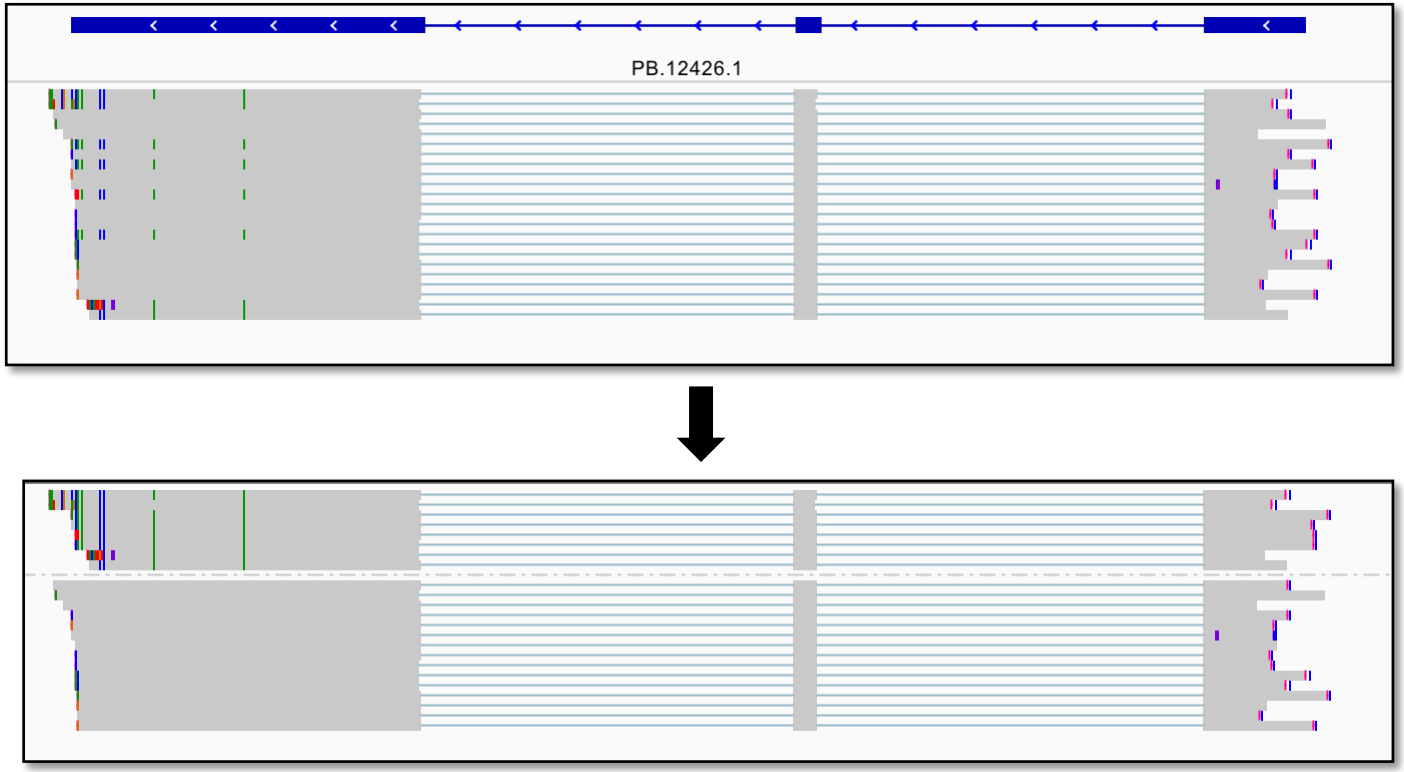
PACIFIC BIOSCIENCES®

IsoPhase: Isoform-Level Phasing

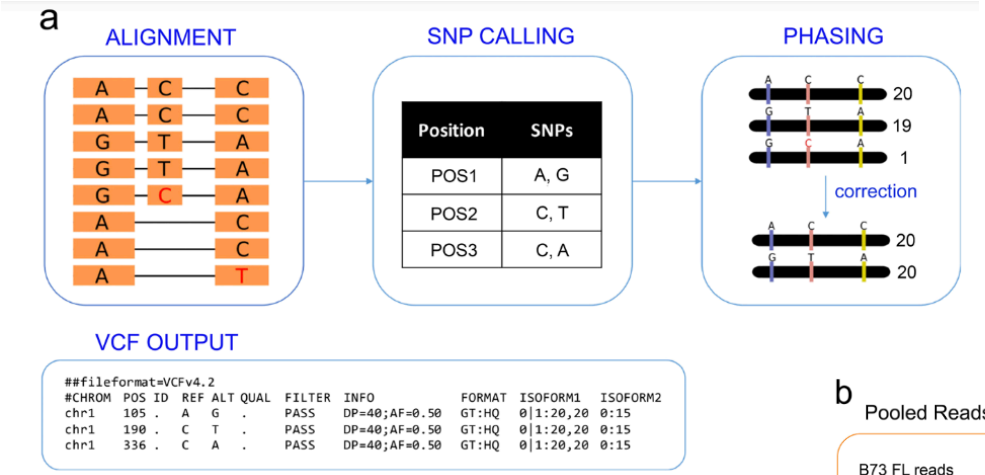
Elizabeth Tseng, Principal Scientist, PacBio

