Supplementary Information for

Mutability of mononucleotide repeats, not oxidative stress, explains the discrepancy between laboratory-accumulated mutations and the natural allele-frequency spectrum in *C. elegans*

Moein Rajaei[1], Ayush Shekhar Saxena[1], Lindsay M. Johnson[1], Michael C. Snyder[1], Timothy A. Crombie[1,2], Robyn E. Tanny[2], Erik C. Andersen[2], Joanna Joyner-Matos[3], and Charles F. Baer[1,4]

1- Department of Biology, University of Florida, Gainesville, FL USA
2- Department of Molecular Biosciences, Northwestern University, Evanston, IL USA
3- Department of Biology, Eastern Washington University, Cheney, WA USA
4- University of Florida Genetics Institute, Gainesville, FL USA

Table of Contents

Section Page

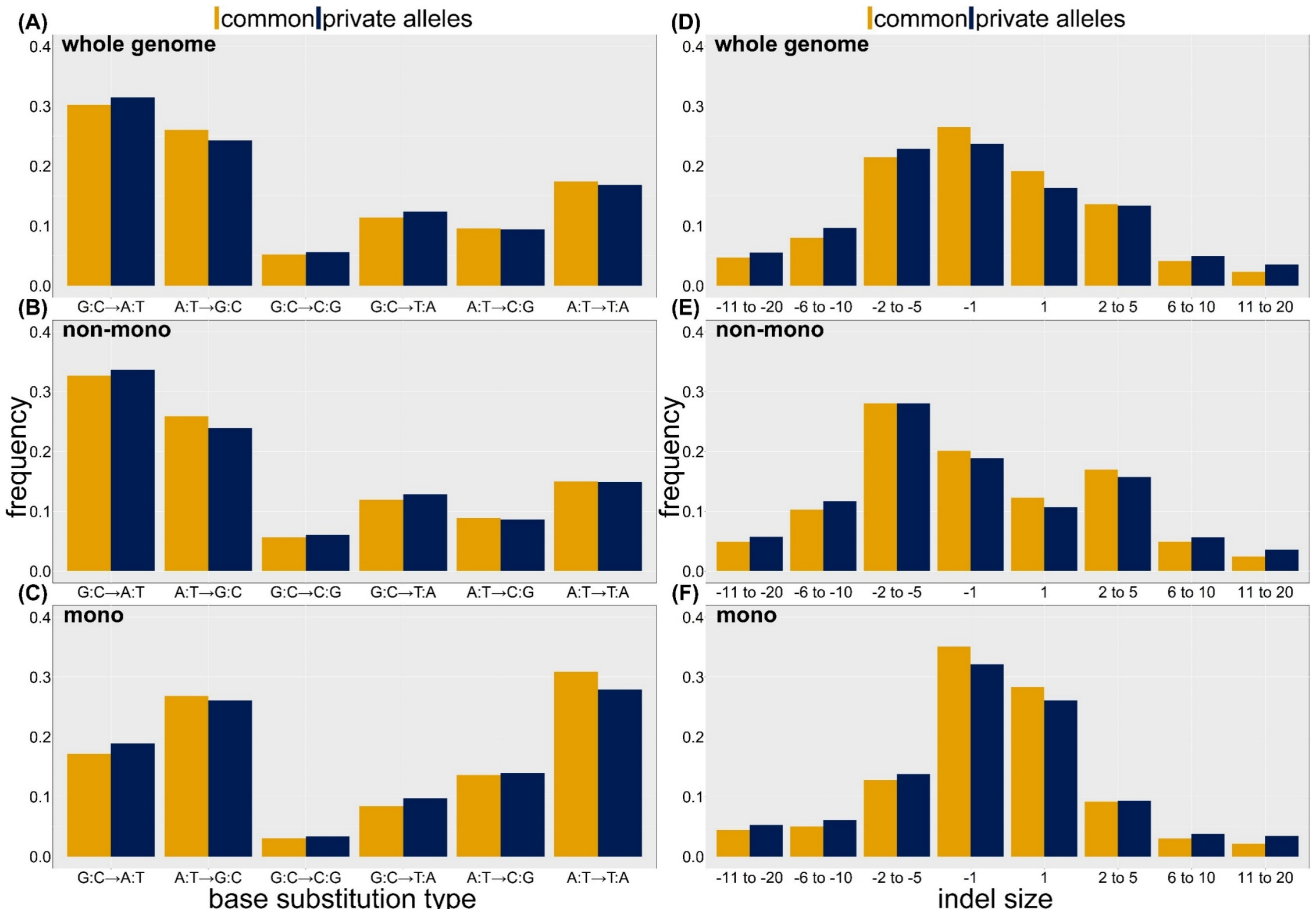

Figure S1. Spectra of private alleles (n=1/N, blue) and common variants (orange). Left (**A-C**), Base-substitution spectra. Right (**D-F**), Indel spectra. Top panels (**A,D**), whole-genome; middle panels (**B,E**), non-mononucleotide sequence; bottom panels (**C,F**), mononucleotide sequence.

