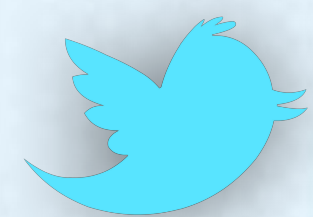
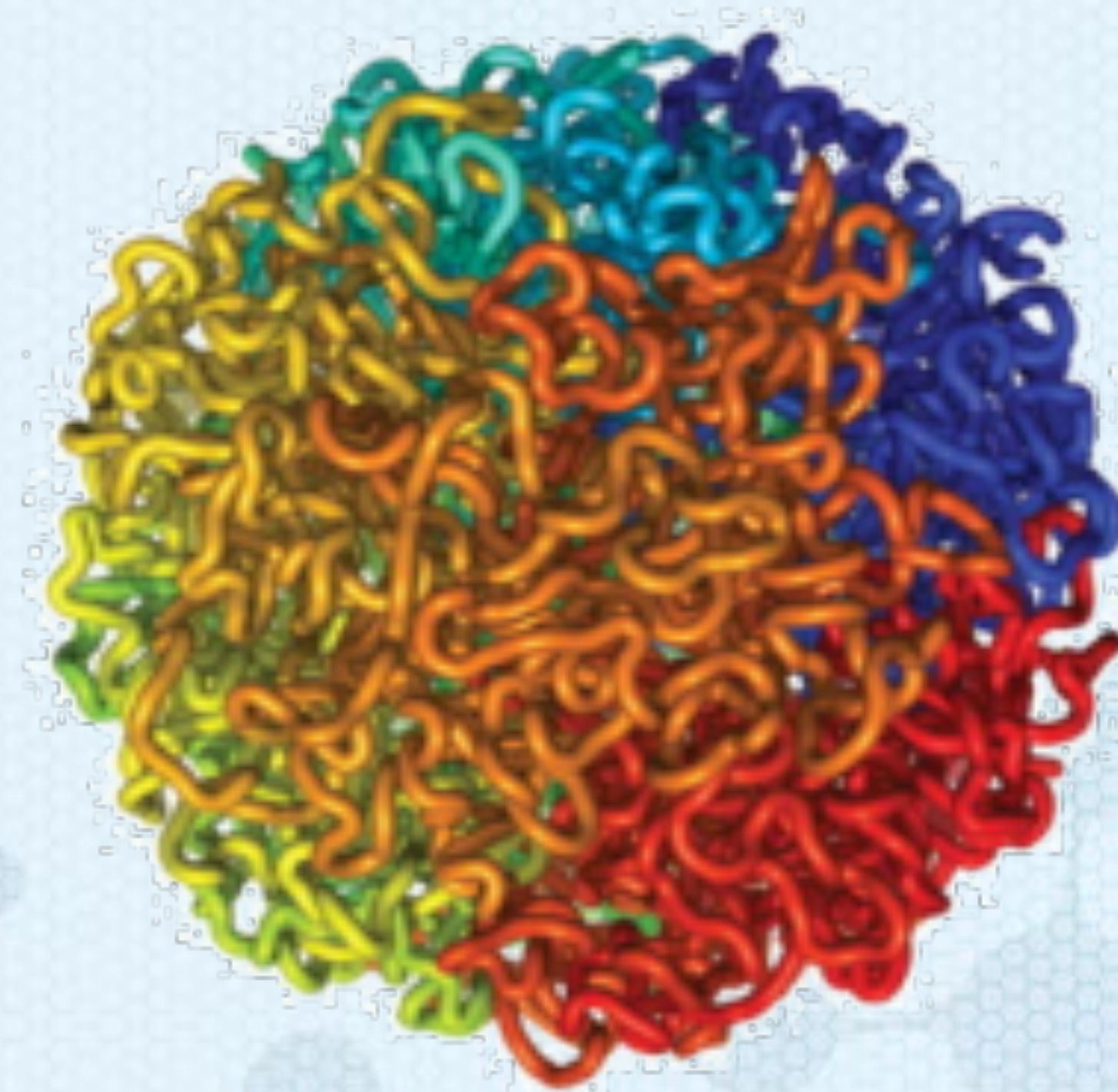


Chromosome-scale Genome Scaffolding using Hi-C Data

Zev Kronenberg Ph.D.



@PhaseGenomics

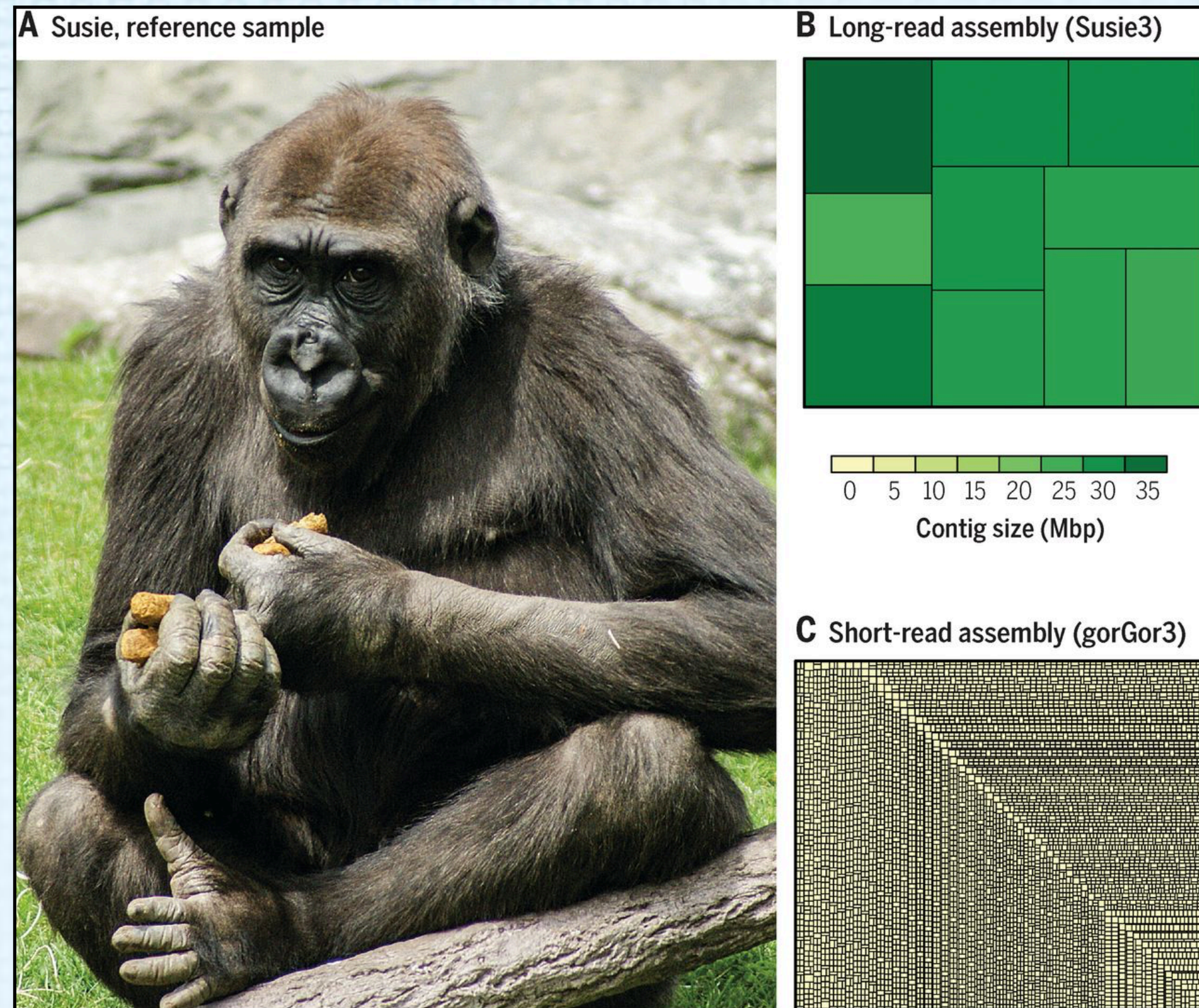
@Zevkronenberg

info@phasegenomics.com

Lieberman-Aiden, *et. al.* Science, 2009

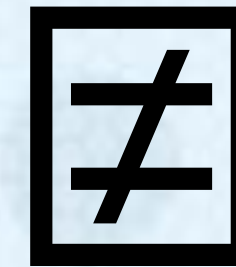
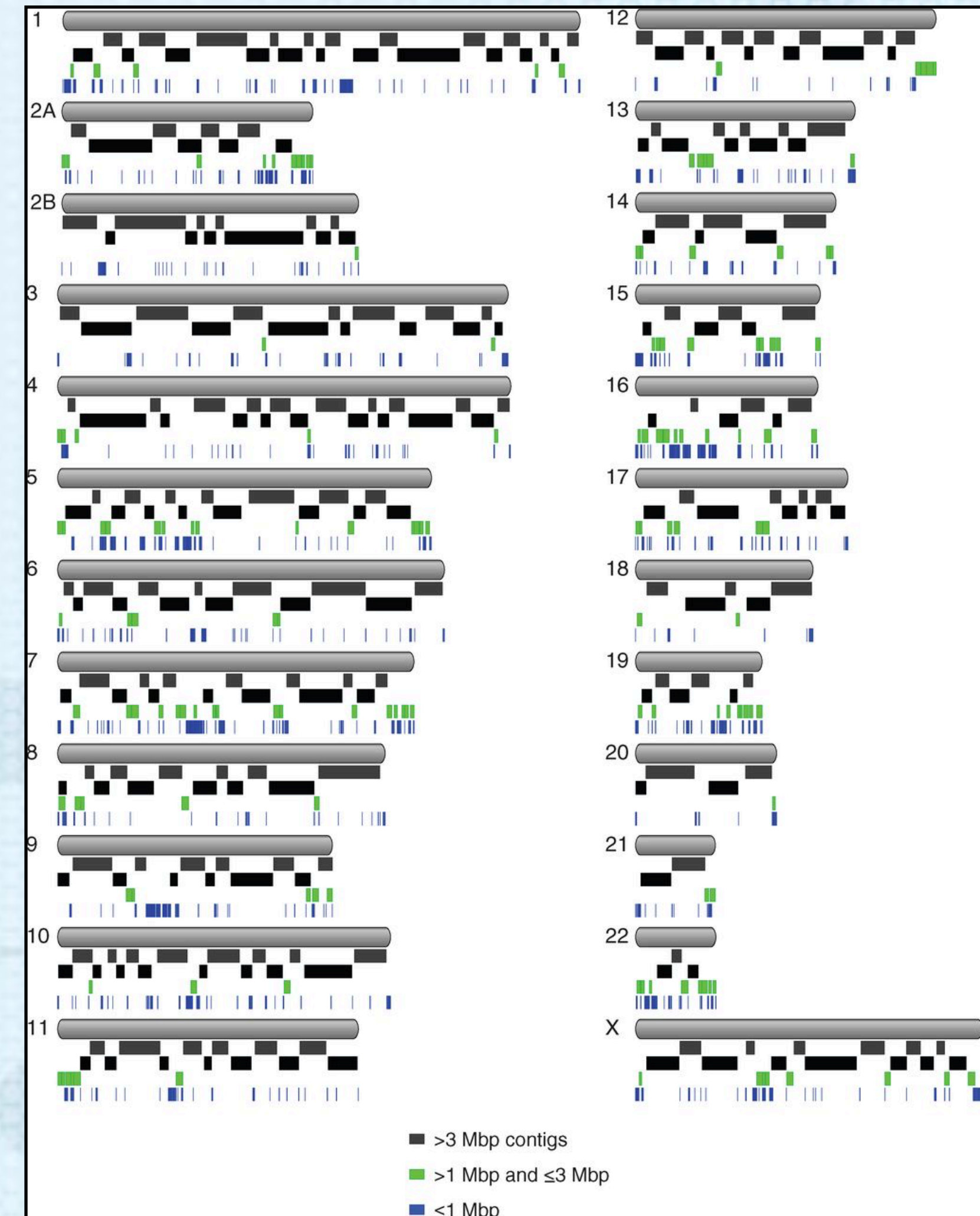
Genome assembly does not yield chromosomes

