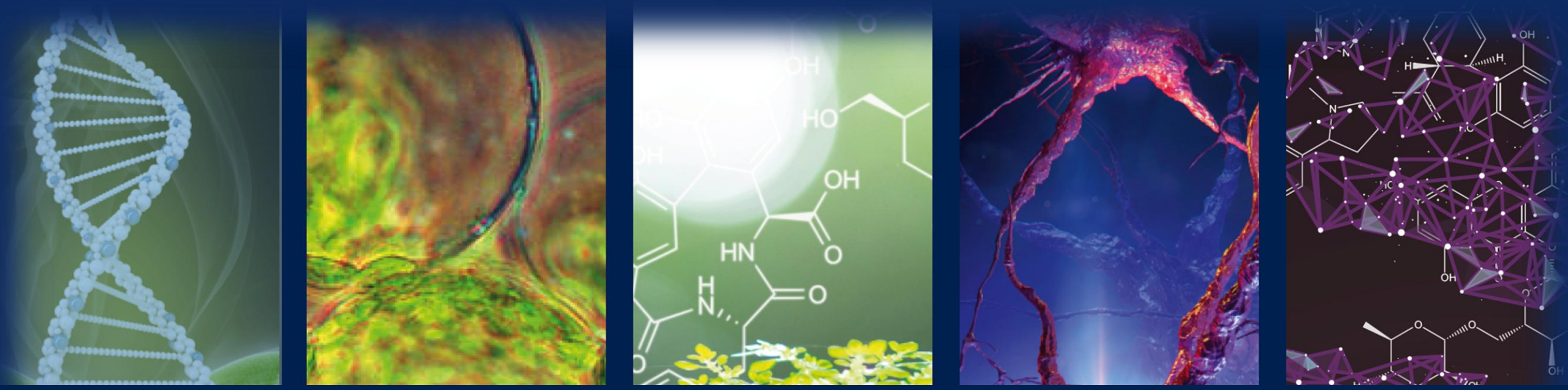


Unlocking the complex genome of *Nicotiana tabacum*

BAT
A BETTER TOMORROW

Dr Parameswari Paul

DISCLAIMER: FOR DISCUSSION PURPOSES ONLY

PRESENTATIONS ARE INTENDED FOR INFORMATIONAL PURPOSES ONLY. STATEMENTS OF FACT AND OPINIONS EXPRESSED ARE THOSE OF THE PRESENTER ONLY, AND ARE NOT NECESSARILY THE OPINIONS, POSITIONS OR POLICIES OF BRITISH AMERICAN TOBACCO (INVESTMENTS) LIMITED OR ITS AFFILIATES ("BAT"). BAT DOES NOT ENDORSE OR APPROVE, AND ASSUMES NO RESPONSIBILITY FOR THE CONTENT, ACCURACY OR COMPLETENESS OF THE INFORMATION PRESENTED. THESE MATERIALS MAY PRESENT CONCEPTS, EXECUTIONS, IDEAS AND THEMES THAT ARE NOT LEGALLY PERMISSIBLE (EITHER IN WHOLE OR IN PART) OR ACCEPTABLE IN CERTAIN JURISDICTIONS, OR WHICH MAY INFRINGE THIRD-PARTY RIGHTS. THIS PRESENTATION MAY CONTAIN REFERENCES TO LAWS AND REGULATIONS THAT MAY CHANGE OVER TIME AND SHOULD BE INTERPRETED IN THAT CONTEXT.

THIS PRESENTATION IS INTENDED FOR SCIENTISTS AND POLICYMAKERS. WE ENCOURAGE PERSONS INTERESTED IN TOBACCO HARM REDUCTION TO CONSIDER ALL POINTS OF VIEW AND SOURCES OF INFORMATION. BAT DOES NOT MAKE HEALTH CLAIMS REGARDING ITS BRANDS. NOTHING CONTAINED HERE SHOULD BE MISCONSTRUED TO THE CONTRARY. TO THE EXTENT THAT THIRD-PARTY SOURCES ARE REFERENCED, NEITHER THE PRESENTER NOR BAT IS RESPONSIBLE FOR THE CONTENT OF REFERENCED SOURCES AND THE VIEWS EXPRESSED MAY NOT REPRESENT THE VIEWS OF THE PRESENTER OR BAT. NO TOBACCO PRODUCT IS SAFE, ALL TOBACCO PRODUCTS CONTAINING NICOTINE ARE ADDICTIVE. YOUTH SHOULD NEVER USE TOBACCO. SMOKERS WHO ARE CONCERNED ABOUT THEIR HEALTH SHOULD QUIT.

THE TOPICS, CONCEPTS AND OTHER INFORMATION DISCUSSED HEREIN ARE NOT FINAL, ARE SUBJECT TO CHANGE AND/OR CANCELLATION AND MAY BE FOR ILLUSTRATIVE OR THEORETICAL PURPOSES ONLY. NO DEFINITIVE PLANS OR COMMITMENTS SHOULD BE INFERRED FROM THESE MATERIALS AND ANY PROPOSED PLANS OR COMMITMENTS ARE SUBJECT IN ALL RESPECTS TO APPLICABLE INTERNAL REVIEW AND GOVERNANCE REQUIREMENTS AND LAWS AND REGULATIONS IN RESPECTIVE JURISDICTIONS.

