

# The Ruminant T2T Consortium

Tim Smith, USDA – Brenda Murdoch, U. Idaho – Stephanie McKay, U. Missouri

**Ben Rosen ([ben.rosen@usda.gov](mailto:ben.rosen@usda.gov))**

**Animal Genomics and Improvement Laboratory  
USDA-ARS, Beltsville, MD**

# Outline

- Why do we need T2T genomes
- Motivations for a ruminant T2T project
- How we assemble T2T genomes
- Draft ruminant T2T genomes

## Ruminant T2T Workshop

### Day 1: February 23, 2023 (8:00 AM – 4:30 PM)

8:00 Transportation from hotel to USMARC  
8:30 – 8:40 Registration and Welcome  
8:40 – 8:45 Welcome (Mark Boggess, Center Director)  
8:45 – 9:00 Setting the stage (Tim Smith)

### Session 1: Lessons from Human T2T

9:00 – 10:00 Adam Phillippy  
10:00 – 10:30 Coffee Break  
10:30 – 11:30 Karen Miga  
11:30 – 12:30 Rachel O'Neill  
12:30 – 1:30 Lunch

### Session 2: Non-Primate Species

1:30 – 2:00 Erich Jarvis (remote)  
2:00 – 2:45 Klaus-Peter Koepfli  
2:45 – 3:15 Yana Safonova  
3:15 – 3:30 Tamara Potapova  
3:30 – 4:30 Discussion and working group assignments  
4:30 Depart for hotel  
6:00 Networking dinner, The Lark (downtown Hastings)

### Day 2: February 24, 2023 (8:00 AM – 11:30)

### Session 3: Defining Ruminant T2T Assembly and Analysis Teams

8:00 Transportation from hotel to USMARC  
8:30 – 8:45 Stephanie McKay – Day 2 welcome and charge  
8:45 – 9:05 Sergey Koren  
9:05 – 9:30 Arang Rhie  
9:30 – 9:45 Ben Rosen  
9:45 – 10:00 Brenda Murdoch  
10:00 – 11:30 Working groups breakout for planning analyses  
11:30 – 11:45 Close out and creating virtual groups  
11:45 Lunch for those who want it before departure



Meat Animal Research Center

# Why we need T2T genomes

- **Human T2T assembly added ~200Mb**
  - Centromeres, segmental duplications, duplicated gene families, rDNA arrays
- **Improved mapping of both long and short read datasets**
  - 100's of thousands new variants/sample, 10's of thousands fewer false variants, ↑ CpG methylation, ↑ chromatin status

## Do we really need that 8%?

- Well, without it ... you die.
- The “Koren” test of assembly quality
  - If I synthesize the assembled genome... do I get the same, living organism back?

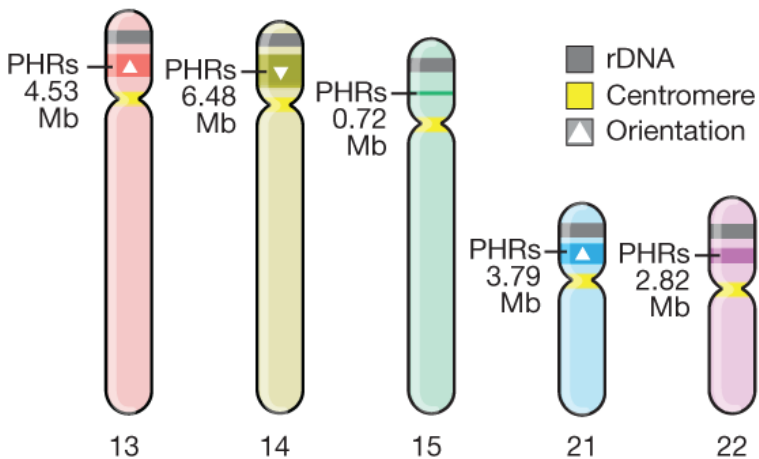


Slide Credit Adam Phillippy

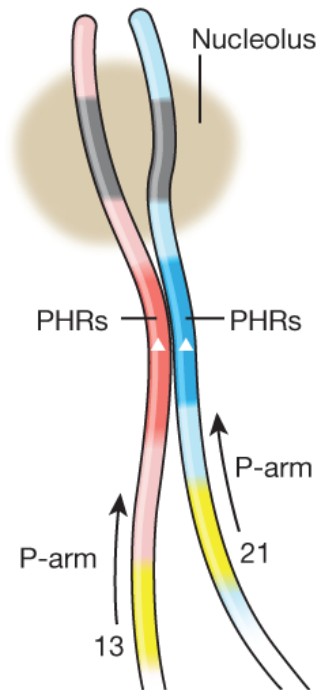
# Why we need T2T genomes

Novel biology revealed : rDNA arrays can drive heterologous chromosome recombination

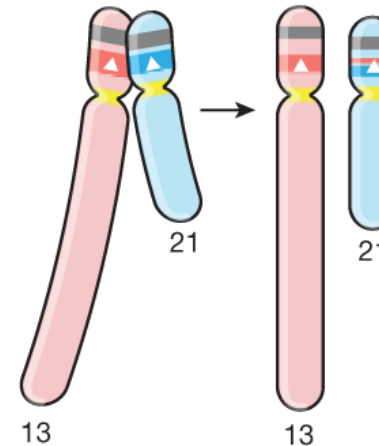
**a Pseudo-Homologous Regions**



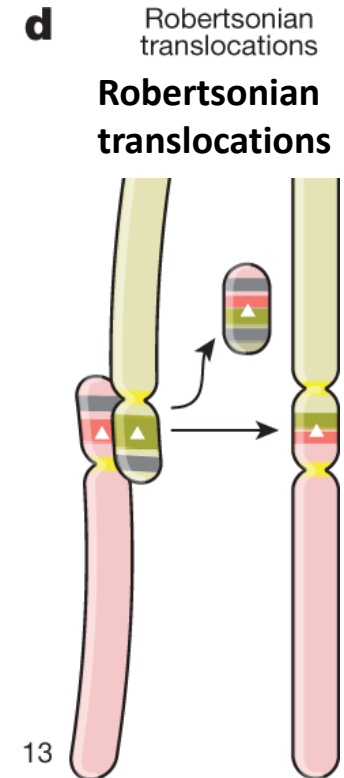
**b Physical proximity**



**c Recombination**



**d**

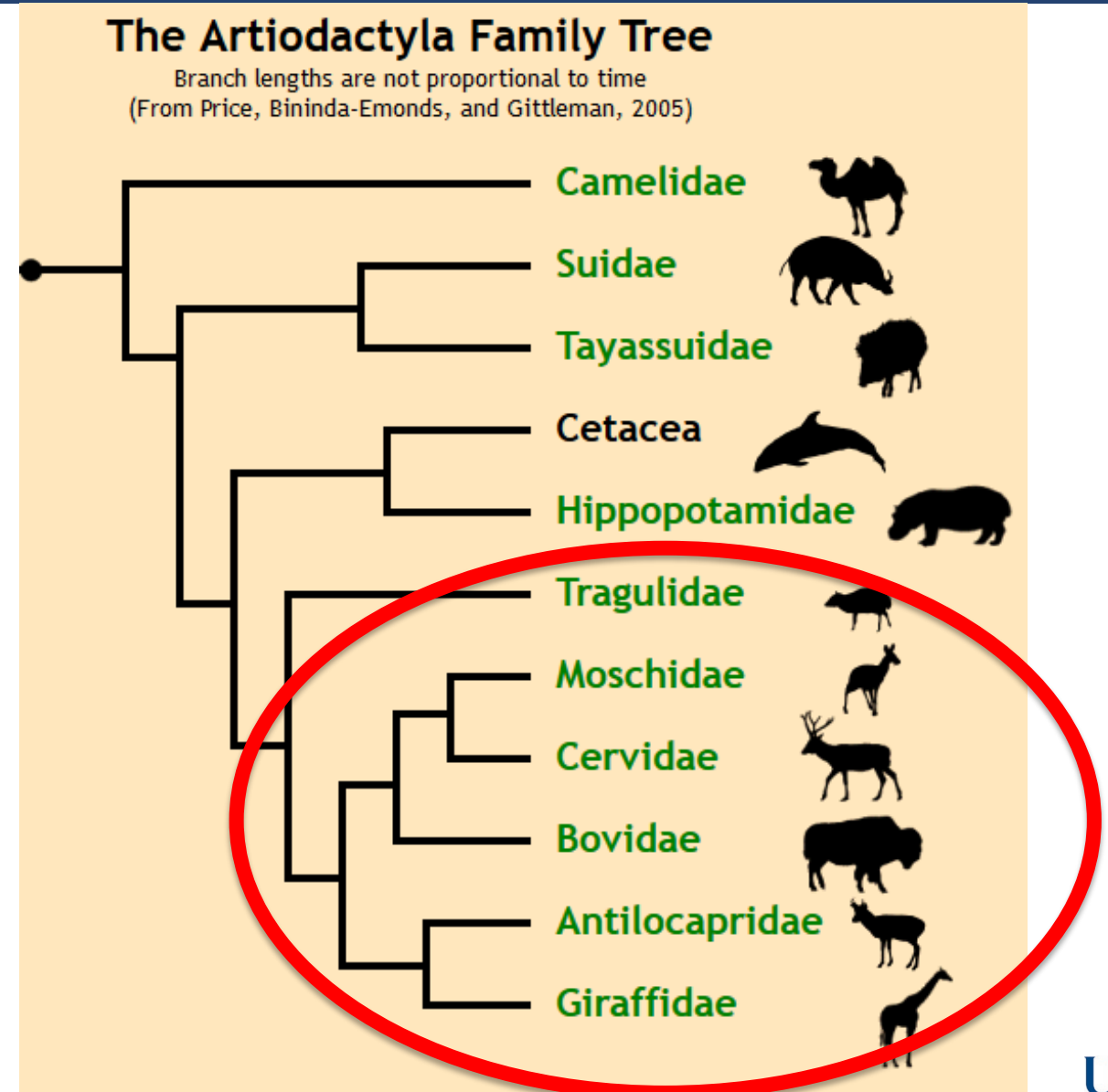


Recombination between heterologous human acrocentric chromosomes

*Nature* **617**, 335–343 (2023)

# Motivations for a ruminant T2T project

- Agricultural production – cattle, sheep and goats



# Motivations for a ruminant T2T project

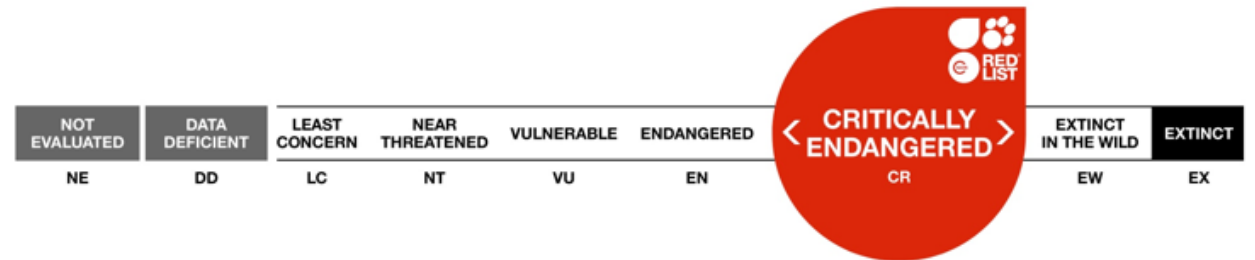
- **Domestication – yak, buffalo, bison, deer**



# Motivations for a ruminant T2T project

- Conservation of endangered species

## Dama Gazelle (Nanger dama)



- Wild: 100 – 200
- Zoos (worldwide): 500 – 600
- Qatar and UAE collections: 192
- Private ranches (Texas): 1,510



Slide Credit Klaus-Peter Koepfli

# Motivations for a ruminant T2T project

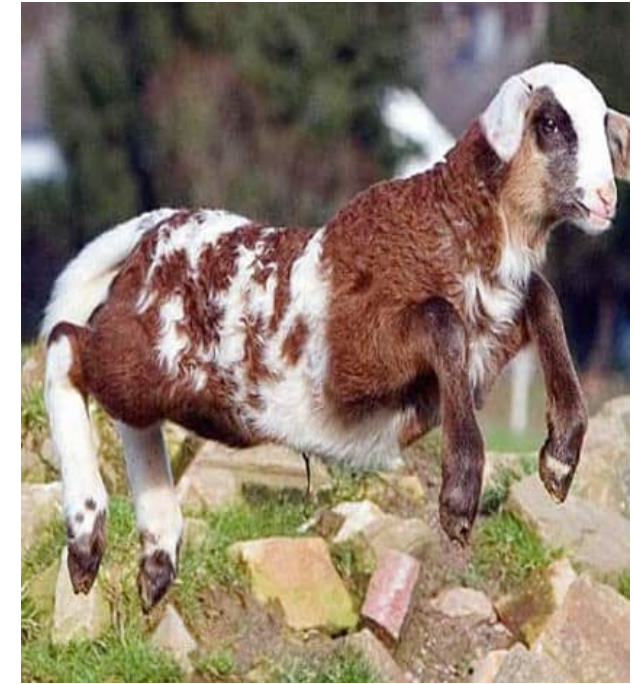
- **Broad evolutionary range (25Mya – interfertile)**



Mithun



Yak



Geep



# Motivations for a ruminant T2T project

- Chromosome biology
- Majority of chromosomes are acro/telocentric
- range  $2n=6/7-70$  (ancestral state  $2n=60$ )

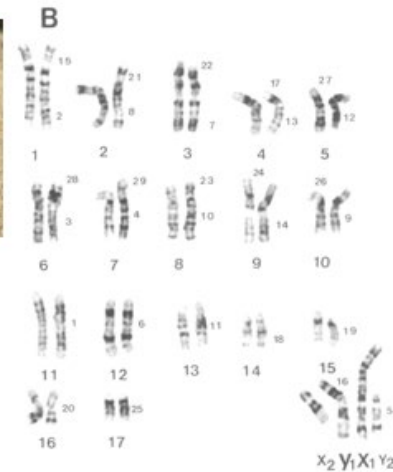
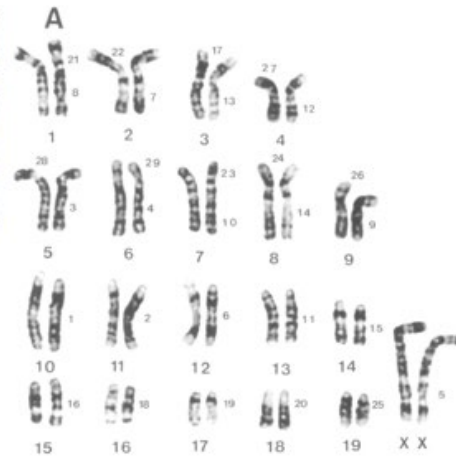