Contigs



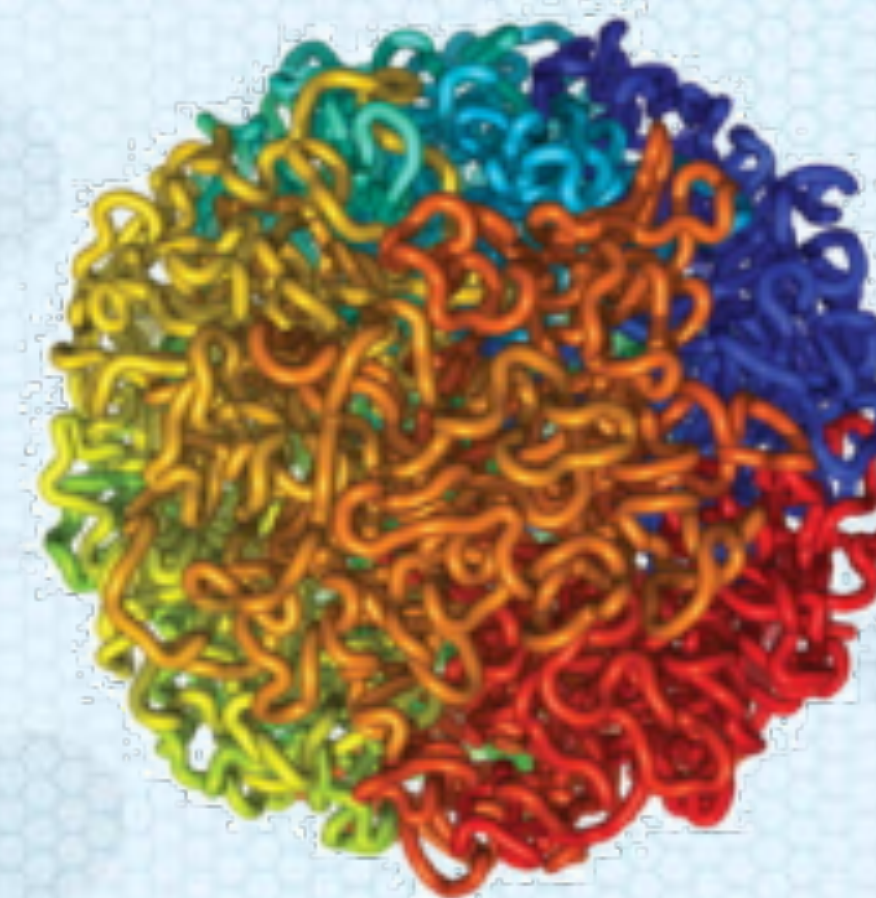
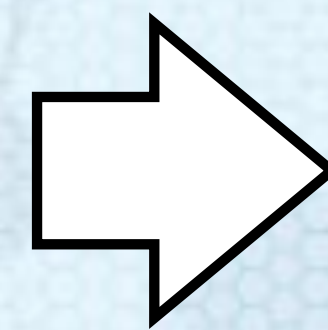
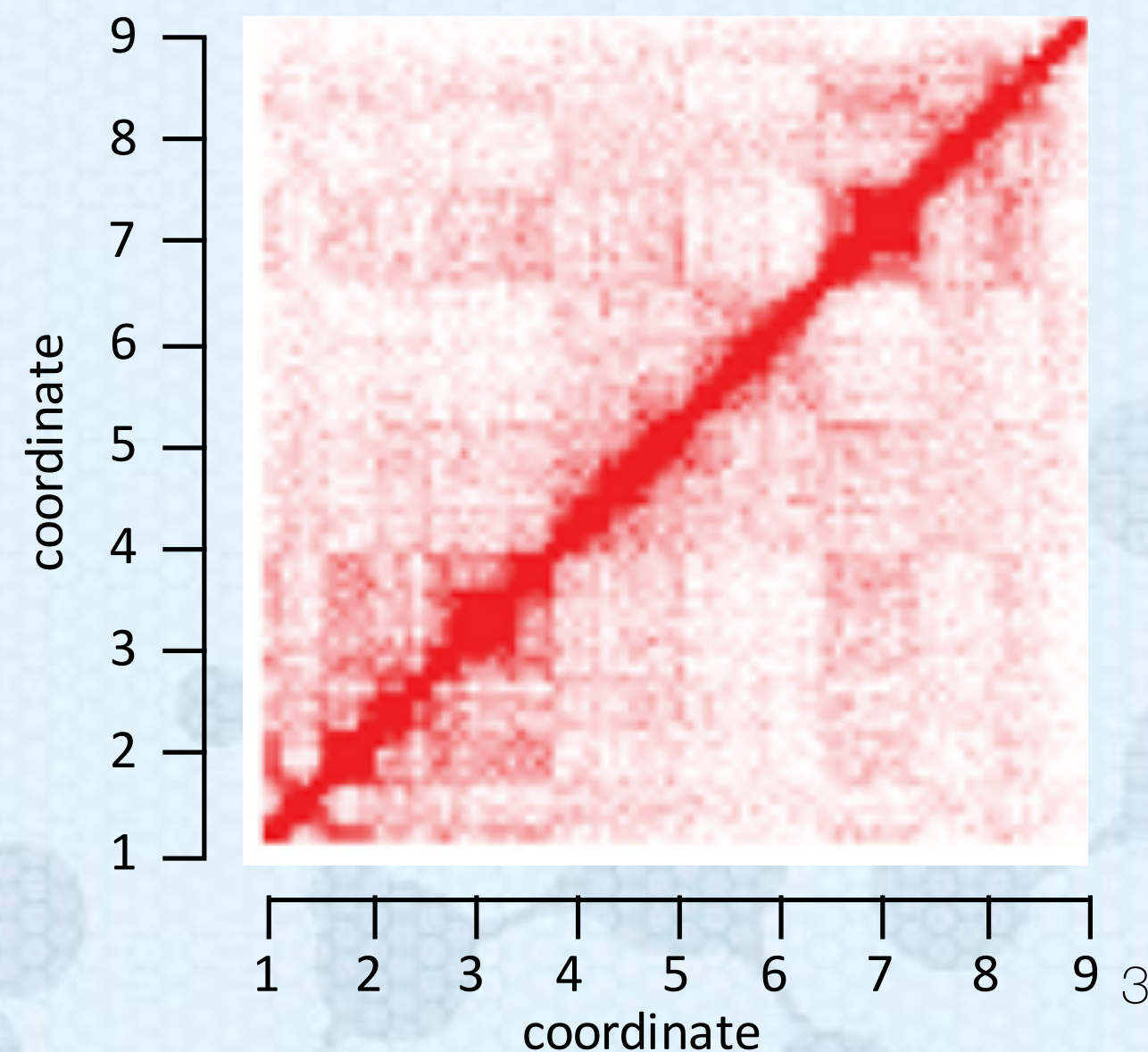
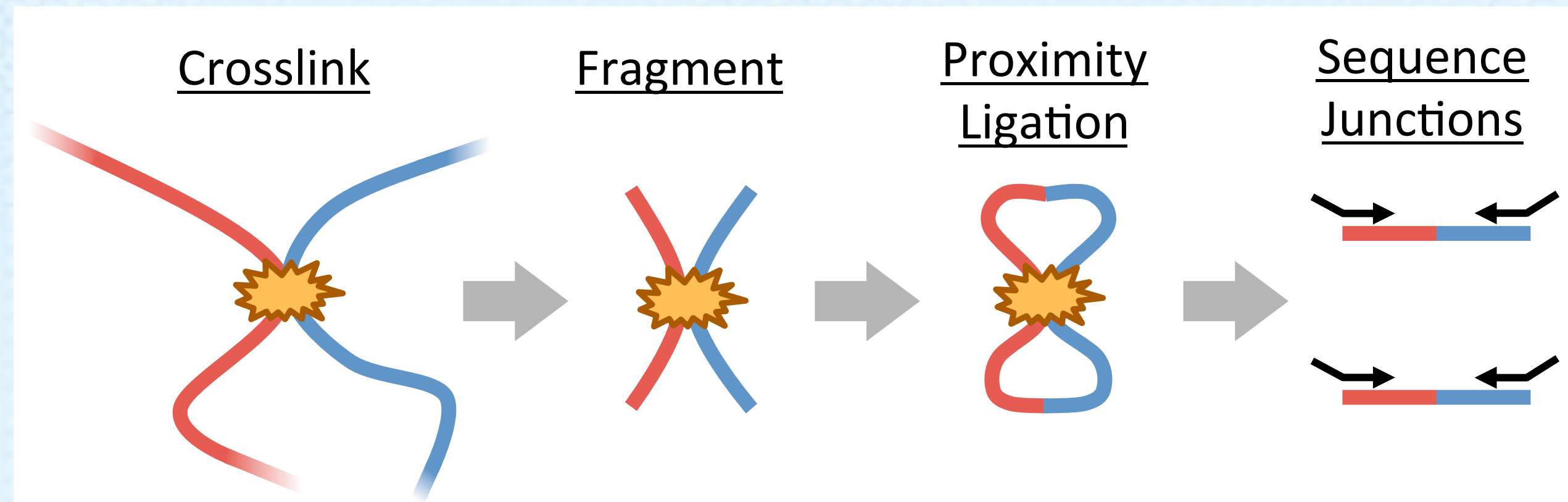
Gordon et al. Science 2016

Chromosomes



How do we go from contigs from chromosomes?

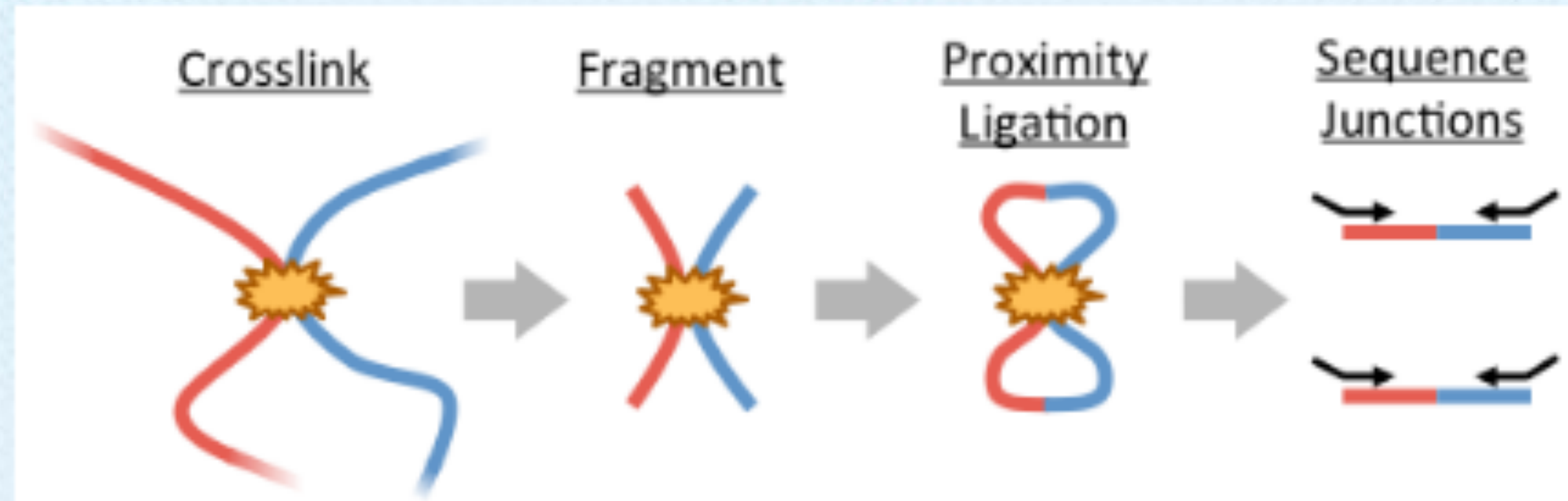
Hi-C captures the 3D structure of Chromosomes



Lieberman-Aiden, *et. al.* Science, 2009

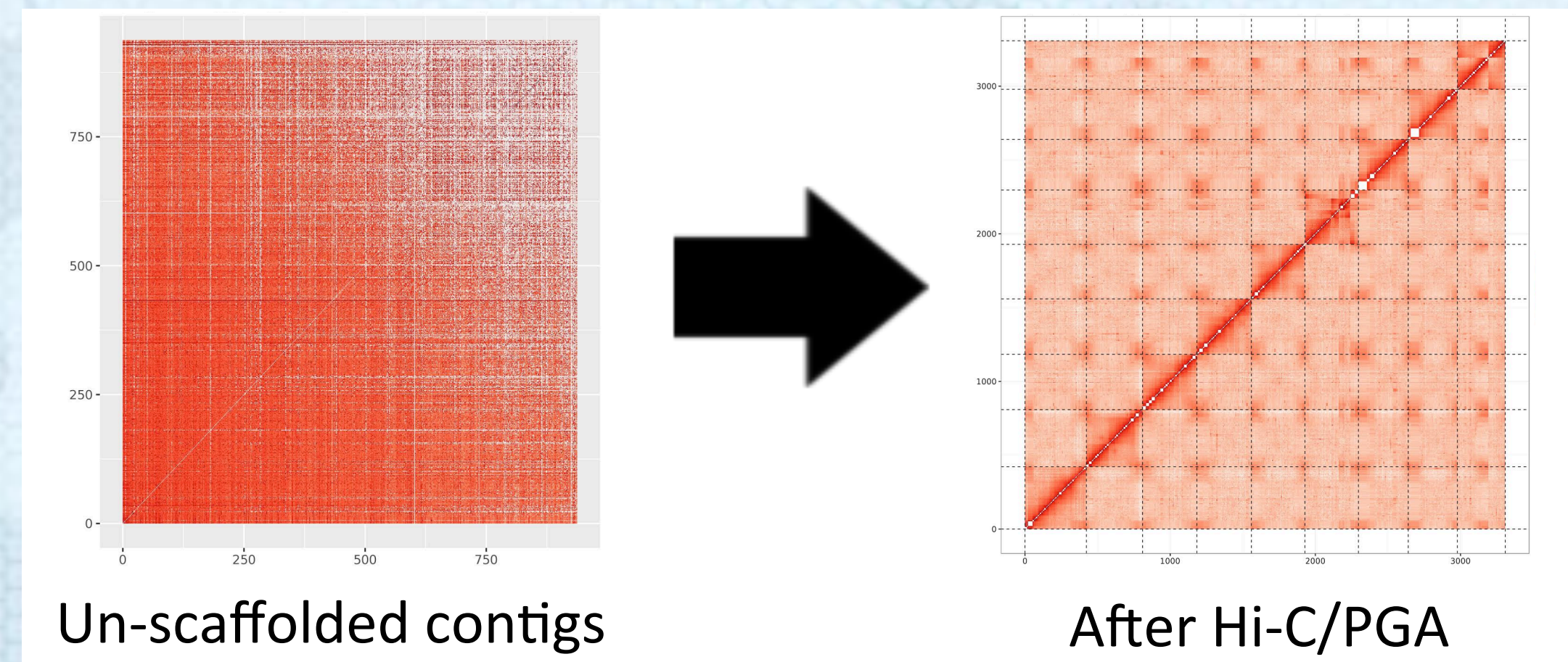
How do we go from contigs from chromosomes?