AGENDA

01 Why Tobacco?

02 Current Statistics of Tobacco Genome

03 What We Lack?

04 Our Aim

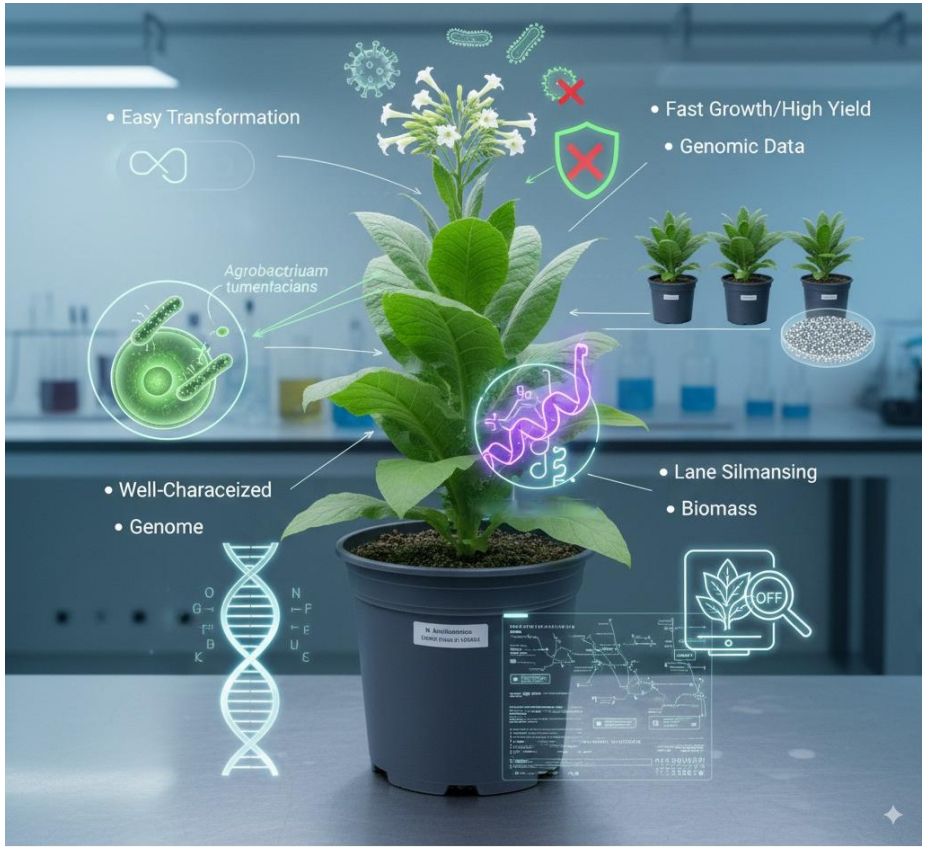
05 NitabV6 genome

06 Conclusion

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

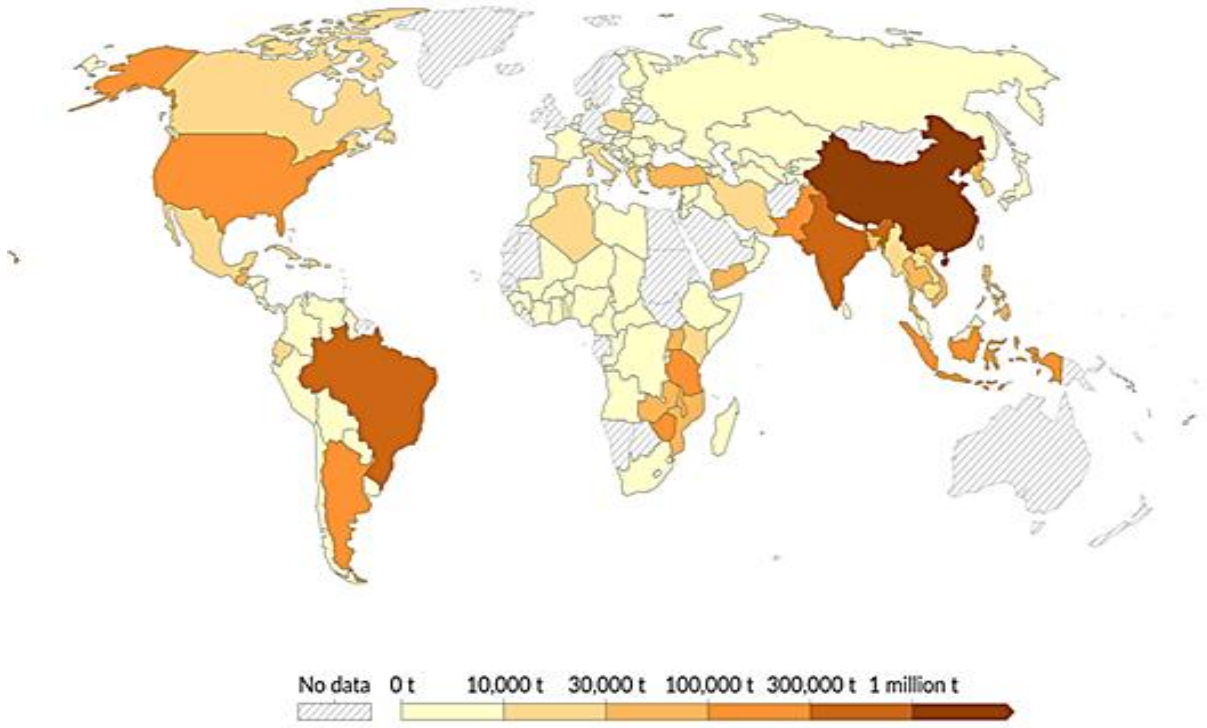
Why Tobacco?

Nicotiana tabacum as a Model Plant



Dependence Worldwide

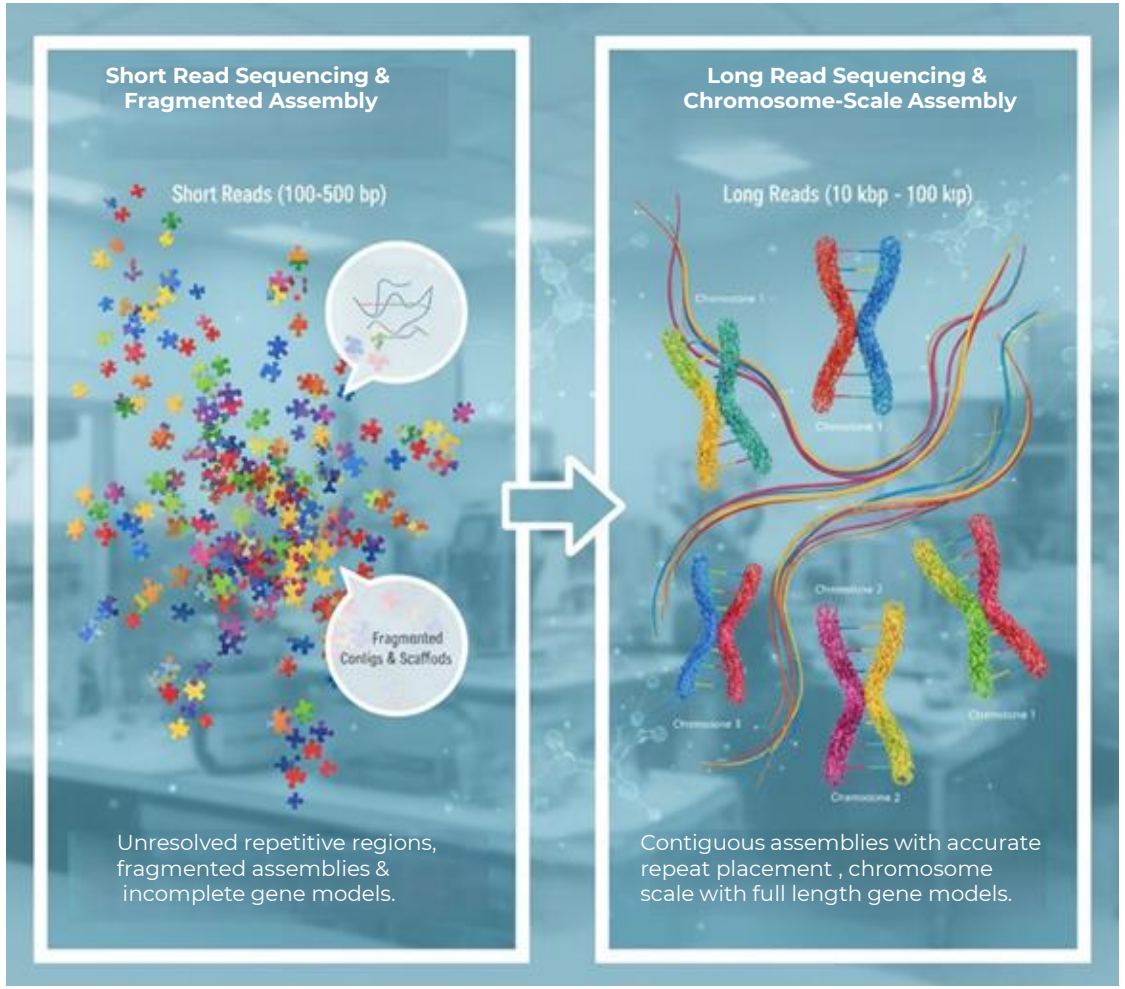
Tobacco production, 2023



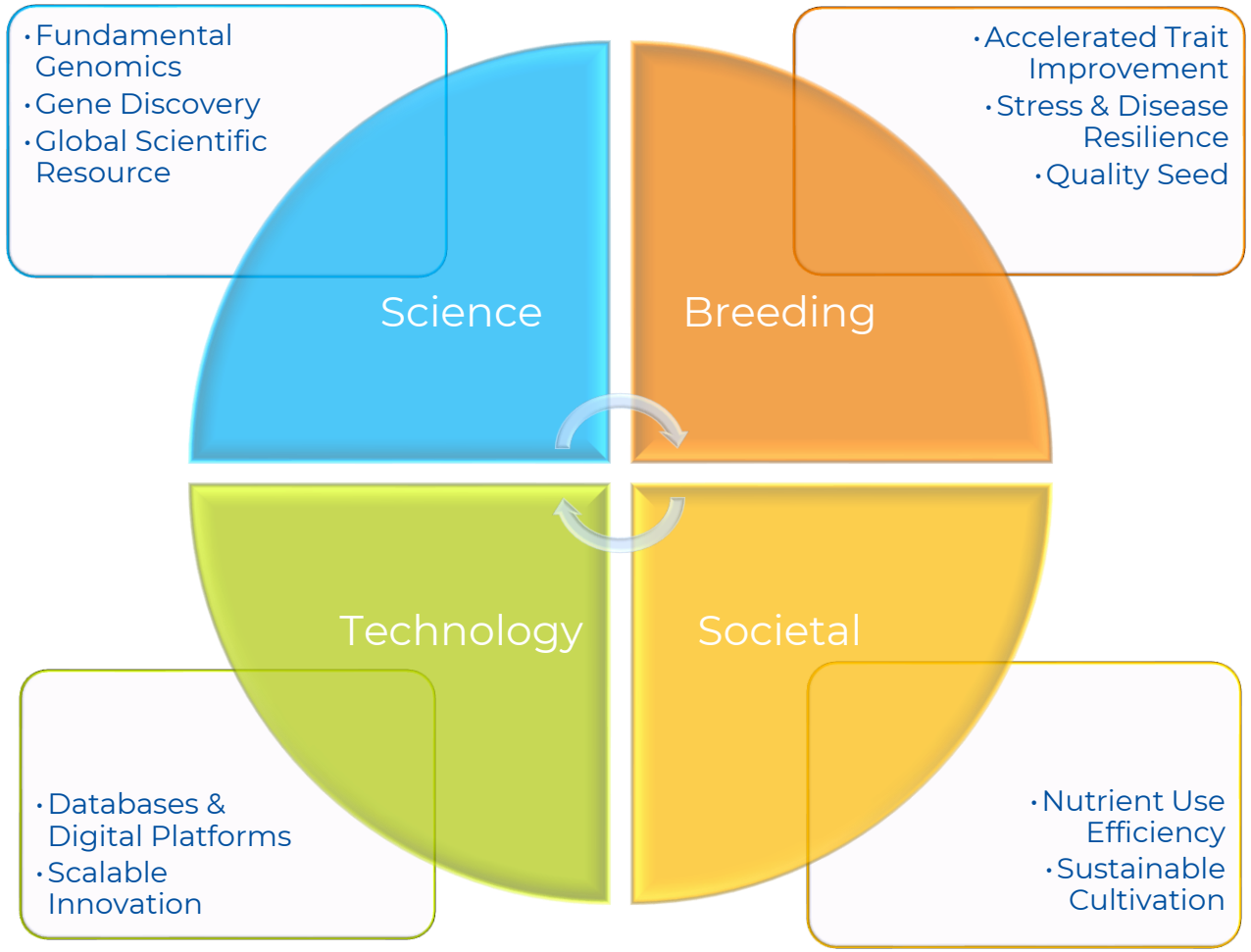
Production measured in tonnes, Source Food and Agriculture Organisation of United Stated 2023.

