The lettuce genome

Presentation by: **Anwar Haidar Alarbash**
Course: intro. To genomics (485) Spring 2020

Outline

- Genome paper
- Interesting facts
- General information about the genome
- Sequencing strategy
- Sequencing method
- Assembly
- Genome outcome
- questions

Genome papers

Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce

Sebastian Reyes-Chin-Wo[1,*], Zhiwen Wang[2,*,†], Xinhua Yang[2,*], Alexander Kozik[1], Siwaret Arikit[3,†], Chi Song[2], Liangfeng Xia[2], Lutz Froenicke[1], Dean O. Lavelle[1], María-José Truco[1], Rui Xia[4], Shilin Zhu[2], Chunyan Xu[2], Huaqin Xu[1], Xun Xu[2], Kyle Cox[1], Ian Korf[1,5], Blake C. Meyers[3,4] & Richard W. Michelmore[1,5,6,7]

Interesting facts

- Lettuce (Lactuca sativa) is an important vegetable crop species and ranks as one of the top 10 most valuable crops in the USA with an annual value of over $2.4 billion , second to potatoes.

- Lettuce was also very important for the ancient Romans and ancient Greeks who believed that lettuce induce sleeping. The roman Emperor Domitian (81-96 AD) served it at the beginning of his feasts, so he could torture his guests by forcing them to stay awake in the presence of the Emperor.

- Iceberg lettuce got it's name from the fact that California growers started shipped it covered with heaps of crushed ice in the 1920s. It had previously been called Crisphead lettuce.

General information about the genome

- L. sativa is diploid with 2n= 2x = 18 chromosomes and an estimated genome size of 2.5 Gb .

- They sequenced the protein coding genes from an interspecific cross of L. sativa x L. serriola acc.

- L. sativa showed an intermediate level of synteny to V. vinifera. Visualization of the syntenic hits between these two genomes suggested a whole-genome triplication event in lettuce since divergence from the grape lineage.
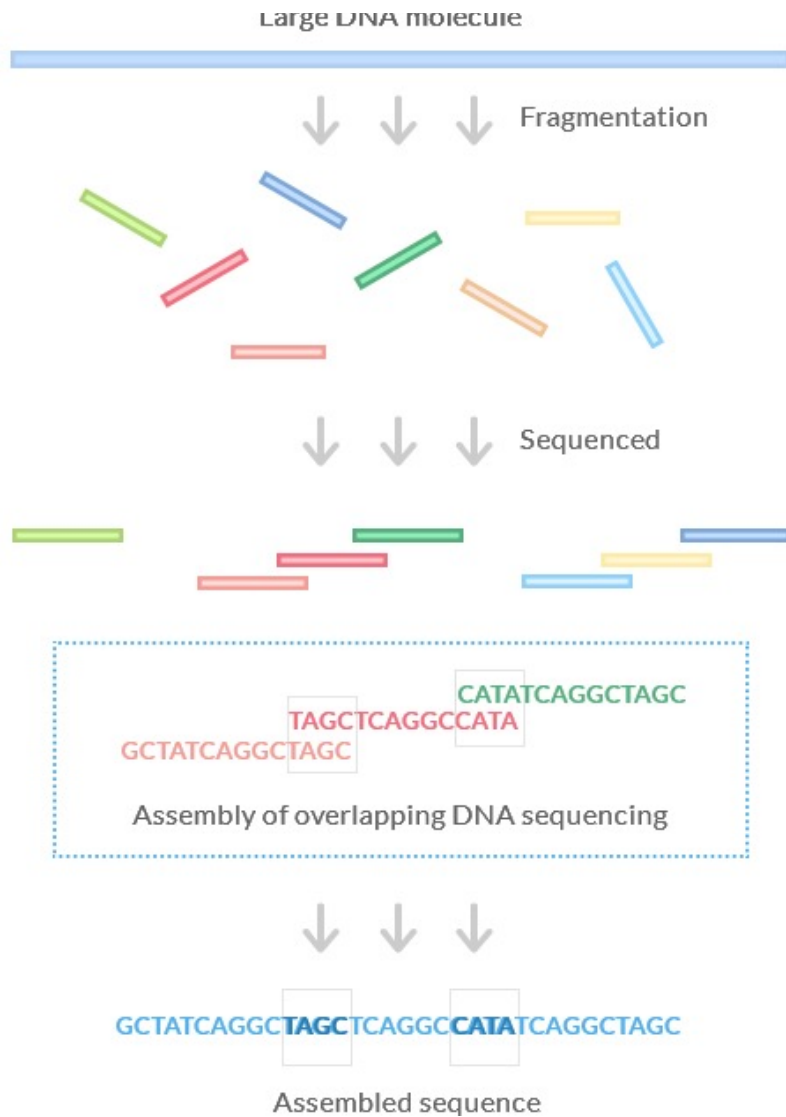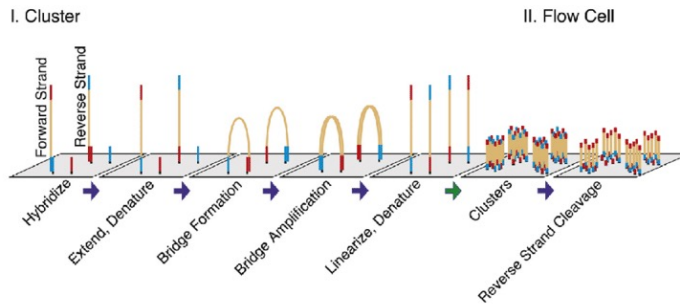
L.Sativia

