

Supercalifragilistic Advances in Plant Genomics

Jeremy Schmutz

HudsonAlpha Institute for Biotechnology
Department of Energy Joint Genome Institute

BERAC Science Talk
October 25th, 2024



Huntsville,
Alabama,
USA



Berkeley,
California,
USA

“Any sufficiently advanced technology is indistinguishable from magic.”

Arthur C. Clarke 1973

In the beginning there was the human genome

- In 1986 Department of Energy announced “Human Genome Initiative” in order to understand how radiation exposure affects DNA and is passed on in the germline

Department of Energy
Office of Health and Environmental Research
SEQUENCING THE HUMAN GENOME
Summary Report of the Santa Fe Workshop
March 3-4, 1986



1986



1996

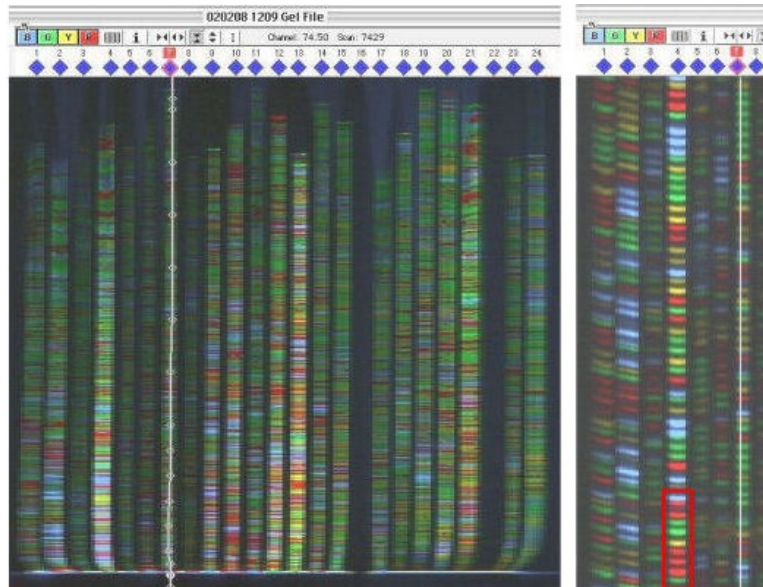
How did we sequence DNA?



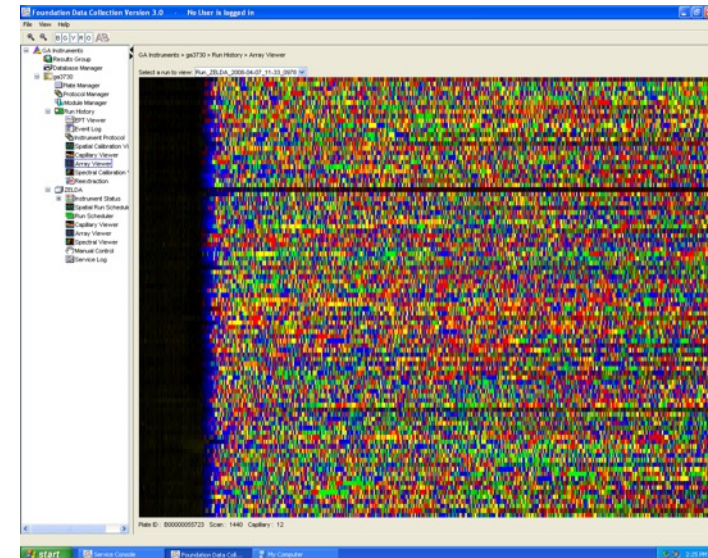
ABI 377 slab gel



ABI 3730 capillary

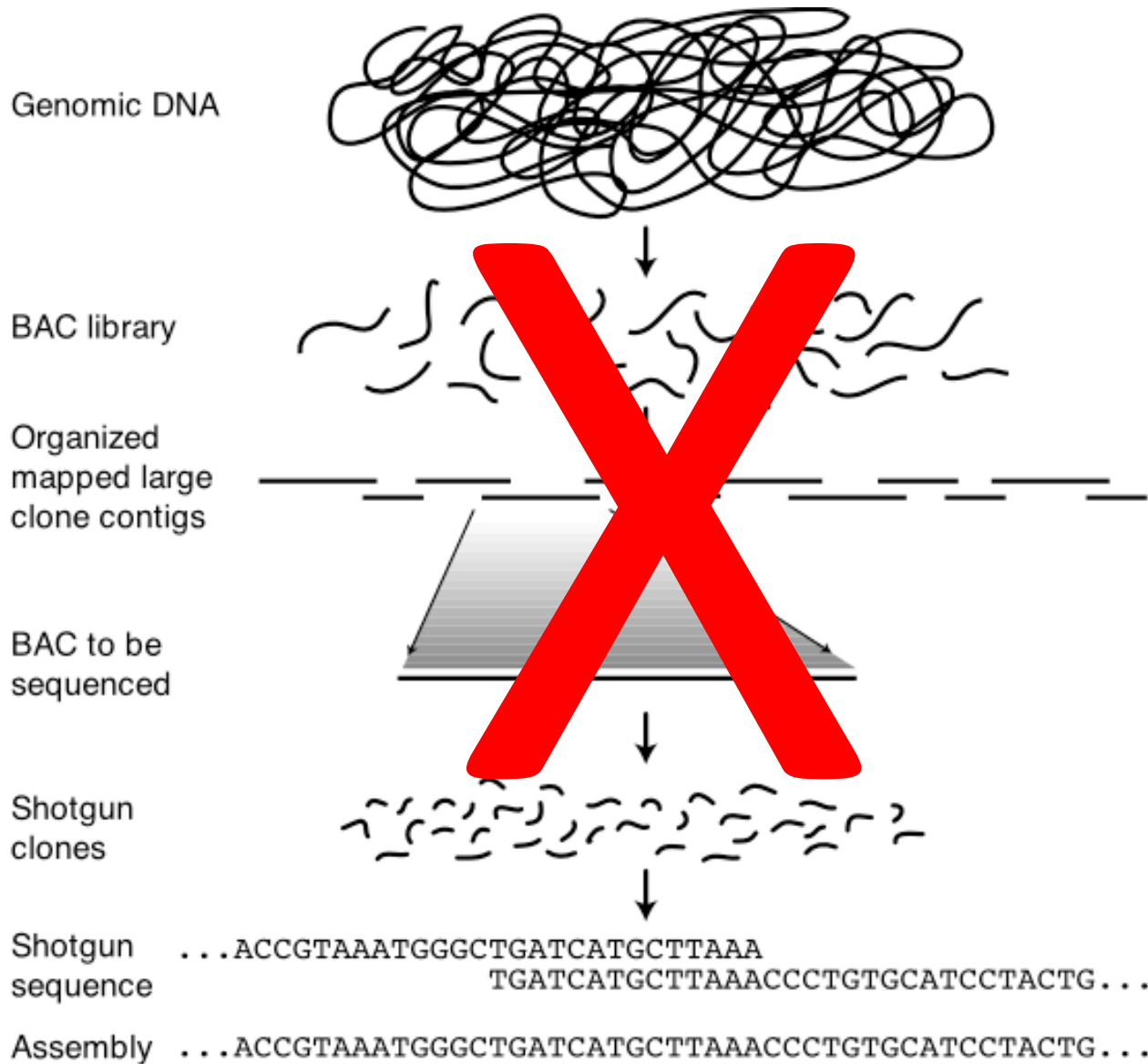


1996



2003

Sequencing a whole genome (BAC vs WGS)?



- WGS requires:
 - As long as reads as possible
 - Positional information from pairs
 - Computational power
- WGS advantages:
 - Many fewer libraries
 - Streamlined production
 - Consistent & measurable quality

Research news | Published: 26 January 2001

Weapons lab to develop Celera's new supercomputer

[Robert Walgate](#)

[Genome Biology](#) 2, Article number: spotlight-20010126-01 (2001) | [Cite this article](#)

733 Accesses | 1 Citations | 4 Altmetric | [Metrics](#)

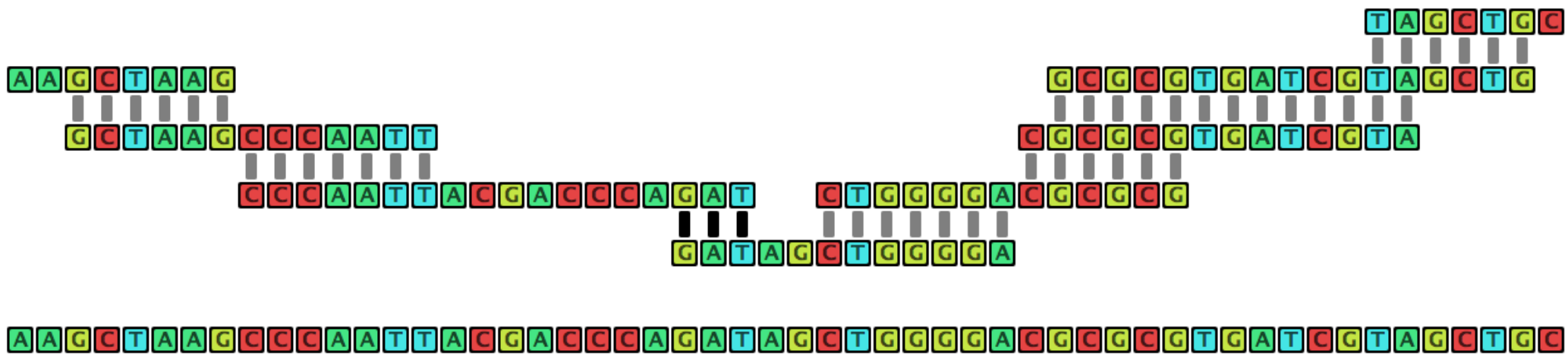
Genome Biology

LONDON Craig Venter, the CEO of [Celera Genomics](#) - which is on the verge of publishing the sequence of the human genome - has signed an agreement with [Sandia National Laboratory](#) in the US to develop the most powerful computer in the world within four years - and it'll be used for biology.

What is assembly?

Shotgun sequence . . .ACCGTAAATGGGCTGATCATGCTTAAA
TGATCATGCTTAAACCCTGTGCATCCTACTG . . .
Assembly . . .ACCGTAAATGGGCTGATCATGCTTAAACCCTGTGCATCCTACTG . . .

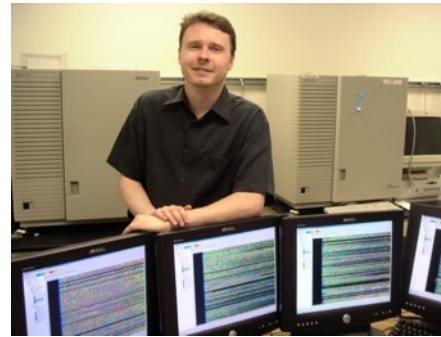
Assembly = Computational reconstruction (aka model) of a linear representation of DNA from repeated sampling of sequence (reads) of variably quality & length



Completing the human genome



Celebrating the human genome
Craig Venter, Ari Patrinos, Francis Collins



Marvin Stodolsky



Marvin Frazier



articles

Initial sequencing and analysis of the human genome

2001

International Human Genome Sequencing Consortium*

* A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

Finishing the euchromatic sequence of the human genome

2004

International Human Genome Sequencing Consortium*

* A list of authors and their affiliations appears in the Supplementary Information

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here we report the result of this finishing process. The current genome

What did JGI do with this sequencing capacity?

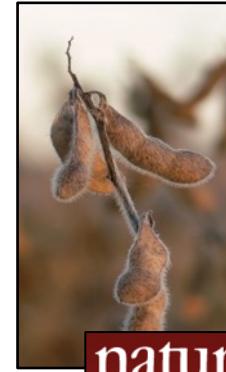
nature

Vol 452|6 March 2008|doi:10.1038/nature06556

LETTERS

The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis

F. Martin¹, A. Aerts², D. Ahren³, A. Brun¹, E. G. J. Danchin⁴, F. Duchaussoy¹, J. Gibon¹, A. Kohler¹, E. Lindquist², V. Pereda¹, A. Salamov⁵, H. J. Shapiro⁶, J. Wu^{1,5}, D. Blaudez⁷, M. Buée¹, P. Brokstein⁸, B. Canbäck⁹, D. Cohen¹, P. E. Courty¹, P. M. Coutinho¹, C. Delaruelle¹, J. C. Detter¹, A. Deveau¹, S. Difazio¹⁰, S. Duplessis¹, L. Fraissinet-Tachet¹, E. Lucic¹, P. Frey-Klett¹, C. Fourrey¹, I. Feussner¹, G. Gay¹, J. Grimwood¹, P. J. Hoegger¹⁰, P. Jain¹¹, S. Kilari¹⁰, J. Labbé¹, Y. C. Lin¹, V. Legué¹, F. Le Tacon¹, R. Marmeisse¹, D. Melayah¹, B. Montanini¹, M. Muratet¹¹, U. Nehls¹², H. Niculita-Hirzel¹³, M. P. Oudot-Le Secq¹, M. Peter^{1,14}, H. G. Posada¹⁵, B. Rajashekar¹, M. Reich^{1,10}, N. Roushier¹, J. Schmutz², T. Yin¹⁶, M. Chalot¹, B. Henrissat¹, U. Kües¹⁰, S. Lucas⁵, Y. Van de Peer⁵, G. K. Podila¹⁷, A. Polle¹⁰, P. J. Pukkila¹⁷, P. M. Richardson¹, P. Rouze^{3,18}, I. R. Sanders¹³, J. E. Stajich¹⁹, A. Tunlid¹, G. Tuskan¹⁶ & I. V. Grigoriev²



nature

Vol 463|14 January 2010|doi:10.1038/nature08760

ARTICLES

Genome sequence of the palaeopolyploid soybean

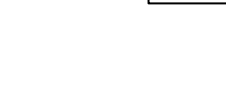
Jeremy Schmutz^{1,2}, Steven B. Cannon¹, Jessica Schluter^{4,5}, Jianxin Ma³, Therese Mitros⁶, William Neill⁷, David L. Hyten⁸, Qijian Song⁹, Jay J. Thelen¹⁰, Jianlin Cheng¹¹, Dong Xu¹¹, Uffe Hellsten¹², Gregory D. Yeiso¹³, Tetsuya Sakurai¹⁴, Taishi Umezawa¹⁵, Madan K. Bhattacharyya¹⁶, Devinder Sandhu¹⁷, Babu Valliyodan¹⁸, Erika Lindquist¹⁹, Myron Peto²⁰, David Grant²¹, Shengqiang Shu²², David Goodstein²³, Ki Montana-Futrell-Griggs²⁴, Brian Abernathy²⁵, Jianchang Du²⁶, Zhishi Tian²⁷, Liucun Zhu²⁸, Navdeep Gill²⁹, Marc Libault³⁰, Anand Sethuraman³¹, Xue-Cheng Zhang³², Kazuo Shinzaki³³, Henry T. Nguyen³⁴, Rod Perry Cregan³⁵, James Specht³⁶, Jane Grimwood³⁷, Dan Rokhsar³⁸, Gary Stacey^{39,40,41}, Randy C. Shoemaker⁴² & Scott A. Jackson⁴³

Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication

G. Albert Wu^{1,2,3,4}, Simon Prochnik^{1,2,3,4}, Jerry Jenkins^{1,2,3,4}, Jerome Salse^{1,2,3,4}, Uffe Hellsten^{1,2,3,4}, Florent Murat^{1,2,3,4}, Xavier Perrier^{1,2,3,4}, Manuel Ruiz^{1,2,3,4}, Simone Scalabrin^{1,2,3,4}, Javier Terol^{1,2,3,4}, Marco Aurelio Tuskis^{1,2,3,4}, Karine Labadie^{1,2,3,4}, Julie Pochain^{1,2,3,4}, Armand Coulouère^{1,2,3,4}, Kamel Jabbari^{1,2,3,4}, Federica Cattaneo^{1,2,3,4}, Cristian Del Fabro^{1,2,3,4}, Sara Pissotto^{1,2,3,4}, Andrea Zuccolò^{1,2,3,4}, Jerrod Chapman^{1,2,3,4}, Jane Grimwood^{1,2,3,4}, Francisco R. Tadeo^{1,2,3,4}, Leandro H. Estrella^{1,2,3,4}, Juan V. Muñoz-Sanz^{1,2,3,4}, Victoria Ibanez^{1,2,3,4}, Amparo Herrero-Ortega^{1,2,3,4}, Pablo Aleza^{1,2,3,4}, Julián Pérez-Pérez^{1,2,3,4}, Daniel Ramon^{1,2,3,4}, Dominique Brunet^{1,2,3,4}, François Laroche^{1,2,3,4}, Chuxian Chen^{1,2,3,4}, William G. Farmerie^{1,2,3,4}, Brian Deany^{1,2,3,4}, Chinnappa Kodira^{1,2,3,4}, Mohammed Mohiuddin^{1,2,3,4}, Tim Harkins^{1,2,3,4}, Karla Fredrickson^{1,2,3,4}, Paul Burns^{1,2,3,4}, Alexandre Lomsadze^{1,2,3,4}, Mark Borodovsky^{1,2,3,4}, Giuseppe Reforgiato^{1,2,3,4}, Juliana Freitas-Astua^{1,2,3,4}, Francis Quetier^{1,2,3,4}, Luis Navarro^{1,2,3,4}, Mikel Roost^{1,2,3,4}, Patrick Winker^{1,2,3,4}, Jeremy Schmutz^{1,2,3,4}, Michele Morgante^{1,2,3,4}, Marcos Antonio Machado^{1,2,3,4}, Manuel Talon^{1,2,3,4}, Olivier Jallou^{1,2,3,4}, Patrick Ollitrault^{1,2,3,4}, Frederic Gmitter^{1,2,3,4} & Daniel Rokhsar^{1,2,3,4}



nature biotechnology



ARTICLES

Genome sequencing and analysis of the model grass *Brachypodium distachyon*

The International Brachypodium Initiative*



nature

RESEARCH ARTICLES

The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray)

G. A. Tuskan^{1,2,3,4}, S. Difazio^{1,2,3,4}, S. Jansson^{1,2,3,4}, J. Bohlmann^{1,2,3,4}, I. Grigoriev^{1,2,3,4}, U. Hellsten^{1,2,3,4}, N. Putnam^{1,2,3,4}, S. Ralph^{1,2,3,4}, S. Rombauts^{1,2,3,4}, A. Salamov^{1,2,3,4}, J. Schein^{1,2,3,4}, T. Sterck^{1,2,3,4}, A. Aerts^{1,2,3,4}, R. R. Bhaleerao^{1,2,3,4}, R. P. Bhaleerao^{1,2,3,4}, D. Blaudez^{1,2,3,4}, W. Boerjan^{1,2,3,4}, A. Brun^{1,2,3,4}, A. Brunner^{1,2,3,4}, V. Busov^{1,2,3,4}, M. Campbell^{1,2,3,4}, J. Carlson^{1,2,3,4}, M. Chalot^{1,2,3,4}, J. Chapman^{1,2,3,4}, G.-L. Chen^{1,2,3,4}, D. Cooper^{1,2,3,4}, P. M. Coutinho^{1,2,3,4}, J. Couturier^{1,2,3,4}, S. Covert^{1,2,3,4}, Q. Cronk^{1,2,3,4}, R. Cunningham^{1,2,3,4}, J. Davis^{1,2,3,4}, S. Degroove^{1,2,3,4}, A. Déjardin^{1,2,3,4}, C. dePamphilis^{1,2,3,4}, J. Detter^{1,2,3,4}, B. Dirks^{1,2,3,4}, I. Dubchak^{1,2,3,4}, S. Duplessis^{1,2,3,4}, J. Ehting^{1,2,3,4}, B. Ellis^{1,2,3,4}, K. Gendler^{1,2,3,4}, D. Goodstein^{1,2,3,4}, M. Grubisov^{1,2,3,4}, J. Grimwood^{1,2,3,4}, A. Grover^{1,2,3,4}, L. Guntler^{1,2,3,4}, B. Hamberger^{1,2,3,4}, B. Heinze^{1,2,3,4}, V. Helariutta^{1,2,3,4}, B. Henrissat^{1,2,3,4}, D. Holligan^{1,2,3,4}, R. Holt^{1,2,3,4}, W. Huang^{1,2,3,4}, N. Islam-Faridi^{1,2,3,4}, S. Jones^{1,2,3,4}, M. Jones-Rhoades^{1,2,3,4}, R. Jorgensen^{1,2,3,4}, C. Joshi^{1,2,3,4}, J. Kangasjärvi^{1,2,3,4}, J. Karlsson^{1,2,3,4}, C. Kelleher^{1,2,3,4}, R. Kirkpatrick^{1,2,3,4}, M. Kirst^{1,2,3,4}, A. Kohler^{1,2,3,4}, U. Kallari^{1,2,3,4}, F. Larimer^{1,2,3,4}, J. Leebens-Mack^{1,2,3,4}, J.-C. Lepie^{1,2,3,4}, P. Locascio^{1,2,3,4}, Y. Lou^{1,2,3,4}, S. Lucas^{1,2,3,4}, F. Martin^{1,2,3,4}, B. Montanini^{1,2,3,4}, C. Napoli^{1,2,3,4}, D. R. Nelson^{1,2,3,4}, C. Nelson^{1,2,3,4}, K. Nieminen^{1,2,3,4}, O. Nilsson^{1,2,3,4}, V. Pereda^{1,2,3,4}, G. Peter^{1,2,3,4}, R. Philippe^{1,2,3,4}, G. Pilate^{1,2,3,4}, A. Pollack^{1,2,3,4}, J. Razumovskaya^{1,2,3,4}, P. Richardson^{1,2,3,4}, C. Rinaldi^{1,2,3,4}, K. Ritland^{1,2,3,4}, P. Rouze^{1,2,3,4}, D. Ryabov^{1,2,3,4}, J. Schmutz^{1,2,3,4}, J. Schradler^{1,2,3,4}, B. Segerman^{1,2,3,4}, H. Shin^{1,2,3,4}, A. Siddiqui^{1,2,3,4}, F. Sterky^{1,2,3,4}, A. Terry^{1,2,3,4}, C. Tsai^{1,2,3,4}, E. Uberbacher^{1,2,3,4}, P. Umberg^{1,2,3,4}, J. Valzela^{1,2,3,4}, K. Wall^{1,2,3,4}, S. Wessler^{1,2,3,4}, G. Yang^{1,2,3,4}, T. Yin^{1,2,3,4}, C. Douglas^{1,2,3,4}, M. Marra^{1,2,3,4}, G. Sandberg^{1,2,3,4}, Y. Van de Peer^{1,2,3,4}, D. Rokhsar^{1,2,3,4}



