

Natural history of a copy number variant in mouse

Andrew P Morgan

28th International Mammalian Genome Conference
Bar Harbor, Maine

26 October 2014



Motivation

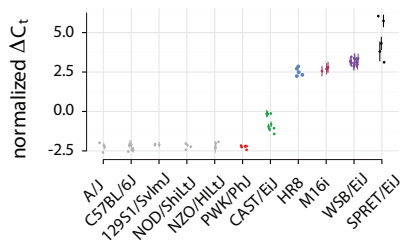
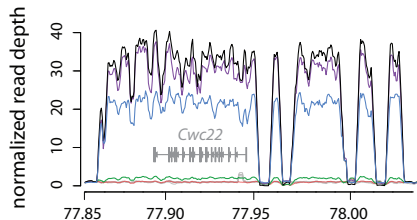
Why study copy-number variation?

- impacts medically-relevant phenotypes
- important driver of genome evolution

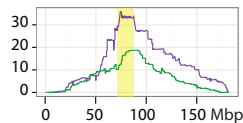
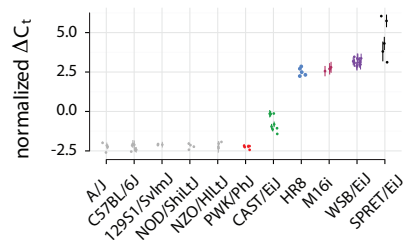
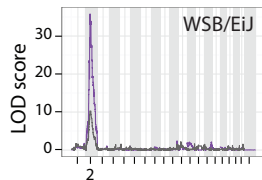
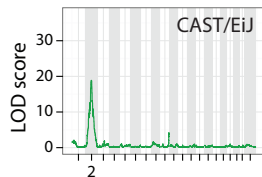
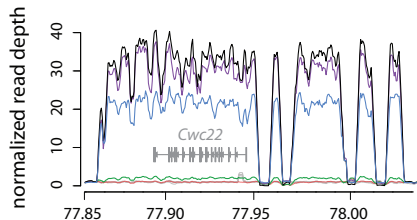
Why study **this locus** (chr2: 77.9 Mbp)?

- controls own segregation in *cis* (**O-1**)
- selective sweeps in lab and in nature (**P-14**)

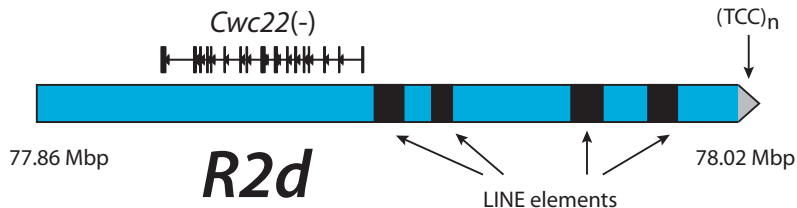
Evidence for a CNV on mouse chr2



Evidence for a CNV on mouse chr2

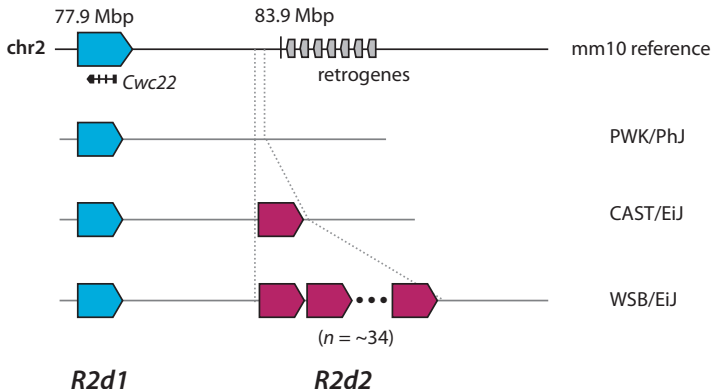
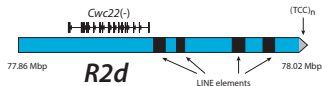


The *R2d* repeat unit



R2d = “responder to (meiotic) drive”