L.Serriola (prickly lettuce)

Sequencing strategy

- Whole-genome shot gun strategy was used was used to sequence and assemble the genome of L. sativa cultivar Salinas from Illumina short reads.
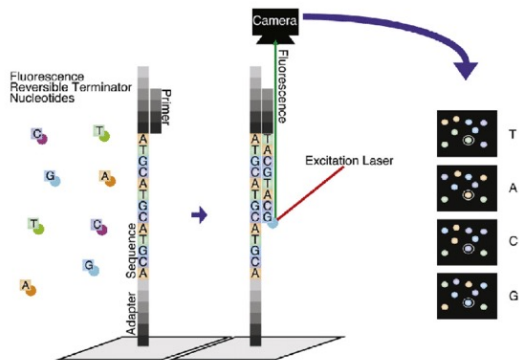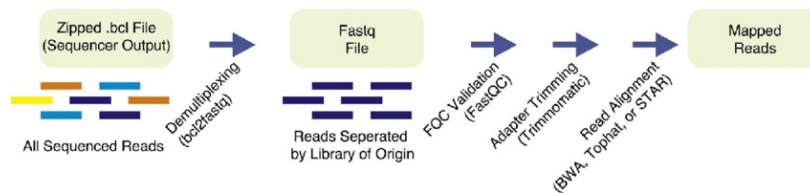
Sequencing methods

Ø the used illumina hiseq 2000 as and hirise scaffolding are sequencing method to prepare two Chicago libraries.

A. Clustering

I. Cluster

II. Flow Cell

Forward Strand

Reverse Strand

Hybridize → Extend, Denature → Bridge Formation → Bridge Amplification → Linearize, Denature → Clusters → Reverse Strand Cleavage

B. High-throughput sequencing

Camera

Fluorescence Reversible Terminator Nucleotides

Primer

Fluorescence

Excitation Laser

T
A
C
G

Adapter   Sequence

C. Demultiplexing samples and read mapping

Zipped .bcl File (Sequencer Output) → Demultiplexing (bcl2fastq) → Fastq File → FQC Validation (FastQC) → Adapter Trimming (Trimmomatic) → Read Alignment (BWA, Tophat, or STAR) → Mapped Reads

All Sequenced Reads   Reads Seperated by Library of Origin

Assembly

➢ SOAPdenovo was used for assembly.

