



***The African Eggplant
(Solanum Aethiopicum)***

Introduction to
Genomics 0497-485

Presented by:
Muneerah.H.AISabah

Spring
2020

Presentation outline

1. The genome paper.
2. Fun facts.
3. General information about the genome.
4. Sequencing strategy.
5. Sequencing method.
6. Assembly.
7. Genome outcome.
8. Questions.



The Genome Paper



GigaScience, 8, 2019, 1–16

doi: 10.1093/gigascience/giz115
Research

RESEARCH

Draft genome sequence of *Solanum aethiopicum* provides insights into disease resistance, drought tolerance, and the evolution of the genome

Bo Song^{1,2,3,†}, Yue Song^{1,3,4,†}, Yuan Fu^{1,2,†}, Elizabeth Balyejusa Kizito⁵, Sandra Ndagire Kamenya^{5,6}, Pamela Nahamya Kabod⁵, Huan Liu^{1,2,3}, Samuel Muthemba⁷, Robert Kariba⁷, Joyce Njuguna⁶, Solomon Maina⁶, Francesca Stomeo^{6,8}, Appolinaire Djikeng^{6,9}, Prasad S. Hendre⁷, Xiaoli Chen^{1,2}, Wenbin Chen^{1,2}, Xiuli Li^{1,2}, Wenjing Sun^{1,2}, Sibow Wang^{1,3}, Shifeng Cheng^{1,2}, Alice Muchugi⁷, Ramni Jamnadass⁷, Howard-Yana Shapiro^{7,10,11}, Allen Van Deynze¹⁰, Huanming Yang^{1,2}, Jian Wang^{1,2}, Xun Xu^{1,2,3}, Damaris Achieng Odeny^{12,*} and Xin Liu^{1,2,3,*}

¹BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China; ²China National GeneBank, BGI-Shenzhen, Jinsha Road, Shenzhen 518120, China; ³State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China; ⁴BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China; ⁵Uganda Christian University, Bishop Tucker Road, Box 4, Mukono, Uganda; ⁶Biosciences Eastern and Central Africa (BeCA) – International Livestock Research Institute (ILRI) Hub, P.O. Box 30709, Nairobi 00100, Kenya; ⁷African Orphan Crops Consortium, World Agroforestry Centre (ICRAF), United Nations Avenue, Nairobi 00100, Kenya; ⁸Present address: European Molecular Biology Laboratory (EMBL), Heidelberg, Germany; ⁹Present address: Centre for Tropical Livestock Genetics and Health (CTLGH), University of Edinburgh, Edinburgh EH25 9RG, UK; ¹⁰University of California, 1 Shields Ave, Davis, CA, USA; ¹¹Mars, Incorporated, 6885 Elm Street, McLean, VA 22101, USA and ¹²International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) –



Fun Facts

African eggplant (*Solanum aethiopicum*) VS Normal eggplant (*Solanum melongena*)



Fun Facts

- African eggplants are commonly known as Mock Tomato, Bitter Tomato, African nightshade, Garden eggs and Scarlet Eggplant, they are a fruiting plant of the genus *Solanum* mainly found in Asia and Tropical Africa.
- These names are a result of its varied morphology, with ripe fruit often looking like a cross between an eggplant and a tomato, the African eggplant was so much confused with the ordinary eggplant.
- African eggplants have a distinctly bitter taste that becomes more bitter as it ripens.
- The bitterness depends on the levels of saponin it contains, some with a sweet flavor and others very bitter. When the berries mature, they turn bright red because of high carotene content.
- It can be used like eggplant (*Solanum melongena*) as a vegetable or as a flavoring for other foods.
- The roots and leaf's are used as a sedative, and to treat colic and high blood pressure.
- African eggplant has been used as a source of disease resistance genes for several commercially grown crops, including *Solanum melongena*(eggplant).

Fun Facts

- African eggplants have been a staple in African culture and cuisine. A popular item sold by street vendors, 80% of the total eggplant production in Africa comes from small family farms. African eggplants are consumed on a regular basis in households and are popularly used in curry dishes. Ancient folklore also mentions this fruit as a blessing of fertility, and African eggplants are often given as a gift during weddings and child naming ceremonies.