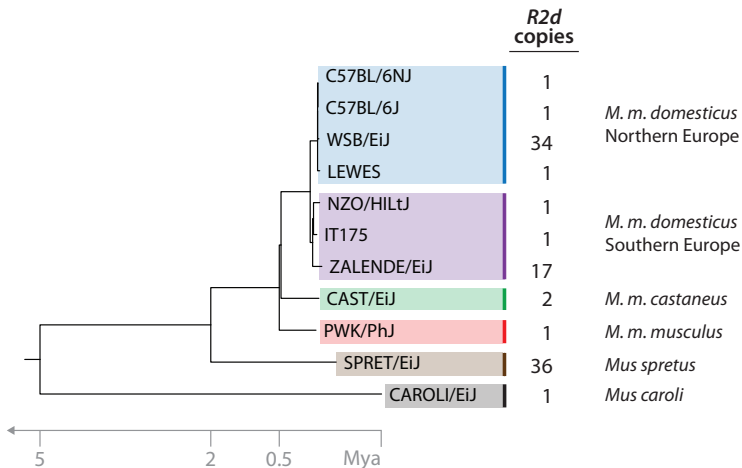
The *R2d1/R2d2* locus



Evolutionary questions

- When was *R2d* duplicated?
- How have *R2d* sequences diverged?
- Has gene conversion occurred between *R2d* copies?
- When did *R2d2* expansion allele(s) arise?

Phylogenetic context



Chevret P *et al.* (2005) *Biol J Linnean Soc*

Didion JP *et al.* (2013) *Mamm Genome*

A new tool for sequence analysis

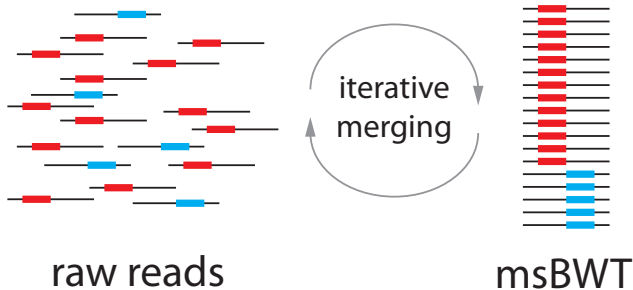
Multi-string Burrows-Wheeler transform (**msBWT**) offers

- compression
- fast sequence searches

A new tool for sequence analysis

Multi-string Burrows-Wheeler transform (**msBWT**) offers

- compression
- fast sequence searches



A new tool for sequence analysis

Query sequence: CTGGCCTGTCCACAGTGTC

```

CTGGCCTGTCCACAGTGTC
--AGAGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGCTAA-----
----AGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGCTTAACA-----
----GCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGCTTAACAT-----
----CATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGCTTAACATA-----
-----AGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGCTTAACATAGTAC-----
-----GCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGCTTAACATAGTACT-----
-----CATTTATCATGCCTTCTGGCCTGTCCACAGTGCTTAACATAGTACTAGAT-----
-----TATCATGCCTTCTGGCCTGTCCACAGTGCTTAACATAGTACTAGATACAC-----
-----TATCATGCCTTCTGGCCTGTCCACAGTGCTTAATATAGTACTAGATACAC-----
-----CATGCCTTCTGGCCTGTCCACAGTGCTTAATATAGTACTAGATACACAAT-----
-----ATGCCTTCTGGCCTGTCCACAGTGCTTAATATAGTACTAGATACACAATG-----
-----TGCCTTCTGGCCTGTCCACAGTGCTTAATATAGTACTAGATACACAATGC-----
-----TGCCTTCTGGCCTGTCCACAGTGCTTAATATAGTACTAGATACACAATGC-----
*
--AGAGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGCTTAAYATAGTACTAGATACACAATGC--
```

We assembled 13.3 kbp of sequence in 8 intervals within *R2d*.

