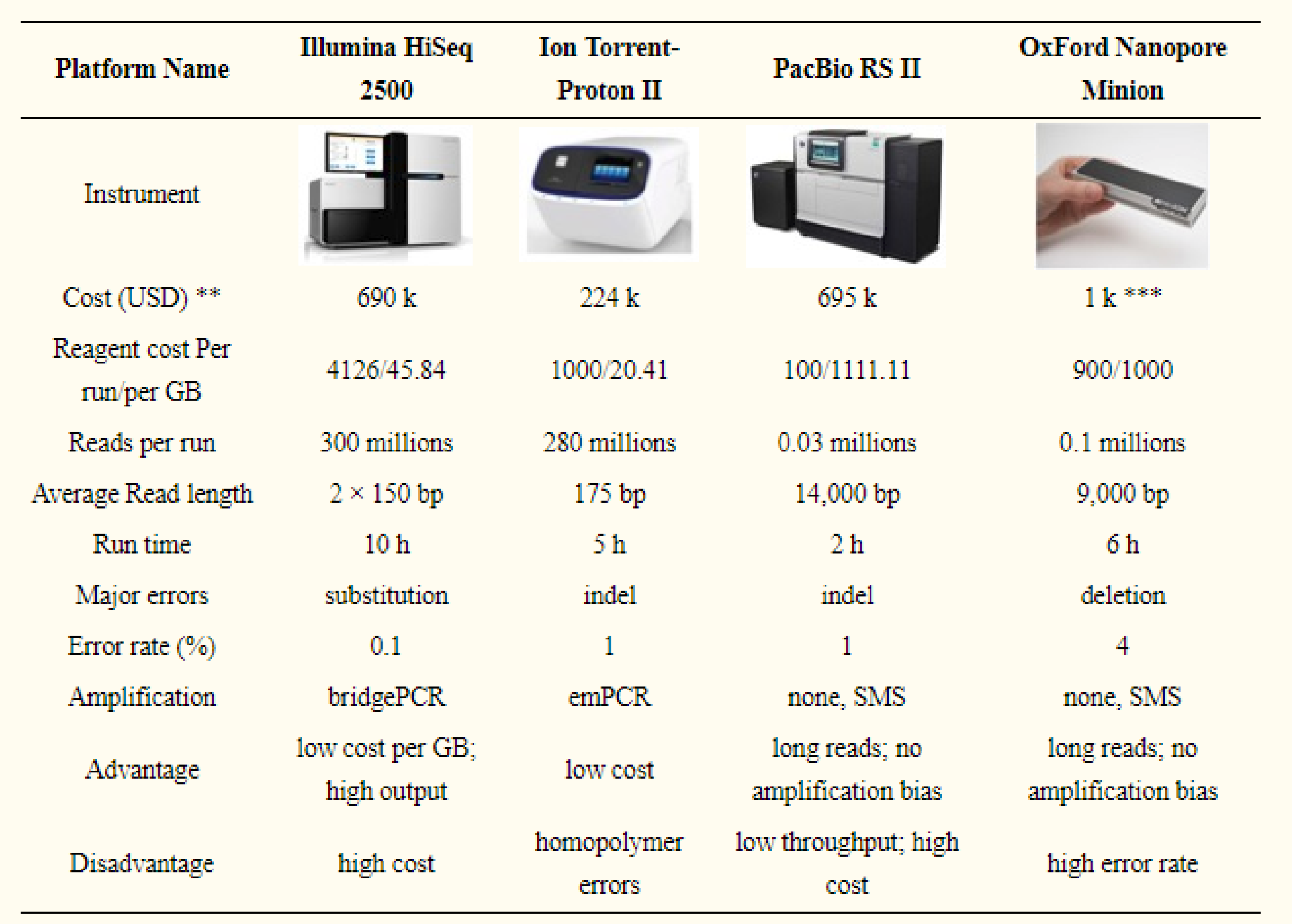
→ → →


data
software
association
model
sequences
nucleotide
based
genome
common
tools
three
research
individuals
genomics
control
rate
applied
variants
enrichment
disease
short
new
provides
approaches
low
multiple
coverage
Next-generation
sample
method
developed
also
large
quality
GAW17
calling
reads
results
NGS
different
methods
individual
whole
samples
genomes
using
DNA
technologies
used
technology
next-generation
error
power
SNPs
genetic
tool
sets
high
important
analysis
genomic
two
analyses
number
many
novel
use
generation
population
next
statistical
well
complex
studies
SWISS-PROT
variation
including
sequence
sequencing
available
single
accuracy
identify
populations
genes
study
gene
within
platforms
read
rare
across
framework
several
information
regions
genotype
human
discovery
identified

- ***short reads*** *short read pairs*
- *Fastq*
-
-



```
@Read_id_1
CTGATGTGCCGCCTCACTTCGGTGGT
+
@@@DDDDDH8<BAHG@BHGIHIII>(
@Read_id_2
TGATGTGCCGCCTCACTACGGTGGTG
+
FHHHHHJIJIJIJIIIJJIIJGIGII
@Read_id_3
...
```

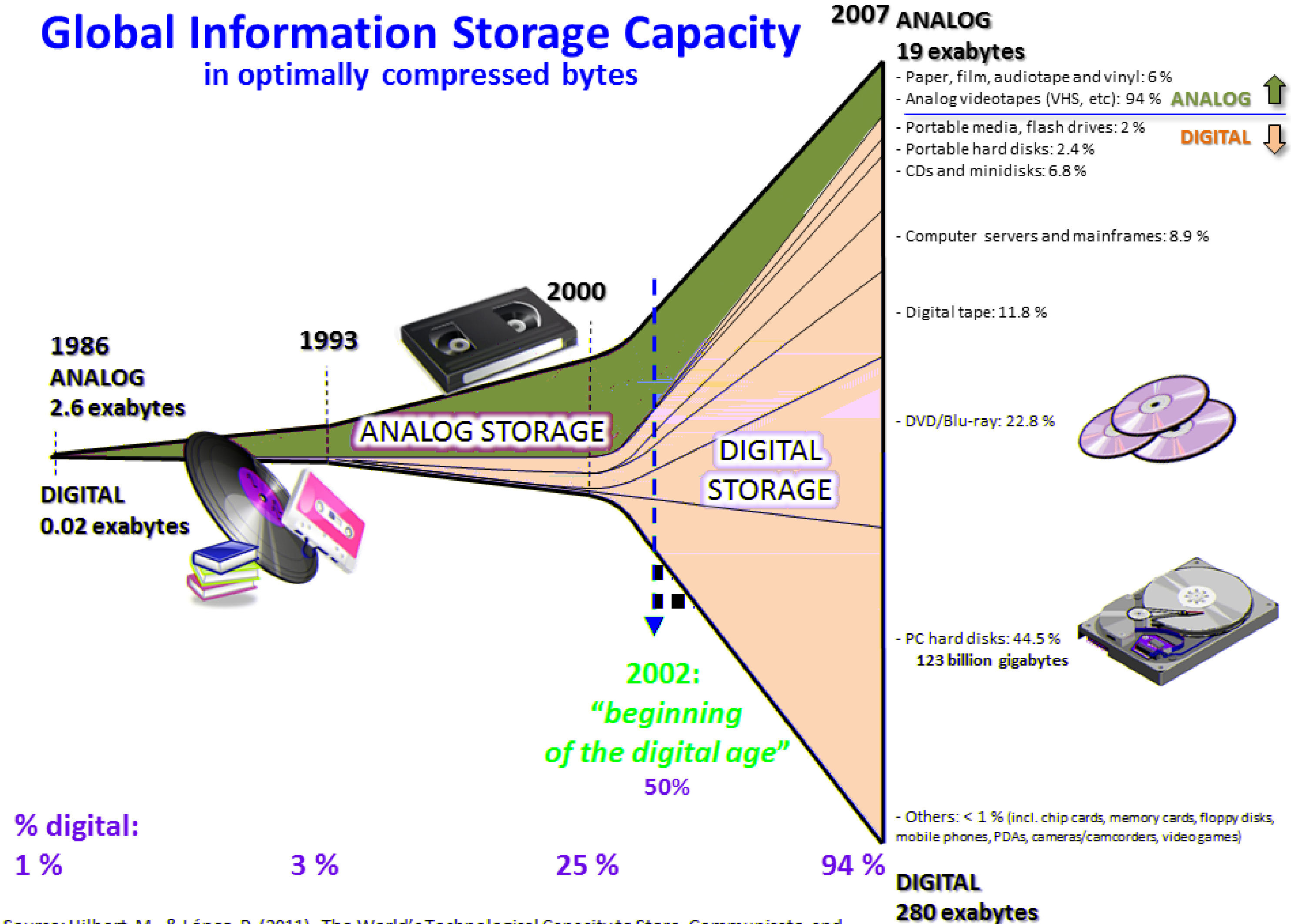


Platform Name	Illumina HiSeq 2500	Ion Torrent-Proton II	PacBio RS II	OxFord Nanopore Minion
Instrument				
Cost (USD) **	690 k	224 k	695 k	1 k ***
Reagent cost Per run/per GB	4126/45.84	1000/20.41	100/1111.11	900/1000
Reads per run	300 millions	280 millions	0.03 millions	0.1 millions
Average Read length	2 × 150 bp	175 bp	14,000 bp	9,000 bp
Run time	10 h	5 h	2 h	6 h
Major errors	substitution	indel	indel	deletion
Error rate (%)	0.1	1	1	4
Amplification	bridgePCR	emPCR	none, SMS	none, SMS
Advantage	low cost per GB; high output	low cost	long reads; no amplification bias	long reads; no amplification bias
Disadvantage	high cost	homopolymer errors	low throughput; high cost	high error rate

Ye H, et al. Pharmaceutics. 2015 Nov 23;7(4):523-41

Global Information Storage Capacity

in optimally compressed bytes



ANALOG
19 exabytes

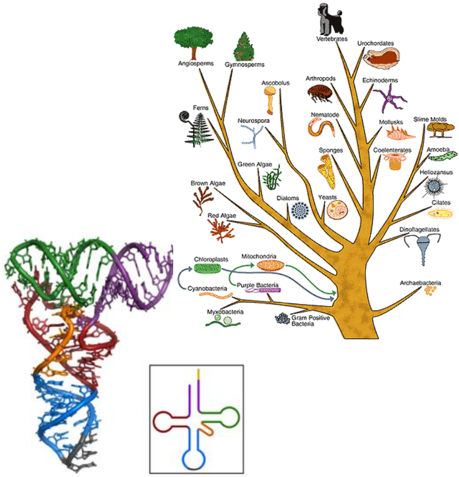
- Paper, film, audiotape and vinyl: 6 %
- Analog videotapes (VHS, etc): 94 % ANALOG ⇧

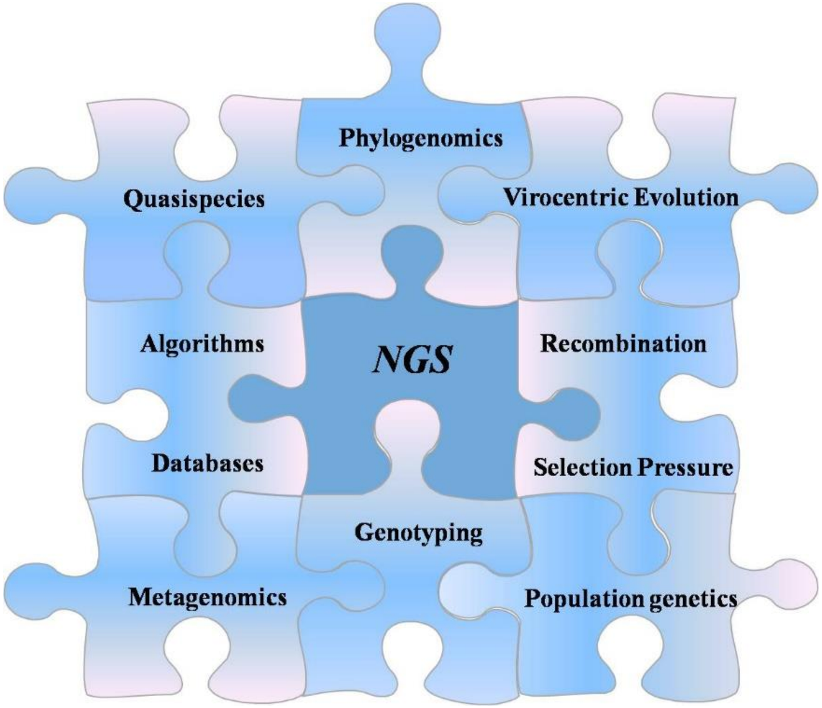
- Portable media, flash drives: 2 % DIGITAL ⇩
- Portable hard disks: 2.4 %
- CDs and minidisks: 6.8 %
- Computer servers and mainframes: 8.9 %
- Digital tape: 11.8 %
- DVD/Blu-ray: 22.8 %
- PC hard disks: 44.5 %
  123 billion gigabytes
- Others: < 1 % (incl. chip cards, memory cards, floppy disks, mobile phones, PDAs, cameras/camcorders, video games)

DIGITAL
280 exabytes

Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

regions
snp
secondary
metabolomics
homology
experiments
fasta
sites
molecular
pfam
storage
phylogenetics
alignment
structural
nlm
algorithm
databases
assembly
pathway
bioinformatics
genes
ngs
information
synteny
www
dna
protein
perl
data
sequence
access
codon
coding
simulations
genbank
annotation


Vertebrates
Urochordates
Angiosperms
Gymnosperms
Arthropods
Echinoderms
Ascobolus
Ferns
Nematode
Neurospora
Molluscs
Slime Molds
Sponges
Coelenterates
Amoeba
Green Algae
Heliozoansus
Brown Algae
Diatoms
Yeasts
Ciliates
Red Algae
Dinoflagellates
Chloroplasts
Mitochondria
Purple Bacteria
Archaebacteria
Cyanobacteria
Myxobacteria
Gram Positive Bacteria


Phylogenomics
Quasispecies
Virocentric Evolution
Algorithms
NGS
Recombination
Databases
Selection Pressure
Genotyping
Metagenomics
Population genetics

[redacted]

- *De novo*
-
-

De Novo

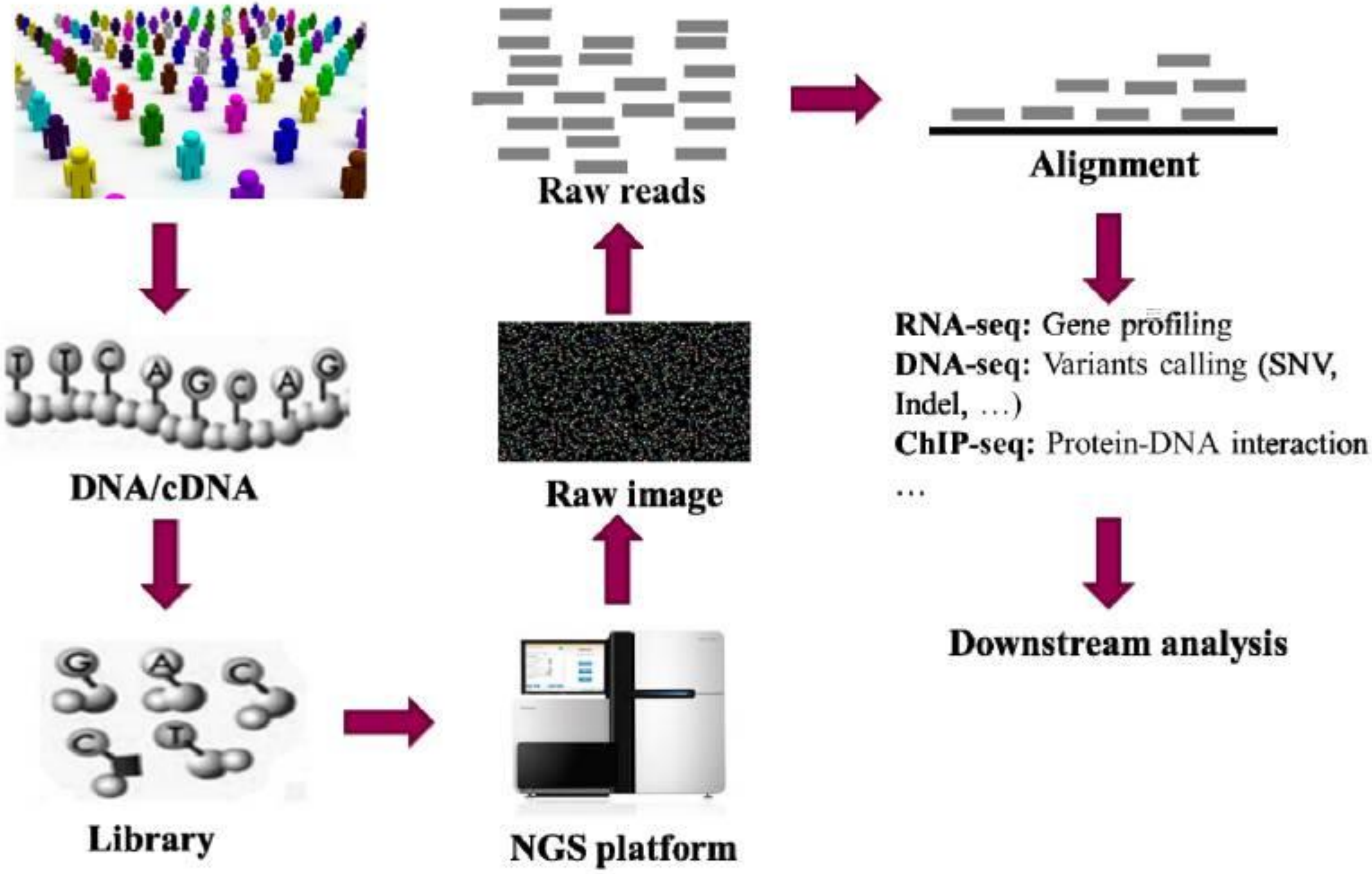
-

- *de novo*

-

-

NEXT=AGRICULTURE
OMIC≡GENERATION



DNA/cDNA

Library

NGS platform

Raw image

Raw reads

Alignment

RNA-seq: Gene profiling

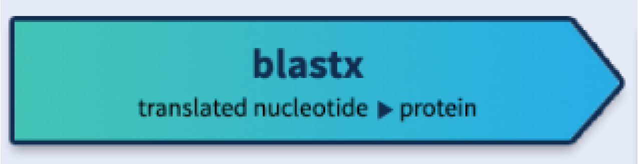
DNA-seq: Variants calling (SNV, Indel, …)

