Table S1. The numbers of TRs in the three populations (A) before and after upper thresholding; (B) separated by motif size and repeat number after upper thresholding.

A.

Population	**Number of TRs**	
	Before upper cut-off	**After upper cut-off ***
YRI	453,854,766	453,500,758
CEU	473,350,569	472,996,810
JPTCHB	473,361,805	473,009,688

* In this column, only mono-, di-, tri-, and tetranucleotide TRs under 10-bp, 18-bp, 24-bp, and 28-bp were considered.

B.

TR class	Repeat Number	Number of TRs		
		YRI	CEU	JPTCHB
Mono	2	246493315	246494715	246495135
	3	82682392	82684184	82684723
	4	28102544	28103294	28104008
	5	9466463	9467098	9467380
	6	2580344	2580726	2580806
	7	996553	996750	996774
	8	357577	356944	356799
	9	142710	143814	144061
	10	87945	88612	89031
Di	2	51119753	62970957	62975747
	3	4294322	5291957	5292472
	4	560433	688397	688420
	5	75393	75398	75357
	6	26218	26164	26030
	7	12616	12676	12576
	8	6929	6939	6904
	9	4367	4364	4365
Tri	2	19270525	24136549	24139464
	3	616995	772052	772051
	4	47164	47121	47147
	5	12949	12917	12938
	6	5358	5389	5353
	7	2527	2472	2466
	8	1327	1327	1290
Tetra	2	6322125	7776436	7778707
	3	165519	203244	203470
	4	27514	27461	27448
	5	11249	11284	11236
	6	5057	5007	4998
	7	2575	2562	2532

Table S2. Proportion of TRs in intergenic+intronic vs. exonic+regulatory regions.

	Intergenic and Intronic regions (%)	**Exonic and Regulatory regions (%)**
Genomic coverage	93.2%	6.8%
Coverage by TRs	92.8%	7.2%

The proportions of all (polymorphic and non-polymorphic) microsatellites in putatively neutral regions (intergenic+intronic, 92.8%) vs. those in regions under selective constraints (exonic+regulatory comprising coding exons, UTR exons, and regulatory regions defined as 5 kb upstream and downstream from transcript boundaries, 7.2%) were found to be consistent with the genomic coverage of the two regions (intergenic+intronic: 2,896,745,931 bp (93.2%), exonic+regulatory: 210,598,511 bp (6.8%)).

Table S3. Linear regression model for (A) log polymorphism incidence (using CEU population data), (B) log polEF as a function of repeat number.

A.

TR class	log(Polymorphism incidence) ~ Repeat number		
	Intercept (*P* value)	Slope (*P* value)	R^2
Mono	-3.347 (1.4E-07)	0.4056 (6.7E-07)	97.55%
Di	-2.872 (4.4E-04)	0.5438 (2.3E-04)	91.13%
Tri	-2.723 (8.3E-03)	0.5874 (4.5E-03)	82.77%
Tetra	-3.050 (0.0194)	0.7095 (0.0134)	81.74%

B.

TR class	log(PolEF) ~ Repeat number		
	Intercept (*P* value)	Slope (*P* value)	R^2
Mono	-5.025 (0.00071)	0.3485 (0.00825)	92.88%
Di	-4.608 (3.3E-06)	0.1973 (0.00260)	86.05%

Table S4. Segmented regression model for log polymorphism incidence as a function of repeat number for mono-, di-, tri-, and tetranucleotide TRs in (A) YRI, (B) JPTCHB.

A.

TR class	log(Polymorphism incidence) ~ Repeat number					
	Change-point	Intercept (*P* value)	Slope below change-point	Slope above change-point	*P* value for difference in slope	R^2
Mono	8.373	-3.986 (0.00023)	0.512	0.295	0.0561	99.54%
Di	4.519	-4.231 (3.8E-05)	1.014	0.282	2.2E-16	99.69%
Tri	3.327	-6.204 (0.00021)	1.921	0.297	2.2E-16	99.83%
Tetra	3.321	-5.176 (0.0040)	1.656	0.310	2.2E-16	99.78%

B.