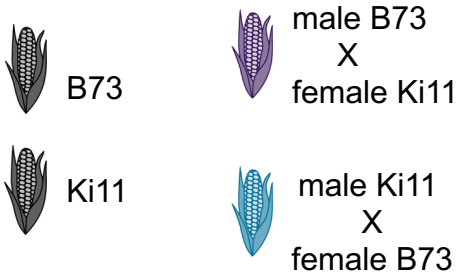
@Magdoll

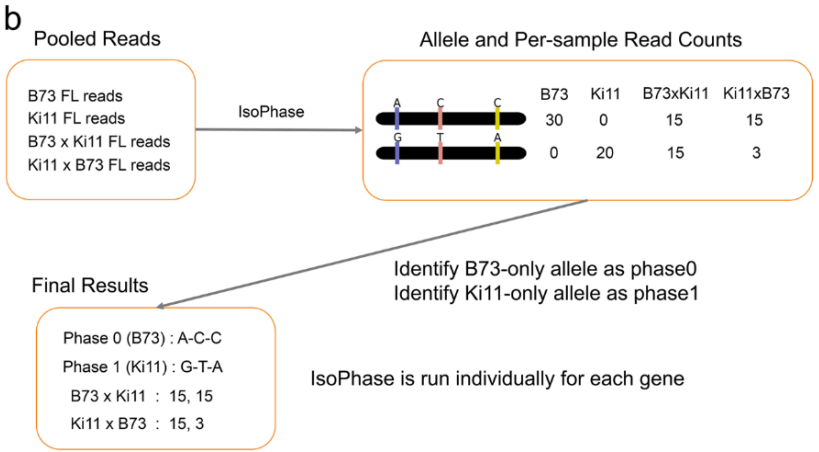
ISOFORM-LEVEL PHASING USING ISO-SEQ READS



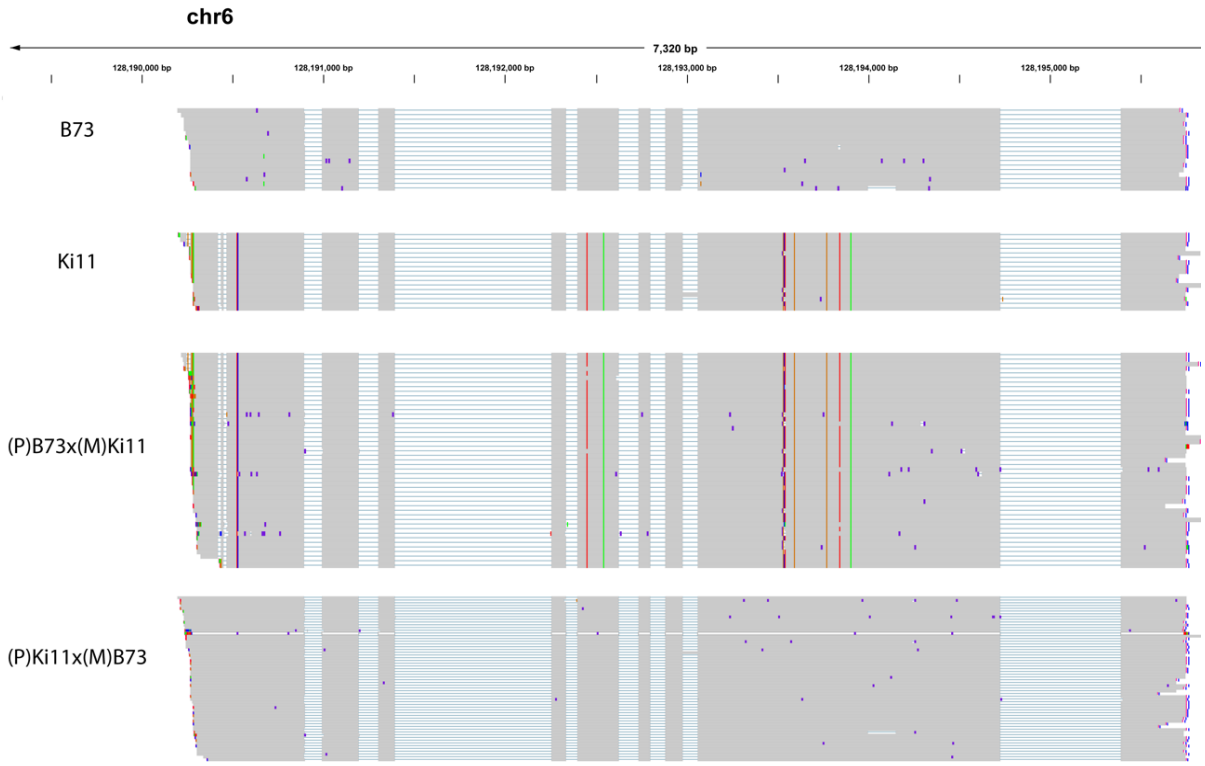
ISOPHASE: ISOFORM-LEVEL PHASING





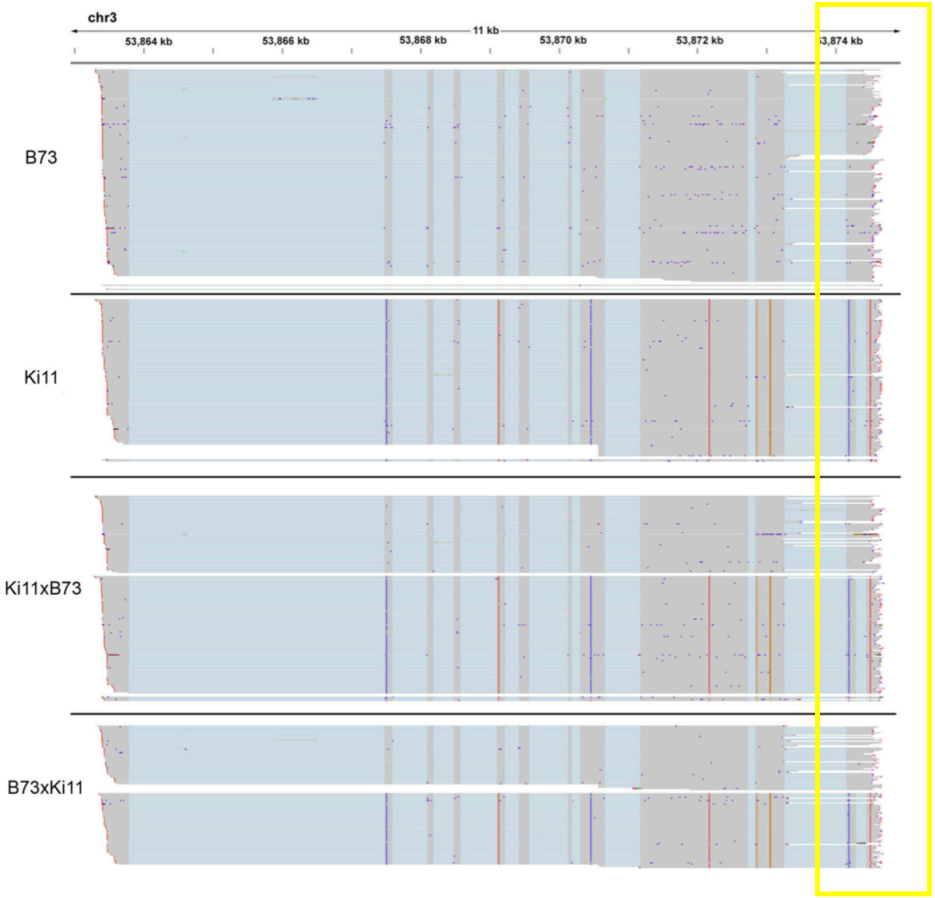


MATERNAL IMPRINTING IN MAIZE



Wang et al “Variant Phasing and Haplotypic Expression from Long-Read Sequencing in Maize.” *Communications Biology* (2020)