A new tool for sequence analysis

Query sequence: CTGGCCTGTCCACAGTGTC

```

                CTGGCCTGTCCACAGTGTC
--AGAGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---TAA-----
---AGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---TAACA-----
---GCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---TAACAT-----
---CATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---TAACATA-----
-----AGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---TAACATAGTAC-----
-----GCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---TAACATAGTACT-----
-----CATTATCATGCCTTCTGGCCTGTCCACAGTGTC---TAACATAGTACTAGAT-----
-----TATCATGCCTTCTGGCCTGTCCACAGTGTC---TAACATAGTACTAGATACAC-----
-----TATCATGCCTTCTGGCCTGTCCACAGTGTC---TAATATAGTACTAGATACAC-----
-----CATGCCTTCTGGCCTGTCCACAGTGTC---TAATATAGTACTAGATACACAAT-----
-----ATGCCTTCTGGCCTGTCCACAGTGTC---TAATATAGTACTAGATACACAATG-----
-----TGCCTTCTGGCCTGTCCACAGTGTC---TAATATAGTACTAGATACACAATGC-----
-----TGCCTTCTGGCCTGTCCACAGTGTC---TAATATAGTACTAGATACACAATGC-----
                *
--AGAGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---TAAYATAGTACTAGATACACAATGC--
```

We assembled 13.3 kbp of sequence in 8 intervals within *R2d*.

A new tool for sequence analysis

Query sequence: CTGGCCTGTCCACAGTGTC

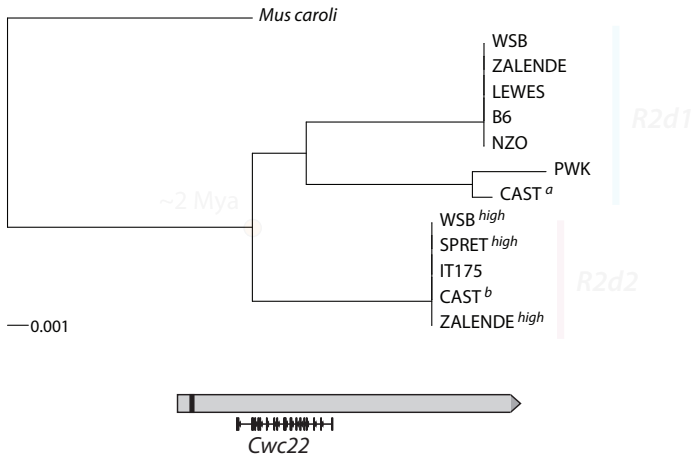
```

CTGGCCTGTCCACAGTGTC
--AGAGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---
---AGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---
---GCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---
---CATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---
-----AGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---
-----GCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---
-----CATTATCATGCCTTCTGGCCTGTCCACAGTGTC---
-----TATCATGCCTTCTGGCCTGTCCACAGTGTC---
-----TATCATGCCTTCTGGCCTGTCCACAGTGTC---
-----CATGCCTTCTGGCCTGTCCACAGTGTC---
-----ATGCCTTCTGGCCTGTCCACAGTGTC---
-----TGCCTTCTGGCCTGTCCACAGTGTC---
-----TGCCTTCTGGCCTGTCCACAGTGTC---
*
--AGAGCATGAGCCTCATTATCATGCCTTCTGGCCTGTCCACAGTGTC---

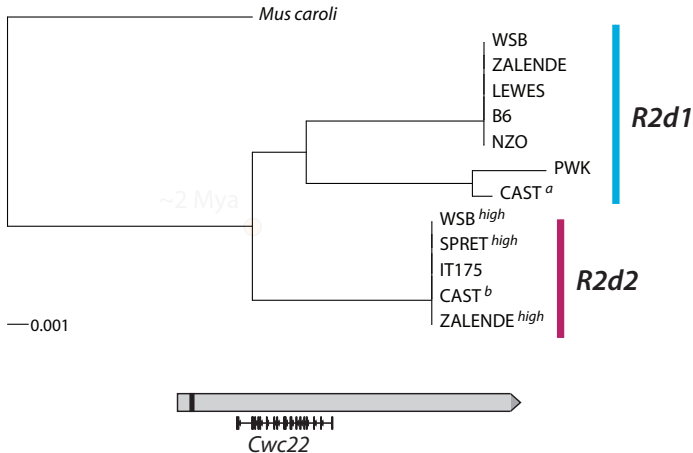
```

We assembled 13.3 kbp of sequence in 8 intervals within *R2d*.