TR class	log(Polymorphism incidence) ~ Repeat number					
	Change-point	Intercept (*P* value)	Slope below change-point	Slope above change-point	*P* value for difference in slope	R^2
Mono	8.308	-4.120 (0.00019)	0.515	0.268	0.0160	99.56%
Di	4.653	-4.490 (2.5E-05)	1.027	0.292	2.2E-16	99.75%
Tri	3.391	-6.488 (0.00012)	1.952	0.293	2.2E-16	99.88%
Tetra	3.356	-5.425 (0.0032)	1.672	0.326	2.2E-16	99.82%

Table S5. (A) Polymorphism Incidence and (B) Theta values for the four TR classes

A.

TR class	Repeat Number	Polymorphism Incidence		
		YRI	CEU	JPTCHB
Mono	2	6.34E-03	5.34E-03	4.82E-03
	3	9.44E-03	8.21E-03	7.35E-03
	4	1.28E-02	1.07E-02	9.76E-03
	5	0.035350423	0.030015425	0.027013951
	6	0.121840212	0.103627062	0.092005076
	7	0.295175867	0.264246267	0.232497578
	8	1.601941086	1.352809271	1.248049496
	9	3.033424427	2.457340732	2.204621653
	10	5.976462562	4.637069471	4.087340365
Di	2	6.00E-03	3.87E-03	3.48E-03
	3	0.070371994	0.046863571	0.042985584
	4	0.639148658	0.423592781	0.394090817
	5	2.356982744	2.080957055	1.924174264
	6	7.243115417	6.413392448	5.758739915
	7	13.10240964	12.14105396	10.98918575
	8	23.56761437	22.19339963	20.32155272
	9	33.40966338	32.14940422	29.43871707
Tri	2	4.35E-03	2.93E-03	2.61E-03
	3	0.362563716	0.251148886	0.23392237
	4	2.149944873	1.878143503	1.832566229
	5	4.849795351	4.157312069	3.988251662
	6	11.83277342	10.68843941	9.770222305
	7	20.8547685	17.43527508	15.85563666
	8	31.80105501	29.5403165	26.82170543
Tetra	2	1.37E-02	2.69E-03	8.29E-03
	3	0.621076734	0.42559682	0.389246572
	4	2.940321291	2.47623903	2.142232585
	5	8.285180905	7.328961361	6.247775009
	6	16.13604904	14.04034352	12.36494598
	7	25.32038835	24.59016393	20.85308057

B.

TR class	Repeat Number	Theta		
		YRI	CEU	JPTCHB
Mono	2	4.36E-05	3.97E-05	3.60E-05
	3	6.65E-05	6.12E-05	5.38E-05
	4	8.15E-05	7.58E-05	6.90E-05
	5	0.00012055	0.000110074	0.000100938
	6	0.000254181	0.000233861	0.000208288
	7	0.00051213	0.000487145	0.000424778
	8	0.00754285	0.006346304	0.006117739
	9	0.021942723	0.018798028	0.016817317
	10	0.036968009	0.033156261	0.029456925
Di	2	3.68E-05	2.70E-05	2.53E-05
	3	0.000470231	0.000337183	0.000308566
	4	0.004371223	0.003131246	0.002841724
	5	0.01679632	0.015591007	0.014215596
	6	0.05042812	0.04691796	0.043438378
	7	0.089110952	0.081329464	0.074679148
	8	0.161969691	0.15080808	0.138290547
	9	0.244308932	0.231465493	0.212395294
Tri	2	2.66E-05	2.06E-05	1.85E-05
	3	0.00272718	0.002005706	0.001851005
	4	0.016372411	0.015149574	0.013980667
	5	0.035023233	0.030721388	0.030328081
	6	0.088289466	0.086815074	0.076652748
	7	0.158221961	0.147703061	0.128793494
	8	0.269163378	0.249760058	0.233731846
Tetra	2	8.94E-05	7.20E-05	5.97E-05
	3	0.004107482	0.003237235	0.002892039
	4	0.021232175	0.018739421	0.01628053
	5	0.064202797	0.056401373	0.050804832
	6	0.120160074	0.102670138	0.09661464
	7	0.224981115	0.210602773	0.183994594

Table S6. Segmented regression model for log theta as a function of repeat number for mono-, di-, tri-, and tetranucleotide TRs in the CEU population.