ChIP-seq: Protein-DNA interaction

…

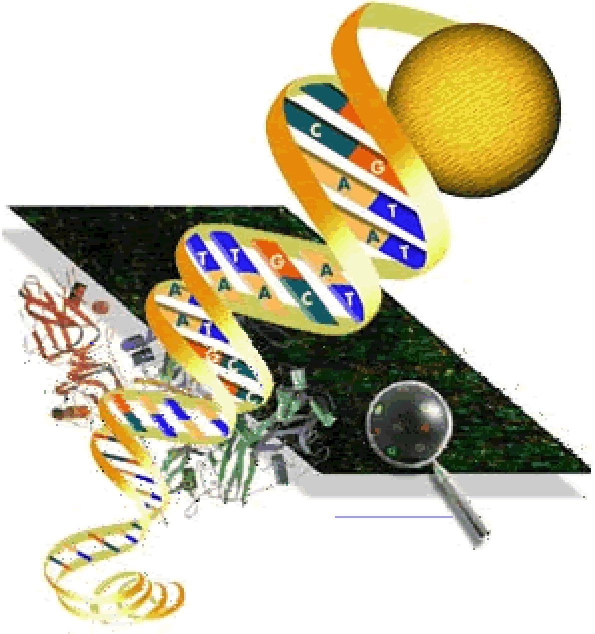
Downstream analysis

Ye H, et al. Pharmaceutics. 2015 Nov 23;7(4):523-41



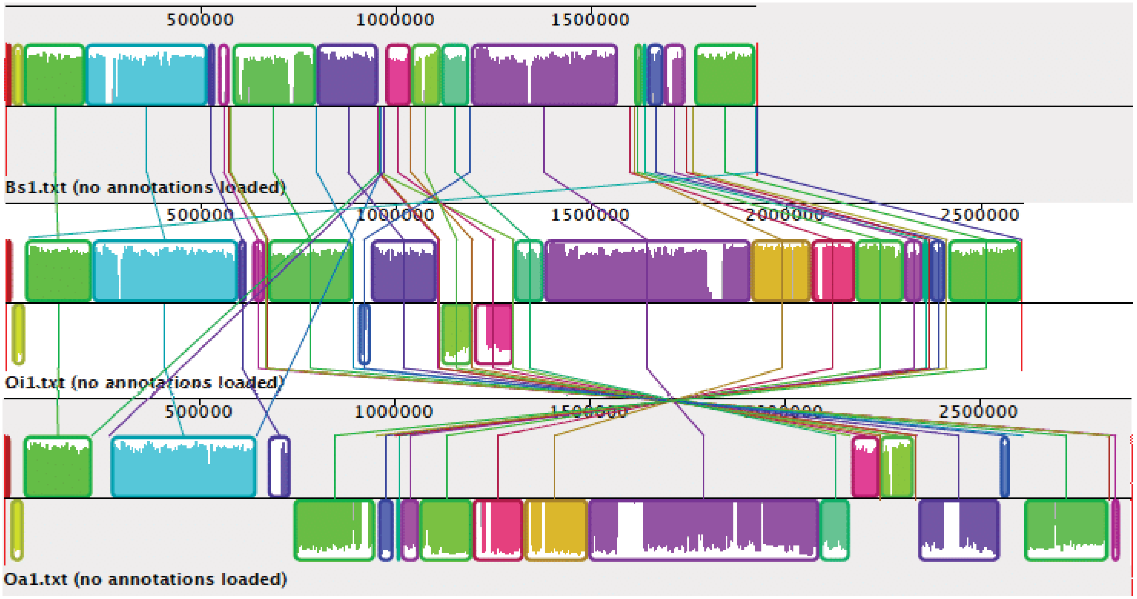
blastx

translated nucleotide ▶ protein



mauve
Multiple Genome Alignment

- rearrangement and inversion

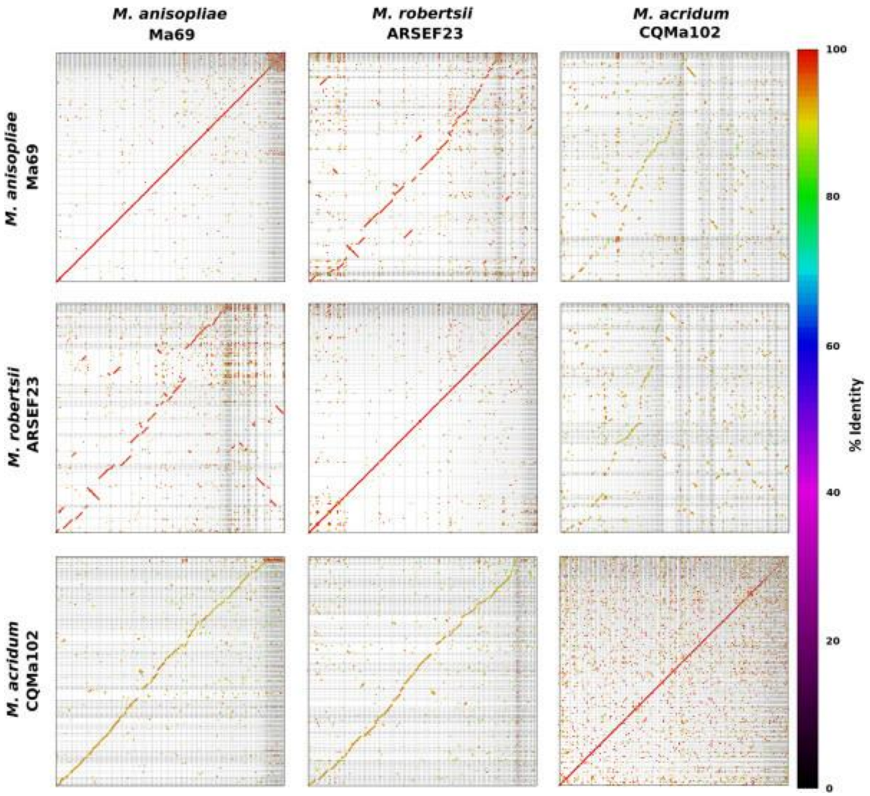


500000
1000000
1500000
Bs1.txt (no annotations loaded)
500000
1000000
1500000
2000000
2500000
Oi1.txt (no annotations loaded)
500000
1000000
1500000
2000000
2500000
Oa1.txt (no annotations loaded)

AMUMMERA3BL

MUMMER 3

MUMMER . 3



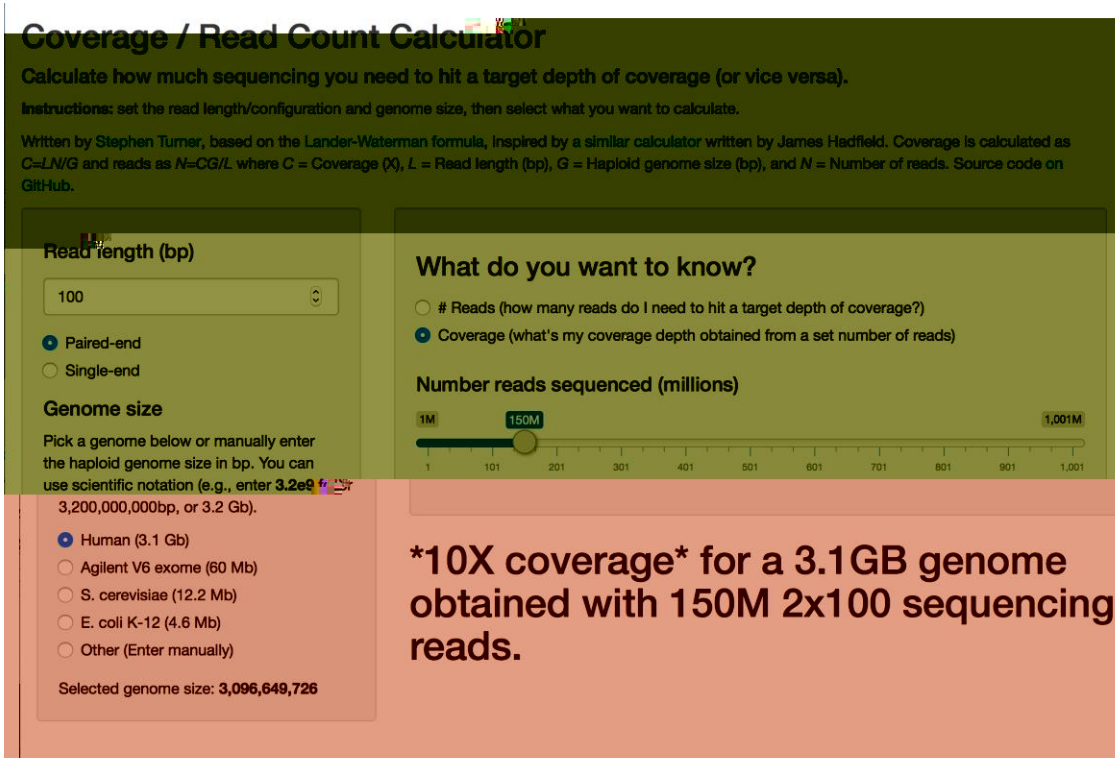
$$C=LN/G$$

C

L

G

N



Coverage / Read Count Calculator

Calculate how much sequencing you need to hit a target depth of coverage (or vice versa).

Instructions: set the read length/configuration and genome size, then select what you want to calculate.

Written by Stephen Turner, based on the Lander-Waterman formula, inspired by a similar calculator written by James Hadfield. Coverage is calculated as *C=LN/G* and reads as *N=CG/L* where *C* = Coverage (X), *L* = Read length (bp), *G* = Haploid genome size (bp), and *N* = Number of reads. Source code on GitHub.

Read length (bp)

100

● Paired-end

○ Single-end

Genome size

Pick a genome below or manually enter the haploid genome size in bp. You can use scientific notation (e.g., enter **3.2e9** for 3,200,000,000bp, or 3.2 Gb).

● Human (3.1 Gb)

○ Agilent V6 exome (60 Mb)

○ S. cerevisiae (12.2 Mb)

○ E. coli K-12 (4.6 Mb)

○ Other (Enter manually)

Selected genome size: **3,096,649,726**

What do you want to know?

○ # Reads (how many reads do I need to hit a target depth of coverage?)

● Coverage (what's my coverage depth obtained from a set number of reads)

Number reads sequenced (millions)

150M

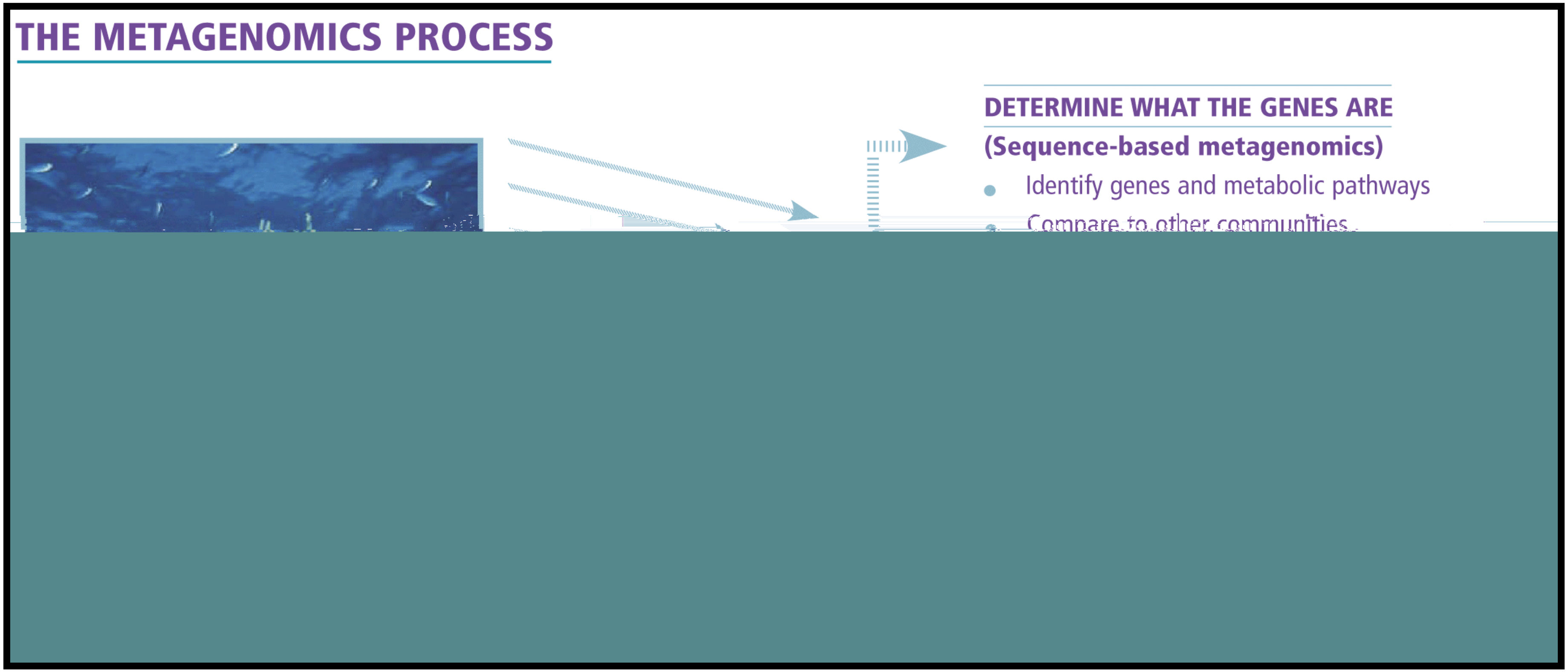
***10X coverage* for a 3.1GB genome obtained with 150M 2x100 sequencing reads.**

[redacted]

●

●

[redacted]



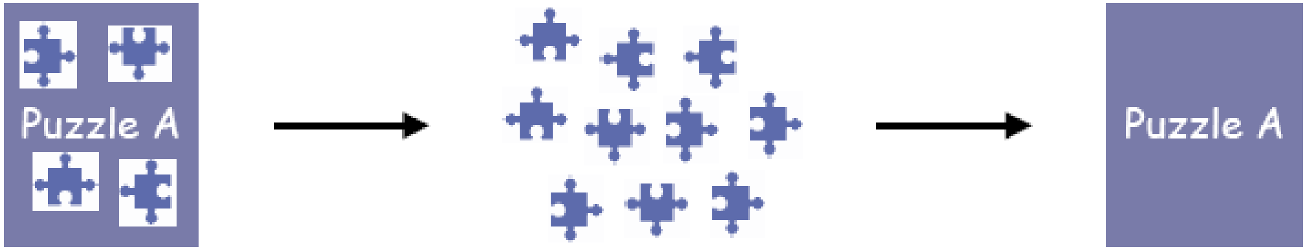
THE METAGENOMICS PROCESS

DETERMINE WHAT THE GENES ARE

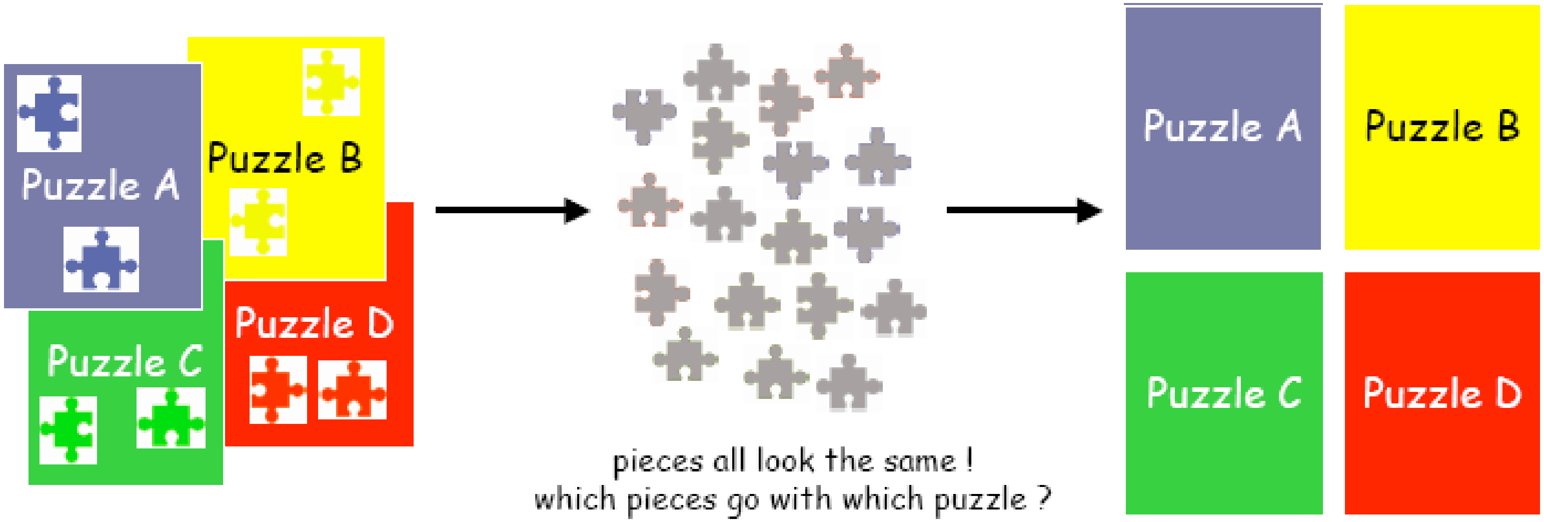
(Sequence-based metagenomics)