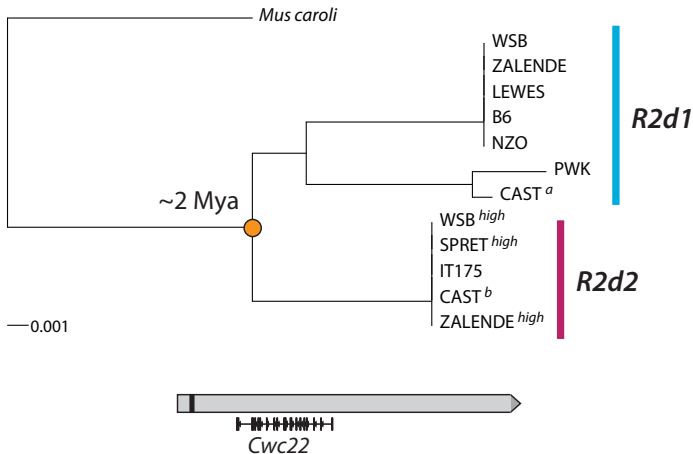
Duplication and divergence



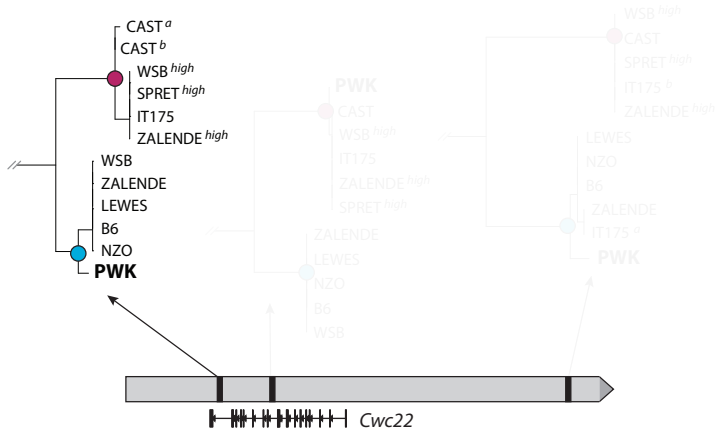
Duplication and divergence



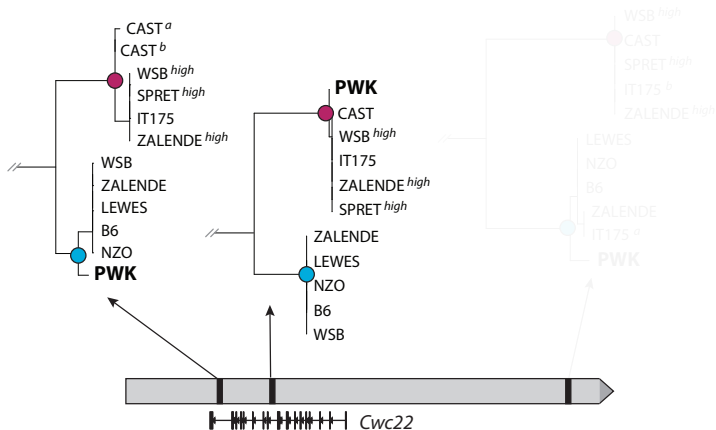
Duplication and divergence



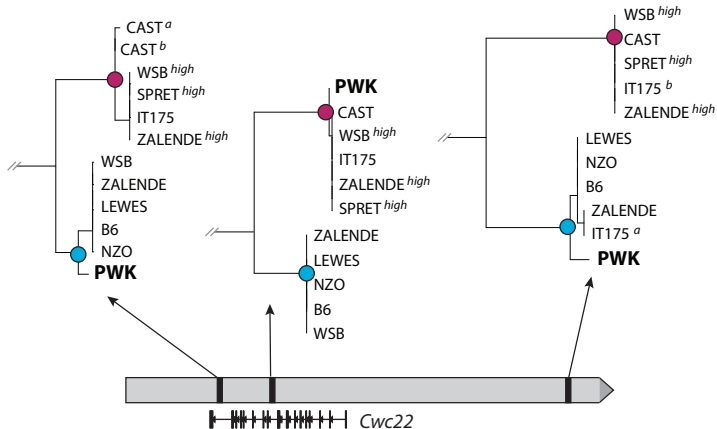
Inter-locus gene conversion



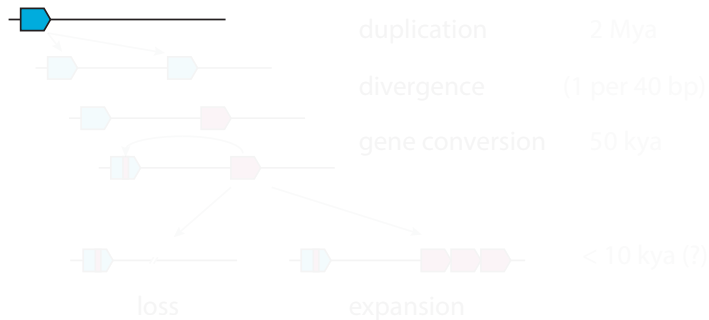
Inter-locus gene conversion



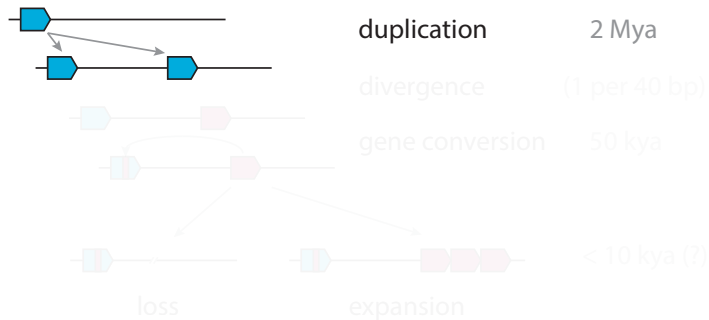
Inter-locus gene conversion



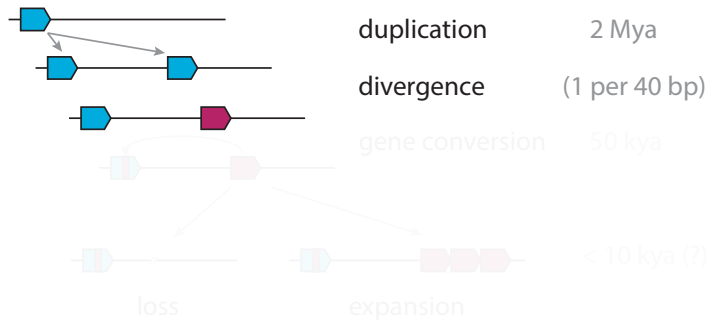
History of *R2d* so far



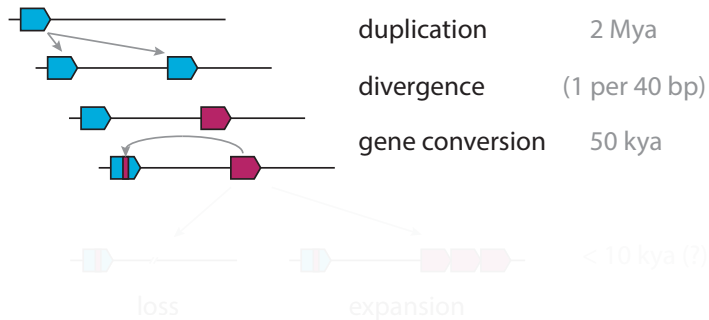
History of *R2d* so far



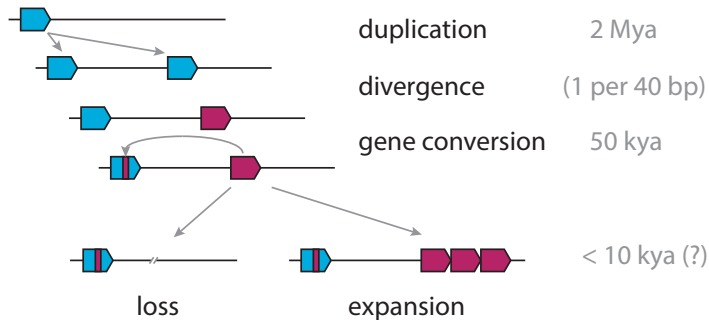
History of *R2d* so far



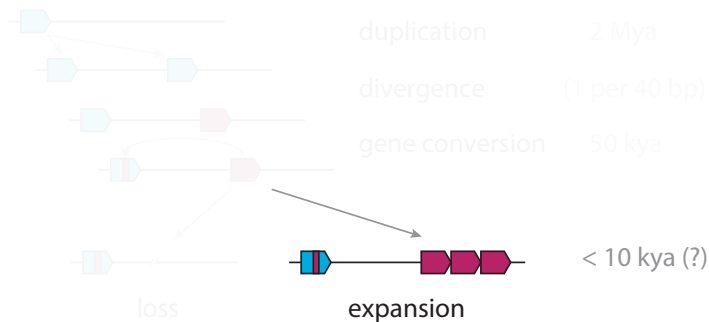
History of *R2d* so far



History of *R2d* so far



History of *R2d* so far



R2d2 expansion alleles are of recent European origin

Classical laboratory mice:

- ALS/LtJ, ALR/LtJ, CHMU/LeJ, NU/J