Science

LETTER

OPEN
doi:10.1038/nature11798

Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres

Andrew H. Paterson¹, Jonathan F. Wendel², Heidrun Gundlach³, Hui Guo⁴, Jerry Jenkins^{4,5}, Dianchuan Jin⁶, Danny Llewellyn⁷, Kurtis C. Showmaker⁸, Shengqiang Shu⁹, Joshua Udall¹⁰, Mi-jeong Yoo¹¹, Robert Byers¹², Wei Chen¹³, Adi Doron-Falgenboim¹⁴, Mary V. Duke¹⁵, Lei Gong¹⁶, Jane Grimwood¹⁷, Corrinne Grover¹⁸, Kara Grupp¹⁹, Guanqing Hua²⁰, Tae-ho Lee²¹, Jingping Li²², Lifeng Lin²³, Tao Liu²⁴, Barry S. Marler²⁵, Justin T. Page²⁶, Allison W. Roberts²⁷, Elisson Romanet²⁸, William S. Sanders²⁹, Emmanuel Szadkowsk³⁰, Xu Tan³¹, Haihao Tang³², Chunming Xu³³, Jimpeng Wang³⁴, Zining Wang³⁵, Dong Zhang³⁶, Lam Zhang³⁷, Hamid Ashraf³⁸, Frank Bedon³⁹, John E. Bowers⁴⁰, Curt L. Brubaker⁴¹, Peng W. Chee⁴², Sayan Das⁴³, Alan R. Gingle⁴⁴, Candace H. Haigler⁴⁵, David Harker⁴⁶, Lucia V. Hoffmann⁴⁷, Ran Hovav⁴⁸, Donald C. Jones⁴⁹, Cornelia Lemke⁵⁰, Shahid Mansoor⁵¹, Mehboob ur Rahman⁵², Lisa N. Rainville⁵³, Aditi Rambani⁵⁴, Umesh K. Reddy⁵⁵, Jun-kang Rong⁵⁶, Yehoshua Saranga⁵⁷, Brian E. Scheffler⁵⁸, Jodi A. Scheffler⁵⁹, David M. Stelly⁶⁰, Barbara A. Triplett⁶¹, Allen Van Deynze⁶², Maiti F. S. Vaslin⁶³, Vijay N. Waghmare⁶⁴, Sally A. Walford⁶⁵, Robert J. Wright⁶⁶, Essam A. Zaki⁶⁷, Tianzhen Zhang⁶⁸, Elizabeth S. Dennis⁶⁹, Klaus F. X. Mayer⁷⁰, Daniel G. Peterson⁷¹, Daniel S. Rokhsar⁷², Xiying Wang⁷³ & Jeremy Schmutz⁷⁴

nature



nature

Vol 456|13 November 2008|doi:10.1038/nature07410

nature

LETTERS

The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes

Chris Bowler^{1,2}, Andrew E. Allen³, Jonathan H. Badger⁴, Jane Grimwood⁵, Kamel Jabbari⁶, Alan Kuo⁷, Uma Maheswari⁸, Cindy Martinez⁹, Florian Maumus¹⁰, Robert P. Olliar¹¹, Edda Rayko¹², Asaf Salamov¹³, Klaus Vandepoel¹⁴, Bank Beszter¹⁵, Ansgar Gruber¹⁶, Marc Heide¹⁷, Michael Katinka¹⁸, Thomas Mock^{19,20}, Klaus Valentin²¹, Frédéric Verret²², John A. Berges²³, Colin Brownlee²⁴, Jean-Paul Cadoret²⁵, Anthony Chiovitti²⁶, Chang Jae Choi²⁷, Sacha Coesel²⁸, Alessandra De Martino²⁹, J. Chris Detter³⁰, Colleen Durkin³¹, Angela Falciatore³², Jérôme Fournet³³, Miyoshi Haruta³⁴, Marie J. Huysman³⁵, Bethany D. Jenkins³⁶, Katerina Jiroukova³⁷, Richard E. Jorgensen³⁸, Yolaïne Joubert³⁹, Aaron Kaplan⁴⁰, Nils Kröger⁴¹, Peter G. Kroth⁴², Julie La Roche⁴³, Erica Lindquist⁴⁴, Markus Lommer⁴⁵, Véronique Martin-Jézéquel⁴⁶, Pascal J. Lopez⁴⁷, Susan Lucas⁴⁸, Manuela Mangogna⁴⁹, Karen McGinnis⁵⁰, Linda K. Medlin⁵¹, Anton Montsant⁵², Marie-Pierre Oudot-Le Secq⁵³, Carolyn Napoli⁵⁴, Miroslav Obornik⁵⁵, Micaela Schnitzler Parker⁵⁶, Jean-Louis Petit⁵⁷, Betina M. Porcel⁵⁸, Nicole Poulsen⁵⁹, Matthew Robison⁶⁰, Leszek Rychlewski⁶¹, Tatiana A. Rynearson⁶², Jeremy Schmutz⁶³, Harris Shapiro⁶⁴, Magali Sinaut⁶⁵, Michele Stanley⁶⁶, Michael R. Sussman⁶⁷, Alison R. Taylor⁶⁸, Assaf Vardi⁶⁹, Peter von Dassow⁷⁰, Wim Vyverman⁷¹, Anusuya Willis⁷², Lucjan S. Wyrzyk⁷³, Daniel S. Rokhsar⁷⁴, Jean Weissenbach⁷⁵, E. Virginia Armbrust⁷⁶, Beverley R. Green⁷⁷, Yves Van de Peer⁷⁸ & Igor V. Grigoriev⁷⁹

ARTICLE

The genome of *Eucalyptus grandis*

Alexander A. Myburg^{1,2}, Dario Grattapaglia^{3,4}, Gerald A. Tuskan^{5,6}, Uffe Hellsten⁷, Richard D. Hayes⁸, Jane Grimwood⁹, Jerry Jenkins¹⁰, Erika Lindquist¹¹, Hope Tice¹², Diane Bauer¹³, David M. Goodstein¹⁴, Inna Dubchak¹⁵, Alexandre Poliakov¹⁶, Eshchar Mizrahi¹⁷, Anand R. K. Kullam¹⁸, Desre Pinard¹⁹, Karen van der Merwe²⁰, Pooga Singh²¹, Ida van Jaarsveld²², Orzenil B. Silva-Junior²³, Roberto C. Togawa²⁴, Marilia R. Pappas²⁵, Danielle A. Faria²⁶, Carolina P. Sansalone²⁷, Cesar D. Petrolí²⁸, Xiaohan Yang²⁹, Priya Ranjan³⁰, Timothy J. Fischplinski³¹, Chu-Yu Yu³², Ting-Li Li³³, Lieven Sterck³⁴, Kevin Vanmeir³⁵, Florent Murat³⁶, Marcel Soler³⁷, Hélène San Clemente³⁸, Najib Salim³⁹, Christophe Dumand⁴⁰, Charles A. Hefer⁴¹, Erich Bornberg-Bauer⁴², Anna R. Kersting⁴³, Kelly Vining⁴⁴, Vindhya Amarasinghe⁴⁵, Martin Rankin⁴⁶, Sushma Naitani⁴⁷, Justin Elser⁴⁸, Alexander E. Boyd⁴⁹, Aaron Liston⁵⁰, Joseph W. Spatafora⁵¹, Palitha Dharmawardhana⁵², Rajani Raja⁵³, Christopher Sullivan⁵⁴, Elisson Romanet^{55,56,57}, Marcio Alves-Ferreira⁵⁸, Carsten Külheim⁵⁹, William Foley⁶⁰, Victor Canocha^{61,62}, Jorge Paiva⁶³, David Kudrna⁶⁴, Sergio H. Brommonschenkel⁶⁵, Giancarlo Pasquati⁶⁶, Margaret Byrne⁶⁷, Philippe Rigault⁶⁸, Josquin Tibbits⁶⁹, Antanas Spokocivicius⁷⁰, Rebecca C. Jones⁷¹, Dorothy A. Steamer⁷², René E. Vaillancourt⁷³, Brad M. Potts⁷⁴, Fourie Joubert⁷⁵, Kerrie Barry⁷⁶, Georgios J. Pappas⁷⁷, Steven H. Strauss⁷⁸, Pankaj Jaswal⁷⁹, Jacqueline Grima-Pettenati⁸⁰, Jérôme Salse⁸¹, Yves Van de Peer⁸², Daniel S. Rokhsar⁸³ & Jeremy Schmutz⁸⁴



nature

Vol 457|29 January 2010|doi:10.1038/nature07723

nature

ARTICLES

The *Sorghum bicolor* genome and the diversification of grasses

Andrew H. Paterson¹, John E. Bowers², Rémy Bruggmann³, Inna Dubchak⁴, Jane Grimwood⁵, Heidrun Gundlach⁶, Georg Haberer⁷, Uffe Hellsten⁸, Therese Mitros⁹, Alexandre Poliakov¹⁰, Jeremy Schmutz¹¹, Manuel Spannagl¹², Haihao Tang¹³, Xiying Wang¹⁴, Thomas Wicker¹⁵, Arvind K. Bhardi¹⁶, Jarrod Chapman¹⁷, F. Alex Fellus¹⁸, Udo Gowik¹⁹, Igor V. Grigoriev²⁰, Eric Lyons²¹, Christopher A. Maher²², Mibaela Martins²³, Apurva Narechania²⁴, Robert P. Olliar²⁵, Bryan W. Penning²⁶, Asaf A. Salamov²⁷, Yu Wang²⁸, Lifang Zhang²⁹, Nicholas C. Carpita³⁰, Michael Freeling³¹, Alan R. Gingle³², C. Thomas Hash³³, Beat Keller³⁴, Patricia Klein³⁵, Stephen Kresovich³⁶, Maureen C. McCann³⁷, Ray Ming³⁸, Daniel G. Peterson³⁹, Mehboob-ur-Rahman⁴⁰, Doreen Ware^{41,42}, Peter Westhoff⁴³, Klaus F. X. Mayer⁴⁴, Joachim Messing⁴⁵ & Daniel S. Rokhsar⁴⁶

Post-Sanger genome sequencing



Solexa GA 2006



Illumina 2500 2012



PacBio RS 2011



PacBio RSII 2013



PacBio Sequel 2016

Short read platforms

- + super inexpensive data
- + low error rate (99.5% accuracy)
- + metric tons of it
- 76 to 150 bp reads
- Short fragments (400bp)

Long read platforms

- expensive data (its all relative)
- high error rate (82% accuracy)
- low production
- + 2,000 - 50,000 bp reads
- + Greatly reduce assembly complexity

New science with long reads



The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*

David J. Bertoli^{1,2,3,30*}, Jerry Jenkins^{4,30}, Josh Clevenger^{1,2,3,30}, Olga Dudchenko⁵, Dongying Gao¹, Guillermo Seijo^{6,7}, Soraya C. M. Leal-Bertoli^{1,2,8}, Longhui Ren⁹, Andrew D. Farmer¹⁰, Manish K. Pandey¹¹, Sergio S. Samoluk^{6,7}, Brian Abernathy¹, Gaurav Agarwal⁸, Carolina Ballén-Taborda², Connor Cameron¹⁰, Jacqueline Campbell¹², Carolina Chavarro^{1,2}, Annapurna Chitkineni¹¹, Ye Chu¹³, Sudhansu Dash¹⁰, Moaine El Baidour^{14,15}, Baozhu Guo¹⁶, Wei Huang¹², Kyung Do Kim^{1,17}, Walid Korani¹, Sophie Lanciano^{15,18,19}, Christopher G. Lui⁵, Marie Mirouze^{15,18,19}, Márcio C. Moretzsohn²⁰, Melanie Pham⁵, Jin Hee Shin^{1,17}, Kenta Shirasawa²¹, Senjuti Sinharoy²², Avinash Sreedasyam⁴, Nathan T. Weeks²³, Xinyou Zhang^{24,25}, Zheng Zheng^{24,25}, Ziqi Sun^{24,25}, Lutz Froenicke²⁶, Erez L. Aiden⁵, Richard Michelmore²⁶, Rajeev K. Varshney¹¹, C. Corley Holbrook²⁷, Ethalinda K. S. Cannon¹², Brian E. Scheffler²⁸, Jane Grimwood⁴, Peggy Ozias-Akins^{2,13}, Steven B. Cannon^{23,31}, Scott A. Jackson^{1,2,3,31*} and Jeremy Schmutz^{4,29,31*}



Inbred Tetraploid (2x)



Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement

Z. Jeffrey Chen^{1,2,14,32}, Avinash Sreedasyam^{3,14}, Atsumi Ando^{1,14}, Qingxin Song^{1,2,14}, Luis M. De Santiago^{4,14}, Amanda M. Hulse-Kemp⁵, Mingquan Ding^{1,4}, Wenxue Ye², Ryan C. Kirkbride¹, Jerry Jenkins³, Christopher Plott³, John Lovell³, Yu-Ming Lin⁴, Robert Vaughn⁴, Bo Liu⁴, Sheron Simpson⁷, Brian E. Scheffler⁷, Li Wen⁸, Christopher A. Sasaki⁸, Corrinne E. Grover⁹, Guanqing Hu⁹, Justin L. Conover⁹, Joseph W. Carlson¹⁰, Shengqiang Shu¹⁰, Lori B. Boston³, Melissa Williams³, Daniel G. Peterson¹¹, Keith McGee¹², Don C. Jones¹³, Jonathan F. Wendel⁹, David M. Stelly⁴, Jane Grimwood^{3,33} and Jeremy Schmutz^{3,30}



Inbred Tetraploid (2x)



Outbred Diploid (2x)
Partial resolved 2nd



ARTICLE
<https://doi.org/10.1038/s41467-021-24328-w> OPEN

Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding

John T. Lovell^{1,15,34}, Nolan B. Bentley^{2,15}, Gaurab Bhattarai^{3,15}, Jerry W. Jenkins^{1,15}, Avinash Sreedasyam^{1,15}, Yanina Alarcon⁴, Clive Bock⁵, Lori Beth Boston¹, Joseph Carlson⁶, Kimberly Cervantes⁷, Kristen Clermont⁸, Sara Duke⁹, Nick Krom⁴, Keith Kubenka¹⁰, Sujan Mamidi¹, Christopher P. Mattison⁸, Maria J. Monteros⁴, Cristina Pisani⁵, Christopher Plott¹, Shanmugam Rajasekar¹¹, Horat Shadgou Rhein⁷, Charles Rohla⁴, Mingzhou Song¹², Rolston St. Hilaire¹³, Shengqiang Shu¹⁰, Lenny Wells¹⁴, Jenell Webber¹, Richard J. Heerema¹², Patricia E. Klein², Patrick Conner¹⁴, Xinwang Wang¹⁰, L. J. Grauke¹⁰, Jane Grimwood¹, Jeremy Schmutz^{1,6,35} & Jennifer J. Randall^{7,35}

Article

Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass

<https://doi.org/10.1038/s41588-020-03127-1>

Received: 1 July 2020

Accepted: 16 December 2020

Published online: 27 January 2021

Open access

Check for updates

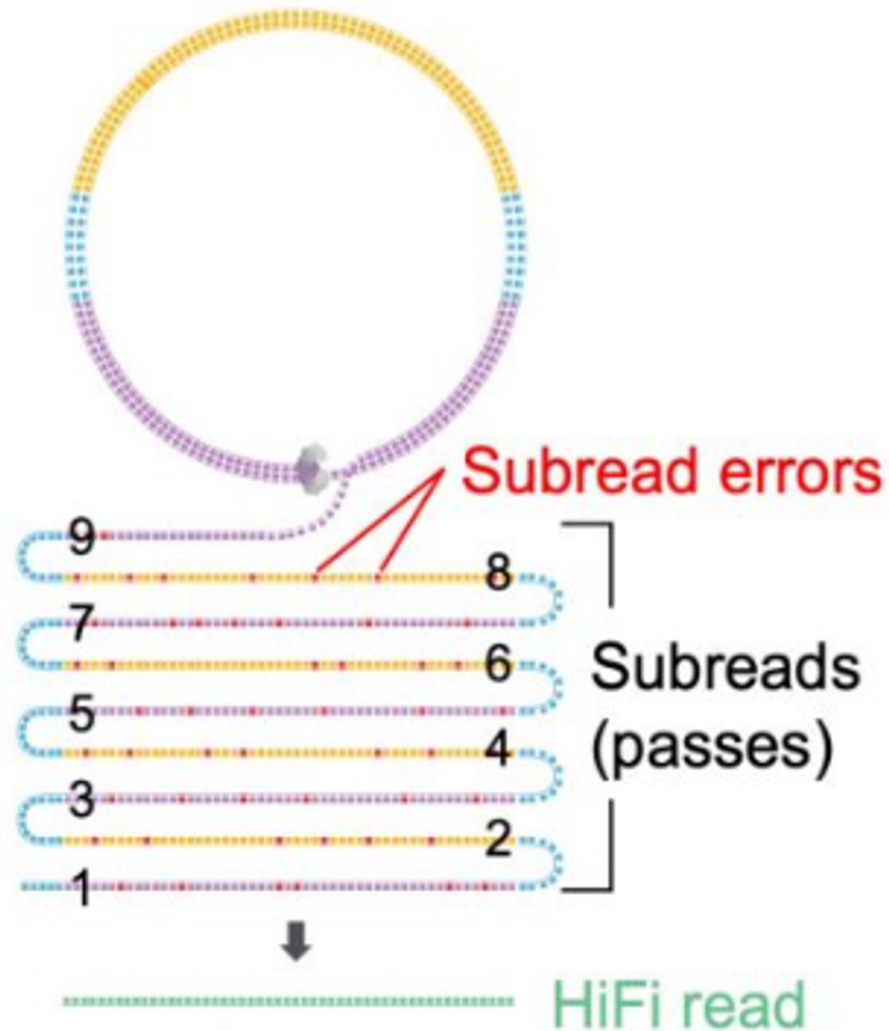
nature

John T. Lovell^{1,36,37}, Alice H. MacQueen^{2,38}, Sujan Mamidi^{1,38}, Jason Bonnette^{2,38}, Jerry Jenkins^{1,38}, Joseph D. Napier¹, Avinash Sreedasyam¹, Adam Healey¹, Adam Session^{3,4}, Shengqiang Shu³, Kerrie Barry³, Stacy Bonos³, LoriBeth Boston¹, Christopher Daum³, Shweta Deshpande³, Aren Ewing³, Paul P. Grabowski³, Taslima Haque³, Melanie Harrison⁴, Jiming Jiang³, Dave Kudrna³, Anna Lipzen³, Thomas H. Pendergast IV^{3,10,31}, Chris Plott³, Peng Qi³, Christopher A. Sasaki³, Eugene V. Shakirov^{2,33}, David Sims³, Manoj Sharma¹⁴, Rita Sharma¹⁵, Ada Stewart¹, Vasanth R. Singan³, Yuhong Tang¹⁶, Sandra Thibivillier¹⁷, Jenell Webber¹⁸, Xiaoyu Weng³, Melissa Williams³, Guohong Albert Wu³, Yuko Yoshinaga³, Matthew Zane³, Li Zhang³, Jiyi Zhang¹⁹, Kathrine D. Behrman², Arvid R. Boe¹⁸, Philip A. Fay¹⁹, Felix B. Fritsch²⁰, Julie D. Jastrow²¹, John Lloyd-Reilley²², Juan Manuel Martínez-Reyna²³, Roser Matamala²¹, Robert B. Mitchell²⁴, Francis M. Rouquette Jr²⁵, Pamela Ronald^{26,27}, Malay Saha¹⁶, Christian M. Tobias²⁸, Michael Udvardi¹⁶, Rod A. Wing², Yanqi Wu²⁹, Laura E. Bartley^{30,31}, Michael Caster^{32,33}, Katrien M. Devos^{34,35,36}, David B. Lowry^{2,35}, Daniel S. Rokhsar^{34,36,37}, Jane Grimwood³, Thomas E. Juenger^{21,38} & Jeremy Schmutz^{1,32,33}

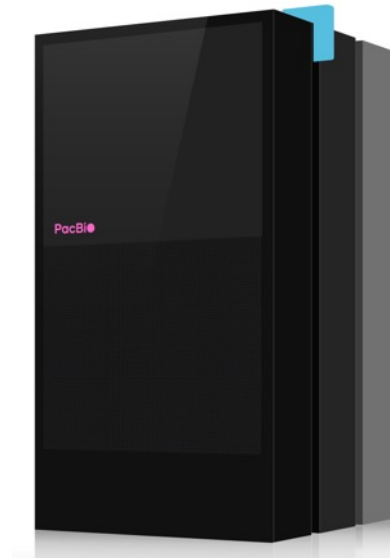


Outbred Tetraploid (4x)
Blended single haplotype

Enter HiFi



PacBio Sequel2 2019



PacBio Revio 2023

- Enabled by longer primary reads (up to 200kb)
- Better data collection on new instruments
- New analysis techniques
- 2023 – new instruments that collect 13x the data for 1/3 cost

DNN model for consensus

DeepConsensus: Gap-Aware Sequence Transformers for Sequence Correction

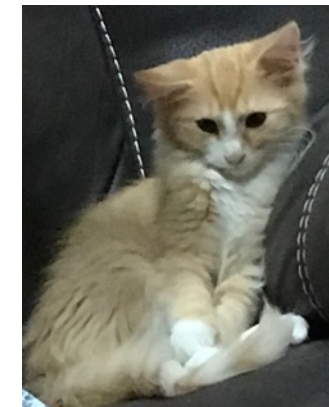
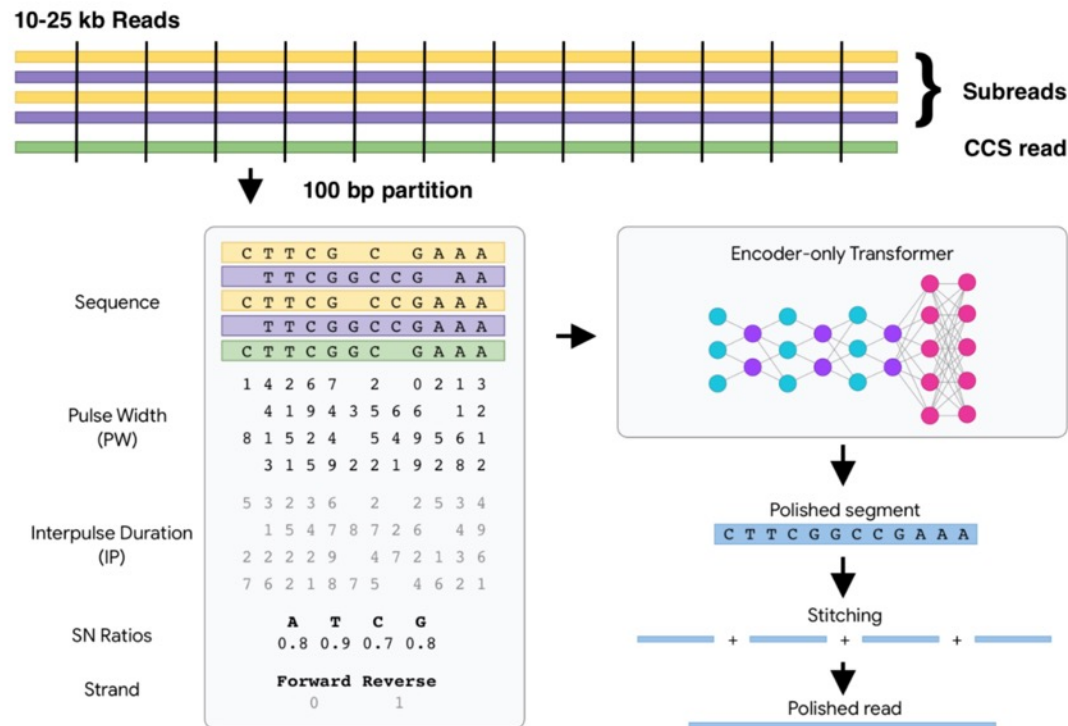
Gunjan Baid^{1*}, Daniel E. Cook^{1*}, Kishwar Shafin¹, Taedong Yun¹, Felipe Llinares-López¹, Quentin Berthet¹, Aaron M. Wenger², William J. Rowell², Maria Nattestad¹, Howard Yang¹, Alexey Kolesnikov¹, Armin Töpfer², Waleed Ammar¹, Jean-Philippe Vert¹, Ashish Vaswani¹, Cory Y. McLean¹, Pi-Chuan Chang^{1^}, Andrew Carroll^{1^}



Pi-Chang Chang



Andrew Carroll



CAT?