- Identify genes and metabolic pathways
- Compare to other communities

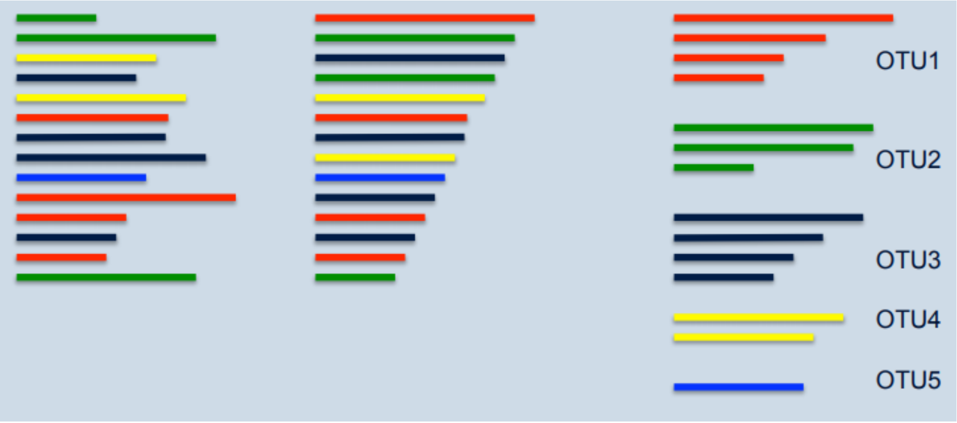
Isolate genome - single source of DNA

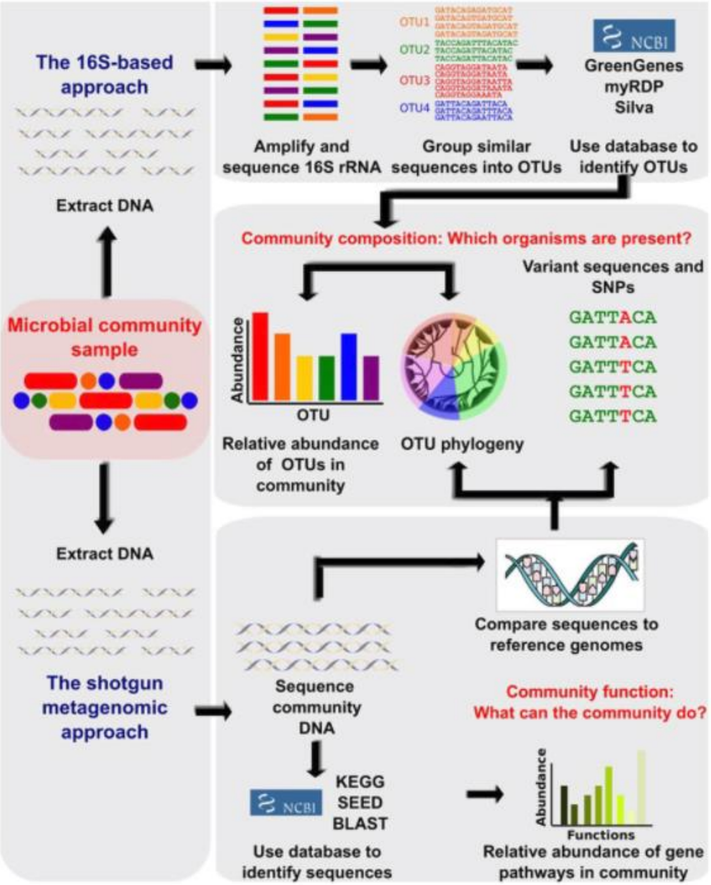


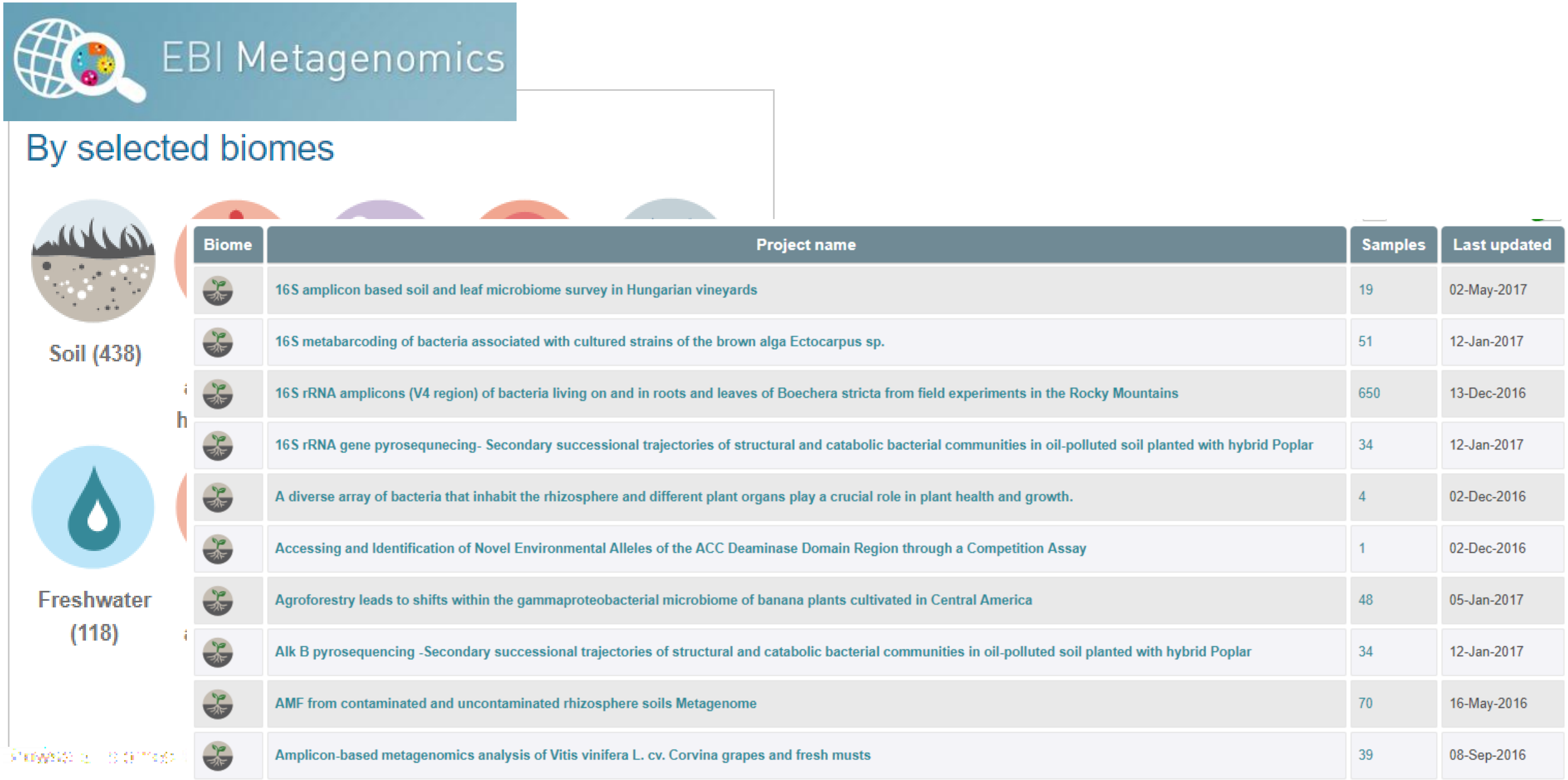
Environmental genome - multiple sources of DNA



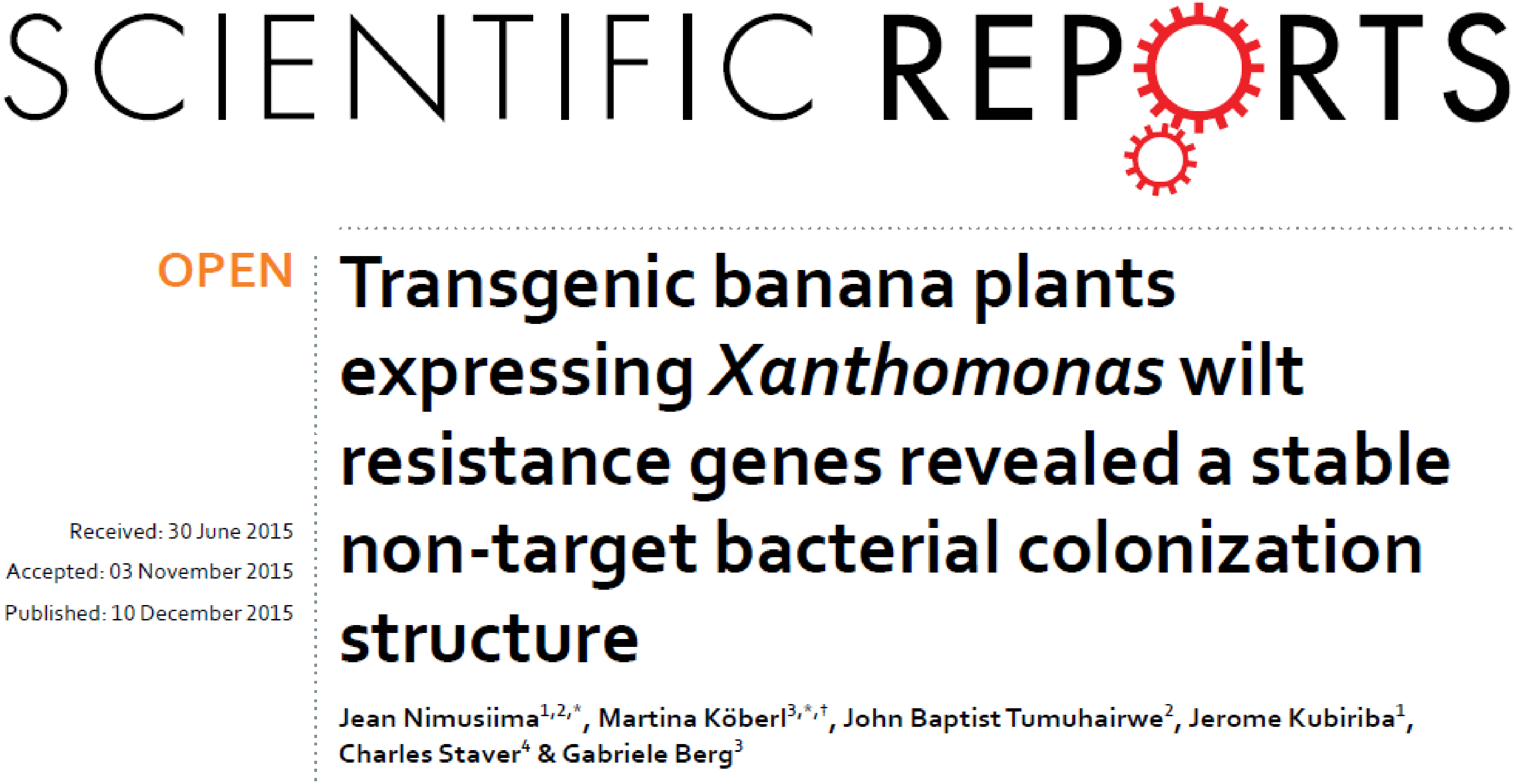
pieces all look the same !
which pieces go with which puzzle ?


OTU1
OTU2
OTU3
OTU4
OTU5


The 16S-based approach
Extract DNA
Amplify and sequence 16S rRNA
OTU1
OTU2
OTU3
OTU4
Group similar sequences into OTUs
NCBI
GreenGenes
myRDP
Silva
Use database to identify OTUs
Community composition: Which organisms are present?
Microbial community sample
Abundance
OTU
Relative abundance of OTUs in community
OTU phylogeny
Variant sequences and SNPs
GATTACA
GATTACA
GATTTCA
GATTTCA
GATTTCA
Extract DNA
The shotgun metagenomic approach
Sequence community DNA
Compare sequences to reference genomes
Community function: What can the community do?
NCBI
KEGG
SEED
BLAST
Use database to identify sequences
Abundance
Functions
Relative abundance of gene pathways in community



Biome	Project name	Samples	Last updated
🌱	16S amplicon based soil and leaf microbiome survey in Hungarian vineyards	19	02-May-2017
🌱	16S metabarcoding of bacteria associated with cultured strains of the brown alga Ectocarpus sp.	51	12-Jan-2017
🌱	16S rRNA amplicons (V4 region) of bacteria living on and in roots and leaves of Boechera stricta from field experiments in the Rocky Mountains	650	13-Dec-2016
🌱	16S rRNA gene pyrosequnecing- Secondary successional trajectories of structural and catabolic bacterial communities in oil-polluted soil planted with hybrid Poplar	34	12-Jan-2017
🌱	A diverse array of bacteria that inhabit the rhizosphere and different plant organs play a crucial role in plant health and growth.	4	02-Dec-2016
🌱	Accessing and Identification of Novel Environmental Alleles of the ACC Deaminase Domain Region through a Competition Assay	1	02-Dec-2016
🌱	Agroforestry leads to shifts within the gammaproteobacterial microbiome of banana plants cultivated in Central America	48	05-Jan-2017
🌱	Alk B pyrosequencing -Secondary successional trajectories of structural and catabolic bacterial communities in oil-polluted soil planted with hybrid Poplar	34	12-Jan-2017
🌱	AMF from contaminated and uncontaminated rhizosphere soils Metagenome	70	16-May-2016
🌱	Amplicon-based metagenomics analysis of Vitis vinifera L. cv. Corvina grapes and fresh musts	39	08-Sep-2016



SCIENTIFIC REPORTS

OPEN

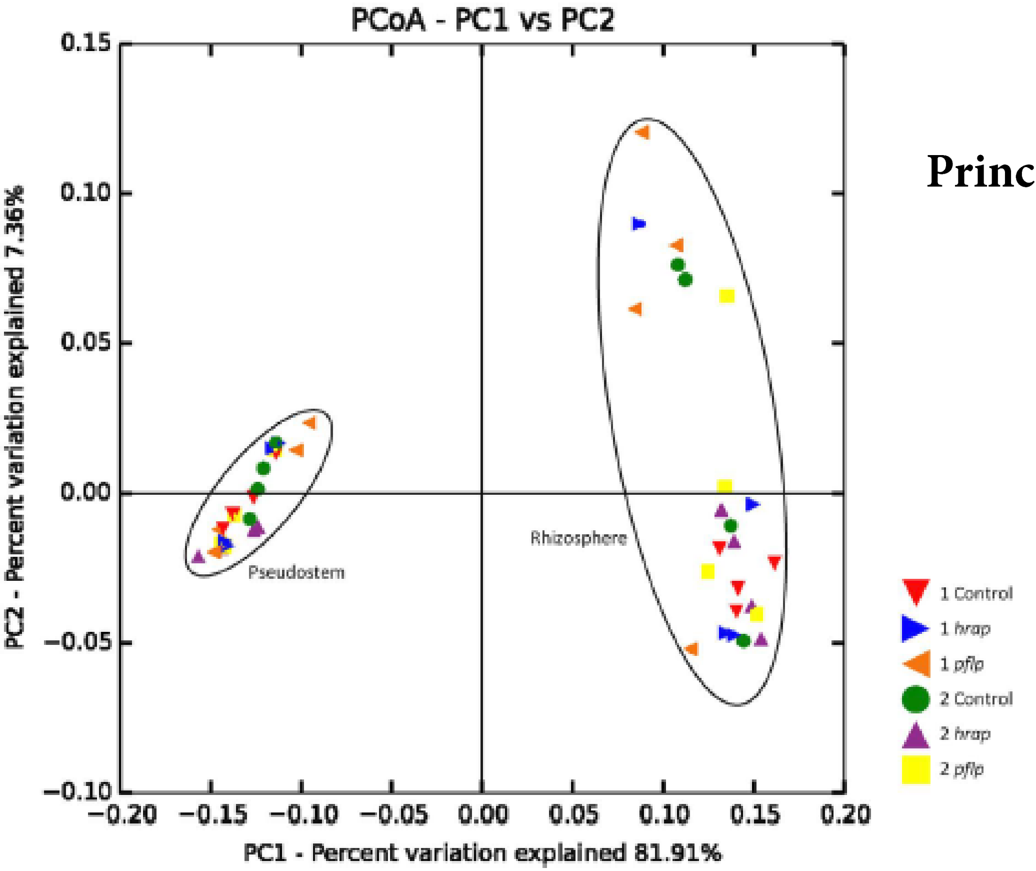
Received: 30 June 2015
Accepted: 03 November 2015
Published: 10 December 2015

Transgenic banana plants expressing *Xanthomonas* wilt resistance genes revealed a stable non-target bacterial colonization structure

Jean Nimusiima[1,2,*], Martina Köberl[3,*,†], John Baptist Tumuhairwe[2], Jerome Kubiriba[1], Charles Staver[4] & Gabriele Berg[3]

-
-
-

-



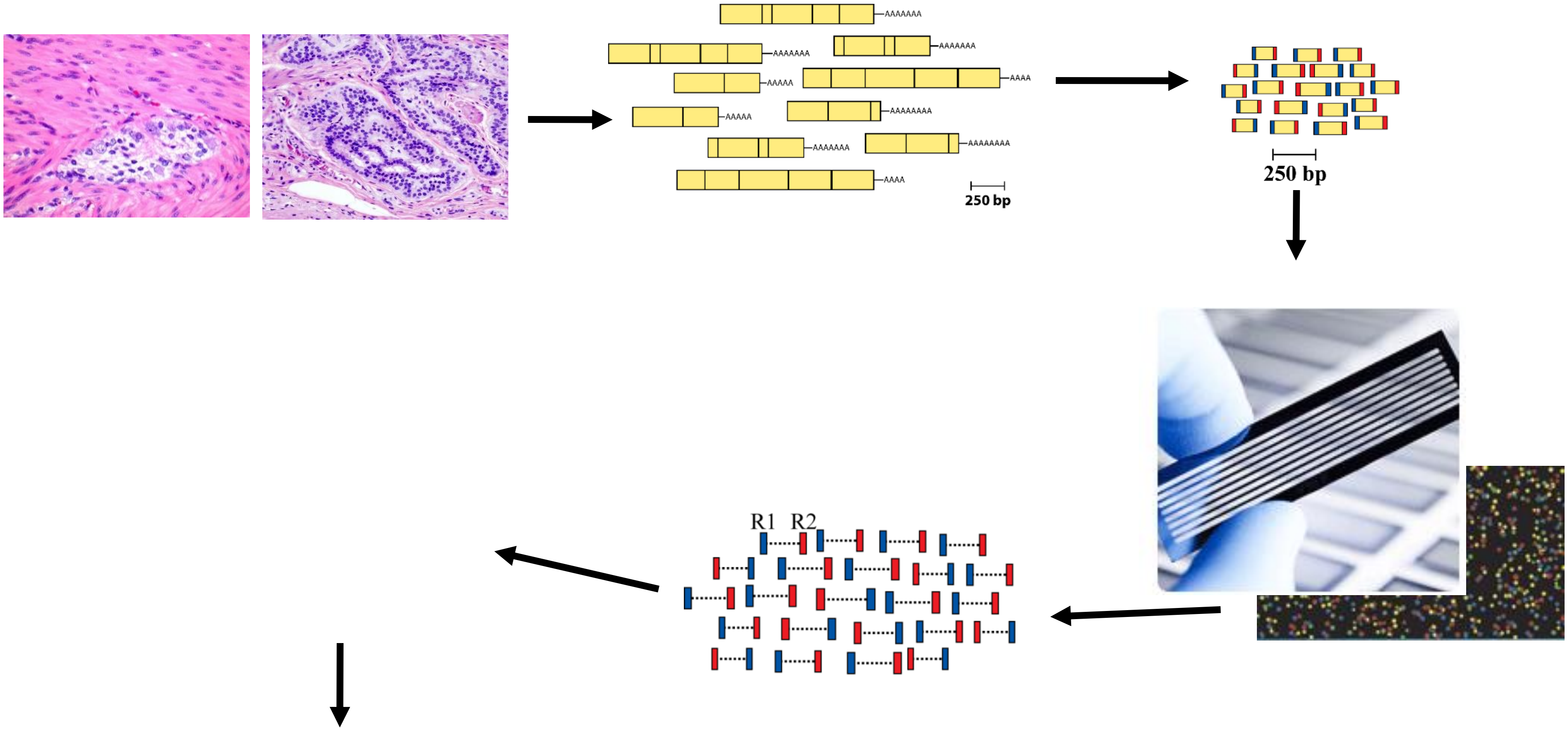
Principal coordinate analysis (PCoA)

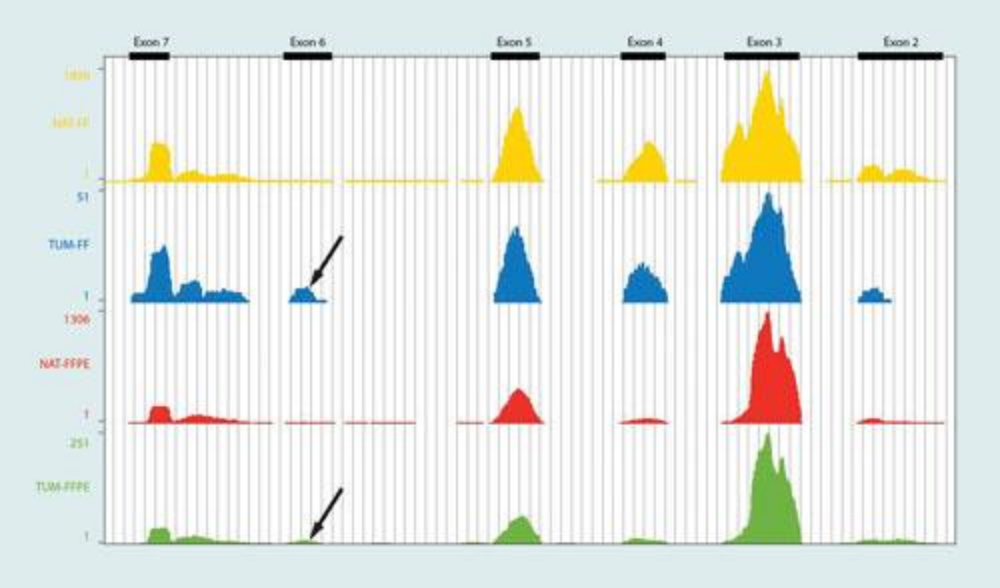
et al.