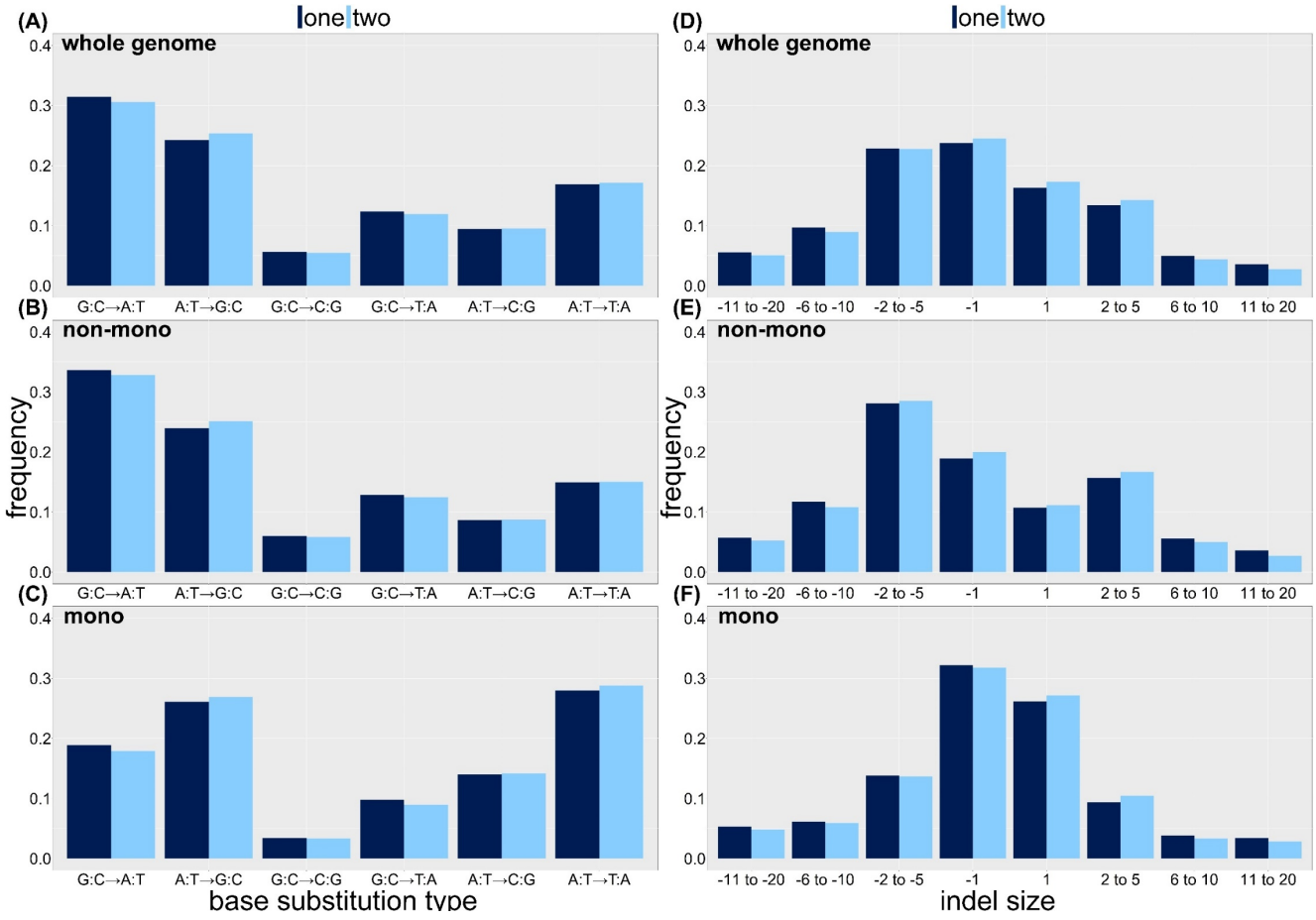

Figure S2. Spectra of segregating singleton (dark blue, n=1/N) and doubleton (light blue, n=2/N) variants. Left (**A-C**), base-substitution spectra. Right (**D-F**), indel spectra. Top panels (**A,D**), whole-genome; middle panels (**B,E**), non-mononucleotide sequence; bottom panels (**C,F**), mononucleotide sequence.

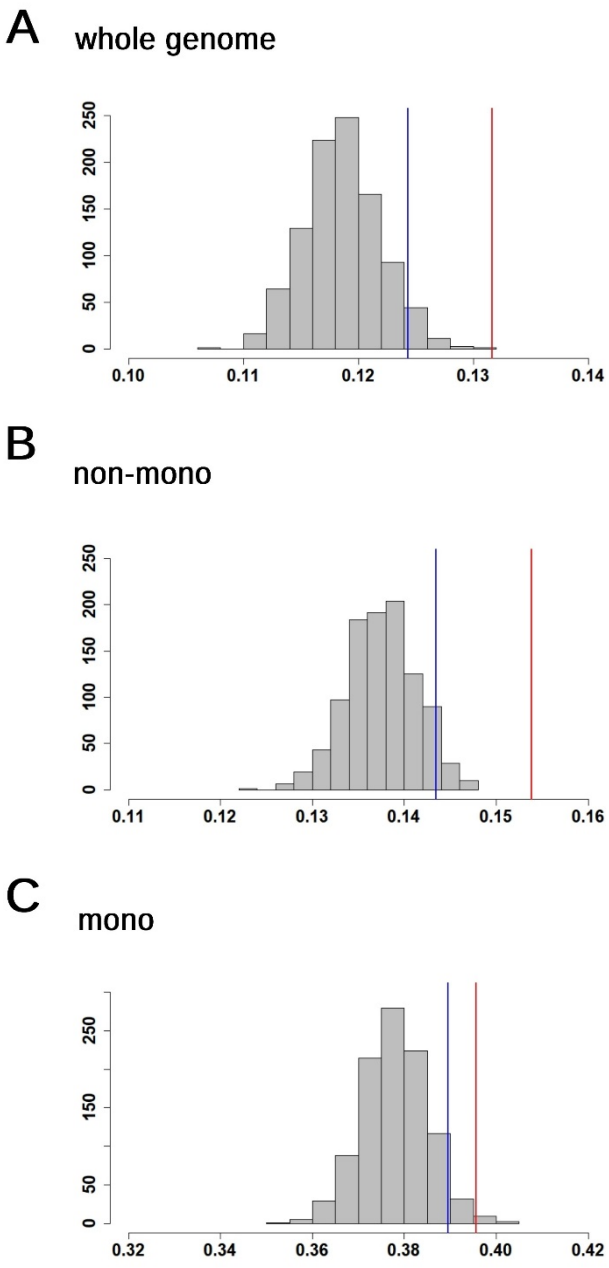

Figure S3. Parametric bootstrap distributions. Red lines show the observed values; blue lines show the upper 95% confidence limit of simulated D(KL).