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

Why Now?

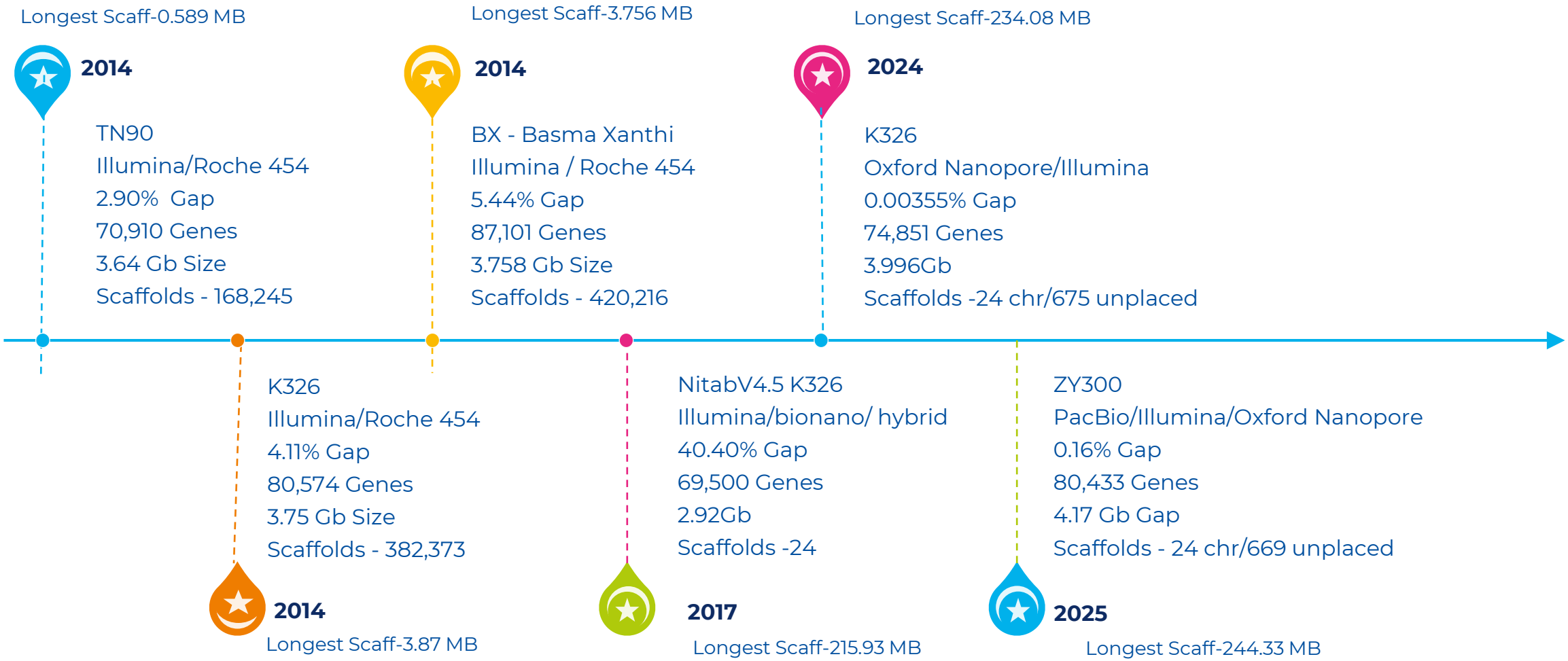


The Impact To Community



Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

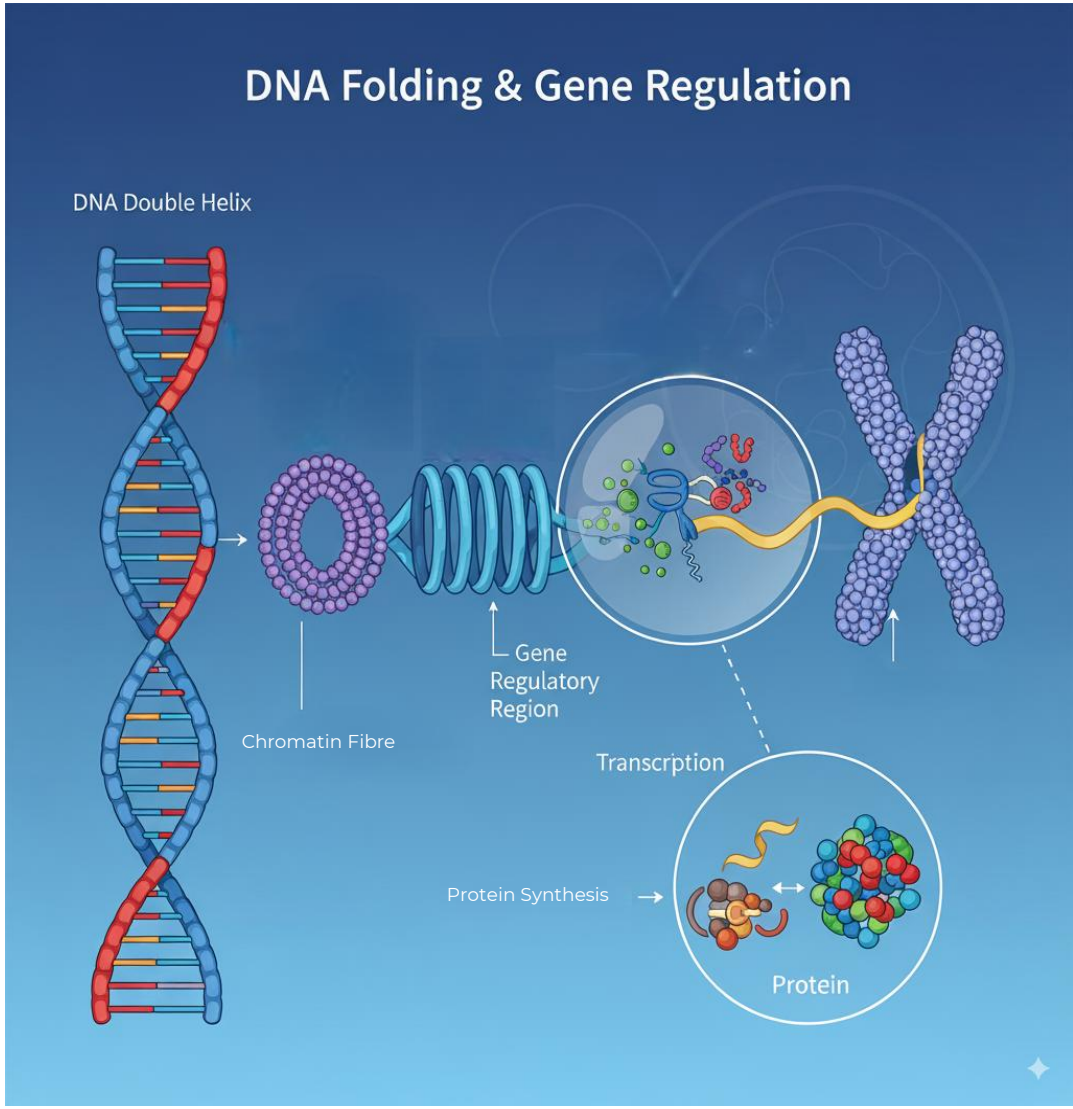
Current Statistics of Tobacco Genome



Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

What is missing?

- Despite decades of research, tobacco lacked a **highly contiguous** reference genome due to its polyploidy and repetitive content.
- Understanding the **3D genome** structure is crucial for deciphering gene regulation and advancing precision breeding in response to climate change.
- A **Haplotype resolved** assembly genome assembly separates maternal and paternal genomes, enabling precise study of traits, plant disease, and inheritance.

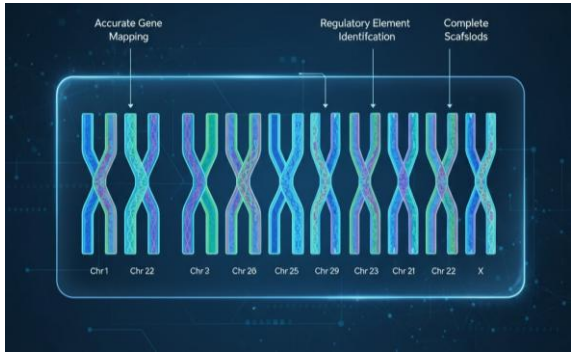


Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

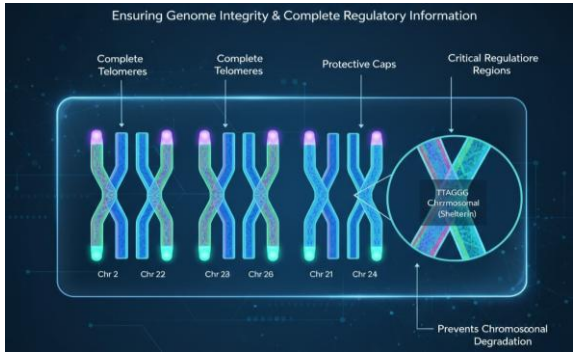
Our Aim

- Generate a near-complete, chromosome-scale genome assembly, resolve haplotypes, and explore 3D genome structure to understand gene regulation.

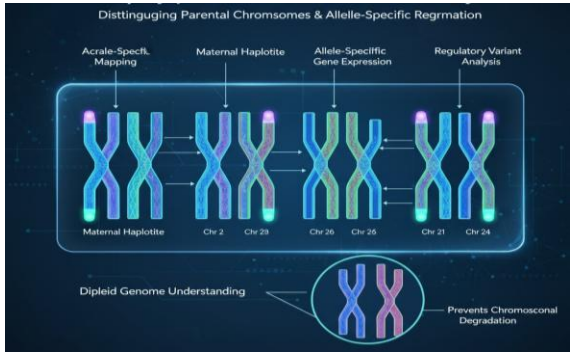
Chromosome Scale



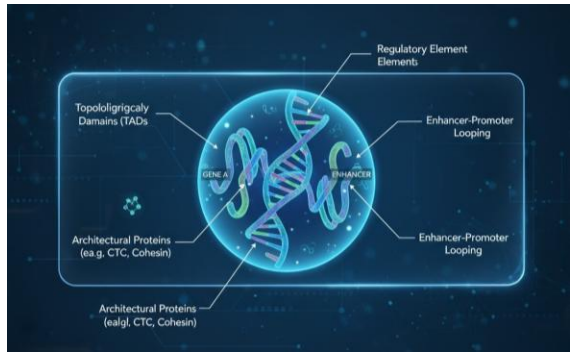
Telomere to Telomere



Haplotype resolved



3D genome



Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

The new *N. tabacum* NitabV6

Our Pipeline

HiFi Sequencing

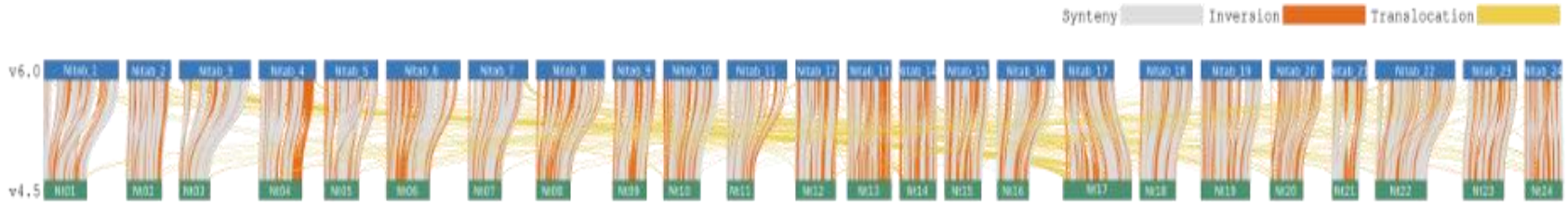
Omni-C, Deep Hi-C Scaffolding

Purge Duplication

Correction using Genetic map

Genome Statistics

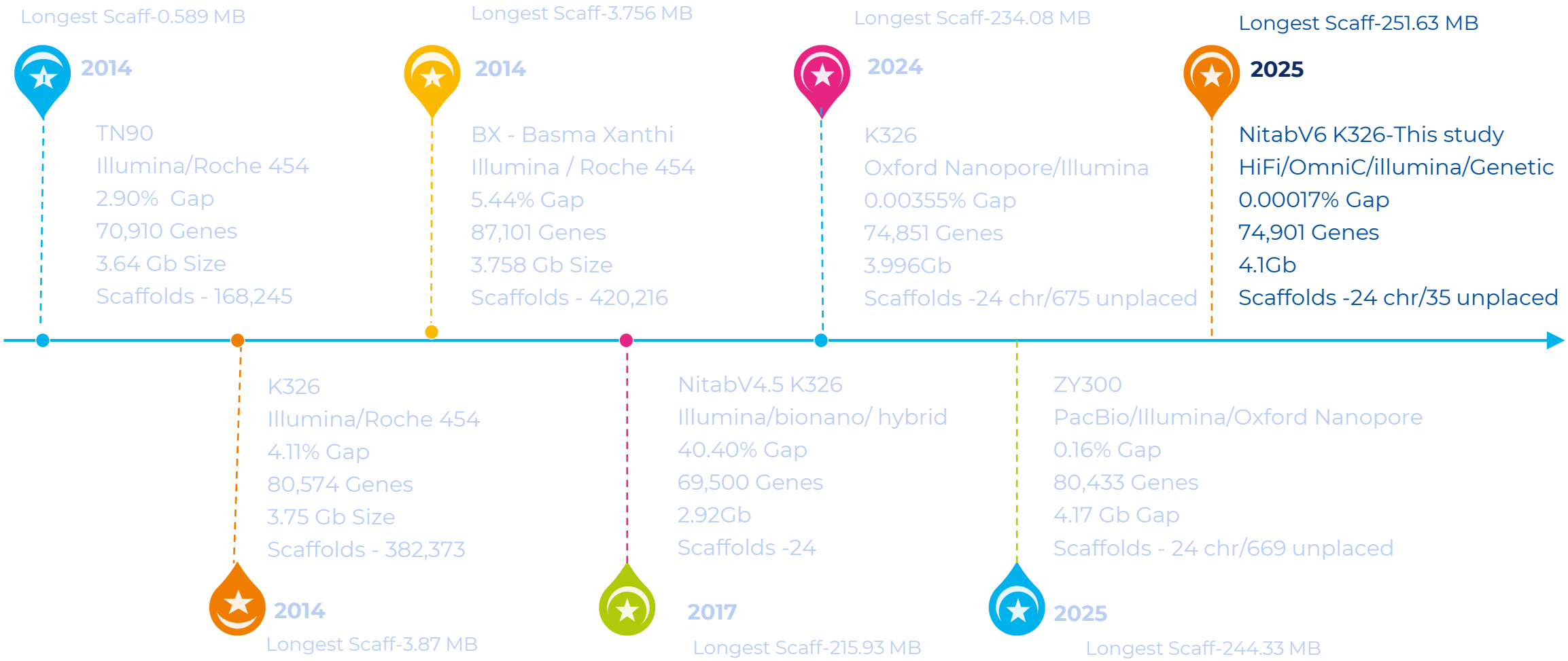
- Chromosome scale genome assembly of NitabV6 with 4.1Gb size.
- Nearly gap free (0.0017%).
- Telomere to Telomere completeness.
- Less fragmented (24 chromosome scale + 35 unplaced).
- Haplotype resolved.
- Longest scaffold 251.65 MB with all the scaffolds >50kb.
- First 3D genome explored.



Comparison of our previous V4.5 genome version (2017) to the current NitabV6 (2025)

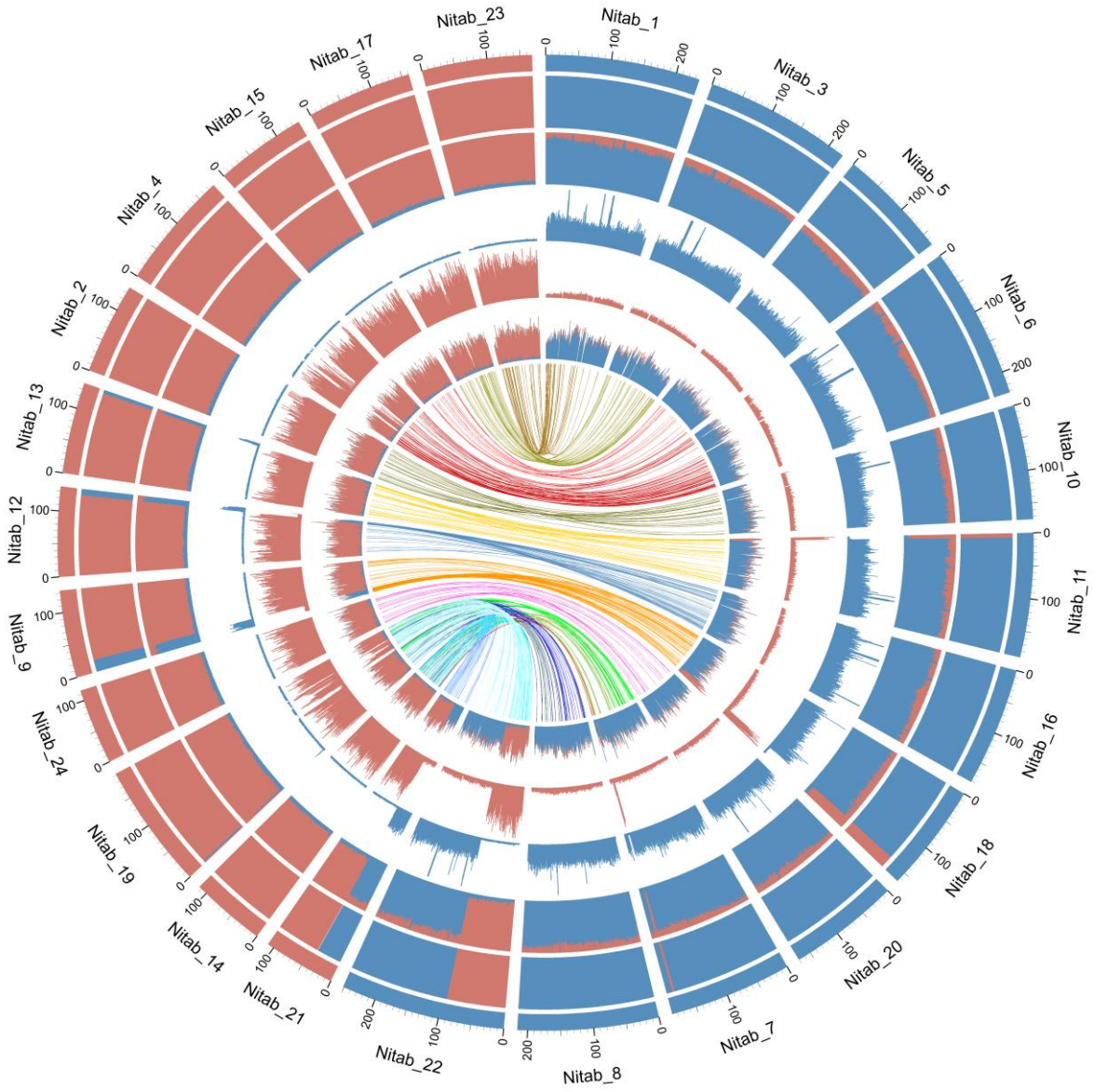
Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

Current Statistics of Tobacco Genome



Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

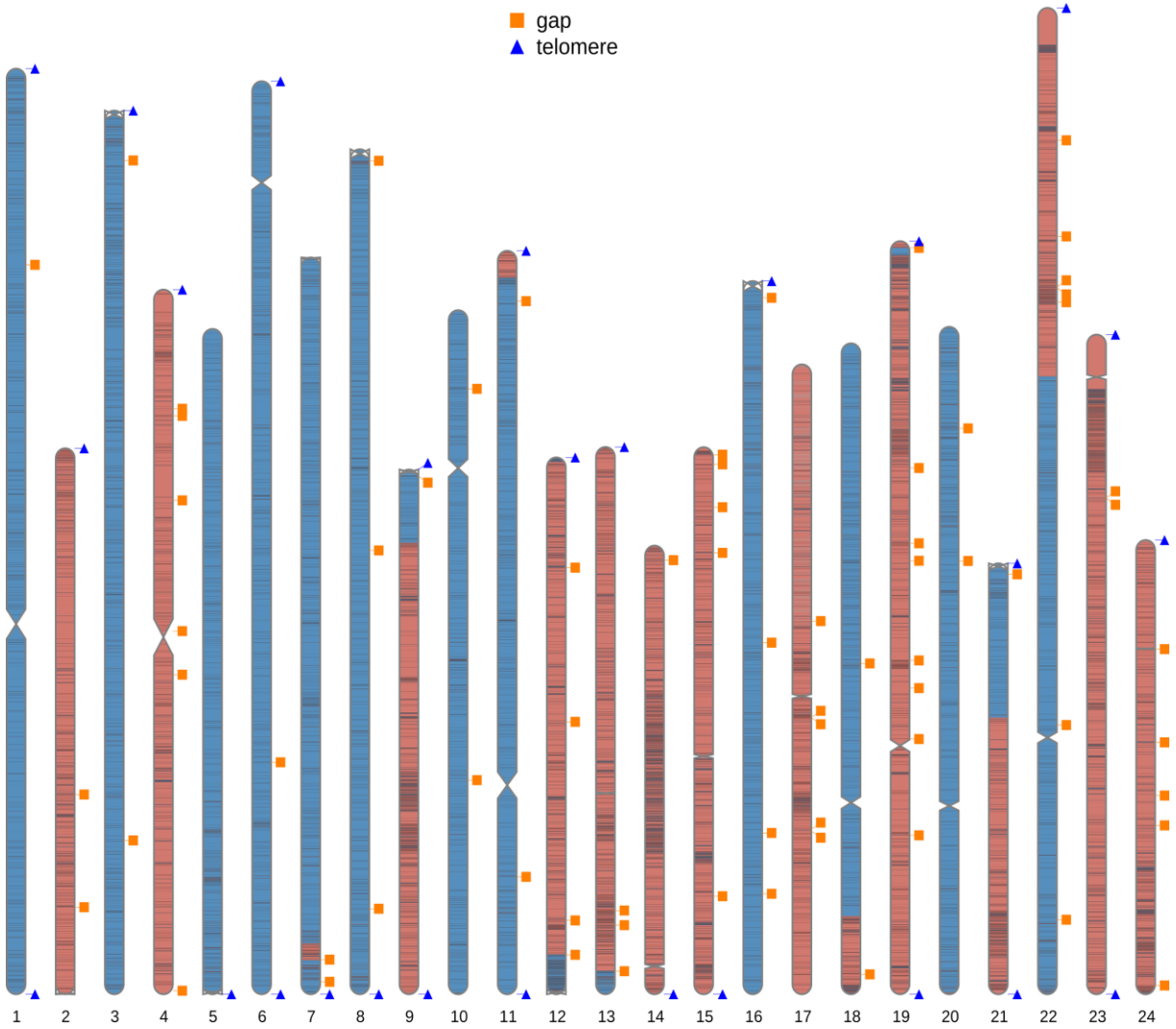
Deep Dive into NitabV6

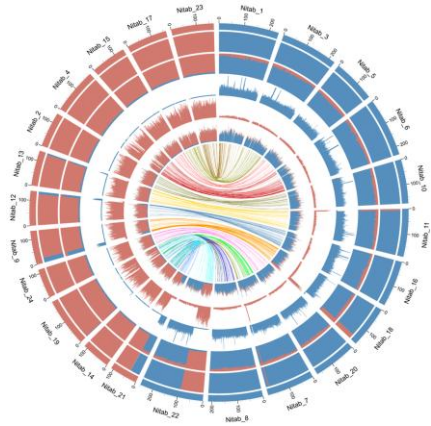


- Circus plot showing NitabV6 assignments with S-subgenome and T-subgenome.
- Approximately 57.4% and 40.9% of *N. tabacum*'s genome originated from *N. sylvestris* and *N. tomentosiformis* respectively.
- stretch of T-subgenome at the tail of Chr11's left arm and Chr07 and Chr18's right arm
- S-subgenome at the tail of Chr12's, Chr13's right arm and Chr19's left arm, which were previously unidentified

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

Deep Dive into NitabV6

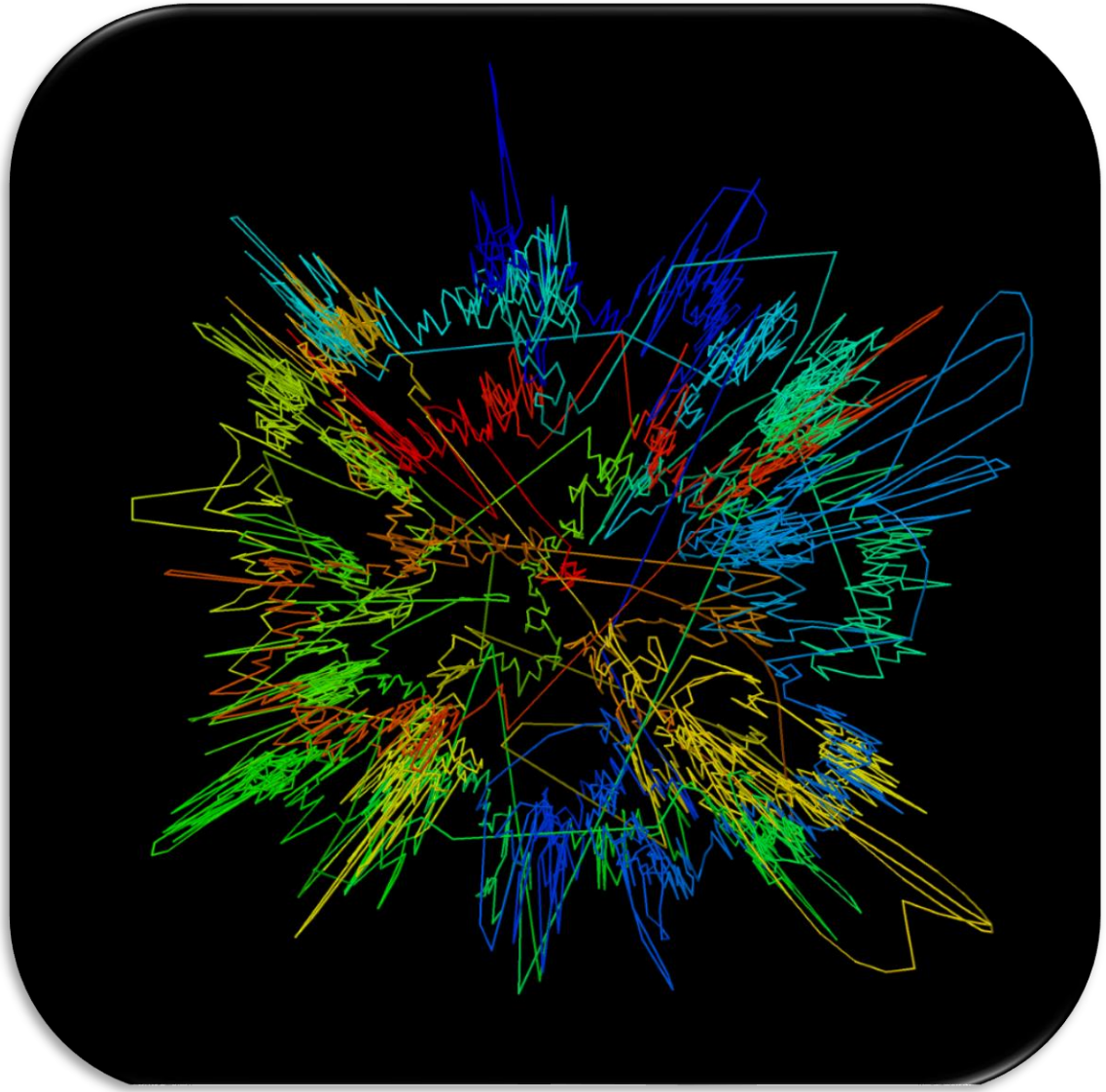




- 12chr with telomeres in one end.
- 4 chr with no telomeric regions.
- 15 candidate centromeres were identified.
- 5 intact chromosomes having telomere ends and centromere candidates.

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

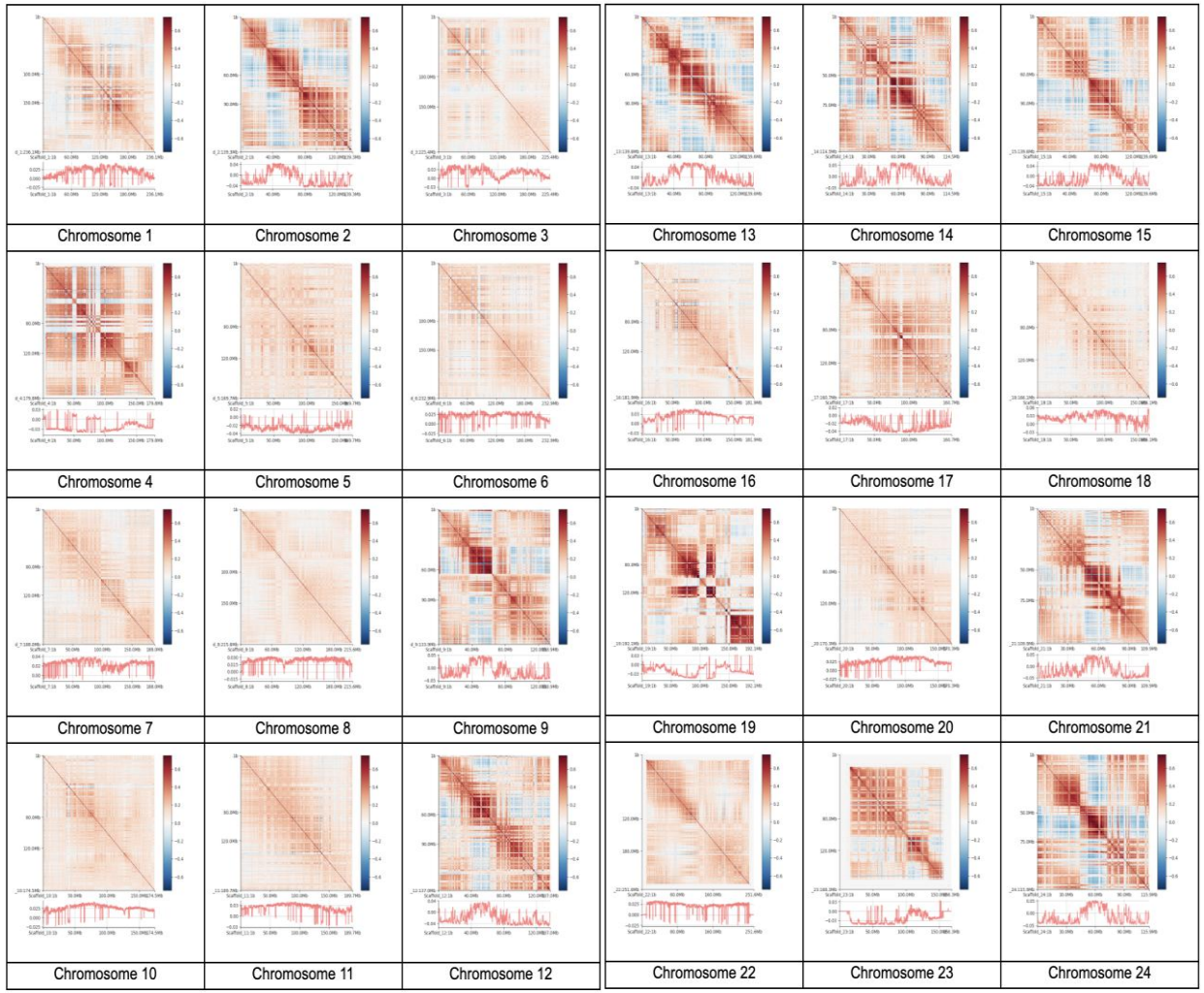
3D genome of NitabV6



- Deep sequencing (120x) carried out to explore the 3D architecture of NitabV6.
- Genome-wide and chromosome reconstruction.
- Representation of 3D organization of NitabV6.
- Each color represents the chromosomes in the genome.
- Reveals how DNA folding brings genes and regulatory elements into contact.

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

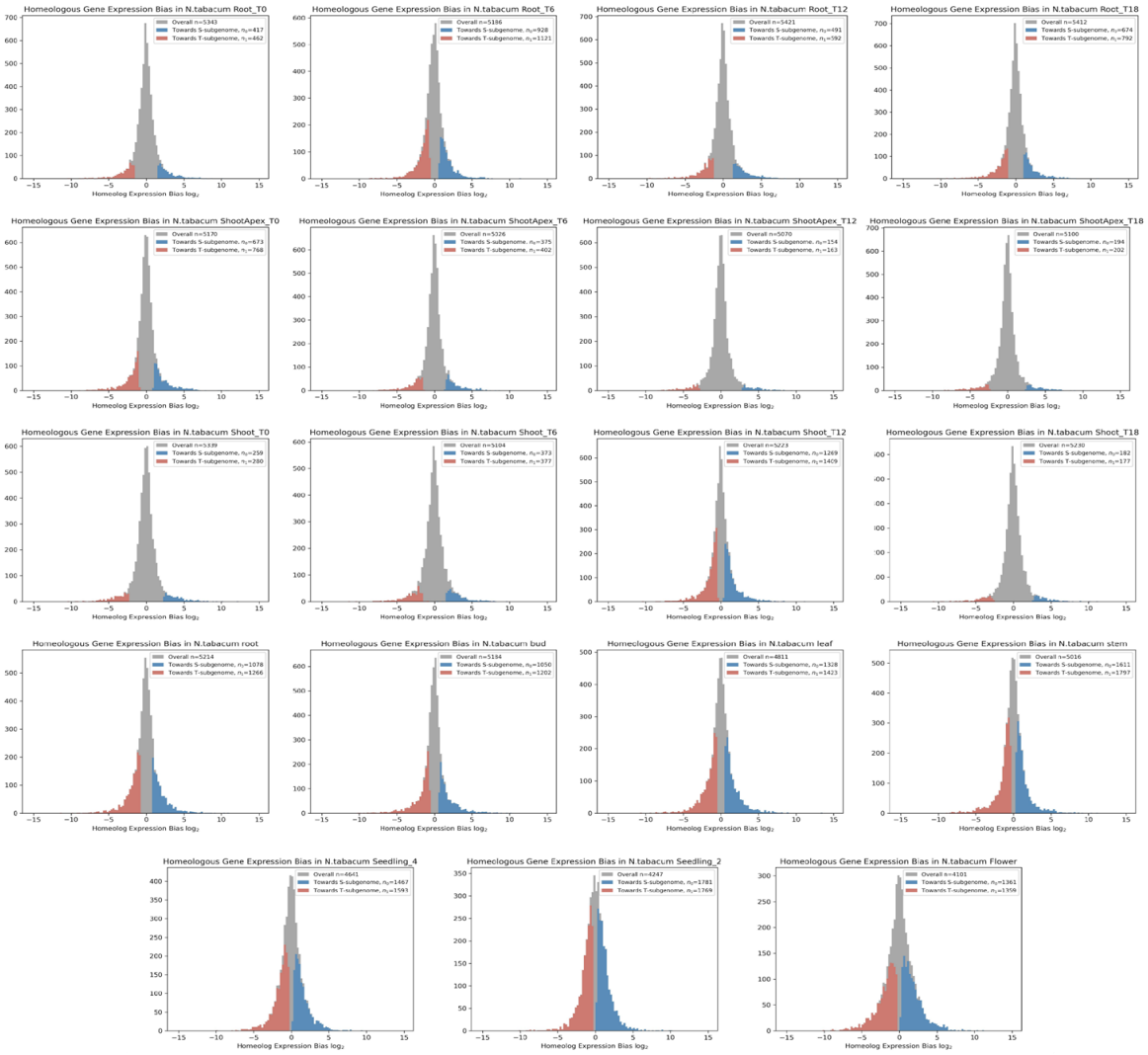
Deep Dive into NitabV6...continued



- Pearson's correlation matrix and the eigenvector's decomposition was calculated for the deep Hi-C data.
- A/B compartment studies were carried to show how the genome folds into active (A) and inactive (B) chromatin regions.
- Chromosomes 5, 6, 8, 10 and 11 were predominantly identified with active regions.
- Chromosomes 2, 4, 9, 12, 13, 14, 15, 19, 21, 23, and 24 displayed dominant regions with active (A) and inactive (B) compartments,