Karen Miga, assistant professor of biomolecular engineering at UC Santa Cruz, was named one of the 100 most influential people of 2022 by TIME. The honor was announced on May 23.
Miga and her colleagues, Adam Phillippy, Evan Eichler, and Michael Schatz, led an international team of scientists — the Telomere-to-Telomere (T2T) Consortium — to complete the first, gapless sequence of the human genome. Parts of the human genome are now available to study for the first time, allowing researchers to better understand genetic diseases, human diversity, and evolution.
The gaps now filled by the new sequence include the entire short arms of five human chromosomes and cover some of the most complex regions of the genome. These include highly repetitive DNA sequences found in and

HOME > SCIENCE > VOL. 376, NO. 6588 > FROM TELOMERE TO TELOMERE: THE TRANSCRIPTIONAL AND EPIGENETIC STATE OF HUMAN REPEAT ELEMENTS

SPECIAL ISSUE RESEARCH ARTICLE | HUMAN GENOMICS



From telomere to telomere: The transcriptional and epigenetic state of human repeat elements

SAVANNAH J. HOYT, JESSICA M. STORER, GABRIELLE A. HARTLEY, PATRICK G. S. GRADY, ARIEL GERSHMAN, LEONARDO G. DE LIMA, CHARLES LIMOUSE, REZA HALABIAN, LUKE WOJENSKI, MATIAS RODRIGUEZ, NICOLAS ALTEMOSE, ARANG RHIE, LEIGHTON J. CORE, JENNIFER L. GERTON, WOJ- CIECH MAKALOWSKI, DANIEL OLSON, JEB ROSEN, ARIAN F. A. SMIT, AARON F. STRAIGHT, MITCHELL R. VOLLGER, TRAVIS J. WHEELER, MICHAEL C. SCHATZ, EVAN E. EICHLER, ADAM M. PHILLIPPY, WINSTON TIMP, KAREN H. MIGA, AND RACHEL J. O'NEILL

fewer Authors

Info & Affiliations

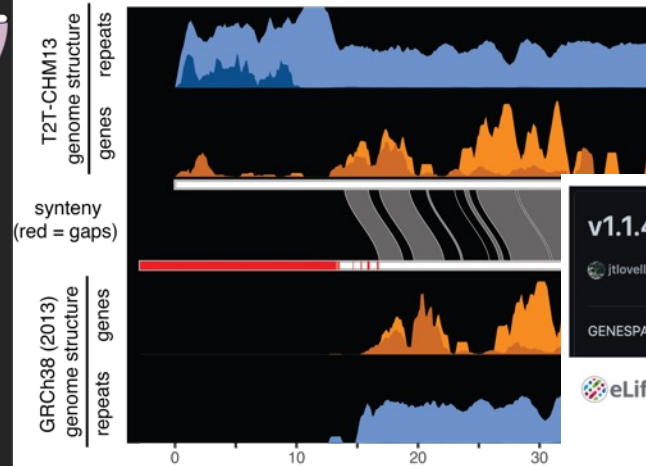
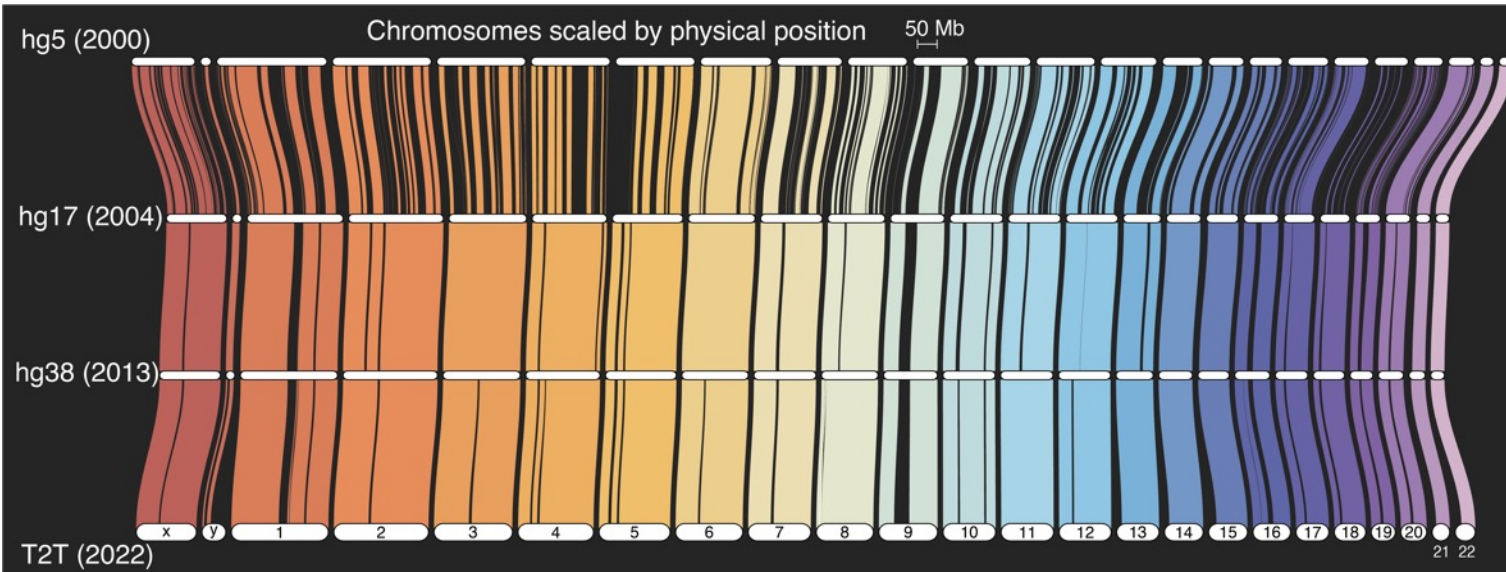
MATTER

Scientists Finish the Human Genome at Last

The complete genome uncovered more than 100 new genes that are probably functional, and many new variants that may be linked to diseases.

Scientists Have Finally Sequenced the Complete Human Genome – And Revealed New Genetic Secrets

TOPICS: DNA Genetics Genome Popular IIC Berkeley



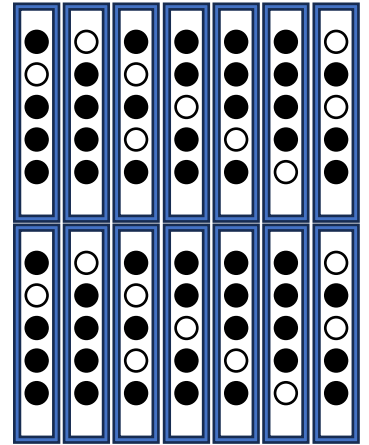
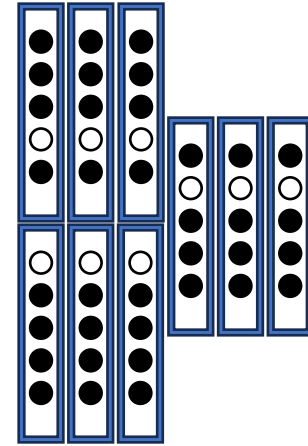
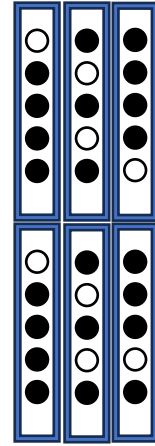
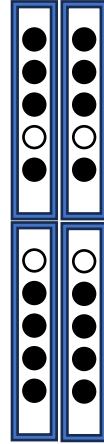
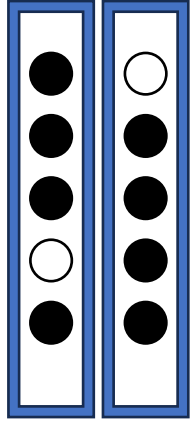
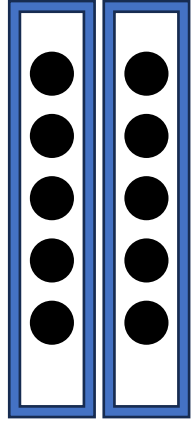
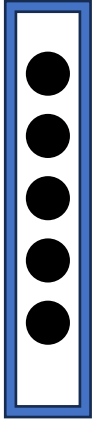
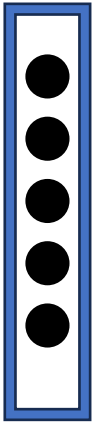
v1.1.4 Latest
jtlovel released this 3 weeks ago · 13 commits to master since this release
GENESPACE v1.1.4 release (updated 6-March 2023)

eLife
GENESPACE tracks regions of interest and gene copy number variation across multiple genomes
John T Lovell^{1,2*}, Avinash Sreedasayam¹, M Eric Schrams¹, Melissa Wilson¹, Joseph W Carlson¹, Alex Harkees¹, David Emma¹, David M Goodstein¹, Jeremy Schmidt¹

Find it on github: jtlovel/GENESPACE

99+% of this is from HIFI alone

Three lines of R to compare multiple complex genomes



Brachypodium
350 Mb x 1
Genes x 1

Sorghum
700 Mb x 1
Genes x 1

Poplar
500 Mb x 2
Genes x 2

Cotton
2,600 Mb x 1
Genes x 2

Switchgrass
1,100 Mb x 2
Genes x 4

Miscanthus
6,100 Mb x 1
Genes x 6

Arundo
1,100 Mb x 2
Genes x 9

Sugarcane
5,000 Mb x 2
Genes x 14

COMPLEXITY

Inbred genome references of small plants

Year	Build	Assembled Size	Contig Number	Contig N50 (MB)
2007	V1 WGS Sanger ¹	688 MB	12,873	0.2
2013	Rio AllPaths Illumina	404 MB	49,137	0.02
2015	V3 WGS Improved ²	675 MB	4,783	1.3
2016	Rio RSII ³	729 MB	3,830	0.4
2018	V4 PacBio Sequel ⁴	721 MB	411	17.7
2020	V5 PacBio Sequel2 ⁴	720 MB	89	50.7
2022	V6 PacBio Hi-Fi ⁴	718 MB	21	55.0



- ¹ Paterson, Nature 2009
- ² McCormick, Plant 2017
- ³ Cooper, BMC genomics 2019
- ⁴ Unpublished

Poplar experiments



2006

RESEARCH ARTICLE

The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray)

G. A. Tuskan^{1,3,*}, S. DiFazio^{1,4,†}, S. Jansson^{5,†}, J. Bohlmann^{6,†}, I. Grigoriev^{9,†}, U. Hellsten^{9,†}, N. Putnam^{9,†}, S. Ralph^{6,†}, S. Rombauts^{10,†}, A. Salamov^{9,†}, J. Schein^{11,†}, L. Sterck^{10,†}, A. Aerts⁹, R. R. Bhalerao⁵, R. P. Bhalerao¹², D. Blaudez¹³, W. Boerjan¹⁰, A. Brun¹³, A. Brunner¹⁴, V. Busov¹⁵, M. Campbell¹⁶, J. Carlson¹⁷, M. Chalot¹³, J. Chapman⁹, G.-L. Chen², D. Cooper⁶, P. M. Coutinho¹³, J. Couturier¹³, S. Covert²⁰, Q. Cronk⁷, R. Cunningham¹, J. Davis²², S. Degroev¹⁰, A. Déjardin²³, C. dePamphilis¹⁸, J. Detter⁹, B. Dirks²⁴, I. Dubchak^{9,25}, S. Duplessis¹³, J. Ehling⁷, B. Ellis⁶, K. Gendler²⁶, D. Goodstein⁹, M. Gribskov²⁷, J. Grimwood²⁸, A. Groover²⁹, L. Gunter¹, B. Hamberger⁷, B. Heinze³⁰, Y. Helariutta^{12,31,33}, B. Henrissat¹⁹, D. Holligan²¹, R. Holt¹¹, W. Huang⁹, N. Islam-Faridi³⁴, S. Jones¹¹, M. Jones-Rhoades³⁵, R. Jorgensen²⁶, C. Joshi¹⁵, J. Kangasjärvi³², J. Karlsson⁵, C. Kelleher⁵, R. Kirkpatrick¹¹, M. Kirst²², A. Kohler¹³, U. Kalluri¹, F. Larimer², J. Leebens-Mack²¹, J.-C. Leplé²³, P. Locascio², Y. Lou⁹, S. Lucas⁹, F. Martin¹³, B. Montanini¹³, C. Napoli²⁶, D. R. Nelson³⁶, C. Nelson³⁷, K. Nieminen³¹, O. Nilsson¹², V. Pereda¹³, G. Peter²², R. Philippe⁶, G. Pilate²³, A. Poliakov²⁵, J. Razumovskaya², P. Richardson⁹, C. Rinaldi¹³, K. Ritland⁸, P. Rouzé¹⁰, D. Ryaboy²⁵, J. Schmutz²⁸, J. Schrader³⁸, B. Segerman⁵, H. Shin¹¹, A. Siddiqui¹¹, F. Sterky³⁹, A. Terry⁹, C.-J. Tsai¹⁹, E. Uberbacher², P. Unneberg²⁹, J. Vahala³², K. Wall¹⁸, S. Wessler²¹, G. Yang²¹, T. Yin¹, C. Douglas^{7,†}, M. Marra^{11,†}, G. Sandberg^{12,†}, Y. Van de Peer^{10,†}, D. Rokhsar^{9,24,†}



- 1,000+ natural genotypes planted in three common gardens
- Resequenced to identify 45M SNPs & phenotyped for many years

Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations:



Mediation of plant–mycorrhizal interaction by a lectin receptor-like kinase

nature plants

A 5-Enolpyruvylshikimate 3-Phosphate Synthase Functions as a Transcriptional Repressor in *Populus*^{OPEN}



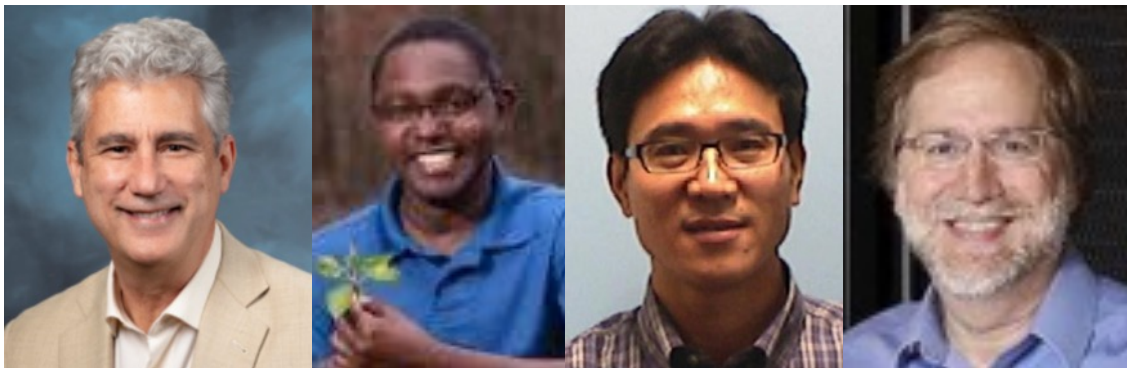
Sequencing and Analysis of the Sex Determination Region of *Populus trichocarpa*



Agave *REVEILLE1* regulates the onset and release of seasonal dormancy in *Populus*

Plant Physiology®

Expression quantitative trait loci mapping identified *PtrXB38* as a key hub gene in adventitious root development in *Populus*



Jerry Tusken
CBI/ORNL

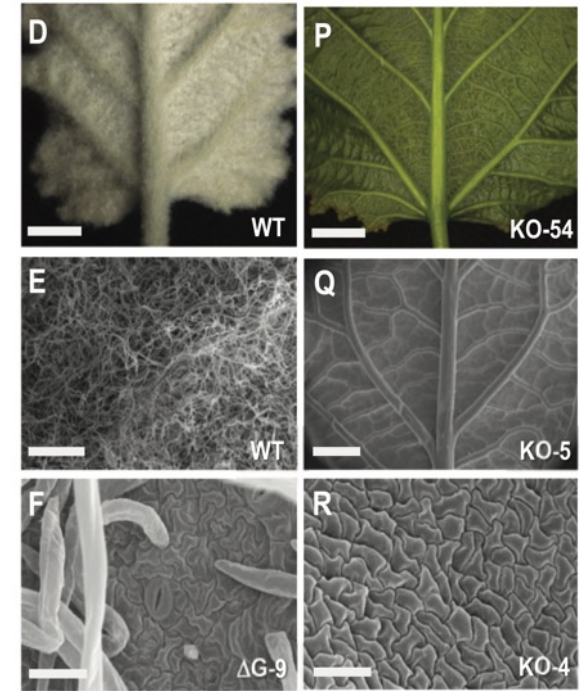
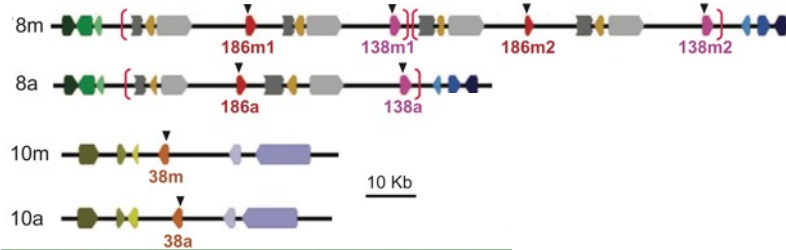
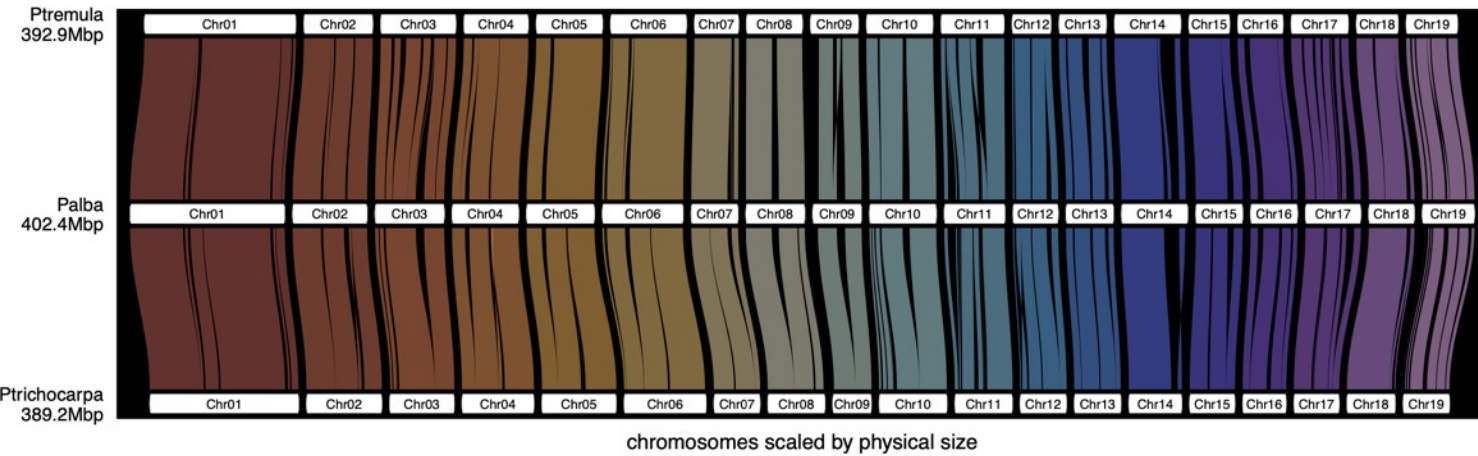
Wellington Muchero
CBI/ORNL

Jay Chen
CBI/ORNL

Dan Jacobsen
CBI/ORNL



Outbred diploid *P. tremula* x *P. alba* = 717-1B4



the plant journal

SEB

The Plant Journal (2023) 116, 1003–1017

doi: 10.1111/tpj.16454

SPECIAL ISSUE ARTICLE

Haplotype-resolved genome assembly of *Populus tremula* x *P. alba* reveals aspen-specific megabase satellite DNA

Ran Zhou^{1,2,3}, Jerry W Jenkins⁴, Yibing Zeng², Shengqiang Shu⁵, Hosung Jang², Scott A. Harding^{1,2,3}, Melissa Williams⁴, Christopher Plott⁴, Kerrie W. Barry⁵, Maxim Koriabine⁵, Mojgan Amirebrahimi⁵, Jayson Talag⁶, Shanmugam Rajasekar⁶, Jane Grimwood⁴, Robert J. Schmitz², R. Kelly Dawe^{2,3}, Jeremy Schmutz^{4,5} and Chung-Jui Tsai^{1,2,3,*}

<https://doi.org/10.1093/plphys/kiac128>

Plant Physiology®

Multiplex knockout of trichome-regulating MYB duplicates in hybrid poplar using a single gRNA

William P. Bewg^{1,2,3}, Scott A. Harding^{1,2,3}, Nancy L. Engle⁴, Brajesh N. Vaidya^{5,*}, Ran Zhou^{1,2,3}, Jacob Reeves^{6,5}, Thomas W. Horn⁶, Nirmal Joshee⁵, Jerry W. Jenkins^{7,8}, Shengqiang Shu⁸, Kerrie W. Barry⁸, Yuko Yoshinaga⁸, Jane Grimwood^{7,8}, Robert J. Schmitz², Jeremy Schmutz^{7,8}, Timothy J. Tschaplinski⁴ and Chung-Jui Tsai^{1,2,3,*†}



C.J. Tsai
UGA & CBI

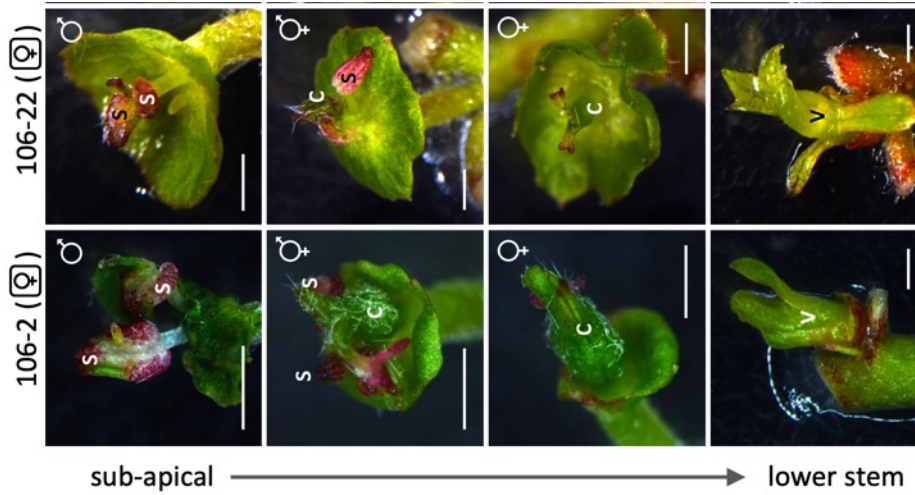
Applying 717



Early Flower Induction



Indeterminate growth after harvest

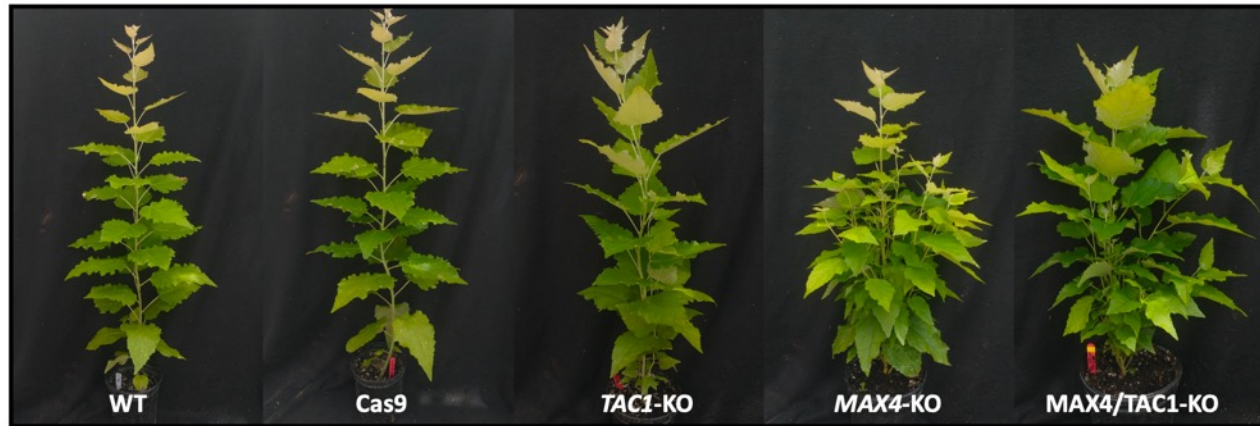


Male
Female
Hermaphrodite
Flowers

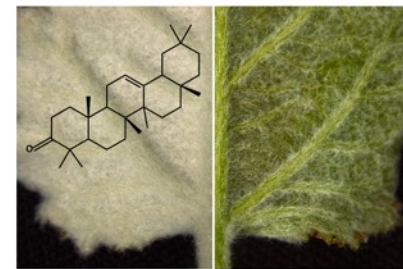
5-7 days after node culture



C.J. Tsai
UGA & CBI

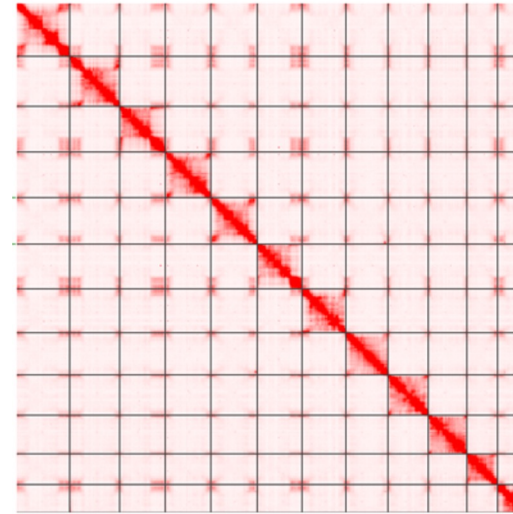
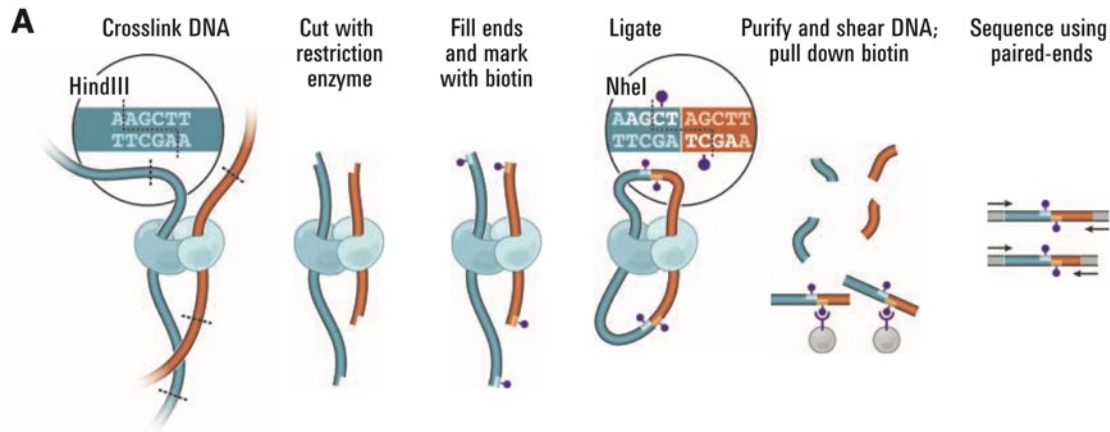


Morphotypes (architecture/trichome)
Chemotypes (bisabolene/cell wall)



Robin Buell, Chris Dardick, Wayne Parrott,
Bob Schmitz, Patrick Shih, CJ Tsai,
Breeanna Urbanowicz





Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

Erez Lieberman-Aiden,^{1,2,3,4*} Nynke L. van Berkum,^{5*} Louise Williams,¹ Maxim Imakaev,² Tobias Ragozy,^{6,7} Agnes Telling,^{6,7} Ido Amit,¹ Bryan R. Lajoie,⁵ Peter J. Sabo,⁸ Michael O. Dorschner,⁸ Richard Sandstrom,⁸ Bradley Bernstein,^{1,9} M. A. Bender,¹⁰ Mark Groudine,^{6,7} Andreas Gnirke,¹ John Stamatoyannopoulos,⁸ Leonid A. Mirny,^{2,11} Eric S. Lander,^{1,12,13†} Job Dekker^{5†}



Dovetail[®] Omni-C[®] Kit

Produce the highest quality genome assembly.

Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm

[Haoyu Cheng](#), [Gregory T. Concepcion](#), [Xiaowen Feng](#), [Haowen Zhang](#) & [Heng Li](#) ✉

Nature Methods **18**, 170–175 (2021) | [Cite this article](#)

Haplotype-resolved assembly of diploid genomes without parental data

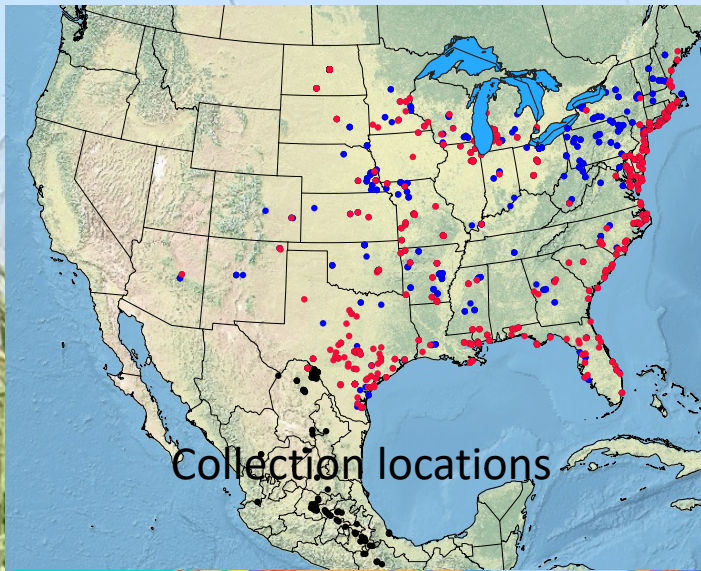
[Haoyu Cheng](#), [Erich D. Jarvis](#), [Olivier Fedrigo](#), [Klaus-Peter Koepfli](#), [Lara Urban](#), [Neil J. Gemmell](#) & [Heng Li](#) ✉

Nature Biotechnology **40**, 1332–1335 (2022) | [Cite this article](#)



Heng Li
Harvard

Also wrote BWA and Minimap



Collection locations



Planting locations

Article
Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass

<https://doi.org/10.1038/s41586-020-03127-1> John T. Lovell^{1,2,3,10}, Alice H. MacQueen^{1,3}, Sujan Mamidi^{1,3}, Jason Bonnette^{3,10}, Jerry Jenkins^{1,3}, Joseph D. Napier¹, Avinash Sreedasyam¹, Adam Healey¹, Adam Session^{1,4}, Shengqiang Shu¹, Kerrie Barry¹, Stacy Bonos¹, LoriBeth Boston¹, Christopher Daum¹, Shweta Deshpande¹, Aren Ewing¹, Paul P. Grabowski¹, Taslima Haque¹, Melanie Harrison¹, Jining Jiang¹, Dave Kudrna¹, Anna Lipzen¹, Thomas H. Pendergast IV^{1,10}, Chris Plott¹, Pang Qi¹, Christopher A. Sasaki¹, Eugene V. Shabirov¹, David Sims¹, Manoj Sharma¹, Rita Sharma¹, Ada Stewart¹, Vasanth R. Singan¹, Yuhong Tang¹, Sandra Thibault¹, Weng¹, Melissa Williams¹, Guohong Albert Wu¹, Yuko Yoshinaga¹, Ji Yi Zhang¹, Kathrine D. Behrman¹, Arvid R. Boe¹, Philip A. Fay¹, J. Jastrow¹, John Lloyd-Reilley¹, Juan Manuel Martínez-Reyna¹, et B. Mitchell¹, Francis M. Rouquette Jr¹, Pamela Ronald^{1,10}, M. Tobias¹, Michael Udvardi¹, Rod A. Wing¹, Yanqi Wu¹, Laura E. Bartley^{1,10}, Michael Casler^{1,10}, Katrien M. Dewo^{1,10}, David B. Lowry^{1,10}, Daniel S. Rokhsar^{1,4,10,11}, Jane Grimwood¹, Thomas E. Juenger^{1,2} & Jeremy Schmutz^{1,10}

Received: 1 July 2020
 Accepted: 16 December 2020
 Published online: 27 January 2021

Open access [Check for updates](#)

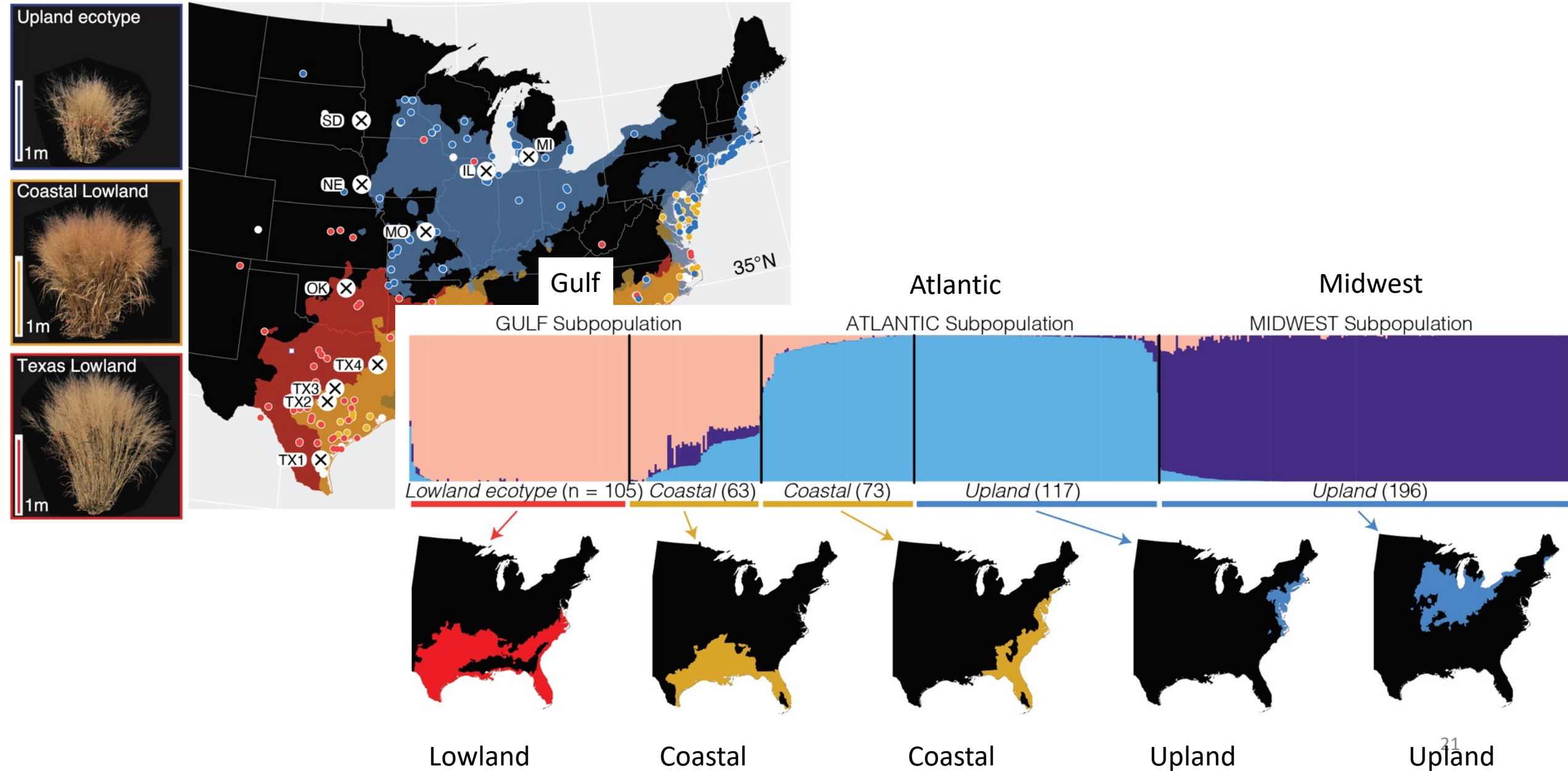
nature



Tom Juenger
UT-Austin



Ecotypes & genetic groups

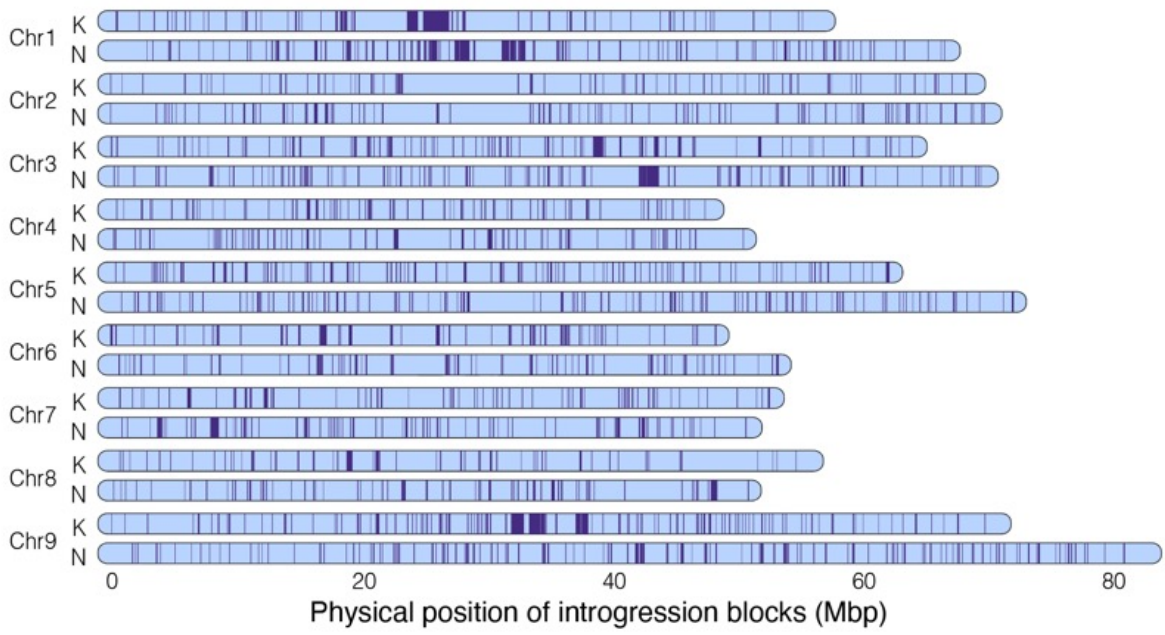




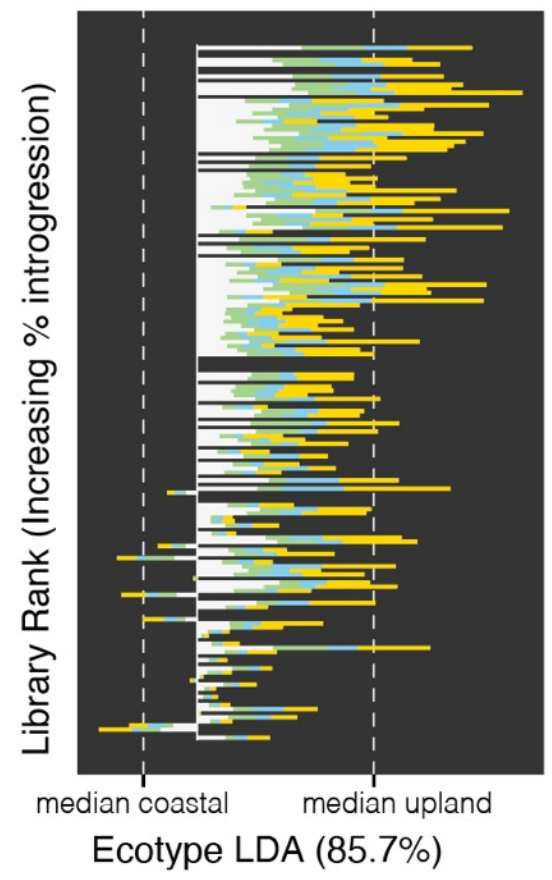
Upland

Coastal

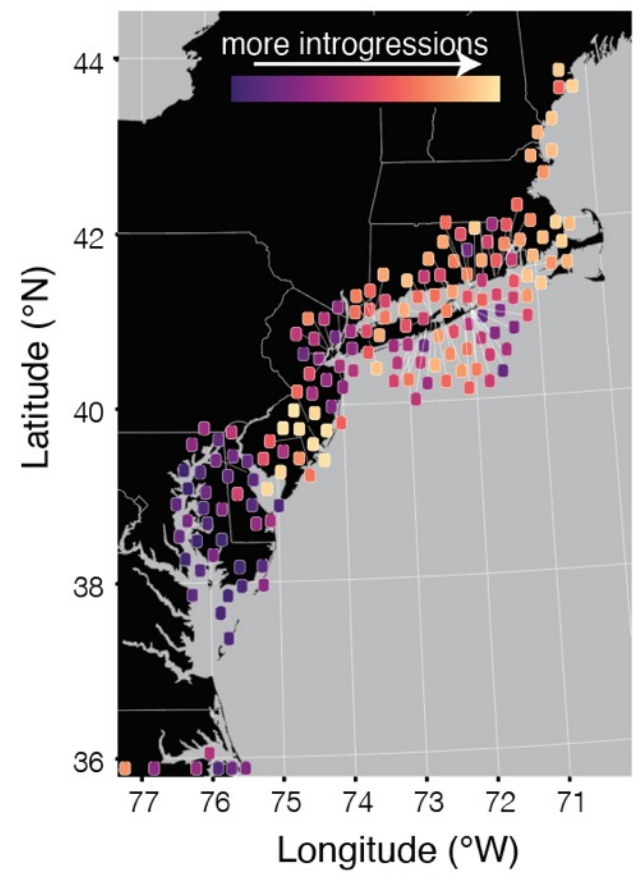
Upland



17 - 34 kyrs, ~8k generations
 19 - 27 kyrs, last glacial maxima



More introgressions = more upland like and greater cold tolerance



Demonstrates/proves how selection and evolution act in a complex native grass.

Today's switchgrass genomes

Release	Size (GB)	Contigs	CN50 (MB)
AP13 V5 CLR	1.10	482	4.3
AP13 V6 H1	1.14	30	55.2
AP13 V6 H2	1.13	33	52.4

4x Upland



4x Lowland₂₃

Article

The complex polyploid genome architecture of sugarcane

nature

<https://doi.org/10.1038/s41586-024-07231-4>
 Received: 24 February 2023
 Accepted: 23 February 2024
 Published online: 27 March 2024
 Open access

A. L. Healey^{1,2,3}, O. Garsmeur^{2,3}, J. T. Lovell^{1,4}, S. Shengquiang⁴, A. Sreedasyam¹, J. Jenkins¹, C. B. Platt¹, N. Piperidis⁵, N. Pompidor^{2,3}, V. Llaca⁶, C. J. Metcalfe⁷, J. Doležel⁸, P. Čápal⁸, J. W. Carlson⁹, J. Y. Hoarau^{2,3,5}, C. Hervouet^{2,3}, C. Zini^{2,3}, A. Dievart^{2,3}, A. Lipzen⁹, M. Williams¹, L. B. Boston¹, J. Webber¹, K. Keymanesh¹, S. Tejomurthula¹, S. Rajasekar¹⁰, R. Suchecki¹¹, A. Furtado¹², G. May¹, P. Parakkal¹³, B. A. Simmons^{2,3,5}, K. Barry¹, R. J. Henry^{2,3,4}, J. Grimwood¹, K. S. Aitken¹, J. Schmutz^{1,4,10} & A. D'Hont^{2,3,10}

JGI CSPs

- 502977 JBEI – Sequencing Sugarcane BACs
- 502967-Understanding polyploidy through the generation of the first sugarcane genome sequence - 2017
- 504319- Genomic Diversity in the *Saccharum* Complex - 2019



Angelique D'Hont



Karen Aiken



Robert Henry



Sugarcane is complicated

S. officinarum
8X polyploid
domesticated



S. spontaneum
5-16X polyploid
wild

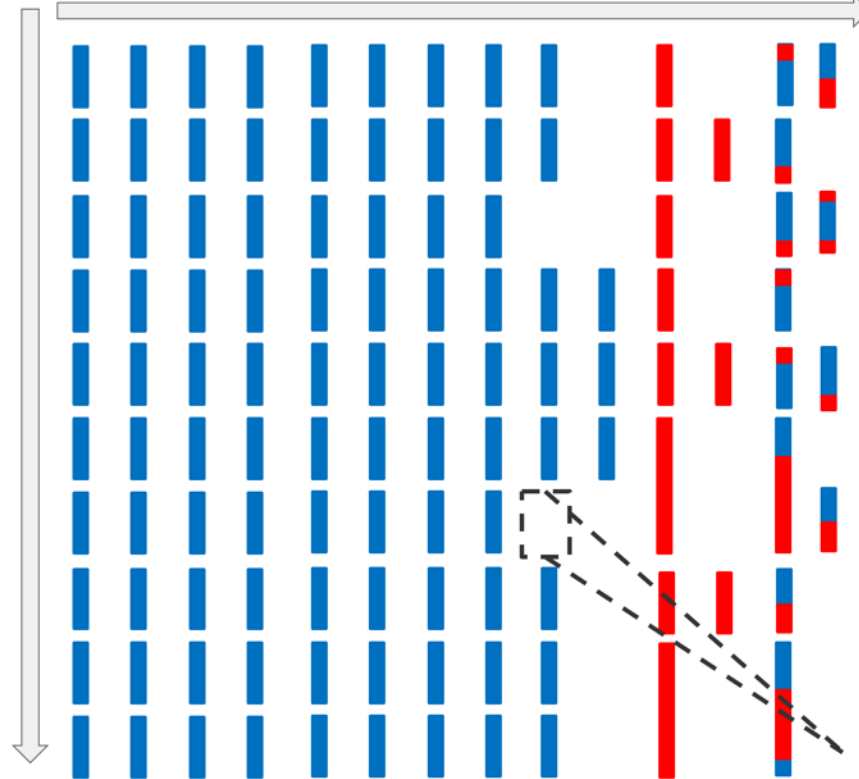


Modern Hybrids
Genotype R570

Chromosome Copy Number -Ploidy

Sorghum

Base Chromosomes



10 Gb
Genome
-110-130

Multiple sequencing attempts



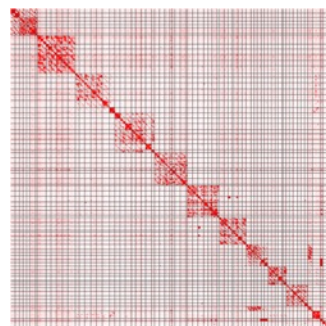
Release	Size	Contigs	Contig N50
Illumina	5.0 Gb	1,700,000	4.4 Kb
Long read	7.4 Gb	33,222	482 kb
HiFi	9.6 Gb	38,822	10 Mb

HiFi version

- Better splitting of haplotypes
- ~20X longer contigs
- ~11,000 more genes captured



HiC for validation

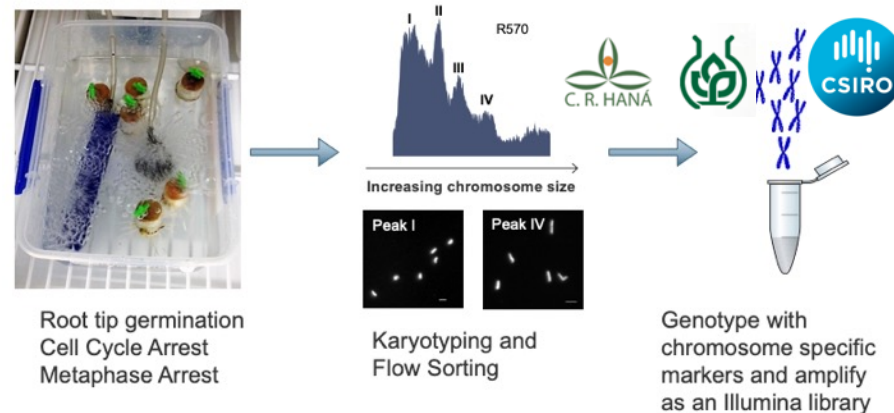


Genetic map (Self1)

ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA
 ACACGCCTAGGCATCCTACGGTTGTACAATAA

Single dose SNP

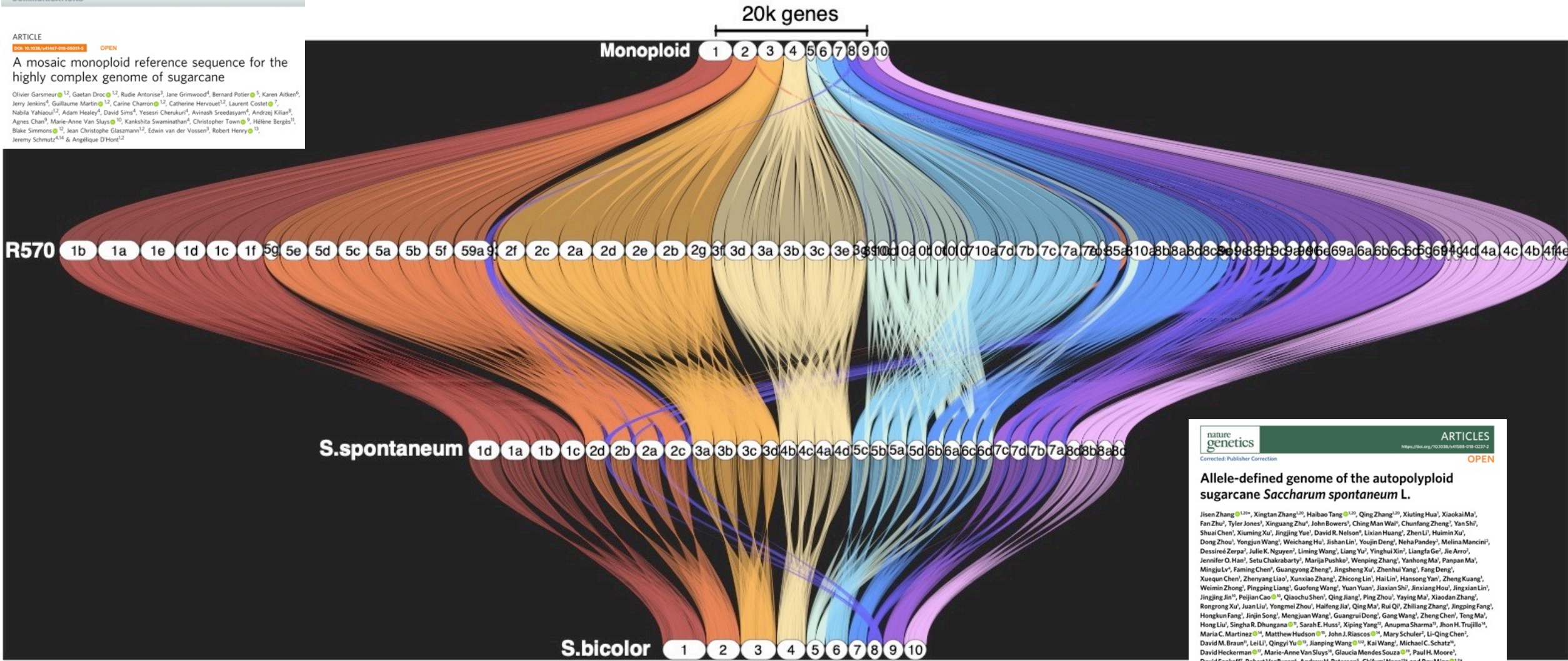
Single chromosome sort libraries (SCL)



BioNano optical map



ARTICLE
 DOI: 10.1038/s41467-018-05201-5 OPEN
A mosaic monoploid reference sequence for the highly complex genome of sugarcane
 Olivier Garsmeur^{1,2}, Gaetan Droc^{1,2}, Rudie Antonisa³, Jane Grimwood⁴, Bernard Potier⁵, Karen Aitken⁶, Jerry Jenkins⁴, Guillaume Martin^{1,2}, Carine Charron^{1,2}, Catherine Hervouet^{1,2}, Laurent Costet⁷, Nabila Yahiaoui^{1,2}, Adam Healey⁸, David Sims⁹, Yesesri Cherukuri⁹, Avinash Sreedhasyam⁹, Andrzej Kilian⁹, Agnes Chan⁹, Marie-Anne Van Sluys¹⁰, Kankshita Swaminathan⁹, Christopher Town⁹, Helène Bergès¹¹, Blake Simmons⁹, Jean Christophe Glazmann¹², Edwin van der Vossen⁷, Robert Henry¹³, Jeremy Schmutz^{1,14} & Angélique D'Hont¹²



nature genetics
 Corrected: Publisher Correction
 ARTICLES
 https://doi.org/10.1038/s41588-018-0232-2
 OPEN

Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L.

Jisen Zhang^{1,20*}, Xingtian Zhang¹⁰, Haibao Tang^{1,20}, Qing Zhang¹⁰, Xiuting Hua¹, Xiaokai Ma¹, Fan Zhu¹, Tyler Jones¹, Xinguang Zhu¹, John Bowers¹, Ching-Man Wai¹, Chunfang Zheng¹, Yan Shi¹, Shuai Chen¹, Xiuming Xu¹, Jingjing Yue¹, David R. Nelson¹, Lixian Huang¹, Zhen Li¹, Huimin Xu¹, Dong Zhou¹, Yongjun Wang¹, Weichang Hu¹, Jishan Lin¹, Youjin Deng¹, Neha Pandey¹, Melina Mancini¹, Désirée Zappa¹, Julie K. Nguyen¹, Liming Wang¹, Liang Yu¹, Yinghui Xin¹, Liangfa Ge¹, Jie Arro¹, Jennifer O'Han¹, Setu Chakrabarty¹, Marija Pushko¹, Wensping Zhang¹, Yanhong Mai¹, Panpan Mai¹, Mingju Lv¹, Faming Chen¹, Guangyong Zheng¹, Jingsheng Xu¹, Zhenhui Yang¹, Fang Deng¹, Xuequn Chen¹, Zhenyang Liao¹, Xunxiao Zhang¹, Zhicong Lin¹, Hai Lin¹, Hansong Yan¹, Zheng Kuang¹, Weimin Zhong¹, Pingping Liang¹, Guofeng Wang¹, Yuan Yuan¹, Jiaxian Shi¹, Jinxiang Hou¹, Jingxian Lin¹, Jingjing Jin¹, Peijian Cao¹⁰, Qiaochu Shen¹, Qing Jiang¹, Ping Zhou¹, Yaying Ma¹, Xiaodan Zhang¹, Rongrong Xu¹, Juan Liu¹, Yongmei Zhou¹, Haifeng Jia¹, Qing Ma¹, Rui Qi¹, Zhiliang Zhang¹, Jingping Fang¹, Hongkun Fang¹, Jinjin Song¹, Mengjuan Wang¹, Guangrui Dong¹, Gang Wang¹, Zheng Chen¹, Teng Ma¹, Hong Liu¹, Singha R. Dhungana¹, Sarah E. Huss¹, Xiping Yang¹, Anupma Sharma¹, Jhon H. Trujillo¹⁰, Maria C. Martinez¹⁰, Matthew Hudson¹⁰, John J. Riascos¹⁰, Mary Schuler¹, Li-Qing Chen¹, David M. Braun¹, Lei Li¹, Qingyi Yu¹, Jianping Wang¹⁰, Kai Wang¹, Michael C. Schatz¹, David Heckerman¹, Marie-Anne Van Sluys⁹, Glauca Mendes Souza¹⁰, Paul H. Moore¹, David Sankoff¹, Robert VanBuren¹, Andrew H. Paterson¹, Chifumi Nagai¹⁰ and Ray Ming^{1,14}

Primary Path
5.1 Gb Genome (Primary Path)
844 Contigs (N50: 15Mb)
68 Chromosomes

Chromosomes scaled by gene rank order

Today: We can do this again with just HiFi and Omni-C

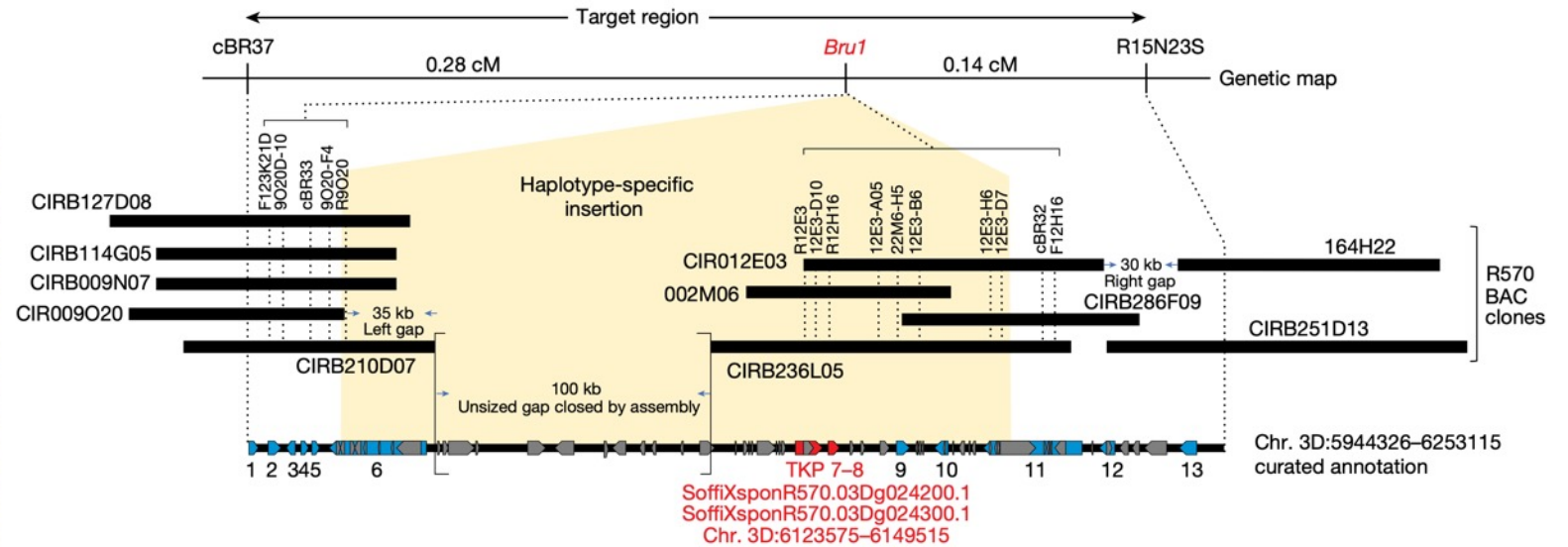
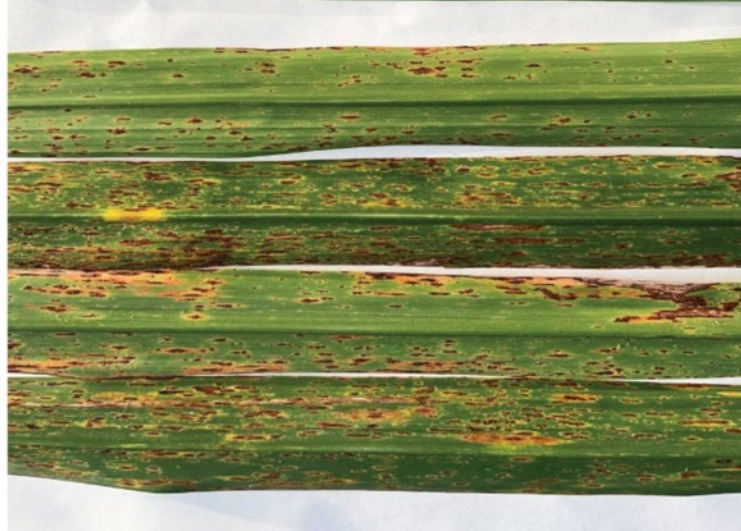
Brown Rust Resistance (*Bru1*) - *Puccinia melanocephala*

R570 Selfed offspring

Resistant



Susceptible



- Before 1980, brown rust caused ~50% yield losses
- From R570, the first resistant cultivar, PCR markers were developed
- Single-copy haplotype-specific insertion that does not recombine
- Candidate gene ended up a two gene, kinase-pseudokinase complex

Terpenes in conifer genomes CSP 503037



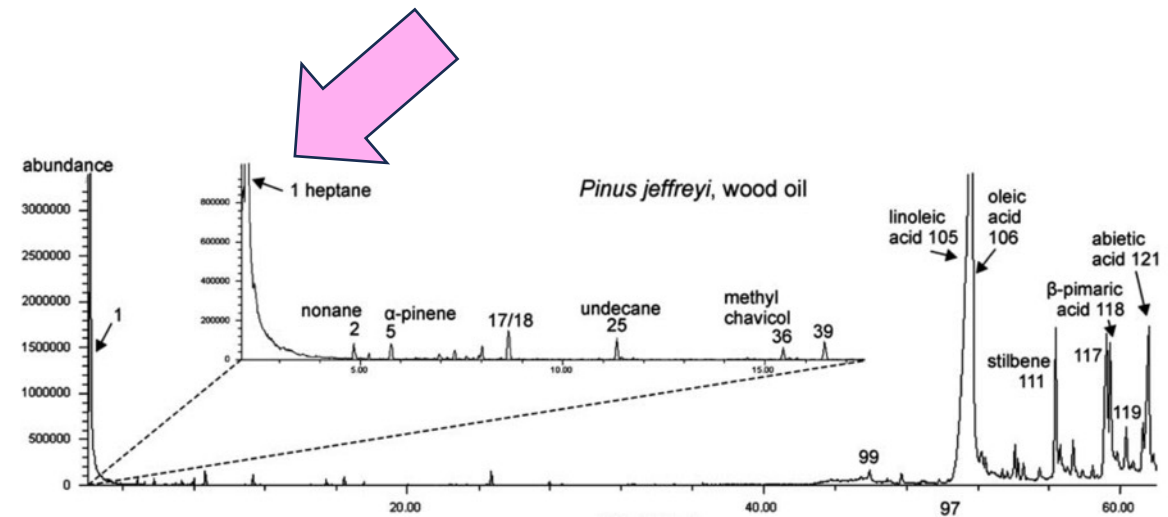
Jorg Bohlmann
UBC



Jay Keasling
LBNL/JBEI



Elizabeth Sattely
Stanford



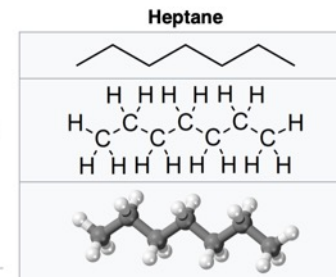
Taxus baccata
English Yew 10GB x 2



Thuja plicata – 12 Gb x 2
Western Red Cedar



Pinus jeffreyi– 25 Gb x 2
Jeffrey's pine



0
Octane Scale

Journal of Essential Oil Research

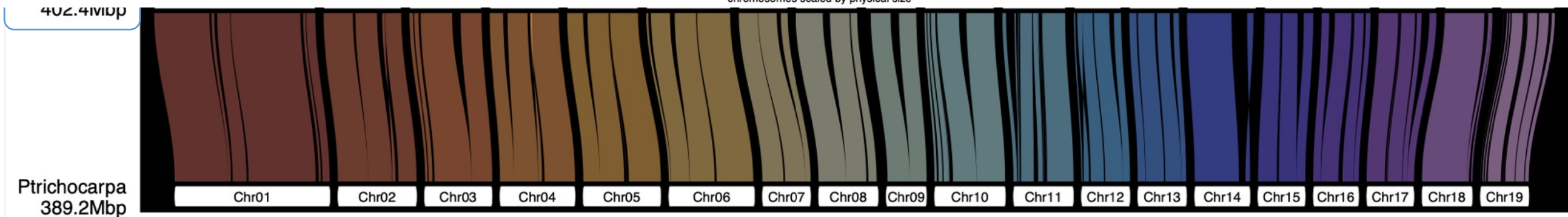
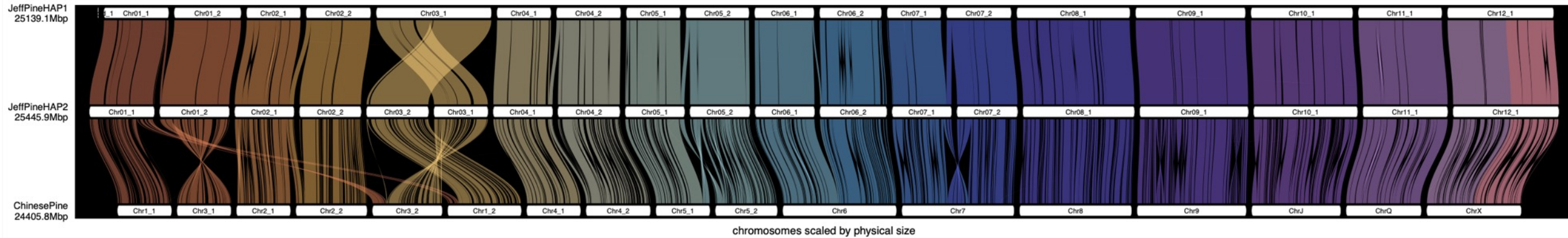
Publication details, including instructions for authors and subscription information:
<http://www.tandfonline.com/loi/tjeo20>

Alkanes and Terpenes in Wood and Leaves of *Pinus jeffreyi* and *P. sabiniana*

Robert P. Adams^a & Jessica W. Wright^b

Comparative genomics in large genomes

DEEPSPACE synteny map



Pinus jeffreyi: 25.2 GBs (63x larger than poplar)

	Technology	Total Contigs	Contig N50 (Mb)
Chinese Pine	Long read	15,385	2.7
Jeff Pine	HiFi	1,508	30.0

Niu, Shihui, et al. "The Chinese pine genome and methylome unveil key features of conifer evolution." *Cell* 185.1 (2022): 204-217.

Great challenge in plant genomics

- What do these genes do?
- What effect does variation in these genes have on our traits of interest?
- What about polyploids?

Genome	No. genes	% GGF
<i>A. thaliana</i> Araport11	27 655	91.33
<i>B. distachyon</i>	34 310	72.56
<i>C. reinhardtii</i>	17 741	43.08
<i>E. grandis</i>	36 349	79.74
<i>G. max</i>	52 872	80.37
<i>K. fedtschenkoi</i>	30 964	82.01
<i>M. truncatula</i>	50 894	67.94
<i>P. hallii</i> var. <i>filipes</i>	33 805	72.65
<i>P. hallii</i> var. <i>hallii</i>	33 263	73.36
<i>P. patens</i>	32 926	55.44
<i>P. trichocarpa</i>	34 699	82.31
<i>P. virgatum</i>	80 278	69.2
<i>S. bicolor</i>	34 129	71.51
<i>S. bicolor</i> Rio	35 490	69.16
<i>S. fallax</i>	25 100	78.31
<i>S. italica</i>	34 584	77
<i>S. viridis</i>	38 334	70.43
<i>L. albus</i>	38 258	78.17



Debbie Laudencia-Chingcuanco, USDA
CSP #1670



PNAS

Grasses use an alternatively wired bHLH transcription factor network to establish stomatal identity

Michael T. Raissig^{a,1}, Emily Abrash^{a,1}, Akhila Bettadapur^b, John P. Vogel^f, and Dominique C. Bergmann^{a,b,2}



Mutations in the predicted DNA polymerase subunit POLD3 result in more rapid flowering of *Brachypodium distachyon*

the plant journal



Original Article Open Access

KARRIKIN INSENSITIVE2 regulates leaf development, root system architecture and arbuscular-mycorrhizal symbiosis in *Brachypodium distachyon*

BMC Plant Biology

RESEARCH

Open Access

BdERECTA controls vasculature patterning and phloem-xylem organization in *Brachypodium distachyon*



Journal of Experimental Botany, Vol. 67, No. 1, pp. 227–237, 2016
doi:10.1093/jxb/erv446 Advance Access publication 3 October 2015
This paper is available online free of access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

RESEARCH PAPER

Mutation in *Brachypodium* caffeic acid O-methyltransferase 6 alters stem and grain lignins and improves straw saccharification without deteriorating grain quality



REPORT

Science

PLANT SCIENCE

Mobile MUTE specifies subsidiary cells to build physiologically improved grass stomata

Michael T. Raissig^{1,a}, Juliana L. Matos¹, M. Ximena Anleu Gil², Ari Kornfeld², Akhila Bettadapur², Emily Abrash¹, Hannah R. Allison¹, Grayson Badgley², John P. Vogel², Joseph A. Berry², Dominique C. Bergmann^{1,2,a}

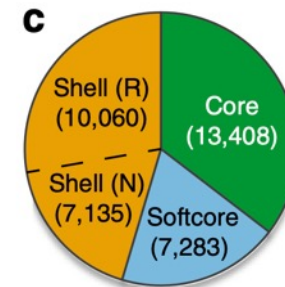
Published online 1 August 2023

Nucleic Acids Research, 2023, Vol. 51, No. 16 8383–8401
<https://doi.org/10.1093/nar/gkad616>

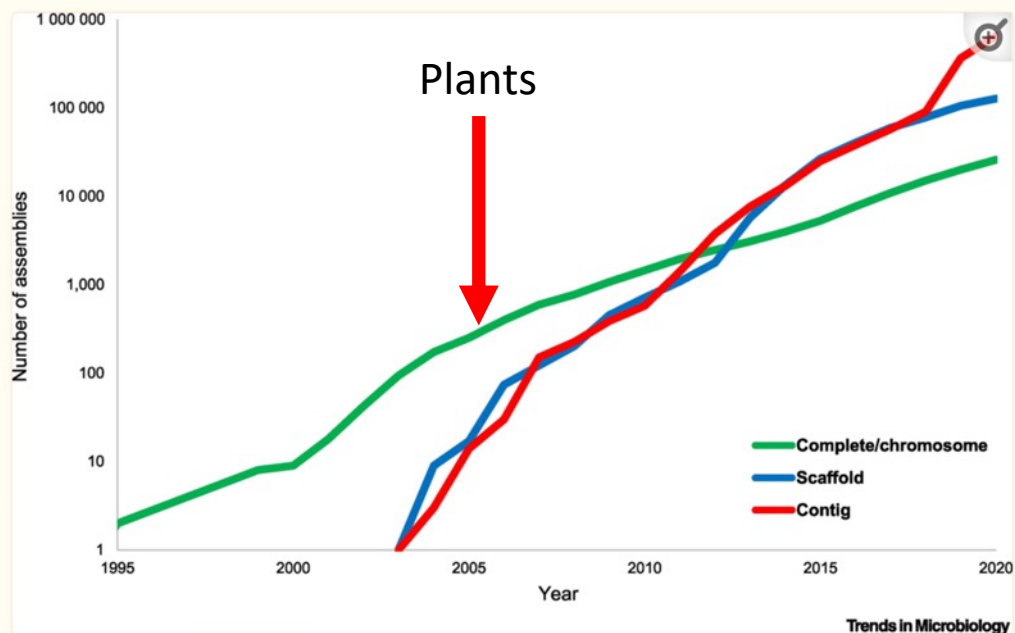
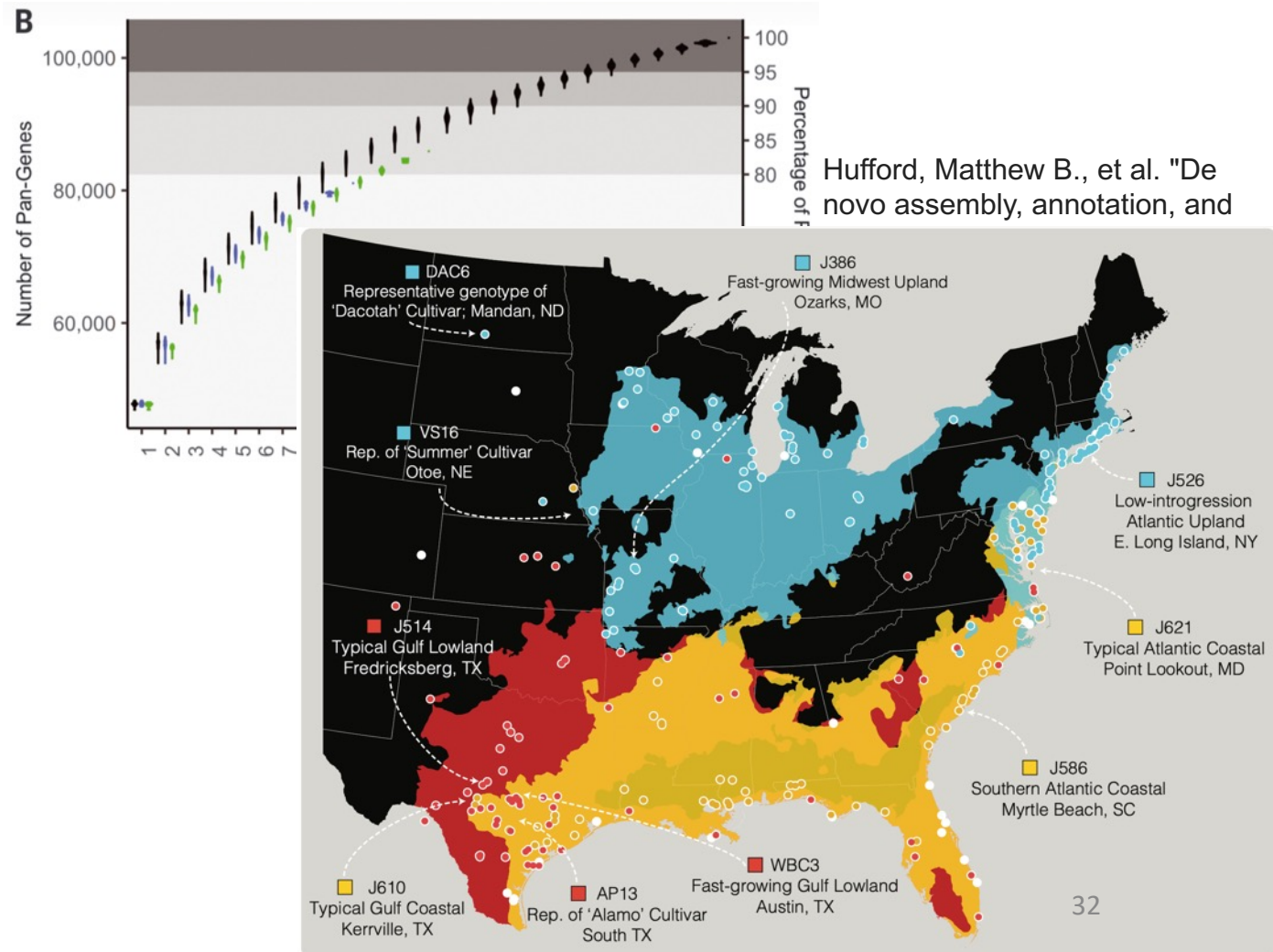
JGI Plant Gene Atlas: an updateable transcriptome resource to improve functional gene descriptions across the plant kingdom

Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure

Sean P. Gordon¹, Bruno Contreras-Moreira^{2,3,4}, Daniel P. Woods^{5,6}, David L. Des Marais^{7,17}, Diane Burgess⁸, Shengqiang Shu¹, Christoph Stritt⁹, Anne C. Roulin⁹, Wendy Schackwitz¹, Ludmila Tyler¹⁰, Joel Martin¹, Anna Lipzen¹, Niklas Dochy¹¹, Jeremy Phillips¹, Kerrie Barry¹, Koen Geuten¹¹, Hikmet Budak¹², Thomas E. Juenger¹³, Richard Amasino^{5,6}, Ana L. Caicedo¹⁰, David Goodstein¹, Patrick Davidson¹, Luis A. J. Mur¹⁴, Melania Figueroa¹⁵, Michael Freeling⁸, Pilar Catalan^{4,16} & John P. Vogel^{1,8}



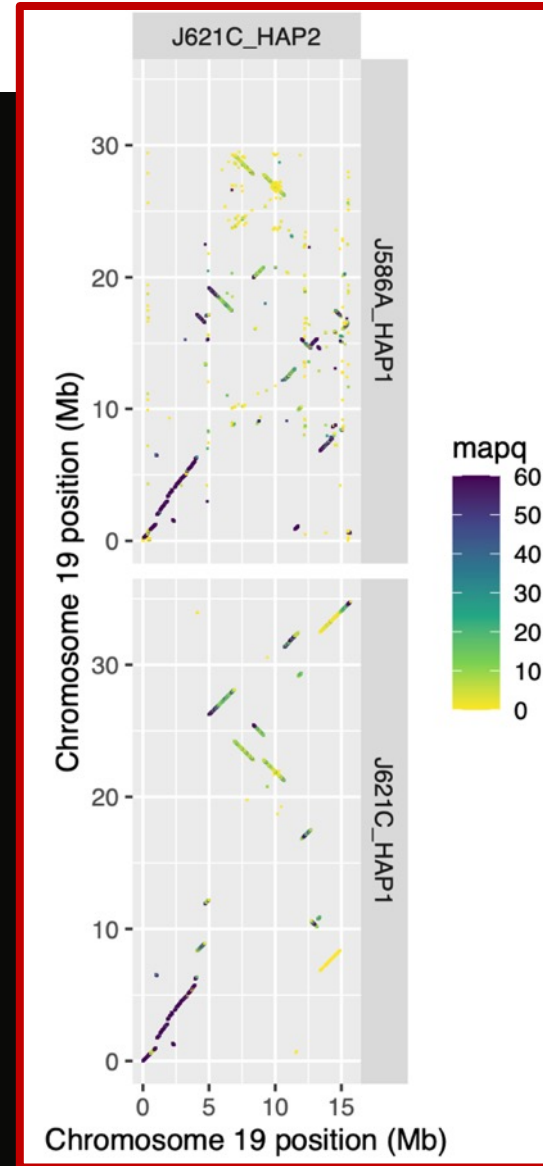
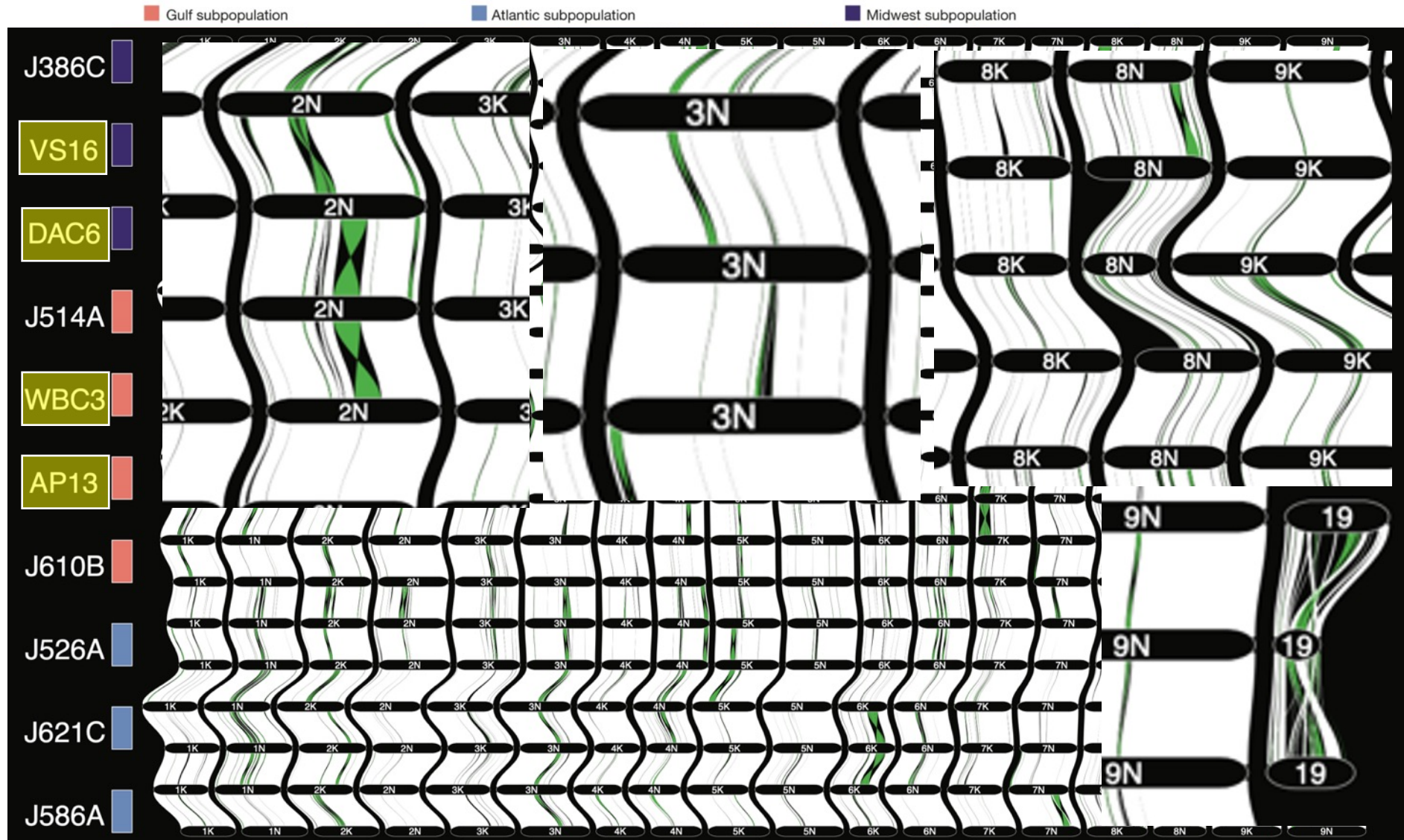
Gene variation in 54 lines



Microbe references

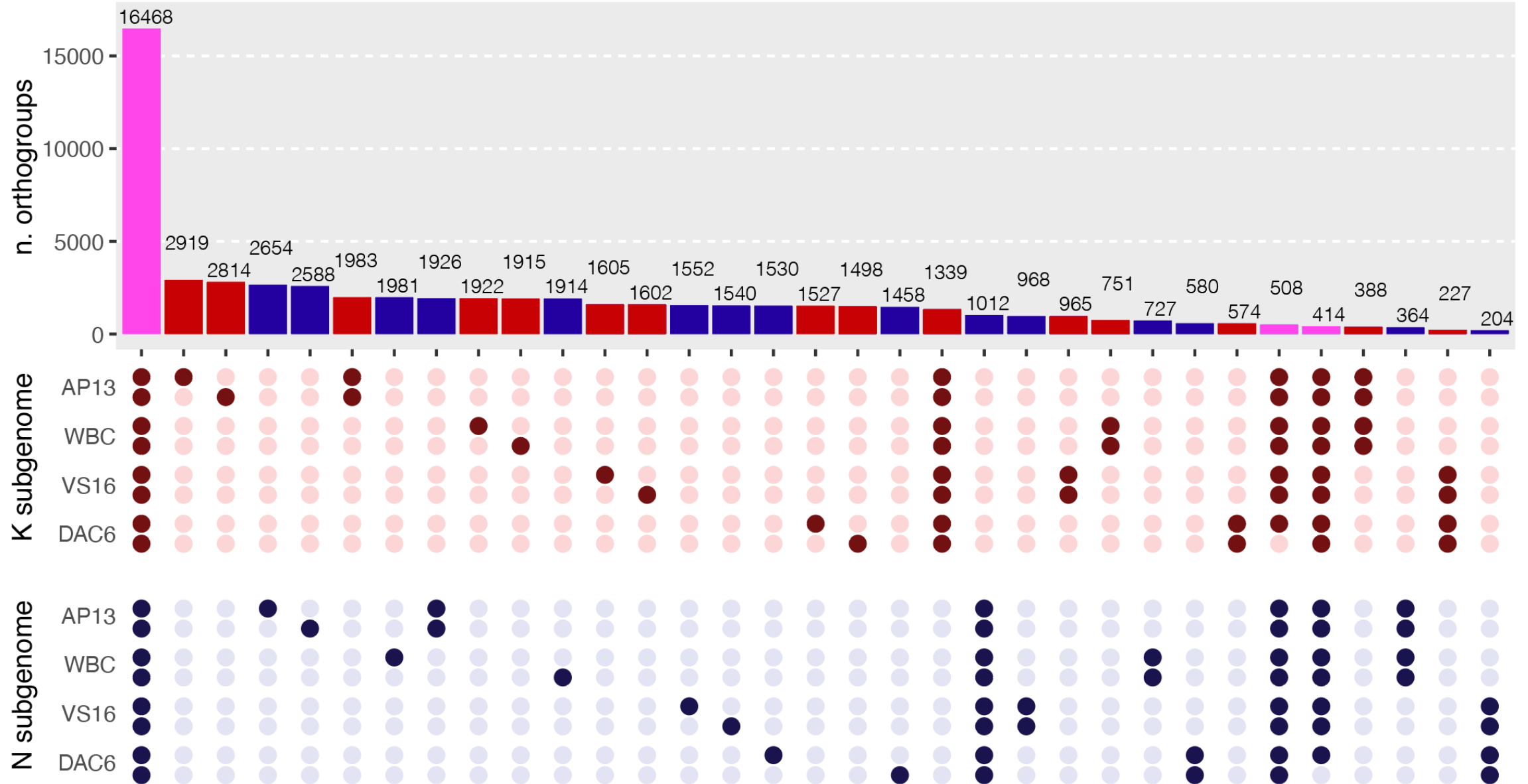
Koonin EV, Makarova KS, Wolf YI. Evolution of Microbial Genomics: Conceptual Shifts over a Quarter Century. Trends Microbiol. 2021 Jul;29(7):582-592.

Switchgrass pangenome



Pangenomes allow us to capture diversity in structural variants that we can't easily identify with short read sequencing³⁸

Switchgrass gene cluster variation



A switchgrass pangenome graph

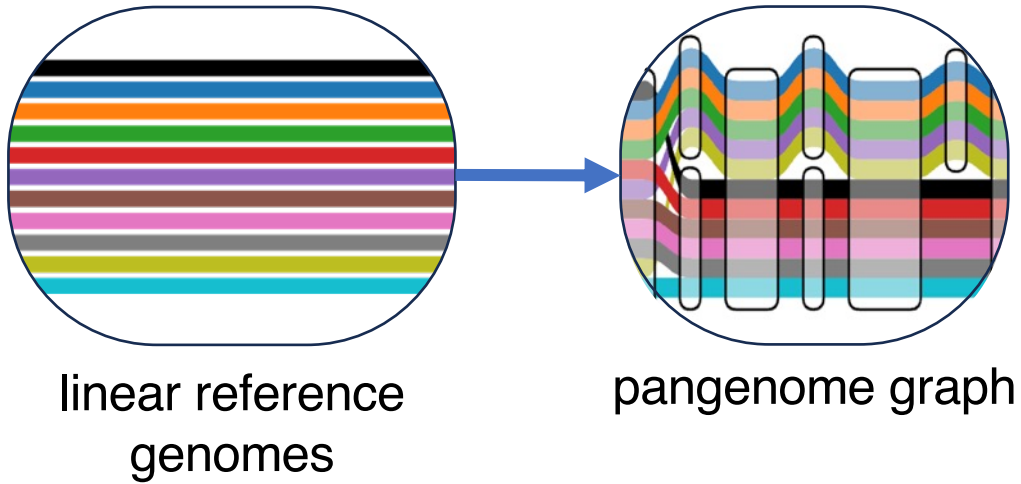


Fig. 1 | References are well represented in final pangenome graph

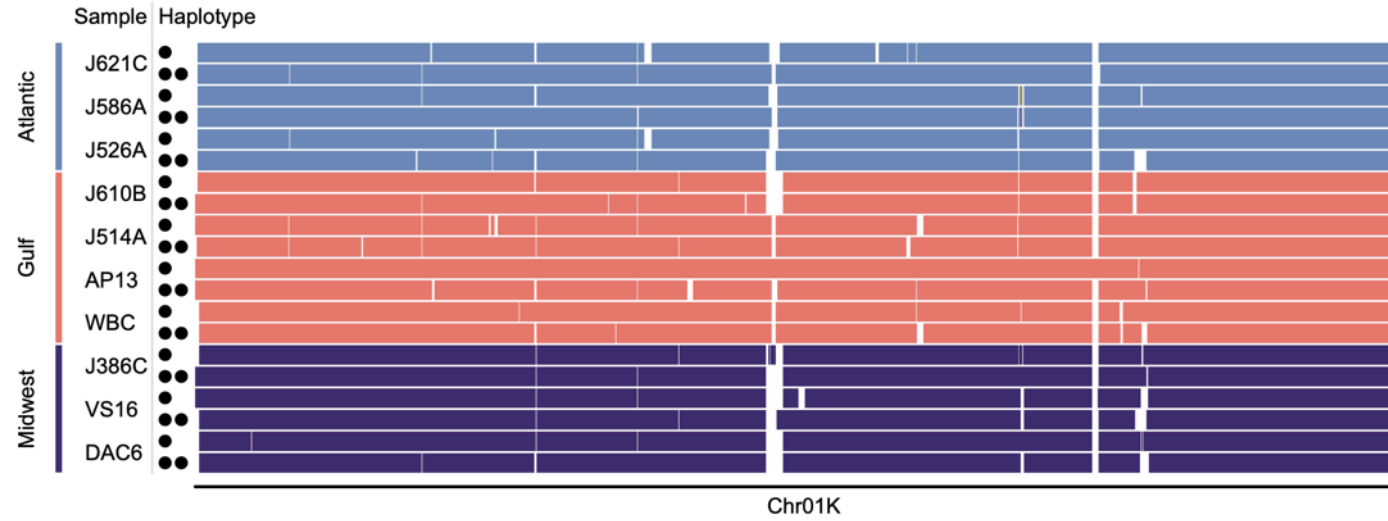
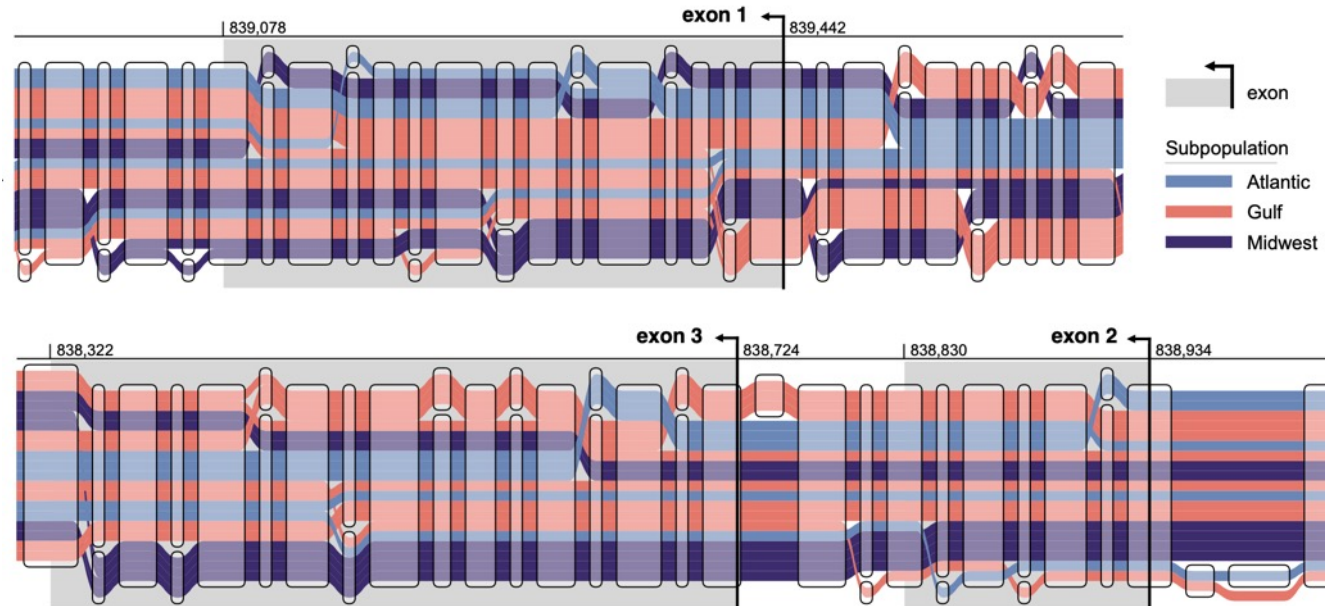


Fig. 3 | Test locus: variation in *ABO3* (AP13 HAP1 Chr08K:838322-839442)



Graph construction: minigraph-cactus
 Read alignment to graph: vg giraffe
 Viz via subway maps: sequenceTubeMap

Hickey, G., Monlong, J., Ebler, J. *et al.* Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nat Biotechnol* (2023).

On average,

89%

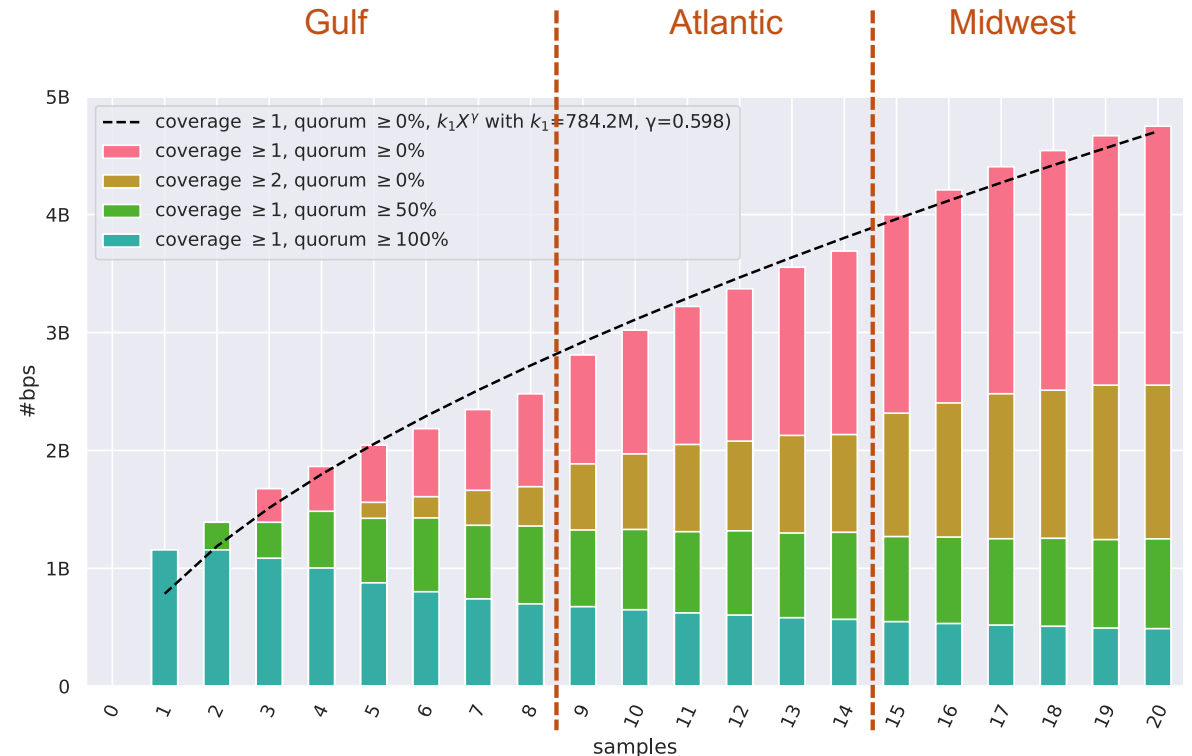
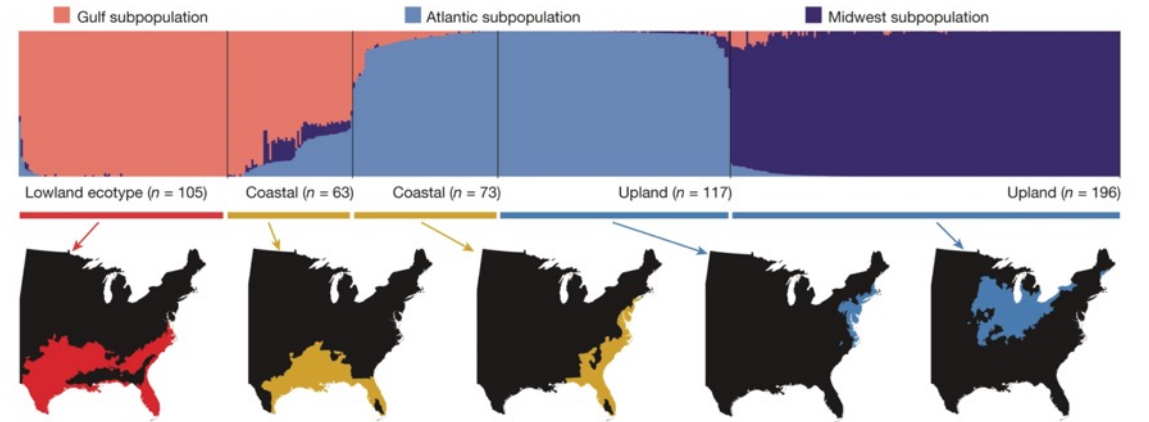
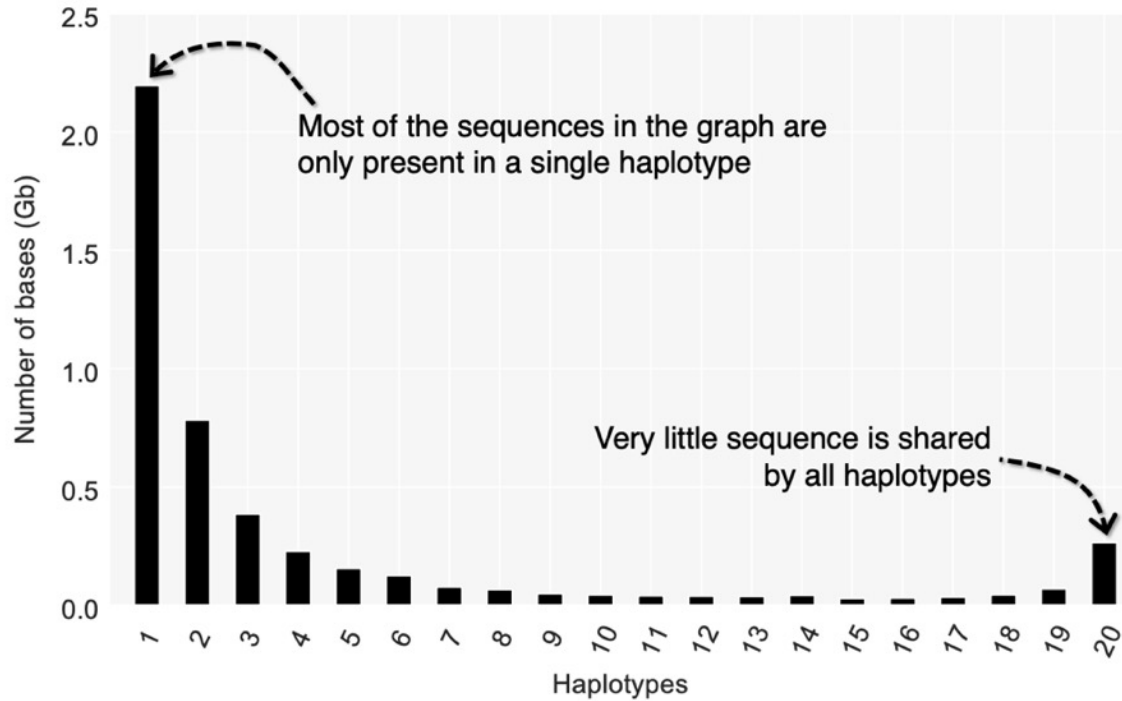
of each input reference genome is retained in the final graph

80%

of clipped sequences are annotated repeats

How much diversity do we cover (10 genomes, 20 haplotypes)?