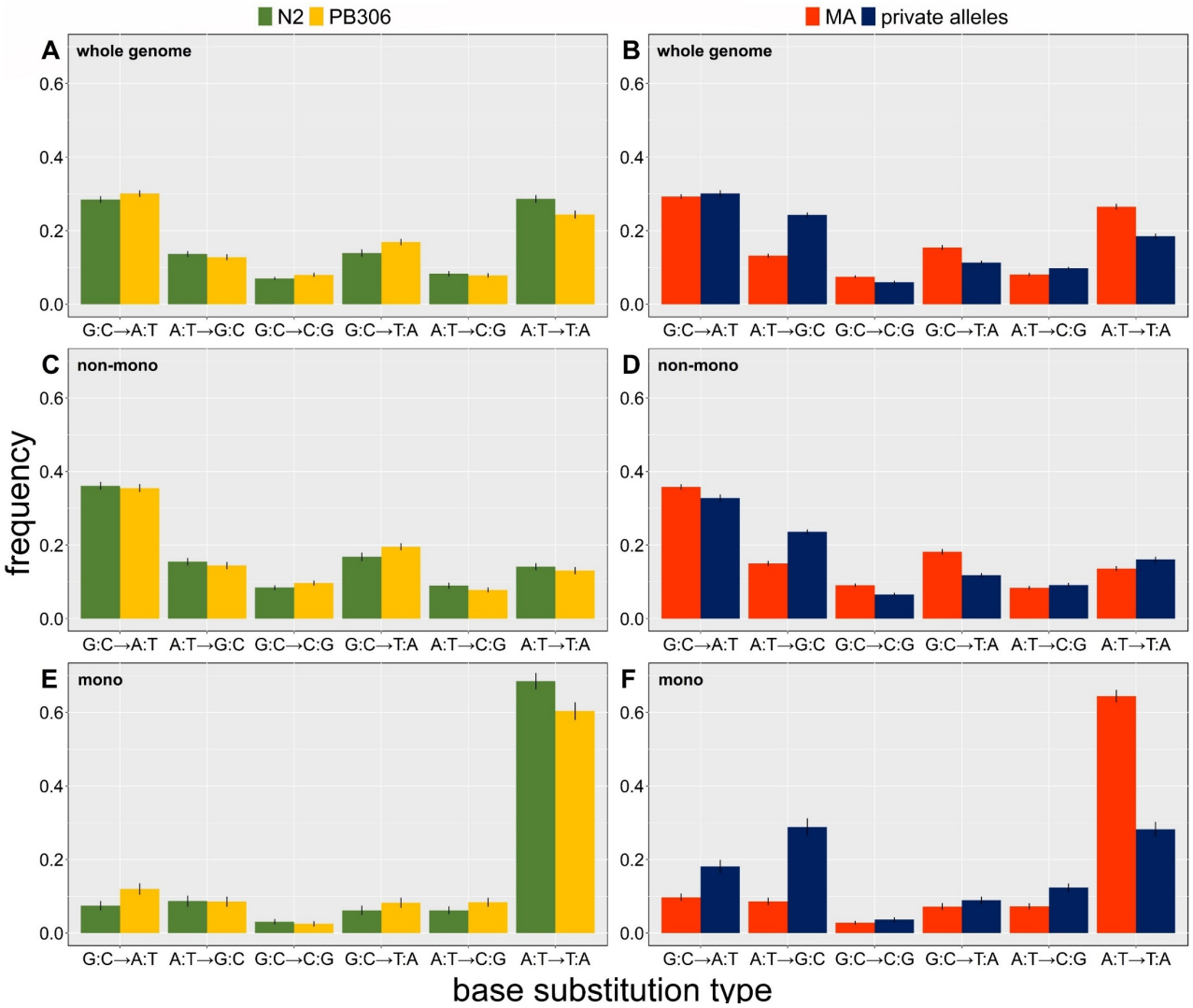

Figure S4. Base-substitution spectra of a subset of N2 and PB306 MA lines (left panel) and MA vs private alleles (right panel). Left panel (**A,C,E**), N2 (green) and PB306 (yellow); Right panel (**B,D,F**), wild isolate private alleles (blue) and MA means (red). Top panels (**A, B**) whole-genome; middle panels (**C, D**) non-mononucleotide sequence; bottom panels (**E, F**) mononucleotide sequence. Error bars show SEM. Description of the samples and the variant-calling pipelines are given in the Supplemental Materials.

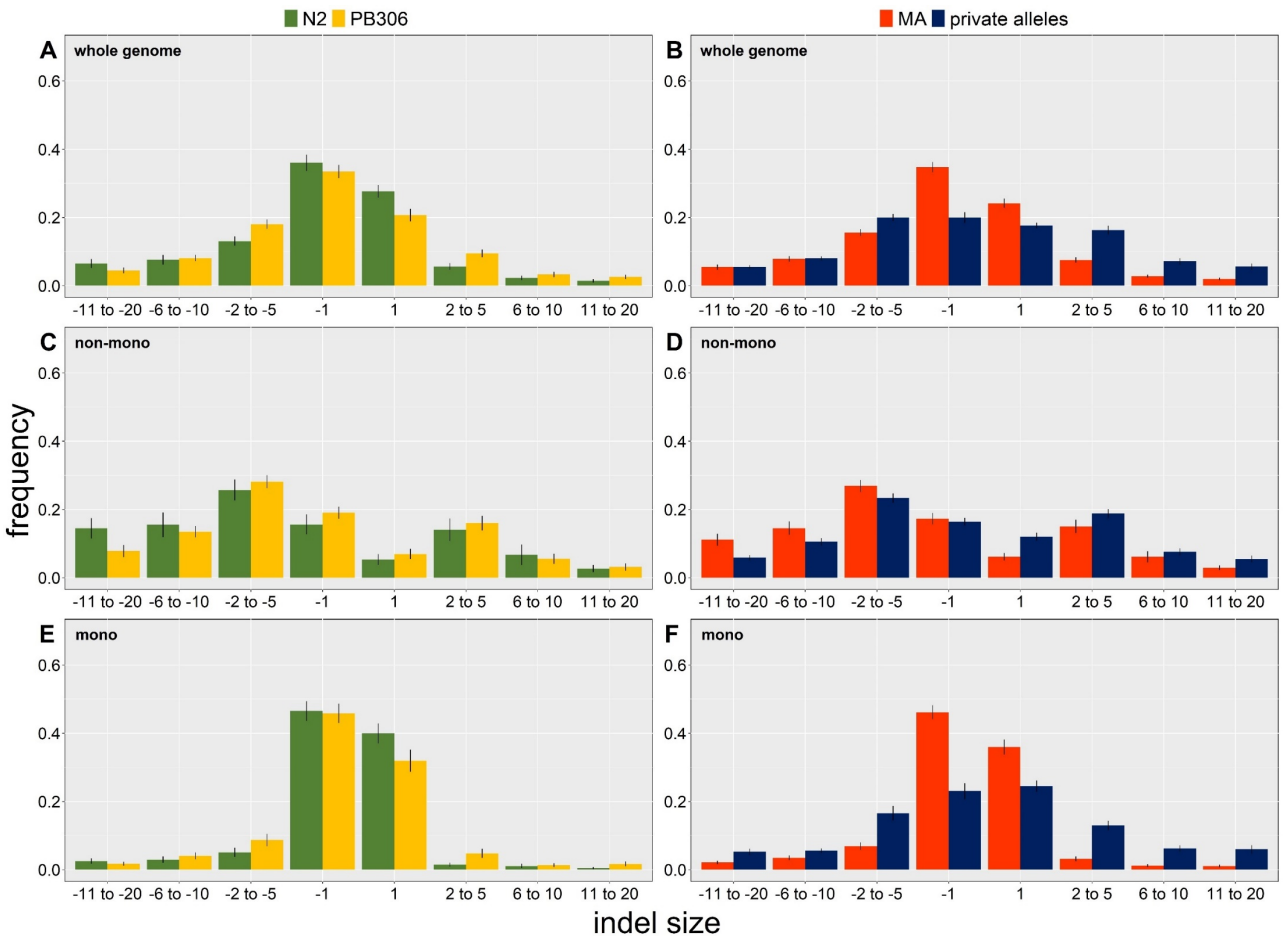

Figure S5. Indel spectra of a subset of N2 and PB306 MA lines (left panel) and MA vs private alleles (right panel). Left panel (**A,C,E**), N2 (green) and PB306 (yellow); Right panel (**B,D,F**), wild isolate private alleles (blue) and MA means (red). Top panels (**A, B**) whole-genome; middle panels (**C, D**) non-mononucleotide sequence; bottom panels (**E, F**) mononucleotide sequence. Error bars show SEM.