Using Hi-C to assemble chromosome-level scaffolds

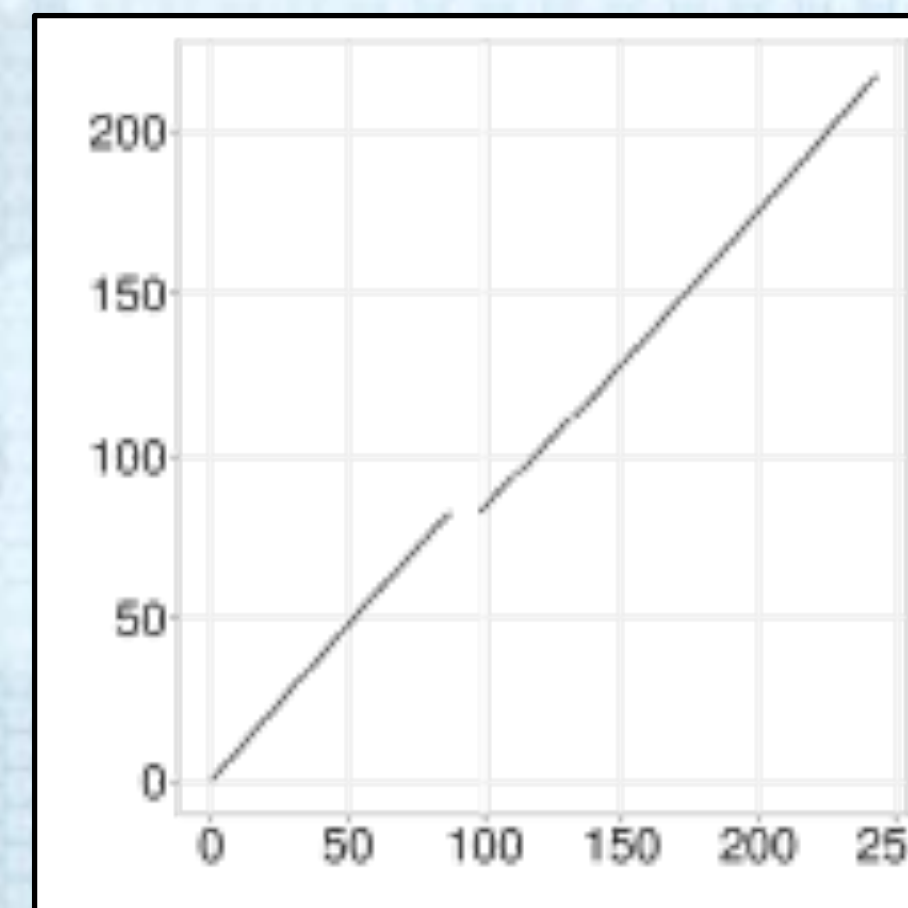
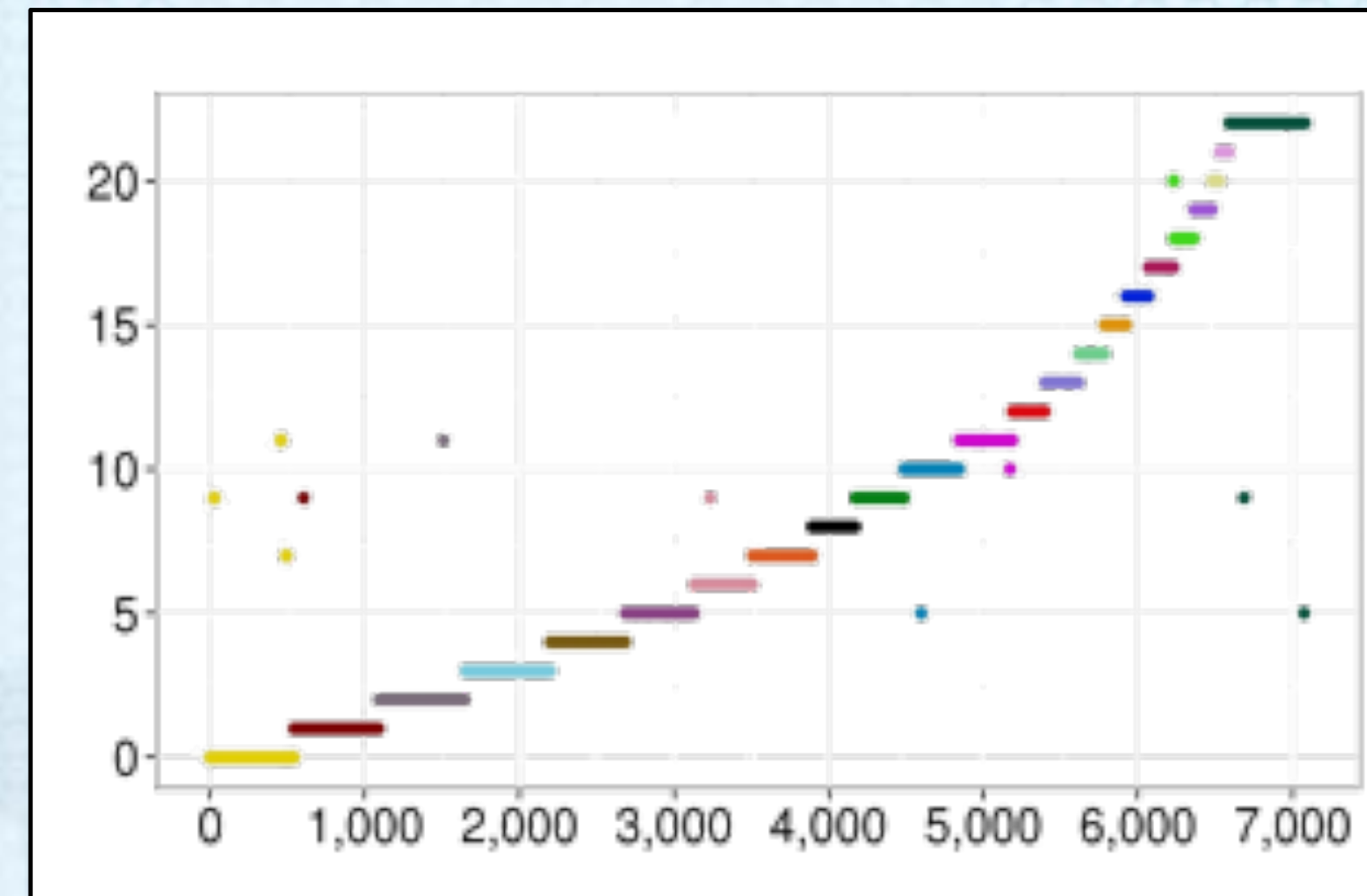
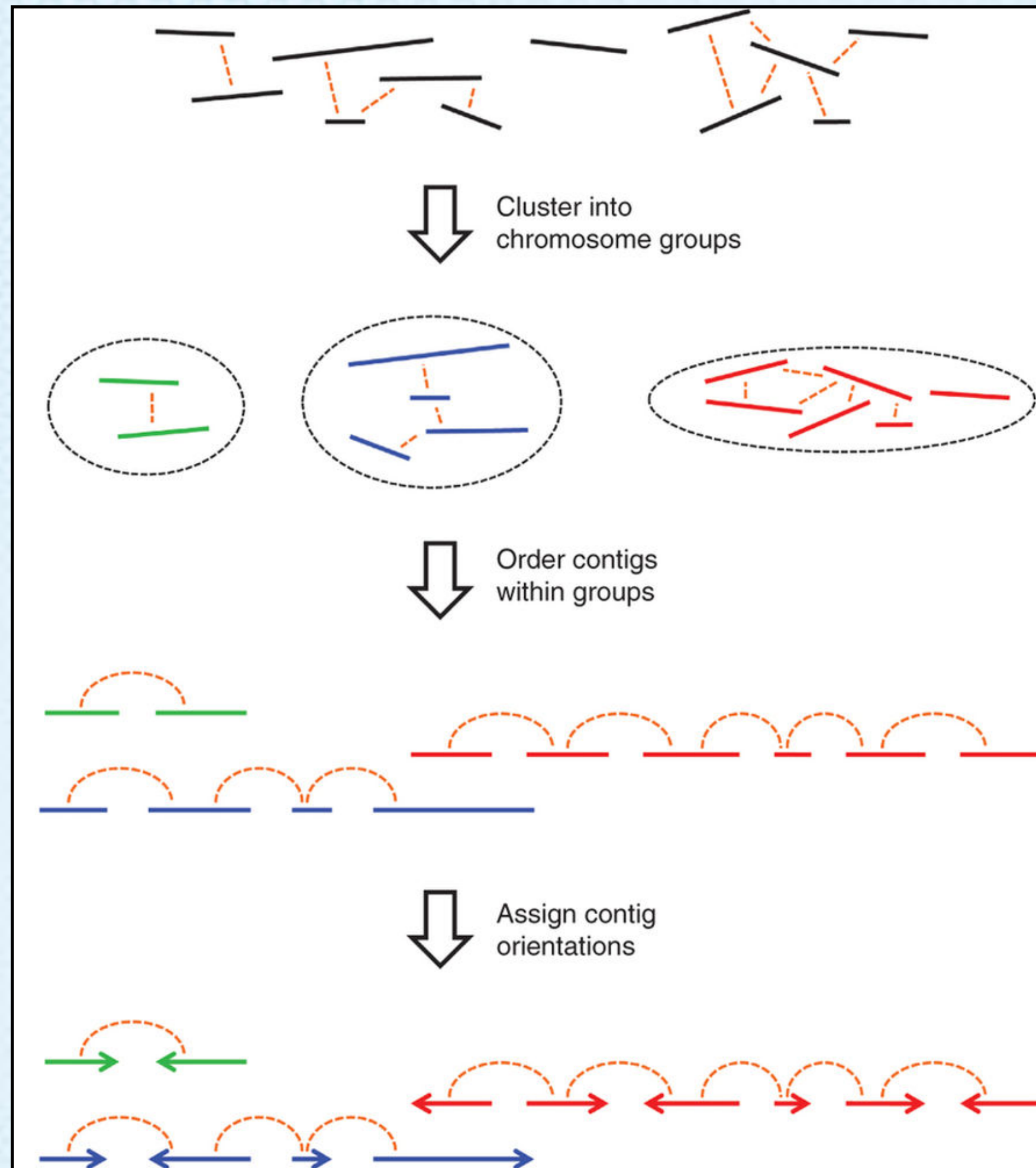


Proximity-Guided Assembly™ :

- Clustering contigs into chromosome groups
- Ordering and orienting the contigs in scaffolds



Proximity-guided assembly is a powerful scaffolding method



Proximo generates high-quality chromosome-level scaffolds

PacBio	+BioNano	+Proximo Hi-C
2.6 Gb total	2.7 Gb total	2.7 Gb total
3110 contigs	1575 scaffolds	31 clusters/chroms
4.7 Mb N50	22.3 Mb N50	91.7 Mb N50
35.6 Mb max	76.3 Mb max	157.3 Mb max
* after Proximo, used PBJelly to close 840 gaps		