PARENTAL-SPECIFIC ISOFORM EXPRESSION IN MAIZE



Parents express only one isoform

F1s express parent-specific isoforms

(3' UTR difference)

Wang et al "Variant Phasing and Haplotypic Expression from Long-Read Sequencing in Maize." *Communications Biology* (2020)

PHASING BEYOND DIPLOID – A CHALLENGE

Tetraploid potato poses phasing challenges

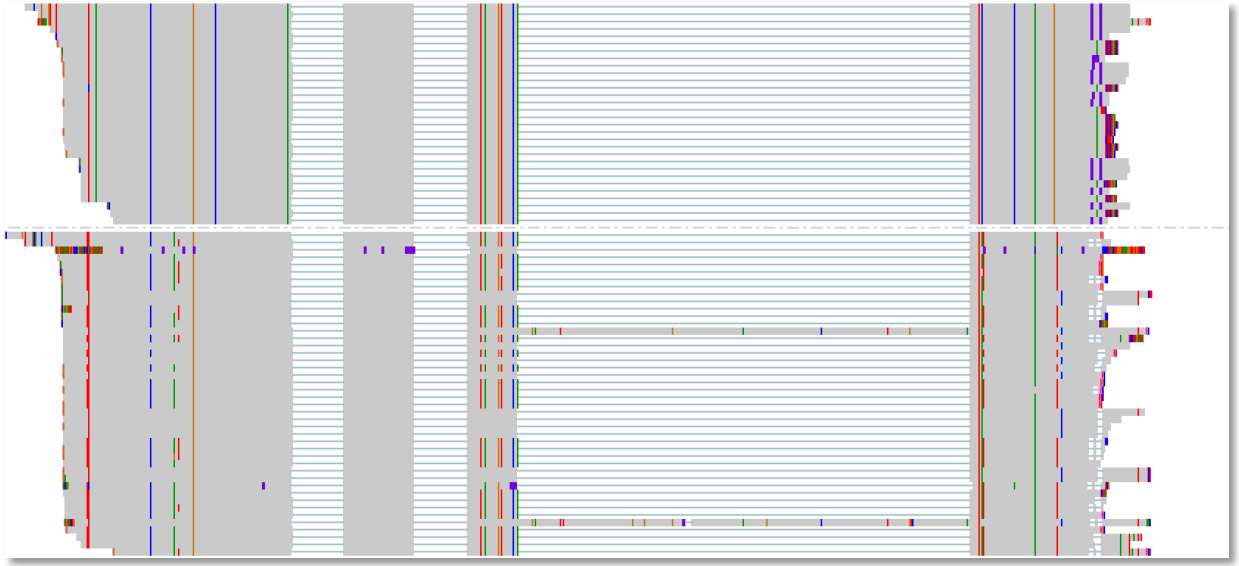
Each variant position has two SNPs

IsoPhase estimated three alleles, not four

Marko Petek

& Kristina Gruden

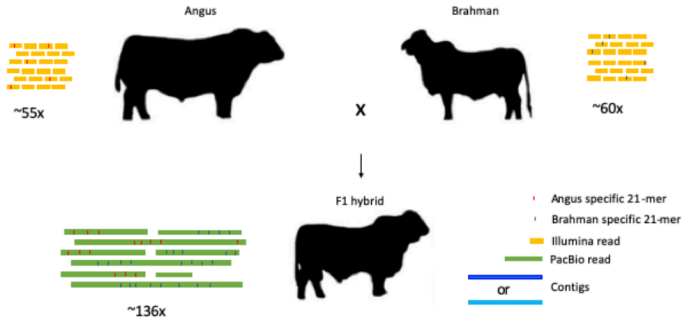
(NIB, Slovenia)



(this figure did not make it to the publication)

Petek et al "Cultivar-Specific Transcriptome and Pan-Transcriptome Reconstruction of Tetraploid Potato." *Scientific Data* (2020)

TISSUE-SPECIFIC ALLELIC EXPRESSION IN CATTLE



Article | Open Access | Published: 29 April 2020

Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle

Wai Yee Low, Rick Tearle, Ruijie Liu, Sergey Koren, Arang Rhie, Derek M. Bickhart, Benjamin D. Rosen, Zev N. Kronenberg, Sarah B. Kingan, Elizabeth Tseng, Françoise Thibaud-Nissen, Fergal J. Martin, Konstantinos Billis, Jay Ghurye, Alex R. Hastie, Joyce Lee, Andy W. C. Pang, Michael P. Heaton, Adam M. Phillippy, Stefan Hiendleder, Timothy P. L. Smith & John L. Williams

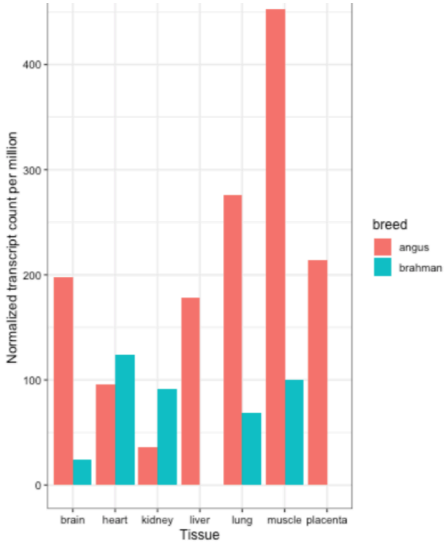
Nature Communications **11**, Article number: 2071 (2020) | Cite this article