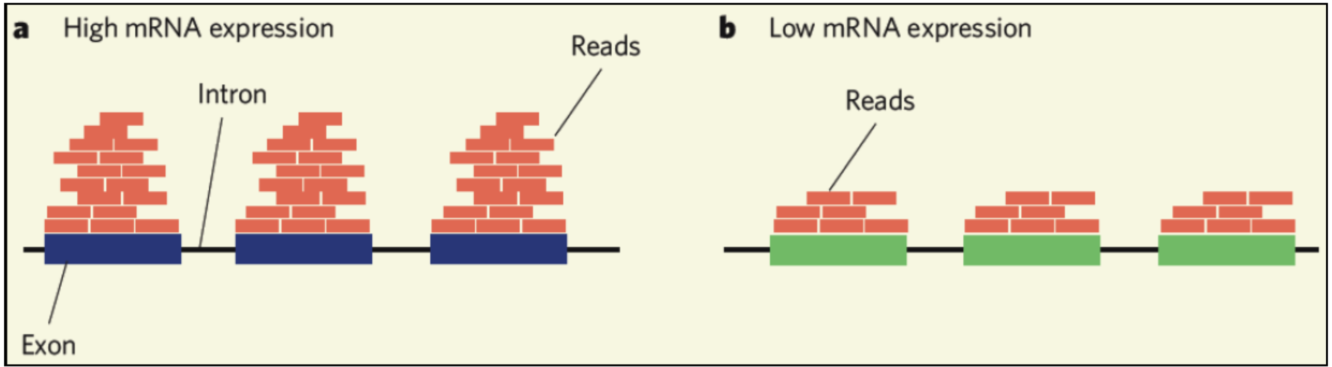
•

•

et al.

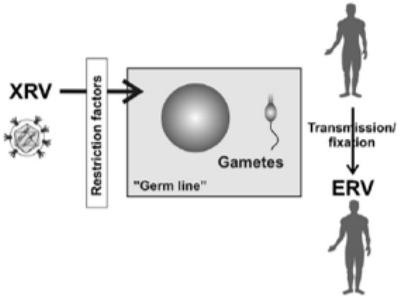

AAAAAAA
AAAAAAA
AAAAAAA
AAAAA
AAAA
AAAAA
AAAAAAAA
AAAAAAA
AAAAAAAA
AAAA
250 bp
250 bp
R1
R2


Exon 7
Exon 6
Exon 5
Exon 4
Exon 3
Exon 2
NAT-FF
1
51
TUM-FF
1
1306
NAT-FFPE
1
251
TUM-FFPE
1


a
High mRNA expression
Reads
Intron
Exon
b
Low mRNA expression
Reads

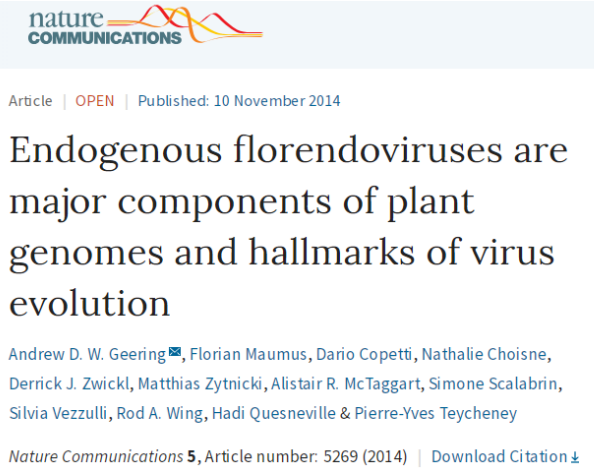
HERV = Human Endogenous Retrovirus

- Endogenous retroviruses (ERVs) = DNA sequences within a genome that are similar to sequences of infectious retroviruses



generally found in the human

HERVs are the remnants of ancient retroviral infections that became fixed in the germ lines.

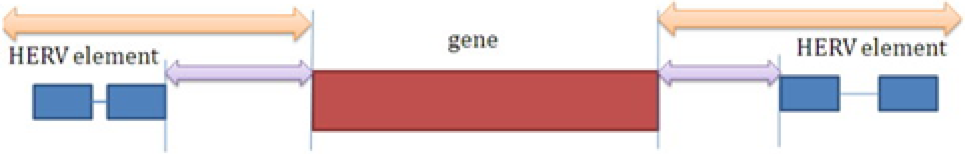


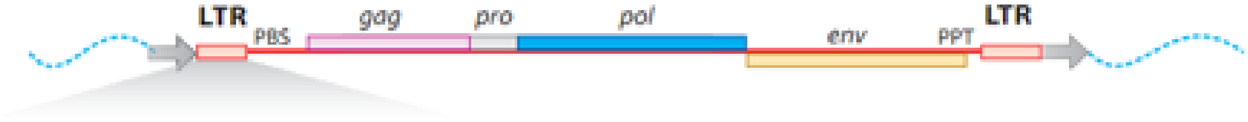
Article | OPEN | Published: 10 November 2014

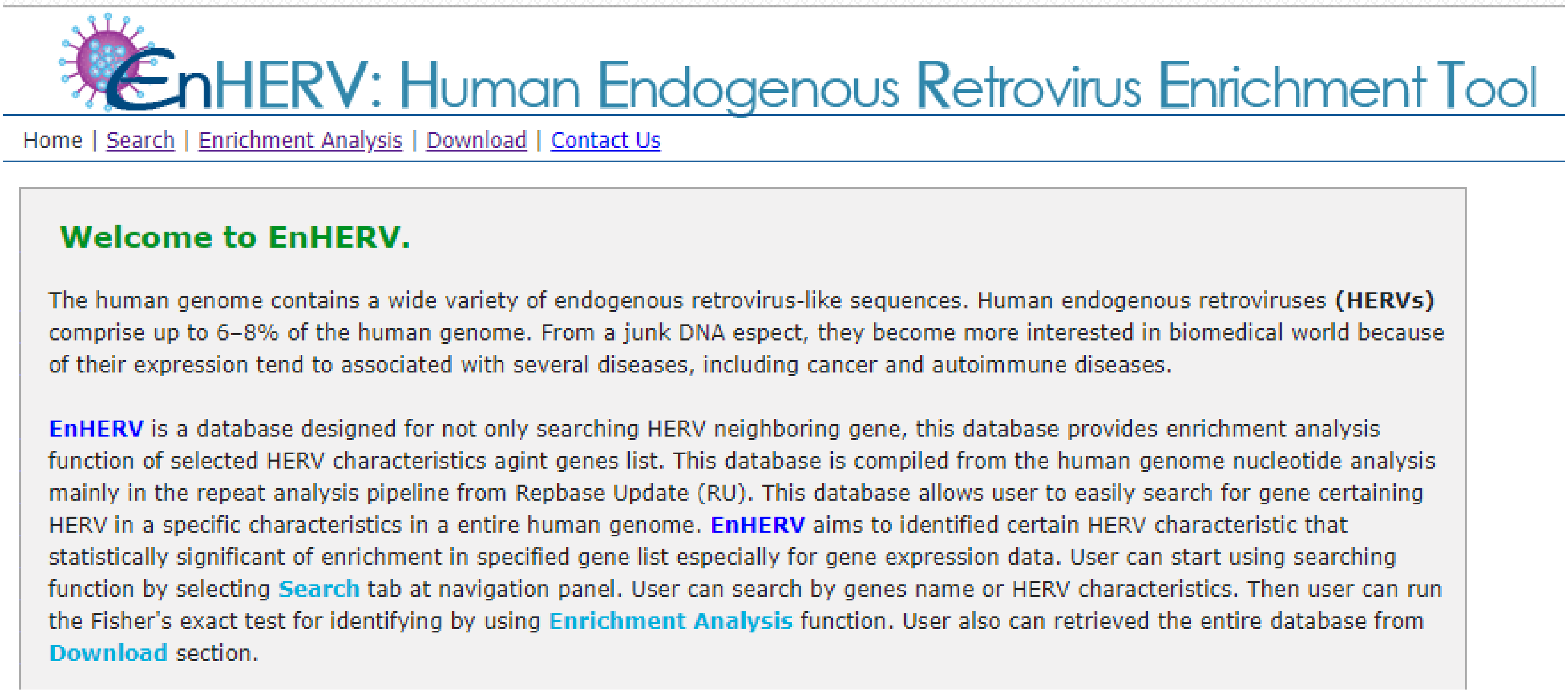
Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution

Andrew D. W. Geering, Florian Maumus, Dario Copetti, Nathalie Choisne, Derrick J. Zwickl, Matthias Zytnicki, Alistair R. McTaggart, Simone Scalabrin, Silvia Vezzulli, Rod A. Wing, Hadi Quesneville & Pierre-Yves Teycheney

Nature Communications **5**, Article number: 5269 (2014) | Download Citation





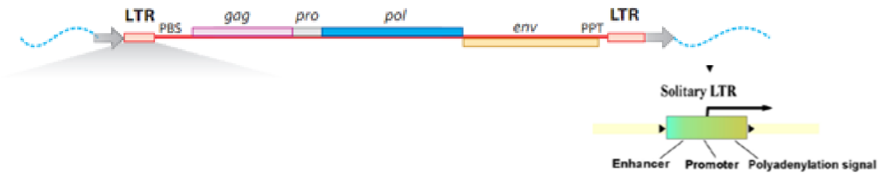


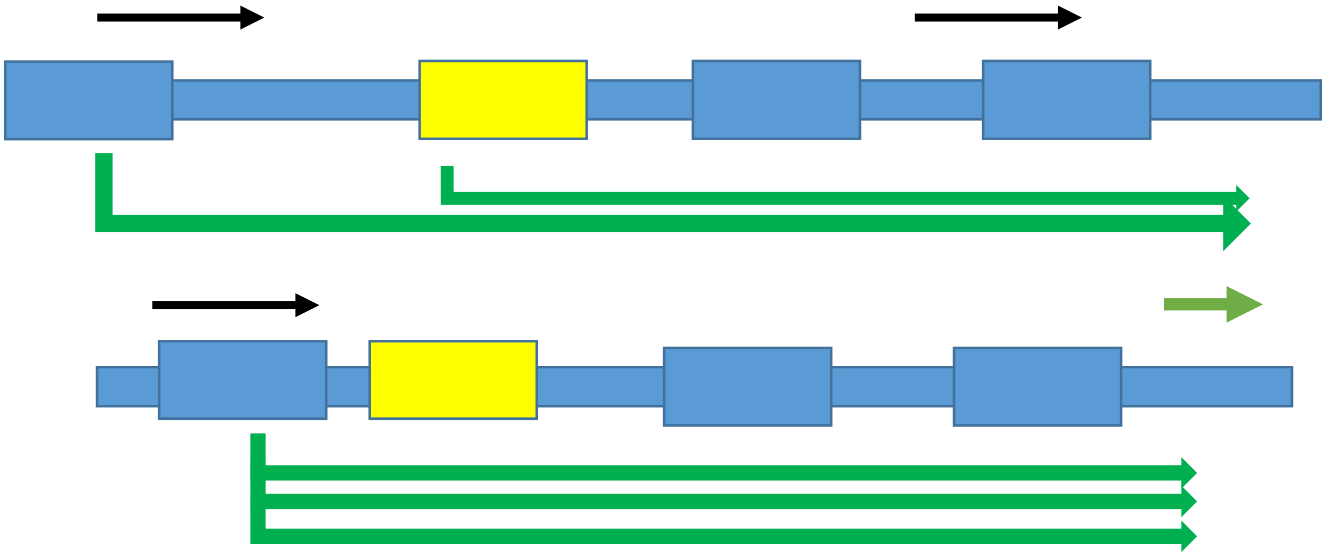
EnHERV: Human Endogenous Retrovirus Enrichment Tool

Welcome to EnHERV.

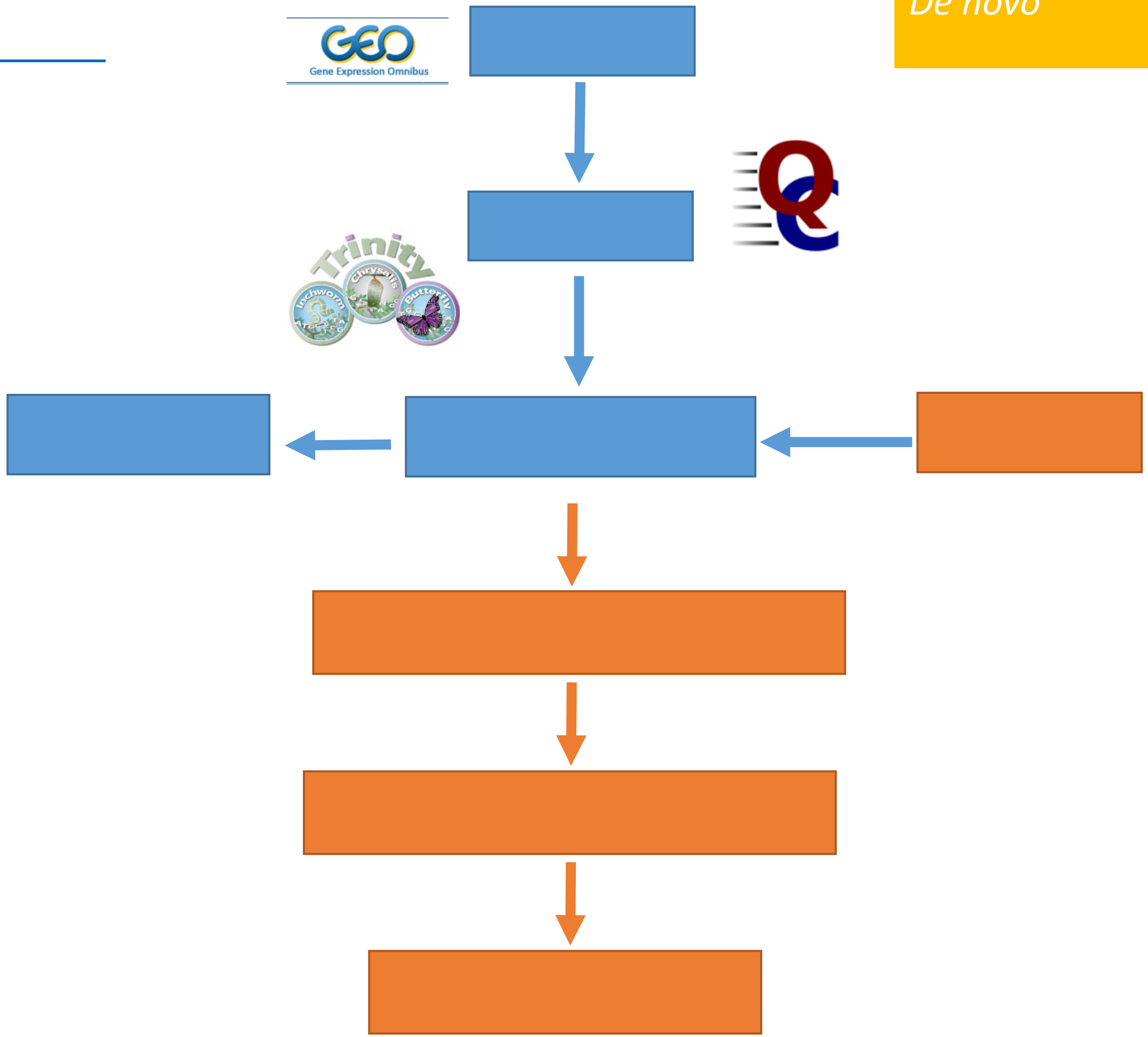
The human genome contains a wide variety of endogenous retrovirus-like sequences. Human endogenous retroviruses **(HERVs)** comprise up to 6–8% of the human genome. From a junk DNA espect, they become more interested in biomedical world because of their expression tend to associated with several diseases, including cancer and autoimmune diseases.

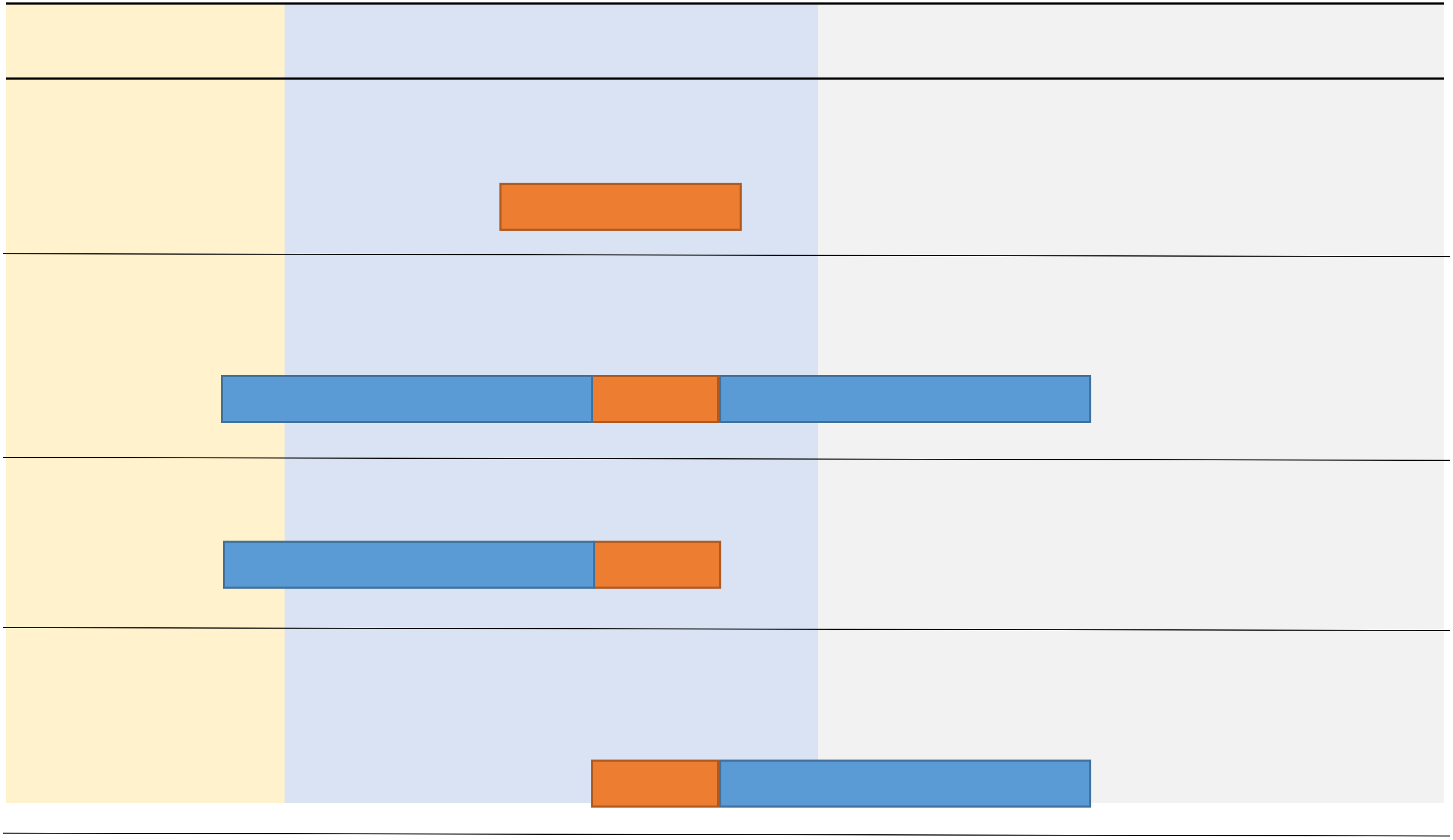
EnHERV is a database designed for not only searching HERV neighboring gene, this database provides enrichment analysis function of selected HERV characteristics agint genes list. This database is compiled from the human genome nucleotide analysis mainly in the repeat analysis pipeline from Repbase Update (RU). This database allows user to easily search for gene certaining HERV in a specific characteristics in a entire human genome. **EnHERV** aims to identified certain HERV characteristic that statistically significant of enrichment in specified gene list especially for gene expression data. User can start using searching function by selecting **Search** tab at navigation panel. User can search by genes name or HERV characteristics. Then user can run the Fisher's exact test for identifying by using **Enrichment Analysis** function. User also can retrieved the entire database from **Download** section.