Chinese Muntjac  $2n=46$   
Photo credit: Rufus46 Wikimedia Commons



Indian Muntjac  $2n=6/7$   
Photo credit: Bernard Dupont

## Dama Gazelle



Vassart *et al.* 1995 *J Heredity*



$2n = 38$

$2n = 39$

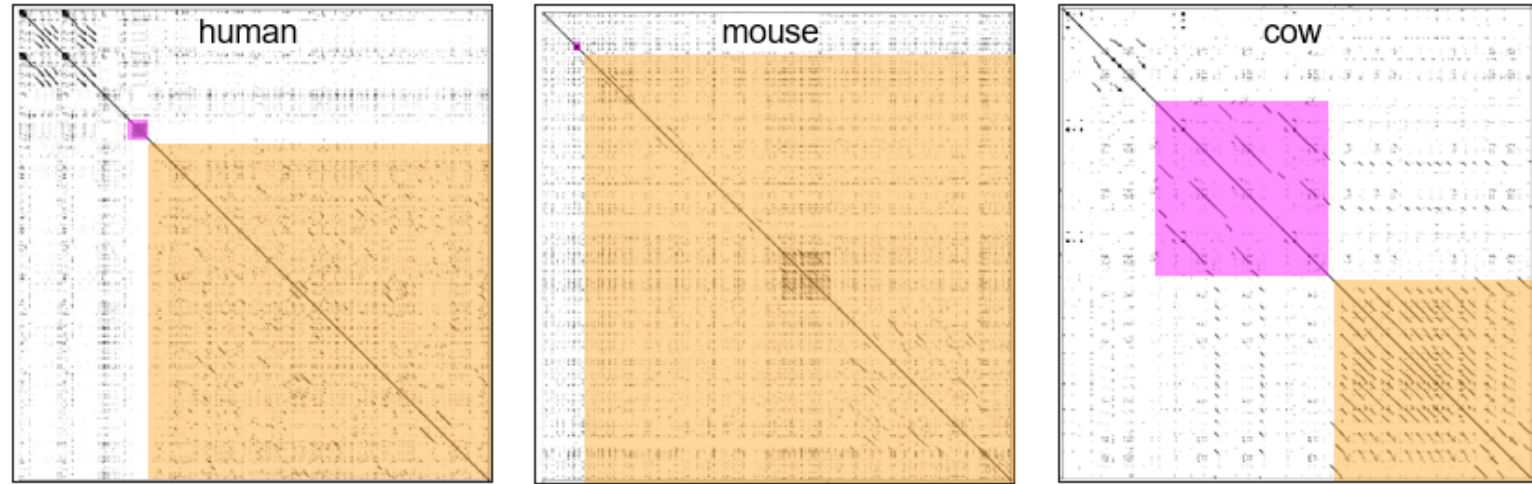
$2n = 40$

$2n = ?$

Slide Credit Klaus-Peter Koepfli

# Motivations for a ruminant T2T project

- **Distinct characteristics of immune system (ultralong antibody genes discovered in cattle)**



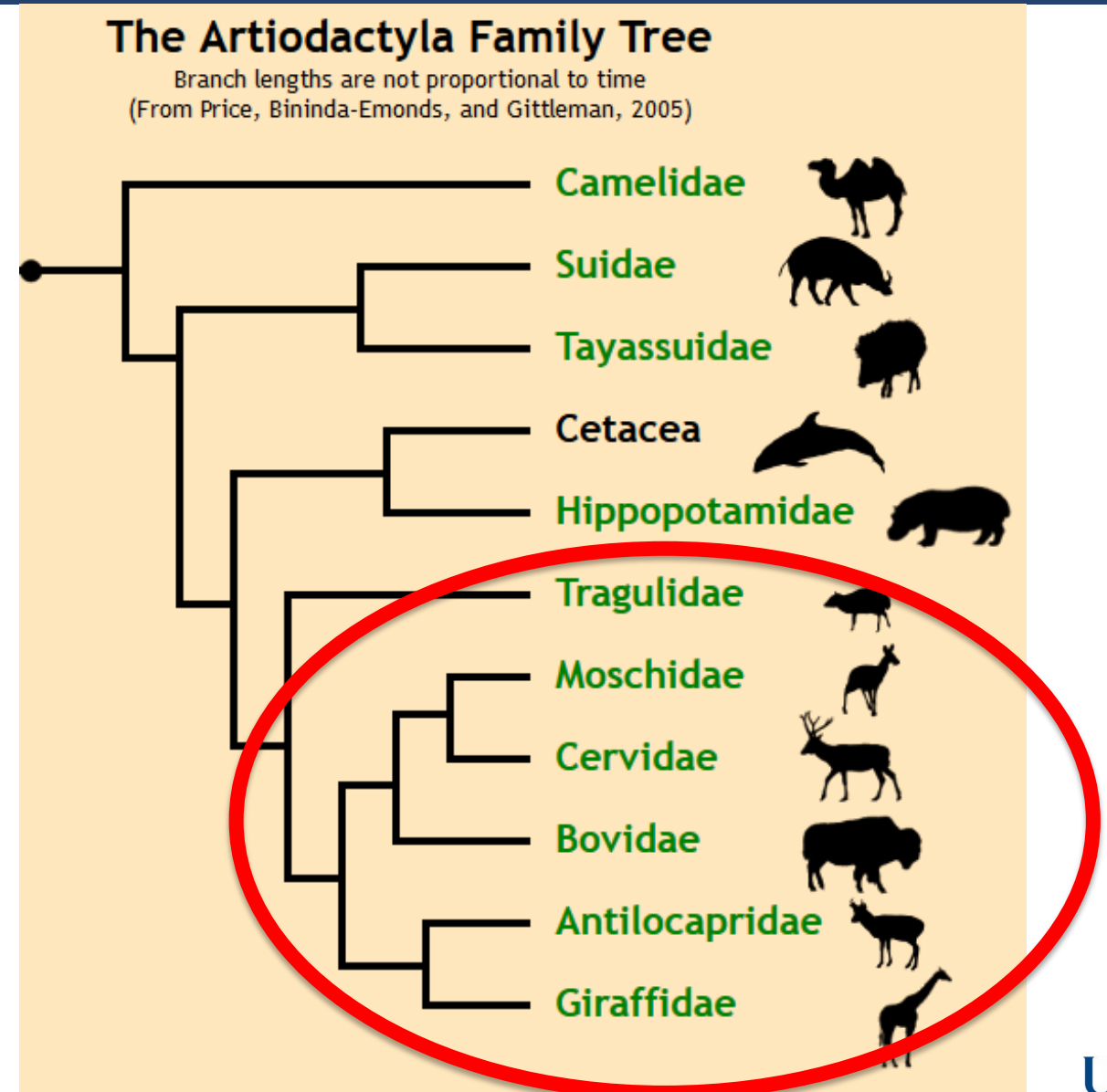
	human	mouse	cow
<b>IGH locus length</b>	1.3 Mbp	2.7 Mbp	0.7 Mbp
<b># V genes</b>	~129	~170	10
<b>D-J region length</b>	60 kbp	90 kbp	200 kbp

D-J region      V region

Slide Credit Yana Safonova

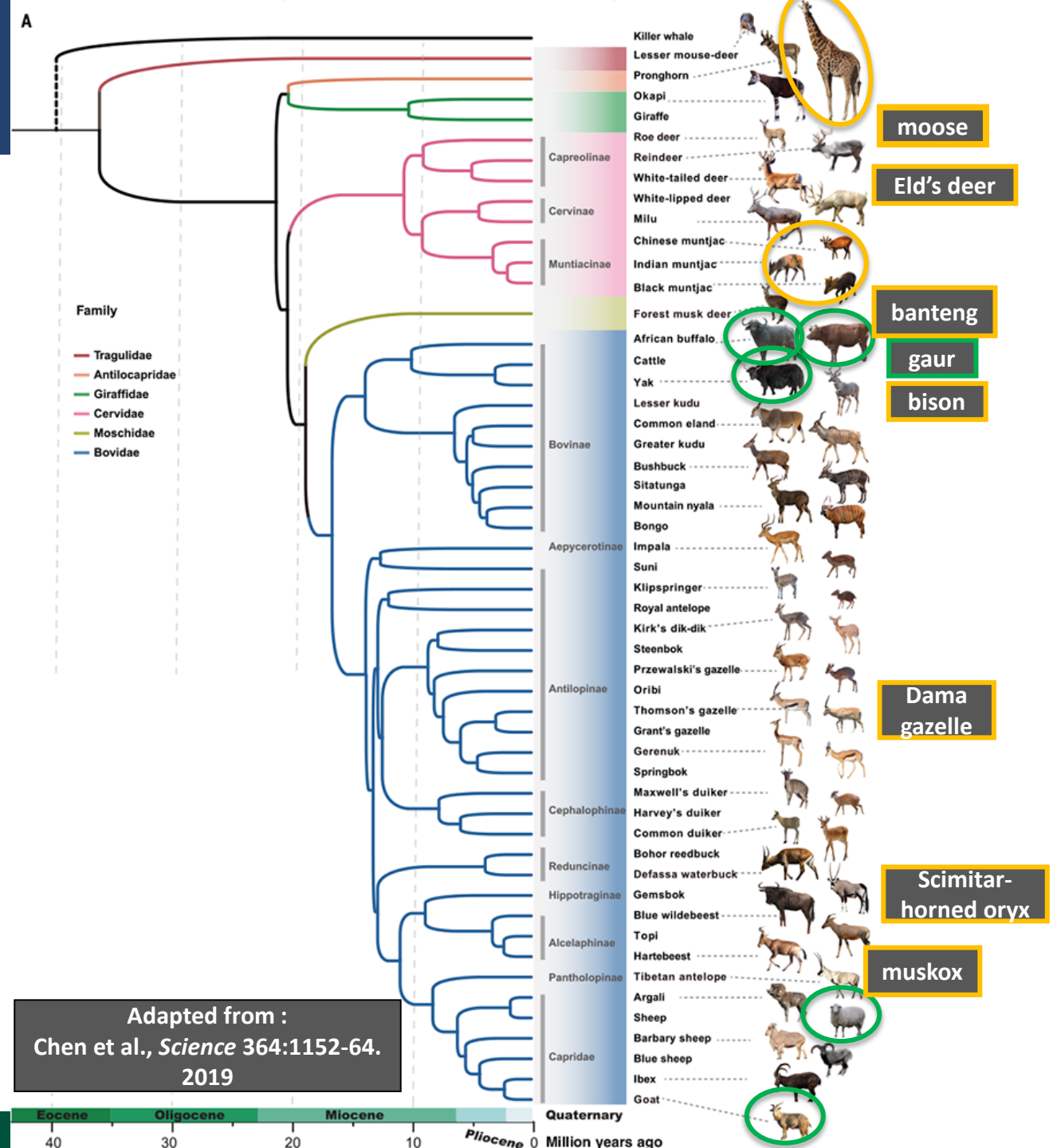
# Motivations for a ruminant T2T project

- Samples from equivalent developmental stage fetus
- Several interspecies crosses (bighorn x sheep; wapiti x red deer; yak, bison, gaur x cattle)
- Hi-C (Micro-C) for multiple tissues from these fetuses
- HiFi and ONT 5mC calling by comparison to bisulfite in cattle, gaur, sheep, goat
- Matching RNAseq, IsoSeq data for fetal tissues



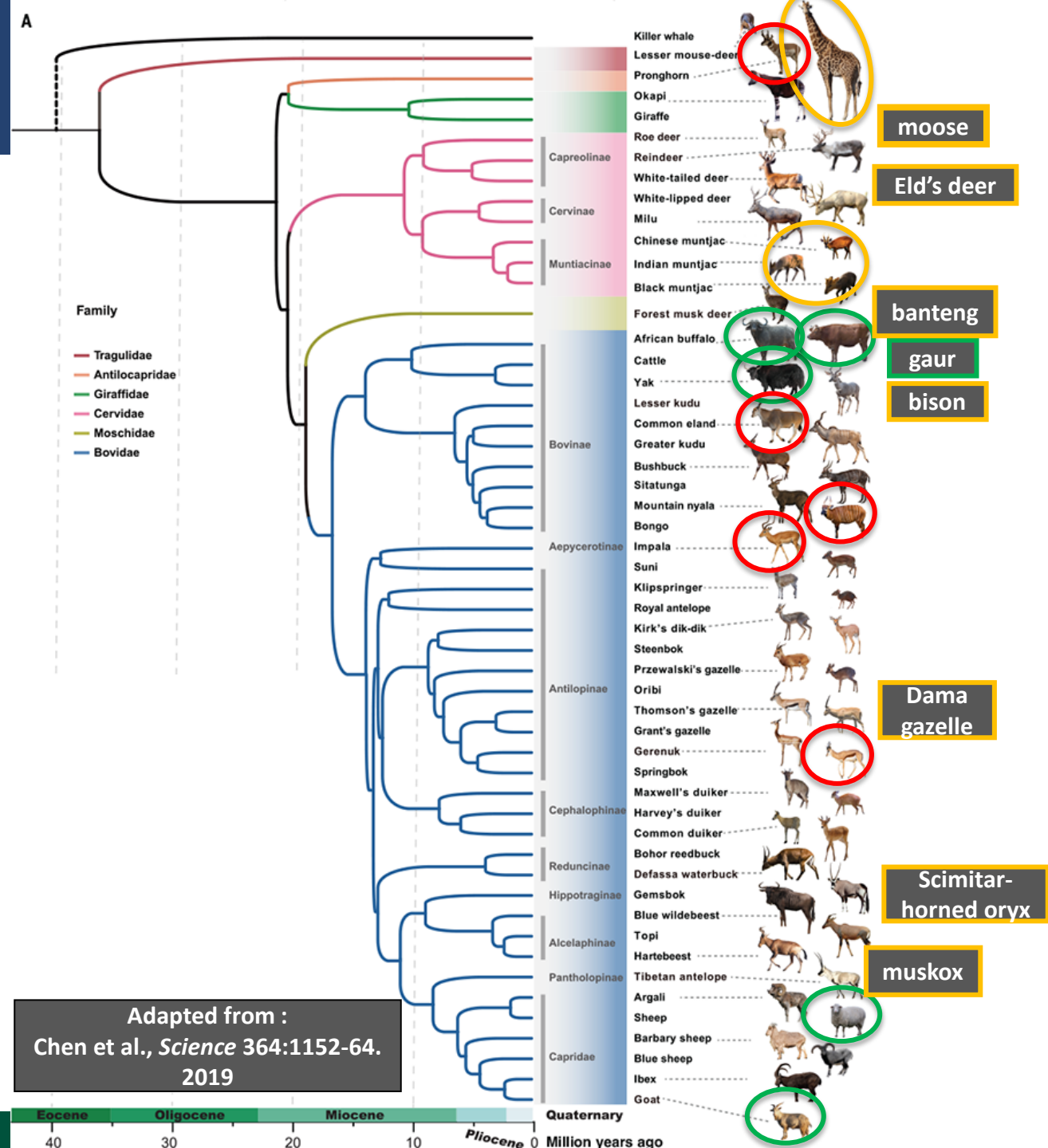
# Target species

- **Artiodactyla**
  - Pig, camel, whale
- **Giraffidae**
  - Giraffe
- **Cervidae**
  - Moose, Eld's deer, muntjac
- **Bovinae**
  - Cattle, gaur x cattle, water buffalo, bison x cattle, banteng
- **Antilopinae**
  - Dama gazelle
- **Hippotraginae**
  - Scimitar-horned oryx, roan antelope
- **Caprinae**
  - Sheep, bighorn x sheep, goat, musk ox



