Supplementary Text 1. Suffix and Longest Common Prefix Arrays

A suffix array is an array of character positions representing a list of all possible suffixes of a string, ordered lexicographically. Consider the sequence "CAGAGA$". A proper suffix array implementation would not enumerate a list of suffixes, but viewing the list helps conceptualize suffix array construction (see Supplementary Fig. 1A and B). The suffix and longest common prefix arrays (with zero-based indexing) for this sequence are shown in Supplementary Fig. 1C. The 6 in position 0 of the suffix array (Supplementary Fig. 1B and C) informs us that the suffix beginning at position 6 (i.e., "$") is lexicographically first. The 5 in position 1 of the suffix array informs us that the suffix beginning at position 5 (i.e., "A$") is lexicographically second. Likewise, the 2 in position 6 of the suffix array informs us that the suffix beginning at position 2 (i.e., "GAGA$") is lexicographically last.

Longest common prefix arrays are arrays of the lengths of the longest common prefix of each adjacent suffix in the suffix array. To illustrate, consider position 3 in the suffix and longest common prefix arrays in Supplementary Fig. 1C. The longest common prefix at this position is 3 (highlighted in red text in Supplementary Fig. 1C), meaning there are three common nucleotides at the beginning of the suffixes starting at positions 1 and 3 (i.e., "AGA"). The longest common prefix array stores the length of the longest common prefix, and the positions of the two suffixes in the original sequence are obtained by looking at the same position in the suffix array (in this example position 3), and the prior position in the suffix array (in this example position 2). This longest common prefix is represented in red nucleotides in Supplementary Fig. 1B. Although the sequence is the same, they are adjacent in the original sequence. These relationships are the basis for our algorithm to find SSRs in a sequence. The longest common prefix array is constructed while creating the suffix array.

Supplementary Text 2. Calculating SSR Length and Position from Suffix and Longest Common Prefix Arrays

Let k equal the length of an SSR repeating unit or period size, r equal the number of times it repeats after the original occurrence, and p equal the position of the first nucleotide of the first period of the SSR. For example, consider the repeating unit "ACG" in the sequence "ACGACGACG". The length of the repeating unit is 3 (k), there are three instances of the unit ($r + 1$), and the SSR begins at position 0 in the sequence (p). So in this example, $k = 3$, $r = 2$ ($r + 1$ is the total number of repeats in the SSR), and $p = 0$. SSRs are identified by calculating k, p, and r from the suffix and longest common prefix arrays. Let i equal the index of any entry in the suffix array (except the first position), where SA and LCPA are the suffix and longest common prefix arrays, respectively:

$$k = |SA_i - SA_{i-1}| \qquad (1)$$

$$r = \left\lfloor \frac{LCPA_i}{k_i} \right\rfloor \qquad (2)$$

$$p = MIN(SA_{i-1}, SA_i) \qquad (3)$$

If $r > 0$, an SSR of length $k * (r + 1)$ exists at position p in the original sequence, otherwise if $r = 0$ there is no SSR at position p. The base unit (e.g. AG in the SSR AGAGAG) of the SSR starts at position p and ends at position $p + (k - 1)$. Thus, by comparing each adjacent element in the suffix array we can find SSRs in a sequence.

Extending the previous example, Fig. 1C shows the values of k, r, and p calculated from the suffix and longest common prefix arrays for "CAGAGA$". Two SSRs, each of length 4, exist at positions 1 and 2 in the original sequence (i.e., "AGAG" and "GAGA") and their locations are shown in Fig. 1D.

Supplementary Figure 1. Suffix and Longest Common Prefix Arrays Example

A

Before Sorting							
0	C	A	G	A	G	A	$
1	A	G	A	G	A	$	
2	G	A	G	A	$		
3	A	G	A	$			
4	G	A	$				
5	A	$					
6	$						

B

After Sorting							
6	$						
5	A	$					
3	A	G	A	$			
1	A	G	A	G	A	$	
0	C	A	G	A	G	A	$
4	G	A	$				
2	G	A	G	A	$		

C

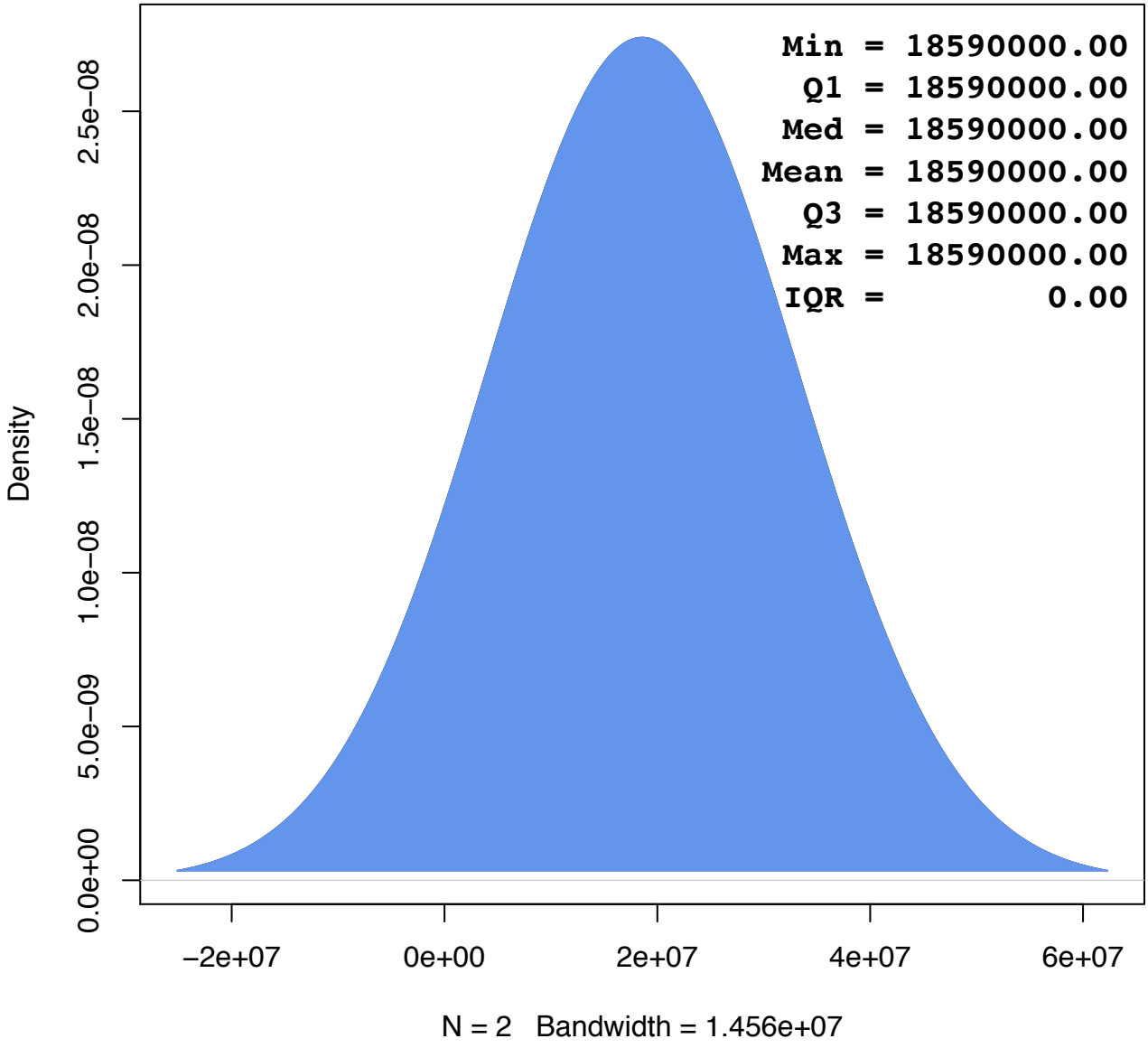
Array Indices:	0	1	2	3	4	5	6
Suffix Array:	6	5	3	1	0	4	2
Longest Common Prefix Array:	-	0	1	3	0	0	2
k:	-	1	2	2	1	4	2
r:	-	0	0	1	0	0	1
p:	-	5	3	1	0	0	2