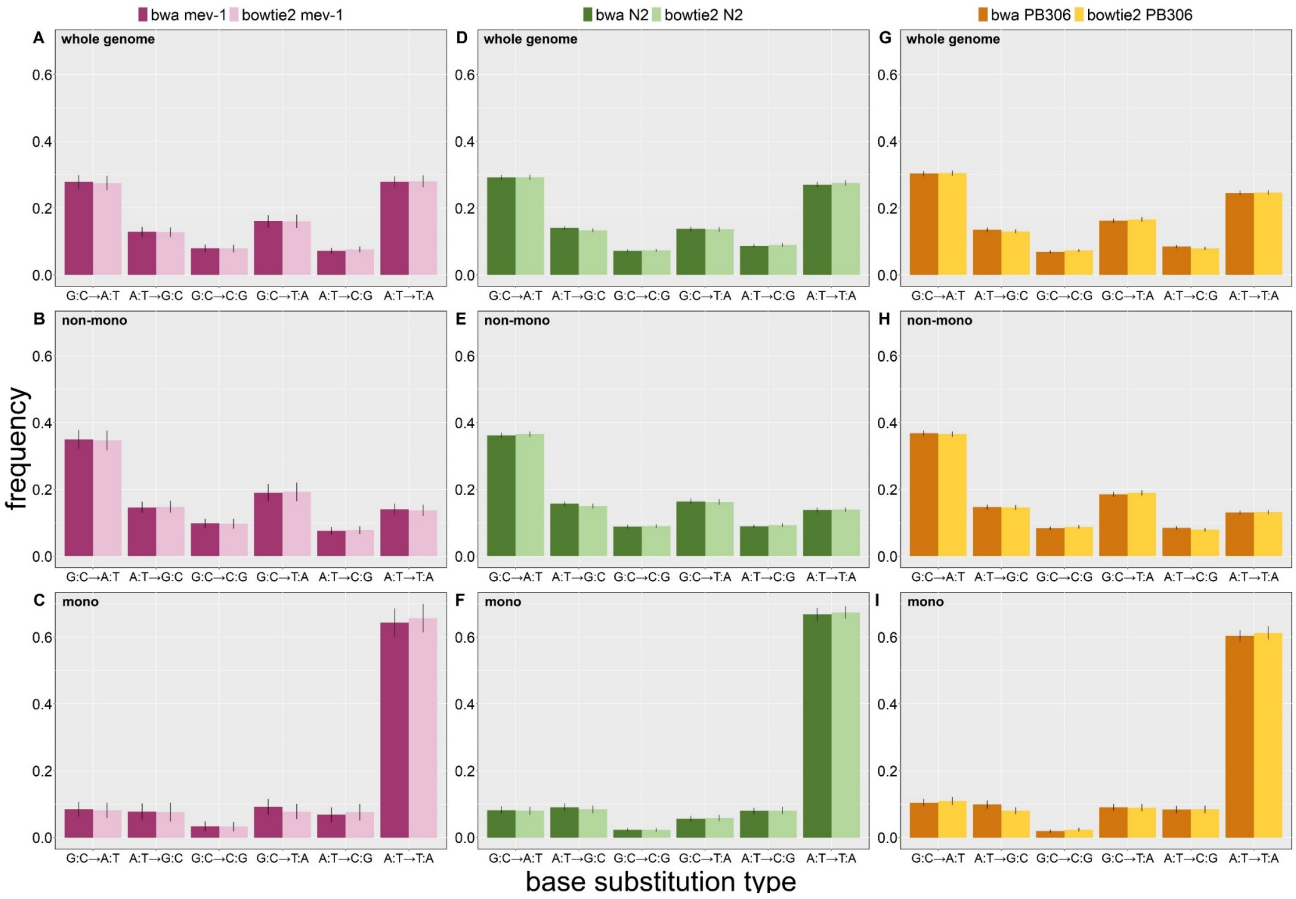

Figure S6. Base-substitution spectra of *mev-1,* N2 and PB306 MA lines using two different mapping programs, Bowtie2 and BWA. Left (**A-C**), *mev-1* MA lines. Middle (**D-F**) N2 MA lines. Right (**G-I**), PB306 MA lines. Top panels (**A, D, G**) whole-genome; middle panels (**B, E, H**) non-mononucleotide sequence; bottom panels (**C, F, I**) mononucleotide sequence. Error bars show SEM.

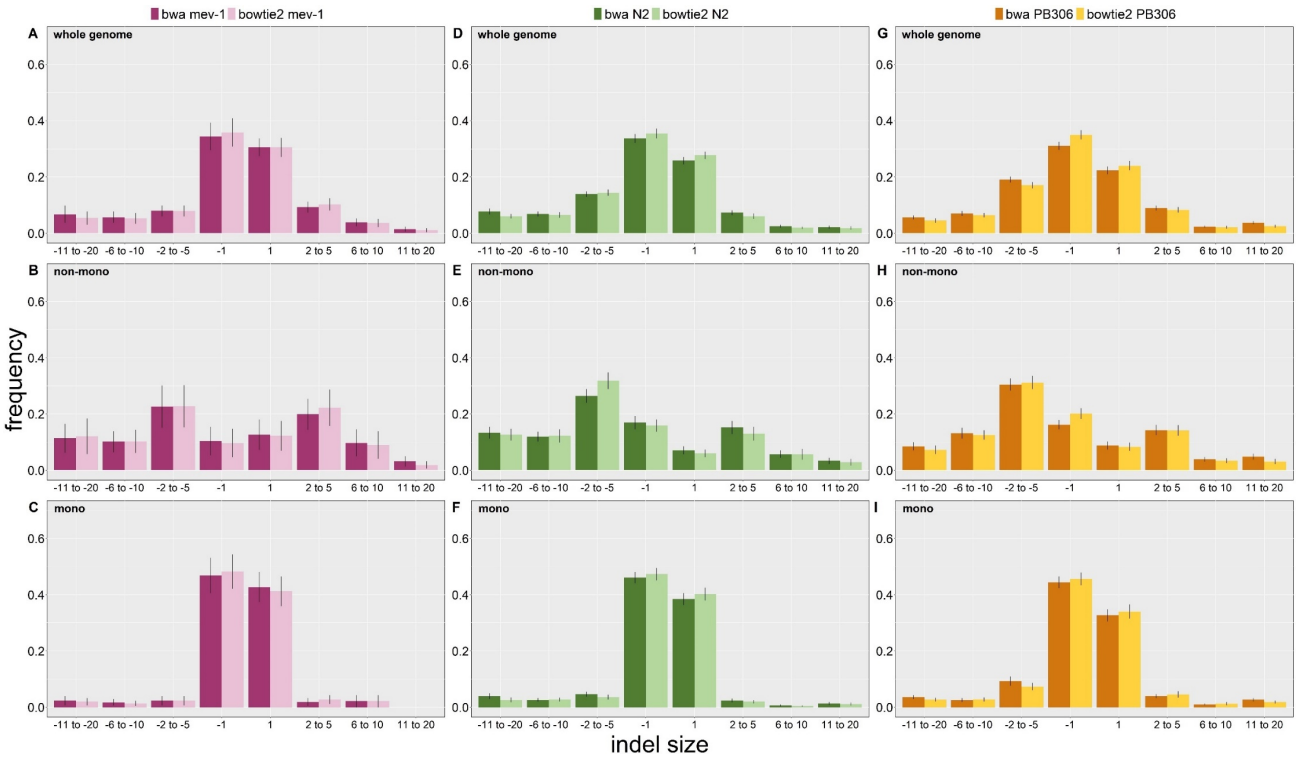

Figure S7. Indel spectra of *mev-1,* N2 and PB306 MA lines using two different mapping programs, Bowtie2 and BWA. Left (**A-C**), *mev-1* MA lines. Middle (**D-F**) N2 MA lines. Right (**G-I**), PB306 MA lines. Top panels (**A, D, G**) whole-genome; middle panels (**B, E, H**) non-mononucleotide sequence; bottom panels (**C, F, I**) mononucleotide sequence. Error bars show SEM.