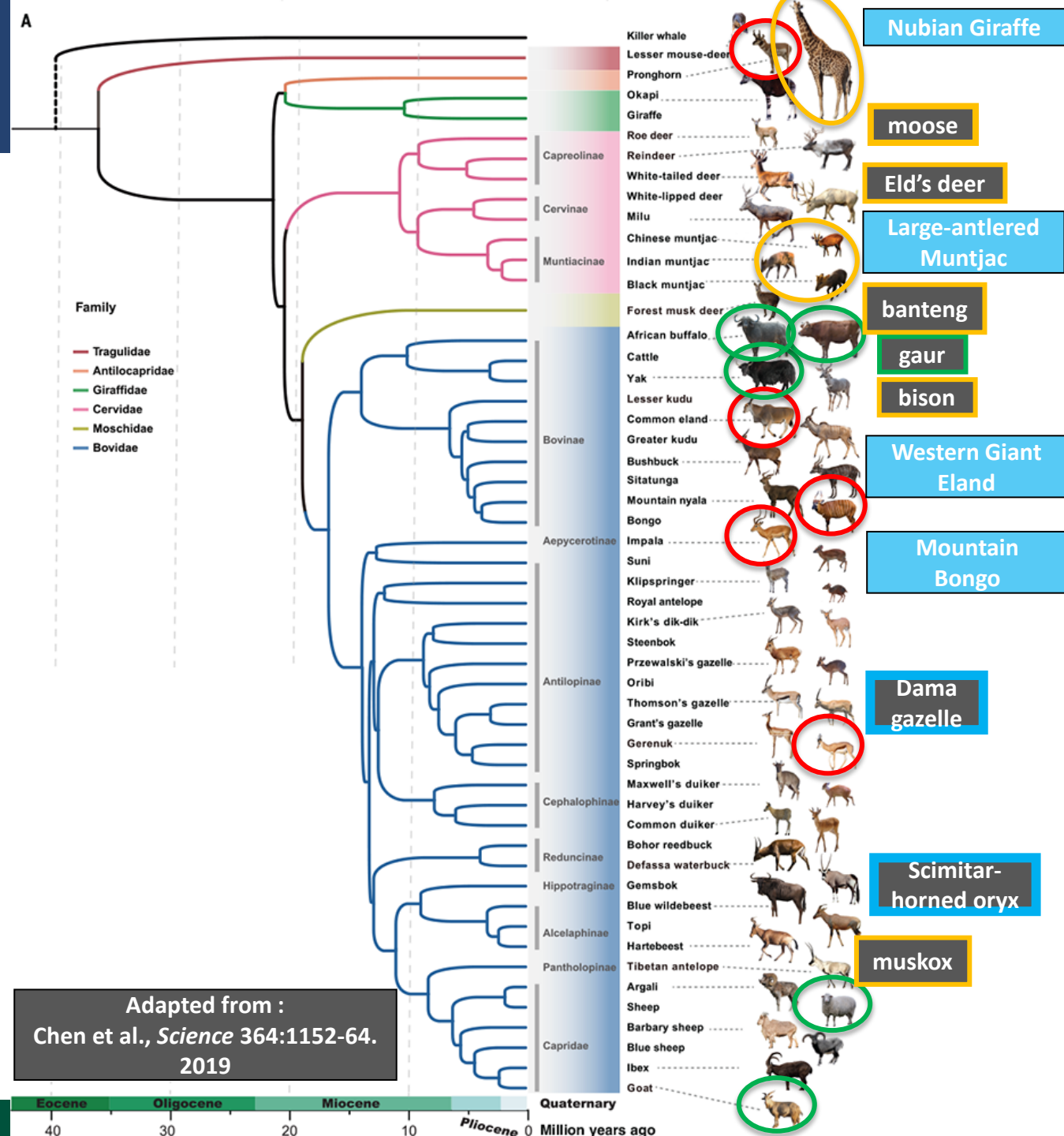
# Target species

- **Artiodactyla**
  - Pig, camel, whale
- **Giraffidae**
  - Giraffe
- **Cervidae**
  - Moose, Eld's deer, muntjac
- **Bovinae**
  - Cattle, gaur x cattle, water buffalo, bison x cattle, banteng
- **Antilopinae**
  - Dama gazelle
- **Hippotraginae**
  - Scimitar-horned oryx, roan antelope
- **Caprinae**
  - Sheep, bighorn x sheep, goat, musk ox



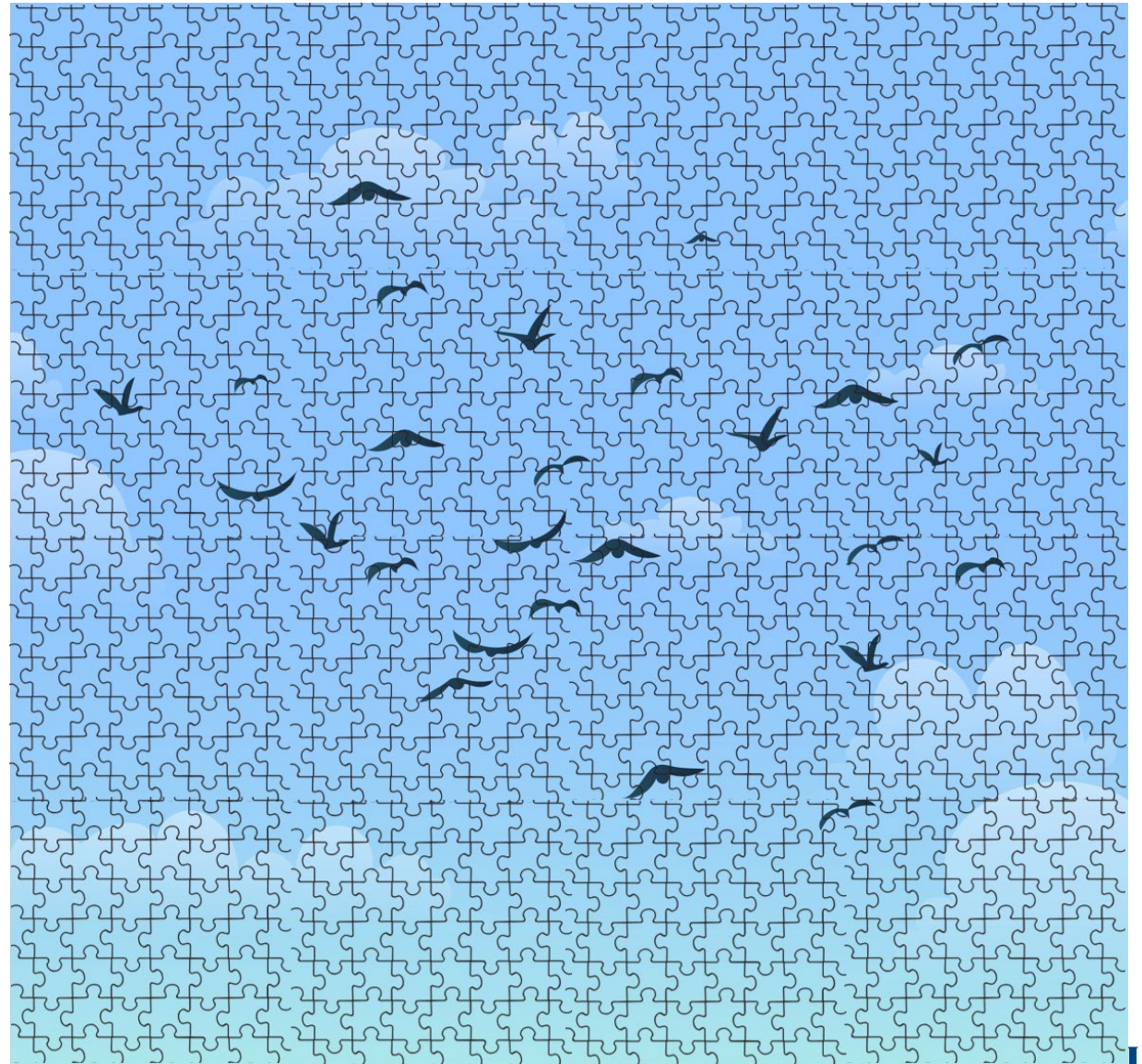
# Target species

- **Artiodactyla**
  - Pig, camel, whale
- **Giraffidae**
  - Giraffe
- **Cervidae**
  - Moose, Eld's deer, muntjac
- **Bovinae**
  - Cattle, gaur x cattle, water buffalo, bison x cattle, banteng
- **Antilopinae**
  - Dama gazelle
- **Hippotraginae**
  - Scimitar-horned oryx, roan antelope
- **Caprinae**
  - Sheep, bighorn x sheep, goat, musk ox



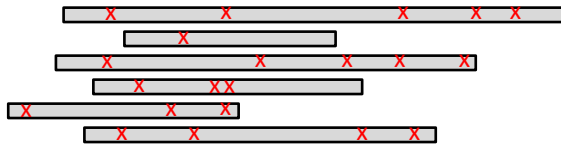
# Assembling T2T genomes

- **Mammalian (and many other) genomes are highly repetitive**
- **Repetitive DNA causes assembly errors when reads are shorter than repeats**

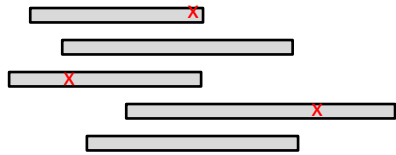


# Assembling T2T genomes

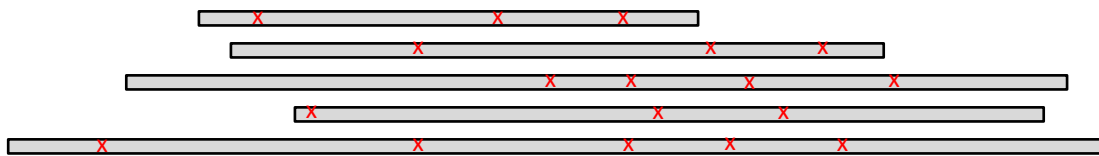
PacBio CLR – 10-80Kb



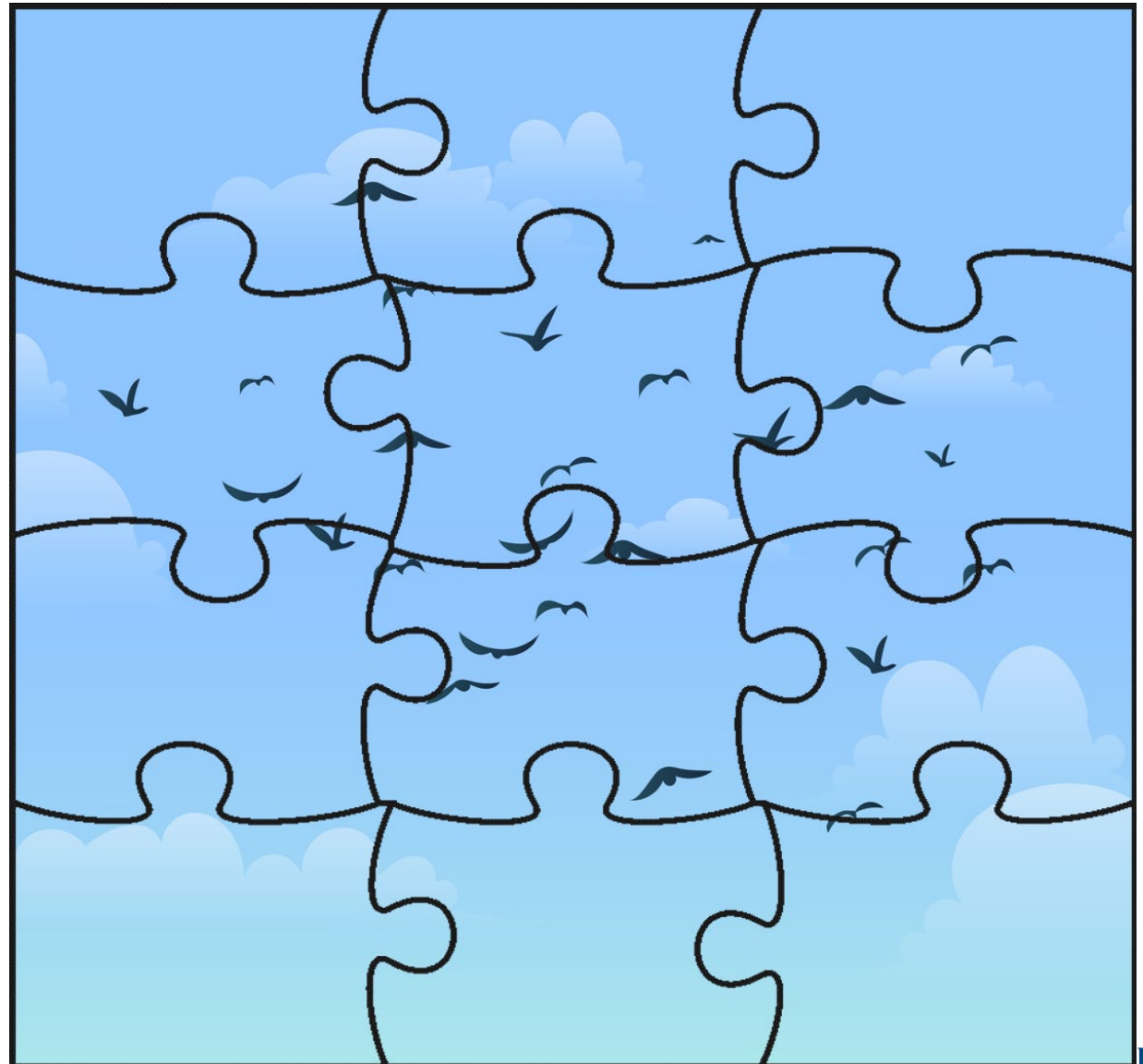
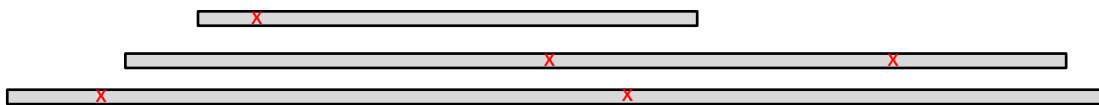
PacBio HiFi – 10-25Kb