Fig. 2 | Pangenome graph captures large amount of sequence diversity

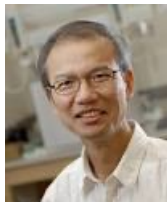


Pangenome graphs in simpler systems

Camelina (12)

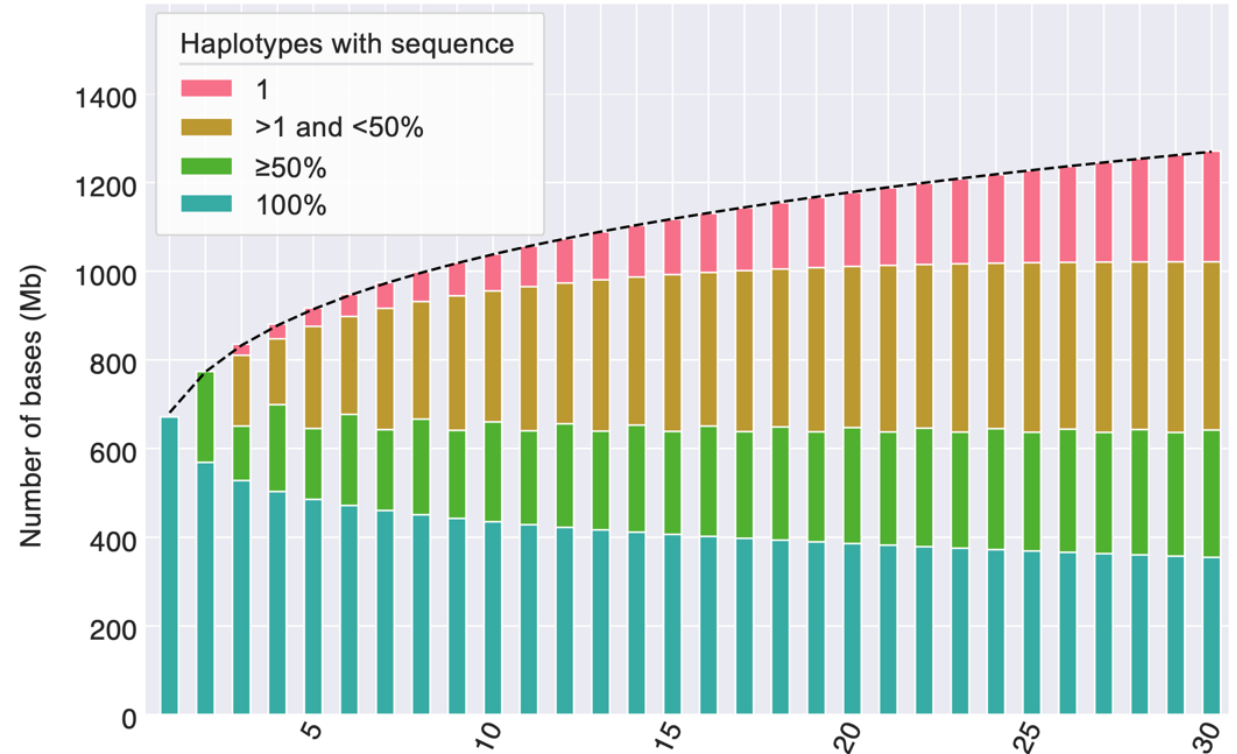


Daniel Kliebenstein
UC Davis
CSP2020 505836



Chaofu Lu
Montana State
BERSS 507613

Sorghum (30)



Todd Mockler
Donald Danforth
CSP2017 503104



John Mullet
Texas A&M



Geoff Morris
Colorado State



Nadia Shakoor
Donald Danforth

Example gene

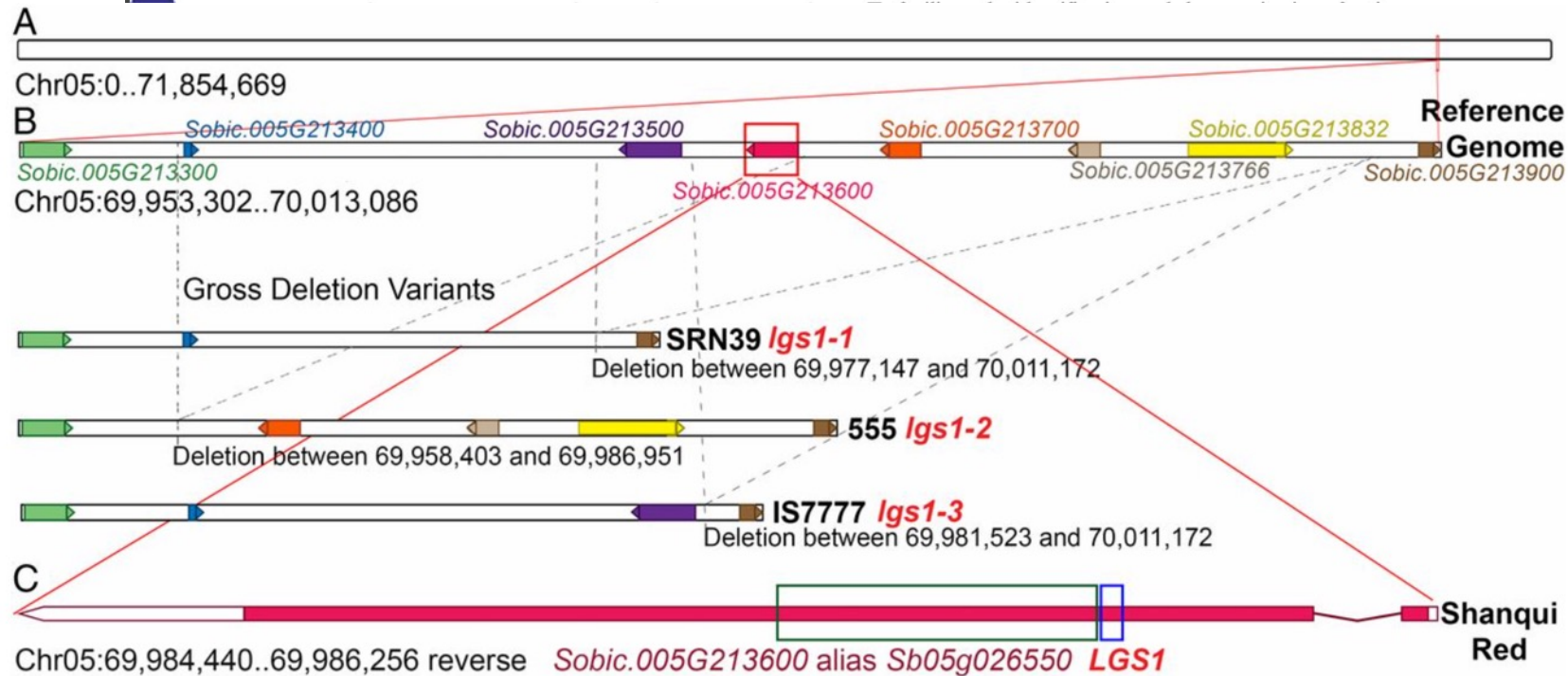
PNAS

Mutation in sorghum *LOW GERMINATION STIMULANT 1* alters strigolactones and causes *Striga* resistance

Daniel Gobena^a, Mahdere Shimels^b, Patrick J. Rich^a, Carolien Ruyter-Spira^b, Harro Bouwmeester^b, Satish Kanuganti^a, Tesfaye Mengiste^c, and Gebisa Ejeta^{a,1}

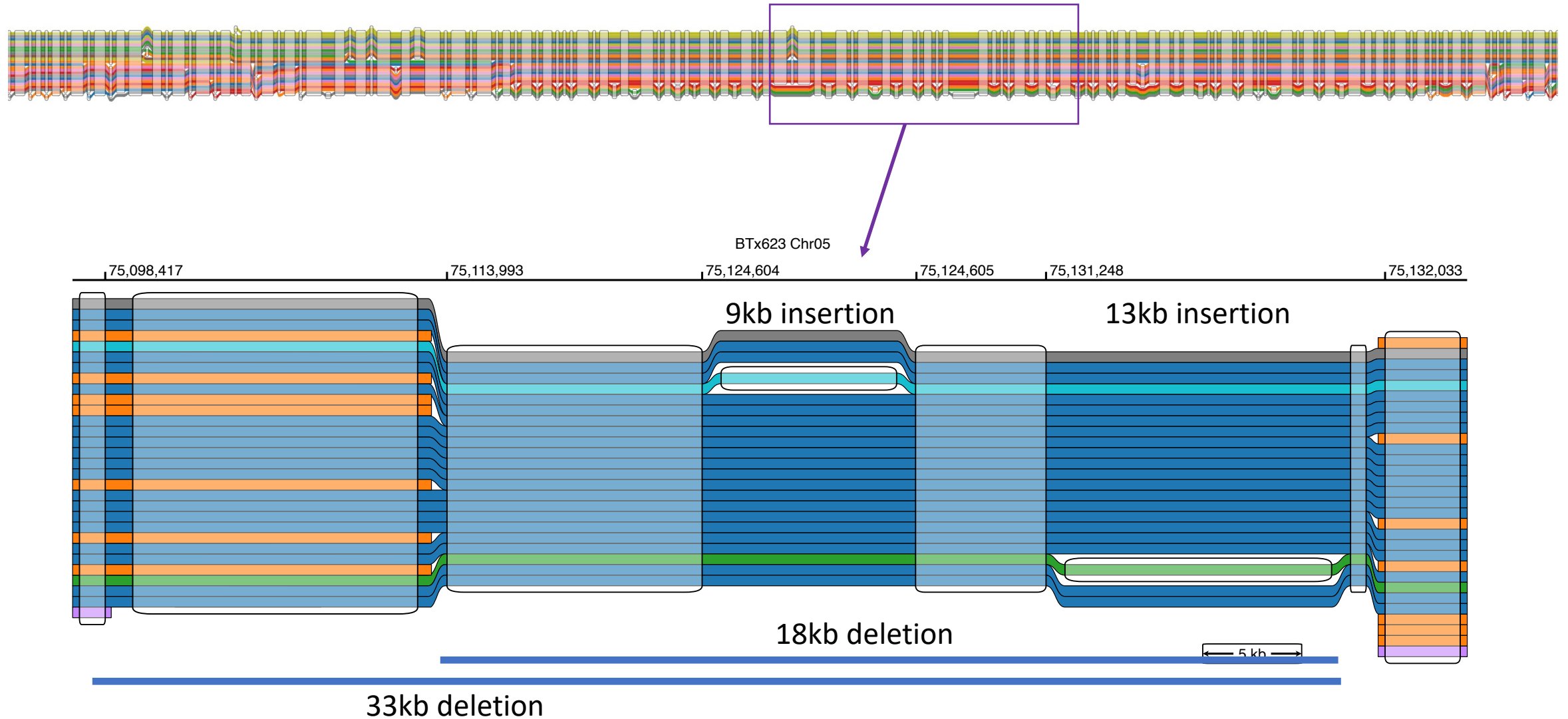
^aDepartment of Agronomy, Purdue University, West Lafayette, IN 47907; ^bLaboratory of Plant Physiology, Wageningen University, 6708 PB Wageningen, The Netherlands; and ^cDepartment of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907

Edited by Ronald L. Phillips, University of Minnesota, St. Paul, MN, and approved March 16, 2017 (received for review November 17, 2016)

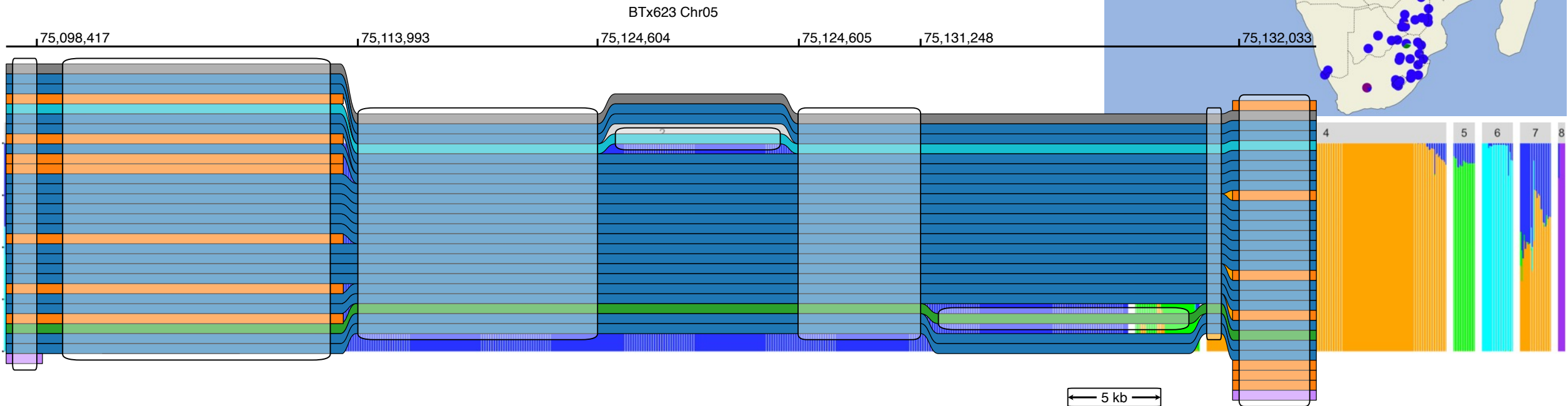
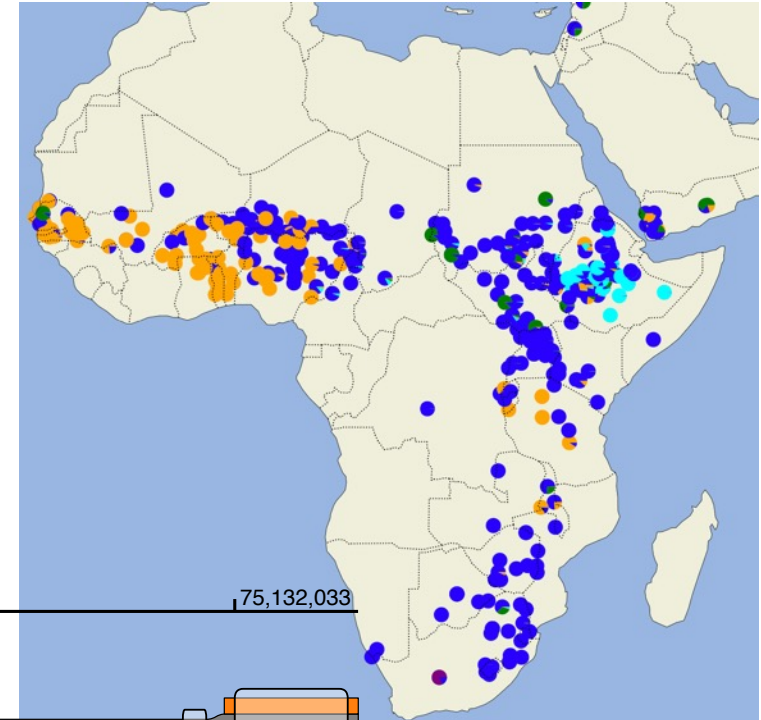
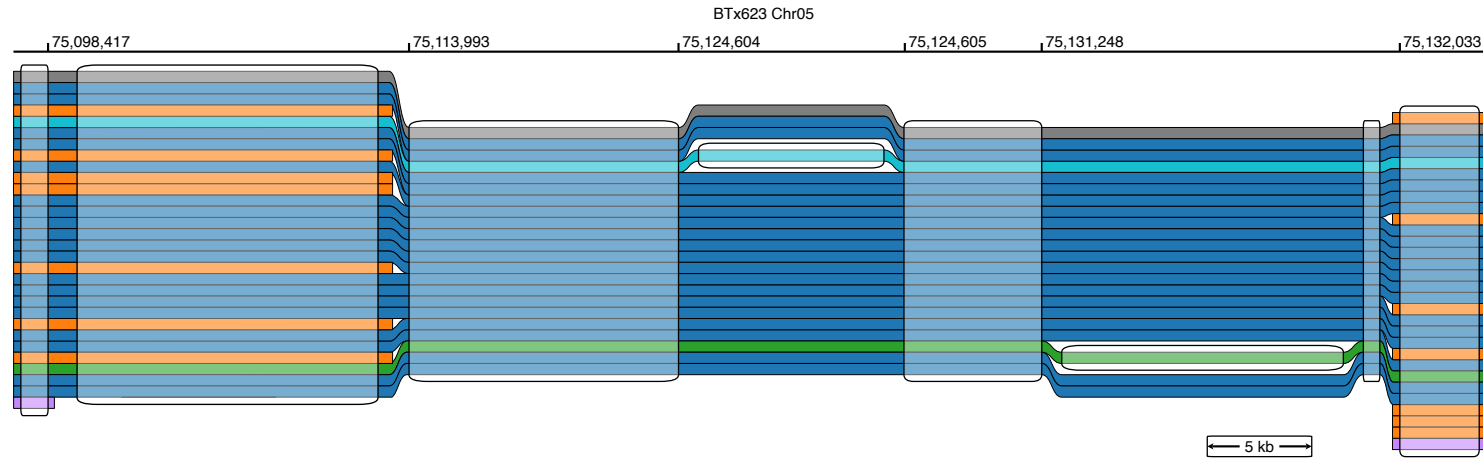


Striga – witchweed *

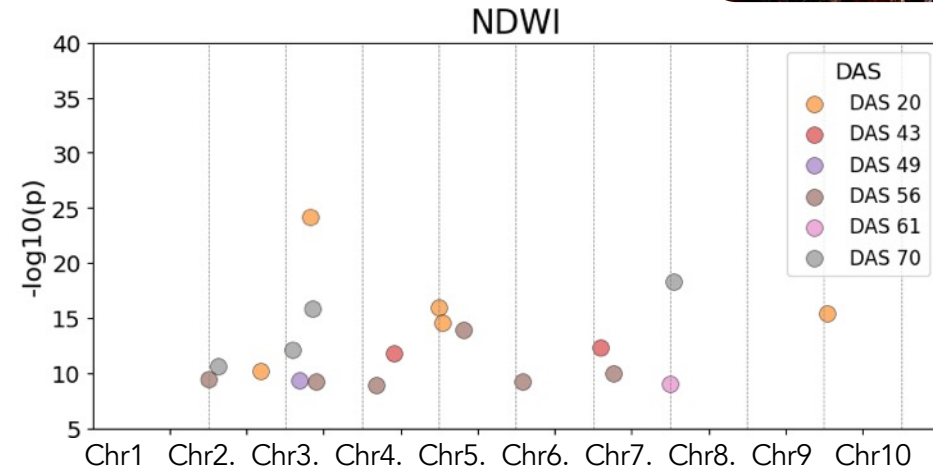
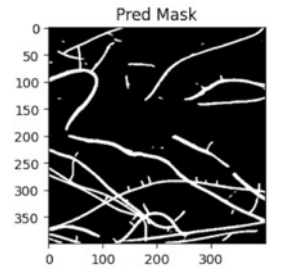
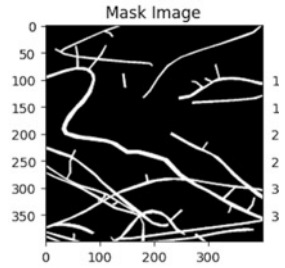
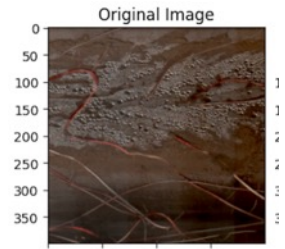
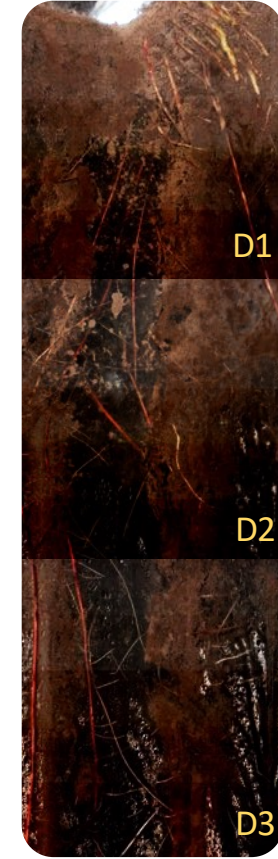
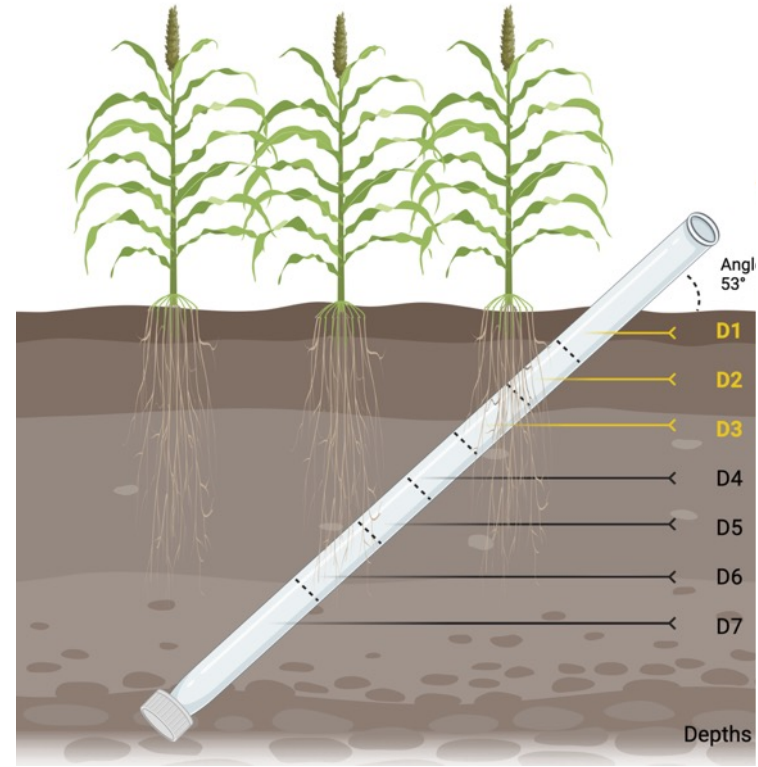
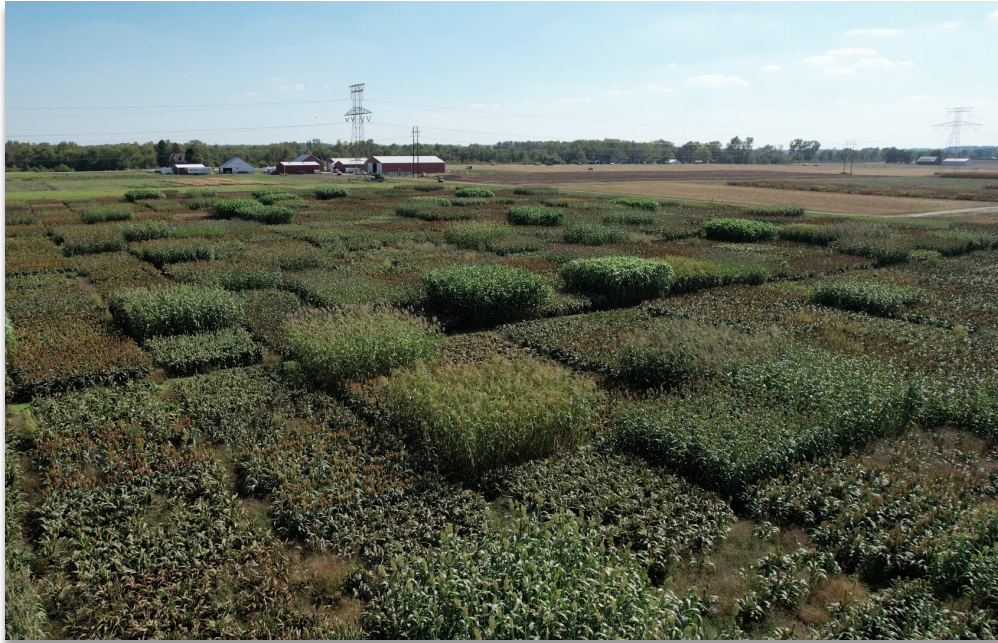
LGS1 in pangenome