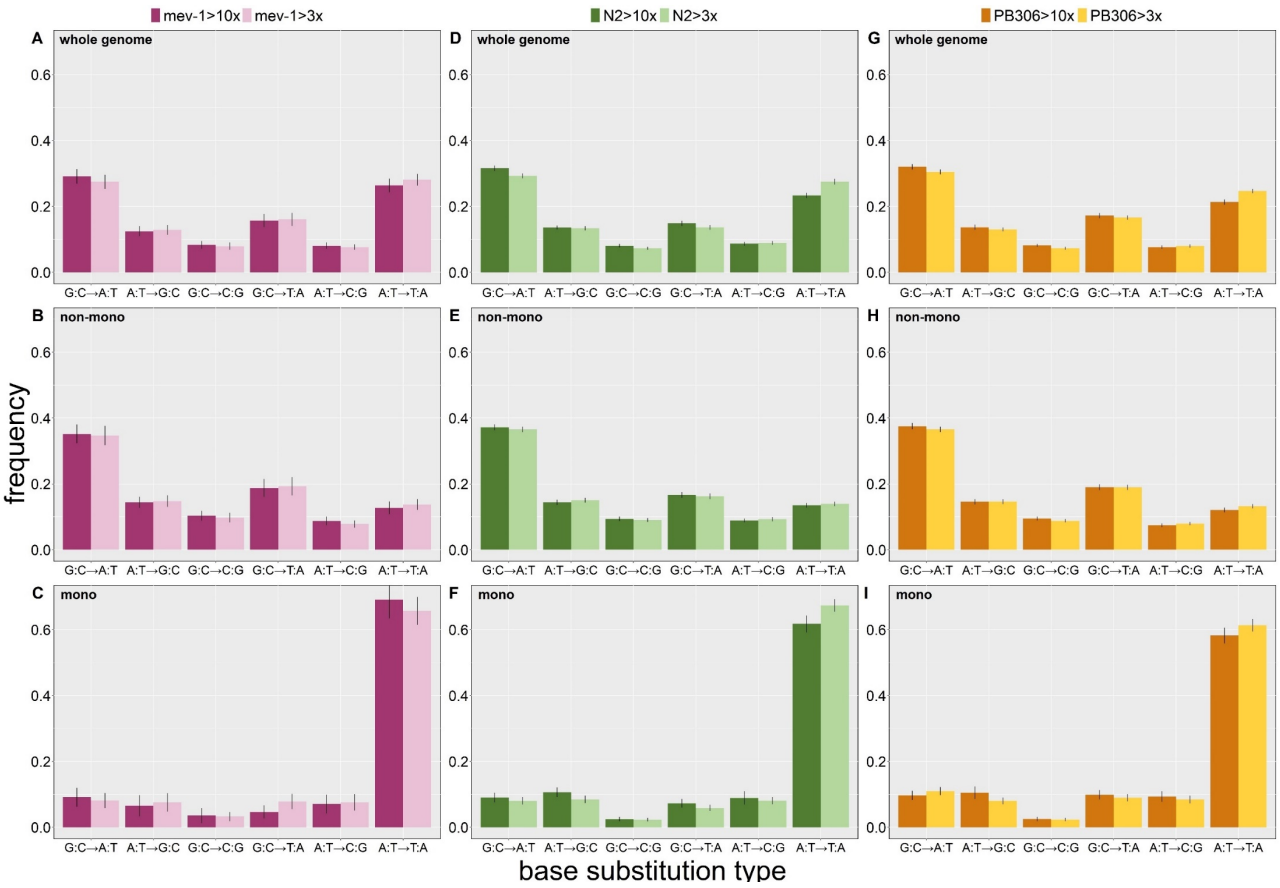

Figure S8. Base-substitution spectra of *mev-1,* N2 and PB306 MA lines using two different coverage threshold, 10x (dark colors) and 3x (light colors). Left (**A-C**), *mev-1* MA lines. Middle (**D-F**) N2 MA lines. Right (**G-I**), PB306 MA lines. Top panels (**A, D, G**) whole-genome; middle panels (**B, E, H**) non-mononucleotide sequence; bottom panels (**C, F, I**) mononucleotide sequence. Error bars show SEM.