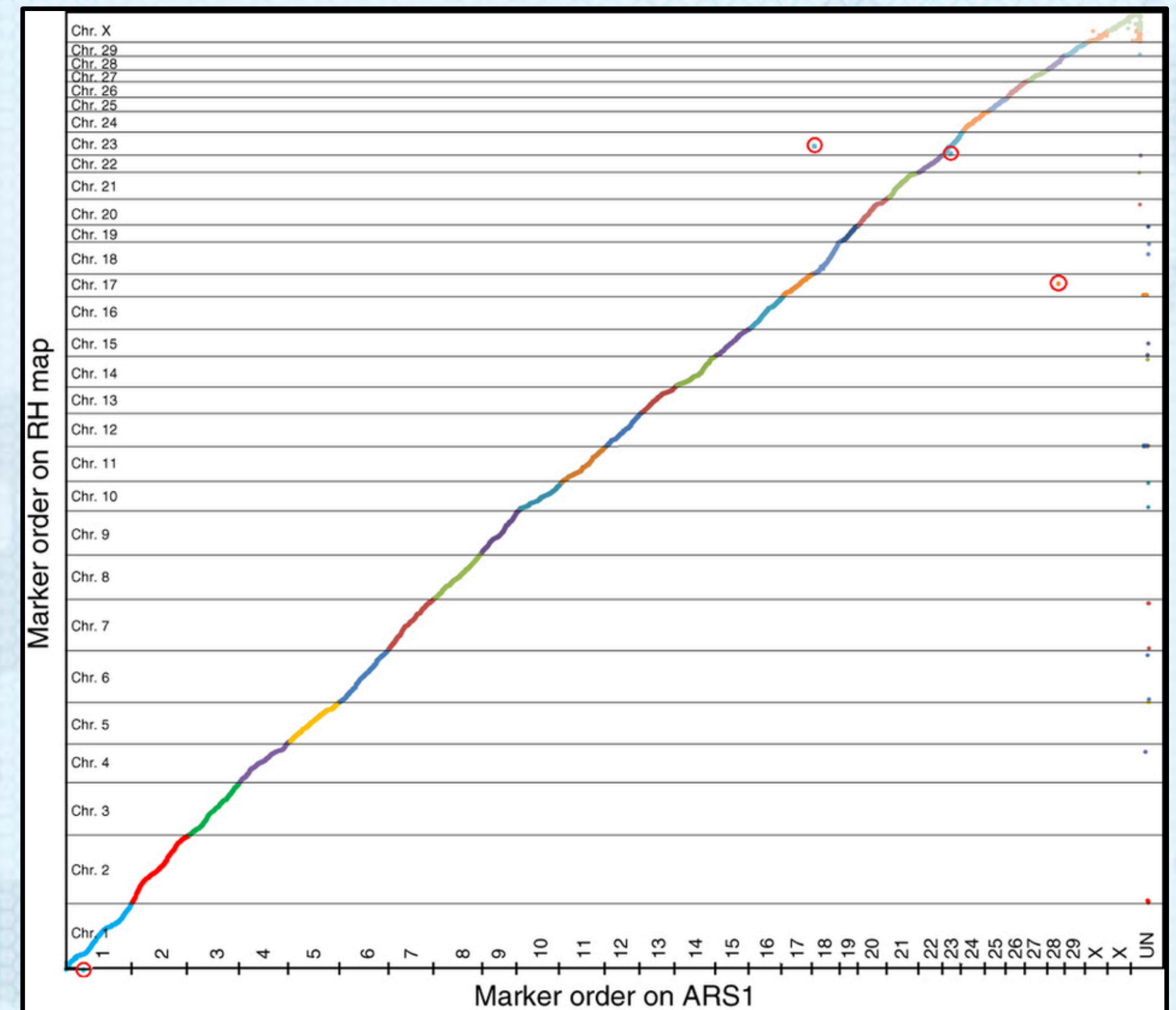
Technical Report | [OPEN](#)

Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome

Derek M Bickhart, Benjamin D Rosen [...] Timothy P L Smith ✉

Nature Genetics **49**, 643–650 (2017)

Received: 02 August 2016



Proximo generates high-quality chromosome-level scaffolds

PacBio	+BioNano	+Proximo Hi-C
2.6 Gb total	2.7 Gb total	2.7 Gb total
3110 contigs	1575 scaffolds	31 clusters/chroms
4.7 Mb N50	22.3 Mb N50	91.7 Mb N50
35.6 Mb max	76.3 Mb max	157.3 Mb max
* after Proximo, used PBJelly to close 840 gaps		

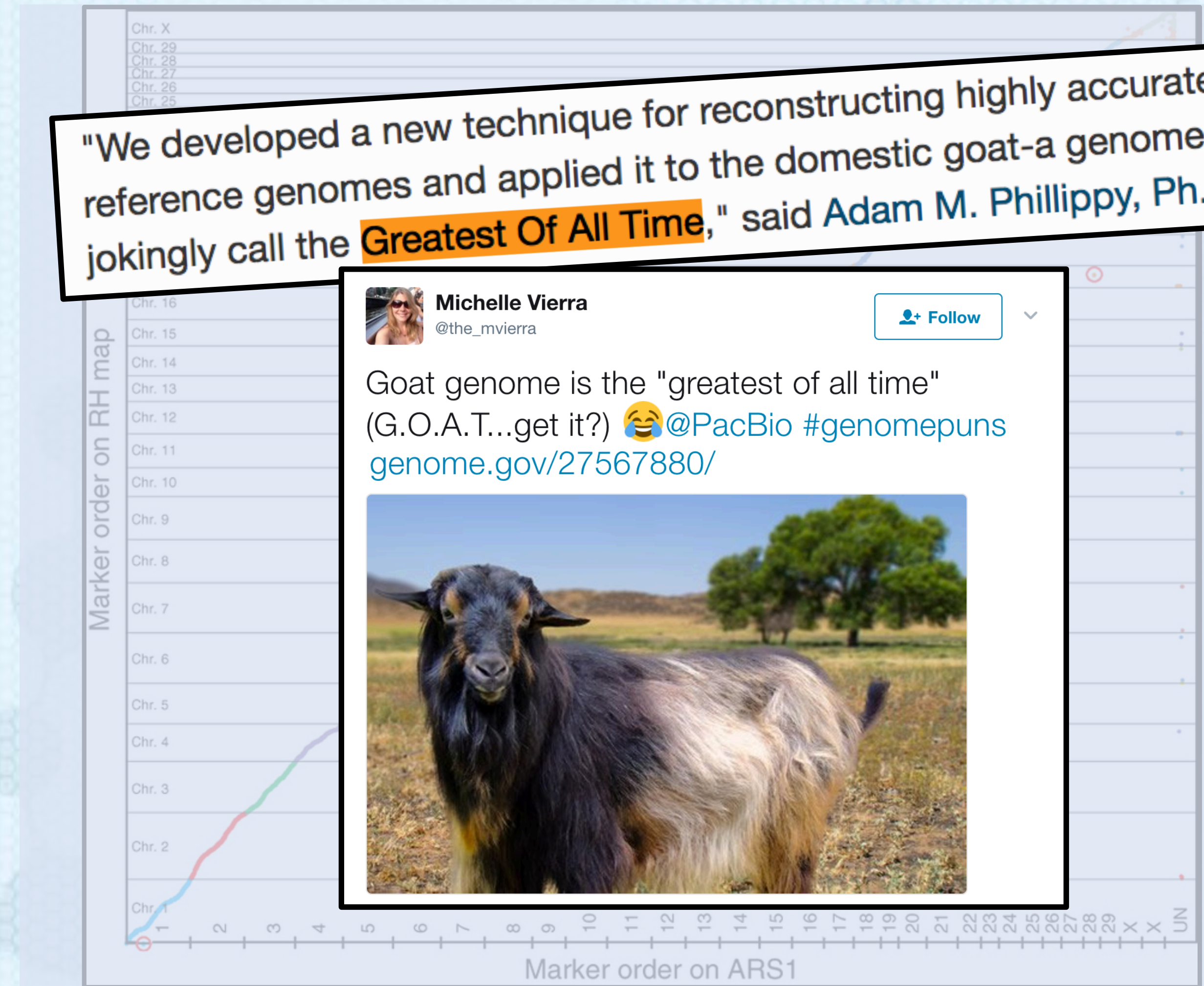
Technical Report | [OPEN](#)

Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome

Derek M Bickhart, Benjamin D Rosen [...] Timothy P L Smith

Nature Genetics **49**, 643–650 (2017)

Received: 02 August 2016



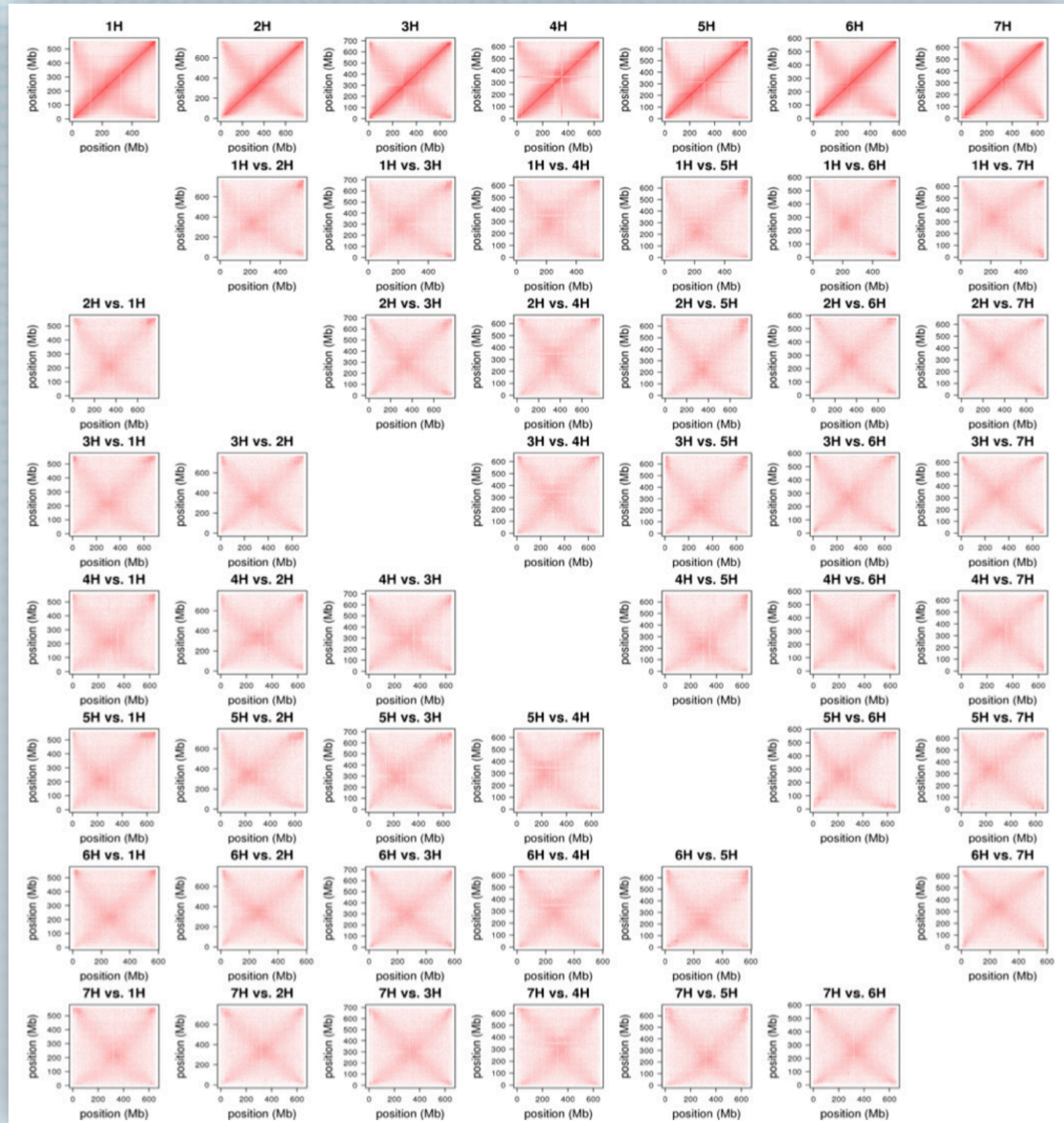
Proximo has been applied to hundreds of plant genomes

Polyploid examples:

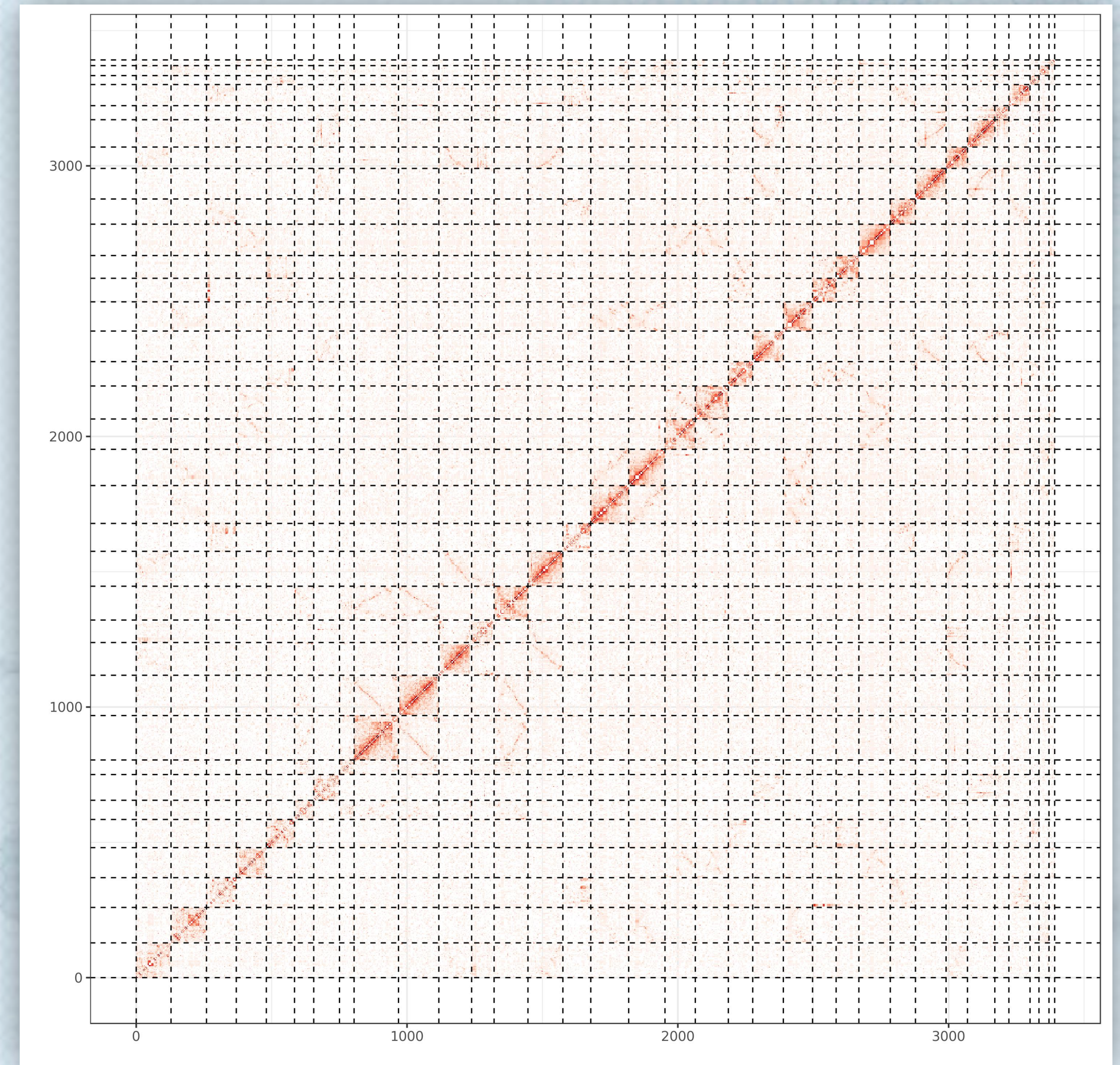
Genome Size (Mb)	Scaffolds (chr #)	Percent in Scaffolds	Original N50	Final N50	Ploidy
2,349	32	95.30%	310,769	77,373,286	tetraploid
740	13	93.03%	4,396,129	53,109,290	tetraploid
1,590	13	97.60%	315,162	126,778,945	variable ploidy
1,302	11	95.97%	294,343	115,898,643	variable ploidy
2,985	24	96.55%	22,539	120,321,936	tetraploid
555	29	87.61%	567,465	16,147,505	hexaploid
2,506	20	91.81%	70,775	111,320,837	tetraploid
1,330	18	95.53%	1,658,863	69,372,900	tetraploid
1,885	40	95.65%	477,316	45,420,557	tetraploid
975	35	99.66%	301,128	26,807,938	hexaploid
533	9	96.22%	1,339,543	58,304,087	variable ploidy

Hi-C data is useful for scaffolding tough polyploids

Barley genome



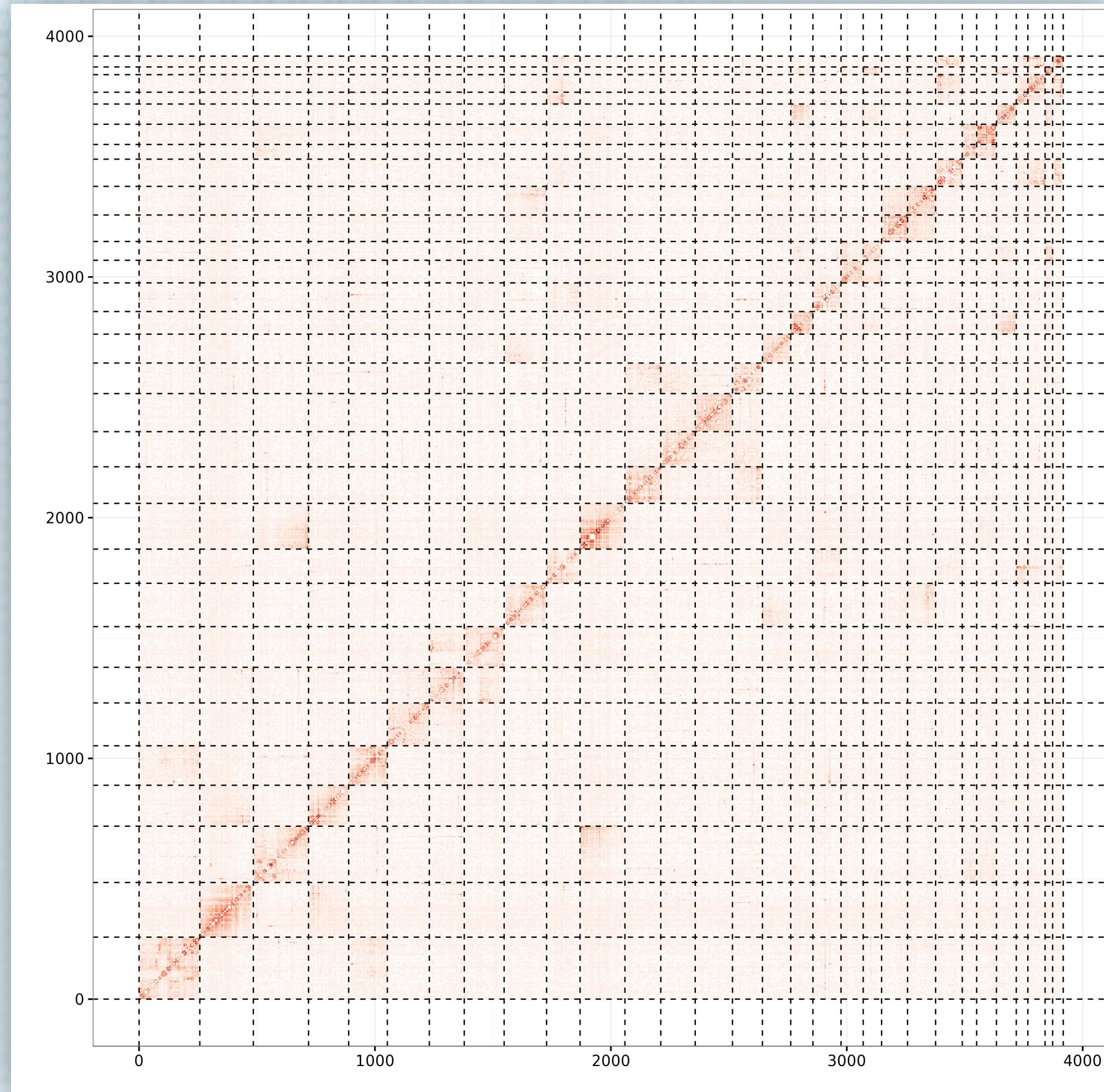
Pentaploid plant genome



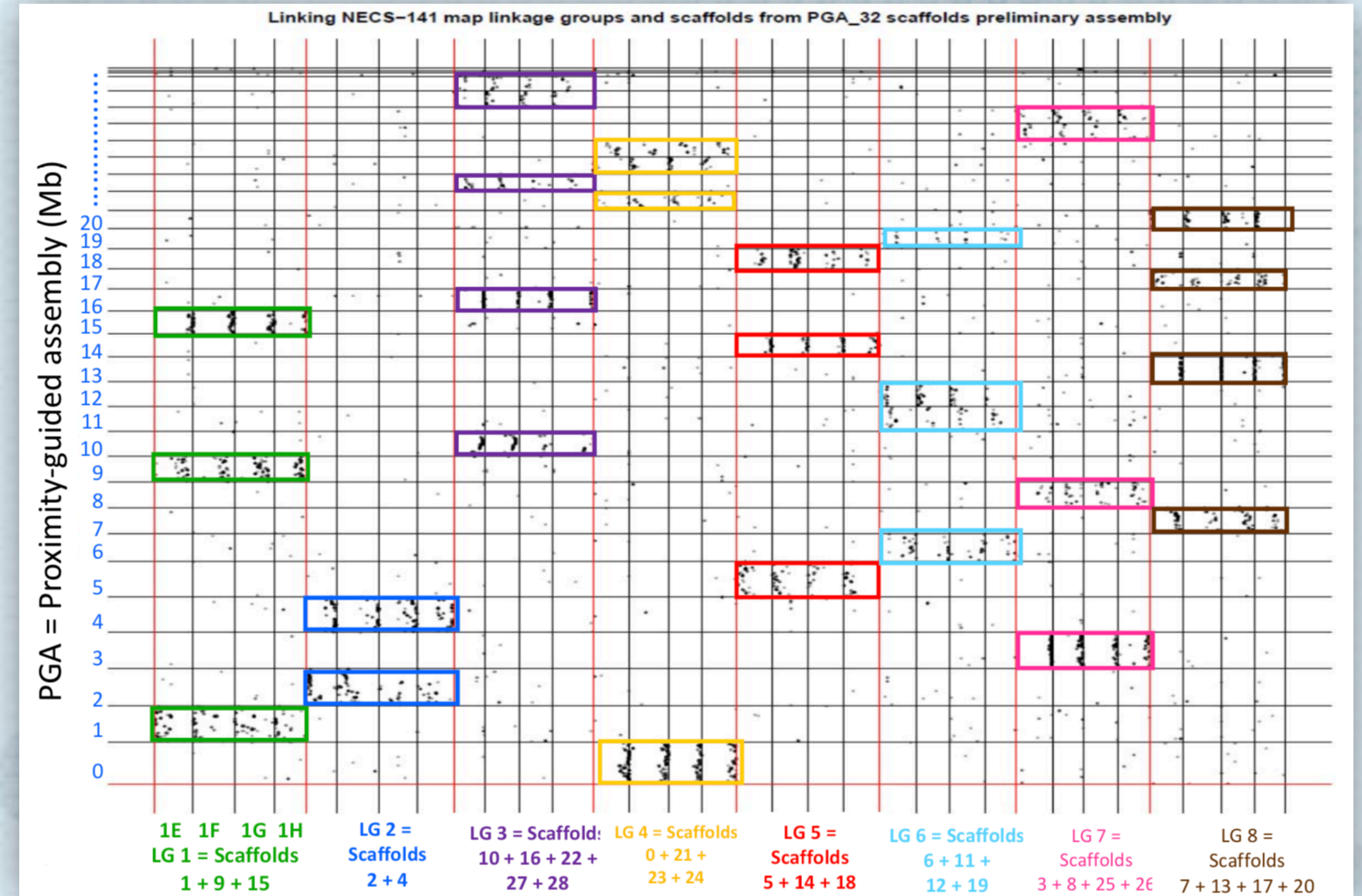
Mascher et al. Nature 2017

Proximo Alfalfa scaffolds are concordant with the genetic map

Proximo 32 chromosomes



Linkage groups vs Proximo



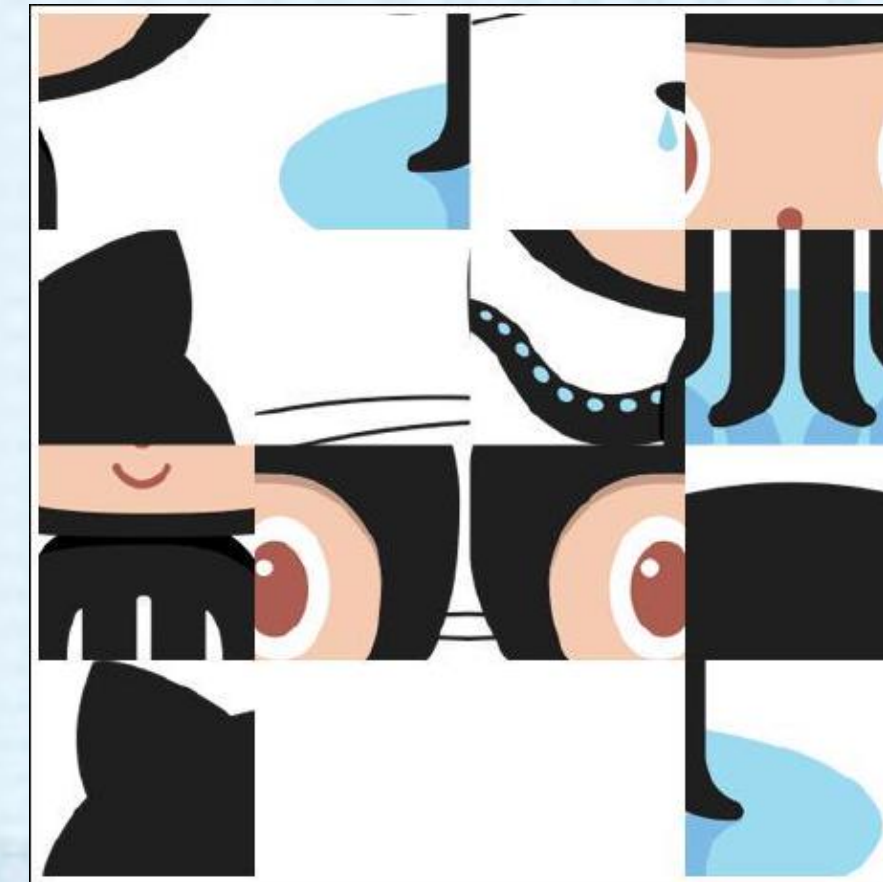
$(2n=4x=32)$

Open Source Scaffolding tools for HiC



SALSA2:

Jay Ghurye



3D-DNA:

Olga Dudchenko

Empowering the Hi-C community

- Hi-C is becoming a common tool for genome scaffolding
- Phase Genomics is committed to making Hi-C more accessible

Here's how we are doing it:

Georg Jander



Open source software



Dependable services

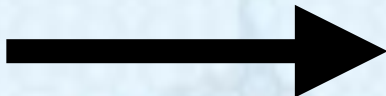
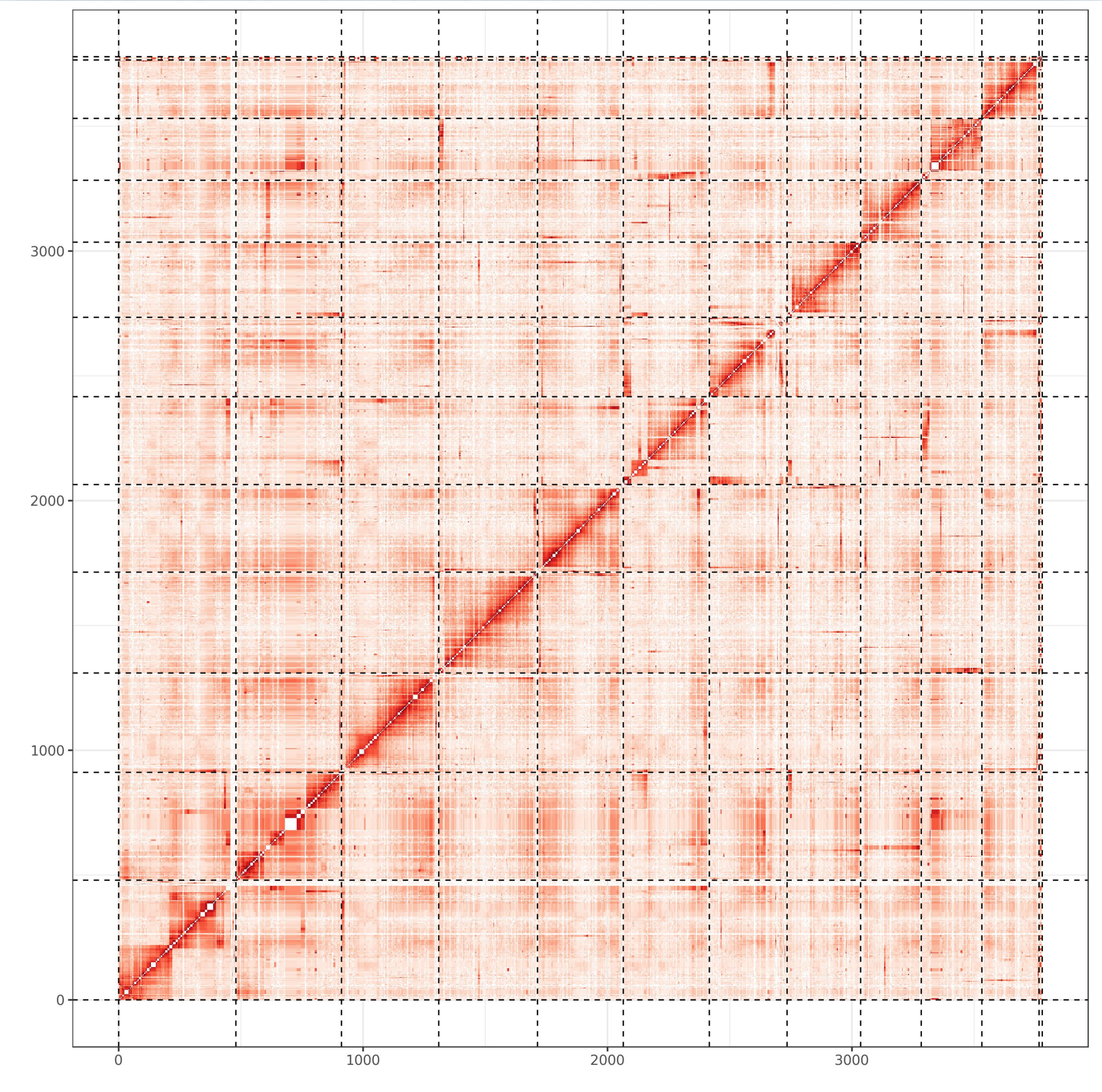


Faster sample preparation

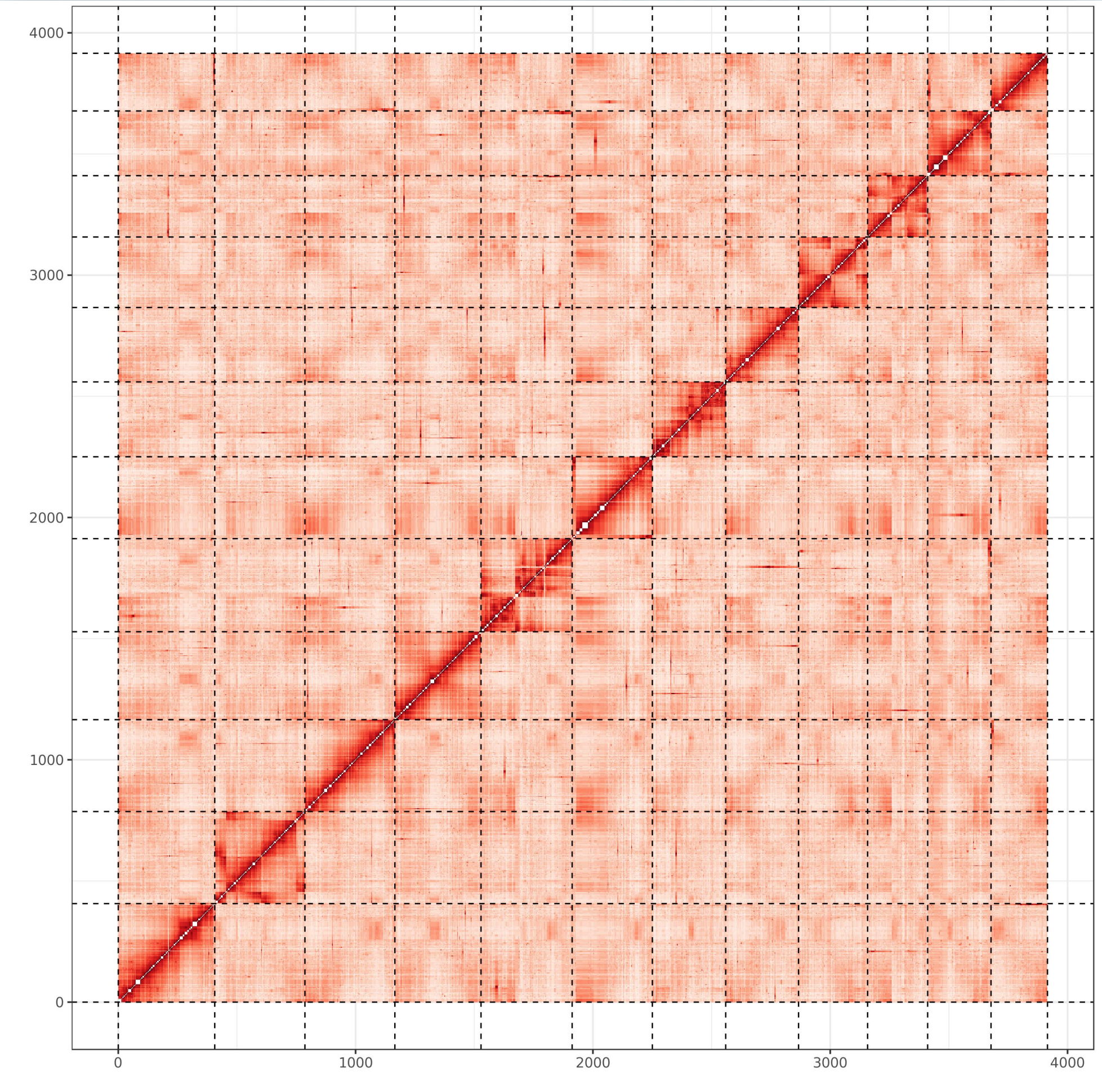
Correcting chimeric contigs



Initial scaffolding (Mint)



Polar Star scaffolding



Polar Star - a tool for breaking chimeric contigs



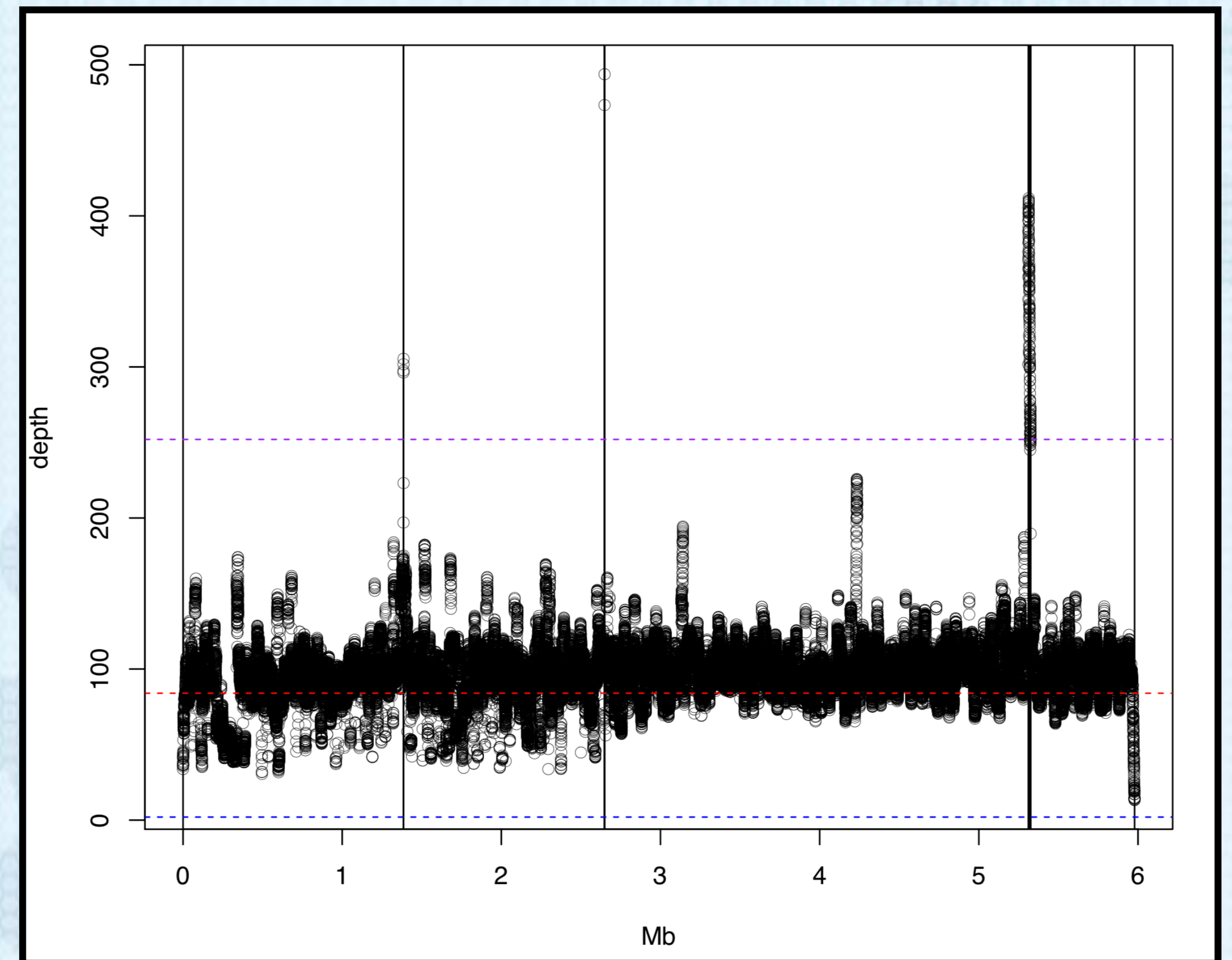
Polar Star



https://github.com/phasegenomics/polar_star

- After running Polar Star
- Majority of chimeric contigs broken
- No truncated chromosomes