ONT – 10-300Kb+



ONT duplex – 10-300Kb+

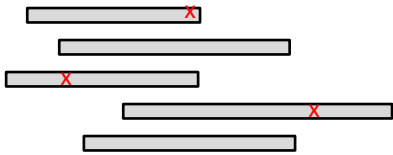




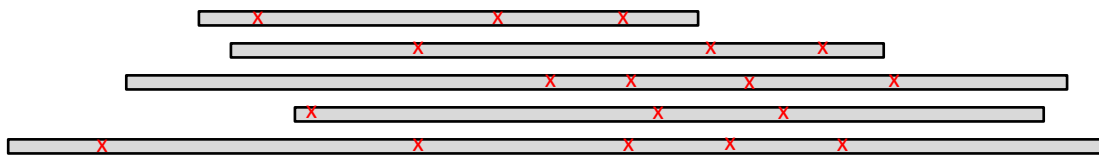
# Assembling T2T genomes

PacBio CLR – 10-80Kb

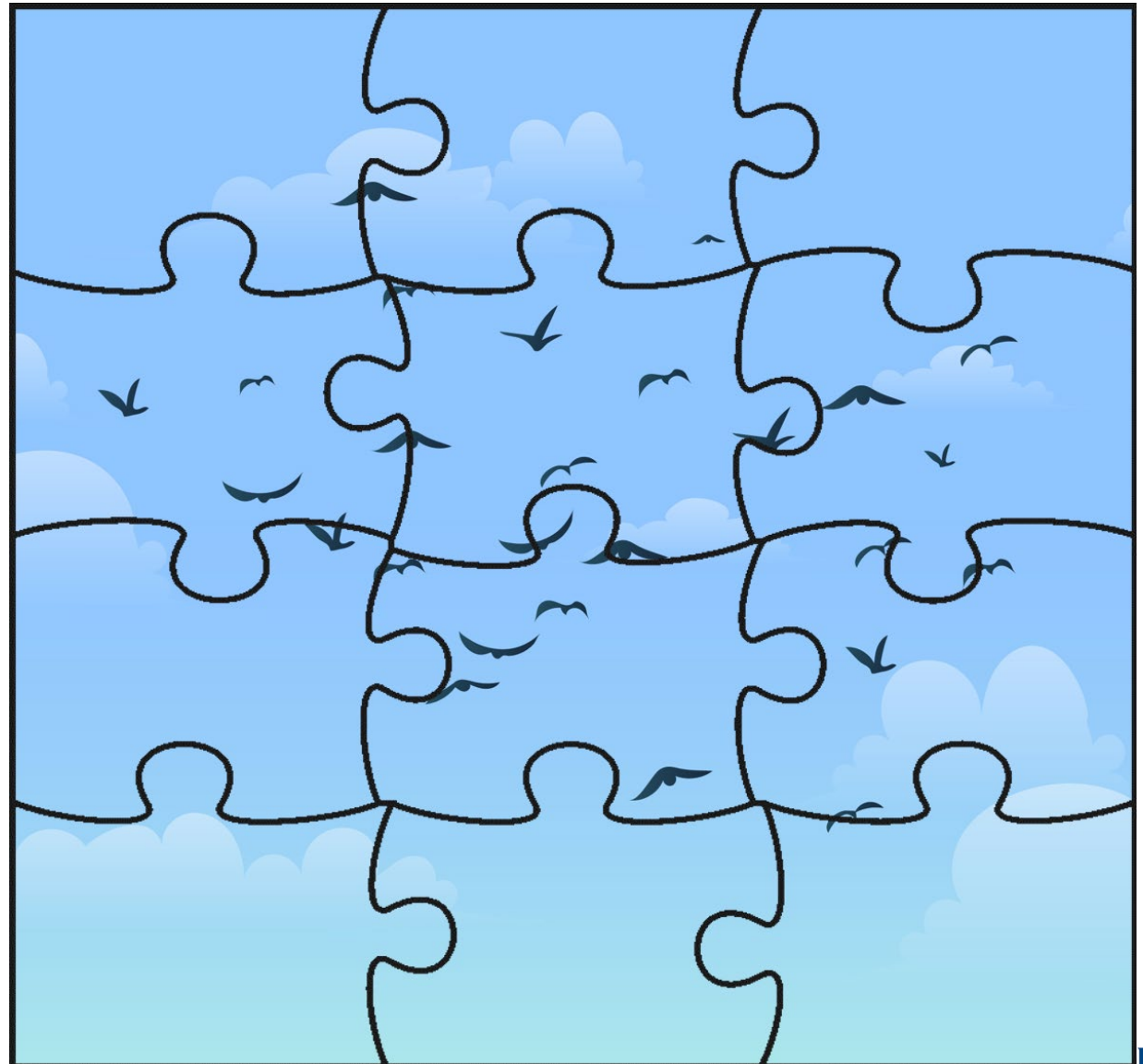
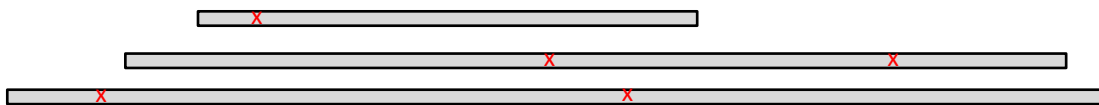
PacBio HiFi – 10-25Kb



ONT – 10-300Kb+



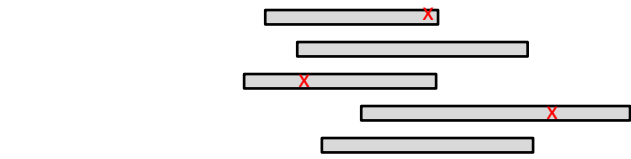
ONT duplex – 10-300Kb+



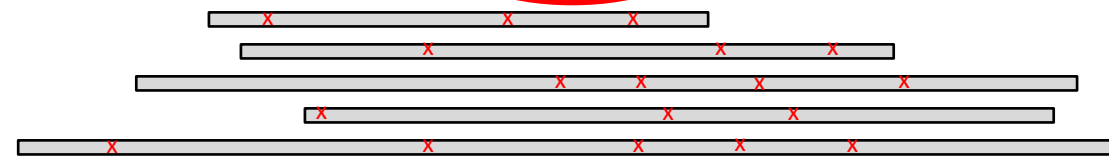
# Assembling T2T genomes

PacBio CLR – 10-80Kb

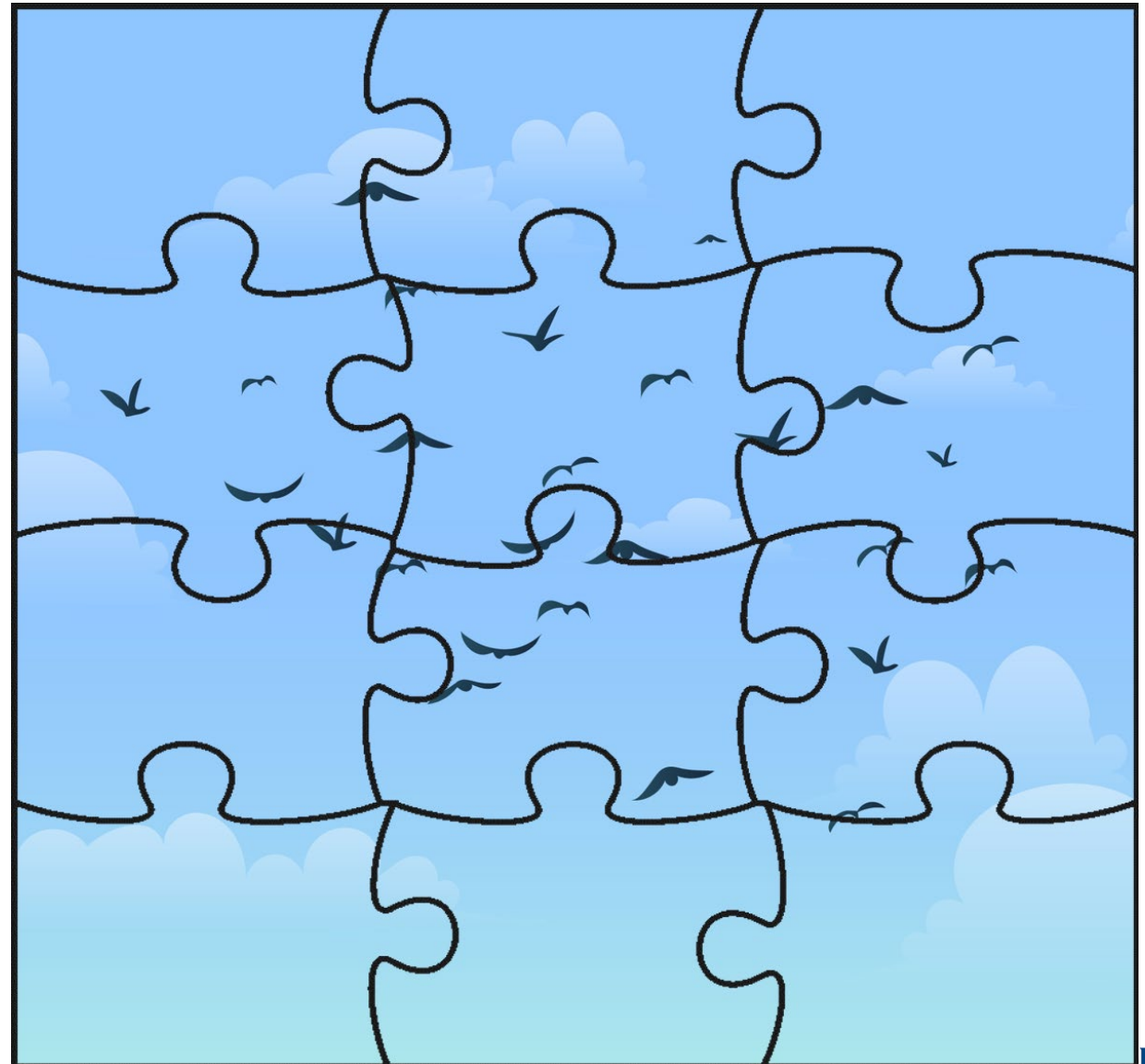
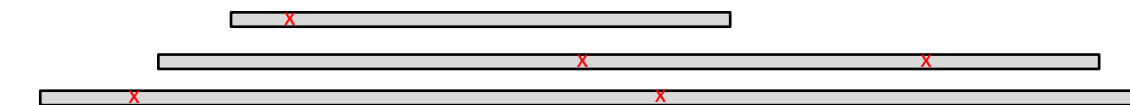
➔ PacBio HiFi – 10-25Kb



ONT – 10-300Kb+

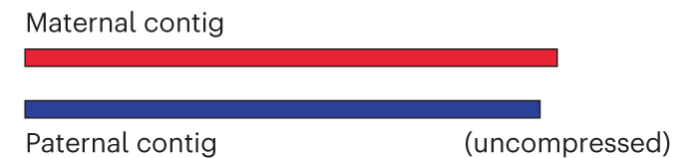
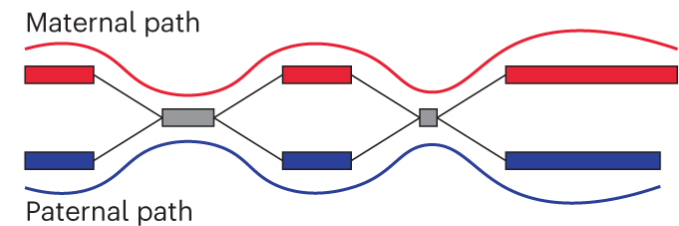
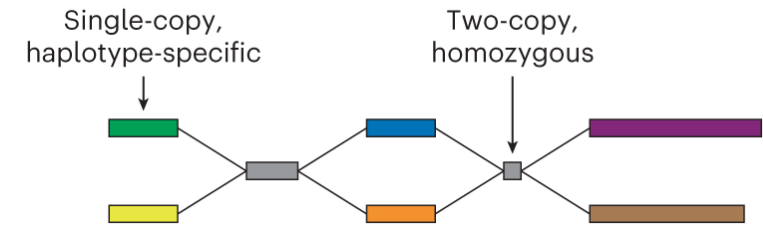
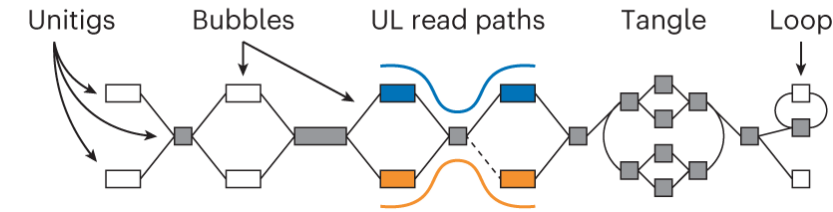
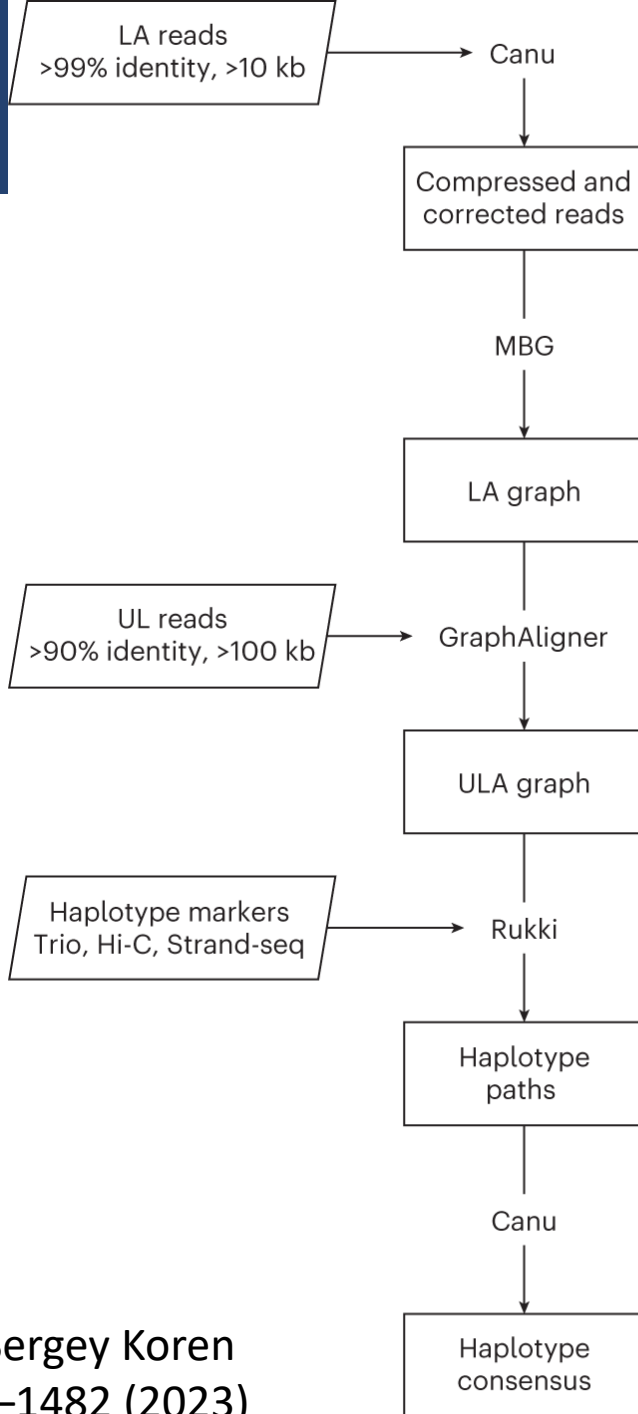


➔ ONT duplex – 10-300Kb+



# Assembling T2T

- 50x coverage of long-accurate data
- 50x coverage of ultra-long ONT
- HiC or accurate reads of parents for phasing



Mikko Rautiainen, ..., & Sergey Koren  
Nature Biotech 41, 1474–1482 (2023)

# Data generation

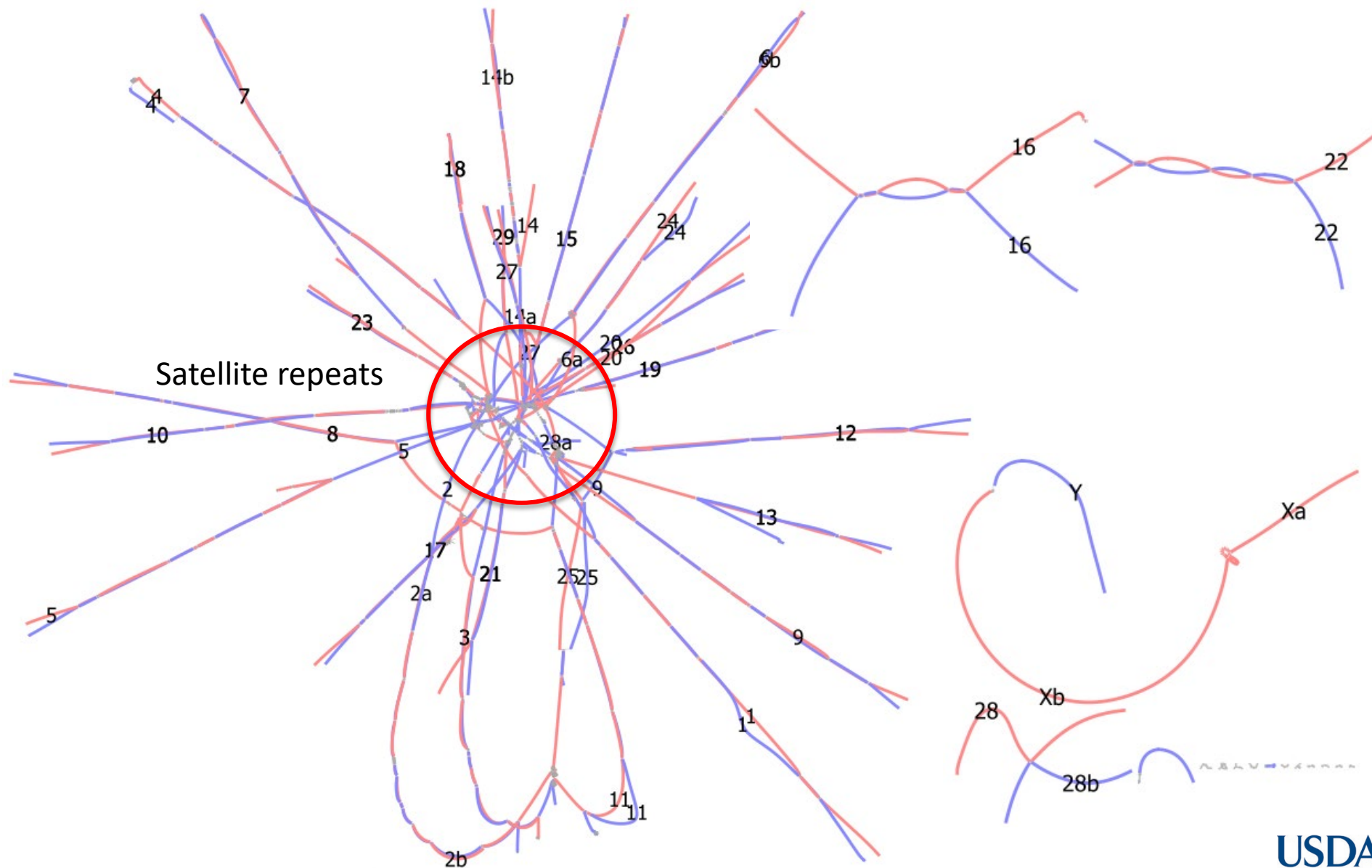
	<u>HiFi</u>	<u>ONT (total/UL)</u>
<b>Pig (Hampshire x composite)</b>	65x	325/29x
<b>Giraffe</b>	56x	86/5x
<b>Moose</b>	18x	195/16x
<b>Cattle (Wagyu x Charolais)</b>	110x	208/43x
<b>Gaur x cattle (Piedmontese)</b>	111x	218/31x
<b>Bison x cattle (Piedmontese)</b>	72x	245/24x
<b>Bison x cattle (Simmental)</b>	74x	284/18x
<b>Banteng</b>	-	61/7x
<b>Water Buffalo</b>	71x	307/27x
<b>Sheep (Churro x Friesian)</b>	86x	214/30x
<b>Bighorn x sheep (Polypay)</b>	60x	245/28x
<b>Goat (Kiko x Saanen)</b>	62x	301/32x
<b>Musk Ox</b>	48x	302/12x