Low et al. 2019. “Haplotype-Resolved Cattle Genomes Provide Insights Into Structural Variation and Adaptation.” *Nat Comm (2020)*

TISSUE-SPECIFIC ALLELIC EXPRESSION IN CATTLE

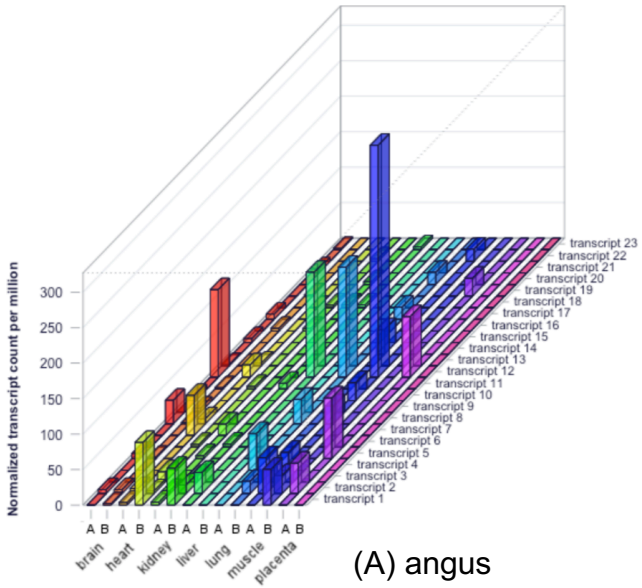
ARIH2

gene level expression



ARIH2

isoform expression by tissue / allele



(A) angus

(B) brahman

Low et al. 2019. “Haplotype-Resolved Cattle Genomes Provide Insights Into Structural Variation and Adaptation.” *Nat Comm (2020)*

ISOPHASE ON GITHUB (PART OF CUPCAKE)

IsoPhase: Haplotyping using Iso Seq data

Elizabeth Tseng edited this page yesterday · 8 revisions

Last Updated: 09/26/2020

https://github.com/Magdoll/cDNA_Cupcake/wiki/IsoPhase:-Haplotyping-using-Iso-Seq-data

PACIFIC BIOSCIENCES®

Cogent: Coding Genome Reconstruction Without A Reference Genome

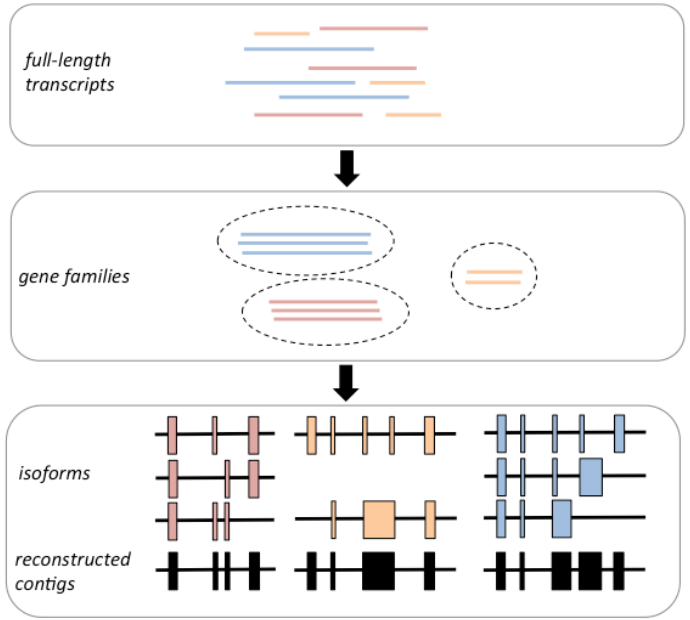
Elizabeth Tseng, Principal Scientist, PacBio

@Magdoll

NO GENOME? NO PROBLEM

COGENT workflow

Using only Iso-Seq data to find gene families and reconstruct a fake “genome”

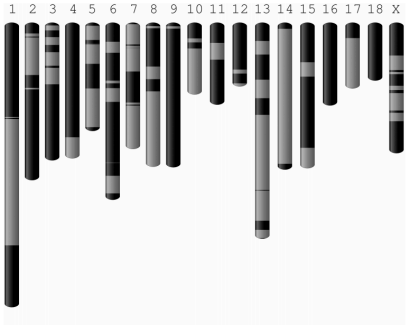


https://github.com/Magdoll/Cogent

Use COGENT results to…

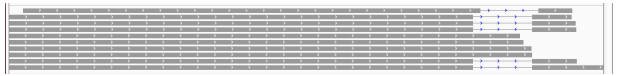
Evaluate genome assemblies

Pig Iso-Seq Cogent rescued 5 missing genes for the new pig assembly



Visualize alternative splicing

You can still see skipped exons!



Warr et al . “An Improved Pig Reference Genome Sequence to Enable Pig Genetics and Genomics Research.” *GigaScience* (2020)

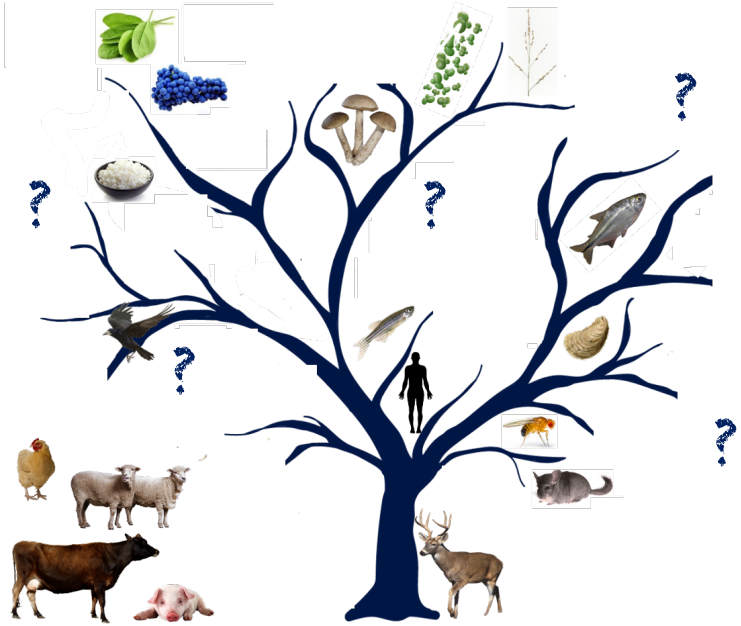
COGENT: WHY?