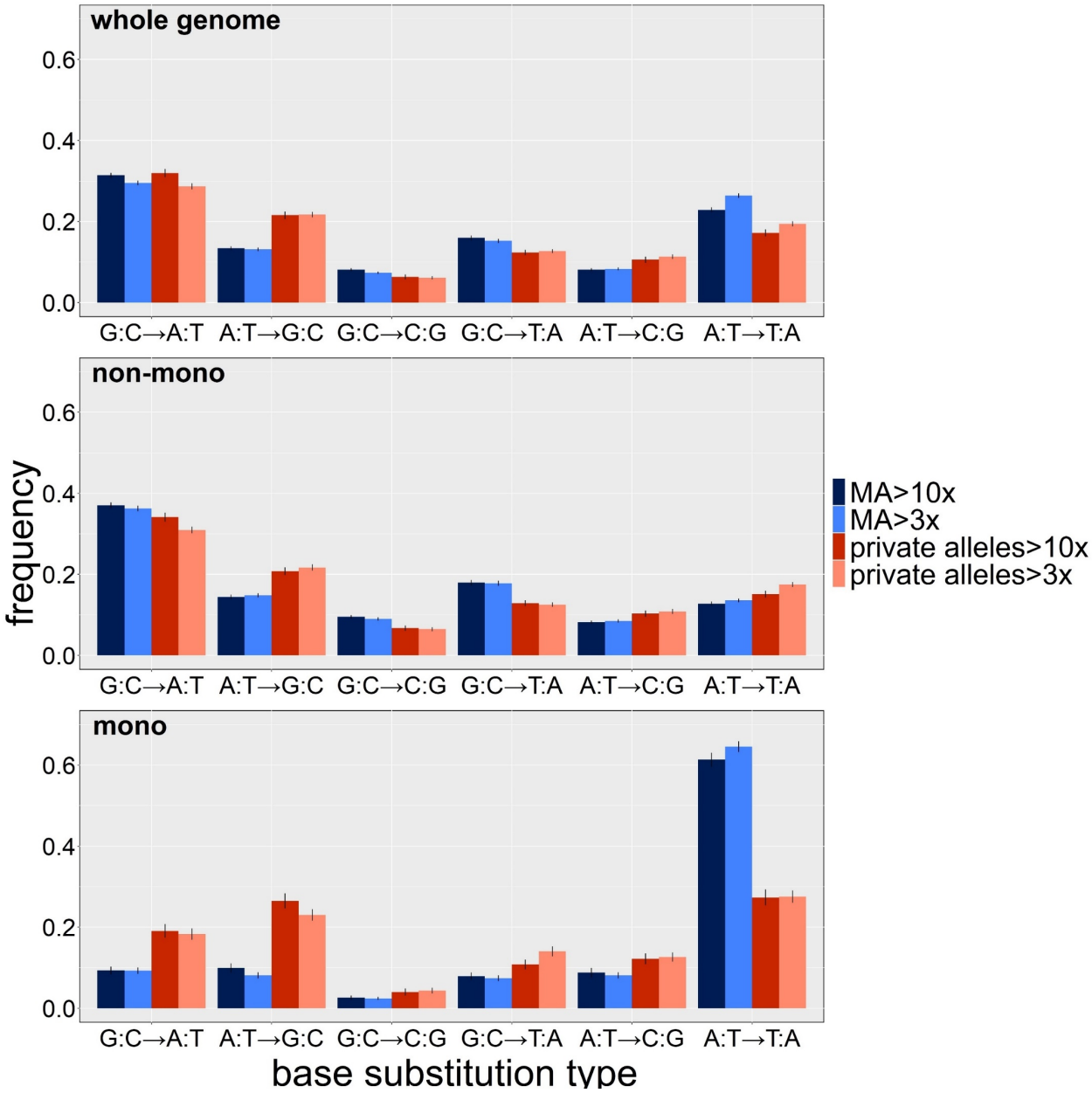

Figure S9. Base-substitution spectra of MA vs private alleles using two different coverage threshold, 10x (dark colors) and 3x (light colors). Top panels whole-genome; middle panels non-mononucleotide sequence; bottom panels mononucleotide sequence. Error bars show SEM.

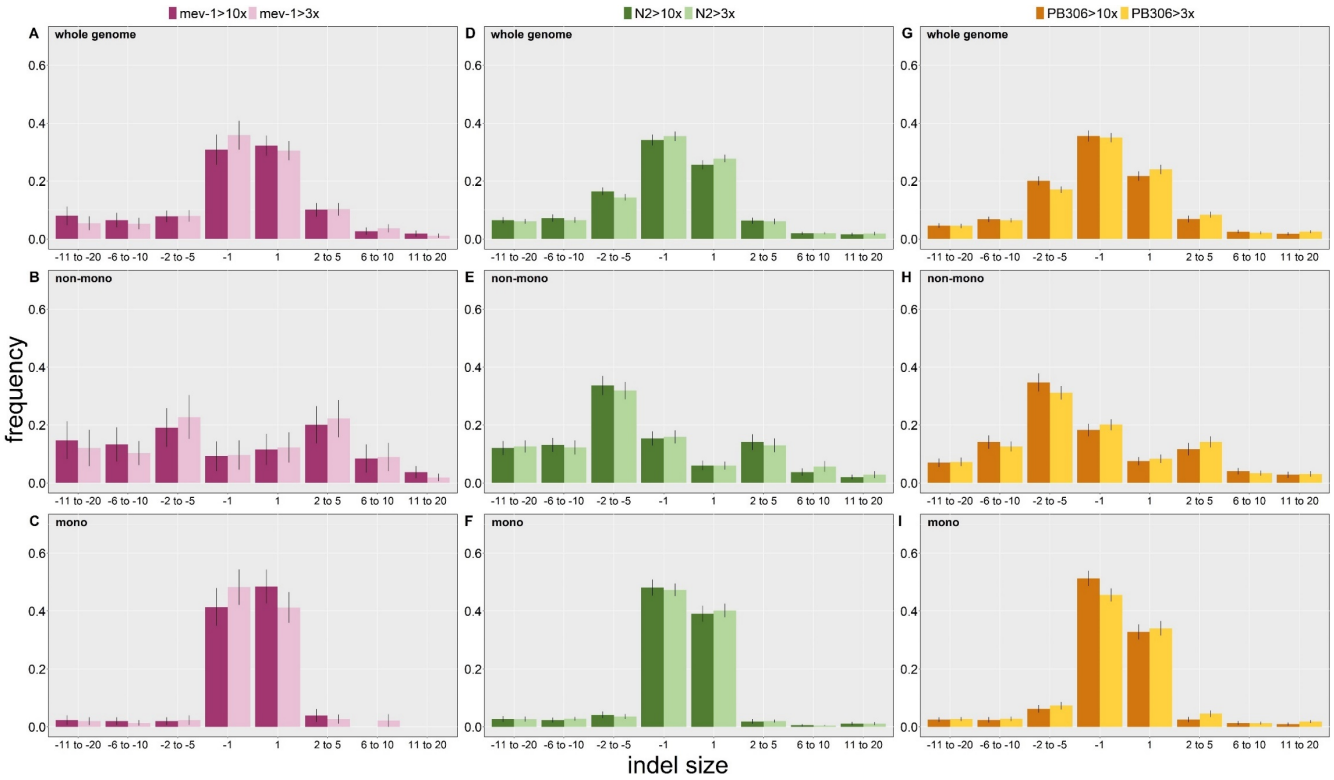

Figure S10. Indel spectra of *mev-1,* N2 and PB306 MA lines using two different coverage threshold, 10x (dark colors) and 3x (light colors). Left (**A-C**), *mev-1* MA lines. Middle (**D-F**) N2 MA lines. Right (**G-I**), PB306 MA lines. Top panels (**A, D, G**) whole-genome; middle panels (**B, E, H**) non-mononucleotide sequence; bottom panels (**C, F, I**) mononucleotide sequence. Error bars show SEM.