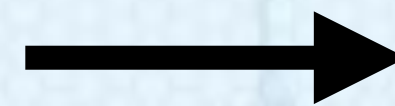
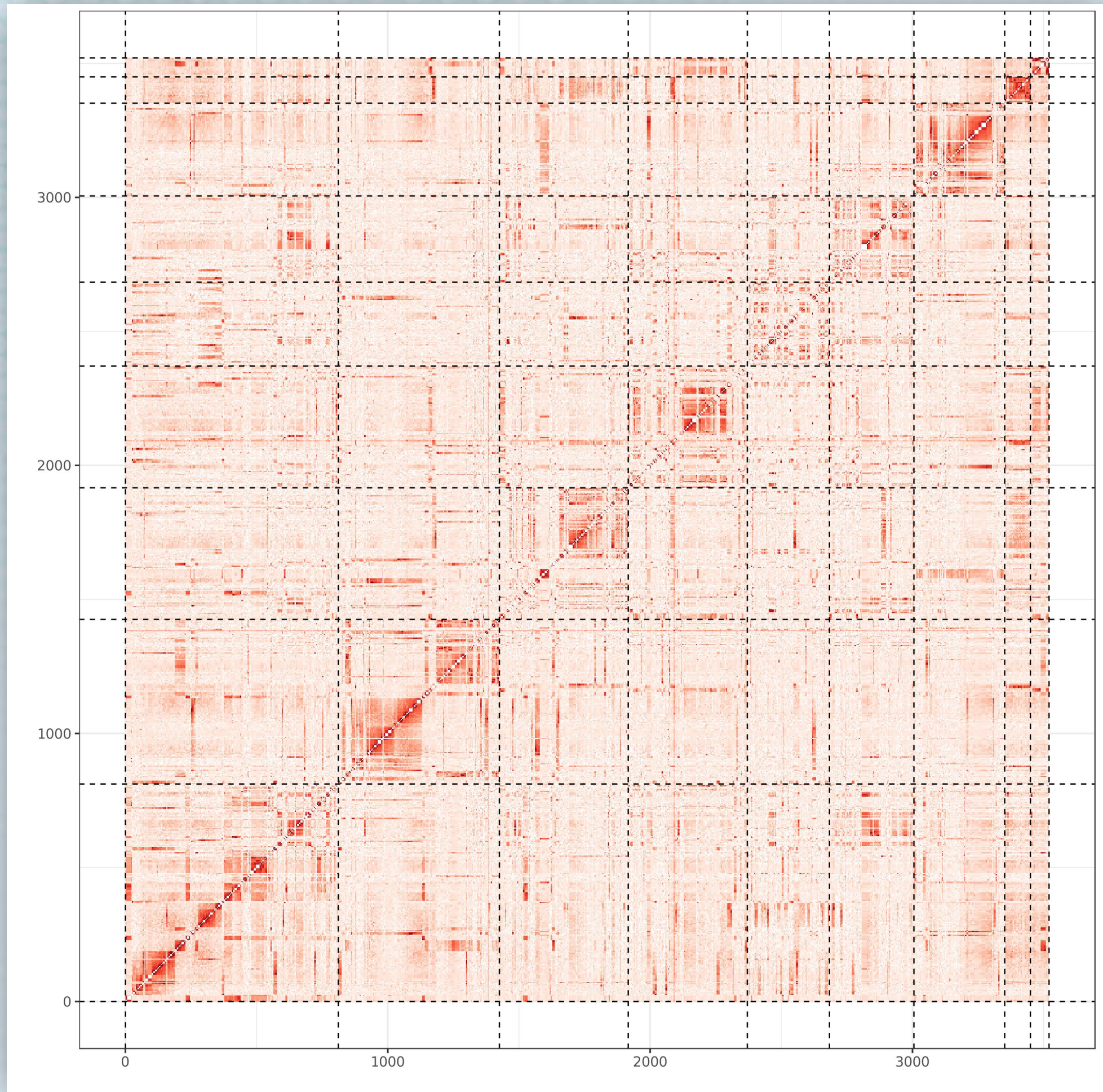
Chimeric contig



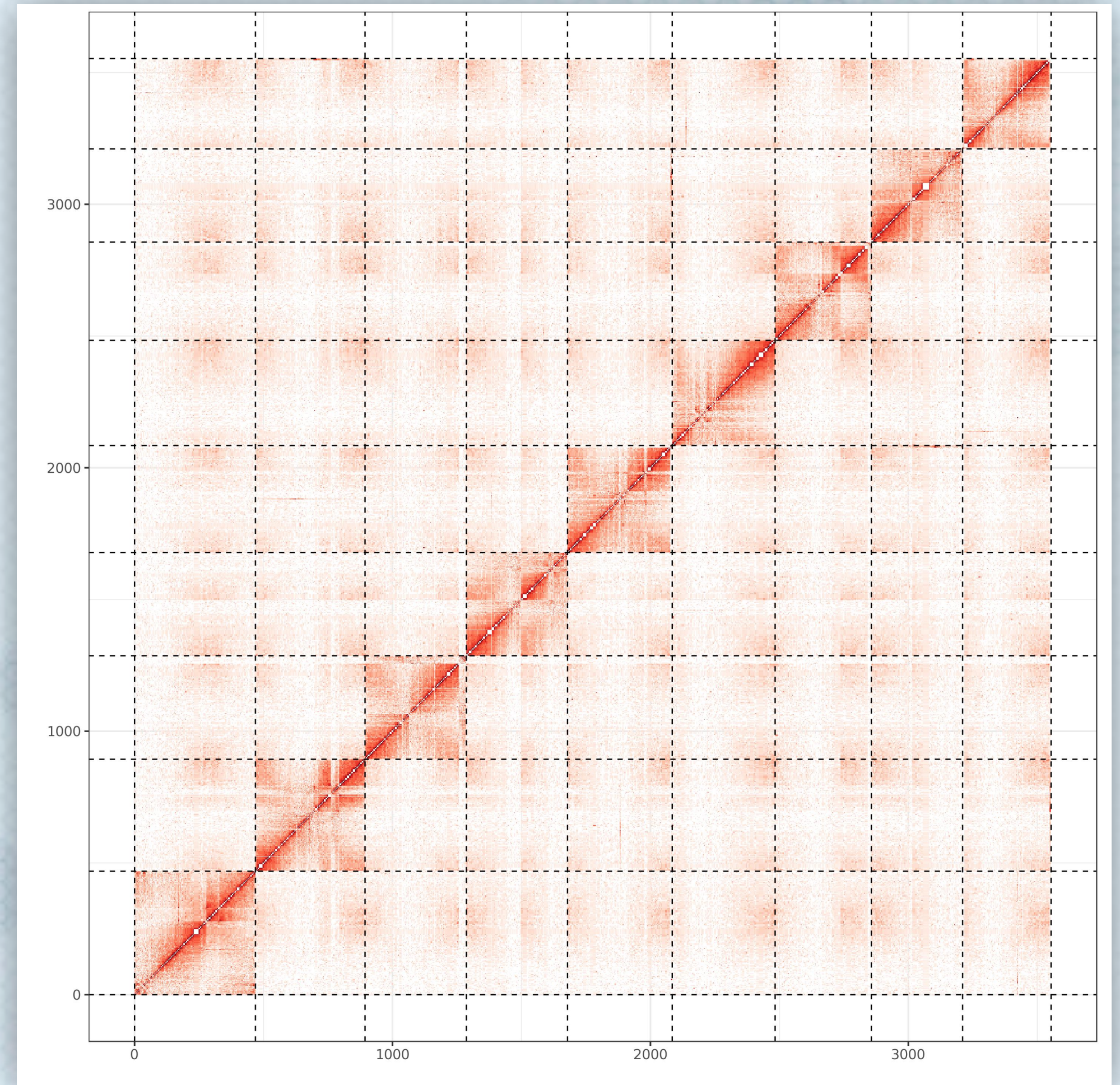


Filtering repeats

Unfiltered scaffolding (Kaniwa)



Post Matlock



Matlock - a multi-tool for working with Hi-C data



Matlock



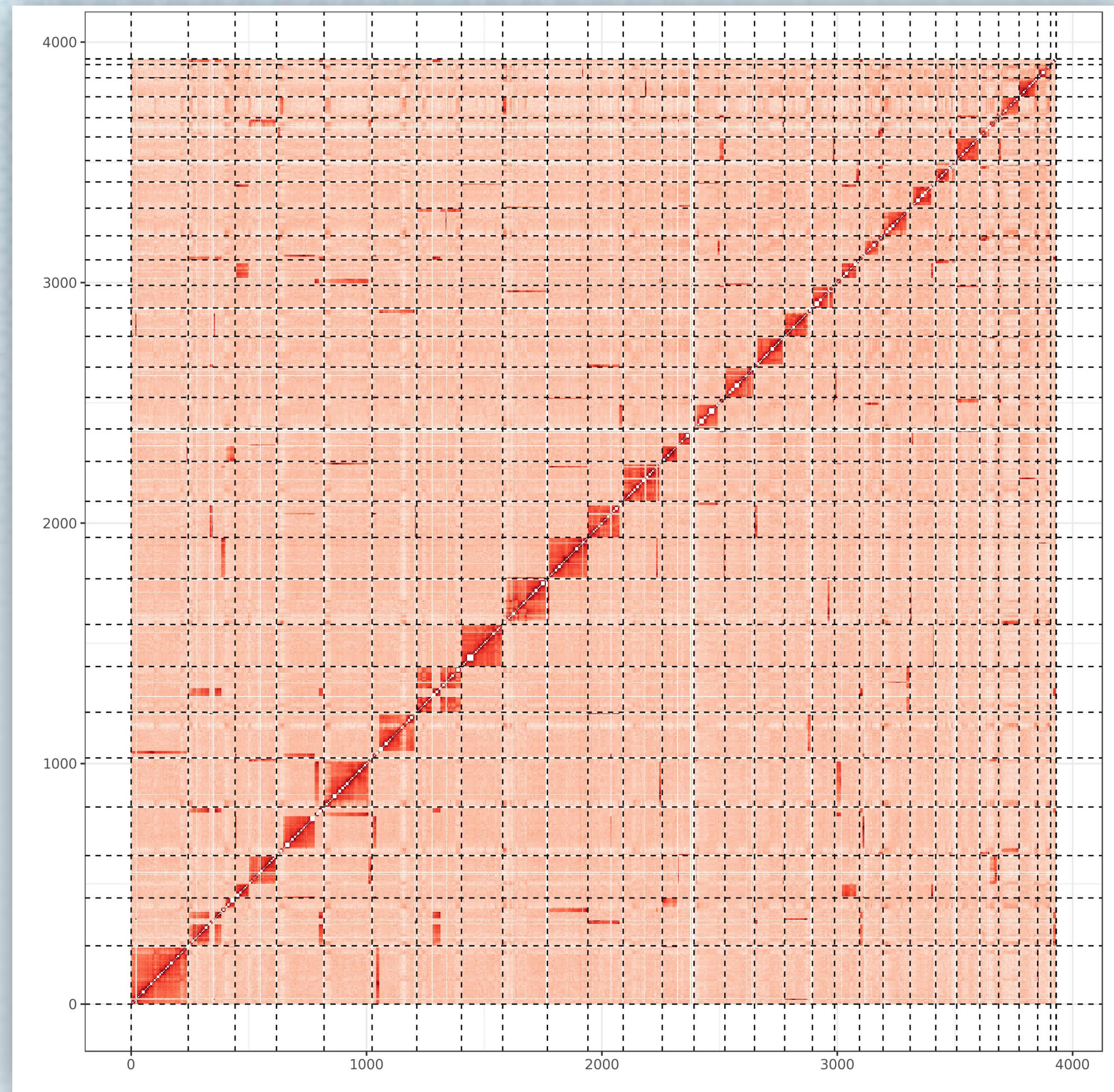
<https://github.com/phasegenomics/matlock>

- **Advanced Hi-C filtering**
- Restriction site counting
- Matrix normalization
- Supports BAM, SAM, and CRAM data
- Written in C for speed
- Compatible Proximo pipeline