- ICR, CD-1 stocks

Wild-derived strains:

- WSA/EiJ, WSB/EiJ, DDO

Strains with Robertsonian karyotypes:

- ZALENDE/EiJ, RBA/DnJ, RBB/DnJ, RBF/DnJ

R2d2 expansion alleles are of recent European origin

Classical laboratory mice:

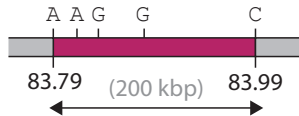
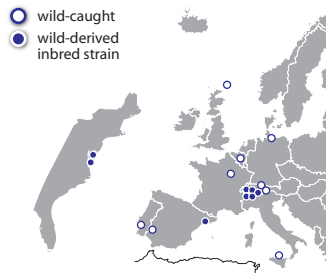
- ALS/LtJ, ALR/LtJ, CHMU/LeJ, NU/J
- ICR, CD-1 stocks

Wild-derived strains:

- WSA/EiJ, WSB/EiJ, DDO

Strains with Robertsonian karyotypes:

- ZALENDE/EiJ, RBA/DnJ, RBB/DnJ, RBF/DnJ



R2d2 expansion alleles are of recent European origin

Classical laboratory mice:

- ALS/LtJ, ALR/LtJ, CHMU/LeJ, NU/J
- ICR, CD-1 stocks

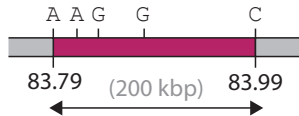
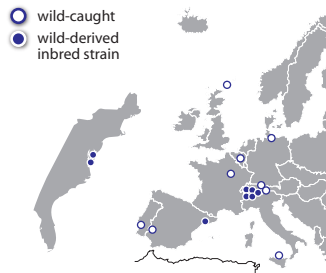
Wild-derived strains:

- WSA/EiJ, WSB/EiJ, DDO

Strains with Robertsonian karyotypes:

- ZALENDE/EiJ, RBA/DnJ, RBB/DnJ, RBF/DnJ

R2d2 expansion alleles have single origin
~ 5-10 kya in *M. m. domesticus*.



Frequent germline *de novo* CNVs at *R2d2*

population	<i>n</i>	new alleles	generations
Diversity Outbred	125	6	$\leq 3?$
ALS/LtJ strain	8	1	$\leq 5?$
WSB/EiJ \times PWK/PhJ*	45	1	1

Mutation rate ~ 0.01 /generation makes *R2d2* among hottest CNV hotspots reported in mouse.

Frequent germline *de novo* CNVs at *R2d2*

population	<i>n</i>	new alleles	generations
Diversity Outbred	125	6	$\leq 3?$
ALS/LtJ strain	8	1	$\leq 5?$
WSB/EiJ \times PWK/PhJ*	45	1	1

Mutation rate ~ 0.01 /generation makes *R2d2* among hottest CNV hotspots reported in mouse.

Implications

Evolution of genome structure in recent past:

- phylogenetic discordance within *M. musculus*
- interspecific introgression into *M. spretus*
- sweeps in favor of an unstable allele

~ 5% of mouse genome has similar signature.