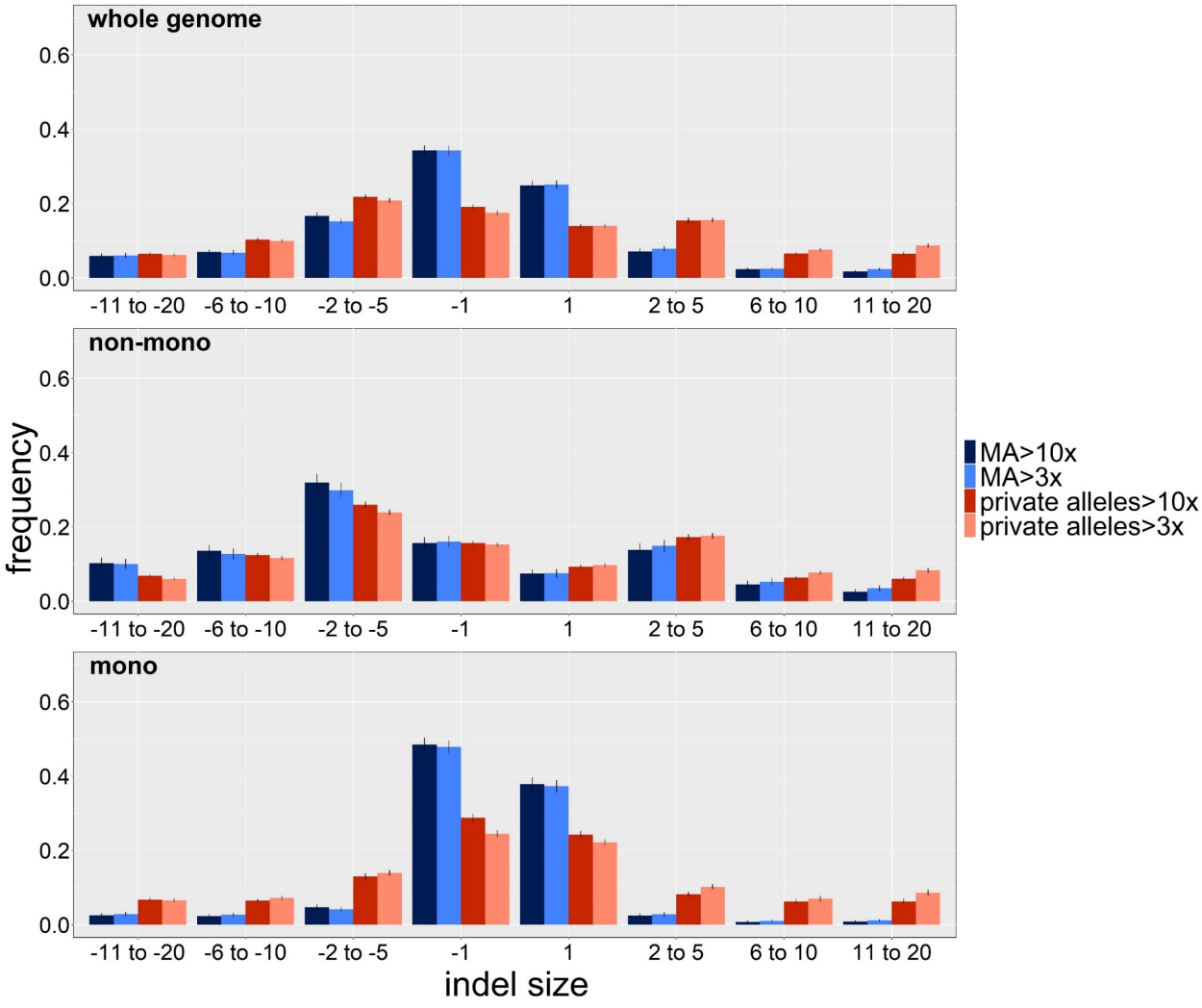

Figure S11. Indel spectra of MA vs private alleles using two different coverage threshold, 10x (dark colors) and 3x (light colors). Top panels whole-genome; middle panels non-mononucleotide sequence; bottom panels mononucleotide sequence. Error bars show SEM.

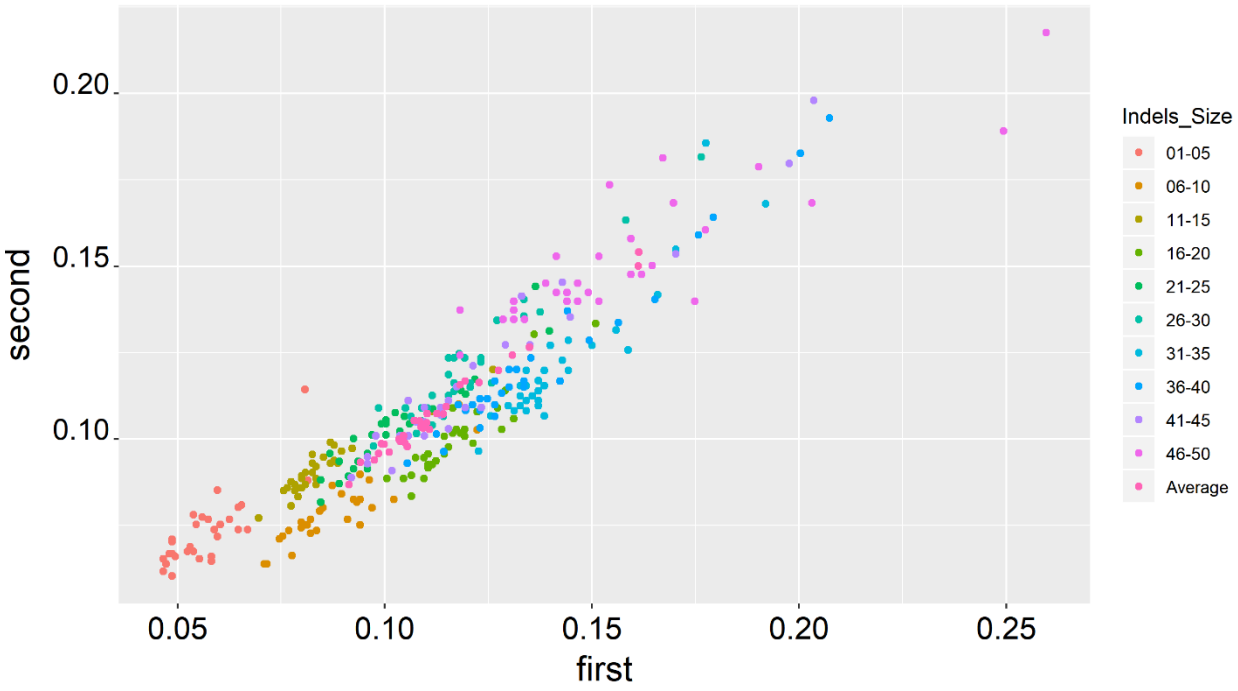

Figure S12. Plot of the failure to recall rates of the two simulated dummy data sets; the first simulation is on the x-axis, the second simulation is on the y-axis ($r=0.93$, $P << 0.0001$). Each data point of a given color represents an individual MA line (n=30 simulated lines). See section 12 for explanation.

Table S1 is submitted as an excel file with this name:
Supplemental_Table_S1.xlsx

Supplemental Table S2. Tests of fixed effects. See Appendix A1.5 for details of the general linear model (GLM) and Table 1 in the main text for trait values. The columns in tables below are: 1. Description of the fixed effect; 2. Degrees of freedom, determined by the Kenward-Roger method; 3. F-statistic; 4. P-value, not corrected for multiple tests. Further descriptions are underneath the tables.

A. Fixed Effect (μ_{BS}, by type)	DF (num, den)	F	Pr>F
Base-substitution type (6)	5,130	267.1	<0.0001
Strain (3)	2,222	5.25	0.0059
Strain x base-sub type	10,198	2.48	0.0081

A. Genome-wide base substitution mutation rates, μ_{BS}, including all three strains of MA lines, the six types of base-substitution, and the interaction.