Not every species has a high-quality reference genome

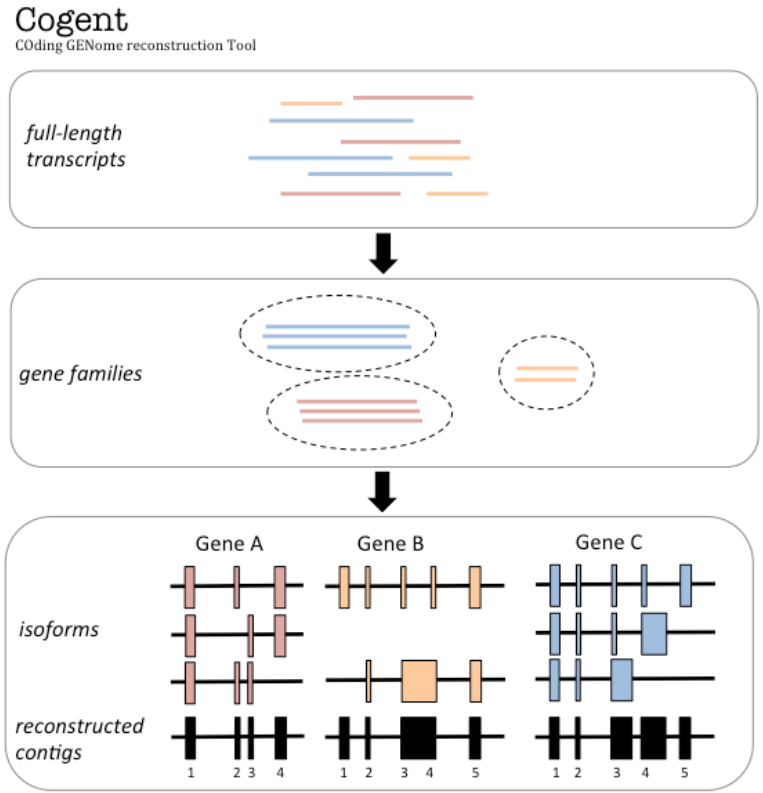
If genome is poor, genome annotation (ab initio prediction & mapping) will suffer

Iso-Seq bioinformatics does not require a genome

There is often enough information in Iso-Seq transcripts itself to identify gene families and the “coding” regions of the genome



COGENT: HOW



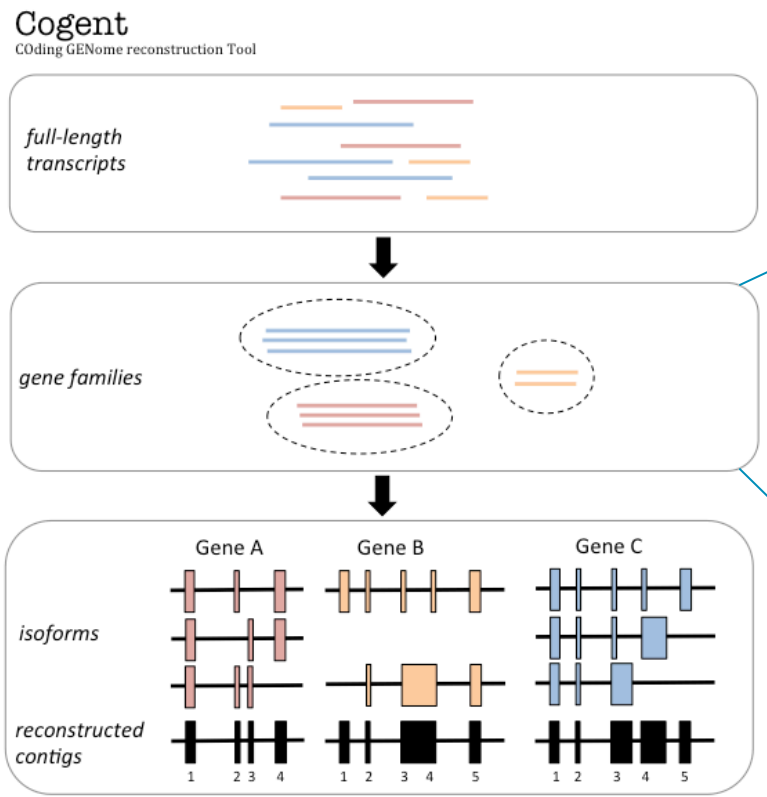
Iso-Seq analysis generates full-length, high-quality (>= 99% accuracy) transcript sequences.

Transcripts are partitioned into gene families based their k-mer similarity.

Each reconstructed contig represents the “union” of all coding bases in a particular gene locus.

(common introns will be invisible)