D

Position:	0	1	2	3	4	5	6
SSR 1:	C	A	G	A	G	A	$
SSR 2:	C	A	G	A	G	A	$

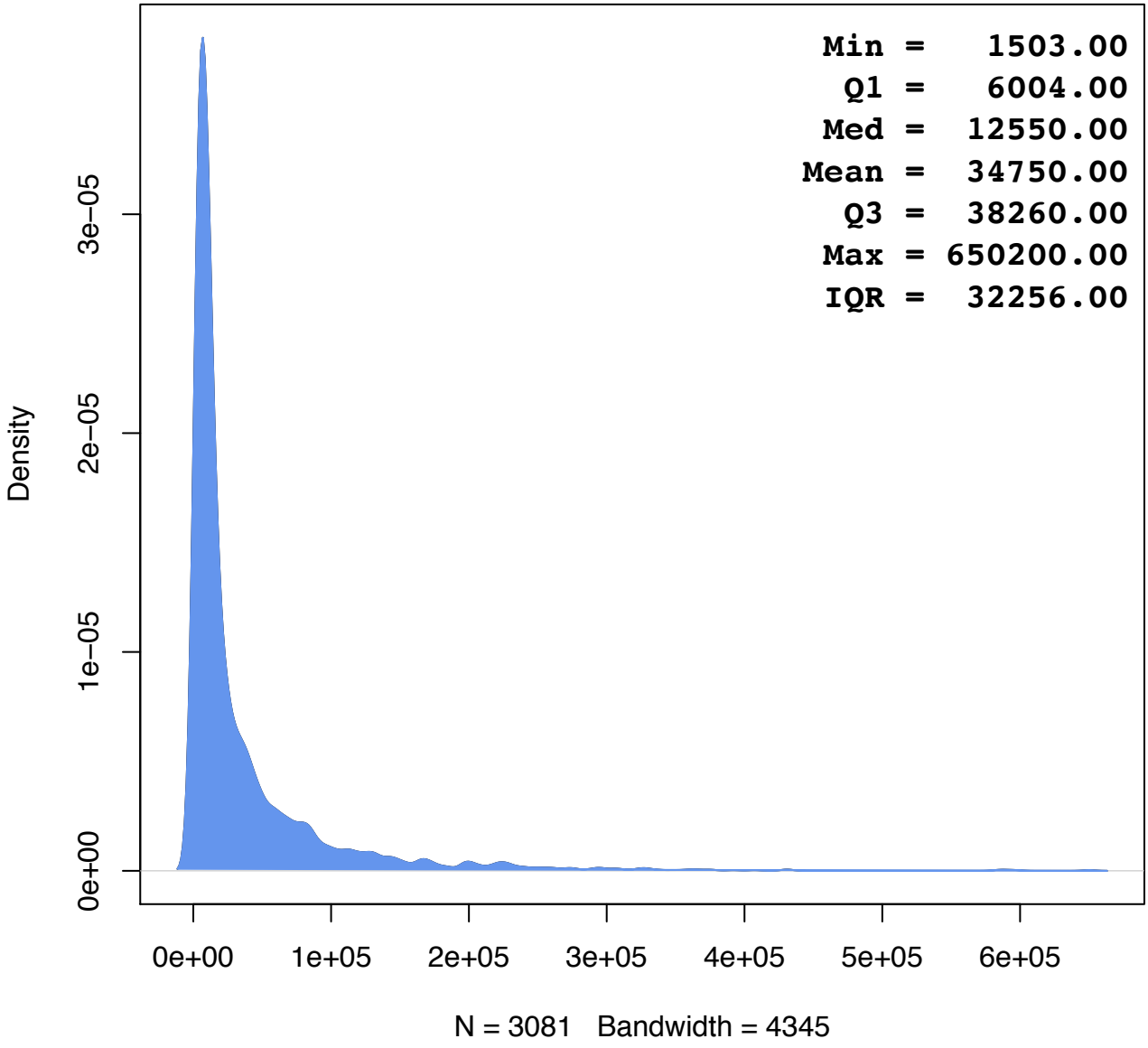
In this figure we demonstrate how to construct a suffix array and its use to identify SSRs. (A) First, all suffixes of "CAGAGA$", are shown here and marked by their beginning position in the original sequence. (B) Next, the set of possible suffixes (part A) are ordered lexicographically, where '$' is the first character in the alphabet, and maintain their start positions in the original sequence. The start positions are the numbers to the left of each suffix. The new ordering of these start positions is the suffix array. (C) Here we show the suffix array, longest common prefix array, and three parameters: k, r, and p (explained in the text). The suffix array stores the ordered start positions determined by ordering possible suffixes (shown in part B). (D) This particular sequence has two SSRs: "AGAG" and "GAGA". In part D we show each of the two SSRs in the original sequence. SSR1 is highlighted blue, and SSR2 is highlighted green. The repeating units of the two SSRs are AG and GA, respectively, and a vertical bar separates each repeating unit in the sequence.

Supplementary Figure 2. *Arabidopsis thaliana* Sequence Length Density Plot



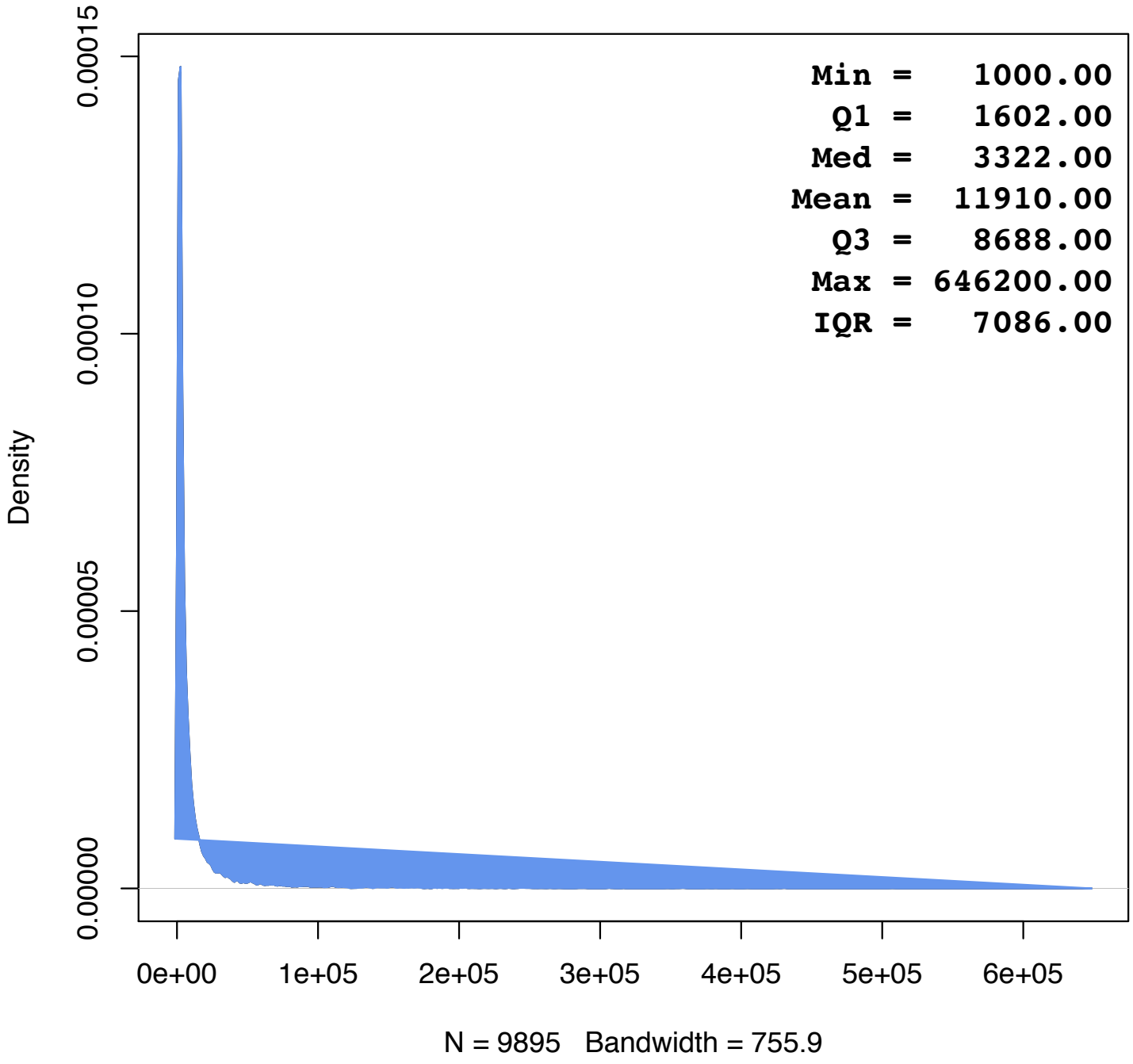
Density plot showing the distribution of sequence lengths for the *Arabidopsis thaliana* chromosome 4. A summary is included in the upper, right-hand corner.