# Assembly graphs

## Cattle

### Wagyu x Charolais

	Wagyu	Charolais
Contig/Scaff N50 Mb	108/108	97/100
Total gaps	17	25
Sat gaps	12	11
rDNA gaps	1	-
near T2T	28/30	28/30
placed rDNA arrays	5	5



# rDNA arrays

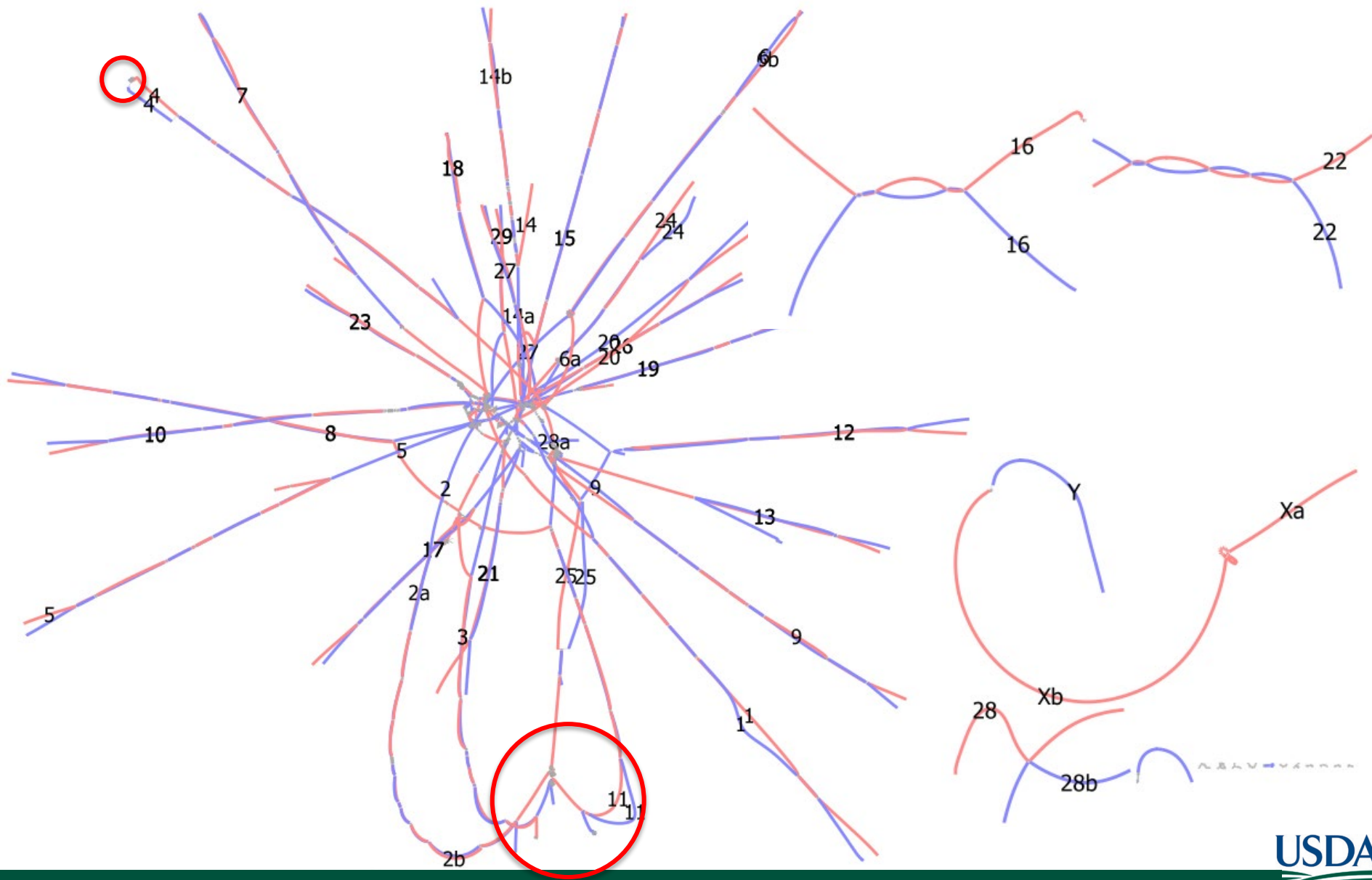
## Cattle

### Wagyu x Charolais

	Wagyu	Charolais
Contig/Scaff N50 Mb	108/108	97/100
Total gaps	17	25
Sat gaps	12	11
rDNA gaps	1	-
near T2T	28/30	28/30
placed rDNA arrays	5	5

chromosomes

2, 3, 4, 11, 25

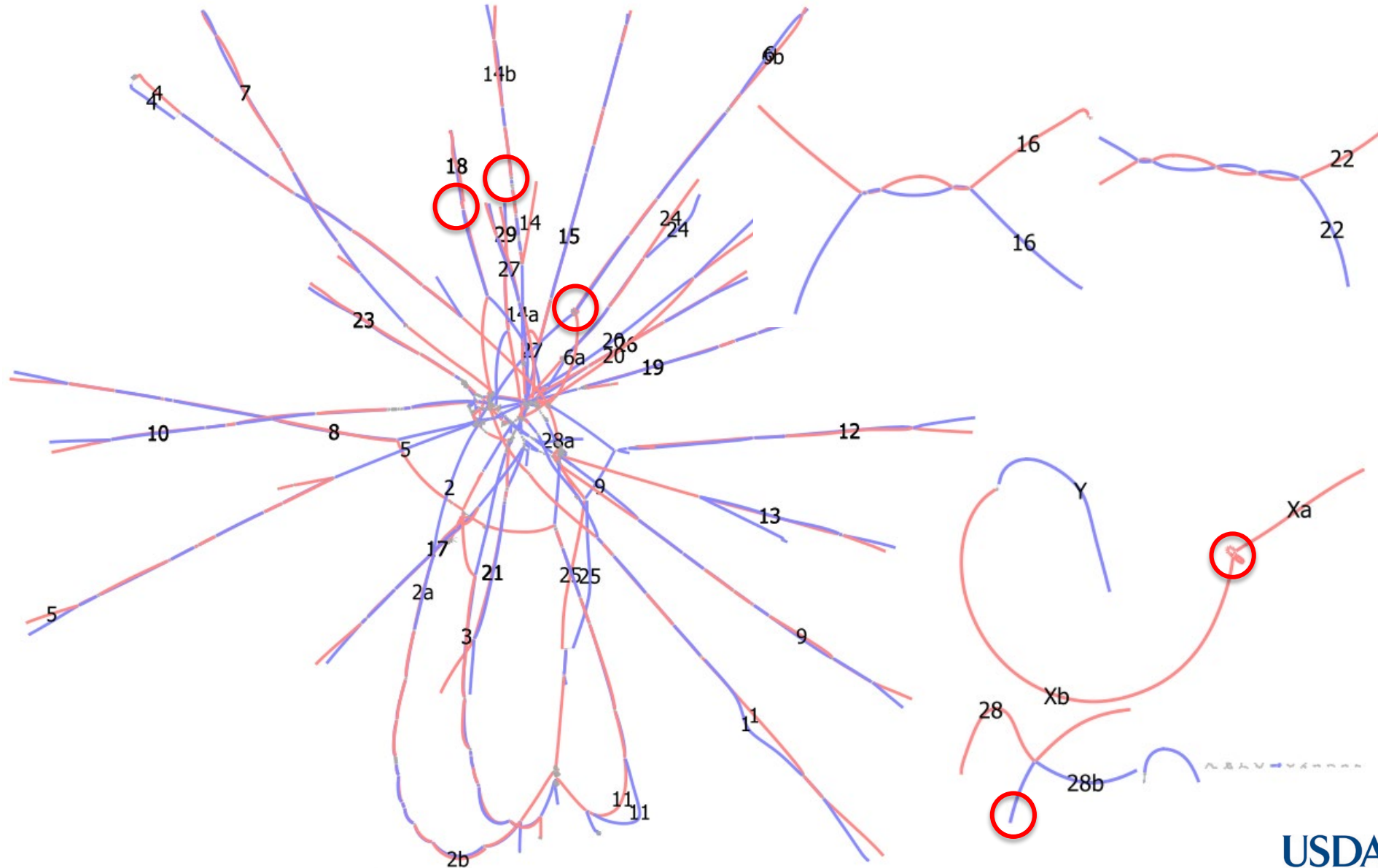


# Gaps

## Cattle

### Wagyu x Charolais

	Wagyu	Charolais
Contig/Scaff N50 Mb	108/108	97/100
Total gaps	17	25
Sat gaps	12	11
rDNA gaps	1	-
near T2T	28/30	28/30
placed rDNA arrays	5	5

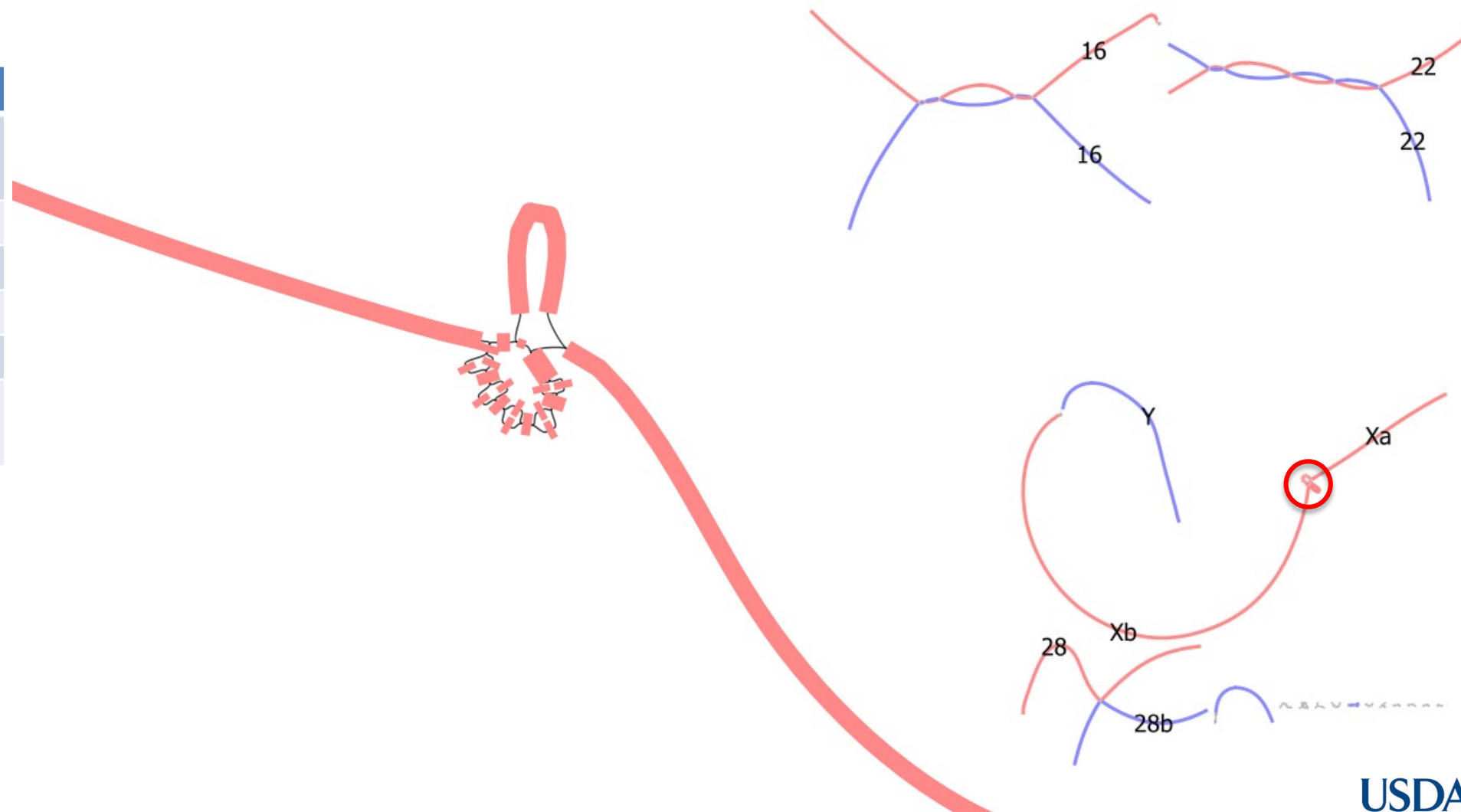


# Gaps

## Cattle

### Wagyu x Charolais

	Wagyu	Charolais
Contig/Scaff N50 Mb	108/108	97/100
Total gaps	17	25
Sat gaps	12	11
rDNA gaps	1	-
near T2T	28/30	28/30
placed rDNA arrays	5	5



