File S1
Modeling strategy for repeat transmission frequency comparisons accounting for paternal CAG repeat differences

To control for potential confounding effects of paternal CAG size on the frequency of unstable transmissions between distinct strains and/or lines, we compared actual transmission frequencies in test datasets with expected frequencies derived from simulated data based on a reference dataset.

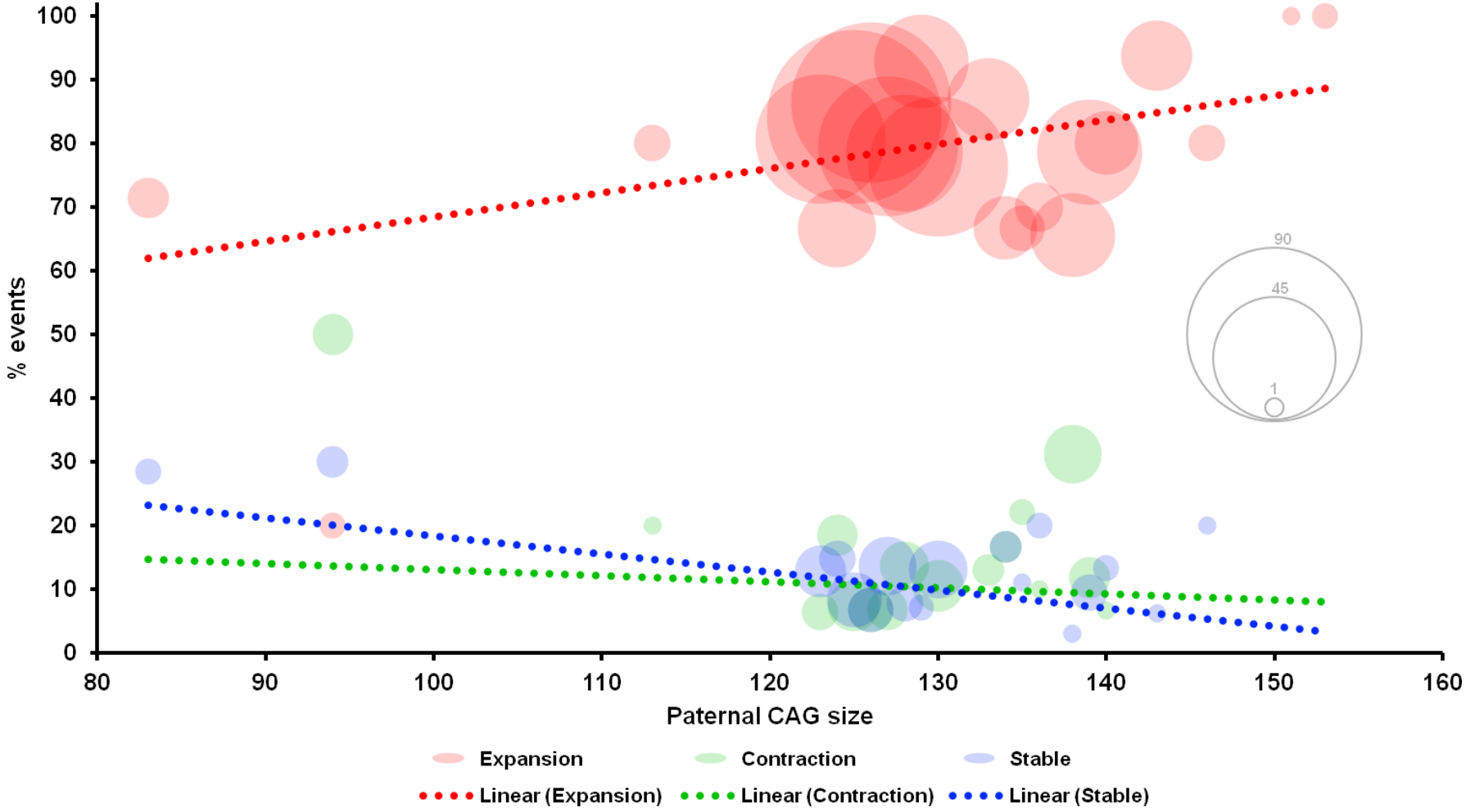
Reference dataset	dataset N	Test dataset(s)	dataset N	Notes
50% B6J (CHGR)	354	50% B6J (CHGR)	353	validation
B6J (CHGR)	707	129 (CHGR) CD1 (CHGR) FVB (CHGR) DBA (CHGR) B6N (CHGR) B6J (CHGR)	213 439 180 64 226 707	strain comparison
CD1^{neo-} (CHGR)	439	CD1^{neo+} (CHGR)	152	neo cassette comparison (I)
Q175^{neo+} (JAX)	9172	Q175^{neo-} (JAX)	256	neo cassette comparison (II)

Notes

Reference datasets always possessed a higher total number of transmissions in order to build optimal frequency vs. paternal CAG models.