COGENT: GENE FAMILY FINDING

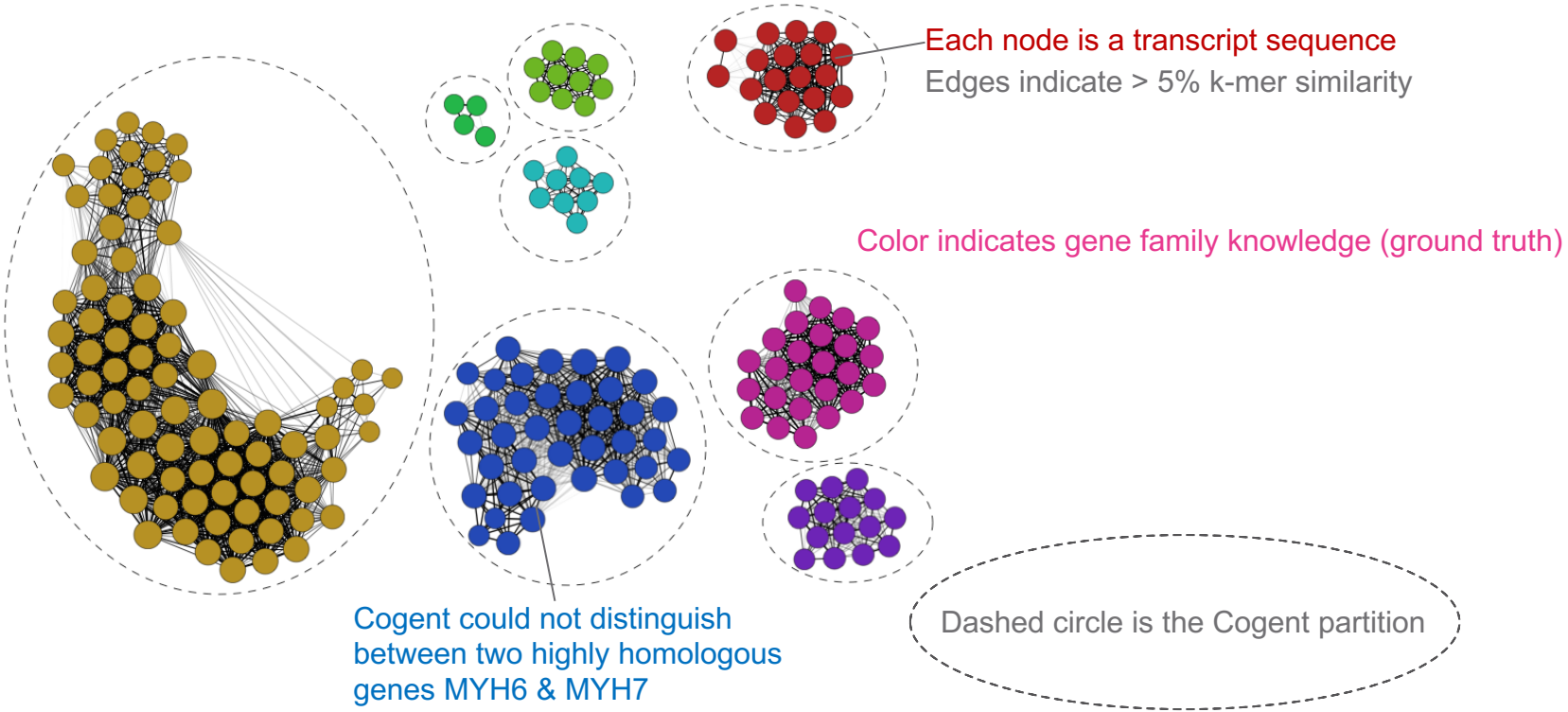


Create k-mer similarity graph

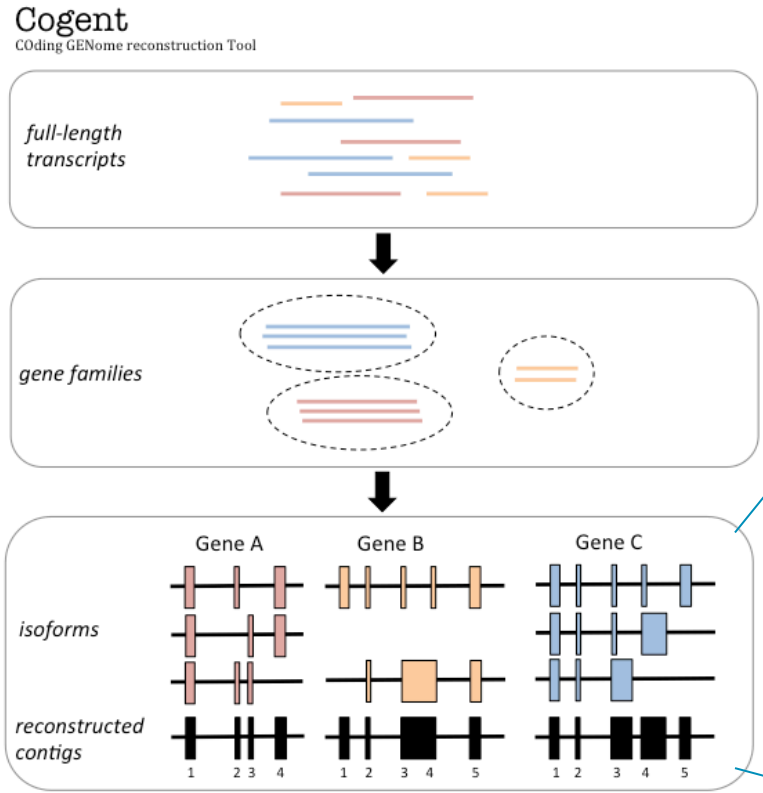
Normalized cut

Each partition is a “gene family”