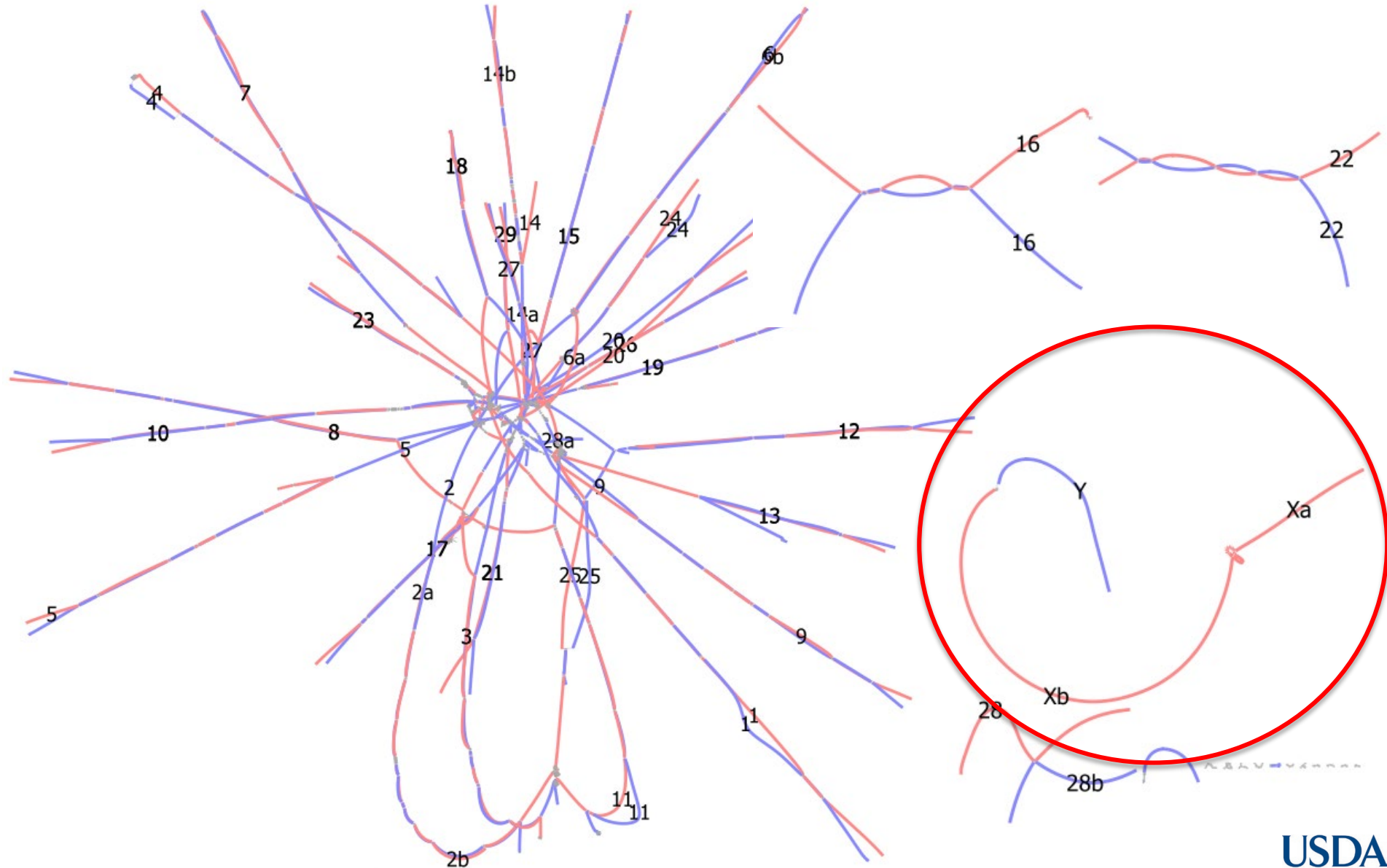


# Sex chromosomes

## Cattle

### Wagyu x Charolais

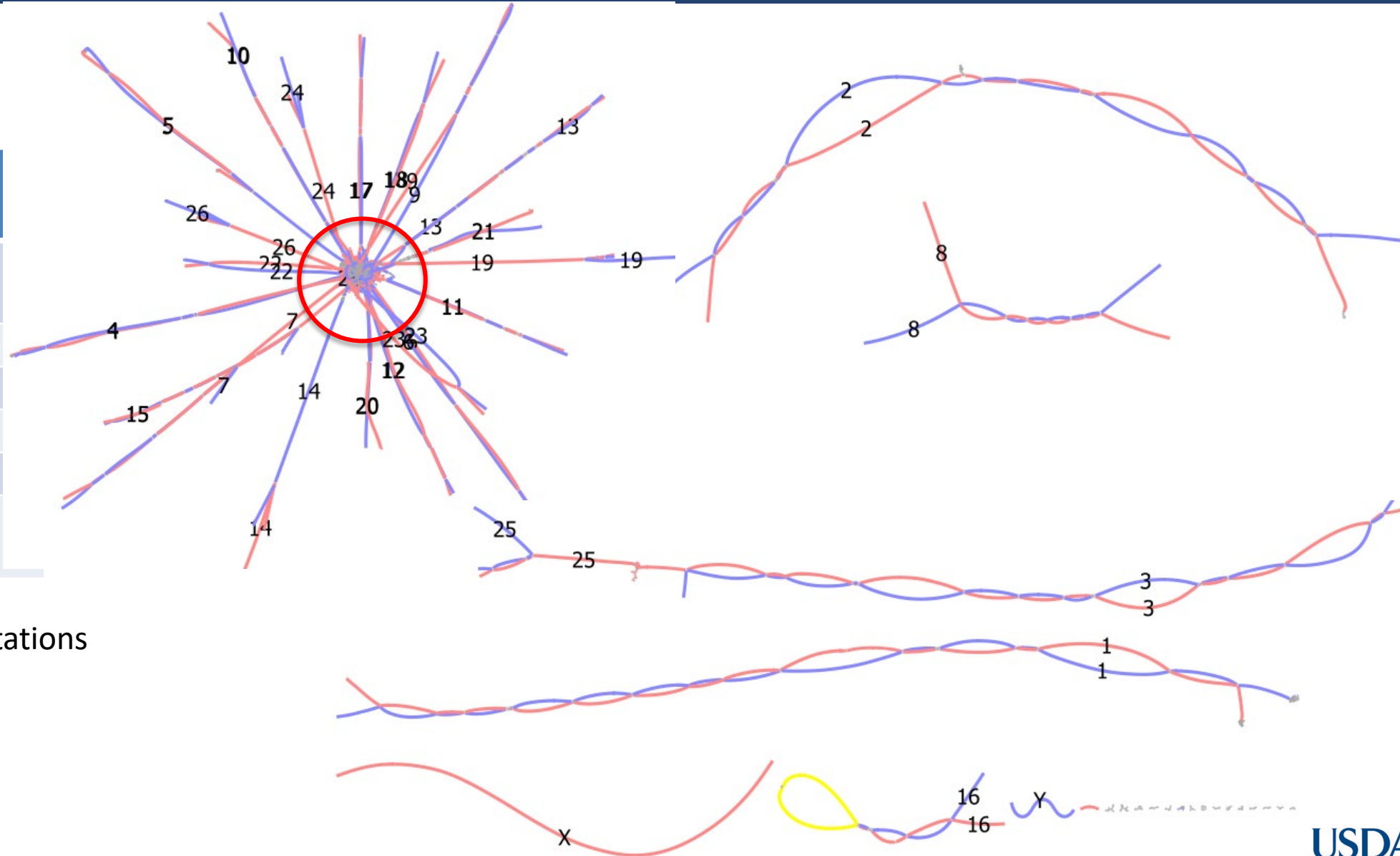
	Wagyu	Charolais
Contig/Scaff N50 Mb	108/108	97/100
Total gaps	17	25
Sat gaps	12	11
rDNA gaps	1	-
near T2T	28/30	28/30
placed rDNA arrays	5	5



# Satellite repeats

## Sheep Churro x Friesian

	Navajo Churro	East Friesian
Contig/Scaff N50 Mb	93/96	97/100
Total gaps	9	16
Sat gaps	5	6
rDNA gaps	1	4
near T2T*	21/27	20/27
placed rDNA arrays	4	5

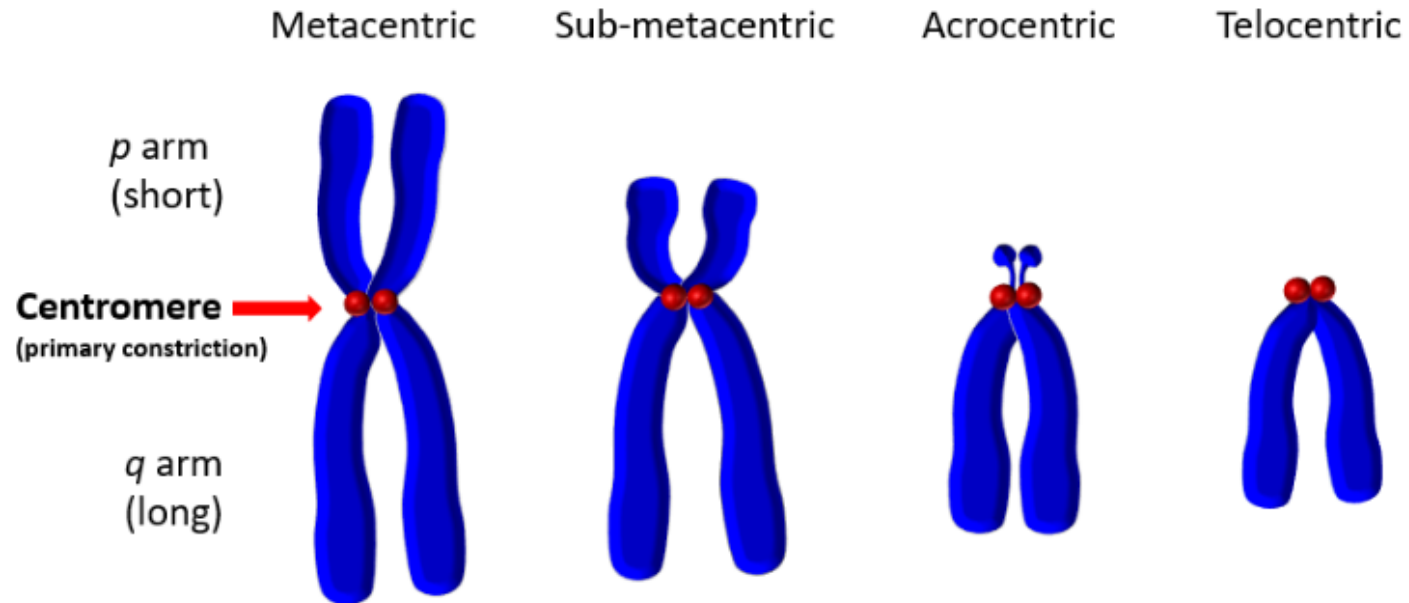


\*missing satellite annotations

# Acrocentric vs Telocentric

Classification of chromosomes based on morphology:

Types of chromosomes based on position of centromere:

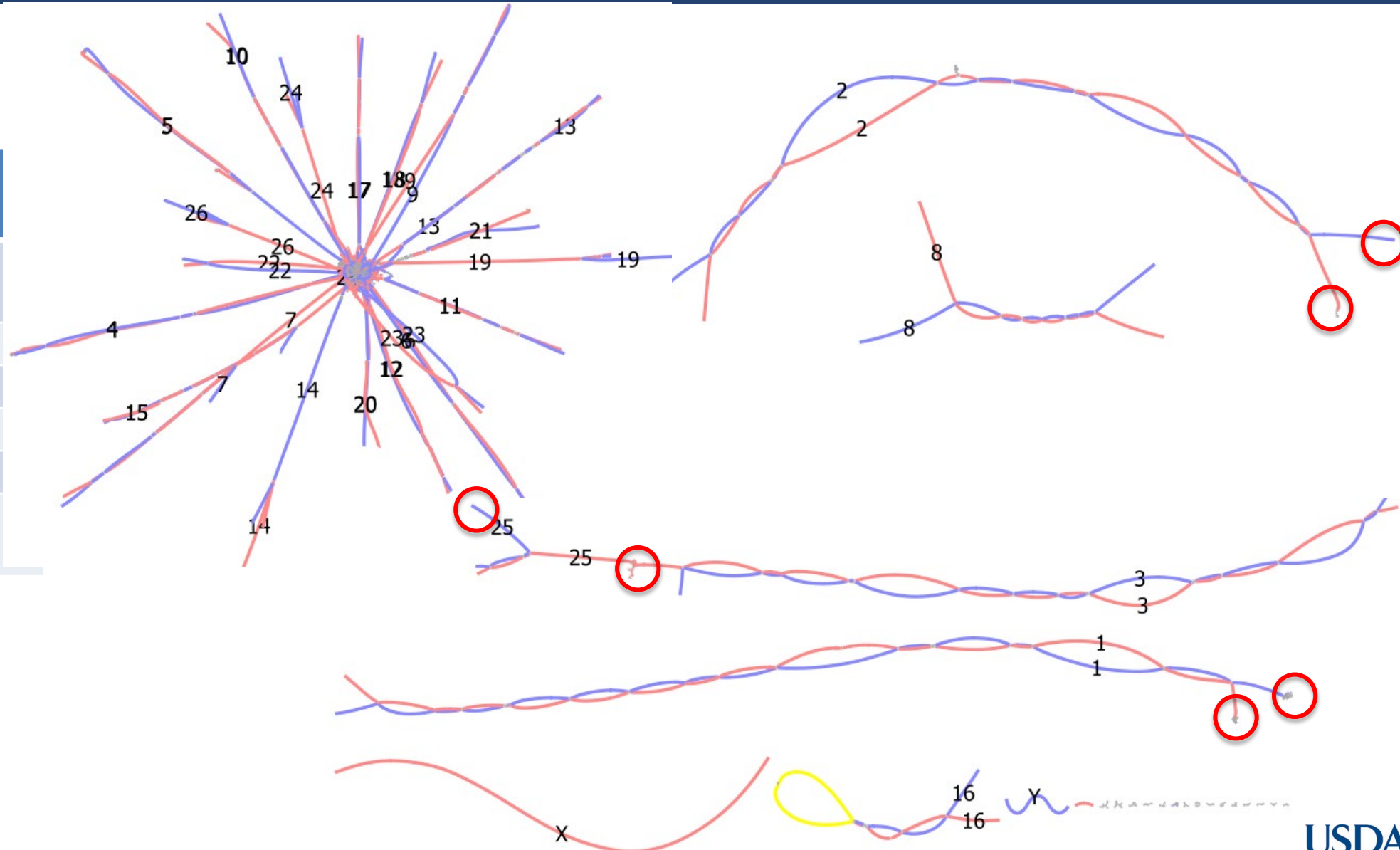


Tamara Potapova  
Stowers Institute

# rDNA arrays

## Sheep Churro x Friesian

	Navajo Churro	East Friesian
Contig/Scaff N50 Mb	93/96	97/100
Total gaps	9	16
Sat gaps	5	6
rDNA gaps	1	4
near T2T	21/27	20/27
placed rDNA arrays	4*	5