Mapping LGS1 variants across Africa



Field phenotyping



Leaf Water Management



Nadia Shakoor
Donald Danforth

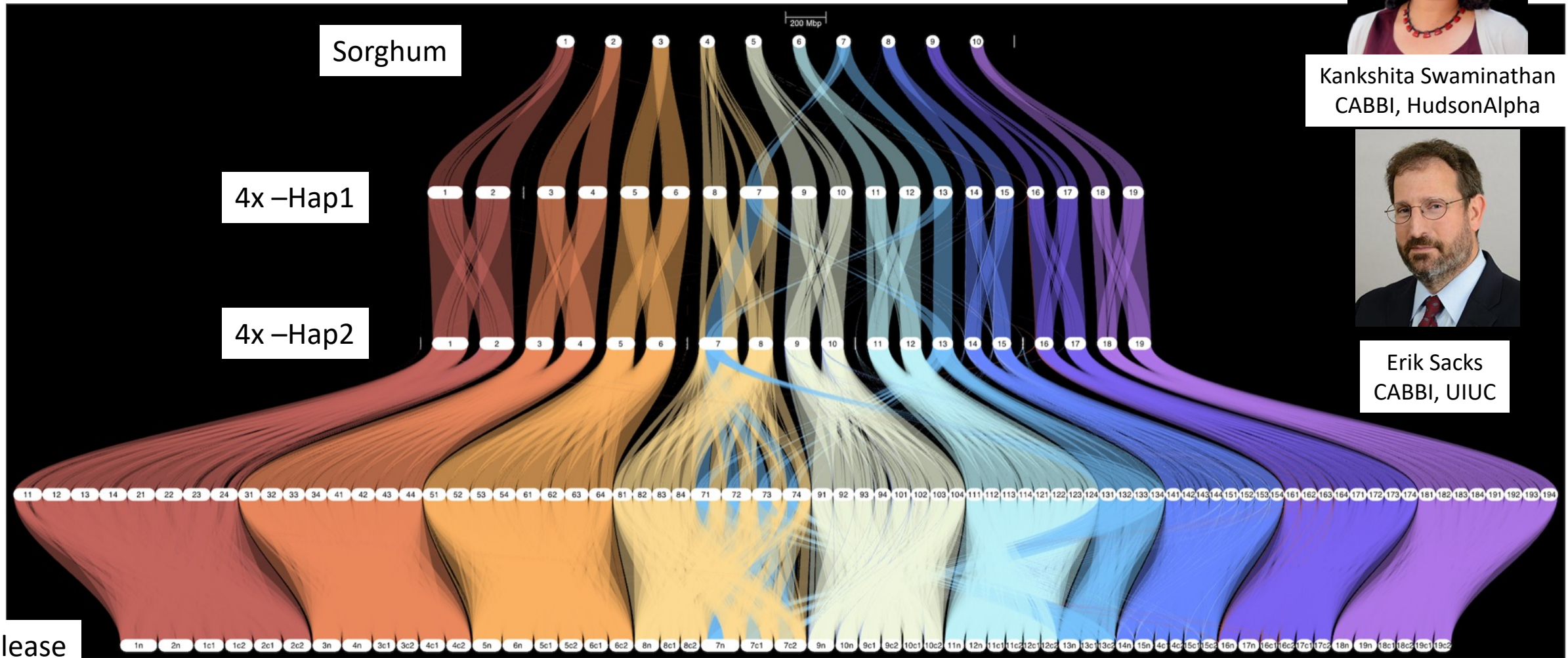
Miscanthus : embracing complexity



Kankshita Swaminathan
CABBI, HudsonAlpha



Erik Sacks
CABBI, UIUC



Chromosomes scaled by physical position

Goals of customization for perennial feedstocks



- Improve field traits: high biomass, improved germination & maturation, lodging resistance, leaf drop, flowering time, reduced fertilizer
- Improve breeding traits: shattering, self compatibility, sterility
- Improve quality/processing traits: fiber length, ash, chemical composition, fiber/lignan ratios, reduce internodes, low pith
- Make into chemical factories

Editing in complex plants



Kankshita Swaminathan



Tony Trieu



Team Swaminathan



wild-type edited line
M. sinensis PMS-014

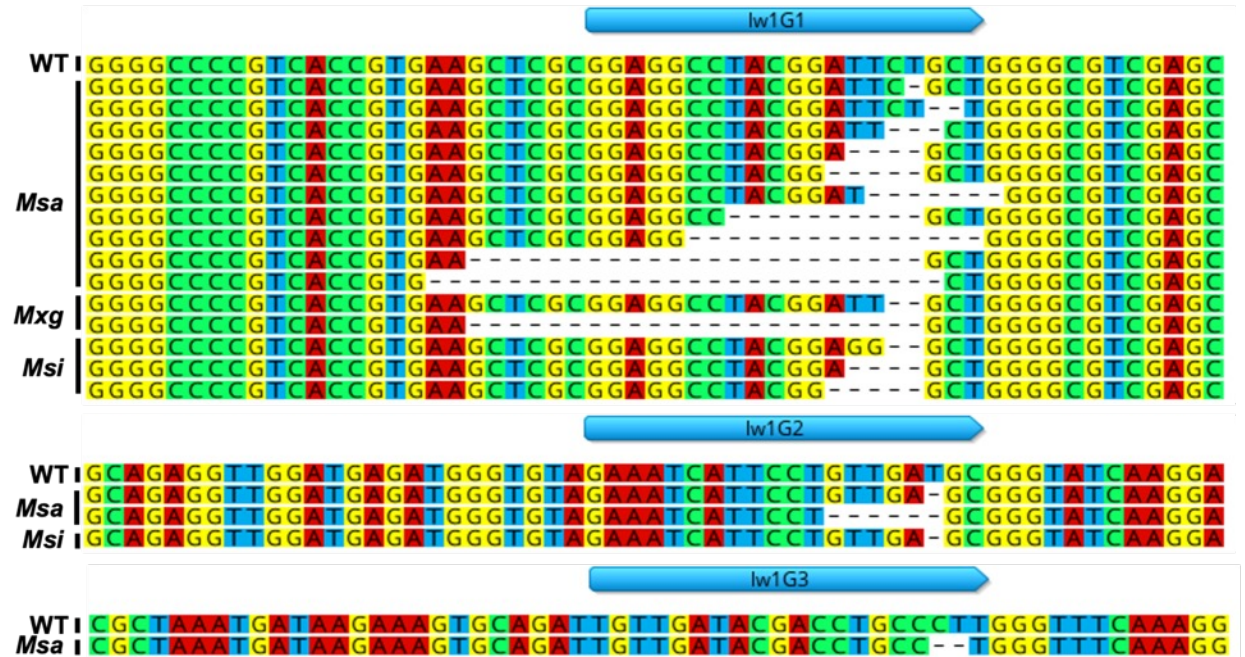


Pradeep Hirannaiah



Joshua Stanley
NSF-BRIDGES

Unpublished



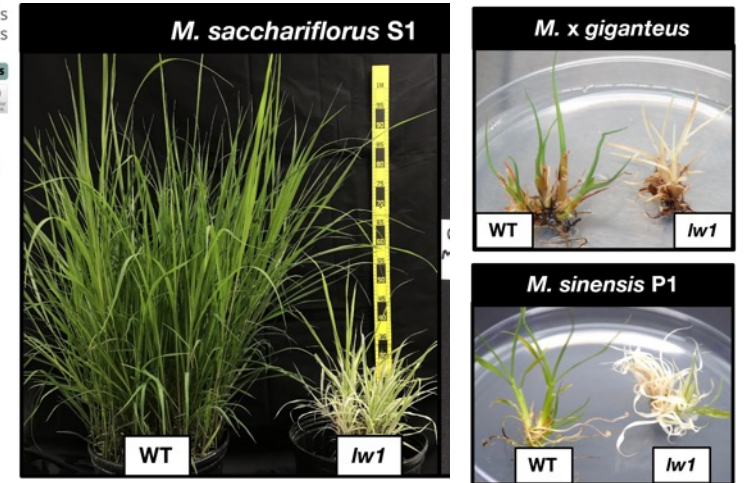
Trieu et al. *Biotechnology for Biofuels and Bioproducts* (2022) 15:148
<https://doi.org/10.1186/s13068-022-02241-8>

Biotechnology for Biofuels and Bioproducts

RESEARCH Open Access

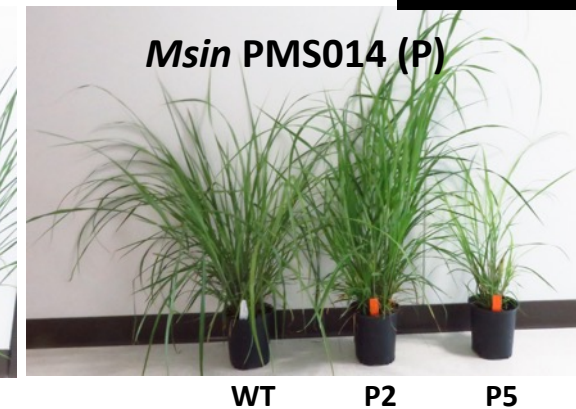
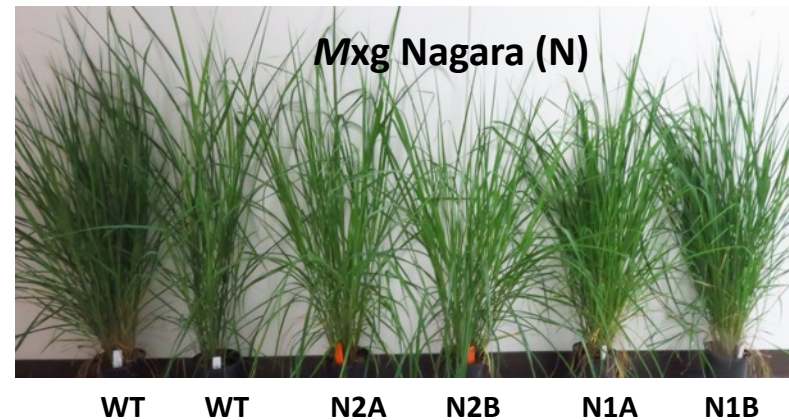
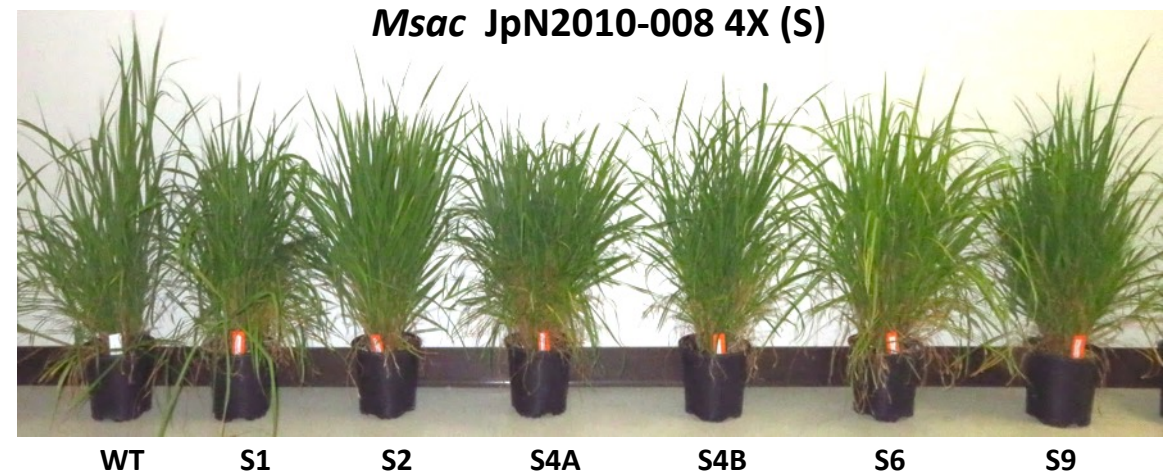
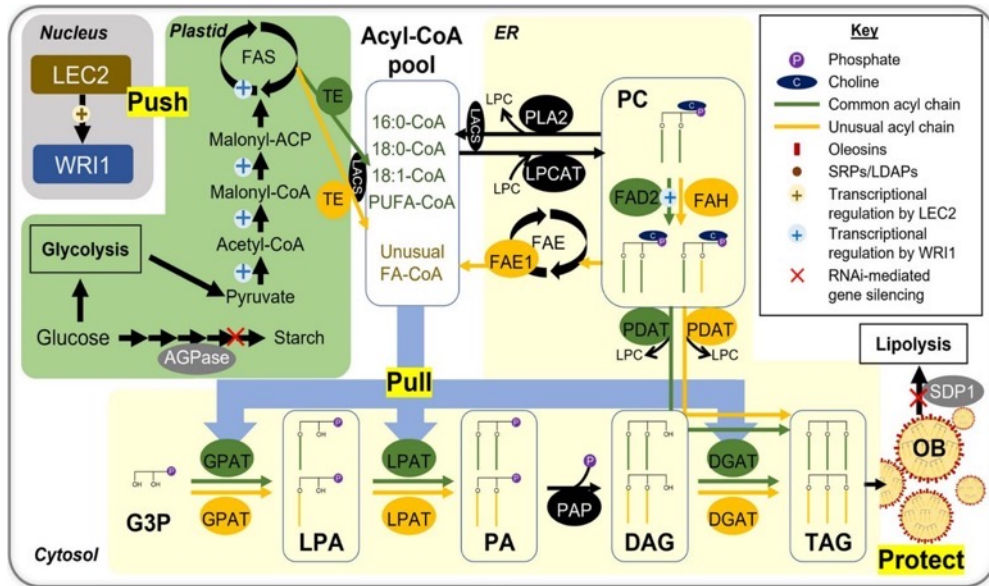
Transformation and gene editing in the bioenergy grass *Miscanthus*

Anthony Trieu^{1,2†}, Mohammad B. Belaffif^{3,4†}, Pradeepa Hirannaiah^{1,5}, Shilpa Manjunatha^{1,5}, Rebekah Wood^{1,5}, Yokshitha Bathula^{1,5}, Rebecca L. Billingsley^{2,5}, Anjali Arpan^{2,5}, Erik J. Sacks^{1,5}, Thomas E. Clemente^{4,5}, Stephen P. Moose^{1,5}, Nancy A. Reichert^{2,5†} and Kankshita Swaminathan^{1,2†}



In tetraploids and hexaploid production genotype

Engineering *Miscanthus* to produce vegetative lipids



“Push” - Photosynthetic carbon into fatty acid biosynthesis

- Wrinkled 1 transcription factor OE

“Pull” - Pull fatty acids into TAG

- DGAT diacylglycerol acyltransferase OE

“Protect” - Block TAG breakdown

- Oleosin OE, lipase suppression

Unpublished data



Our legacy – JGI impact on plant genomes

2007

Genome Name	Species	Release Date
Sorghum bicolor v5.1	sweet sorghum	Feb 1, 2023
Sorghum bicolor Wray v1.1	sweet sorghum Wray	Feb 1, 2023
Setaria viridis v4.1	green foxtail millet	Feb 1, 2023
Gossypium hirsutum Coker	upland cotton	Feb 1, 2023
Gossypium hirsutum DeltaPearl	upland cotton	Feb 1, 2023
Gossypium hirsutum FM958	upland cotton	Feb 1, 2023
Diphasiastrum complanatum v3.1		Sep 29, 2022

Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, provides JGI users and the broader plant science community a hub for accessing, visualizing and analyzing JGI-sequenced plant genomes, as well as selected genomes and datasets that have been sequenced elsewhere. By integrating this large collection of plant genomes into a single resource and performing comprehensive and uniform annotation and analyses, Phytozome facilitates accurate and insightful comparative genomics studies.

As of release v13, Phytozome hosts 411 assembled and annotated genomes, from 174 Archaeplastida species, and contains the 54 Brachypodium distachyon lines from the **BrachyPan** pan-genome study, the 20 species included in the **Brassicales Man Alignment Project** and 8 cowpea (*Vigna unguiculata*) genomes

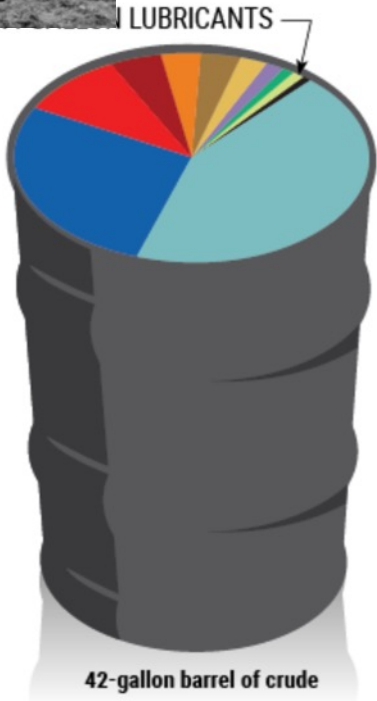
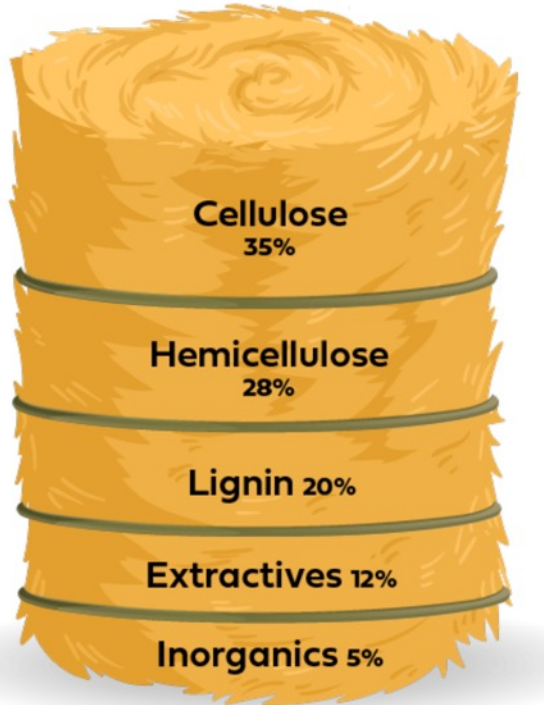
2009

1. Choose genomes by selecting from tree or type genus/species/commo 1 genomes selected

2. find genes by keyword search by BLAST get standard data files build custom o

- Panicum virgatum var. DAC6 HAP2 v1.1 (i) RST JGI
- Panicum virgatum var. VS16 HAP1 v1.1 (i) RST JGI
- Panicum virgatum var. VS16 HAP2 v1.1 (i) RST JGI
- Panicum virgatum var. WBC HAP1 v1.1 (i) RST JGI
- Panicum virgatum var. WBC HAP2 v1.1 (i) RST JGI
- Paspalum vaginatum v3.1 (i) UNRST JGI
- Saccharum officinarum x spontaneum R570 v2.1 (i) RST JGI
- Setaria italica v2.2 (i) UNRST GeneAtlas JGI
- Setaria viridis v1.1 (i) UNRST GeneAtlas JGI
- Setaria viridis v2.1 (i) UNRST GeneAtlas JGI
- Setaria viridis v4.1 (i) RST JGI
- Urochloa fusca v1.1 (i) RST JGI
- Maize
 - Zea mays RefGen_V4 (i) UNRST EXT
 - Zea mays Zm-B73-REFERENCE-NAM-5.0.55 (i) UNRST EXT
 - Zea mays B84 v1.2 (i) UNRST JGI
 - Zea mays LH145 v1.2 (i) UNRST JGI
 - Zea mays NKH8431 v1.2 (i) UNRST JGI
 - Zea mays PH207 v1.1 (i) UNRST EXT
 - Zea mays PHB47 v1.1 (i) UNRST JGI
 - Zea mays PHB47 v1.2 (i) UNRST JGI

2024



- 19.4 gallons gasoline
- 12.5 gallons diesel/home heating oil
- 4.4 gallons kerosene-type jet fuel
- 2.1 gallons coke
- 1.6 gallons still gas
- 1.5 gallons liquefied refinery gases
- 0.9 gallons heavy, residual fuel oils
- 0.8 gallon asphalt and road oil
- 0.5 gallon petrochemical feedstocks
- 0.4 gallon lubricants
- 0.4 gallon other

These are only examples of discoveries in plant genomics- 1000's of pathways and variant for traits of interest are needed to customize and build genotypes for end uses.

A Unified Data Infrastructure for Biological and Environmental Research

Report from the BER Advisory Committee

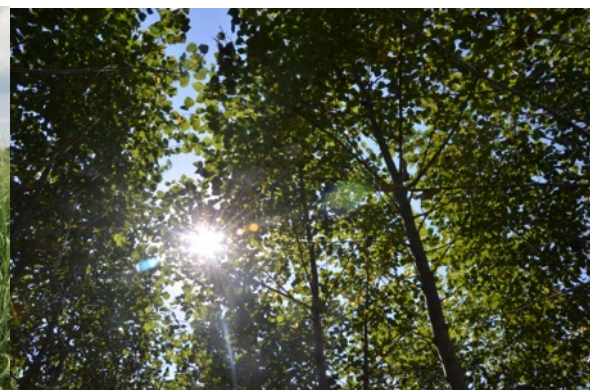


Overcoming Barriers in Plant Transformation

A Focus on Bioenergy Crops



BER is uniquely positioned to accelerate these effort with a combination of climate modeling for global biogeochemistry, BRCs for deconstruction/customization of feedstocks, and the combined efforts of JGI, national and academic labs to acquire and develop the biological, environmental knowledge to make this possible.





U.S. DEPARTMENT OF
ENERGY

Office of
Science

Biological and Environmental Research

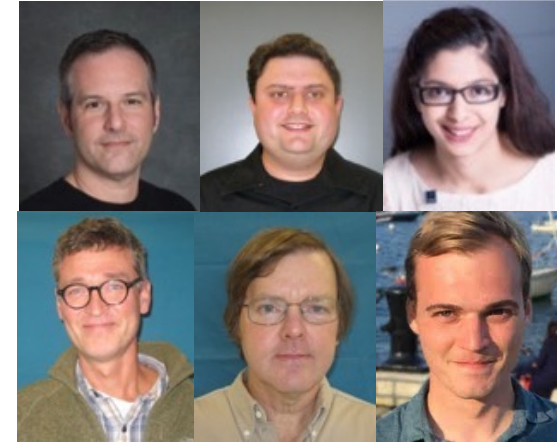
Contributors from the JGI



HudsonAlpha Experimental Group
Jane Grimwood & Mellissa Williams



HudsonAlpha Computational Group
John Lovell & Jerry Jenkins



Phytozome & Annotation Group
David Goodstein



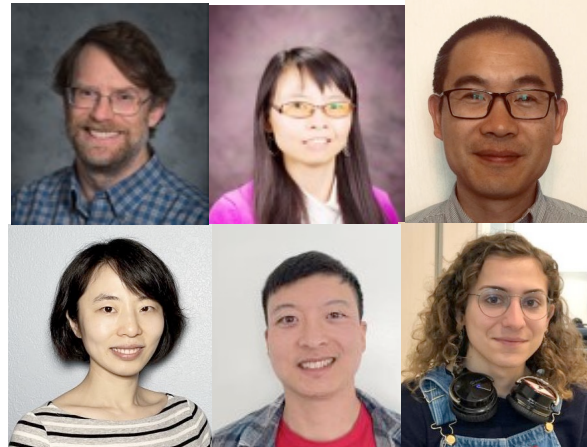
PMO
Kerrie Barry



Computational
Genomics
Dan Rokhsar



JGI Production Groups
Yuko Yoshinaga & Chris Daum



Plant Functional Genomics
John Vogel



Arizona Genomic Partner Group
Jayson Talag & Dario Copetti

Thank you to our users and partners