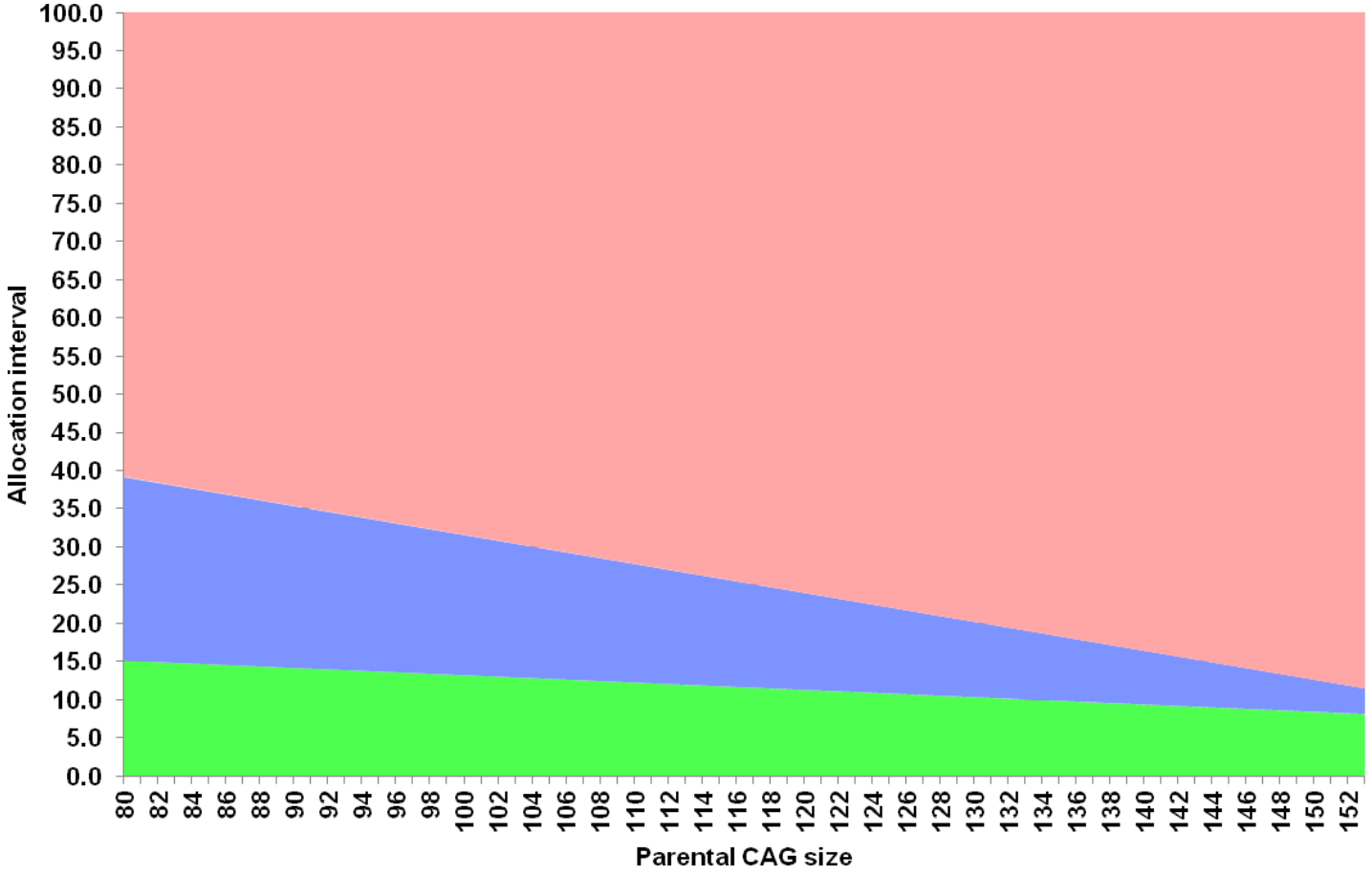
Figures in this supplementary file are representations of B6J (CHGR) as the reference dataset and CD1 (CHGR) as the test dataset.

