

# False Positives structural deletions

M. Mielczarek<sup>1,2</sup>, M. Frąszczak<sup>1</sup>, J. Szyda<sup>1,2</sup>

<sup>1</sup>Biostatistics Group, Department of Genetics Wrocław University of Environmental and Life Sciences , Kozuchowska 7, 51-631 Wrocław, Poland, <sup>2</sup>National Research Institute of Animal Production; Krakowska 1, 32-083 Cracow-Balice, Poland; magda.mielczarek@upwr.edu.pl

## Objective

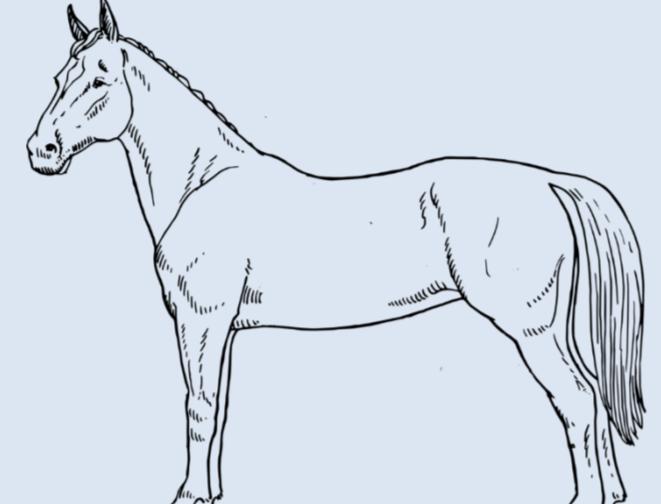
How many structural deletions in DGVa are false positives caused by gaps in the reference genomes

## Conclusions

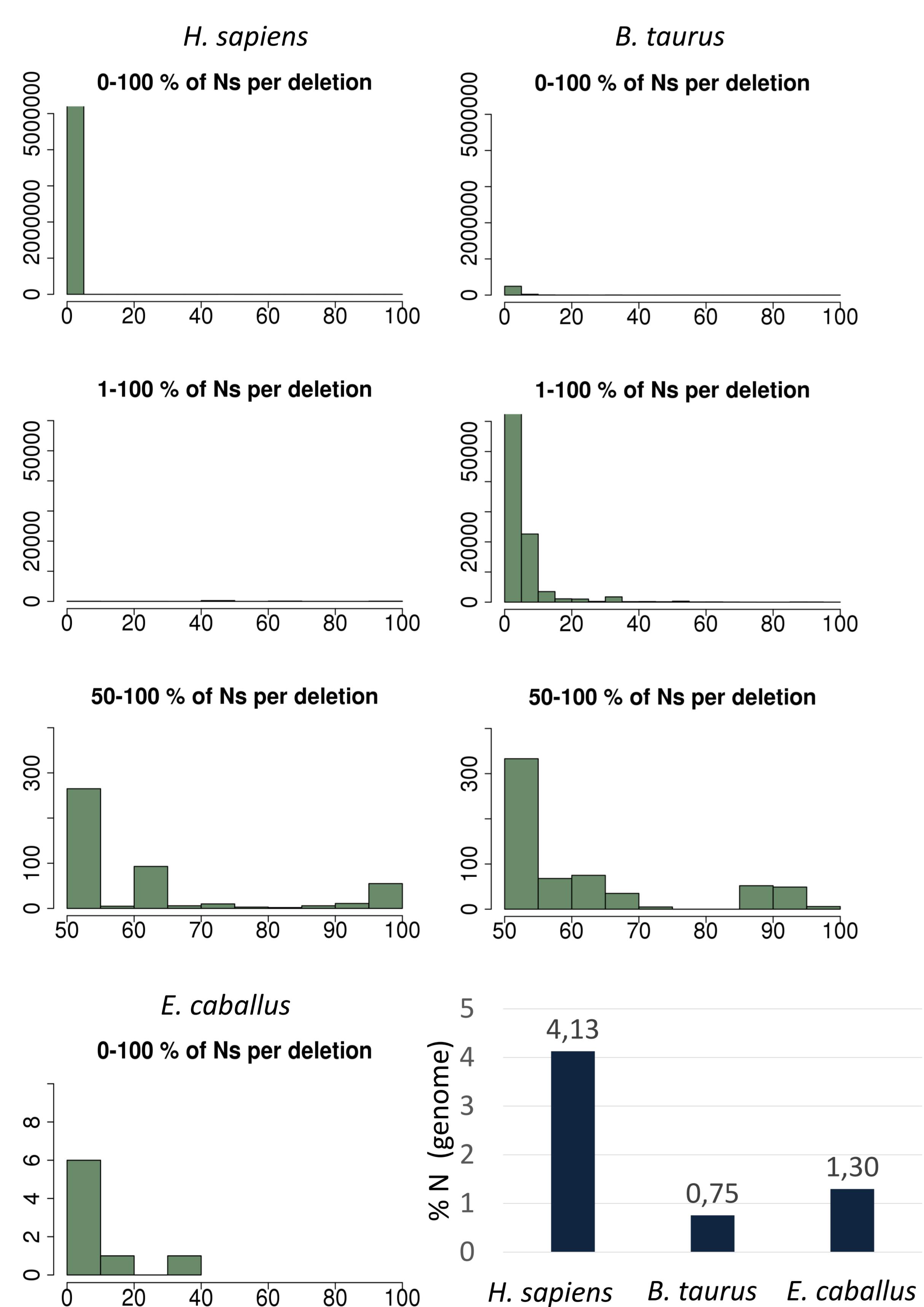
- The most of deletions are not effected by gaps in the reference genome
- False positive deletions may result from unfavorable features of the reference

## Data

- Reference genomes ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov))



- Structural deletions ([www.ebi.ac.uk/dgva](http://www.ebi.ac.uk/dgva))



## Methods

### I. Data filtering:

- Deletions
  - NGS based
  - Minimal length = 50 bp
- N content analysis of structural deletions
  - N content analysis of reference genomes

## Results

species	study	# all CNV	# del	# del with at least 90 % of N	CNV detection methods*	validation/filtration methods
human	1000 GP Consortium <i>et al.</i> 2015	15,963,503	6,410,414	0	BreakDancer (PEM), DELLY (PEM + SR), CNVnator (RD), GenomeSTRiP (PEM + RD), VariationHunter (PEM)	merging of the CNV callsets
	Alsmadi <i>et al.</i> 2014	88,344	30,220	45 ** (0.15 %)	HugeSeq pipeline: BreakDancer (PEM), Pindel (SR), CNVnator (PEM), BreakSeq (JM)	merging of the CNV callsets
	Thareja <i>et al.</i> 2015	23,868	7,402	21 *** (0.28 %)	HugeSeq pipeline	merging of the CNV callsets
cattle	Bickhart <i>et al.</i> 2012	13,769	219	0	WSSD pipeline (RD)	aCGH, qPCR, FISH
	Boussaha <i>et al.</i> 2015	25,365	12,651	56 **** (0.44 %)	Pindel (SR)	keeping CNV supported by min. 3 reads
	Keel <i>et al.</i> 2016	284,780	262,366	0	CN.MOPS (RD)	determining a strategy for CNV detection - 3 algorithms comparison on benchmark CNV datasets
horse	Doan <i>et al.</i> 2012	387,250	8	0	FREEC (RD)	telomeric regions removing

\* PEM - paired-end, SR - split-read, RD - read-depth, JM - junction mapping / \*\* 20 of 45 deletions contain 100 % of Ns / \*\*\* 11 of 21 deletions contain 100 % of Ns / \*\*\*\* 1 deletion shared by 43 of 62 animals