chromosomes

1, 2, 3\*, 4, 25

# Resolving rDNA

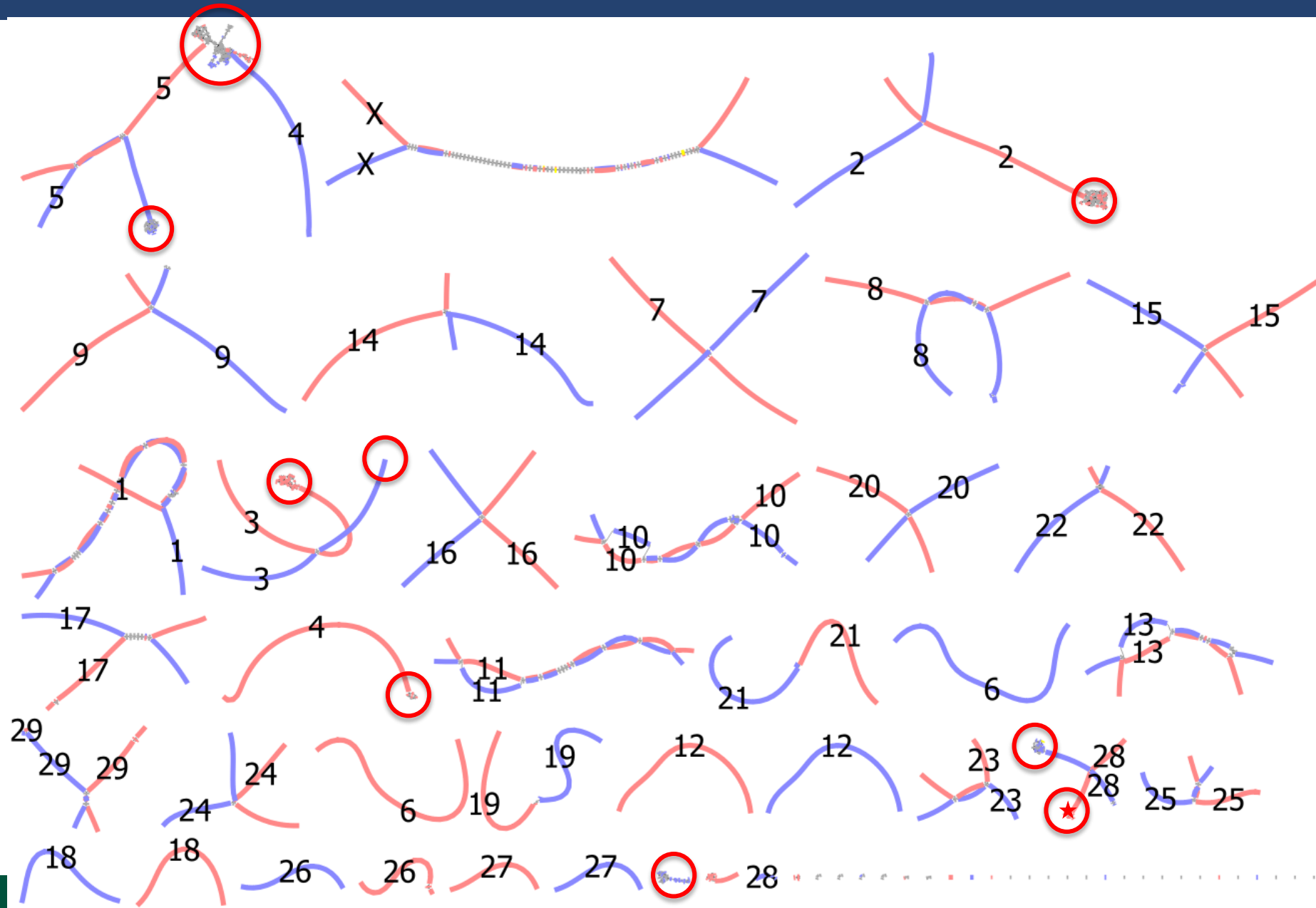
## Goat

### Kiko x Saanen

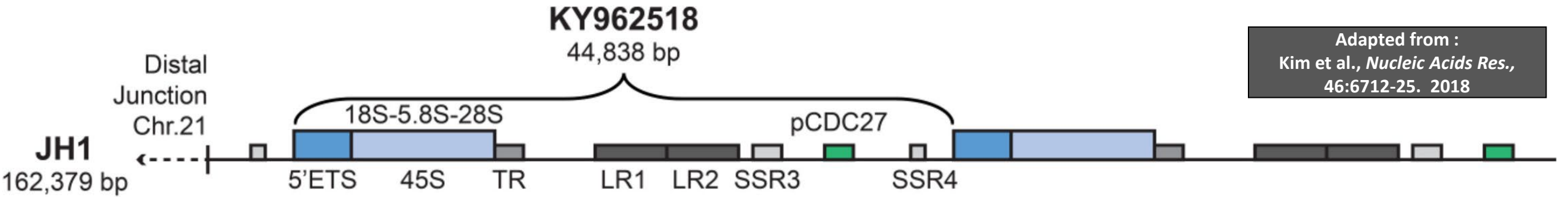
	Kiko	Saanen
Contig/Scaff N50 Mb	101/101	101/101
Total gaps	5	6
Sat gaps	-	2
rDNA gaps	2	2
near T2T	27/30	25/30
placed rDNA arrays	4	5 (1T2T)★

### chromosomes

2, 3, 4, 5, 28



# rDNA



haplotypel-0000007	127278482	8362	13296	LSU-rRNA_Hsa	rRNA	(0)	5035	1	5.3
haplotypel-0000007	127278482	16081	17949	SSU-rRNA_Hsa	rRNA	(0)	1869	1	0.5
haplotypel-0000007	127278482	46841	51775	LSU-rRNA_Hsa	rRNA	(0)	5035	1	5.3
haplotypel-0000007	127278482	54571	56439	SSU-rRNA_Hsa	rRNA	(0)	1869	1	0.5
haplotypel-0000007	127278482	85712	90646	LSU-rRNA_Hsa	rRNA	(0)	5035	1	5.3
haplotypel-0000007	127278482	93417	95285	SSU-rRNA_Hsa	rRNA	(0)	1869	1	0.6

← → ↻ <https://github.com/maickrau/ribotin>

Product Solutions Open Source

maickrau / ribotin Public

<> Code Issues Pull requests Actions Projects Security Insights

Mikko Rautiainen

# Resolving rDNA

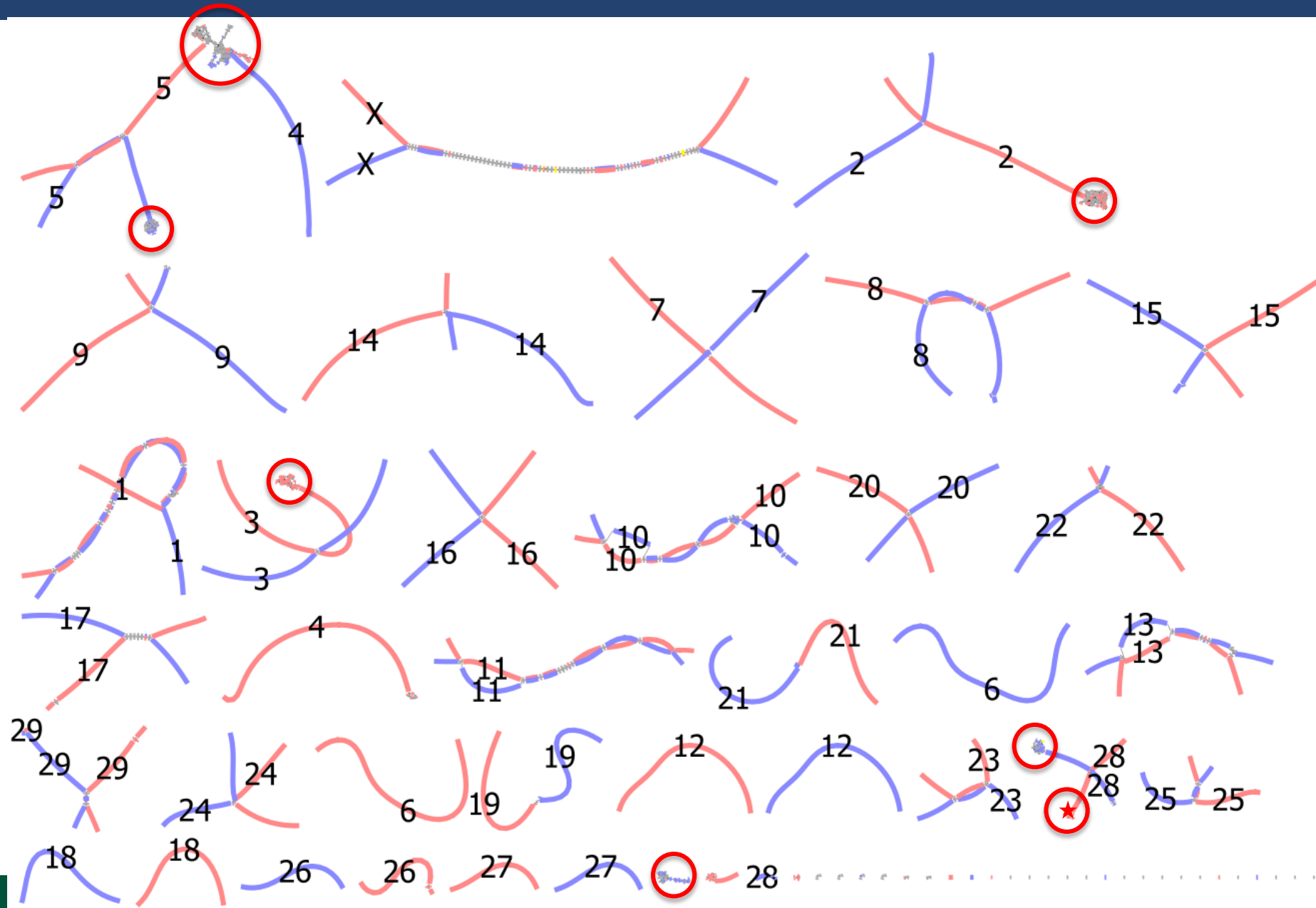
## Goat Kiko x Saanen

	Kiko	Saanen
Contig/Scaff N50 Mb	101/101	101/101
Total gaps	5	6
Sat gaps	-	2
rDNA gaps	2	2
near T2T	27/30	25/30
placed rDNA arrays	4	5 (1T2T)*

chromosomes

2, 3, 4, 5, 28

**Ribotin identified  
7 clusters**



# Resolving rDNA

## Goat

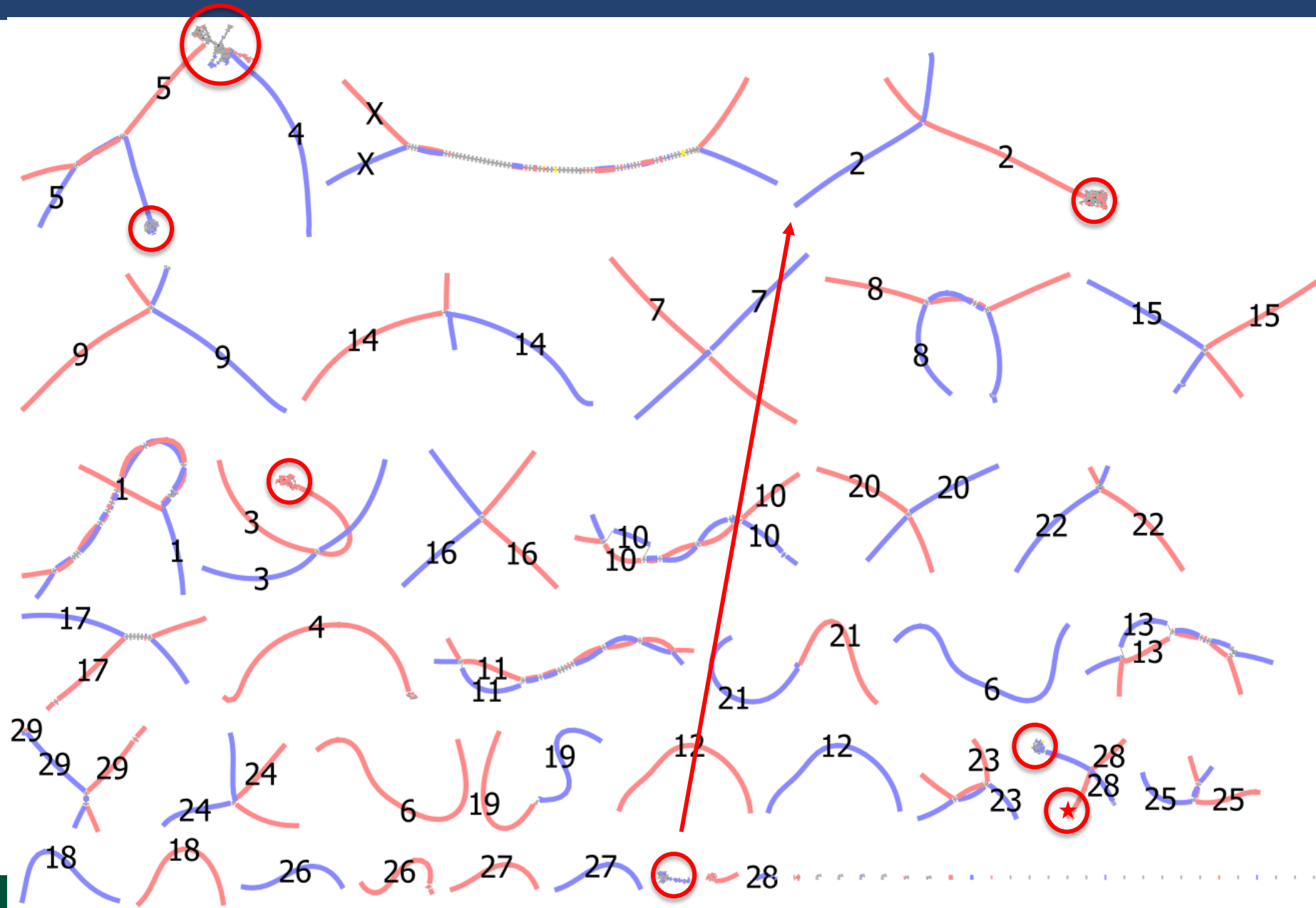
### Kiko x Saanen

	Kiko	Saanen
Contig/Scaff N50 Mb	101/101	101/101
Total gaps	5	6
Sat gaps	-	2
rDNA gaps	2	2
near T2T	27/30	25/30
placed rDNA arrays	4	5 (1T2T)*

### chromosomes

2, 3, 4, 5, 28

**Ribotin identified  
8/10 clusters**





# Resolving rDNA

## Goat

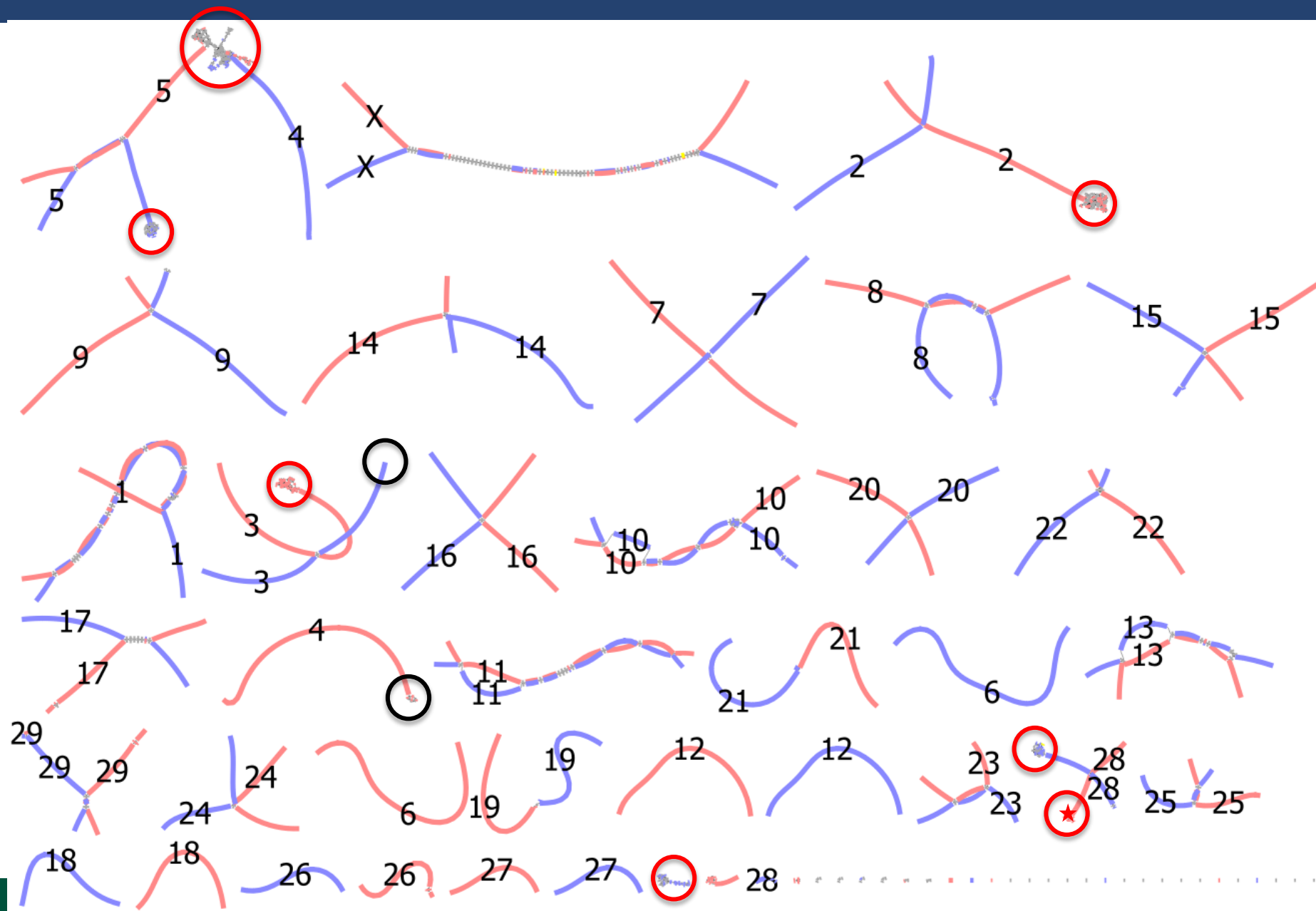
### Kiko x Saanen

	Kiko	Saanen
Contig/Scaff N50 Mb	101/101	101/101
Total gaps	5	6
Sat gaps	-	2
rDNA gaps	2	2
near T2T	27/30	25/30
placed rDNA arrays	4	5 (1T2T)*