Supplementary Figure 3. *Caenorhabditis elegans* Sequence Lengths Density Plot



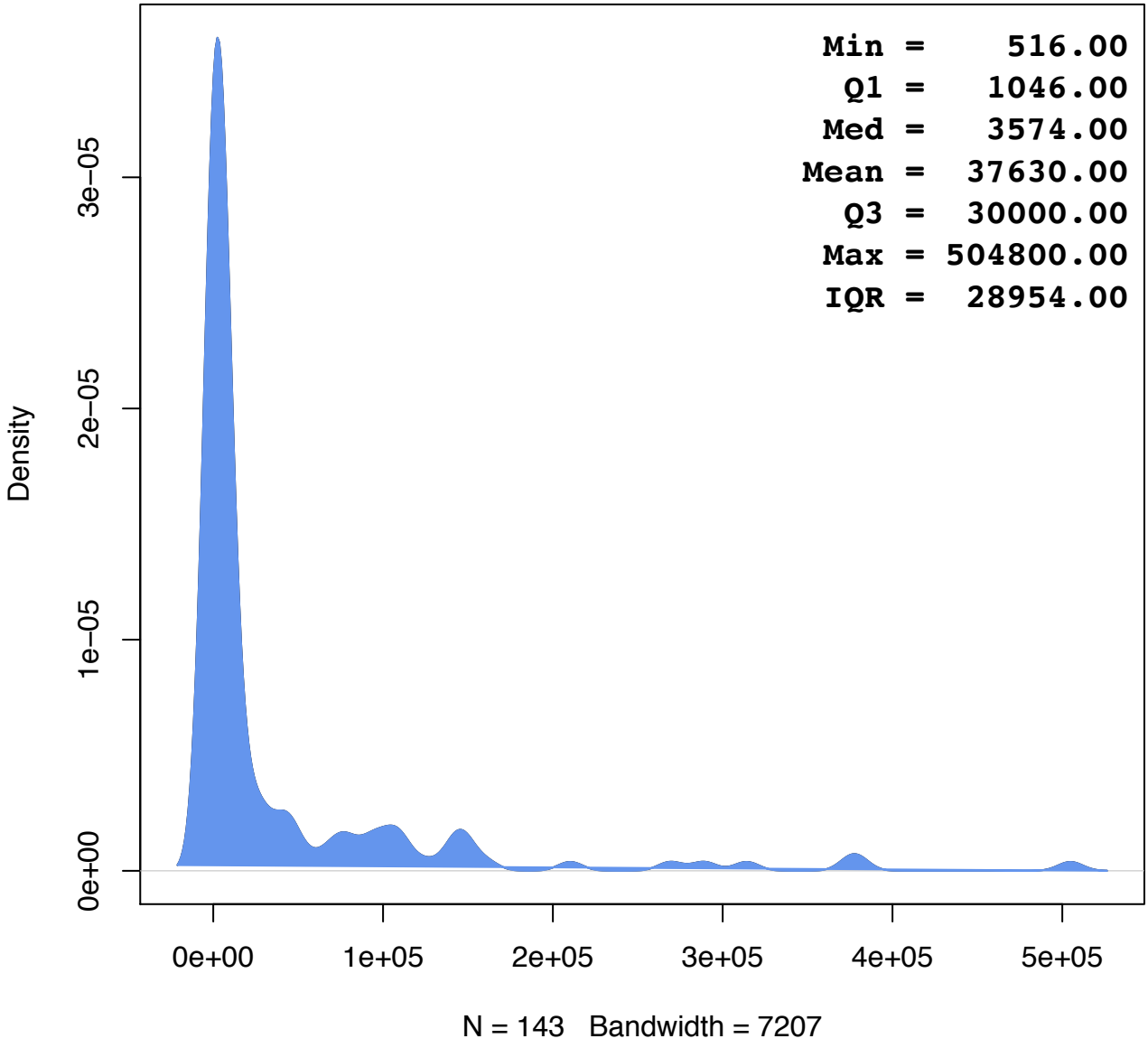
Density plot showing the distribution of sequence lengths for the *Caenorhabditis elegans* genome. A summary is included in the upper, right-hand corner.

Supplementary Figure 4. *Drosophila melanogaster* Sequence Lengths Density Plot



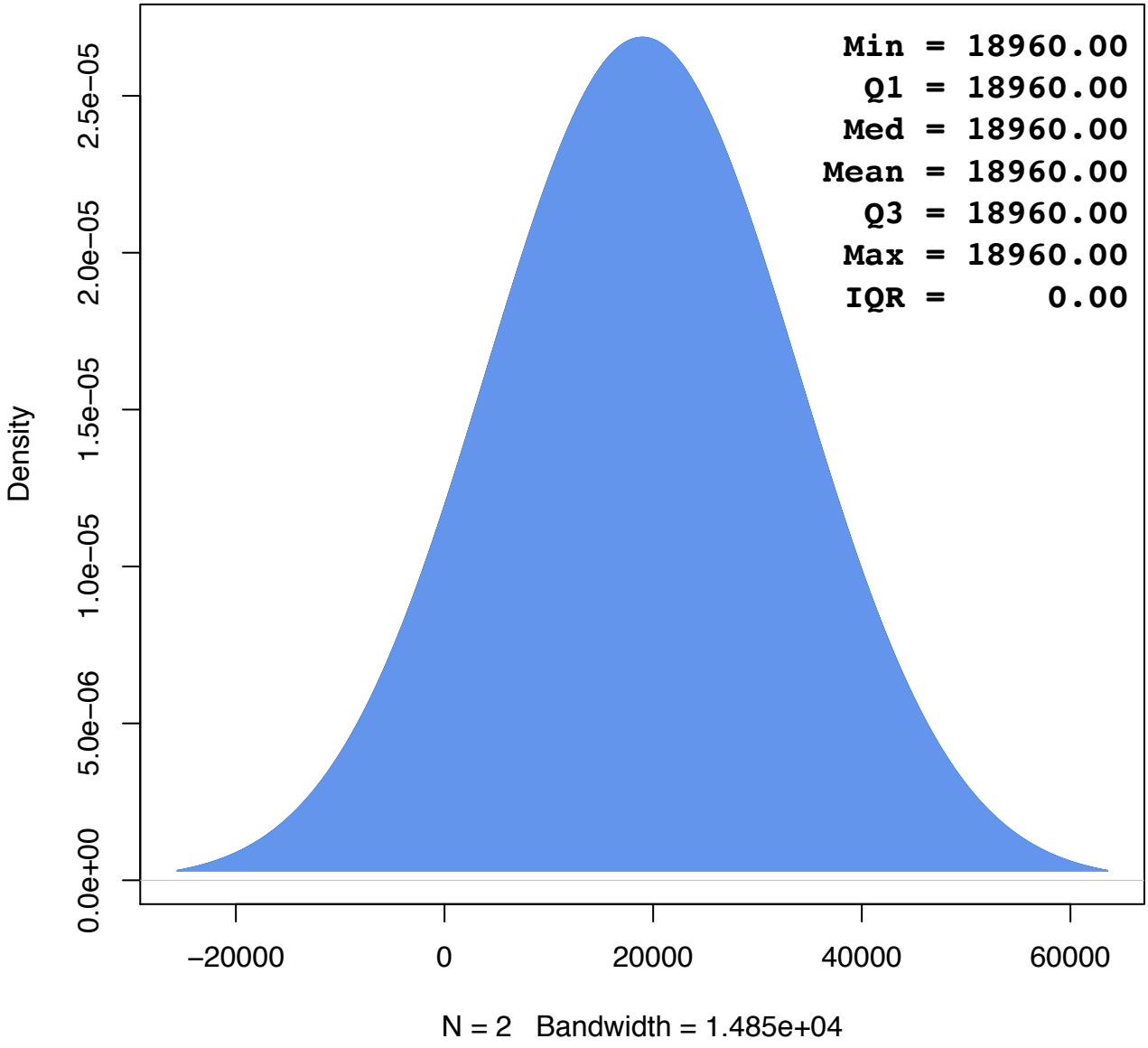
Density plot showing the distribution of sequence lengths for the *Drosophila melanogaster* genome. A summary is included in the upper, right-hand corner.