Molecular biology of mutational processes:

- inter-locus gene conversion
- segmental duplication/deletion

Acknowledgments

Dept of Genetics

Rachel McMullan

Amelia Clayshulte

John Didion

Tim Bell

Fernando Pardo-Manuel de Villena

Dept of Computer Science

J Matthew Holt

Leonard McMillan

Funding

P50 GM076468

F30 MH103925

Data

Sanger Center/Wellcome Trust

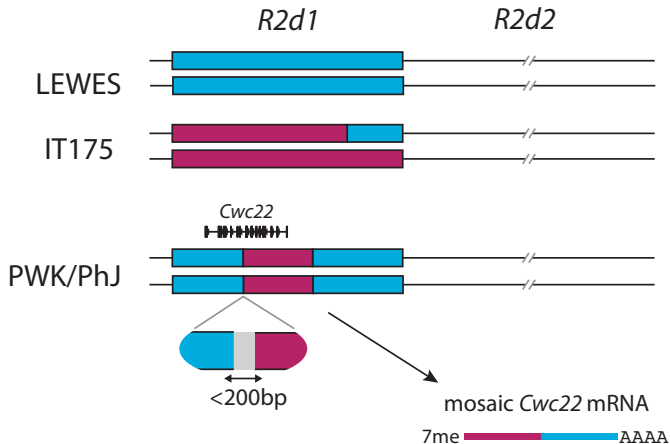
The Jackson Lab

Wild mice

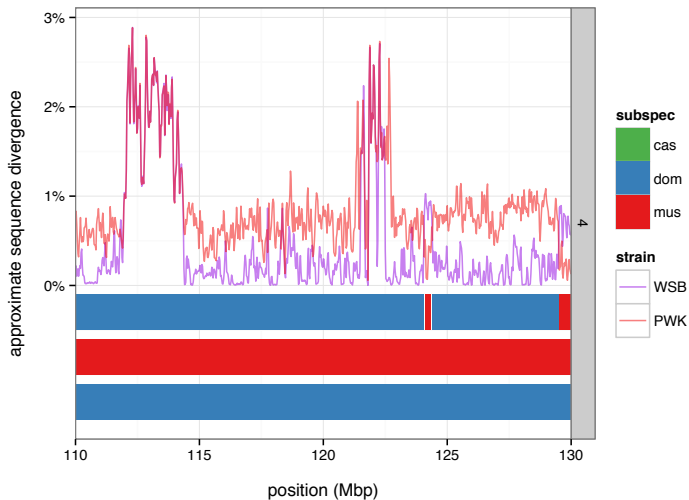
Jeremy Searle, Francois Bonhomme, Pierre Boursot, Janice Britton-Davidian, Ricardo Castiglia, Eva Giagia-Athan-asopolou, Sofia Gabriel, Silvia Garagna, Sofia Grize, Isla Gündüz, Bettina Harr, Heidi Hauffe, Jeremy Herman, Leon Kontrimavicius, Anna Lindholm, Maria de Luz Mathias, George Mistainas, Jaroslav Pialek, Priscilla Tucker, Jacint Ventura and Jan Wojcik.

Supplementary material

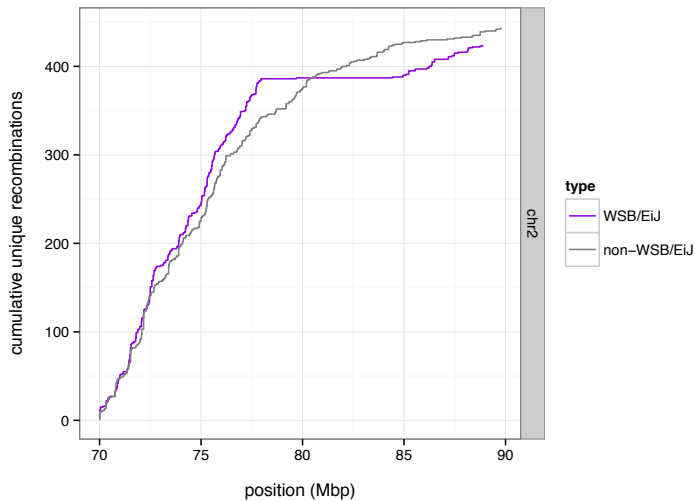
Inter-locus gene conversion



Signature of paralogous variation



Absence of recombination between *R2d1*, *R2d2*



Olfactory receptor cluster distal to *R2d2*