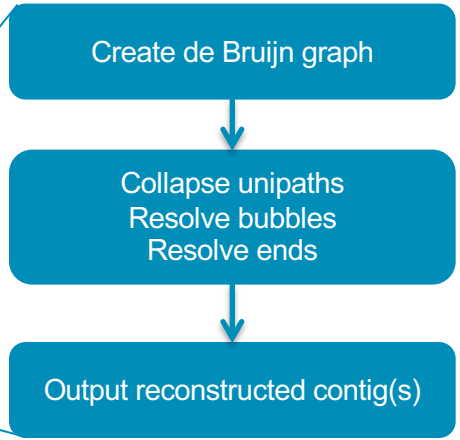
GENE FAMILY PARTITIONING OF 9 HUMAN GENES



COGENT: GENOME RECONSTRUCTION

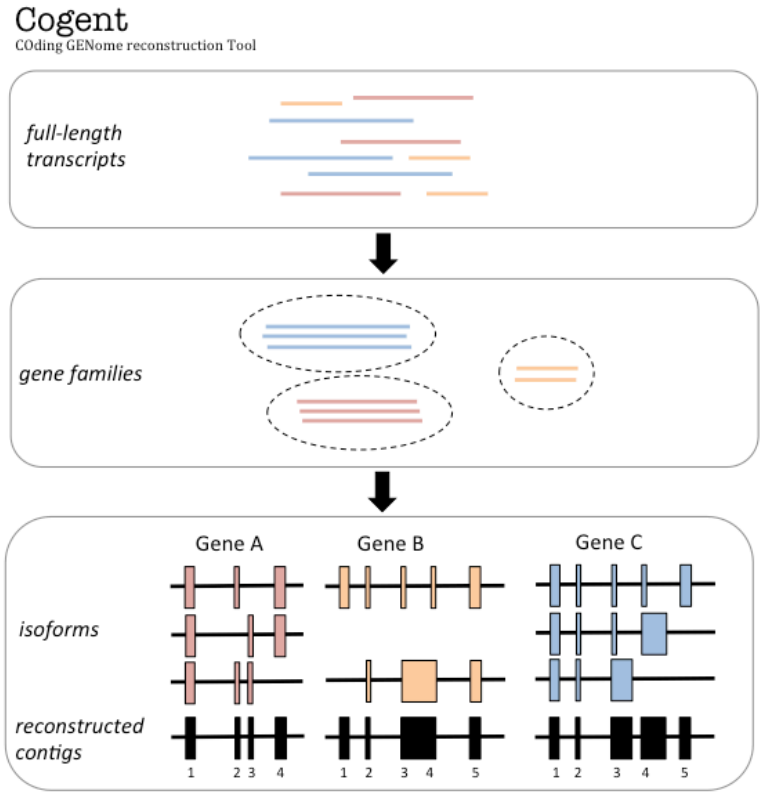


for each gene family (partition):



Full Cogent algorithm at https://github.com/Magdoll/Cogent

COGENT: GENOME RECONSTRUCTION

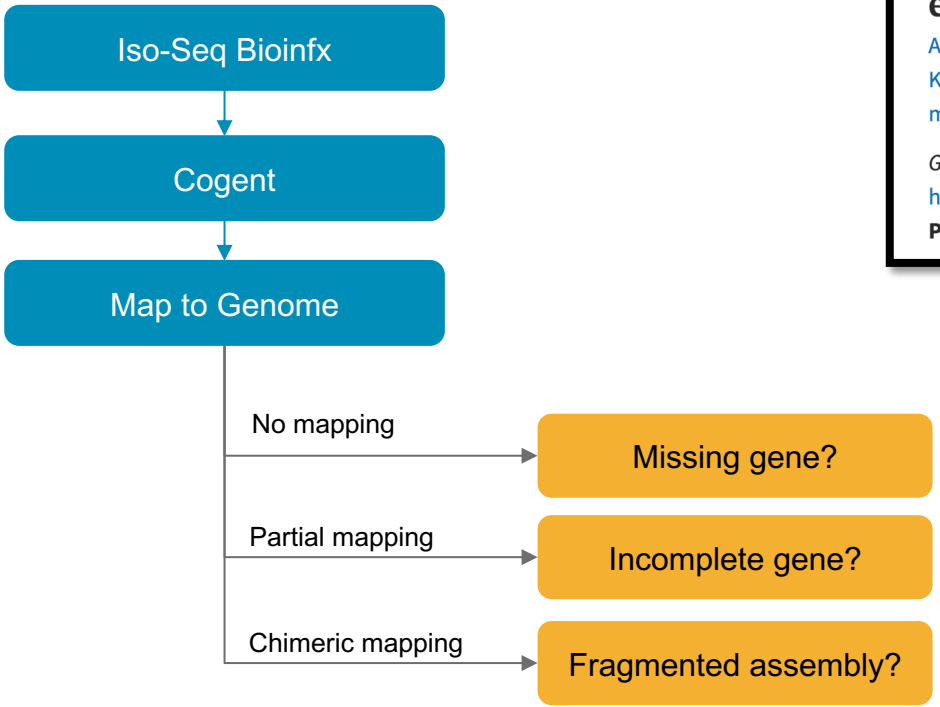


Shortfalls:

May not separate paralogs

May reconstruct multiple contigs (case of gene C)

COGENT CASE STUDY: PIG GENOME EVALUATION



An improved pig reference genome sequence to enable pig genetics and genomics research

Amanda Warr, Nabeel Affara, Bronwen Aken, Hamid Beiki, Derek M Bickhart, Konstantinos Billis, William Chow, Lel Eory, Heather A Finlayson, Paul Flicek ... Show more

GigaScience, Volume 9, Issue 6, June 2020, giaa051,
https://doi.org/10.1093/gigascience/giaa051
Published: 16 June 2020 **Article history ▾**

Warr et al . "An Improved Pig Reference Genome Sequence to Enable Pig Genetics and Genomics Research." *GigaScience* (2020)

COGENT CASE STUDY: PIG GENOME EVALUATION

Cogent Family	# of Iso-Seq Transcripts	Comment
6667_0	11	Missing CHAMP1 gene
16614_0	2	rotovirus
8757_0	30	Missing ERLIN1 gene
15567_0	8	Missing IRLIN1 gene
17496_0	17	Missing MB gene
17631_0	2	Missing PSD4 gene
17631_1	3	Missing PSD4 gene

An improved pig reference genome sequence to enable pig genetics and genomics research

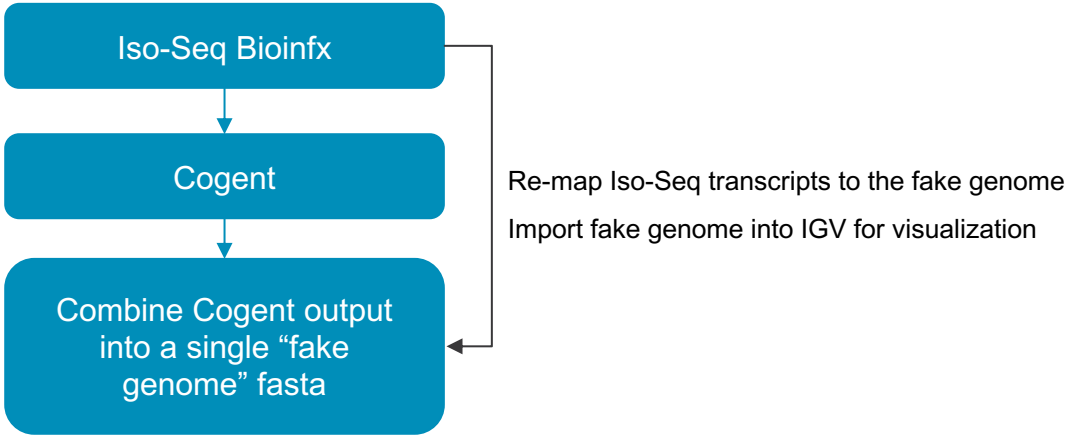
Amanda Warr, Nabeel Affara, Bronwen Aken, Hamid Beiki, Derek M Bickhart, Konstantinos Billis, William Chow, Lel Eory, Heather A Finlayson, Paul Flicek ... Show more