Step 1. Linear weighted trend lines for percent of events (expansions, contractions and stable transmissions) per paternal CAG size were calculated for the reference dataset (e. g. CHGR B6J strain).



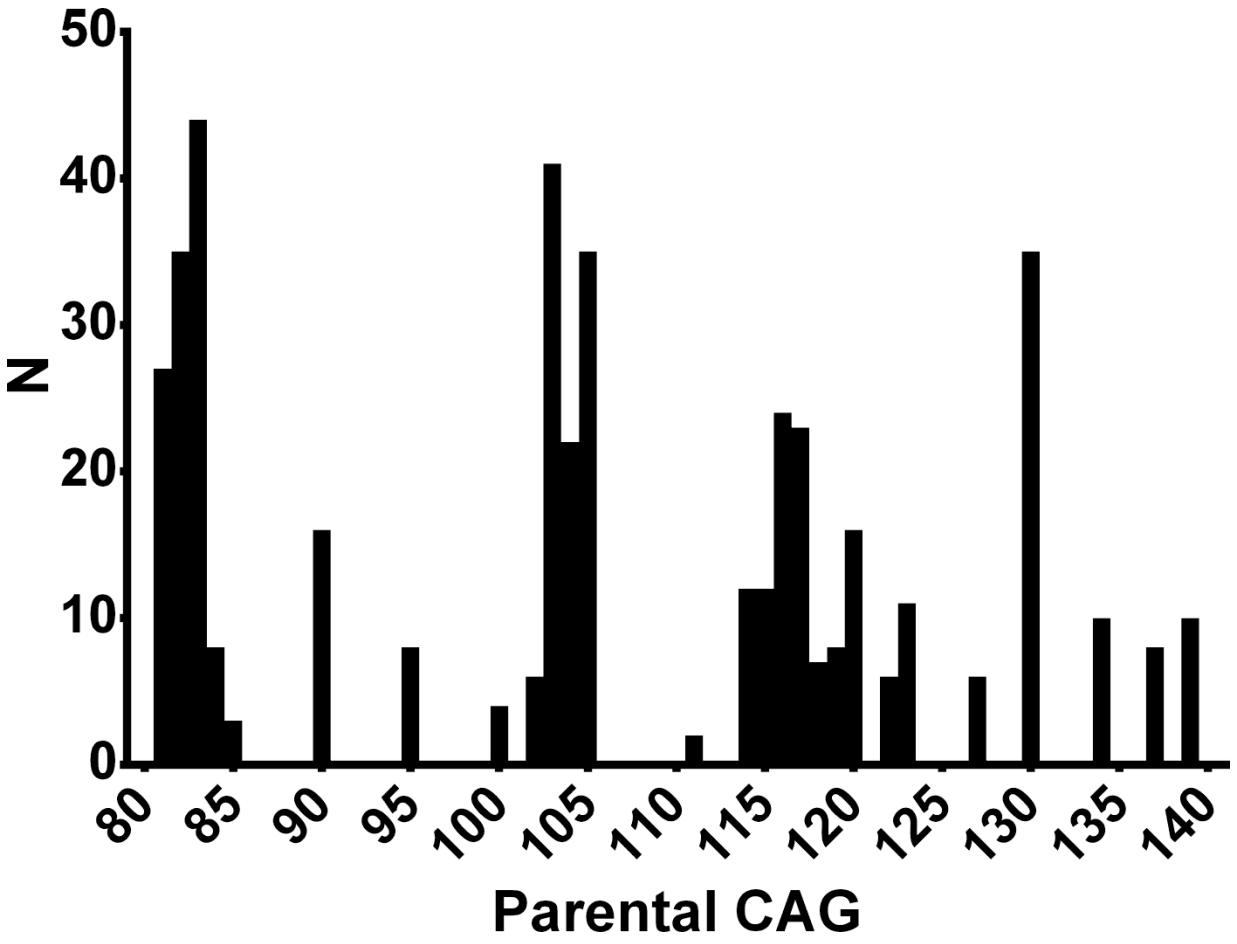
Events with null frequency (N=0) are considered for trend line weighing but are not depicted as bubbles.

Step 2. Based on the weighted trend lines we determined frequency intervals for each event over the range of CAG sizes in the reference dataset.