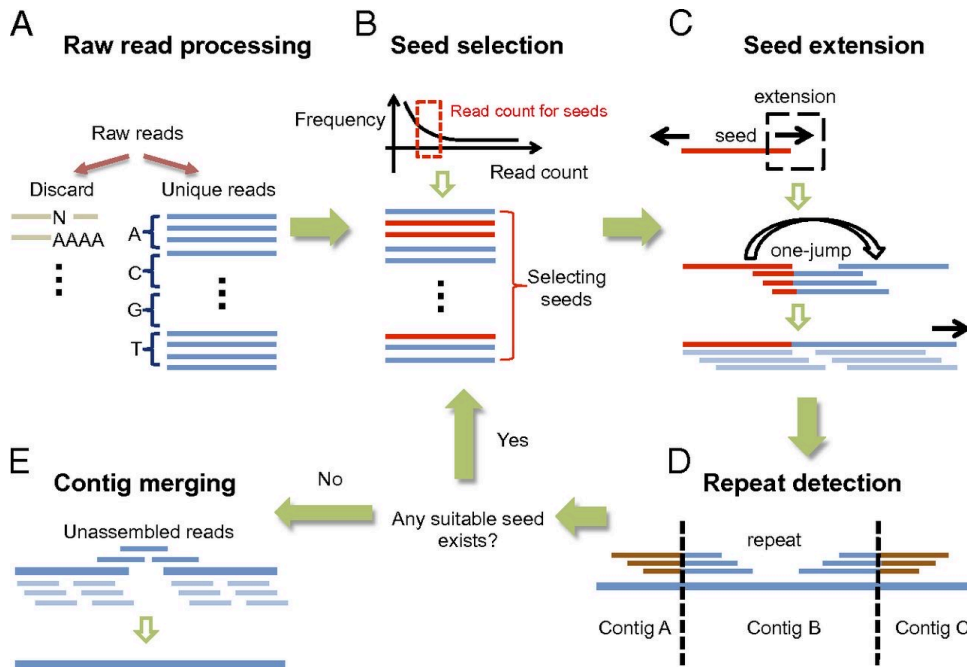**Table 1 | Assembly statistics for the genome of *L. sativa* cv. Salinas.**

| Genome assembly metrics		SOAPDenovo	HiRise (2 lanes)
Contigs			
N50 (size/number)		36 kb/21,116	—
Largest		253 kb	—
Total size		2.21 Gb	—
Total number		153,952	—
Scaffolds			
N50 (size/number)		476 kb/1,445	1.8 Mb
N90 (size/number)		118 kb/5,237	360 kb/1,520
Largest		3.1 Mb	12.2 Mb
Total size		2.38 Gb	2.38 Gb
Total number		21,686	11,474

Genome annotation*			
		Family	Total Length
Transposable elements		Retroelements	1.5 Gb (61.5%)
		DNA elements	29.5 Mb (1.2%)
		MITEs	103.7 Mb (4.4%)
		Others	115.3 kb (<1%)
		Unknown	152.9 Mb (6.3%)
		Total	1.8 Gb (74.2%)
		Type	Copies
Non-coding RNA		rRNAs	2,587
		tRNAs	1,347
		Predicted miRNAs	483
		Detected miRNAs	86
		snRNAs	1,514
Protein coding genes		Total number	38,919
		Annotated transcripts	31,348
		Average CDS length	1.05 kb

MITE, Miniature Inverted-Repeat Transposable Elements.
*Annotation provided for HiRise assembly.