Supplementary Figure 5. *Escherichia coli* Sequence Lengths Density Plot



Density plot showing the distribution of sequence lengths for the *Escherichia coli* genome. A summary is included in the upper, right-hand corner.

Supplementary Figure 6. *Zaire ebolavirus* Sequence Length Density Plot



Density plot showing the distribution of sequence lengths for the *Zaire ebolavirus* genome. A summary is included in the upper, right-hand corner.

Supplementary Table 1. Algorithms Included in Comparisons

Algorithm
GMATo (Wang, et al., 2013)
MREPS (Kolpakov, et al., 2003)
PRoGeRF (Lopes, et al., 2015)
QDD (Meglécz, et al., 2014)
SSR-Pipeline (Miller, et al., 2013)
SSRIT (Temnykh, et al., 2001)
TRF (Benson, 1999)

We compared our algorithm to existing algorithms that (a) were capable of processing the *Drosophila melanogaster genome* dataset (see the main text), (b) had a non-interactive, Linux, command-line interface, (c) were freely available for immediate download, and (d) had 10 or more citations per year (based on publication date and Google Scholar citation count) or were published in the last three years. A few other algorithms met our requirements, but were rendered unusable due to antiquated shared libraries, compile- or run-time errors, or other issues.

Supplementary Table 2. Performance Comparison

								Comparison with SA-SSR		
		CPU Time[a] (mm:ss)	Real Time[a] (mm:ss)	SSRs Reported	SSRs In Range[b]	Number Correct[c]	Percent Correct	SSRs Unique to Software[d]	SSRs Unique to SA-SSR	SSRs Shared
Arabidopsis thaliana (chr 4)	GMATo	312:29	312:29	27,511,385	8,667	0	0	0	2,265	0
	MREPS	386:15	386:15	4,201	1,608	1,608	100	2	668	1,597
	PRoGeRF	9:23	9:23	4,116,484	1,599	1,599	100	2	698	1,567
	QDD	2:02	2:02	3,965	1,100	1,100	100	0	1,167	1,098
	SA-SSR	28,066:12	2,338:47	2,265	2,265	2,265	100	NA	NA	NA
	SSR-Pipeline	1,395:04	1,395:04	4,754,929	1,580	1,580	100	2	769	1,496
	SSRIT	0:10	0:10	900	900	900	100	0	1,372	893
	TRF	0:47	0:47	135,135	7,505	1,527	20.35	1	862	1,403
Caenorhabditis elegans	GMATo	9:39	9:39	22,889,822	5,551	5,551	100	20	4,663	3,260
	MREPS	4:34	4:34	18,958	7,440	7,440	100	26	567	7,356
	PRoGeRF	744:21	744:21	531,822	99	99	100	0	7,826	97
	QDD	10:32	10:32	11,720	3,379	3,379	100	6	4,560	3,363
	SA-SSR	645:54	60:31	7,923	7,923	7,923	100	NA	NA	NA
	SSR-Pipeline	13:14	13:14	26,475,821	7,305	7,305	100	24	904	7,019
	SSRIT	0:57	0:57	2,374	2,374	2,374	100	3	5,566	2,357
	TRF	7:20	7:20	1,029,051	31,500	6,174	19.60	6	2,649	5,274
Drosophila melanogaster	GMATo	6:31	6:31	21,180,679	1,053	1,053	100	0	27,294	586
	MREPS	1:47	1:47	52,347	28,009	28,009	100	43	104	27,776
	PRoGeRF	2,436:55	2,436:55	470,382	576	566	98.26	0	27,324	556
	QDD	11:11	11:11	37,525	12,931	12,931	100	4	15,012	12,868
	SA-SSR	52:58	4:52	27,880	27,880	27,880	100	NA	NA	NA
	SSR-Pipeline	1:47	1:47	29,015,430	27,513	27,513	100	42	1,354	26,526
	SSRIT	1:02	1:02	9,943	9,943	9,943	100	2	17,993	9,887
	TRF	4:01	4:01	856,363	108,070	26,156	24.20	5	3,911	23,969
Escherichia coli	GMATo	0:39	0:39	1,127,792	13	13	100	0	15	5
	MREPS	0:26	0:26	46	19	19	100	0	1	19
	PRoGeRF	3:36	3:36	334,091	4	4	100	0	16	4
	QDD	0:32	0:32	38	8	8	100	0	20	0
	SA-SSR	55:07	12:21	20	20	20	100	NA	NA	NA
	SSR-Pipeline	1:15	1:15	93,025	0	0	NA	0	20	0
	SSRIT	0:03	0:03	0	0	0	NA	0	20	0
	TRF	0:06	0:06	15,107	209	19	9.09	0	1	19
Zaire ebolavirus	GMATo	0:00	0:00	4,180	0	0	NA	NA	NA	NA
	MREPS	0:00	0:00	0	0	0	NA	NA	NA	NA
	PRoGeRF	0:03	0:03	4,350	0	0	NA	NA	NA	NA
	QDD	0:00	0:00	0	0	0	NA	NA	NA	NA
	SA-SSR	0:01	0:01	0	0	0	NA	NA	NA	NA
	SSR-Pipeline	0:01	0:01	4,862	0	0	NA	NA	NA	NA
	SSRIT	0:00	0:00	0	0	0	NA	NA	NA	NA
	TRF	0:00	0:00	59	0	0	NA	NA	NA	NA
Combined	GMATo	329:18	329:18	72,713,858	15,284	6,617	43.29	20	34,237	3,851
	MREPS	393:02	393:02	75,552	37,076	37,076	100	71	1,340	36,748
	PRoGeRF	3,194:18	3,194:18	5,457,129	2,278	2,268	99.56	2	35,864	2,224
	QDD	24:17	24:17	53,248	17,418	17,418	100	10	20,759	17,329
	SA-SSR	28,820:12	2,416:32	38,088	38,088	38,088	100	NA	NA	NA
	SSR-Pipeline	1,411:21	1,411:21	60,344,067	36,398	36,398	100	68	3,047	35,041
	SSRIT	2:12	2:12	13,217	13,217	13,217	100	5	24,951	13,137
	TRF	12:14	12:14	2,035,715	147,284	33,876	23.00	12	7,423	30,665

[a] MREPS timing includes the pre- and post-processing time for each genome necessary to adjust positions to account for removing "incorrect symbols" and Ns. The additional times are an average of multiple approaches.

[b] We only considered SSRs with period sizes 1-7 (inclusive) and lengths of at least 16 nucleotides (nt). The difference between the number of SSRs in range and reported is due exclusively to SSR length (less than 16 nt) and period size (greater than 7).

[c] Whenever possible, we salvaged correct SSRs that were inside incorrect SSRs reported by other software packages. For example, in *Drosophila melanogaster*, we recovered three for PRoGeRF and 8,408 for TRF. To illustrate, in sequence JXOZ01000043.1, TRF reports a CT repeated 36 times at position 2,171. While TRF does correctly identify a low-complexity region with many CT repeats, there are not 36 perfect repeats in a row. In this case, we salvaged two perfect CT regions, each repeating 8 times.

[d] Detailed pairwise comparisons can be found in Supplementary Tables 4-31.

Supplementary Table 3. Features of Software for Finding SSRs

	Op. Sys.					Format					Complexity						
	MS Win	Mac OS X	Linux	CLI	GUI	Input	Output	Language	Algorithm	Type	Time	Space	Period	Repeats	Multi-threaded	Ignore Characters	Search for Specific SSRs
SA-SSR			X	X		FASTA	TSV	C++	Combinatorial	Exact	O(n)	O(n)	1+	2+	X	Yes (Configurable)	X
GMATo	X	X	X	X	X	FASTA	TSV	Perl & Java	Regular Expressions	Exact	?	?	1-10	2+		Yes (default)	
MREPS			X	X		FASTA	Text	C	Combinatorial	Inexact	O(nk · log(n/k) + S)	?	1+	2+		Yes (only some Ns)	
PRoGeRF			X	X	Web	FASTA	TSV	Perl	?	Inexact	?	?	1-12	2+		Yes (default)	
QDD	X		X	X		FASTA	SCSV	Perl	?	Exact	?	?	?	2+		Yes (default)	
SSR-Pipeline	X	X	X	X		FASTA	FASTA	Python	?	Exact	?	?	2-25	2+		Yes (default)	
SSRIT			X	X		FASTA	TSV	Perl	Regular Expressions	Exact	?	?	?	2+		Yes (default)	
TRF	X	X	X	X	X	FASTA	Text	?	Heuristic	Inexcat	$O(n^2 \cdot \text{polylog}(n))$	?	1+	2+		Yes (default)	

Supplementary Table 4. SA-SSR compared with GMATo for *Arabidopsis thaliana*

	1	2	3	4	5	6	7	Total
GMATo	0	0	0	0	0	0	0	**0**
SA-SSR	660	721	343	126	60	245	110	**2265**
Shared	0	0	0	0	0	0	0	**0**

The number of SSRs in the *Arabidopsis thaliana* chromosome 4 found unique to GMATo, unique to SA-SSR, and shared between the two using the following parameter set: -l 1 -L 18600000 -m 1 -M 7 -n 16 -r 1 –i D,M,N. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Supplementary Table 5. SA-SSR compared with MREPS for *Arabidopsis thaliana*

		1	2	3	4	5	6	7	Total
Normal	MREPS	0	5	1	2	1	2	0	**11**
	SA-SSR	660	5	1	1	0	1	0	**668**
	Shared	0	716	342	125	60	244	110	**1597**
Overlap	MREPS	0	0	0	0	1	1	0	**2**
	SA-SSR	2742	6064	2171	322	134	553	535	**12521**
	Shared	0	721	343	127	60	245	110	**1606**
Exhaustive	MREPS	0	0	0	0	0	0	0	**0**
	SA-SSR	2752	8824	3761	9867	1029	10115	1194	**37542**
	Shared	0	721	343	127	61	246	110	**1608**

The number of SSRs in the *Arabidopsis thaliana* chromosome 4 found unique to MREPS, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 18600000 -m 1 -M 7 -n 16 -r 1 –i D,M,N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 11 SSRs that MREPS found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 9 of the 11 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 2 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 6. SA-SSR compared with ProGeRF for *Arabidopsis thaliana*

		1	2	3	4	5	6	7	Total
Normal	ProGeRF	0	5	6	3	2	16	0	**32**
	SA-SSR	660	7	21	5	1	4	0	**698**
	Shared	0	714	322	121	59	241	110	**1567**
Overlap	ProGeRF	0	0	0	0	1	15	0	**16**
	SA-SSR	2742	6066	2186	325	134	556	535	**12544**
	Shared	0	719	328	124	60	242	110	**1583**
Exhaustive	ProGeRF	0	0	0	0	0	0	0	**0**
	SA-SSR	2752	8826	3776	9870	1029	10104	1194	**37551**
	Shared	0	719	328	124	61	257	110	**1599**

The number of SSRs in the *Arabidopsis thaliana* chromosome 4 found unique to ProGeRF, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 18600000 -m 1 -M 7 -n 16 -r 1 –i D,M,N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 32 SSRs that ProGeRF found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 16 of the 32 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. 14 of the remaining 16 SSRs were also found by SA-SSR, but SA-SSR correctly reported shorter period lengths than ProGeRF did. Obviously, reporting a longer period length than is strictly necessary to describe the SSR is misleading and certainly incorrect. AAAAAAAAA has a period size of one repeated nine times, not three repeated three times. Likewise, ATATATAT has a period size of two repeated four times, not four repeated two times. The last 2 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 7. SA-SSR compared with QDD for *Arabidopsis thaliana*

		1	2	3	4	5	6	7	Total
Normal	QDD	0	2	0	0	0	0	0	**2**
	SA-SSR	660	2	1	99	55	240	110	**1167**
	Shared	0	719	342	27	5	5	0	**1098**
Overlap	QDD	0	0	0	0	0	0	0	**0**
	SA-SSR	2742	6064	2172	422	189	793	645	**13027**
	Shared	0	721	342	27	5	5	0	**1100**

The number of SSRs in the *Arabidopsis thaliana* chromosome 4 found unique to QDD, unique to SA-SSR, and shared between the two using two different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 18600000 -m 1 -M 7 -n 16 -r 1 –i D,M,N. The overlap set was identical to normal with the following addition: -o. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 2 SSRs that QDD found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. Both were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result.

Supplementary Table 8. SA-SSR compared with SSR-Pipeline for *Arabidopsis thaliana*

		1	2	3	4	5	6	7	Total
Normal	SSR-Pipeline	0	47	16	7	1	7	6	**84**
	SA-SSR	660	59	26	9	0	8	7	**769**
	Shared	0	662	317	117	60	237	103	**1496**
Overlap	SSR-Pipeline	0	0	0	0	1	2	0	**3**
	SA-SSR	2742	6076	2181	325	134	556	536	**12550**
	Shared	0	709	333	124	60	242	109	**1577**
Exhaustive	SSR-Pipeline	0	0	0	0	0	0	0	**0**
	SA-SSR	2752	8836	3771	9870	1029	10117	1195	**37570**
	Shared	0	709	333	124	61	244	109	**1580**

The number of SSRs in the *Arabidopsis thaliana* chromosome 4 found unique to SSR-Pipeline, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 18600000 -m 1 -M 7 -n 16 -r 1 –i D,M,N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 84 SSRs that SSR-Pipeline found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 81 of the 84 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. One of the remaining 3 SSRs was just a different SSR base, but covering essentially the same SSR (AATAAA vs AAAATA). The remaining 2 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 9. SA-SSR compared with SSRIT for *Arabidopsis thaliana*

		1	2	3	4	5	6	7	Total
Normal	SSRIT	0	5	1	1	0	0	0	**7**
	SA-SSR	660	198	1	98	60	245	110	**1372**
	Shared	0	523	342	28	0	0	0	**893**
Overlap	SSRIT	0	0	0	0	0	0	0	**0**
	SA-SSR	2742	6257	2171	420	194	798	645	**13227**
	Shared	0	528	343	29	0	0	0	**900**

The number of SSRs in the *Arabidopsis thaliana* chromosome 4 found unique to SSRIT, unique to SA-SSR, and shared between the two using two different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 18600000 -m 1 -M 7 -n 16 -r 1 –i D,M,N. The overlap set was identical to normal with the following addition: -o. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 7 SSRs that SSRIT found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. All 7 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result.

Supplementary Table 10. SA-SSR compared with TRF for *Arabidopsis thaliana*

		1	2	3	4	5	6	7	Total
Normal	TRF	0	48	26	9	2	14	25	**124**
	SA-SSR	660	67	41	13	3	42	36	**862**
	Shared	0	654	302	113	57	203	74	**1403**
Overlap	TRF	0	1	3	5	0	3	1	**13**
	SA-SSR	2742	6084	2189	332	135	584	547	**12613**
	Shared	0	701	325	117	59	214	98	**1514**
Exhaustive	TRF	0	1	3	0	0	1	1	**6**
	SA-SSR	2752	8844	3779	9872	1031	10145	1206	**37629**
	Shared	0	701	325	122	59	216	98	**1521**

The number of SSRs in the *Arabidopsis thaliana* chromosome 4 found unique to TRF, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 18600000 -m 1 -M 7 -n 16 -r 1 –i D,M,N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 124 SSRs that TRF found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 111 of the 124 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 13 SSRs were also found by SA-SSR and they fall into three different categories. The categories are overstated period size, finding different numbers of repeats, and special cases requiring the exhaustive approach by SA-SSR. 6 of the 13 are cases where TRF overstated the period size (e.g., calling ATATATAT a 4-mer instead of a 2-mer). Obviously, reporting a longer period length than is strictly necessary to describe the SSR is misleading and certainly incorrect. AAAAAAAAA has a period size of one repeated nine times, not three repeated three times. Likewise, ATATATAT has a period size of two repeated four times, not four repeated two times. Of the remaining 7, the 6 that were not found even under the exhaustive approach were actually found by SA-SSR, but SA-SSR correctly reported a larger number of repeats. So, while it appeared that SA-SSR didn't find them, it actually did. For these 6, both are correct, but SA-SSR is more complete. Finally, the last of the 7 was found during the exhaustive approach and is a special, rare case involving the specific sequence and suffix sort. Of course, the number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 11. SA-SSR compared with GMATo for *Caenorhabditis elegans*

		1	2	3	4	5	6	7	Total
Normal	GMATo	0	687	220	248	55	807	274	**2291**
	SA-SSR	522	866	428	601	130	1551	565	**4663**
	Shared	0	1032	415	393	50	1097	273	**3260**
Overlap	GMATo	0	3	0	5	0	12	1	**21**
	SA-SSR	1862	13378	2802	4084	661	16224	5361	**44372**
	Shared	0	1716	635	636	105	1892	546	**5530**
Exhaustive	GMATo	0	0	0	0	0	0	0	**0**
	SA-SSR	1862	15261	3803	21089	1258	32453	5858	**81584**
	Shared	0	1719	635	641	105	1904	547	**5551**

The number of SSRs in the *Caenorhabditis elegans* genome found unique to GMATo, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 700000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 2291 SSRs that GMATo found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 2270 of the 2291 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. One of the remaining 21 SSRs were also found by SA-SSR, but SA-SSR correctly reported a greater number of repeats than GMATo did. Finally, the last 20 were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 12. SA-SSR compared with MREPS for *Caenorhabditis elegans*

		1	2	3	4	5	6	7	Total
Normal	MREPS	0	11	3	16	0	39	15	**84**
	SA-SSR	522	6	0	8	0	22	9	**567**
	Shared	0	1892	843	986	180	2626	829	**7356**
Overlap	MREPS	0	5	1	8	0	14	2	**30**
	SA-SSR	1862	13196	2592	3726	586	15465	5065	**42492**
	Shared	0	1898	845	994	180	2651	842	**7410**
Exhaustive	MREPS	0	0	0	0	0	0	0	**0**
	SA-SSR	1862	15077	3592	20728	1183	31692	5561	**79695**
	Shared	0	1903	846	1002	180	2665	844	**7440**

The number of SSRs in the *Caenorhabditis elegans* genome found unique to MREPS, unique to SA-SSR, and shared between the two using the three different parameter sets. The normal parameter set was as follows: -l 1 -L 700000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 84 SSRs that MREPS found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 54 of the 84 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. Four of the remaining 30 SSRs were also found by SA-SSR, but SA-SSR reported a different repeating unit than MREPS did (e.g., GT vs TG). Finally, the last 26 were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 13. SA-SSR compared with ProGeRF for *Caenorhabditis elegans*

		1	2	3	4	5	6	7	Total
Normal	ProGeRF	0	0	0	0	0	1	1	**2**
	SA-SSR	522	1871	833	971	179	2620	830	**7826**
	Shared	0	27	10	23	1	28	8	**97**
Overlap	ProGeRF	0	0	0	0	0	1	0	**1**
	SA-SSR	1862	15067	3427	4697	765	18088	5898	**49804**
	Shared	0	27	10	23	1	28	9	**98**

The number of SSRs in the *Caenorhabditis elegans* genome found unique to ProGeRF, unique to SA-SSR, and shared between the two using two different sets of parameters for SA-SSR. The normal parameter set was as follows: -l -L 700000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 2 SSRs that ProGeRF found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 1 of the 2 was also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining SSR was also found by SA-SSR, but SA-SSR correctly reported shorter period lengths than ProGeRF did. Obviously, reporting a longer period length than is strictly necessary to describe the SSR is misleading and certainly incorrect. AAAAAAAAA has a period size of one repeated nine times, not three repeated three times. Likewise, ATATATAT has a period size of two repeated four times, not four repeated two times.

Supplementary Table 14. SA-SSR compared with QDD for *Caenorhabditis elegans*

		1	2	3	4	5	6	7	Total
Normal	QDD	0	8	1	4	0	3	0	**16**
	SA-SSR	522	4	0	715	141	2340	838	**4560**
	Shared	0	1894	843	279	39	308	0	**3363**
Overlap	QDD	0	5	1	0	0	1	0	**7**
	SA-SSR	1862	13197	2594	4437	727	17806	5907	**46530**
	Shared	0	1897	843	283	39	310	0	**3372**
Exhaustive	QDD	0	0	0	0	0	0	0	**0**
	SA-SSR	1862	15078	3594	21447	1324	34046	6405	**83756**
	Shared	0	1902	844	283	39	311	0	**3379**

The number of SSRs in the *Caenorhabditis elegans* genome found unique to QDD, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 700000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 16 SSRs that QDD found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 9 of the 16 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. One of the remaining 7 was a case where the two programs correctly reported different repeating units (e.g., GT vs TG). The remaining 6 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 15. SA-SSR compared with SSR-Pipeline for *Caenorhabditis elegans*

		1	2	3	4	5	6	7	Total
Normal	SSR-Pipeline	0	116	31	38	1	87	13	**286**
	SA-SSR	522	141	53	56	3	115	14	**904**
	Shared	0	1757	790	938	177	2533	824	**7019**
Overlap	SSR-Pipeline	0	5	1	5	0	14	2	**27**
	SA-SSR	1862	13226	2617	3749	588	15510	5072	**42624**
	Shared	0	1868	820	971	178	2606	835	**7278**
Exhaustive	SSR-Pipeline	0	0	0	0	0	0	0	**0**
	SA-SSR	1862	15107	3617	20754	1185	31737	5568	**79830**
	Shared	0	1873	821	976	178	2620	837	**7305**

The number of SSRs in the *Caenorhabditis elegans* genome found unique to SSR-Pipeline, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 700000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 286 SSRs that SSR-Pipeline found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 259 of the 286 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. Three of the remaining 27 were cases where the two programs correctly reported different repeating units (e.g., GT vs TG). The remaining 24 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 16. SA-SSR compared with SSRIT for *Caenorhabditis elegans*

		1	2	3	4	5	6	7	Total
Normal	SSRIT	0	8	3	6	0	0	0	**17**
	SA-SSR	522	662	0	716	180	2648	838	**5566**
	Shared	0	1236	843	278	0	0	0	**2357**
Overlap	SSRIT	0	2	1	0	0	0	0	**3**
	SA-SSR	1862	13852	2592	4436	766	18116	5907	**47531**
	Shared	0	1242	845	284	0	0	0	**2371**
Exhaustive	SSRIT	0	0	0	0	0	0	0	**0**
	SA-SSR	1862	15736	3592	21446	1363	34357	6405	**84761**
	Shared	0	1244	846	284	0	0	0	**2374**

The number of SSRs in the *Caenorhabditis elegans* genome found unique to SSRIT, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 700000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 17 SSRs that SSRIT found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 14 of the 17 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 3 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 17. SA-SSR compared with TRF for *Caenorhabditis elegans*

		1	2	3	4	5	6	7	Total
Normal	TRF	0	99	46	77	11	537	130	**900**
	SA-SSR	522	144	66	165	26	1443	283	**2649**
	Shared	0	1754	777	829	154	1205	555	**5274**
Overlap	TRF	0	9	8	10	3	17	2	**49**
	SA-SSR	1862	13250	2622	3824	604	16391	5224	**43777**
	Shared	0	1844	815	896	162	1725	683	**6125**
Exhaustive	TRF	0	8	7	2	3	5	1	**26**
	SA-SSR	1862	15135	3622	20826	1201	32620	5721	**80987**
	Shared	0	1845	816	904	162	1737	684	**6148**