General Information About The Genome

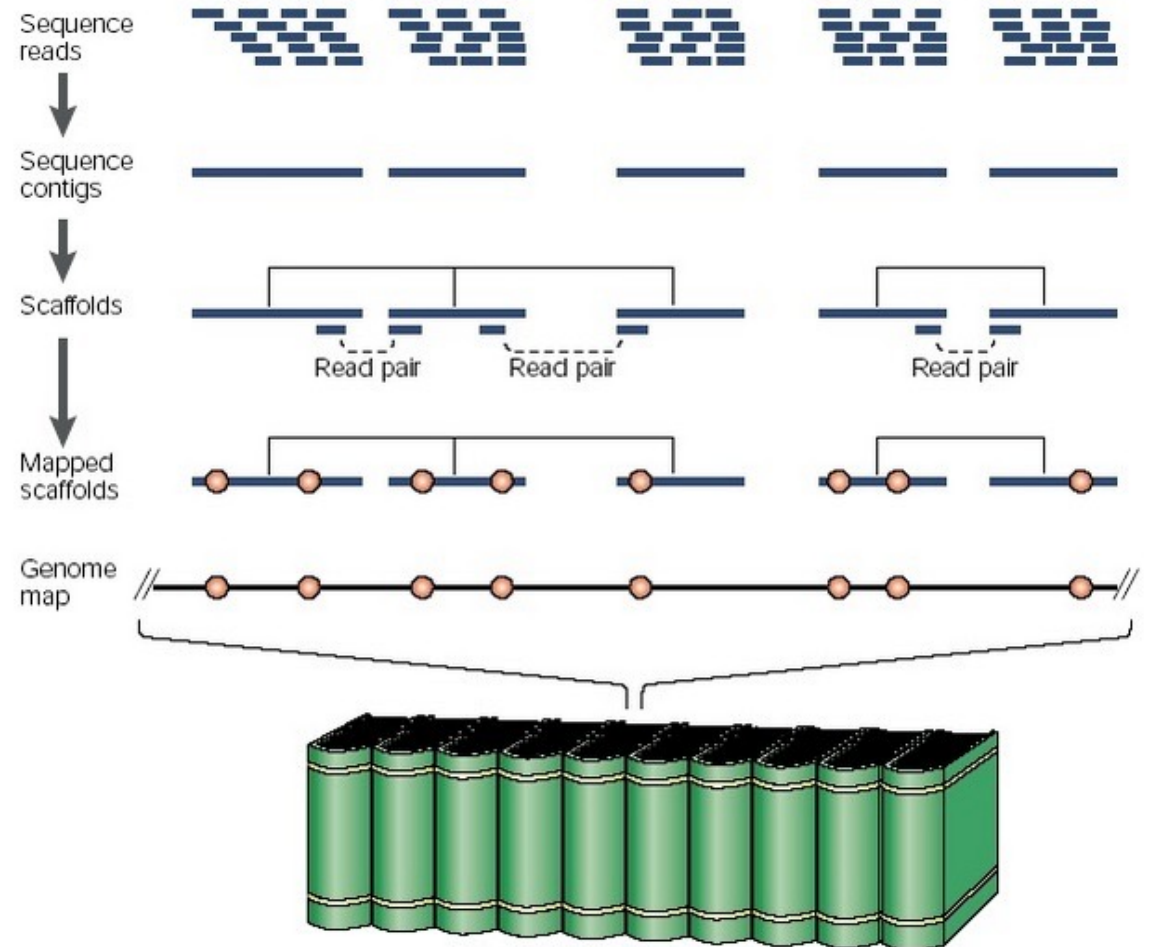
- Genome Size: 1.17 Gb
- Level of ploidy: Diploid Homozygous
- Chromosome Number: 12 **chromosomes** ($2n = 24$)
- Protein Coding Genes: 34.906
- GC Ratio: 33.12%
- Repetitive Sequences: 78.9%
- 2 rounds of amplification of long terminal repeat retrotransposons (LTR-Rs)
- Number of genes: 34.906
- Average/total coding sequence length: 1104.3 bp/38.5 Mb
- Source of sample: young leaves of 14-day-old seedlings of *Solanum aethiopicum* (African eggplant).



Sequencing strategy

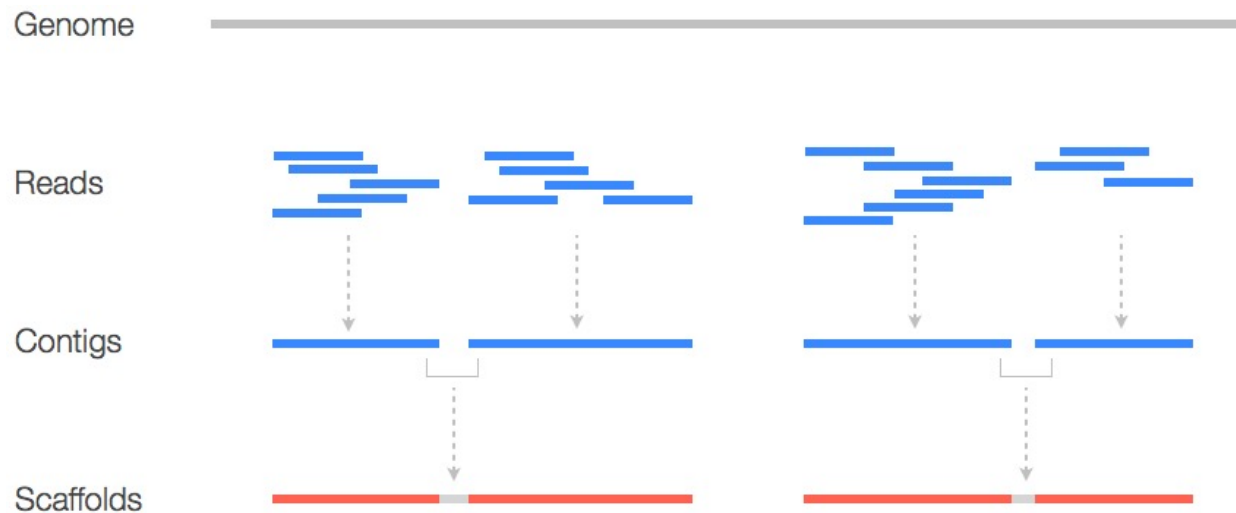
- Whole genome shotgun approach was used: which is a method that entails sequencing many overlapping DNA fragments in parallel and then using a computer to assemble the small fragments into larger contigs and, eventually, chromosomes
- A total of 242.61 Gb raw reads were generated