GigaScience, Volume 9, Issue 6, June 2020, giaa051,
https://doi.org/10.1093/gigascience/giaa051
Published: 16 June 2020 **Article history ▾**

Five missing genes in the assembly were manually put back after Cogent evaluation

Warr et al . "An Improved Pig Reference Genome Sequence to Enable Pig Genetics and Genomics Research." *GigaScience* (2020)

COGENT CASE STUDY: MAKING A “FAKE CODING” GENOME

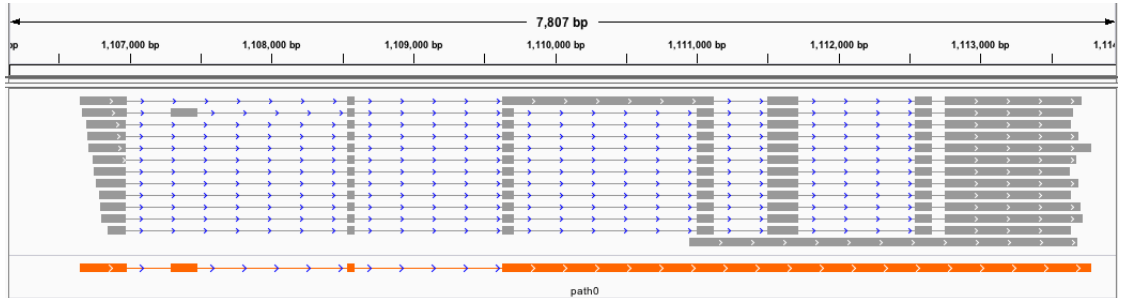


https://github.com/Magdoll/Cogent/wiki/Tutorial%3A-Using-Cogent-to-collapse-redundant-transcripts-in-absence-of-genome

COGENT CASE STUDY: MAKING A “FAKE CODING” GENOME

Genome-based View

All introns are visible

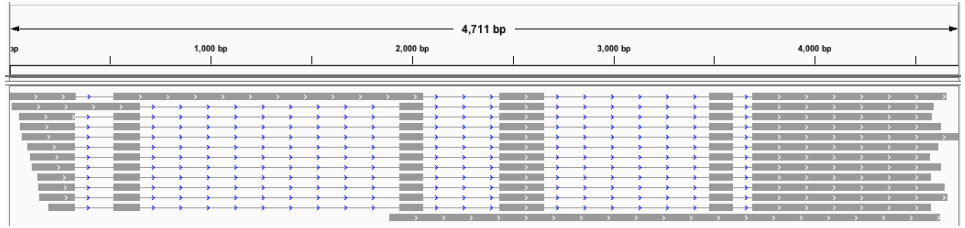


Iso-Seq Transcripts

Cogent Output

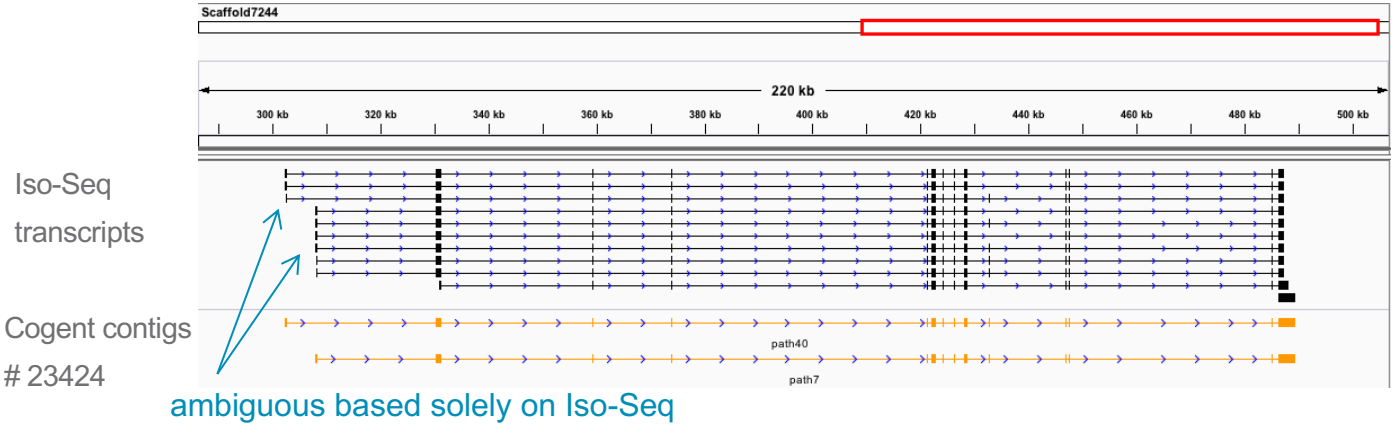
Cogent-based View

Introns that are never transcribed are not visible



https://github.com/Magdoll/Cogent/wiki/Tutorial%3A-Using-Cogent-to-collapse-redundant-transcripts-in-absence-of-genome

WHY COGENT MAY SOMETIMES OUTPUT >1 CONTIGS



- Lack of connectivity information between exon 1 and 2 based solely on transcripts
- Cogent outputs two contigs, one with exon 1 - 3, one with exon 2 – 3
- Mapping back to genome shows that the reconstruction is correct
- This is a case where genome information can be used to order exon 1 and 2

Coming soon: Special Cogent parameter to make “best guess” for ambiguity

PACBIO®

www.pacb.com