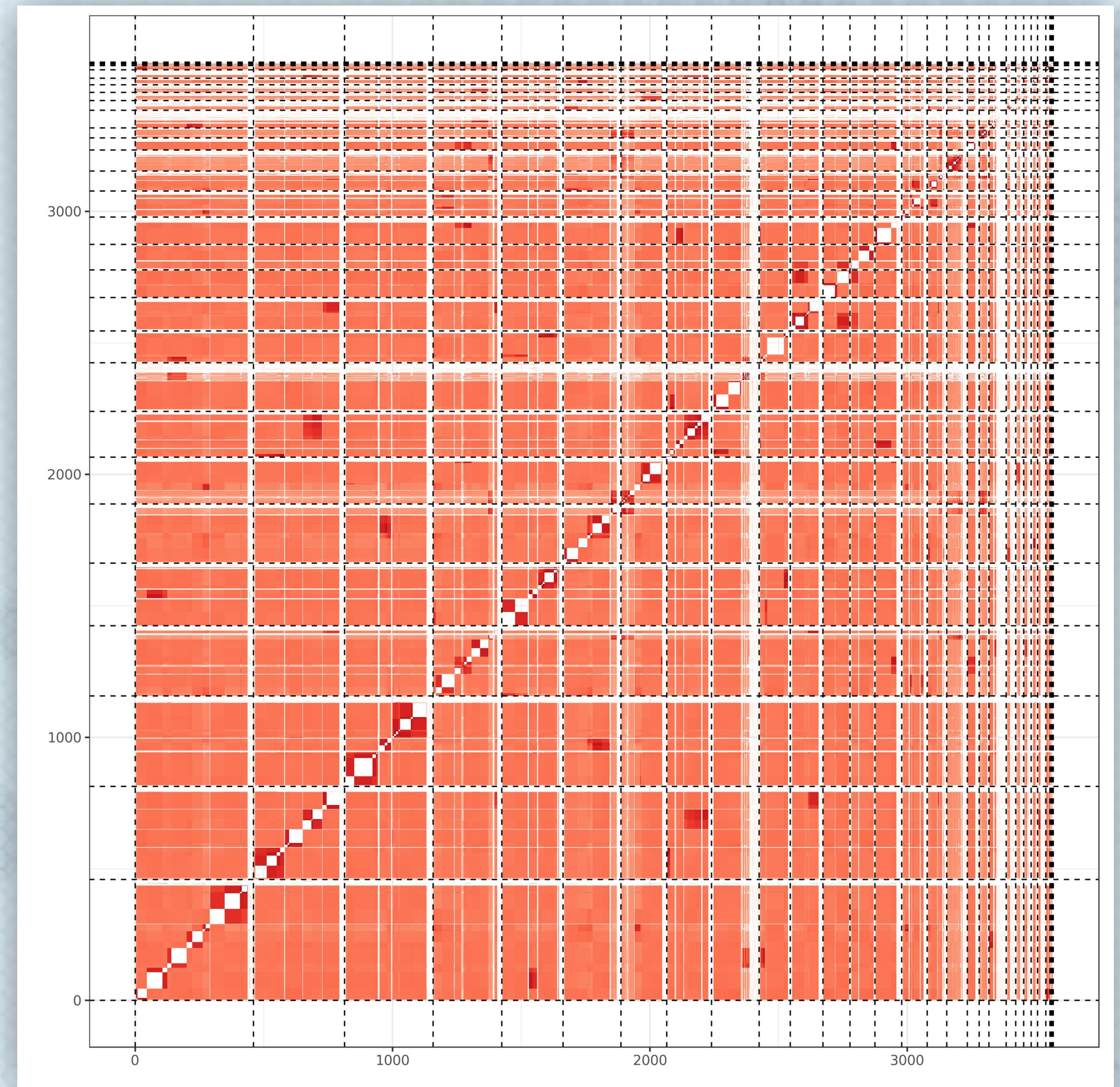
Using Matlock to partition Hi-C data prior to Proximo



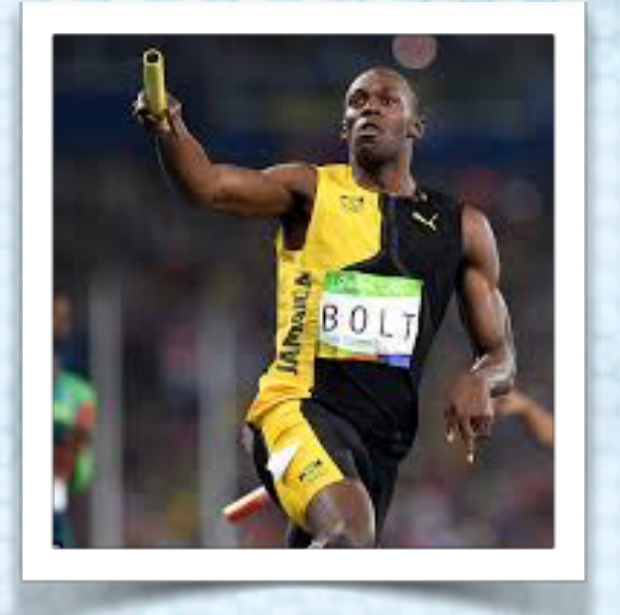
Falcon: Primary Contigs



Falcon: haplotigs

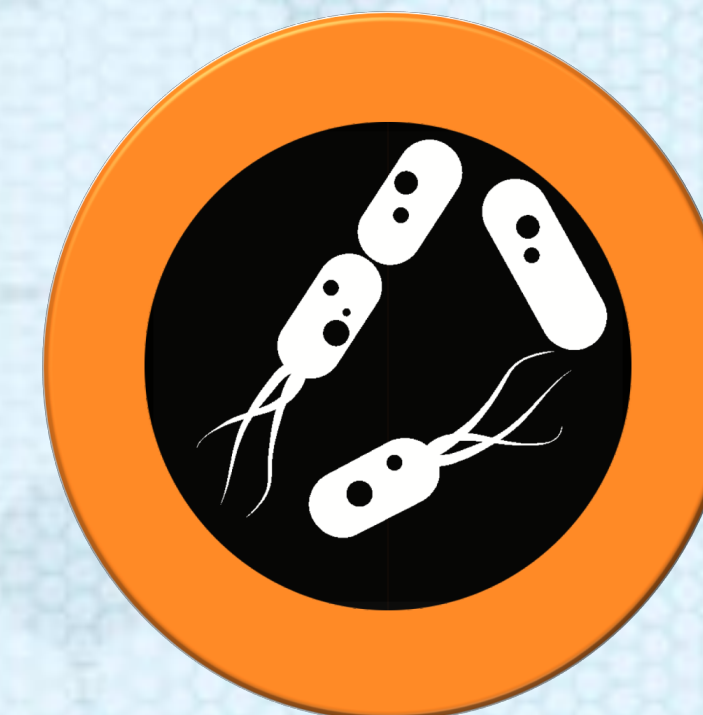


Phase Genomics now offers a Hi-C kit

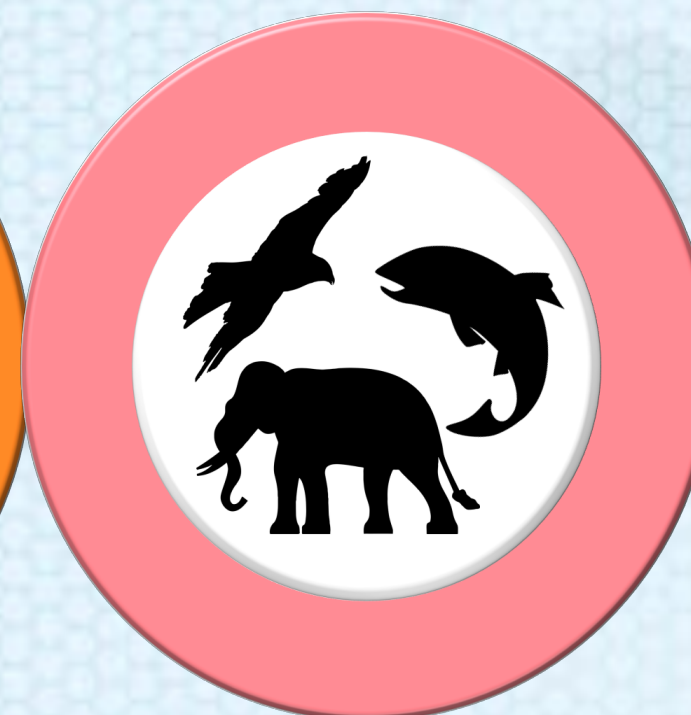


- **Fast (1 day)**
- **Easy to use**
- **Affordable**
- **Simple sample requirements**
- **No HMW DNA**

Metagenomics



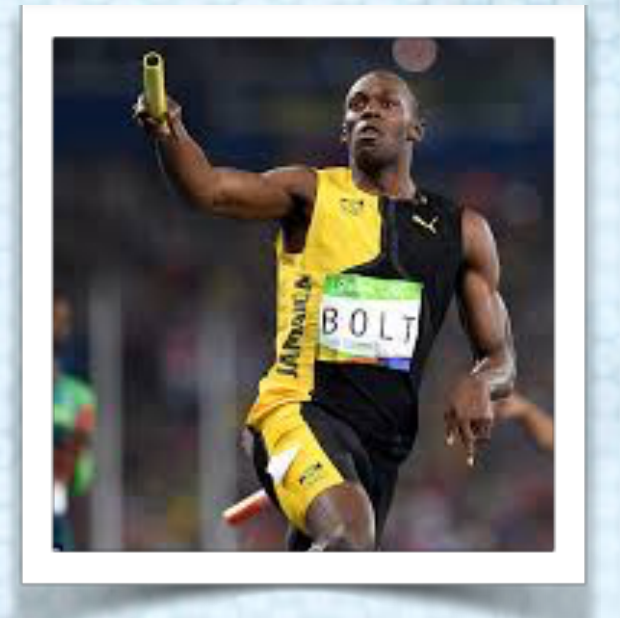
Animals



Plants

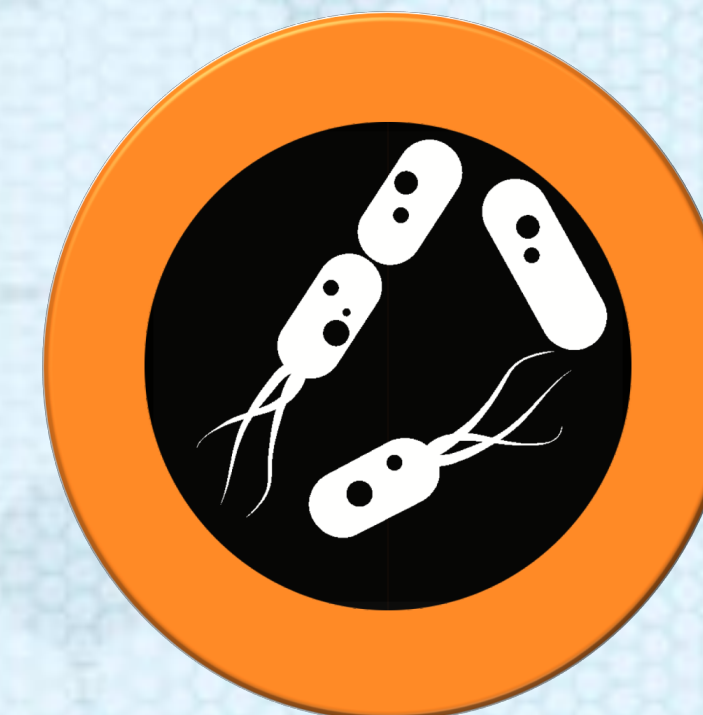


Phase Genomics now offers a Hi-C kit

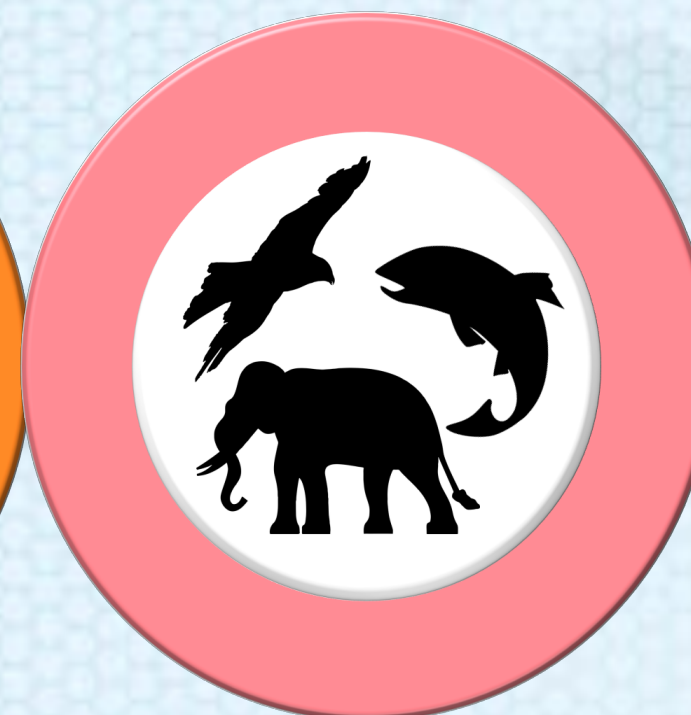


- **Fast (1 day)**
- **Easy to use**
- **Affordable**
- **Simple sample requirements**
- **No HMW DNA**

Metagenomics



Animals



Plants



Summary

- Hi-C scaffolding creates chromosome scale genomes
- Tease apart polyploidy and large structural changes
- Does not require high molecular weight DNA
- Both services and open access software are available
- **Now in an easy to use Kit!**

Acknowledgements

Nobel Research Institute

- Maria Monteros



Phase Genomics

- Shawn Sullivan
- Ivan Liachko
- Kayla Young

Oregon State University

- Kelly Vining



Brigham Young University

- Hayley Hansen
- Jeff Maughan