TR class	log(theta) ~ Repeat number					
	Change point	**Intercept (*P* value)**	**Slope below change point**	**Slope above change point**	***P* value for difference in slope**	**R^2**
Mono	8.761	-8.0644 (0.126)	0.7168	0.2465	0.4348	97.12%
Di	4.658	-6.6106 (3.9E-06)	1.032	0.285	2.2E-16	99.78%
Tri	3.356	-8.664 (6.9E-05)	1.989	0.312	2.2E-16	99.86%
Tetra	3.363	-7.448 (0.0012)	1.653	0.341	2.2E-16	99.87%

Table S7. Primary Pol EF data for (A) mono-, and (B) dinucleotide TRs
A.

2 unit monos	**Target Location**	**TR mutants**	**Pol EF**
AA	127-128	4	5.96E-03
	161-162	1	
	164-165	0	
TT	94-95	2	
	111-112	62	
	142-143	16	
CC	99-100	3	
	120-121	5	
	129-130	1	
	153-154	12	
	167-168	63	
GG	102-103	0	
	118-119	1	
	138-139	3	
	155-156	7	
2 unit monos	**Target Location**	**TR mutants**	**Pol EF**
TT	127-128	8	5.17E-03
	161-162	9	
	164-165	55	
AA	94-95	0	
	111-112	2	
	142-143	1	
GG	99-100	0	
	120-121	6	
	129-130	0	
	153-154	0	
	167-168	0	
CC	102-103	60	
	118-119	11	
	138-139	18	
	155-156	21	
total all mutants			1134
total TR mutants		371	
avg Pol EF			5.57E-03
TR EF			1.82E-03
TR EF per site			6.07E-05
3 unit monos	**Target Location**	**TR mutants**	**Pol EF**
CCC	147-149	18	7.98E-03
	180-182	6	
	196-198	1	
	269-271	0	
GGG	209-211	1	
	232-234	1	
	273-275	0	
TTT	212-214	2	
3 unit monos	**Target Location**	**TR mutants**	**Pol EF**
GGG	147-149	0	1.34E-02
	180-182	1	
	196-198	0	
	269-271	0	
CCC	209-211	0	
	232-234	1	
	273-275	3	
AAA	212-214	0	
avg Pol EF			1.07E-02
total all mutants			414
total STR mutants			34
TR Pol EF			8.76E-04
TR Pol EF per site			5.47E-05
4 unit monos	**Target Location**	**TR mutants**	**Pol EF**
AAAA	190-194	2	7.98E-03
	241-244	6	
CCCC	226-229	2	
GGGG	237-240	9	
4 unit monos	**Target Location**	**TR mutants**	**Pol EF**
TTTT	190-194	6	1.34E-02
	241-244	12	
GGGG	226-229	0	
CCCC	237-240	1	
avg EF			1.07E-02
total TR mutants			38
total all mutants			414
TR Pol EF			9.79E-04
TR Pol EF per site			1.22E-04
8 unit monos	**Target Location**	**TR mutants**	**Pol EF**
T8	TR site	26	3.10E-02
A8	TR site	33	4.72E-02
avg pol EF			3.91E-02
total TR mutants			59
total all mutants			62
TR Pol EF			3.72E-02
TR Pol EF per site			1.86E-02
10 unit monos	**Target Location**	**N mutants**	**Pol EF**
G10	TR site	57	4.60E-02
C10	TR site	22	2.47E-02
avg EF			3.54E-02
total TR mutants			79
total all mutants			95
TR Pol EF			2.94E-02
TR Pol EF per site			1.47E-02

B.