B. Fixed Effect (μ_{BS}, pooled)	DF (num, den)	F	Pr>F
Strain (*mev-1* vs N2)	1,56.8	15.4	0.0002

B. Pairwise-test of pooled genome-wide base-substitution rate difference between *mev-1* and N2.

C. Fixed Effect (μ_{BS}, pooled)	DF (num, den)	F	Pr>F
Strain (*mev-1* vs PB306)	1,49.1	7.88	0.0072

C. Pairwise-test of pooled genome-wide base-substitution rate difference between *mev-1* and PB306.

D. Fixed Effect (μ_{BS}, pooled)	DF (num, den)	F	Pr>F
Strain (N2 vs PB306)	1,131	2.10	0.1494

D. Pairwise-test of pooled genome-wide base-substitution rate difference between N2 and PB306.

E. Fixed Effect ($\mu_{GC\rightarrow TA}$)	DF (num, den)	F	Pr>F
Strain (all 3)	2,55.1	4.09	0.0221

E. Planned comparison among strains of the GC→TA transversion rate.

F. Fixed Effect ($\mu_{AT\rightarrow TA}$)	DF (num, den)	F	Pr>F
Strain (all 3)	2,58.7	4.25	0.0189

F. Post hoc comparison among strains of AT→TA transversion rate.

Fixed Effect (μ_{INS})	DF (num, den)	F	Pr>F
G. Strain (all 3)	2, 53.9	5.62	0.0061

G. Comparison among strains of the genome-wide insertion rate.

H. Fixed Effect (μ_{DEL})	DF (num, den)	F	Pr>F
Strain (all 3)	2,60.8	5.74	0.0052

H. Comparison among strains of the genome-wide deletion rate.

I. Fixed Effect (μ_{INS})	DF (num, den)	F	Pr>F
Strain (*mev-1* vs N2)	1, 26.9	8.06	0.0085

I. Pairwise test of genome-wide insertion rate difference between *mev-1* and N2.

J. Fixed Effect (μ_{DEL})	DF (num, den)	F	Pr>F
Strain (*mev-1* vs N2)	1, 34.4	0.26	0.6129

J. Pairwise test of genome-wide deletion rate difference between *mev-1* and N2.

K. Fixed Effect (μ_{INS})	DF (num, den)	F	Pr>F
Strain (*mev-1* vs PB306)	1, 31.2	2.42	0.1295

K. Pairwise test of genome-wide insertion rate difference between *mev-1* and PB306.

L. Fixed Effect (μ_{DEL})	DF (num, den)	F	Pr>F
Strain (*mev-1* vs PB306)	1, 50.1	4.13	0.0474

L. Pairwise test of genome-wide deletion rate difference between *mev-1* and PB306.

M. Fixed Effect (μ_{INS})	DF (num, den)	F	Pr>F
Strain (N2 vs PB306)	1, 122	5.37	0.0221

M. Pairwise test of genome-wide insertion rate difference between N2 and PB306.

N. Fixed Effect (μ_{DEL})	DF (num, den)	F	Pr>F
Strain (N2 vs PB306)	1, 115	11.48	0.0010

N. Pairwise test of genome-wide insertion rate difference between N2 and PB306.

O. Fixed Effect (μ_{BS})	DF (num, den)	F	Pr>F
Strain (all 3)	2, 272	0.89	0.4132
sequence type (non-mono vs. mono)	**1, 243**	**120.92**	**<0.0001**
Strain x seq. type	2, 272	0.15	0.8611
base-sub type (all 6)	5, 140	146.34	<0.0001
Strain x base-sub type	10, 228	1.95	0.0399
seq. type x base-sub type	**5, 140**	**86.49**	**<0.0001**
Strain x seq. type x base-sub type	10, 228	1.16	0.3202

O. Test of effects of sequence type (non-mononucleotide vs. mononucleotide), strain, base-substitution type, and all interactions on the base-substitution rate (μ_{BS}). The important comparisons are the ones highlighted in bold font.

P. Fixed Effect ($\mu_{AT\rightarrow TA}$)	DF (num, den)	F	Pr>F
Strain (all 3)	2, 61.3	4.83	0.0113
sequence type (non-mono vs. mono)	**1, 51.6**	**464.99**	**<0.0001**
Strain x seq. type	2, 61.3	3.99	0.0235

P. Post-hoc test of the effect of sequence type (non-mononucleotide vs. mononucleotide), strain, and their interaction on the rate of AT→TA transversions.

Q. Fixed Effect (μ_{1BP_INDEL})	DF (num, den)	F	Pr>F
Strain (all 3)	2, 105	2.15	0.1217
sequence type (non-mono vs. mono)	**1, 78.3**	**490.15**	**<0.0001**
Strain x seq. type	2, 105	1.94	0.1482
Indel type (+/- 1 bp)	1, 78.3	8.68	0.0042
Strain x indel type	2, 105	0.66	0.5172
seq. type x indel type	1, 78.3	7.02	0.0097
Strain x seq. type x indel type	2, 105	0.60	0.5525

Q. Test of the effect of sequence type (non-mononucleotide vs. mononucleotide, strain, indel type (insertion vs. deletion), and their interactions on the rate of +/- 1 bp indels.

R. Fixed Effect (*5'-ttA-3'*)	DF (num, den)	F	Pr>F
Strain (all 3)	2, 57.6	4.78	0.0120
sequence type (non-mono vs. mono)	**1, 48.5**	**210.49**	**<0.0001**
Strain x seq. type	2, 57.6	4.43	0.0162

R. Test of the effect of sequence type (non-mono vs. mono), strain, and their interaction on the rate of 5'-ttA-3' mutation.

Table S3 is submitted as an excel file with this name:
Supplemental_Table_S3.xlsx

Table S4 is submitted as an excel file with this name:
Supplemental_Table_S4.xlsx

Table S5 is submitted as an excel file with this name:
Supplemental_Table_S5.xlsx

Supplemental Table S6. Correlations between base-substitution (SNP) and indel mutation rates. (a) *mev-1*; (b) N2; (c) PB306. Correlations are reported for each strain separately because the best-fit linear model includes the among-line covariance estimated separately for each strain. See Methods for details.

x

(**a**) ***mev-1***	Del	Ins	SNP
Del		-0.21	0.21
Ins			-0.025

(**b**) **N2**	Del	Ins	SNP
Del		-0.065	0.24
Ins			-0.07

(**c**) **PB306**	Del	Ins	SNP
Del		0.35	0.32
Ins			0.09

Supplemental Table S7. Correlation matrix of the six type-specific base-substitution mutation rates. Data are pooled across strains because the best-fit linear model includes a single (pooled) estimate of the among-line covariance. See Methods for details. Two outliers were removed prior to analysis.

x

Mut type	AT>GC	AT>TA	GC>AT	GC>CG	GC>TA	Row Ave
AT>CG	0.27	0.10	0.19	-0.03	-0.06	0.09
AT>GC		0.26	0.20	0.07	-0.14	0.13
AT>TA			0.17	0.06	-0.03	0.11
GC>AT				0.00	-0.01	0.11
GC>CG					0.09	0.04
GC>TA						-0.03

Table S8 is submitted as an excel file with this name:
Supplemental_Table_S8.xlsx

Table S9 is submitted as an excel file with this name:
Supplemental_Table_S9.xlsx

Table S10 is submitted as an excel file with this name:
Supplemental_Table_S10.xlsx

Table S11 is submitted as an excel file with this name:
Supplemental_Table_S11.xlsx

Table S12. Size distribution of the dummy indels introduced into the reference genome. See section 10 for explanation.

Bin (bp)	Deletions	Insertions	Total Indels
1-5	668	707	1375
6-10	660	680	1340
11-20	1034	1130	2164
21-50	1898	1922	3820
>50bp	639	652	1291