### chromosomes

2, 3, 4, 5, 28

**Only missed low copy number arrays**



# Resolving rDNA

## Goat

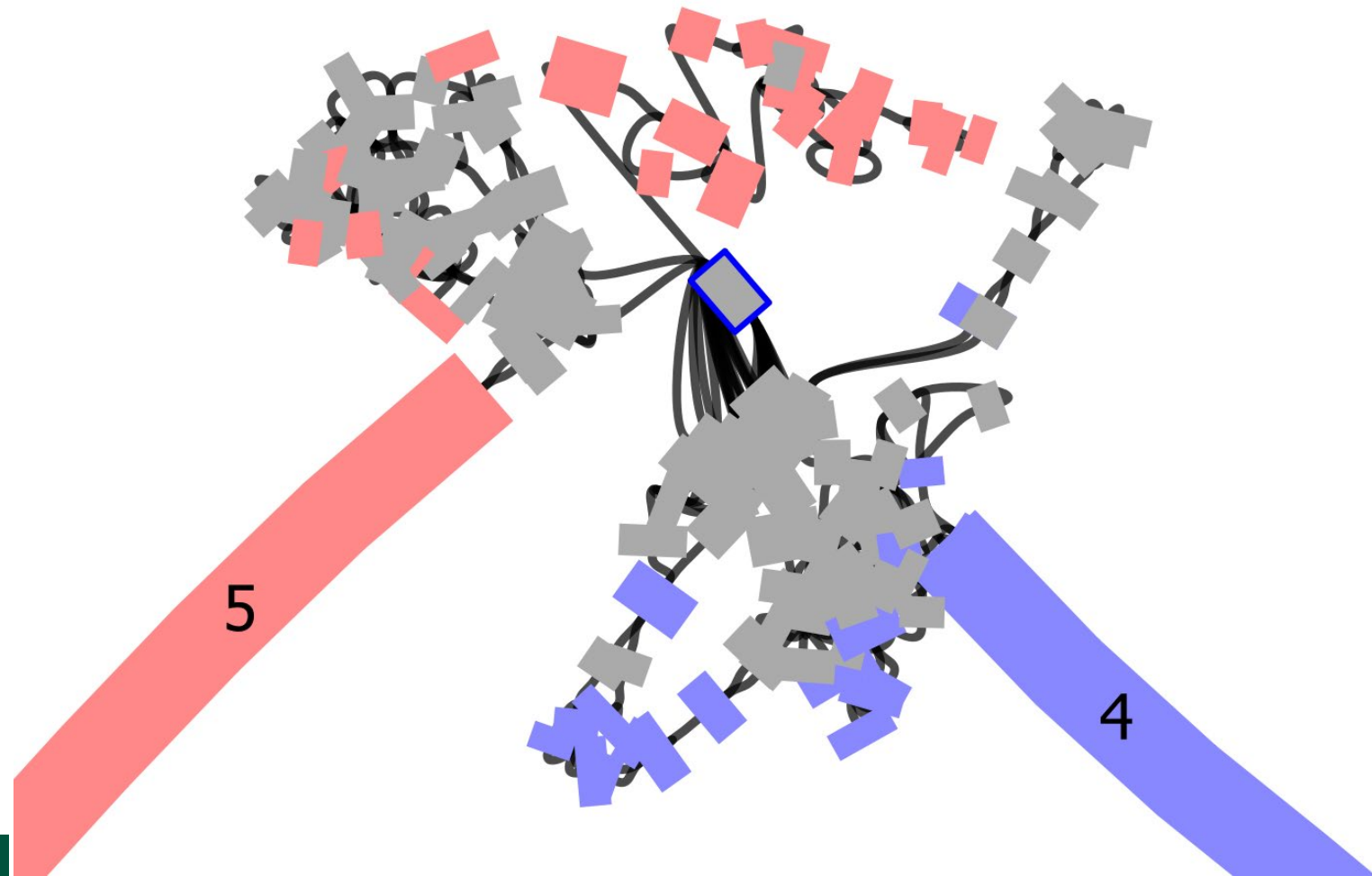
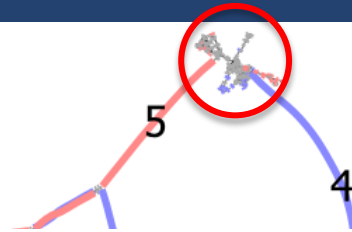
### Kiko x Saanan

	Kiko	Saanen
Contig/Scaff N50 Mb	101/101	101/101
Total gaps	5	6
Sat gaps	-	2
rDNA gaps	2	2
near T2T	27/30	25/30
placed rDNA arrays	4	5 (1T2T)*

chromosomes

2, 3, 4, 5, 28

**Manual splitting**

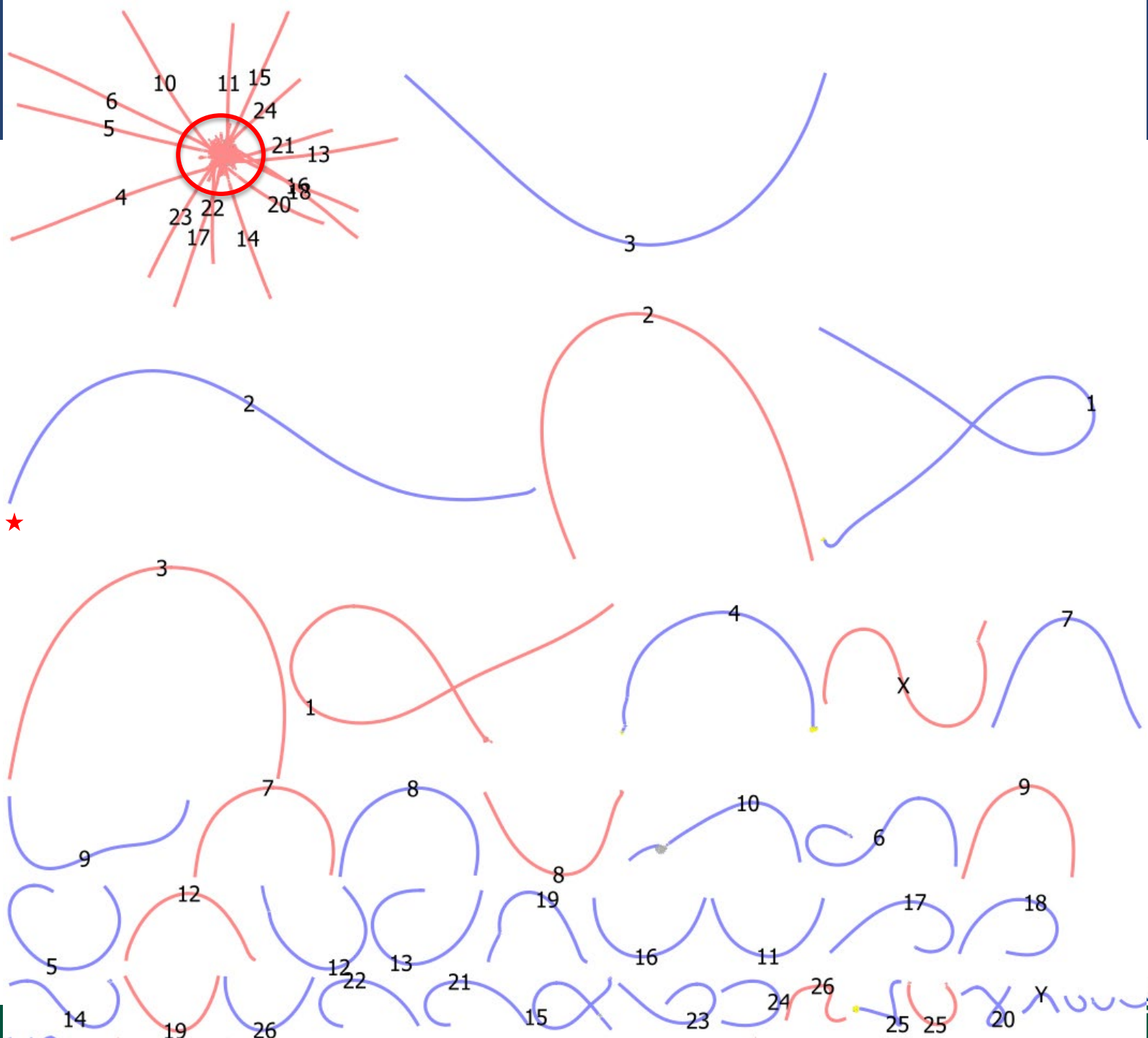


# Resolving Sats

## Bighorn x sheep (Polypay)

	Bighorn	Polypay
Contig/Scaff N50 Mb	103/103	95/95
Total gaps	8	16
Sat gaps	5	10
rDNA gaps	1	2
near T2T*	19/27	23/27
placed rDNA arrays	4	5 (1T2T)★

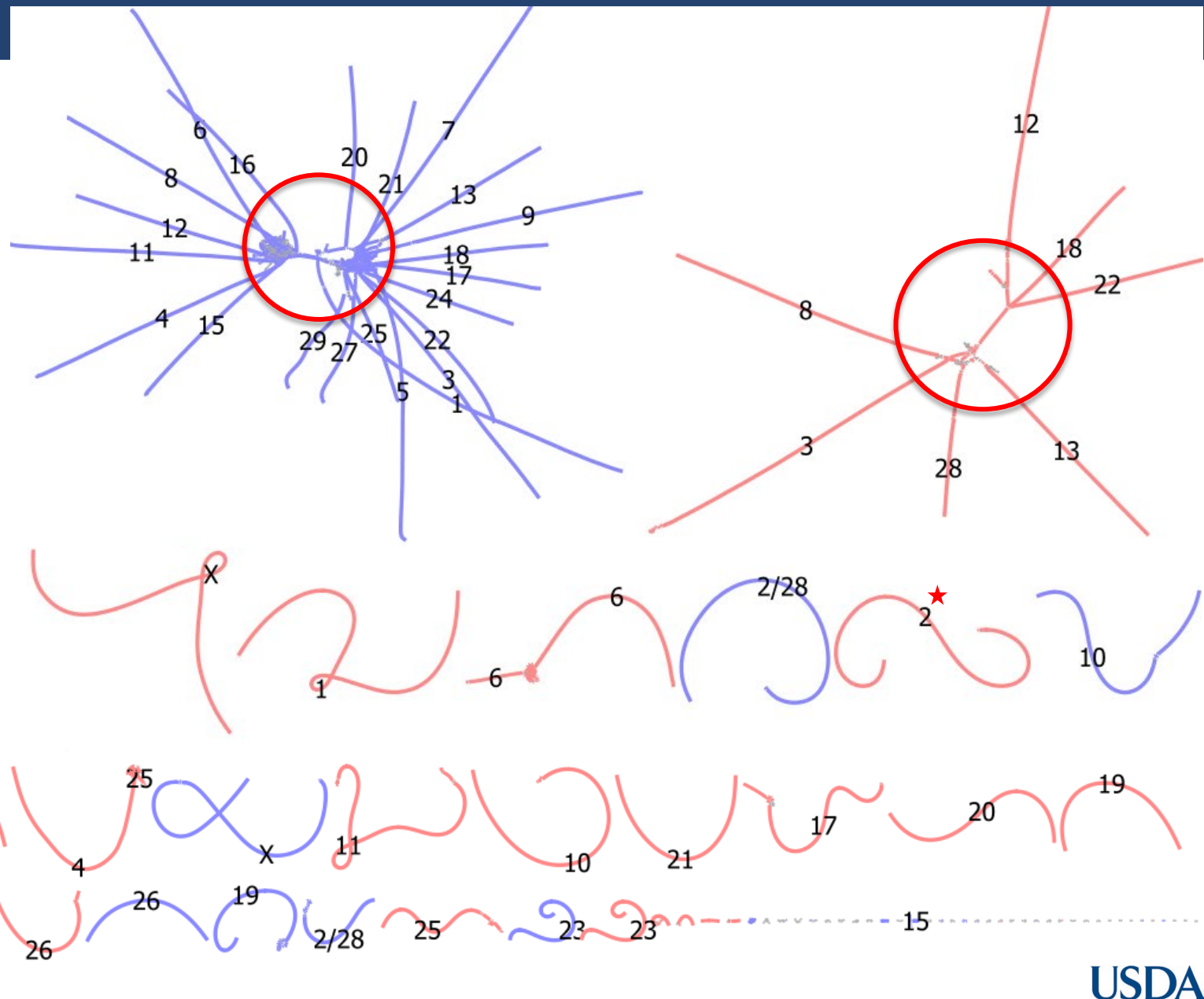
\*missing satellite annotations



# Resolving Sats

## Gaur x cattle (Piedmontese)

	Gaur	Pied
Contig/Scaff N50 Mb	88/108	112/112
Total gaps	14	16
Sat gaps	8	7
rDNA gaps	-	3
near T2T	27/29	28/30
placed rDNA arrays	4	5 (1T2T) <sup>★</sup>



# THANKS!

Tim Smith, USDA – Brenda Murdoch, U. Idaho – Stephanie McKay, U. Missouri

Adam Phillippy

Sergey Koren

Arang Rhie

Brandon Pickett

Mikko Rautiainen

Rachel O'Neill

Patrick Grady

Klaus-Peter Koepfli

Karen Miga

Tamara Potapova

Emily Clark

Noelle Cockett

Tracy Hadfield

Michelle Mousel

Mike Heaton

Brian Vander Ley

Derek Bickhart

Yana Safonova

Bob Schnabel

Jim Reecy

Chris Elsik

James Koltes

Darren Hagen

George Liu

Curt Van Tassell

Ted Kalbfleisch

Chris Tuggle

Temitayo Olagunju

Morgan Stegemiller

Shangqian Xie

Andrew Hess

Mazdak Salavati

Shannon Clarke

Rudiger Brauning

Erich Jarvis

Holly Neibergs

Wai Yee Low

Paulene Pineda

John Williams

Paolo Ajmone Marsan

Alex Leonard

Hubert Pausch

Mention of trade names or commercial products in this presentation is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture