

Polymorphism of retrotransposons in *Bos taurus*

Bernt Guldbrandtsen Goutam Sahana Mogens Sandø Lund

Center for Quantitative Genetics and Genomics
Aarhus University
Denmark

Transposons

- Type of “jumping genes” (McClintock, 1956)
- Mobile genetic elements – genetic parasites
- Contain “machinery” for own mobilization

Transposons

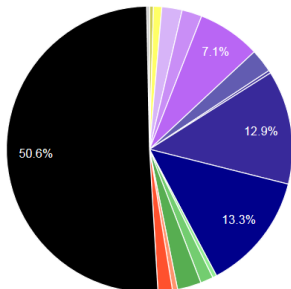
- Type of “jumping genes” (McClintock, 1956)
- **Mobile genetic elements – genetic parasites**
- Contain “machinery” for own mobilization

Transposons

- Type of “jumping genes” (McClintock, 1956)
- Mobile genetic elements – genetic parasites
- Contain “machinery” for own mobilization

Prevalence

- Min. 45% of human genome (Mills *et al.*, 2007)
- Alu1 and LINE1 make up about 25% of human genome (Smit, 1996)
- In cattle:

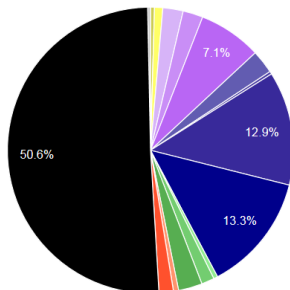


Bos taurus

Source: <http://www.repeatmasker.org>

Prevalence

- Min. 45% of human genome (Mills *et al.*, 2007)
- Alu1 and LINE1 make up about 25% of human genome (Smit, 1996)
- In cattle:

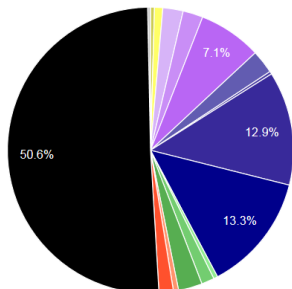


Bos taurus

Source: <http://www.repeatmasker.org>

Prevalence

- Min. 45% of human genome (Mills *et al.*, 2007)
- Alu1 and LINE1 make up about 25% of human genome (Smit, 1996)
- In cattle:

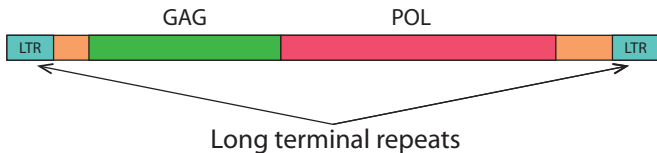


Bos taurus

Source: <http://www.repeatmasker.org>

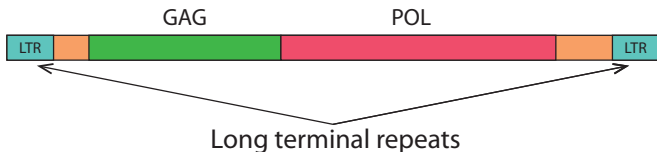
Transposons

- Several classes
- Some (retrovirus-like) are flanked by *long terminal repeats* (LTR)
- Lifecycle depends on reverse transcriptase (RNA→DNA)



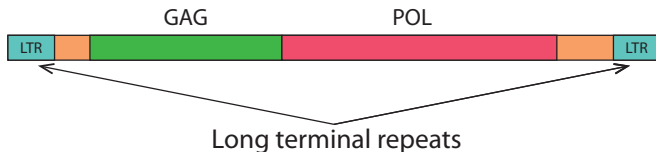
Transposons

- Several classes
- Some (retrovirus-like) are flanked by *long terminal repeats* (LTR)
- Lifecycle depends on reverse transcriptase (RNA→DNA)



Transposons

- Several classes
- Some (retrovirus-like) are flanked by *long terminal repeats* (LTR)
- Lifecycle depends on reverse transcriptase (RNA→DNA)



Transposition

- May contain “machinery” for own mobilization (autonomous)
- Other copies are incomplete (non-autonomous)
- Accumulate mutations
- Most copies presumably defective
- Prevalence underestimated – mutated beyond recognition
- New integration hard to detect with NGS data

Transposition

- May contain “machinery” for own mobilization (autonomous)
- Other copies are incomplete (non-autonomous)
- Accumulate mutations
- Most copies presumably defective
- Prevalence underestimated – mutated beyond recognition
- New integration hard to detect with NGS data

Transposition

- May contain “machinery” for own mobilization (autonomous)
- Other copies are incomplete (non-autonomous)
- **Accumulate mutations**
- Most copies presumably defective
- Prevalence underestimated – mutated beyond recognition
- New integration hard to detect with NGS data

Transposition

- May contain “machinery” for own mobilization (autonomous)
- Other copies are incomplete (non-autonomous)
- Accumulate mutations
- **Most copies presumably defective**
- Prevalence underestimated – mutated beyond recognition
- New integration hard to detect with NGS data

Transposition

- May contain “machinery” for own mobilization (autonomous)
- Other copies are incomplete (non-autonomous)
- Accumulate mutations
- Most copies presumably defective
- **Prevalence underestimated – mutated beyond recognition**
- New integration hard to detect with NGS data

Transposition

- May contain “machinery” for own mobilization (autonomous)
- Other copies are incomplete (non-autonomous)
- Accumulate mutations
- Most copies presumably defective
- Prevalence underestimated – mutated beyond recognition
- **New integration hard to detect with NGS data**

Mutagenesis

- **Insertion of transposons may damage important sequence**
- 10% of all spontaneous mutations in mouse due to retrotransposition (Kazazian, 1998)
- Sequence identity may lead to non-homologous recombination between elements (Kadri *et al.*, 2014).

Mutagenesis

- Insertion of transposons may damage important sequence
- 10% of all spontaneous mutations in mouse due to retrotransposition (Kazazian, 1998)
- Sequence identity may lead to non-homologous recombination between elements (Kadri *et al.*, 2014).

Mutagenesis

- Insertion of transposons may damage important sequence
- 10% of all spontaneous mutations in mouse due to retrotransposition (Kazazian, 1998)
- Sequence identity may lead to non-homologous recombination between elements (Kadri *et al.*, 2014).

Scanning the *Bos taurus* genome

- **GenomeTools' ltrharvest (Ellinghaus *et al*, 2008)**
- Detect LTR flanked transposons – one class of transposons
- $\approx 4.6\%$ of the cattle genome
(<http://www.repeatmasker.org/species/bosTau.html>)
- University of Maryland Assembly 3.1 (Zimin *et al*, 2009) – 2 Hereford individuals: 2/3 a cow, 1/3 a bull
- Detects 87,310 putative LTR transposons in *Bos taurus* genome

Scanning the *Bos taurus* genome

- GenomeTools' ltrharvest (Ellinghaus *et al*, 2008)
- **Detect LTR flanked transposons – one class of transposons**
- $\approx 4.6\%$ of the cattle genome
(<http://www.repeatmasker.org/species/bosTau.html>)
- University of Maryland Assembly 3.1 (Zimin *et al*, 2009) – 2 Hereford individuals: 2/3 a cow, 1/3 a bull
- Detects 87,310 putative LTR transposons in *Bos taurus* genome

Scanning the *Bos taurus* genome

- GenomeTools' ltrharvest (Ellinghaus *et al*, 2008)
- Detect LTR flanked transposons – one class of transposons
- **≈ 4.6% of the cattle genome**
(<http://www.repeatmasker.org/species/bosTau.html>)
- University of Maryland Assembly 3.1 (Zimin *et al*, 2009) – 2 Hereford individuals: 2/3 a cow, 1/3 a bull
- Detects 87,310 putative LTR transposons in *Bos taurus* genome

Scanning the *Bos taurus* genome

- GenomeTools' ltrharvest (Ellinghaus *et al*, 2008)
- Detect LTR flanked transposons – one class of transposons
- $\approx 4.6\%$ of the cattle genome
(<http://www.repeatmasker.org/species/bosTau.html>)
- **University of Maryland Assembly 3.1 (Zimin *et al*, 2009) – 2 Hereford individuals: 2/3 a cow, 1/3 a bull**
- Detects 87,310 putative LTR transposons in *Bos taurus* genome

Scanning the *Bos taurus* genome

- GenomeTools' ltrharvest (Ellinghaus *et al*, 2008)
- Detect LTR flanked transposons – one class of transposons
- $\approx 4.6\%$ of the cattle genome
(<http://www.repeatmasker.org/species/bosTau.html>)
- University of Maryland Assembly 3.1 (Zimin *et al*, 2009) – 2 Hereford individuals: 2/3 a cow, 1/3 a bull
- **Detects 87,310 putative LTR transposons in *Bos taurus* genome**

Scanning Sequenced Animals

- Detect polymorphism in sequenced animals
- 172 animals with NGS data
- 46 Holstein, 27 Jersey, 83 Nordic Red and 16 Brown Swiss
- About 10X, about 300 bp inserts

Scanning Sequenced Animals

- Detect polymorphism in sequenced animals
- **172 animals with NGS data**
- 46 Holstein, 27 Jersey, 83 Nordic Red and 16 Brown Swiss
- About 10X, about 300 bp inserts

Scanning Sequenced Animals

- Detect polymorphism in sequenced animals
- 172 animals with NGS data
- 46 Holstein, 27 Jersey, 83 Nordic Red and 16 Brown Swiss
- About 10X, about 300 bp inserts

Scanning Sequenced Animals

- Detect polymorphism in sequenced animals
- 172 animals with NGS data
- 46 Holstein, 27 Jersey, 83 Nordic Red and 16 Brown Swiss
- **About 10X, about 300 bp inserts**

Method

Scan for pairs with reads mapping left and right of putative transposon to detect absence of putative LTR transposons detected in the assembly



Classifying

1. More than 100 hits in animal → discard hit
2. More than 30 hits in animal → discard animal for hit
3. Less than 5 animals left in breed → discard breed for hit
4. 0-1 hits → animal carries putative transposon
5. More than 1 hit → animal has at least one chromatid without transposon

Classifying

1. More than 100 hits in animal → discard hit
2. **More than 30 hits in animal → discard animal for hit**
3. Less than 5 animals left in breed → discard breed for hit
4. 0-1 hits → animal carries putative transposon
5. More than 1 hit → animal has at least one chromatid without transposon

Classifying

1. More than 100 hits in animal → discard hit
2. More than 30 hits in animal → discard animal for hit
3. **Less than 5 animals left in breed → discard breed for hit**
4. 0-1 hits → animal carries putative transposon
5. More than 1 hit → animal has at least one chromatid without transposon

Classifying

1. More than 100 hits in animal → discard hit
2. More than 30 hits in animal → discard animal for hit
3. Less than 5 animals left in breed → discard breed for hit
4. **0-1 hits → animal carries putative transposon**
5. More than 1 hit → animal has at least one chromatid without transposon

Classifying

1. More than 100 hits in animal → discard hit
2. More than 30 hits in animal → discard animal for hit
3. Less than 5 animals left in breed → discard breed for hit
4. 0-1 hits → animal carries putative transposon
5. More than 1 hit → animal has at least one chromatid without transposon

Lots of Polymorphism

Status	BSW	HOL	JER	RDC
n	16	46	27	83
Retained	76,320	73,101	75,499	71,650
“Absent”	163	166	680	14
Polymorphic	4,550	7,862	4,958	9,476
NA	108	1	4	1

Lots of Polymorphism

Status	BSW	HOL	JER	RDC
n	16	46	27	83
Retained	76,320	73,101	75,499	71,650
“Absent”	163	166	680	14
Polymorphic	4,550	7,862	4,958	9,476
NA	108	1	4	1

Lots of Polymorphism

Status	BSW	HOL	JER	RDC
n	16	46	27	83
Retained	76,320	73,101	75,499	71,650
“Absent”	163	166	680	14
Polymorphic	4,550	7,862	4,958	9,476
NA	108	1	4	1

Lots of Polymorphism

Status	BSW	HOL	JER	RDC
n	16	46	27	83
Retained	76,320	73,101	75,499	71,650
“Absent”	163	166	680	14
Polymorphic	4,550	7,862	4,958	9,476
NA	108	1	4	1

Lots of Polymorphism

Status	BSW	HOL	JER	RDC
n	16	46	27	83
Retained	76,320	73,101	75,499	71,650
“Absent”	163	166	680	14
Polymorphic	4,550	7,862	4,958	9,476
NA	108	1	4	1

Distribution

- 80,314 present in all dairy breeds
- 70,046 present in all animals in all dairy animals
- 736 “absent” in at least one dairy breed
- 13 “absent” from all dairy breeds
- 11,080 polymorphic in at least one dairy breed
- 2,514 polymorphic in all dairy breeds

Distribution

- 80,314 present in all dairy breeds
- **70,046 present in all animals in all dairy animals**
- 736 “absent” in at least one dairy breed
- 13 “absent” from all dairy breeds
- 11,080 polymorphic in at least one dairy breed
- 2,514 polymorphic in all dairy breeds

Distribution

- 80,314 present in all dairy breeds
- 70,046 present in all animals in all dairy animals
- **736 “absent” in at least one dairy breed**
- 13 “absent” from all dairy breeds
- 11,080 polymorphic in at least one dairy breed
- 2,514 polymorphic in all dairy breeds

Distribution

- 80,314 present in all dairy breeds
- 70,046 present in all animals in all dairy animals
- 736 “absent” in at least one dairy breed
- **13 “absent” from all dairy breeds**
- 11,080 polymorphic in at least one dairy breed
- 2,514 polymorphic in all dairy breeds

Distribution

- 80,314 present in all dairy breeds
- 70,046 present in all animals in all dairy animals
- 736 “absent” in at least one dairy breed
- 13 “absent” from all dairy breeds
- **11,080 polymorphic in at least one dairy breed**
- 2,514 polymorphic in all dairy breeds

Distribution

- 80,314 present in all dairy breeds
- 70,046 present in all animals in all dairy animals
- 736 “absent” in at least one dairy breed
- 13 “absent” from all dairy breeds
- 11,080 polymorphic in at least one dairy breed
- **2,514 polymorphic in all dairy breeds**

Interpretation

- Elements private to UMD3.1 animals must have arisen since separation of dairy and beef or were polymorphic in ancestral population
- High degree of within and between breed polymorphism, means many fairly recent new insertions

Interpretation

- Elements private to UMD3.1 animals must have arisen since separation of dairy and beef or were polymorphic in ancestral population
- High degree of within and between breed polymorphism, means many fairly recent new insertions

Conclusions

- Evidence of extensive mobilization of LTR transposons
- Source of genetic variation that is hard to follow with current NGS data processing pipelines
- We may be missing a major type of genetic variants in sequence analysis and gene mapping

Conclusions

- Evidence of extensive mobilization of LTR transposons
- Source of genetic variation that is hard to follow with current NGS data processing pipelines
- We may be missing a major type of genetic variants in sequence analysis and gene mapping

Conclusions

- Evidence of extensive mobilization of LTR transposons
- Source of genetic variation that is hard to follow with current NGS data processing pipelines
- We may be missing a major type of genetic variants in sequence analysis and gene mapping