2 unit dinucleotides	**Pol EF**	**TR mutants**	**Total mutants**
GCGC	2.90E-03	1	61
CGCG	4.10E-03	1	
avg EF	3.50E-03		
sum		2	61
STR EF			1.15E-04
per site			5.74E-05
3 unit dinucleotides	**Pol EF**	**TR mutants**	**Total mutants**
TA	7.99E-03	2	19
TA	3.10E-03	6	11
TC	3.00E-03	0	18
TC	1.80E-03	0	17
AG	2.40E-03	0	12
AG	2.80E-03	0	12
avg EF	3.52E-03		
sum		8	89
avg STR EF			3.16E-04
avg STR EF per site			1.05E-04
4 unit dinucleotides	**Pol EF**	**TR mutants**	**Total mutants**
TC	3.70E-03	5	20
TC	2.90E-03	3	22
AG	2.20E-03	2	13
AG	3.30E-03	1	15
GT	2.90E-03	0	40
GT	4.10E-03	2	21
CA	1.60E-03	1	11
CA	1.80E-03	1	13
TA	1.20E-02	14	19
TA	1.83E-03	8	11
avg EF	3.63E-03		
TOTAL		37	185
avg STR EF			7.27E-04
avg STR EF per site			1.45E-04
5 unit dinucleotides	**Pol EF**	**TR mutants**	**Total mutants**
TA	1.20E-02	6	11
TA	4.68E-03	18	22
TC	5.40E-03	2	23
TC	3.80E-03	5	20
AG	4.00E-03	2	16
AG	5.00E-03	2	18
GT	2.10E-03	0	20
GT	2.80E-03	1	20
CA	3.50E-04	5	10
CA	2.00E-03	0	14
avg EF	4.21E-03		
sum		41	174
avg STR EF			9.93E-04
avg STR EF per site			1.99E-04
6 unit dinucleotides		**TR mutants**	**Total mutants**
TA	1.99E-02	10	12
TA	9.69E-03	16	18
GT	4.10E-03	6	21
GT	4.10E-03	7	21
CA	5.10E-03	5	24
CA	5.20E-03	4	20
avg EF	8.02E-03		
sum		48	116
avg STR EF			3.32E-03
avg STR EF per site			1.11E-03
8 unit dinucleotides	**Pol EF**	**TR mutants**	**Total mutants**
GT	2.60E-03	8	27
GT	3.30E-03	1	21
CA	3.30E-03	9	14
CA	4.70E-03	5	13
TC	1.20E-02	3	11
TC	8.30E-03	2	17
AG	2.00E-03	15	16
AG	3.20E-03	16	19
TA	2.12E-02	26	31
TA	7.30E-03	35	40
avg EF	6.79E-03		
sum		120	209
avg STR EF			3.90E-03
avg STR EF per site			7.80E-04
9 unit dinucleotides	**Pol EF**	**TR mutants**	**Total mutants**
GT	5.40E-03	10	29
GT	5.90E-03	11	19
CA	3.30E-03	15	19
CA	3.60E-03	12	20
avg EF	4.55E-03		
sum		48	87
avg STR EF			2.51E-03
avg STR EF per site			1.26E-03

Table S8. Linear regression models for polymorphism incidence as a function of recombination rate for mono-, di-, tri-, and tetranucleotide TRs.

TR class	**Polymorphism incidence ~ Recombination rate**		
	Intercept (*P* value)	**Slope (*P* value)**	**R^2**
Mono	-0.0139 (0.517)	0.3891 (<2E-16)	4.62%
Di	-0.0058 (0.79)	0.1626 (4.0E-05)	0.81%
Tri	8E-04 (<2E-16)	7.2E-05 (2.3E-08)	1.48%
Tetra	0.0015 (<2E-16)	1.4E-04 (7.7E-07)	1.17%

Table S9. Two sample mean t-test for recombination rate of polymorphic and non-polymorphic mono-, di-, tri-, and tetranucleotide TRs

TR class	Mean recombination rate		*P* value for difference in mean recombination rate between P and NP > 0
	Polymorphic (P)	Non-polymorphic (NP)	
Mono	0.977778	0.959090	2.2E-16
Di	0.983590	0.968106	1.409E-07
Tri	0.990975	0.966578	8.358E-09
Tetra	0.999976	0.968900	2.076E-08