LTR
PBS
gag
pro
pol
env
PPT
LTR
Solitary LTR
Enhancer
Promoter
Polyadenylation signal

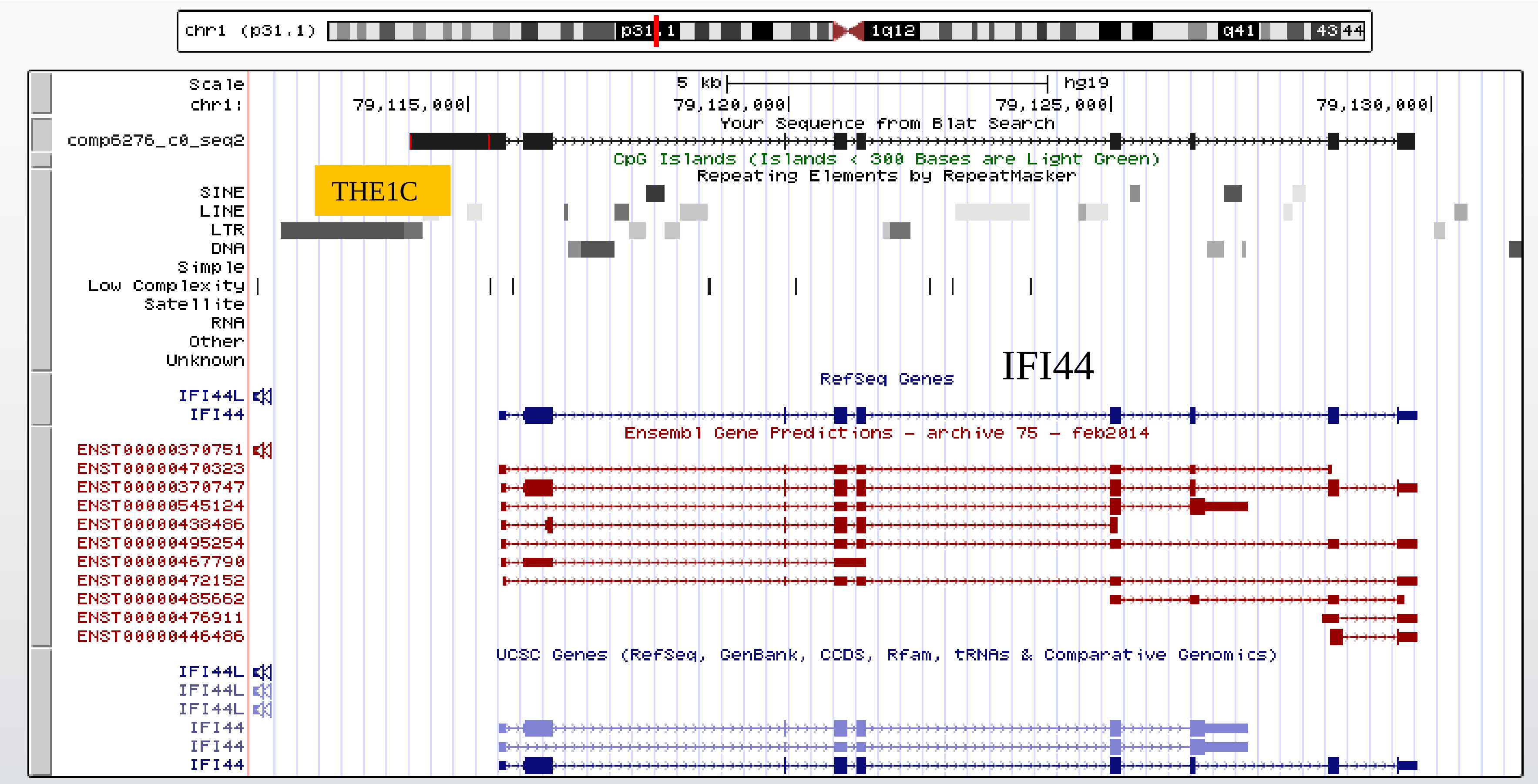


De novo


GEO
Gene Expression Omnibus
QC
Trinity
Chrysalis
Inchworm
Butterfly



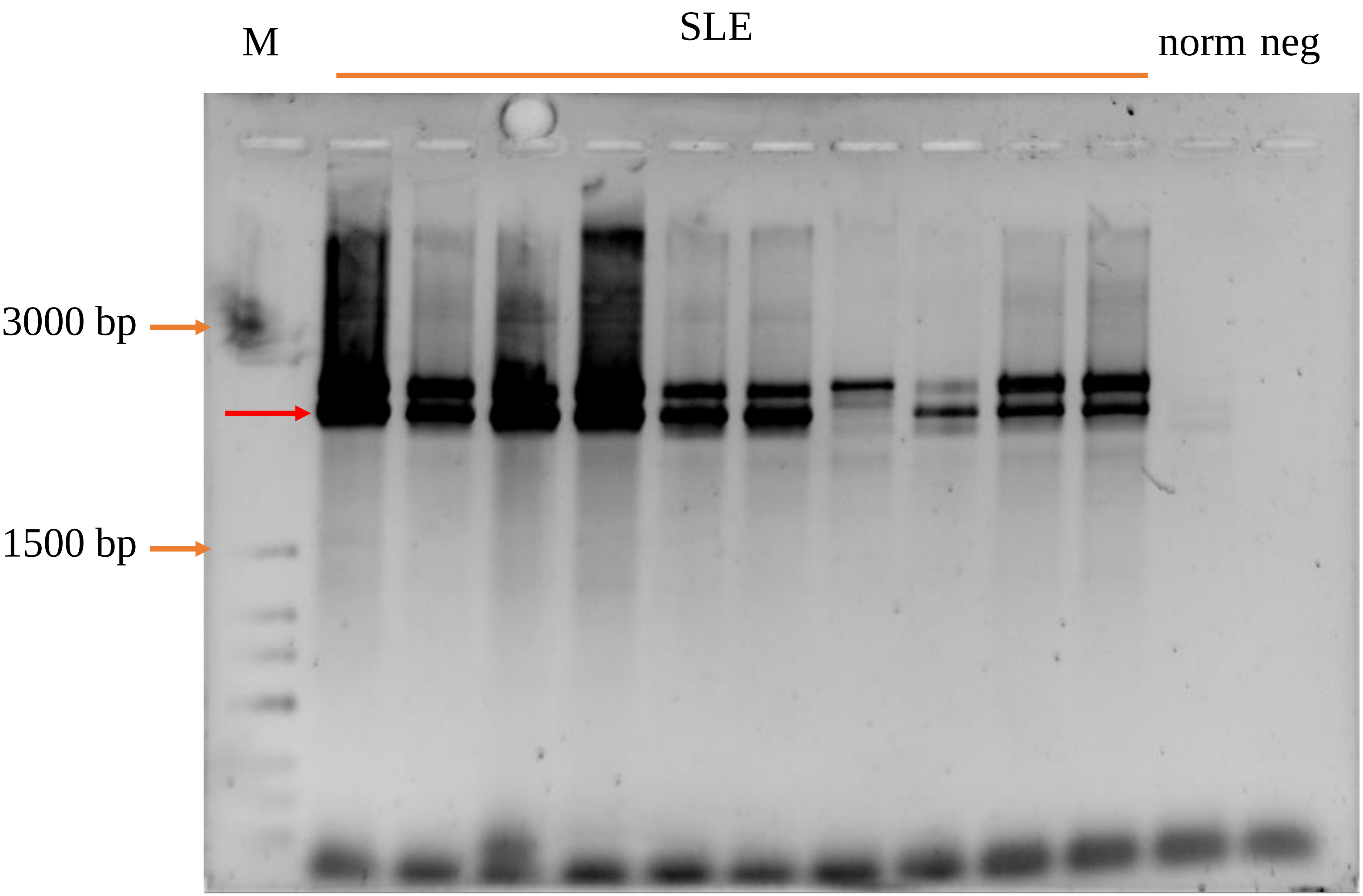
THE1C



KA Kirou. *et al.* 2004. Arthritis Res Ther. 6(Suppl 3):91

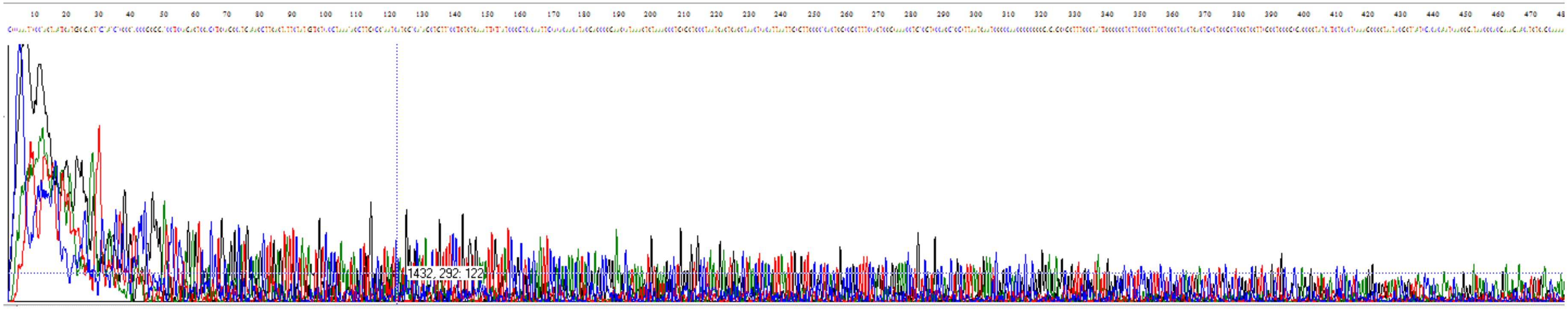
IFI44-THE1C chimeric amplicons.

The expected size is 1740 bps.

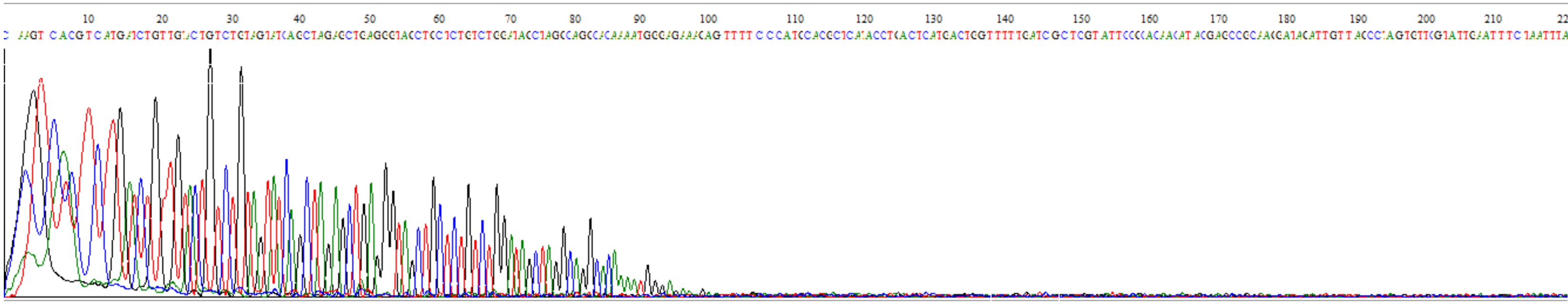


IFI44-THE1B chimeric amplicons.

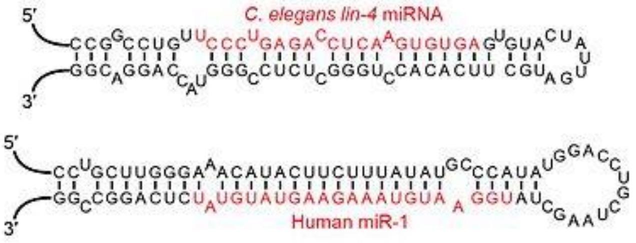
THE1C-forward amplicon



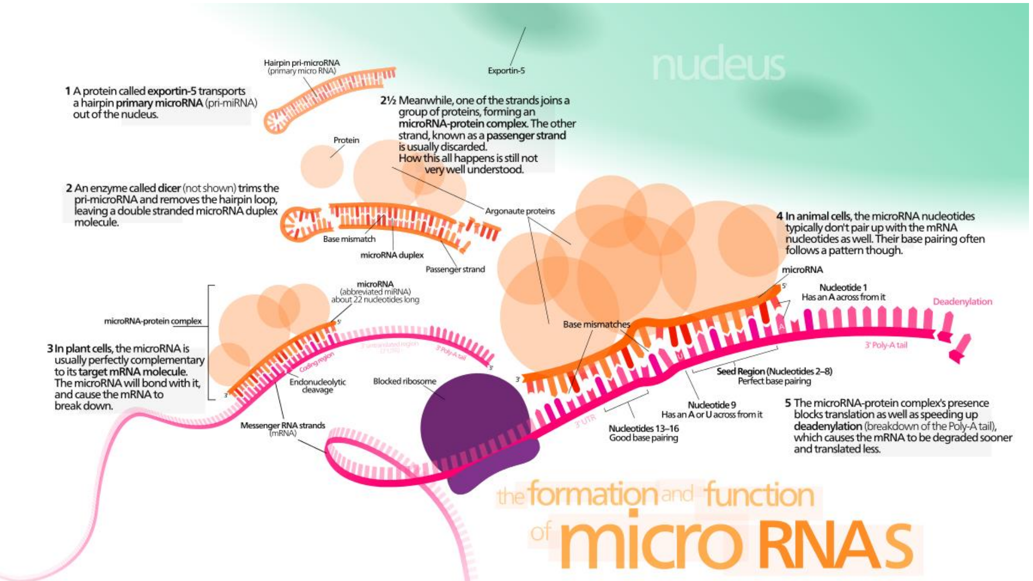
IFI44-reverse amplicon



-



5′
C. elegans lin-4 miRNA
CCGGCCUGUUCCCUGAGACCUCAAGUGUGAGUGUACUA
GGCAGGACCAUGGGCCUCUCGGGUCCACACUUCGUAGUU
3′
5′
CCUGCUUGGGAAACAUACUUCUUUAUAUGCCCAUAUGGACCU
GGCCGGACUCUAUGUAUGAAGAAAUGUAAGGUAUCGAAUCG
Human miR-1
3′



1 A protein called **exportin-5 transports** a hairpin **primary microRNA** (pri-miRNA) out of the nucleus.

2 An enzyme called **dicer** (not shown) **trims the** pri-microRNA and removes the hairpin loop, leaving a double stranded microRNA duplex molecule.

2½ Meanwhile, one of the strands joins a group of proteins, forming an **microRNA-protein complex.** The other strand, known as a **passenger strand** is usually discarded. How this all happens is still not very well understood.

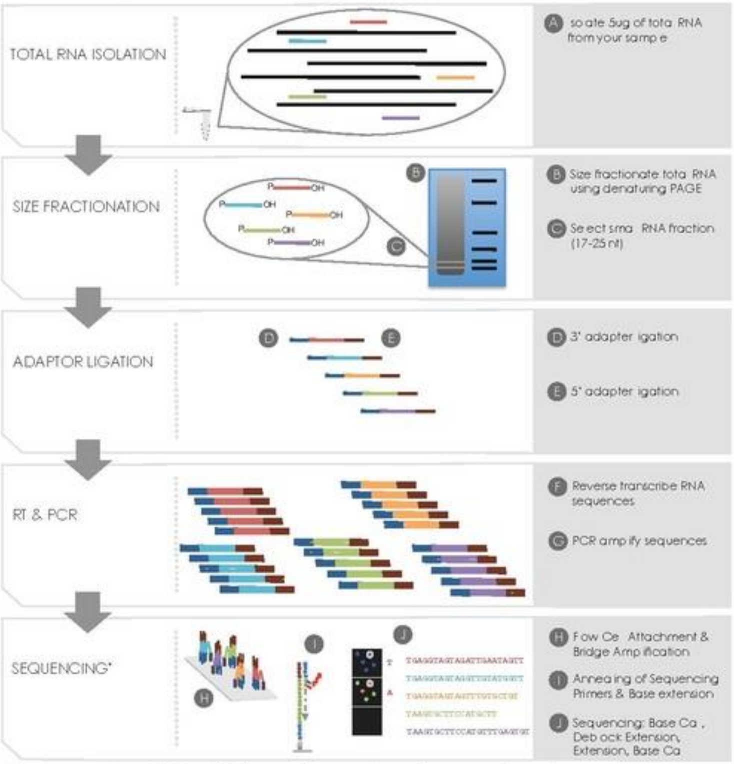
3 In plant cells, the microRNA is usually perfectly complementary to its **target mRNA molecule.** The microRNA will bond with it, and cause the mRNA to break down.

4 In animal cells, the microRNA nucleotides typically don't pair up with the mRNA nucleotides as well. Their base pairing often follows a pattern though.

5 The microRNA-protein complex's presence blocks translation as well as speeding up **deadenylation** (breakdown of the Poly-A tail), which causes the mRNA to be degraded sooner and translated less.

[redacted]

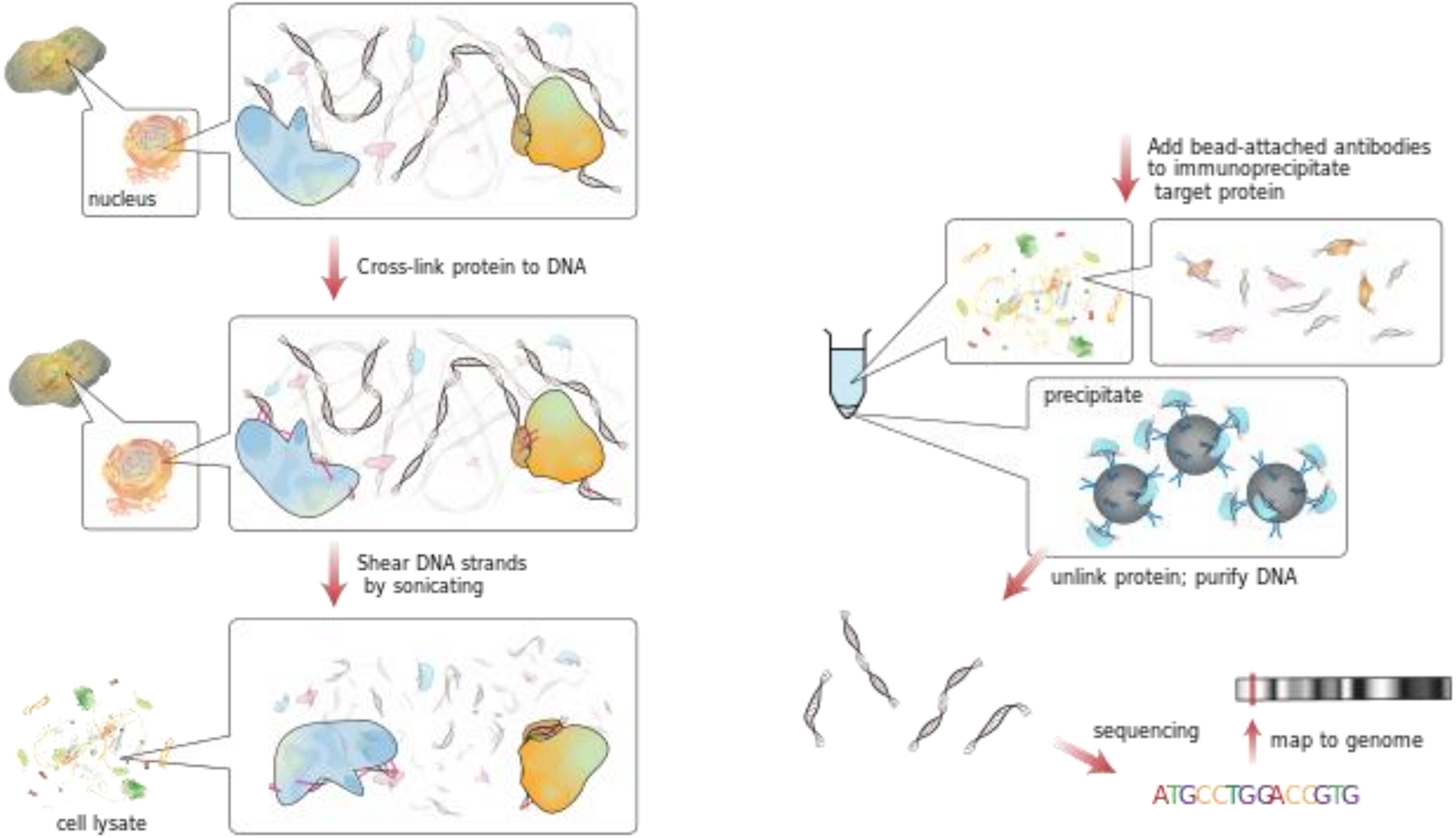
MIRNA-SEQ LIBRARY PREPARATION


TOTAL RNA ISOLATION
A so ate 5ug of tota RNA from your samp e
SIZE FRACTIONATION
P OH
B Size fractionate tota RNA using denaturing PAGE
C Se ect sma RNA fraction (17-25 nt)
ADAPTOR LIGATION
D 3' adapter igation
E 5' adapter igation
RT & PCR
F Reverse transcribe RNA sequences
G PCR amp ify sequences
SEQUENCING*
H F ow Ce Attachment & Bridge Amp ification
I Annea ing of Sequencing Primers & Base extension
J Sequencing: Base Ca , Deb ock Extension, Extension, Base Ca

* umina sequencing method depicted however other sequencing p atforms can a so be used.