b

Distribution of sequence divergence for syntelog pairs
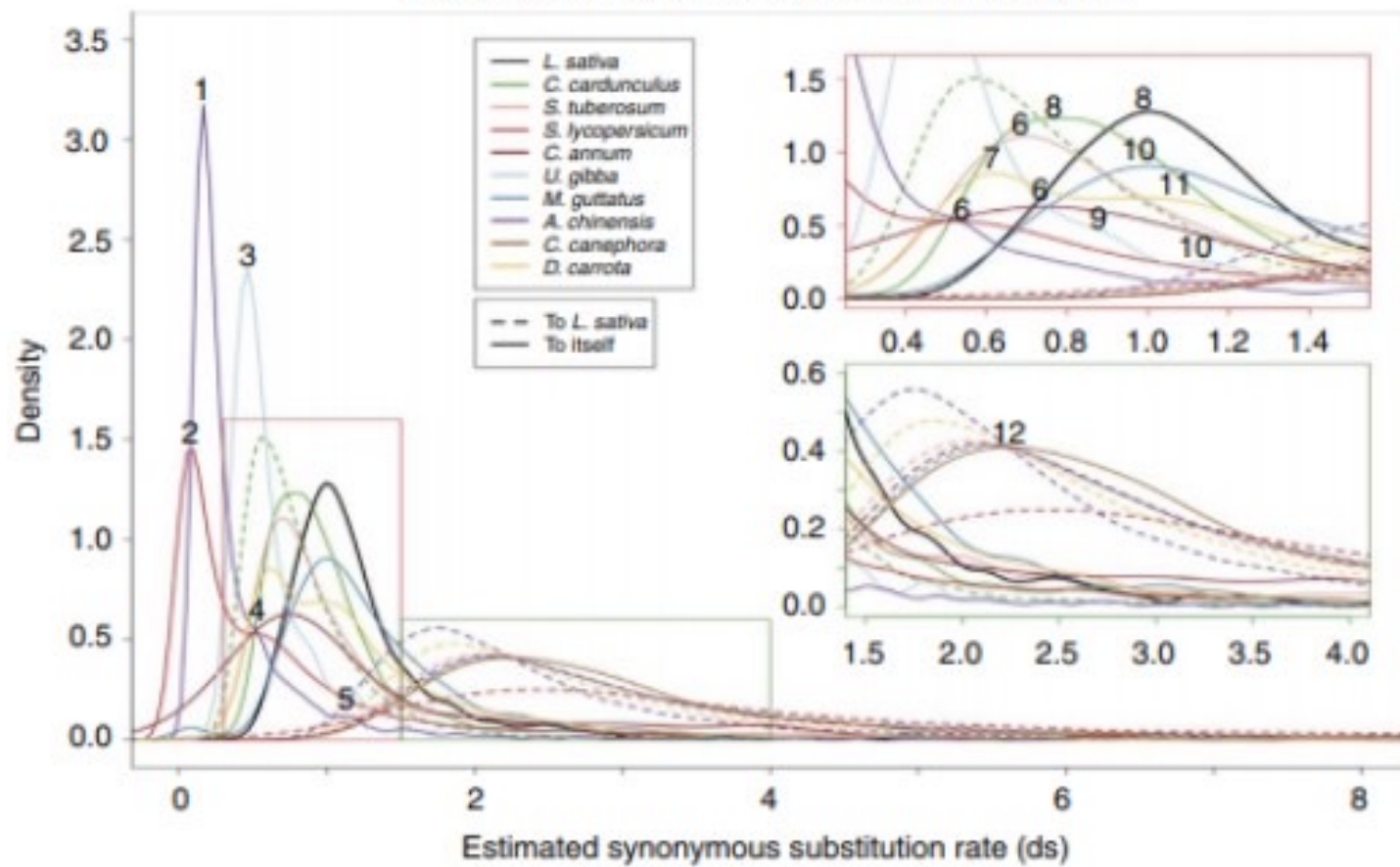
Genome outcome

➢ Genome size: 2.38Gb
➢ Covering almost 88% of the genome done
➢ GC% : between 35% to 38%
➢ 1.8 Gb from 2.38Gb assembled genome is comprised of repetitive elements.
➢ LINE : 15,767
➢ LTR : 1,584,480 (specifically the Gypsy (33.9%) and Copia (24.9%) subfamilies)
➢ SINE: 9,260
➢ Protein coding genes : 38,919

Questions

1- What is the substance in lettuce that induces sleep?

- Lactucin

2- what is '' hirise scaffolding '' ?

-it is new software pipeline that can identify poor quality joins and produce accurate, long-range sequence scaffolds it is used to construct a highly accurate de novo assembly and scaffolding .