	Scaffolds	Contigs
Number	162,187	231,821
Total length	1.02 Gb	936 Mb
N50	516.1 kb	25.2 kb
Longest	2.94 Mb	366.2 kb



Sequencing strategy

- A **scaffold** is a portion of the genome sequence reconstructed from end-sequenced whole-genome shotgun clones. **Scaffolds** are composed of **contigs** and gaps. A **contig** is a contiguous length of genomic sequence in which the order of bases is known to a high confidence level.



	Scaffolds	Contigs
Number	162,187	231,821
Total length	1.02 Gb	936 Mb
N50	516.1 kb	25.2 kb
Longest	2.94 Mb	366.2 kb

Sequencing Method

- Illumina HiSeq 2000 platform, which is a computer system that uses the method based on sequencing-by-synthesis (SBS), and reversible dye-terminators that enable the identification of single bases as they are introduced into DNA strands.
- analysis of the sequence revealed the *S. aethiopicum* genome to be diploid and homozygous, with an estimated genome size of 1.17 Gb . “
- Clean reads” amounting to 127.83 Gb (~109×) were used to assemble the genome.

	Scaffolds	Contigs
Number	162,187	231,821
Total length	1.02 Gb	936 Mb
N50	516.1 kb	25.2 kb
Longest	2.94 Mb	366.2 kb



Assembly

- Platanus: is a novel *de novo* sequence assembler that can reconstruct genomic sequences of diploids from massively parallel shotgun sequencing data.

- Clean reads amounting to 127.83 Gb (~109×) were used to assemble the genome using Platanus de nova.
- A final assembly of 1.02 Gb in size was obtained, containing 162,187 scaffolds with N50 contig and scaffold values of 25.2 and 516.15 kb

	Scaffolds	Contigs
Number	162,187	231,821
Total length	1.02 Gb	936 Mb
N50	516.1 kb	25.2 kb
Longest	2.94 Mb	366.2 kb

Genome Outcome

- a 1.02-Gb draft genome of *S. aethiopicum*, which contained predominantly repetitive sequences (78.9%), was assembled
- 37,681 gene models, including 34,906 protein-coding genes, was annotated.
- Expansion of disease resistance genes was observed via 2 rounds of amplification of long terminal repeat retrotransposons (LTR-Rs which may have occurred ~1.25 and 3.5 million years ago.
- By resequencing *S. aethiopicum*, 18,614,838 single-nucleotide polymorphisms were identified, of which 34,171 were located within disease resistance genes.
- Analysis of domestication and demographic history revealed active selection for genes involved in drought tolerance in *S. aethiopicum*
- A pan-genome of *S. aethiopicum* was assembled, containing 51,351 protein-coding genes; 7,069 of these genes were missing from the reference genome.
- The results reveal that the *S. aethiopicum* genome is larger than that of other *Solanum* genomes

	Scaffolds	Contigs
Number	162,187	231,821
Total length	1.02 Gb	936 Mb
N50	516.1 kb	25.2 kb
Longest	2.94 Mb	366.2 kb

Genome Outcome

- *Solanum aethiopicum* is cross-compatible with *S. melongena* and is routinely used as a donor of disease resistance genes to its close relative
- Genomic analysis of *S. aethiopicum* revealed higher LTR-mediated expansion of resistance gene families than its other close relatives.
- The development and expansion of resistance genes is usually accompanied by the amplification of long terminal repeat retrotransposons (LTR-Rs)
- LTR amplification is one of the major forces driving genome evolution.
- The genome sequence of *S. aethiopicum* enhances our understanding of its biotic and abiotic resistance.
- The single-nucleotide polymorphisms identified are immediately available for use by breeders.
- The information provided here will accelerate selection and breeding of the African eggplant, as well as other crops within the Solanaceae family.

	Scaffolds	Contigs
Number	162,187	231,821
Total length	1.02 Gb	936 Mb
N50	516.1 kb	25.2 kb
Longest	2.94 Mb	366.2 kb

Questions

1. what is the major forces driving genome evolution?

LTR's

2. What are **Scaffolds** and **contigs** ?

A **scaffold** is a portion of the genome sequence reconstructed from end-sequenced whole-genome shotgun clones. **Scaffolds** are composed of **contigs** and gaps. A **contig** is a contiguous length of genomic sequence in which the order of bases is known to a high confidence level.