[redacted]

-
-
-
-



Cross-link protein to DNA

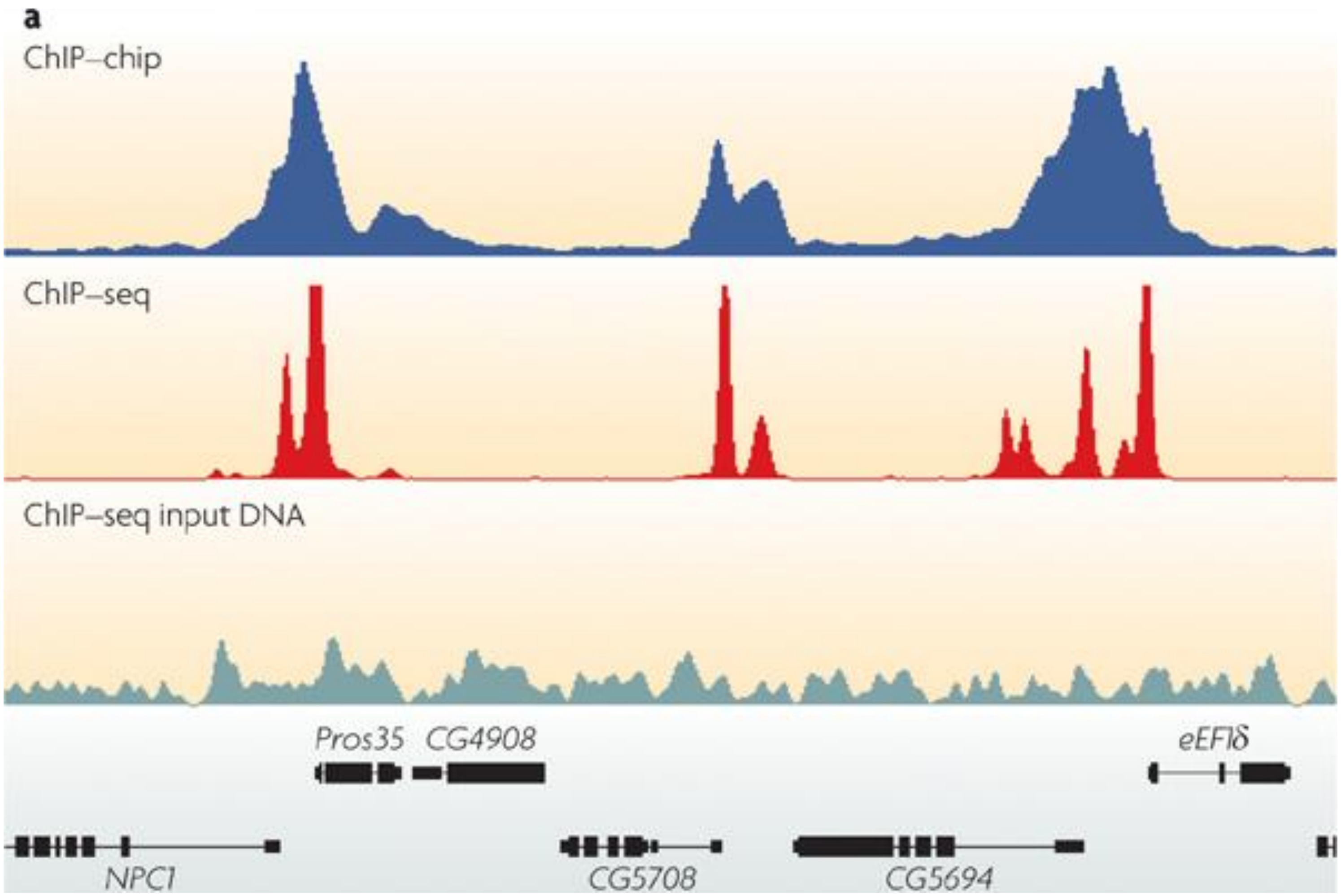
Shear DNA strands by sonicating

Add bead-attached antibodies to immunoprecipitate target protein

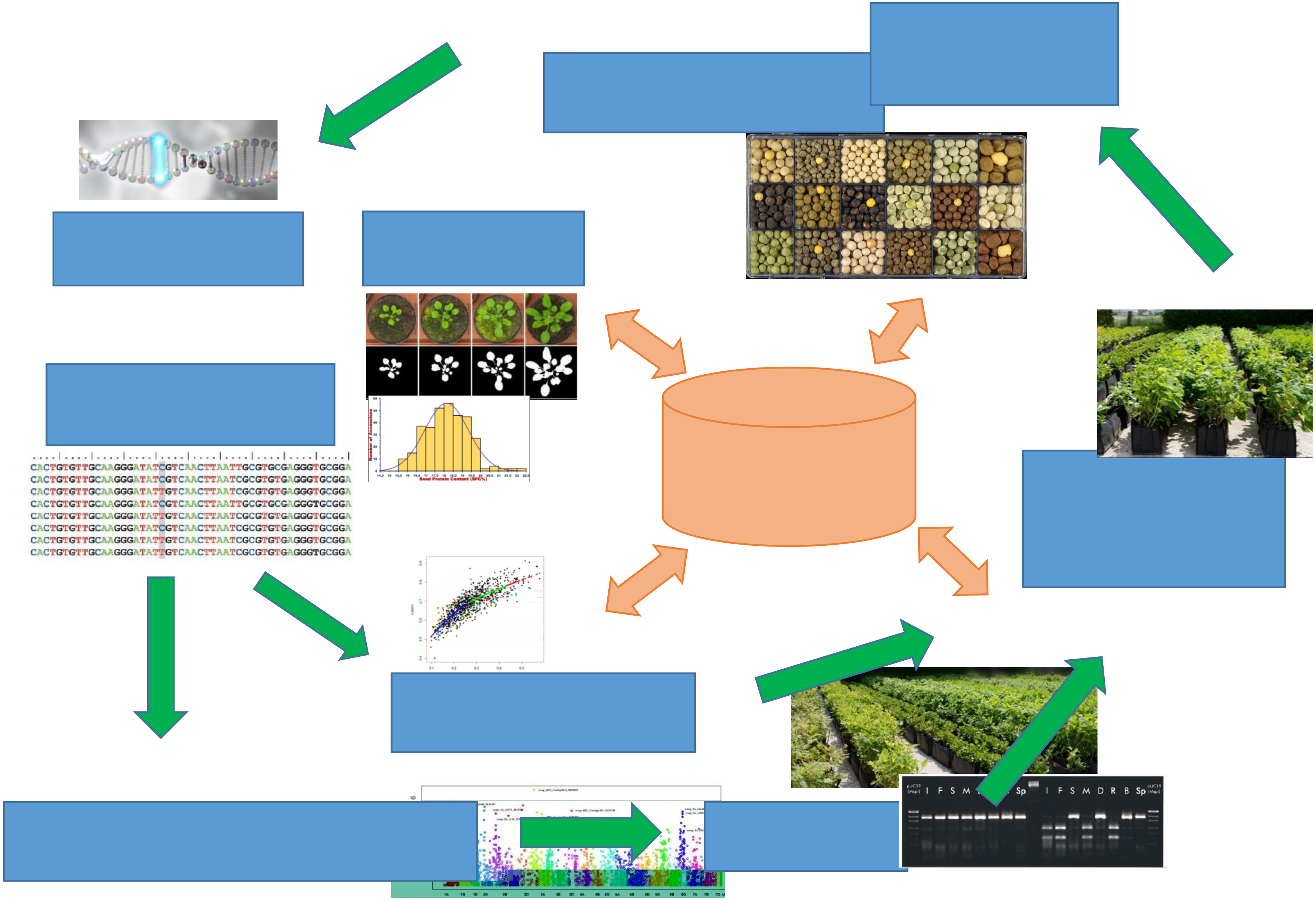
unlink protein; purify DNA

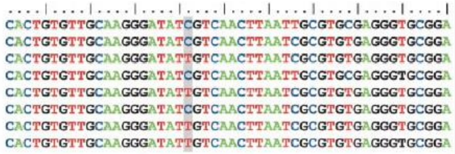
sequencing

map to genome

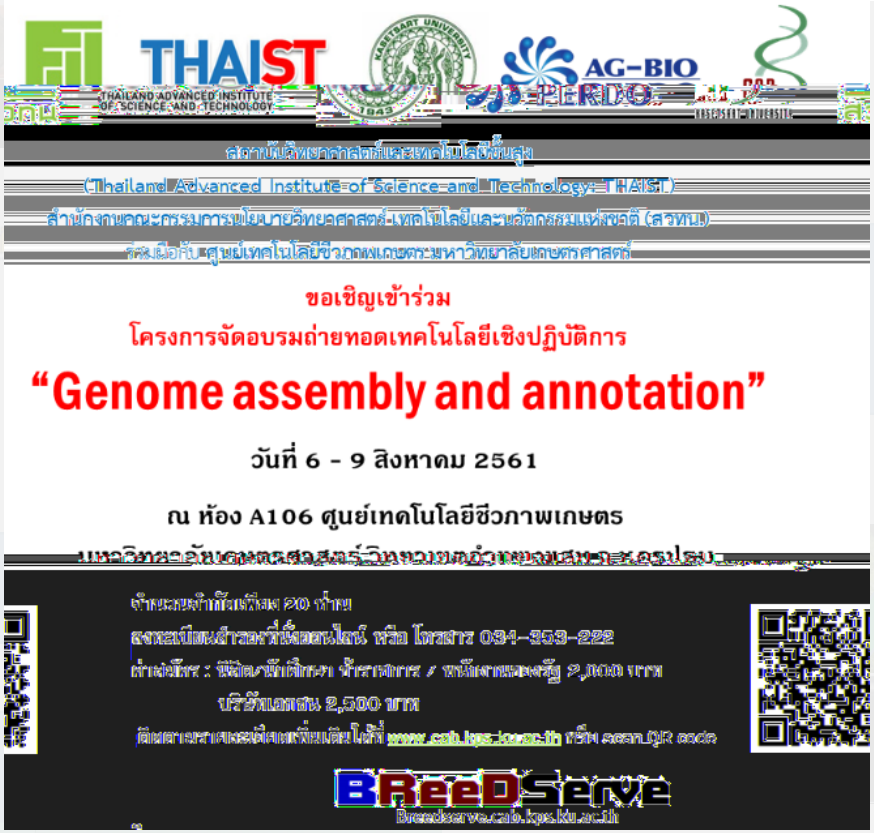

a
ChIP–chip
ChIP–seq
ChIP–seq input DNA
Pros35
CG4908
eEF1δ
NPC1
CG5708
CG5694

BReeDServe
breedserve.cab.kps.ku.ac.th





```
....|....|....|....|....|....|....|....|....|....|
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
```


THAIST
THAILAND ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY
KASETSART UNIVERSITY
1943
AG-BIO
PERDO

สถาบันวิทยาศาสตร์และเทคโนโลยีขั้นสูง

(Thailand Advanced Institute of Science and Technology: THAIST)

สำนักงานคณะกรรมการนโยบายวิทยาศาสตร์ เทคโนโลยีและนวัตกรรมแห่งชาติ (สวทน.)

ร่วมมือกับ ศูนย์เทคโนโลยีชีวภาพเกษตร มหาวิทยาลัยเกษตรศาสตร์

ขอเชิญเข้าร่วม

โครงการจัดอบรมถ่ายทอดเทคโนโลยีเชิงปฏิบัติการ

“Genome assembly and annotation”

วันที่ 6 - 9 สิงหาคม 2561

ณ ห้อง A106 ศูนย์เทคโนโลยีชีวภาพเกษตร

มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน จ.นครปฐม

จำนวนจำกัดเพียง 20 ท่าน

ลงทะเบียนสำรองที่นั่งออนไลน์ หรือ โทรสาร 034-353-222

ค่าสมัคร : นิสิต/นักศึกษา ข้าราชการ / พนักงานของรัฐ 2,000 บาท

บริษัทเอกชน 2,500 บาท

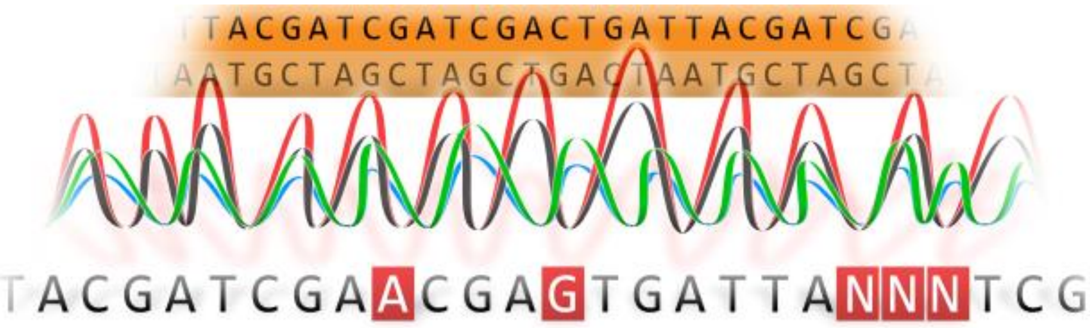
ติดตามรายละเอียดเพิ่มเติมได้ที่ www.cab.kps.ku.ac.th หรือ scan QR code