The number of SSRs in the *Caenorhabditis elegans* genome found unique to TRF, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 700000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 900 SSRs that TRF found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 851 of the 900 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 49 SSRs were also found by SA-SSR and they fall into three different categories. The categories are overstated period size, finding different numbers of repeats, and special cases requiring the exhaustive approach by SA-SSR. 10 of the 49 are cases where TRF overstated the period size (e.g., calling ATATATAT a 4-mer instead of a 2-mer). Obviously, reporting a longer period length than is strictly necessary to describe the SSR is misleading and certainly incorrect. AAAAAAAAA has a period size of one repeated nine times, not three repeated three times. Likewise, ATATATAT has a period size of two repeated four times, not four repeated two times. Of the remaining 38, the 26 that were not found even under the exhaustive approach were actually found by SA-SSR. For 25 of the 26, SA-SSR correctly reported a larger number of repeats. So, while it appeared that SA-SSR didn't find them, it actually did. For these 25, both are correct, but SA-SSR is more complete. The last of the 26 was also found by SA-SSR, but SA-SSR correctly stated a shorter period size (another example where ATATATAT should be a 2-mer, not a 4-mer). This leaves us with 13 unaccounted for. 7 were more cases where TRF and SA-SSR either reported different SSRs (e.g., GT vs TG) or reported different number of repeats. Finally, the last 6 were found during the exhaustive approach and is a special, rare case involving the specific sequence and suffix sort. Of course, the number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 18. SA-SSR compared with GMATo for *Drosophila melanogaster*

		1	2	3	4	5	6	7	Total
Normal	GMATo	0	0	0	15	20	151	281	**467**
	SA-SSR	4734	8094	3286	3328	1088	5557	1207	**27294**
	Shared	0	0	0	15	25	228	318	**586**
Overlap	GMATo	0	0	0	1	1	6	9	**17**
	SA-SSR	31700	47110	16452	14537	4328	25154	6006	**145287**
	Shared	0	0	0	29	44	373	590	**1036**

The number of SSRs in the *Drosophila melanogaster* genome found unique to GMATo, unique to SA-SSR, and shared between the two using two different sets of parameters for SA-SSR. The normal parameter set was as follows: -L 1000000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 467 SSRs that GMATo found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 450 of the 467 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 17 SSRs were also found by SA-SSR, but SA-SSR correctly reported longer SSRs than GMATo did (e.g., in sequence JXOZ01000280.1, SA-SSR reported CAGGGAC repeated 7 times beginning at position 73168 while GMATo reported the same repeating only 4 times).

Supplementary Table 19. SA-SSR compared with MREPS for *Drosophila melanogaster*

		1	2	3	4	5	6	7	Total
Normal	MREPS	6	21	33	56	17	90	9	**232**
	SA-SSR	1	11	19	16	10	42	5	**104**
	Shared	4733	8083	3267	3327	1103	5743	1520	**27776**
Overlap	MREPS	2	2	0	36	3	1	0	**44**
	SA-SSR	26963	39008	13152	11219	3255	19695	5067	**118359**
	Shared	4737	8102	3300	3347	1117	5832	1529	**27964**
Exhaustive	MREPS	0	0	0	0	0	0	0	**0**
	SA-SSR	26963	70718	36560	90090	21713	91821	21709	**359574**
	Shared	4739	8104	3300	3383	1120	5833	1529	**28008**

The number of SSRs in the *Drosophila melanogaster* genome found unique to MREPS, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -L 1000000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 232 SSRs that MREPS found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 188 of the 232 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. 43 of the remaining 44 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The last SSR was a case where SA-SSR and MREPS simply reported a slightly different SSR (e.g., AT vs TA). The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 20. SA-SSR compared with ProGeRF for *Drosophila melanogaster*

		1	2	3	4	5	6	7	Total
Normal	ProGeRF	1	1	4	0	1	3	0	**10**
	SA-SSR	4651	7930	3233	3271	1095	5659	1485	**27324**
	Shared	83	164	53	72	18	126	40	**556**
Overlap	ProGeRF	0	1	2	0	0	1	0	**4**
	SA-SSR	31616	46946	16397	14494	4353	25399	6556	**145761**
	Shared	84	164	55	72	19	128	40	**562**

The number of SSRs in the *Drosophila melanogaster* genome found unique to ProGeRF, unique to SA-SSR, and shared between the two using two different sets of parameters for SA-SSR. The normal parameter set was as follows: -L 1000000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 10 SSRs that ProGeRF found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 6 of the 10 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 4 SSRs were also found by SA-SSR, but SA-SSR correctly reported shorter period lengths than ProGeRF did (e.g., in sequence JXOZ01000073.1, SA-SSR reported A repeated 19 times beginning at position 136707 while ProGeRF reported AAA repeating 6 times at the same position). Obviously, reporting a longer period length than is strictly necessary to describe the SSR is misleading and certainly incorrect. AAAAAAAAA has a period size of one repeated nine times, not three repeated three times. Likewise, ATATATAT has a period size of two repeated four times, not four repeated two times.

Supplementary Table 21. SA-SSR compared with QDD for *Drosophila melanogaster*

		1	2	3	4	5	6	7	Total
Normal	QDD	0	25	22	8	6	2	0	**63**
	SA-SSR	4734	18	15	2246	880	5594	1525	**15012**
	Shared	0	8076	3271	1097	233	191	0	**12868**
Overlap	QDD	0	2	0	2	0	0	0	**4**
	SA-SSR	31700	39011	13159	13463	4133	25334	6596	**133396**
	Shared	0	8099	3293	1103	239	193	0	**12927**
Exhaustive	QDD	0	0	0	0	0	0	0	**0**
	SA-SSR	31702	70721	36567	92368	22594	97461	23238	**374651**
	Shared	0	8101	3293	1105	239	193	0	**12931**

The number of SSRs in the *Drosophila melanogaster* genome found unique to QDD, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -L 1000000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 63 SSRs that QDD found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 59 of the 63 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 4 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 22. SA-SSR compared with SSR-Pipeline for *Drosophila melanogaster*

		1	2	3	4	5	6	7	Total
Normal	SSR-Pipeline	6	386	207	152	45	166	25	**987**
	SA-SSR	1	473	271	190	70	298	51	**1354**
	Shared	4733	7621	3015	3153	1043	5487	1474	**26526**
Overlap	SSR-Pipeline	2	2	0	36	2	1	0	**43**
	SA-SSR	26963	39105	13230	11297	3286	19875	5097	**118853**
	Shared	4737	8005	3222	3269	1086	5652	1499	**27470**
Exhaustive	SSR-Pipeline	0	0	0	0	0	0	0	**0**
	SA-SSR	26963	70815	36638	90168	21745	92001	21739	**360069**
	Shared	4739	8007	3222	3305	1088	5653	1499	**27513**

The number of SSRs in the *Drosophila melanogaster* genome found unique to SSR-Pipeline, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -L 1000000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 987 SSRs that SSR-Pipeline found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 944 of the 987 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. 42 of the remaining 43 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The last SSR was a case where SA-SSR and SSR-Pipeline simply reported a slightly different SSR (e.g., AT vs TA). The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 23. SA-SSR compared with SSRIT for *Drosophila melanogaster*

		1	2	3	4	5	6	7	Total
Normal	SSRIT	0	12	32	12	0	0	0	**56**
	SA-SSR	4734	2570	18	2248	1113	5785	1525	**17993**
	Shared	0	5524	3268	1095	0	0	0	**9887**
Overlap	SSRIT	0	0	0	2	0	0	0	**2**
	SA-SSR	31700	41574	13152	13461	4372	25527	6596	**136382**
	Shared	0	5536	3300	1105	0	0	0	**9941**
Exhaustive	SSRIT	0	0	0	0	0	0	0	**0**
	SA-SSR	31702	73286	36560	92366	22833	97654	23238	**377639**
	Shared	0	5536	3300	1107	0	0	0	**9943**

The number of SSRs in the *Drosophila melanogaster* genome found unique to SSRIT, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -L 1000000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 56 SSRs that SSRIT found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 54 of the 56 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 2 SSRs were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 24. SA-SSR compared with TRF for *Drosophila melanogaster*

