

# On the subspecific origin of the laboratory mouse

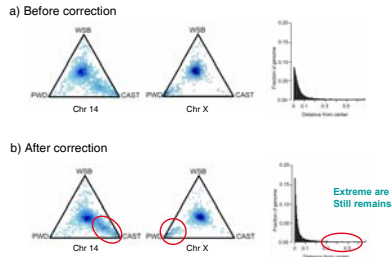


Hyuna Yang<sup>1</sup>, Timothy A Bell<sup>2</sup>,  
Gary A. Churchill<sup>1</sup> and Fernando Pardo-Manuel de Villena<sup>2</sup>  
1, The Jackson Laboratory, Bar Harbor, ME 2, Department of Genetics, UNC-Chapel Hill, Chapel Hill, NC.  
Center for Genome Dynamics : <http://www.genomedynamics.org>

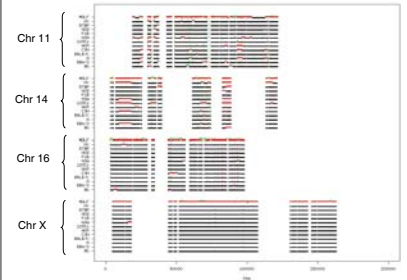
## Abstract

The genome of the laboratory mouse is thought to be a mosaic of regions with distinct subspecific origins. We have developed a high-resolution map of the origin of the laboratory mouse by generating 25,400 phylogenetic trees in 100 kb intervals spanning the genome. On average 92% of the genome is of *M. m. domesticus* origin and the distribution of diversity is strikingly non random among the chromosomes. There are large regions of extremely low diversity, representing blind spots for studies of natural variation and complex traits, as well as hot spots of diversity. In contrast with the mosaic model we found that the majority of the genome has intermediate levels of variation of intraspecific origin. Finally, the wild-derived mouse strains that are supposed to represent different mouse subspecies show substantial intersubspecific introgression. This has serious implications for evolutionary studies that assume these are pure representatives of a given subspecies.

## Correct the Minor Allele Frequency effect



## Assignment of intersubspecific origin in 72% of the genome



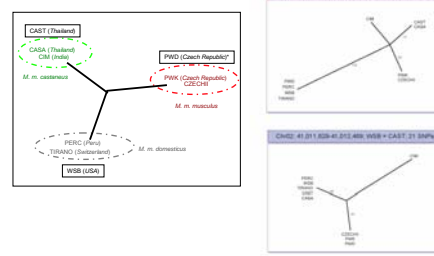
## The NIEHS Resequencing Project

Strains	BTBR T+tl/J KK/HJ NZW/LacJ	"Strains maximizing genetic diversity"
129S1/SvImJ A/J AKR/J BALB/cByJ C3H/HeJ DBA/2J FVB/NJ NOD/LLJ C57BL/6J	CAST/EiJ MOLF/EiJ PWD/PhJ WSB/EiJ	"Strains derived from the major taxonomic groups"
	Reference strain	

**Data**

# of SNPs (total)	8,322,543
# of SNPs (mapped)	8,262,814
Coverage (Mb)	1440.91

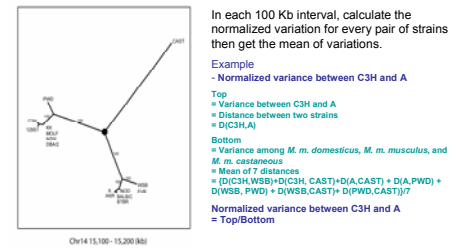
## Introgression in the reference strains: Direct confirmation



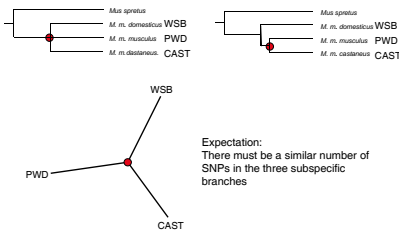
## Subspecific contribution in each strain

	B6	DBA/2J	A	BALB	C3H	AKR	129S1	NZW	FVB	NOD	BTBR	KK	MOLF
domesticus	0.92	0.91	0.94	0.95	0.92	0.94	0.91	0.87	0.96	0.93	0.92	0.86	0.11
musculus	0.07	0.07	0.05	0.04	0.07	0.05	0.08	0.11	0.03	0.06	0.06	0.12	0.74
castaneus	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.15

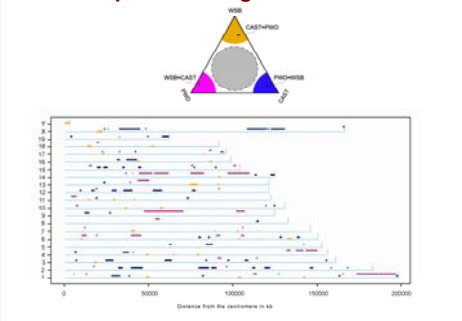
## Normalized variation in classical strains



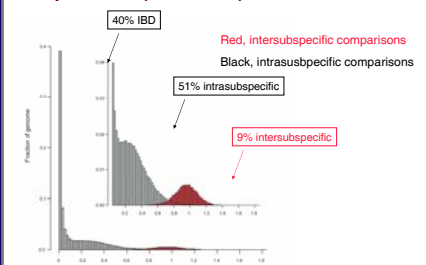
## Evolution - Expectation



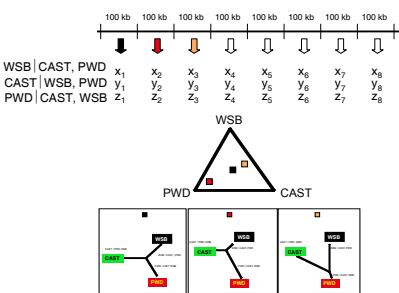
## Intersubspecific introgression



## Intraspecific and intersubspecific variation in classical mouse strains: Pairwise comparisons (≈1 million)



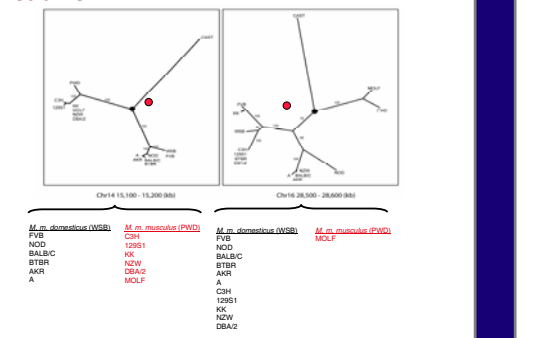
## Diagnostic SNPs: WSB, PWD and CAST



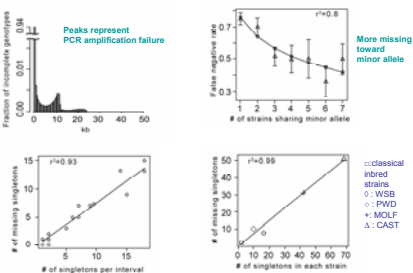
## Subspecific origin in classical and hybrid strains

1. Generate phylogenetic trees for 24,500 100-kb intervals.
2. Correct for effect of MAF on branch length.
3. Identify regions without introgression (72% of the genome).
4. Determine the origin based on topology.

## Subspecific origin in classical and hybrid strains



## Data - systemic biases



## Average variation among 11 classical inbred strains