BReeDServe
Breedserve.cab.kps.ku.ac.th

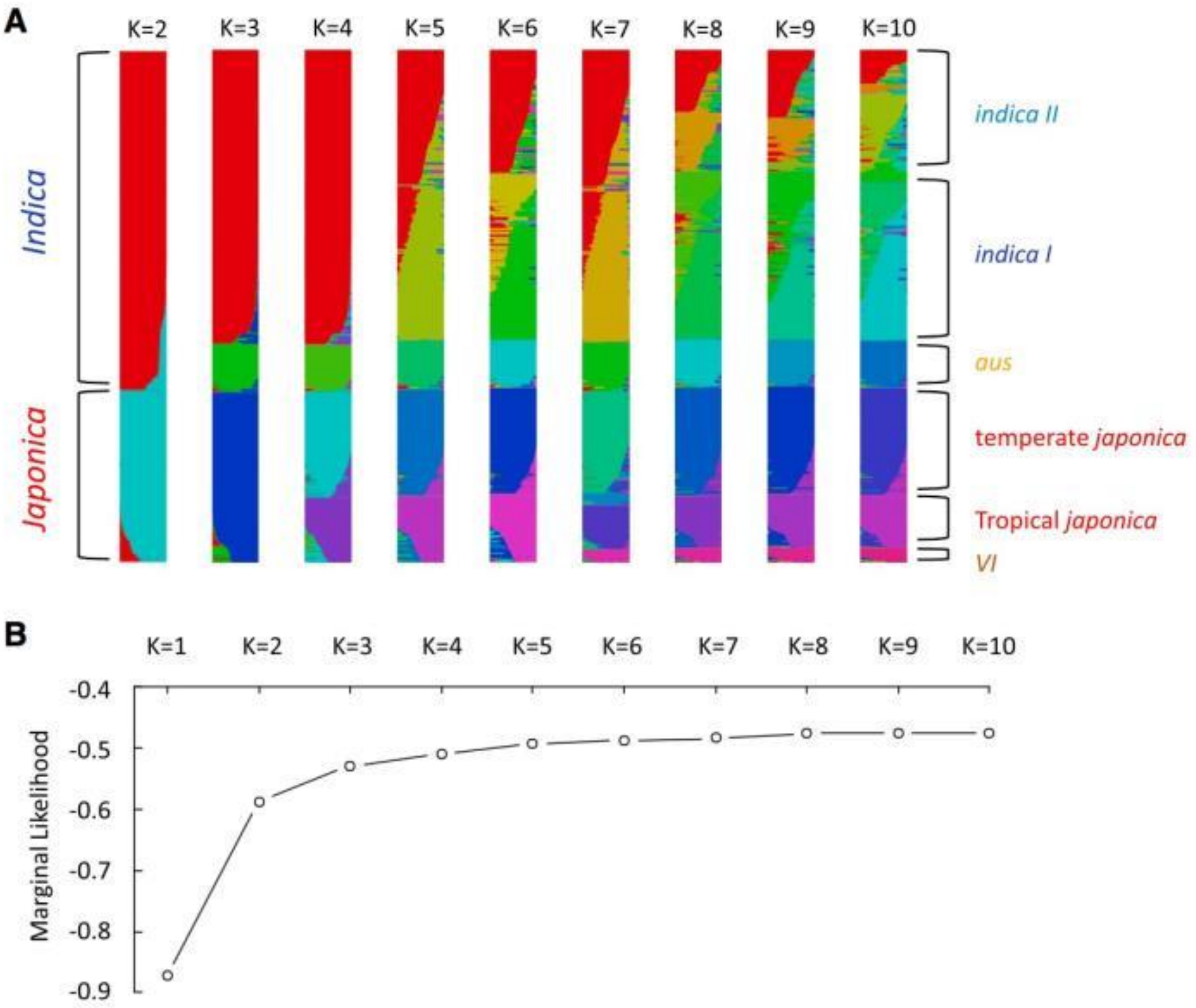
-
-

- *De novo*
-

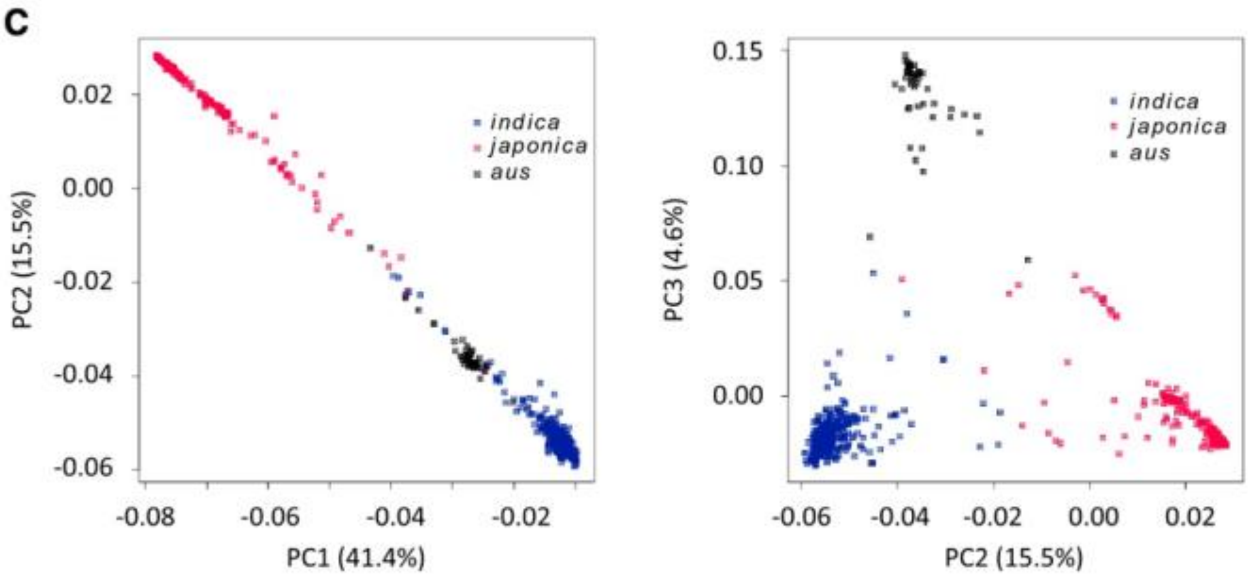
- *de novo*
-
-

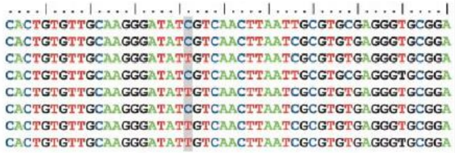

TACGATCGATCGACTGATTACGATCG
AATGCTAGCTAGCTGACTAATGCTAGCT
TACGATCGAACGAGTGATTANNNTCG

K



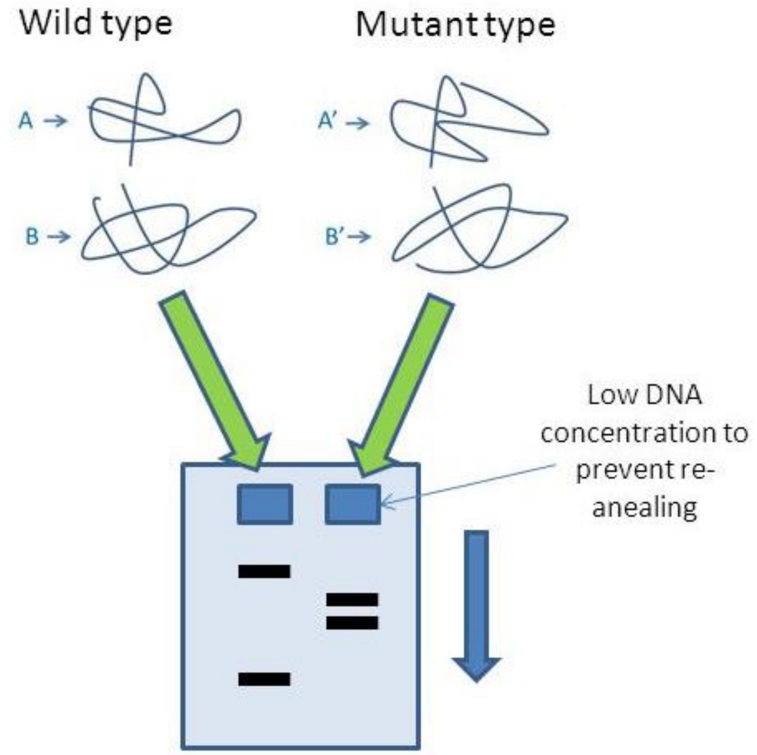
- Detects the underlying genetic population among a set of individuals genotyped at multiple markers
- Computes the proportion of the genome of an individual originating from each inferred population (quantitative clustering method)
- Calculate K: when approaching a plateau or continues increasing slightly
- ***For the TRUE value of K, find the smallest value of K that captures the major structure in the data***


c
0.02
0.00
-0.02
-0.04
-0.06
PC2 (15.5%)
-0.08
-0.06
-0.04
-0.02
PC1 (41.4%)
indica
japonica
aus
0.15
0.10
0.05
0.00
PC3 (4.6%)
-0.06
-0.04
-0.02
0.00
0.02
PC2 (15.5%)
indica
japonica
aus

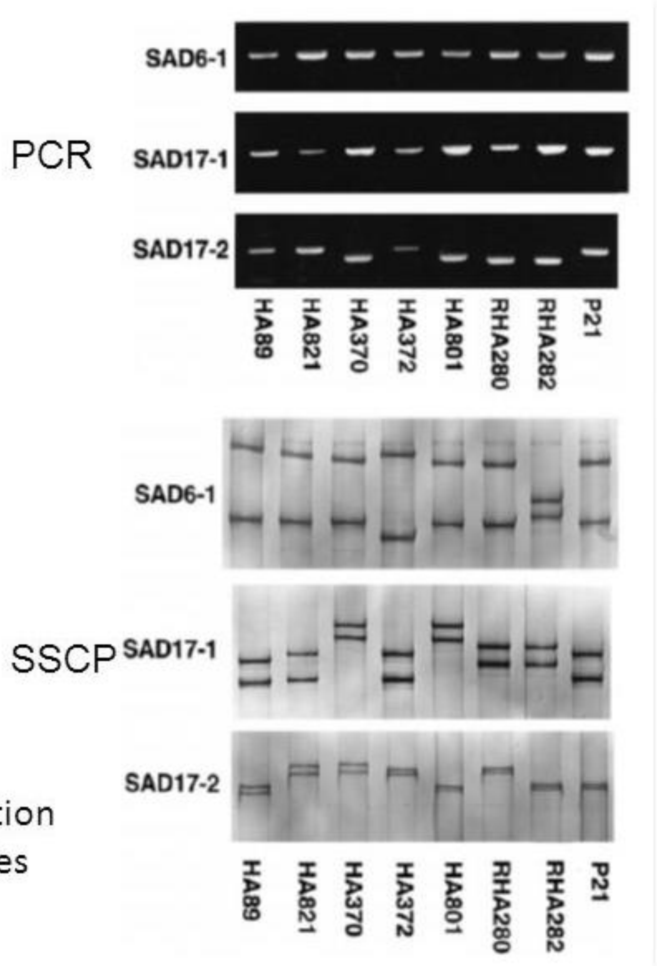


```
....|....|....|....|....|....|....|....|....|....|
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
```

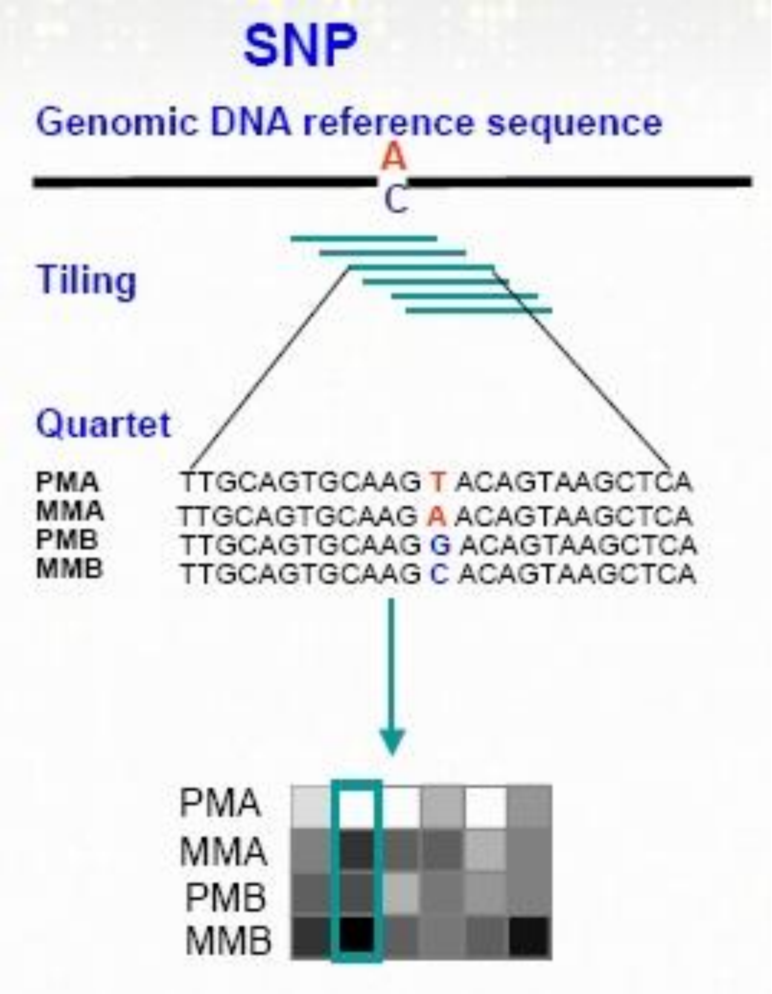
single-strand conformation polymorphism (SSCP)

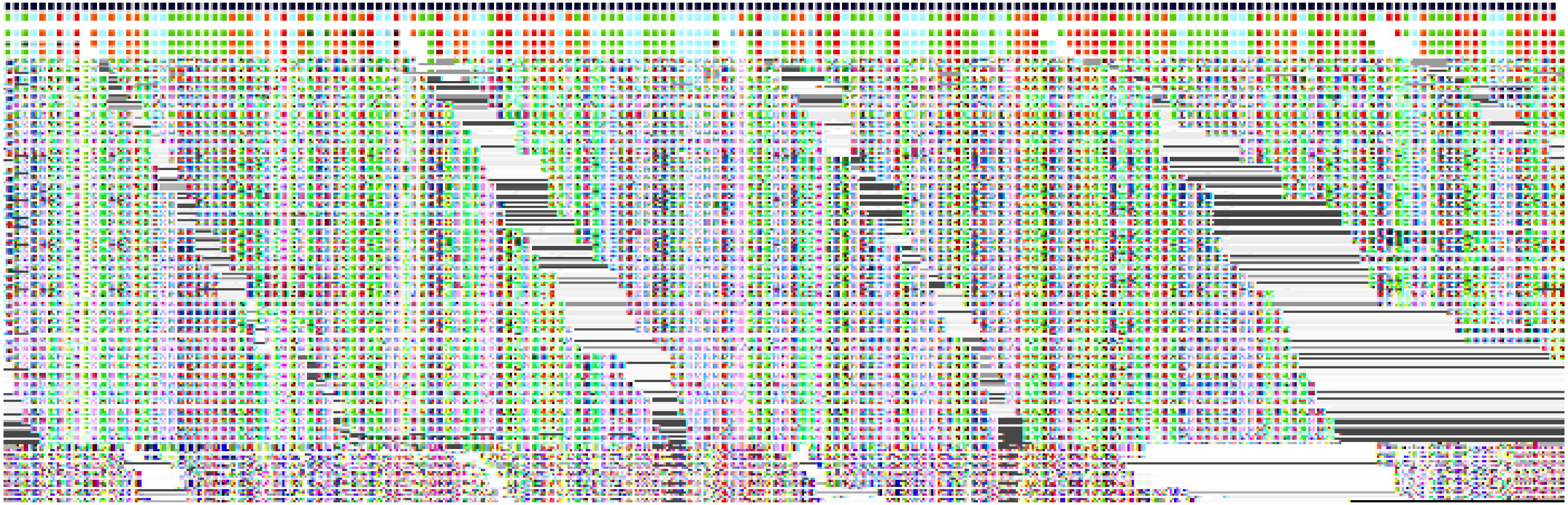


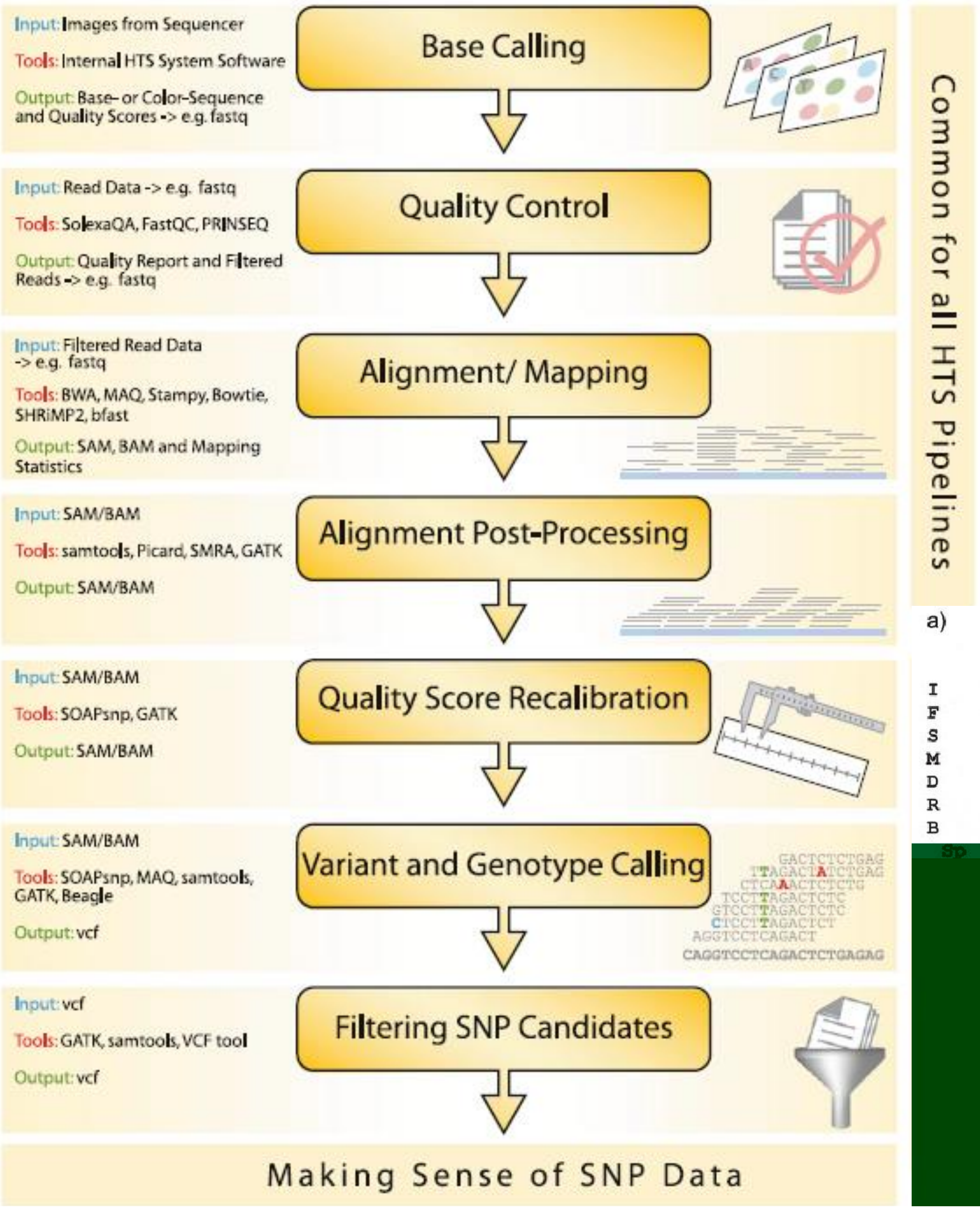
The electrophoretic mobility of separation is a function of the shape of the folded, single-stranded molecules



(Hongtrakul et al. 1998)


SNP
Genomic DNA reference sequence
A
C
Tiling
Quartet
PMA TTGCAGTGCAAG T ACAGTAAGCTCA
MMA TTGCAGTGCAAG A ACAGTAAGCTCA
PMB TTGCAGTGCAAG G ACAGTAAGCTCA
MMB TTGCAGTGCAAG C ACAGTAAGCTCA
PMA
MMA
PMB
MMB





Input: Images from Sequencer

Tools: Internal HTS System Software

Output: Base- or Color-Sequence and Quality Scores -> e.g. fastq

Base Calling

Input: Read Data -> e.g. fastq

Tools: SolexaQA, FastQC, PRINSEQ

Output: Quality Report and Filtered Reads -> e.g. fastq

Quality Control

Input: Filtered Read Data -> e.g. fastq

Tools: BWA, MAQ, Stampy, Bowtie, SHRiMP2, bfast

Output: SAM, BAM and Mapping Statistics

Alignment/ Mapping

Input: SAM/BAM

Tools: samtools, Picard, SMRA, GATK

Output: SAM/BAM

Alignment Post-Processing

Input: SAM/BAM

Tools: SOAPsnp, GATK

Output: SAM/BAM

Quality Score Recalibration

Input: SAM/BAM

Tools: SOAPsnp, MAQ, samtools, GATK, Beagle

Output: vcf

Variant and Genotype Calling

Input: vcf

Tools: GATK, samtools, VCF tool

Output: vcf

Filtering SNP Candidates

Making Sense of SNP Data

Common for all HTS Pipelines

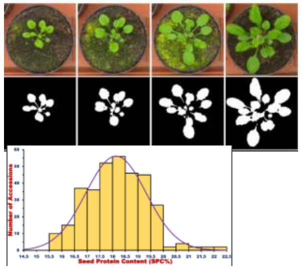
a)

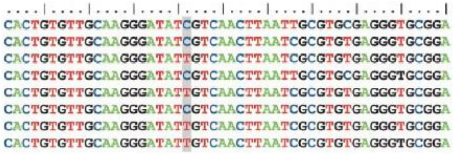
```
           140       150       160       170       180
    ....|....|....|....|....|....|....|....|....|....|
I   CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
F   CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
S   CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
M   CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
D   CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
R   CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
B   CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
```

-

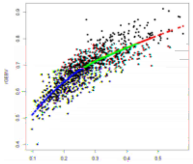
-

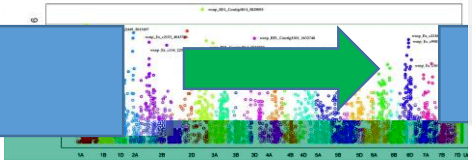
-





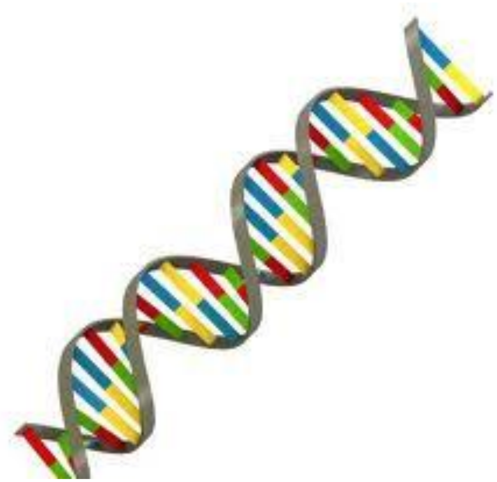
```
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
```

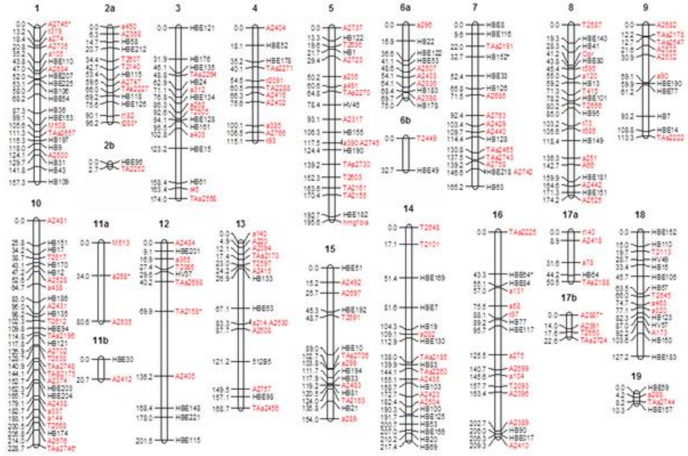




Quantitative Trait Loci (QTL)

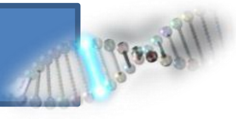
Trends in Plant Science



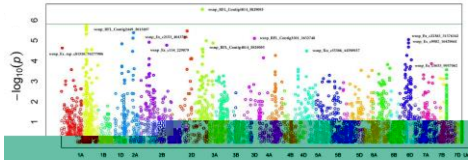


Souza LM, et al. 2013.