		1	2	3	4	5	6	7	Total
Normal	TRF	5	769	373	323	61	528	128	**2187**
	SA-SSR	22	1042	551	544	210	1224	318	**3911**
	Shared	4712	7052	2735	2799	903	4561	1207	**23969**
Overlap	TRF	1	53	14	54	9	36	2	**169**
	SA-SSR	26984	39342	13358	11498	3417	20474	5263	**120336**
	Shared	4716	7768	3094	3068	955	5053	1333	**25987**
Exhaustive	TRF	0	52	13	15	8	13	2	**103**
	SA-SSR	26985	71053	36765	90366	21877	92578	21905	**361529**
	Shared	4717	7769	3095	3107	956	5076	1333	**26053**

The number of SSRs in the *Drosophila melanogaster* genome found unique to TRF, unique to SA-SSR, and shared between the two using three different sets of parameters for SA-SSR. The normal parameter set was as follows: -L 1000000 -m 1 -M 7 -n 16 -r 1 –i N. The overlap set was identical to normal with the following addition: -o. The exhaustive set was identical to overlap with the following addition: -e. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 2187 SSRs that TRF found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. 2018 of the 2187 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result. The remaining 169 SSRs were also found by SA-SSR and they fall into three different categories. The categories are overstated period size, finding different numbers of repeats, and special cases requiring the exhaustive approach by SA-SSR. 60 of the 169 are cases where TRF overstated the period size (e.g., in sequence JXOZ01000843.1, TRF reports an AGAG repeating 4 times at position 109312 while SA-SSR correctly reports an AG repeated 8 times at the same position). 2 of these appear again in the 103 that SA-SSR didn't appear to find using the exhaustive parameter set, but SA-SSR did find them, it just reported the correct period size. Obviously, reporting a longer period length than is strictly necessary to describe the SSR is misleading and certainly incorrect. AAAAAAAAA has a period size of one repeated nine times, not three repeated three times. Likewise, ATATATAT has a period size of two repeated four times, not four repeated two times.
The remaining 111 cases fall into the other two categories. 104 of the 169 are cases where TRF and SA-SSR reported different SSRs (e.g., AT vs TA) or TRF reported less repeats of the same SSR (e.g., in sequence JXOZ01001169.1, TRF reports a TTTCGA repeated 3 times at position 83483 while SA-SSR reports the same repeated 4 times). 101 of these also appear not to be found using the exhaustive parameter set because SA-SSR correctly reported SSRs with more repeats. The remaining 5 were also found by SA-SSR, but only when using the exhaustive approach because of a special, rare case involving the specific sequence and suffix sort order. The number of unique SSRs found by SA-SSR as reported using the exhaustive parameter set is also inflated.

Supplementary Table 25. SA-SSR compared with GMATo for *Escherichia coli*

		1	2	3	4	5	6	7	Total
Normal	GMATo	0	0	0	0	0	7	1	**8**
	SA-SSR	1	0	0	0	0	13	1	**15**
	Shared	0	0	0	1	0	4	0	**5**
Overlap	GMATo	0	0	0	0	0	0	0	**0**
	SA-SSR	5	0	0	2	0	287	36	**330**
	Shared	0	0	0	1	0	11	1	**13**

The number of SSRs in the *Escherichia coli* genome found unique to GMATo, unique to SA-SSR, and shared between the two using two different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 600000 -m 1 -M 7 -n 16 -r 1. The overlap set was identical to normal with the following addition: -o. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 8 SSRs that GMATo found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. All 8 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result.

Supplementary Table 26. SA-SSR compared with MREPS for *Escherichia coli*

	1	2	3	4	5	6	7	Total
MREPS	0	0	0	0	0	0	0	**0**
SA-SSR	1	0	0	0	0	0	0	**1**
Shared	0	0	0	1	0	17	1	**19**

The number of SSRs in the *Escherichia coli* genome found unique to MREPS, unique to SA-SSR, and shared between the two using the following parameter set: -l 1 -L 600000 -m 1 -M 7 -n 16 -r 1. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Supplementary Table 27. SA-SSR compared with ProGeRF for *Escherichia coli*

	1	2	3	4	5	6	7	Total
ProGeRF	0	0	0	0	0	0	0	**0**
SA-SSR	1	0	0	1	0	13	1	**16**
Shared	0	0	0	0	0	4	0	**4**

The number of SSRs in the *Escherichia coli* genome found unique to ProGeRF, unique to SA-SSR, and shared between the two using the following parameter set: -l 1 -L 600000 -m 1 -M 7 -n 16 -r 1. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Supplementary Table 28. SA-SSR compared with QDD for *Escherichia coli*

		1	2	3	4	5	6	7	Total
Normal	QDD	0	0	0	0	0	8	0	**8**
	SA-SSR	1	0	0	1	0	17	1	**20**
	Shared	0	0	0	0	0	0	0	**0**
Overlap	QDD	0	0	0	0	0	0	0	**0**
	SA-SSR	5	0	0	3	0	290	37	**335**
	Shared	0	0	0	0	0	8	0	**8**

The number of SSRs in the *Escherichia coli* genome found unique to QDD, unique to SA-SSR, and shared between the two using two different sets of parameters for SA-SSR. The normal parameter set was as follows: -l 1 -L 600000 -m 1 -M 7 -n 16 -r 1. The overlap set was identical to normal with the following addition: -o. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Why did SA-SSR not find the 8 SSRs that QDD found uniquely? By default, SA-SSR reports only one SSR when multiple may be found in an overlapping location. All 8 were also found by SA-SSR when this default behavior is changed to report every SSR, even though they overlap. Naturally, the number of unique SSRs found by SA-SSR as reported using the overlap parameter set is inflated as a result.

Supplementary Table 29. SA-SSR compared with SSR-Pipeline for *Escherichia coli*

	1	2	3	4	5	6	7	Total
SSR-Pipeline	0	0	0	0	0	0	0	**0**
SA-SSR	1	0	0	1	0	17	1	**20**
Shared	0	0	0	0	0	0	0	**0**

The number of SSRs in the *Escherichia coli* genome found unique to SSR-Pipeline, unique to SA-SSR, and shared between the two using the following parameter set: -l 1 -L 600000 -m 1 -M 7 -n 16 -r 1. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Supplementary Table 30. SA-SSR compared with SSRIT for *Escherichia coli*

	1	2	3	4	5	6	7	Total
SSRIT	0	0	0	0	0	0	0	**0**
SA-SSR	1	0	0	1	0	17	1	**20**
Shared	0	0	0	0	0	0	0	**0**

The number of SSRs in the *Escherichia coli* genome found unique to SSRIT, unique to SA-SSR, and shared between the two using the following parameter set: -l 1 -L 600000 -m 1 -M 7 -n 16 -r 1. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Supplementary Table 31. SA-SSR compared with TRF for *Escherichia coli*

	1	2	3	4	5	6	7	Total
TRF	0	0	0	0	0	0	0	**0**
SA-SSR	1	0	0	0	0	0	0	**1**
Shared	0	0	0	1	0	17	1	**19**

The number of SSRs in the *Escherichia coli* genome found unique to TRF, unique to SA-SSR, and shared between the two using the following parameter set: -l 1 -L 600000 -m 1 -M 7 -n 16 -r 1. Any SSRs with period size greater than 7, with total length less than 16nt, or that were incorrect were excluded from this comparison.

Supplementary References

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences, *Nucleic acids research*, **27**, 573.

Kolpakov, R., Bana, G. and Kucherov, G. (2003) mreps: efficient and flexible detection of tandem repeats in DNA, *Nucleic acids research*, **31**, 3672-3678.

Lopes, R.d.S.*, et al.* (2015) ProGeRF: Proteome and Genome Repeat Finder Utilizing a Fast Parallel Hash Function, *BioMed research international*, **2015**.

Meglécz, E.*, et al.* (2014) QDD version 3.1: a user‐friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate, *Molecular ecology resources*, **14**, 1302-1313.

Miller, M.P.*, et al.* (2013) SSR_pipeline: A bioinformatic infrastructure for identifying microsatellites from paired-end Illumina high-throughput DNA sequencing data, *Journal of Heredity*, est056.

Temnykh, S.*, et al.* (2001) Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential, *Genome research*, **11**, 1441-1452.

Wang, X., Lu, P. and Luo, Z. (2013) GMATo: A novel tool for the identification and analysis of microsatellites in large genomes, *Bioinformation*, **9**, 541-544.