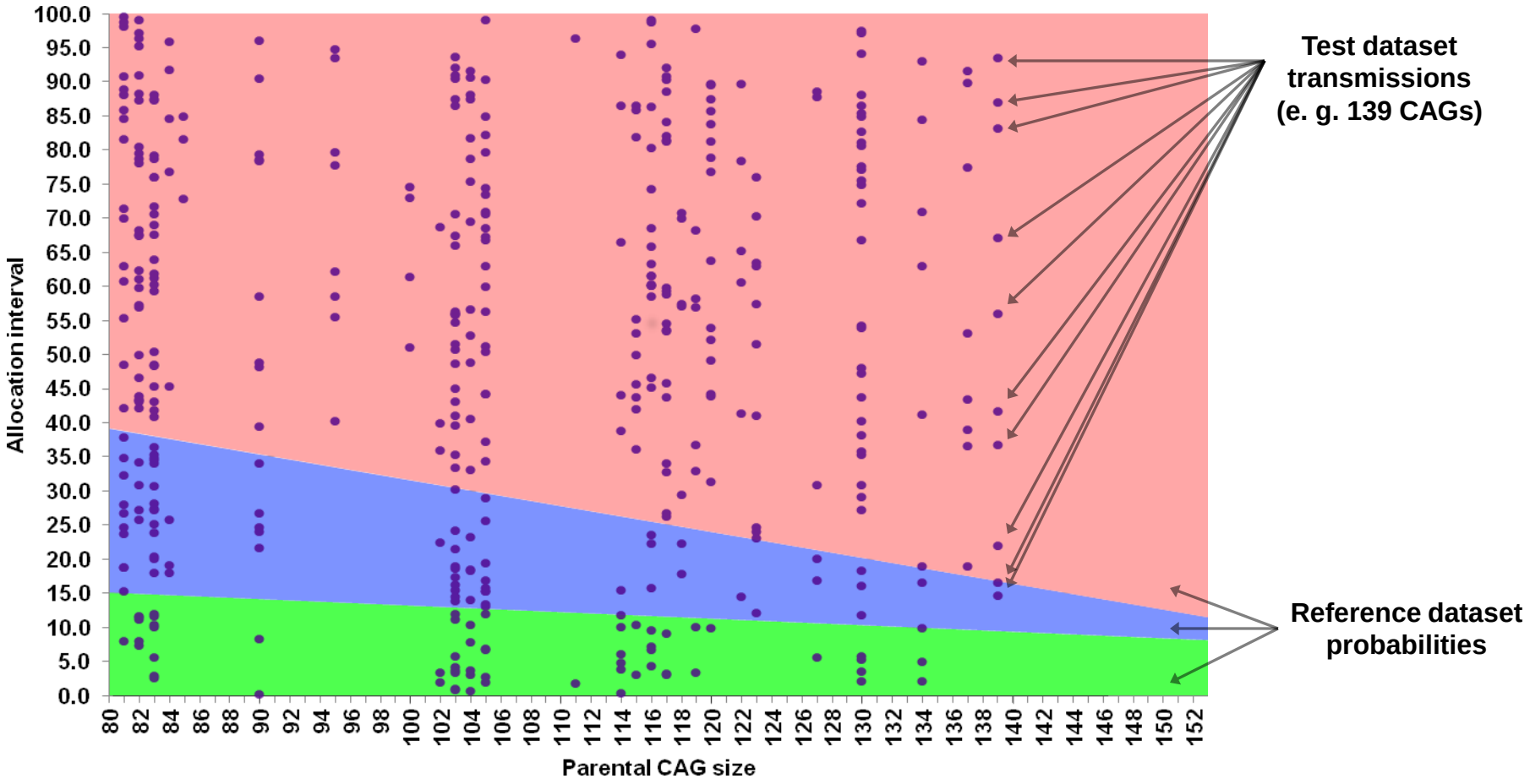
Expansion – red; Stable – blue; Contraction – green

Paternal CAG repeat size distribution in the CHGR CD1 test dataset.