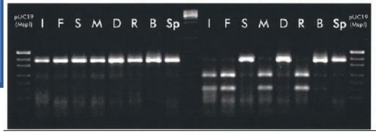
Association Analysis



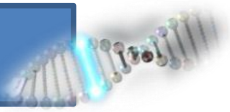
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA

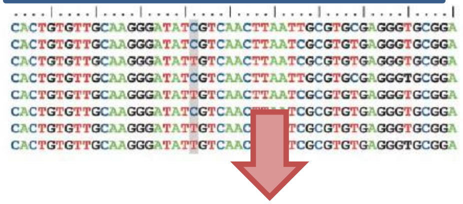


Marker Assisted Selection



Genomic Selection



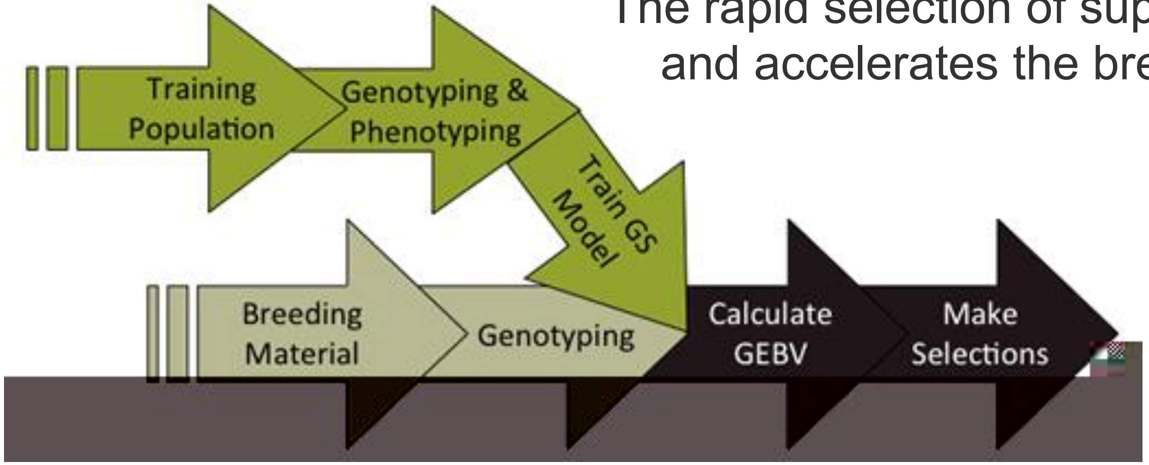


```
....|....|....|....|....|....|....|....|....|....|
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAACTTAATTGCGTGCGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAACTTAATCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATCGTCAAC    TCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAAC    TCGCGTGTGAGGGTGCGGA
CACTGTGTTGCAAGGGATATTGTCAAC    TCGCGTGTGAGGGTGCGGA
```

ASreml®

R

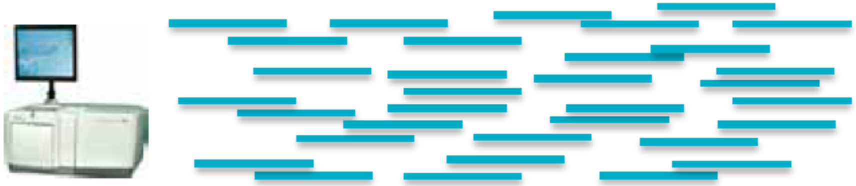
“The rapid selection of superior genotypes and accelerates the breeding cycle”

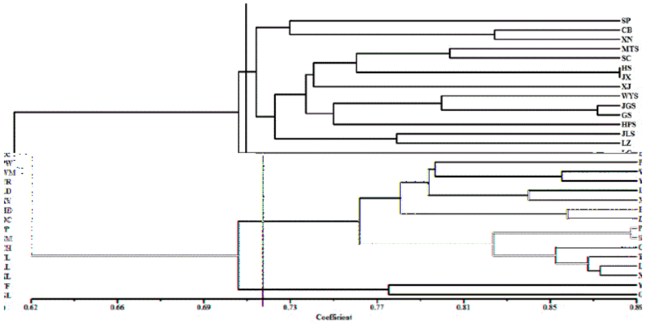


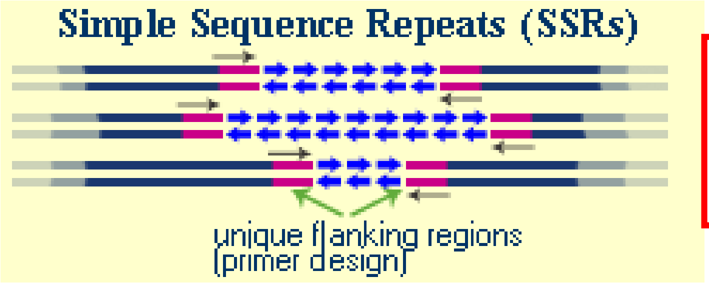
Genomic Estimated Breeding Value (GEBV)

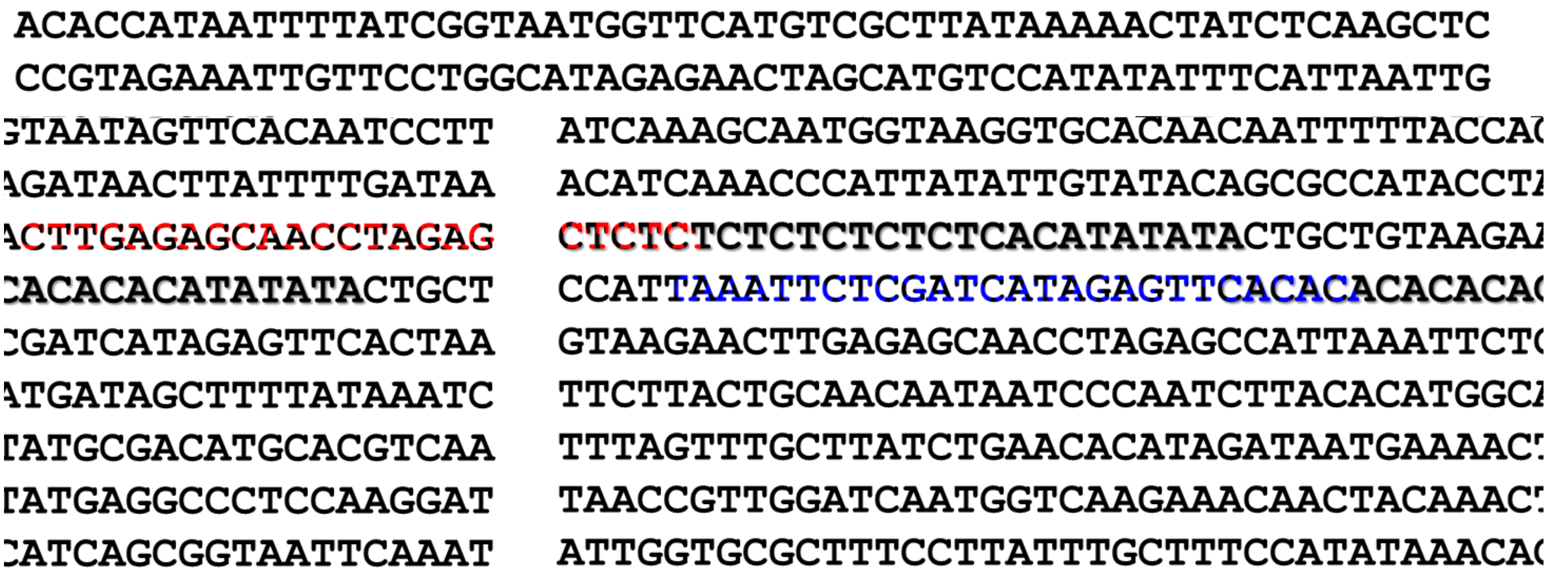
Crossa J. et al. Trends Plant Sci. 2017 Nov;22(11)

BReeDServe
breedserve.cab.kps.ku.ac.th




SP
CB
XN
MTS
SC
HS
JX
XJ
WYS
JGS
GS
HPS
JLS
LZ
0.62
0.66
0.69
0.73
0.77
0.81
0.85
Coefficient


Simple Sequence Repeats (SSRs)
unique flanking regions
(primer design)

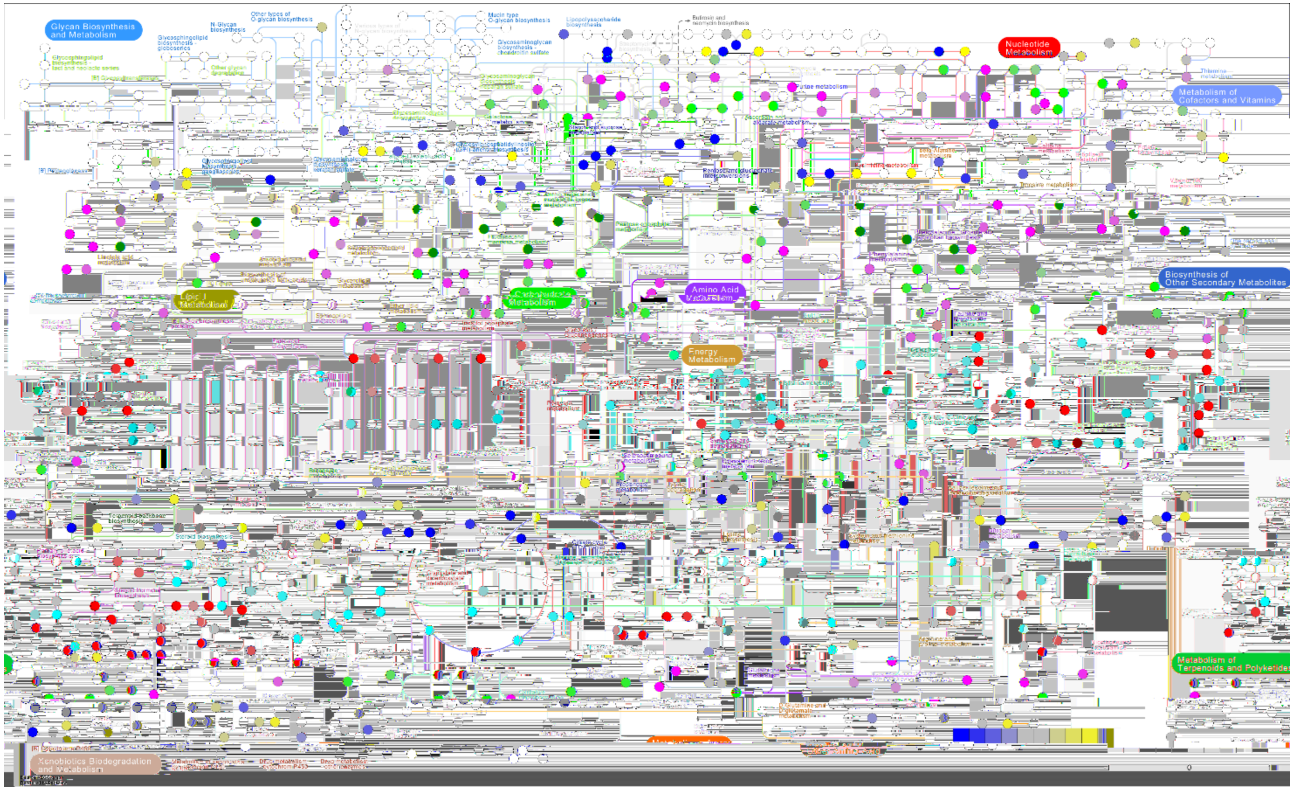


```
ACACCATAATTTTATCGGTAATGGTTCATGTCGCTTATAAAAACTATCTCAAGCTC
CCGTAGAAATTGTTCCTGGCATAGAGAACTAGCATGTCCATATATTTCATTAATTG
GTAATAGTTCACAATCCTT    ATCAAAGCAATGGTAAGGTGCACAACAATTTTTACCAC
AGATAACTTATTTTGATAA    ACATCAAACCCATTATATTGTATACAGCGCCATACCTA
ACTTGAGAGCAACCTAGAG    CTCTCTCTCTCTCTCTCACATATATACTGCTGTAAGAA
CACACACATATATACTGCT    CCATTAAATTCTCGATCATAGAGTTCACACACACACAC
CGATCATAGAGTTCACTAA    GTAAGAACTTGAGAGCAACCTAGAGCCATTAAATTCTC
ATGATAGCTTTTATAAATC    TTCTTACTGCAACAATAATCCCAATCTTACACATGGCA
TATGCGACATGCACGTCAA    TTTAGTTTGCTTATCTGAACACATAGATAATGAAAACT
TATGAGGCCCTCCAAGGAT    TAACCGTTGGATCAATGGTCAAGAAACAACTACAAACT
CATCAGCGGTAATTCAAAT    ATTGGTGCGCTTTCCTTATTTGCTTTCCATATAAACAC
```

Sequences in Red: repeat motif - (CT)8(CA)2(TA)3; Sequences in Blue: repeat motif – (CA)10(TA)3

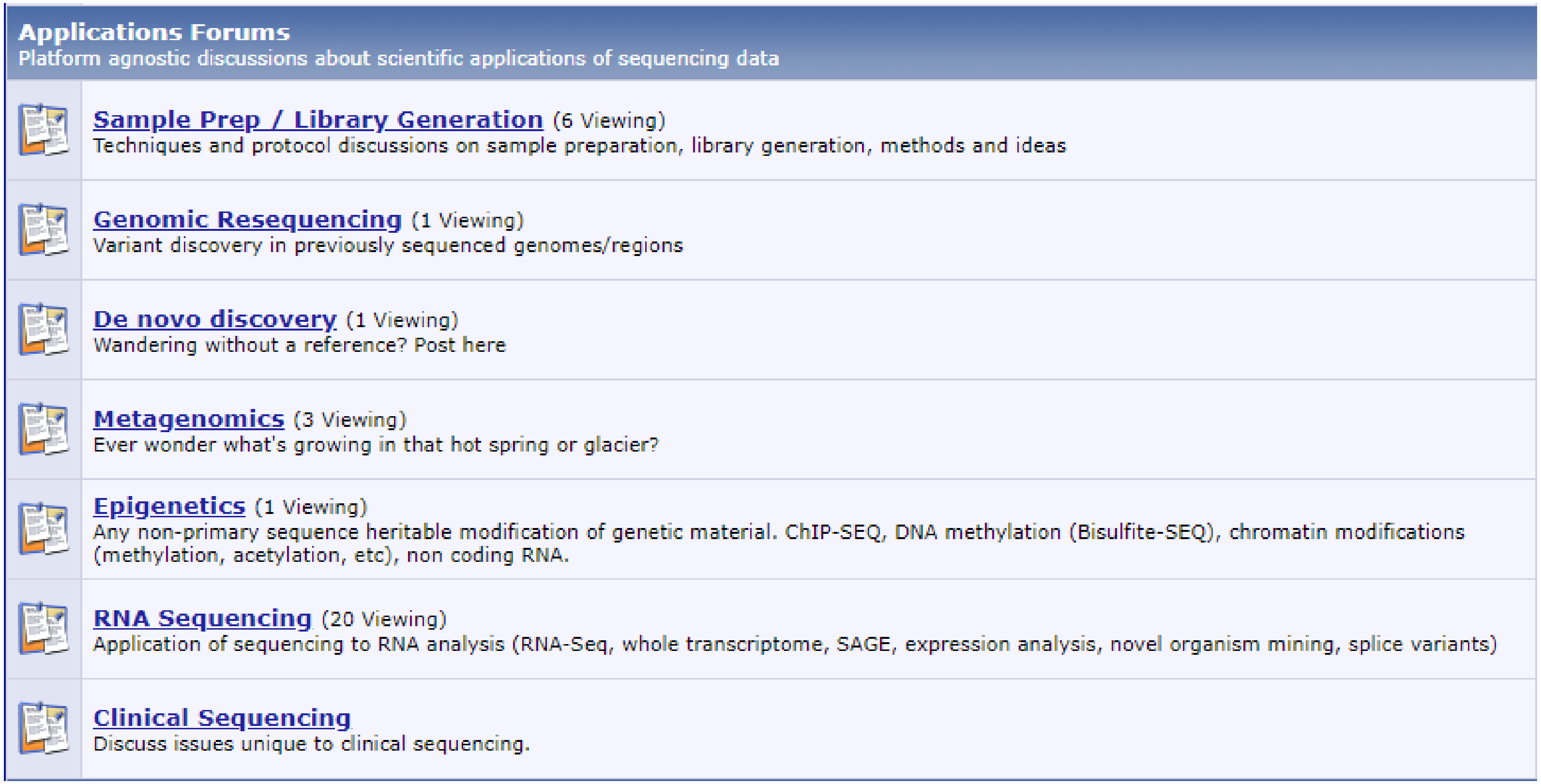
●

●


Glycan Biosynthesis and Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Biosynthesis of Other Secondary Metabolites
Energy Metabolism
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
N-Glycan biosynthesis
Other types of O-glycan biosynthesis
Mucin type O-glycan biosynthesis
Lipopolysaccharide biosynthesis
Glycosphingolipid biosynthesis - globoseries
Glycosaminoglycan biosynthesis - chondroitin sulfate
Thiamine metabolism

SEQanswers
the next generation sequencing community

Biostars
BIOINFORMATICS EXPLAINED



Applications Forums
Platform agnostic discussions about scientific applications of sequencing data

Sample Prep / Library Generation (6 Viewing)
Techniques and protocol discussions on sample preparation, library generation, methods and ideas

Genomic Resequencing (1 Viewing)
Variant discovery in previously sequenced genomes/regions

De novo discovery (1 Viewing)
Wandering without a reference? Post here

Metagenomics (3 Viewing)
Ever wonder what's growing in that hot spring or glacier?

Epigenetics (1 Viewing)
Any non-primary sequence heritable modification of genetic material. ChIP-SEQ, DNA methylation (Bisulfite-SEQ), chromatin modifications (methylation, acetylation, etc), non coding RNA.

RNA Sequencing (20 Viewing)
Application of sequencing to RNA analysis (RNA-Seq, whole transcriptome, SAGE, expression analysis, novel organism mining, splice variants)

Clinical Sequencing
Discuss issues unique to clinical sequencing.

HikingArtist.com