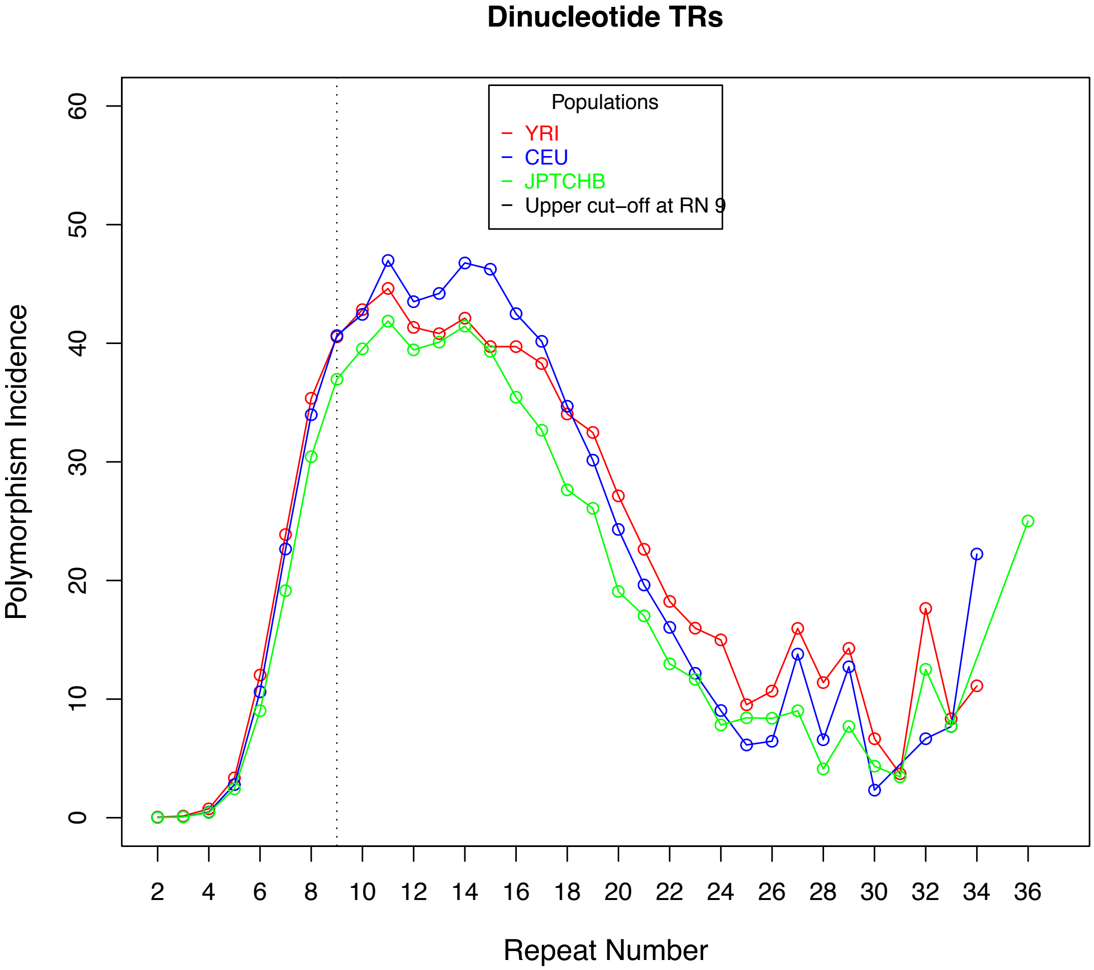
Table S10. Segmented regression model for log polymorphism incidence as a function of repeat number for motif-composition/structure based sub-classes of mono-, di-, tri-, and tetranucleotide TRs in the CEU population.

TR class	**Motif-composition/structure sub-class**	**log(Polymorphism incidence) ~ Repeat number**					
		Change point	**Intercept (*P* value)**	**Slope below change point**	**Slope above change point**	***P* value for difference in slope**	**R^2**
Mono	A	n.s.	-3.362 (0.292)	0.317	n.s.	0.792 (n.s.)	96.53%
	G	n.s.	-5.740 (0.015)	0.7562	n.s.	0.1351 (n.s.)	98.46%
Di	AC	4.532	-4.414 (1.3E-05)	1.003	0.3187	2.2E-16	99.82%
	AG	4.483	-4.6681 (1.9E-06)	1.093	0.2851	2.2E-16	99.76%
	AT	5.491	-4.0730 (1.2E-06)	0.9166	0.2005	2.2E-16	99.92%
	GC	NA					
Tri	No secondary structure	3.467	-5.839 (9.8E-05)	1.736	0.289	2.2E-16	99.91%
	Triplex	3.423	-6.359 (8.7E-05)	1.915	0.406	2.2E-16	99.94%
	Hairpin	3.288	-6.729 (0.0472)	2.051	0.319	2.2E-16	99.82%
Tetra	No secondary structure	3.390	-5.050 (0.00528)	1.654	0.295	2.2E-16	99.69%
	Triplex	3.481	-4.507 (0.00304)	1.385	0.346	1.2E-14	99.86%
	Hairpin	3.505	-6.188 (0.00176)	1.821	0.341	2.2E-16	99.90%

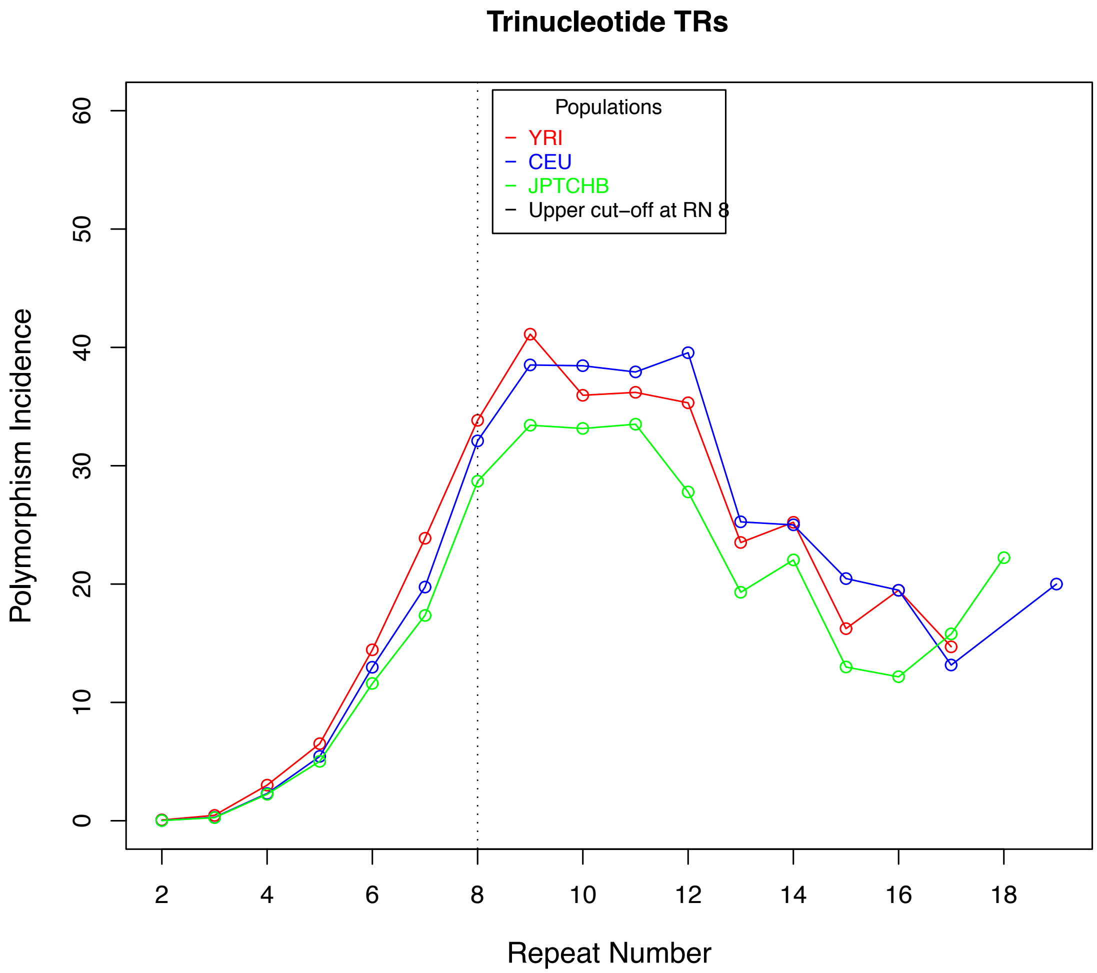
NA = data not available at all repeat numbers
n.s.: no statistical significance

Figure S1. Proportion of polymorphic TRs by Tract length with proposed upper cut-offs

A. Dinucleotide TRs



B. Trinucleotide TRs



C. Tetranucleotide TRs