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

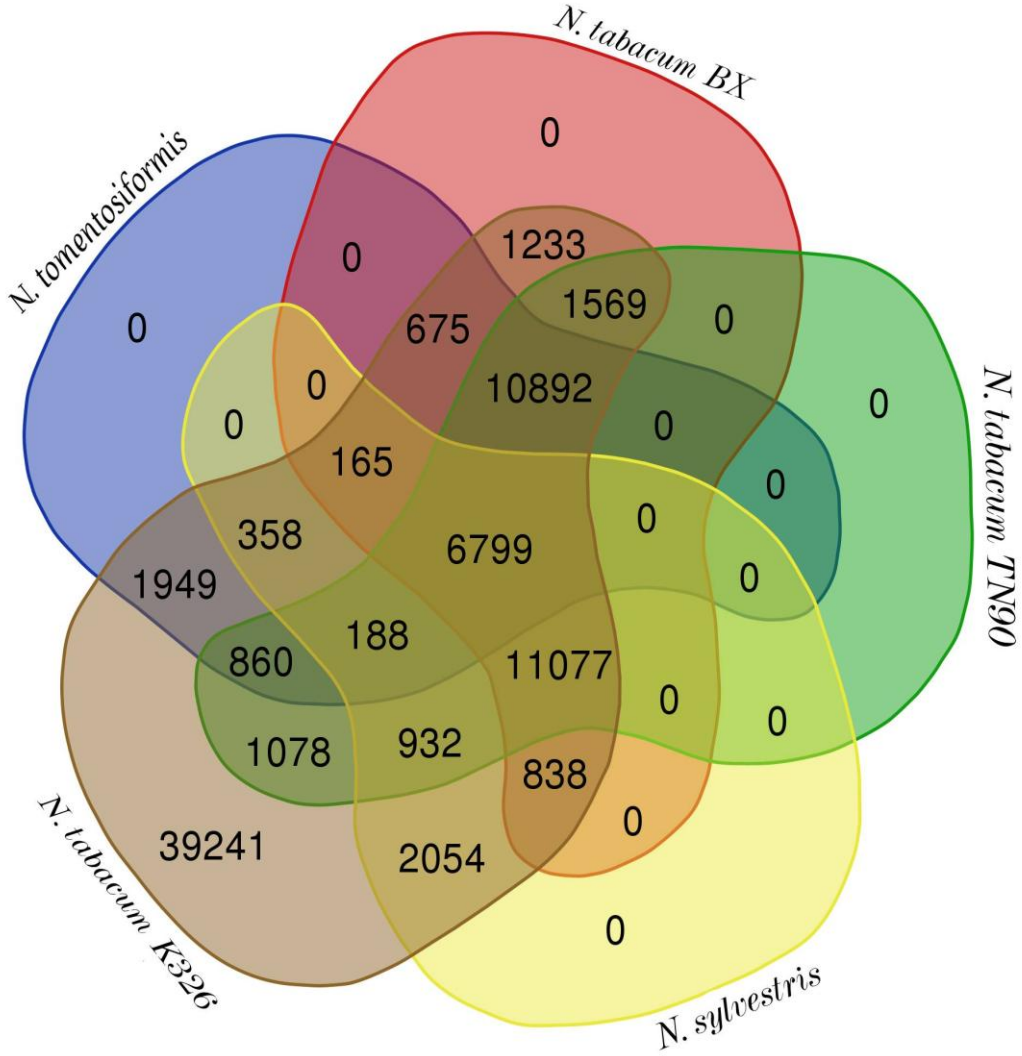
Deep Dive into NitabV6...continued



- 8,042 homoeologous gene pairs were identified and their expression profiles were assessed across 19 tissues.
- Tissues in different developmental stages, including shoot apex, shoot, root, bud, leaf, stem, seedling, and flower were utilized in the study.
- Expression bias between the S and T subgenomes was quantified using the homoeologous expression bias (HEB) framework (FDR 0.05).
- A strong T-subgenome bias was observed in 16 of 19 tissues, most notably in root, shoot apex, bud, leaf, stem.
- In contrast, three tissues flower, 2-week-old seedlings, and shoot a weaker S-subgenome bias was observed.
- S-subgenome dominates genome, T-subgenome dominates transcriptionally.
- Suggesting that **subgenome dominance** may **emerge from higher-order nuclear organization** rather than solely genetic or evolutionary factors.

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

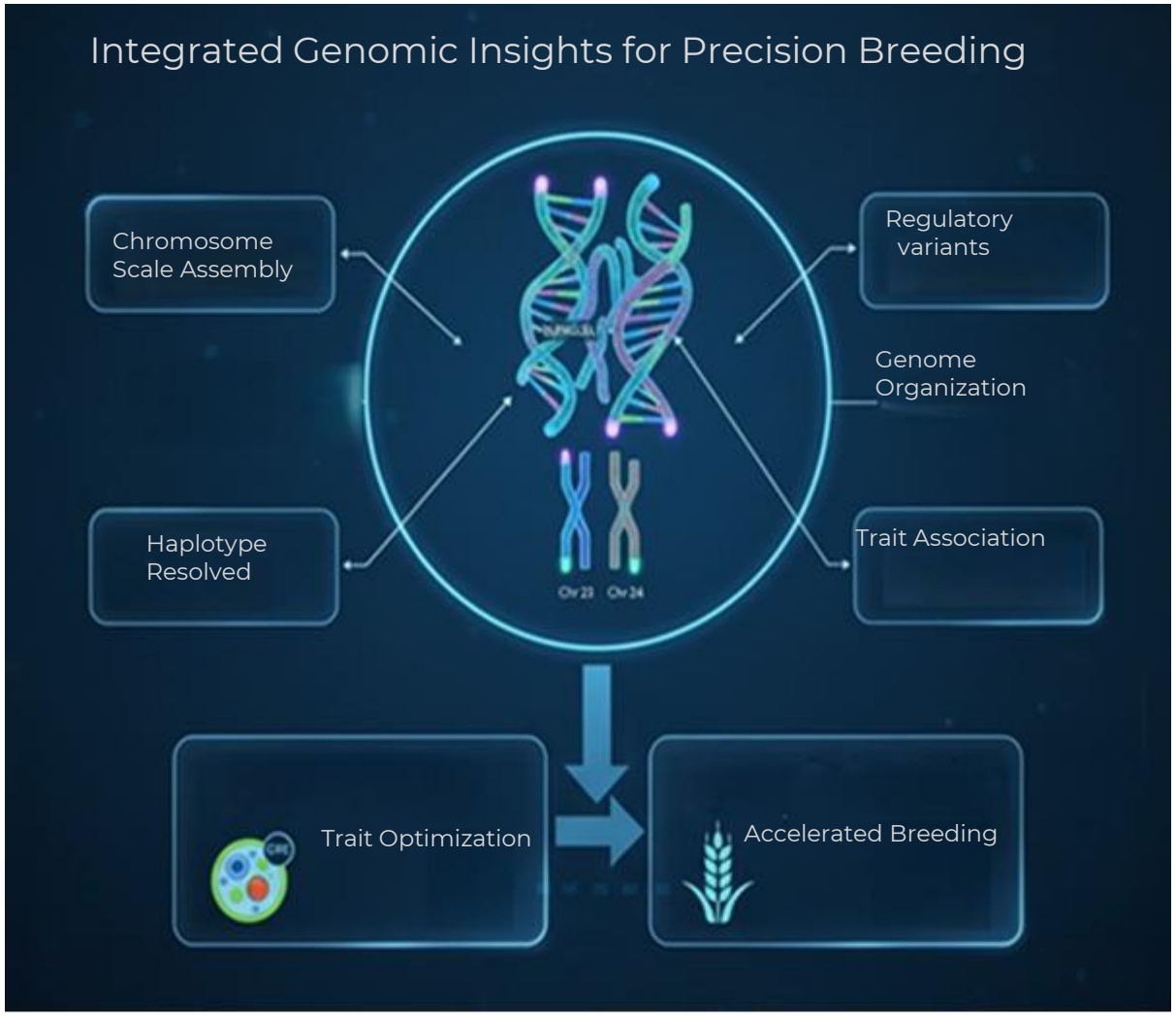
Deep Dive into NitabV6...continued



- NitabV6, *N. sylvestris, N. tomentiformis, N. tabacum* TN90 and *N. tabacum* Bx.
- The results highlighted that 93.2% of the genes were orthologous, while 6.8% remained unassigned to the any of the orthogroups.
- We observed 888 gene families to be expanded in *N. tabacum*.
- 18 gene families were found to be contracted.
- The genome shares 6,799 core genes with the other genomes (TN90, Bx, *N. tomentiformis and N. sylvestris*) with 39,241 genes unique to the plant.

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

Conclusion



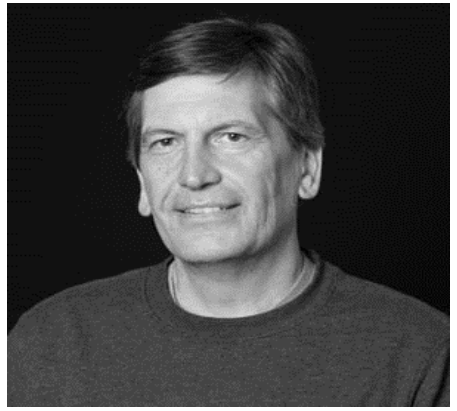
The chromosome-scale tobacco genome can accelerate plant breeding by powering marker-assisted pipelines and predictive trait selection, directly addressing industry demands for quality, climate resilience, and sustainability.

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

The Team

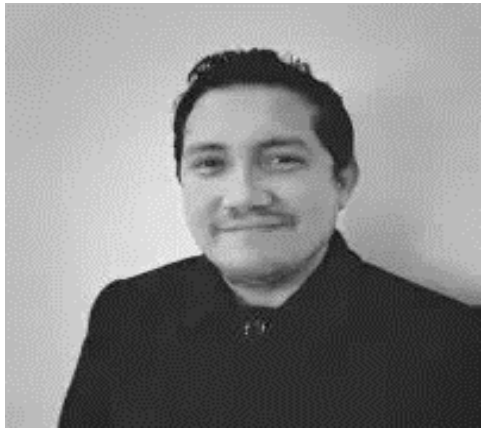
Dr. Parameswari Paul
Bioinformatics,
British American Tobacco (Investments) Ltd

Jillian DeWalt
Head of Biotech,
British American Tobacco (Investments) Ltd



Dr. Lukas A Mueller.
Adjunct Prof,
Boyce Thomson Institute

Mirella Flores Gonzalez
.Bioinformatics,
Boyce Thomson Institute



Dr. David Vega Herrera
Biochemist, British American Tobacco (Investments) Ltd

Dr. Alvaro Montiel Jorda.
Functional Genomics,
British American Tobacco (Investments) Ltd

Adrian Federico Powell,
Senior Research Associate,
Boyce Thomson Institute

Jing Zhang,
Bioinformatics,
Boyce Thomson Institute

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials

BAT
A BETTER TOMORROW

Exclusively for discussion purposes. Should not be copied or shared without prior consent. Not intended as promotional materials