N_{total} = 439

Step 3. A random number between 0.0 and 100.0 was generated for each transmission and, based on the paternal CAG length, was allocated to contraction / expansion / stable transmission according to the frequency intervals defined by the reference dataset.



iteration 1

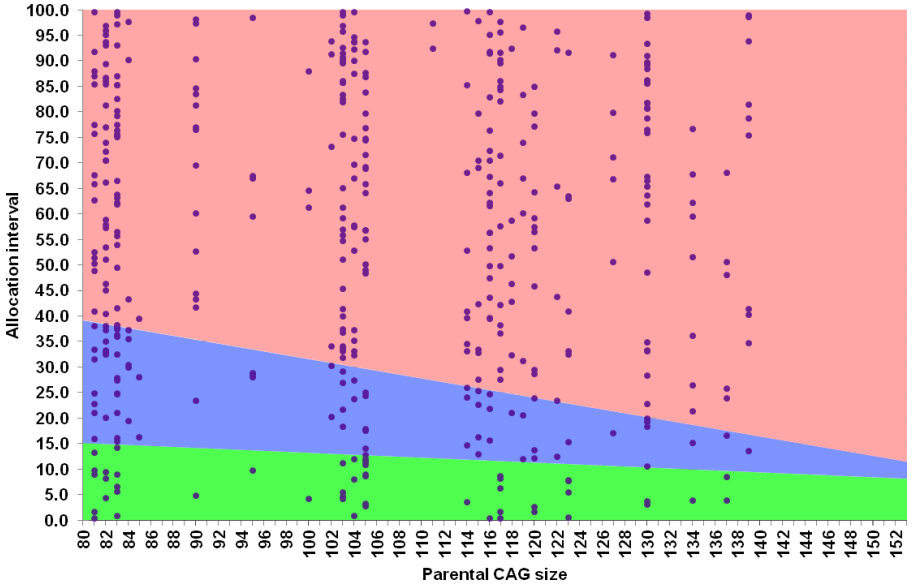
Overall distribution

%Expansion	68.8
%Stable	17.5
%Contraction	13.7

Each dot represents a unique transmission (N=439)

Step 4. 1,000 iterations of the random number generation and event allocation were performed.

iteration 2

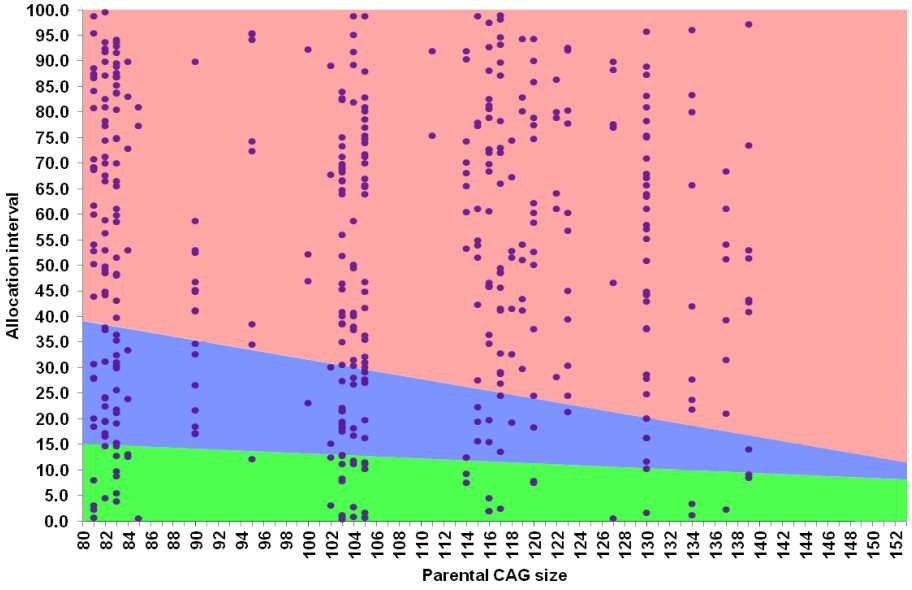


%Expansion 71.5
%Stable 16.9
%Contraction 11.6

…

…

iteration 1000

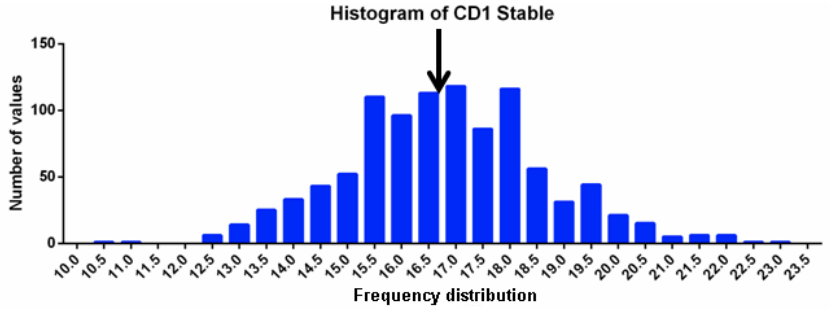


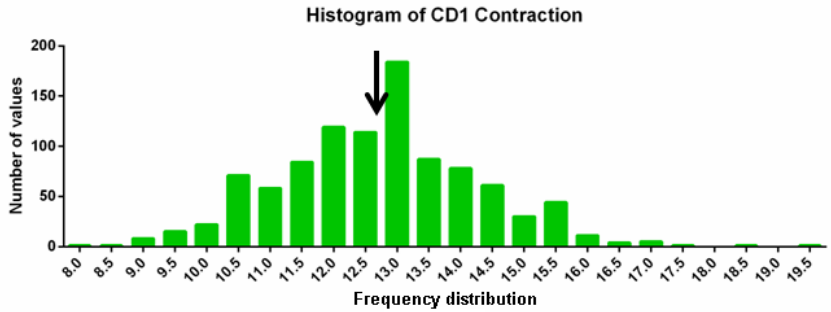
%Expansion 70.8
%Stable 17.8
%Contraction 11.4

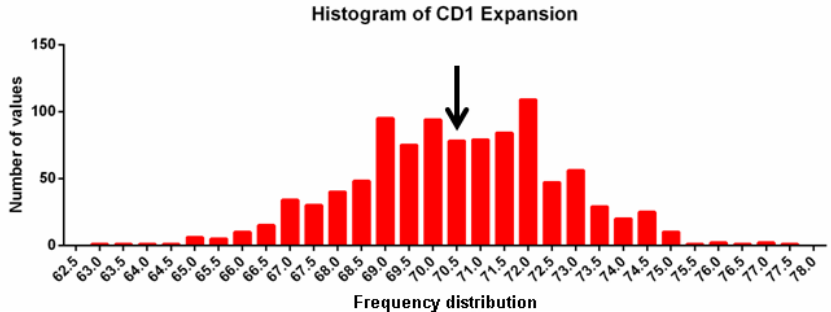
Step 5. The average dataset was determined and characterized.

Line	Event	Mean	St. dev.	95% CI
CD1	%Stable	16.8	1.8	0.11
	%Contraction	12.7	1.6	0.10
	%Expansion	70.5	2.2	0.13

Average dataset for the CD1 strain

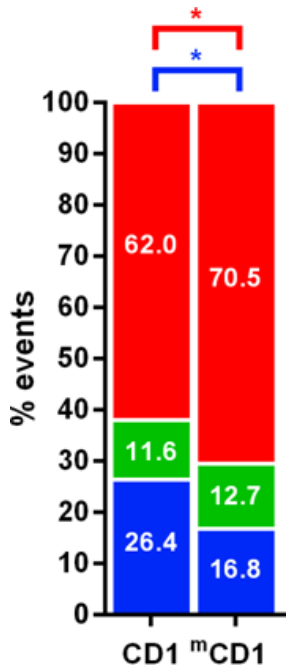






Frequency distributions (binned at 0.5% intervals for figure simplification) for stable transmissions, contractions and expansions across the 1,000 simulated datasets. Arrows indicate the value of the average dataset.

Step 6. Statistical analyses were performed to determine significant differences between observed and expected frequencies, as detailed in the Methods section. Validation and results are present in Figure S2, Figure 6A, Figure 7A and C.



Comparison of expansions, contractions and stable transmissions frequencies in the CD1 strain. Observed frequencies (left) and modeled(^{m})/expected frequencies (right).

(Partial reproduction of Figure 6A)

Trend lines determined for the reference datasets:

Validation [50% B6J (CHGR), N=354]
%Expansions = 0.214 × Parental CAG + 52.929
%Contractions = 0.053 × Parental CAG + 3.095
%Stable = -0.267 × Parental CAG + 43.976

Strain comparison [B6J (CHGR)]
%Expansions = 0.380 × Parental CAG + 30.423
%Contractions = -0.096 × Parental CAG + 22.77
%Stable = -0.284 × Parental CAG + 46.807

neo cassette comparison (I) [CD1^{neo-} (CHGR)]
%Expansions = 0.596 × Parental CAG - 0.998
%Contractions = -0.156 × Parental CAG + 28.08
%Stable = -0.44 × Parental CAG + 72.918

neo cassette comparison (II) [Q175^{neo+} (JAX)]
%Expansions = -0.581 × Parental CAG + 193.207
%Contractions = 0.524 × Parental CAG - 88.907
%Stable = 0.057 × Parental CAG - 4.3