1 SUPPLEMENTARY METHODS

1.1 Exhaustive enumeration of words over the IUPAC alphabet

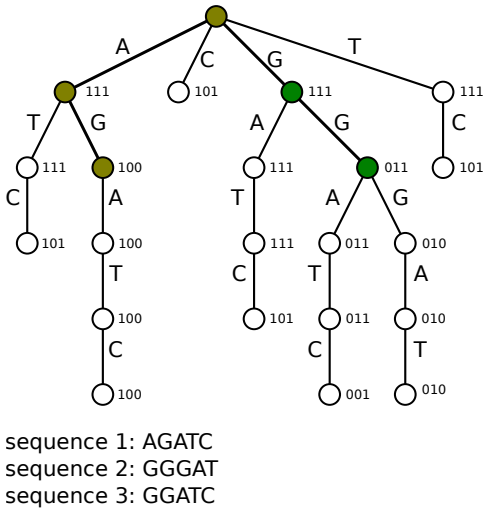

Fig. 1. Example of a generalized suffix tree representing three sequences (for simplicity, a trie is depicted).

1.1.1 Algorithm In alignment-free mode, words with a length between $k_{\min}$ and $k_{\max}$ that occur in a promoter sequence (including its reverse complement) of a gene family are exhaustively spelled in the IUPAC alphabet using a generalized suffix tree (GST). For example, the GST in Fig. 1 represents three sequences. For simplicity, the reverse complement of the sequences is not represented in this example. Each substring contained by any of the sequences is 'spelled out' along a path that originates from the root of the tree. A bitvector at each node of the GST represents the sequences that contain the substring implied by the path from the root to that node, e.g., '101' denotes occurrences in sequences 1 and 3 but not in sequence 2.

To enumerate all words that exist in the sequences, we rely on a modified version of the algorithm by Marsan and Sagot (2000), similar to what is used in the Mosdi tool (Marschall and Rahmann, 2009). A depth-first traversal of the GST is performed, examining a single word during each step. For example, assume the word RG is currently being processed (with R being the IUPAC character representing A or G). The algorithm will then have tracked paths AG and GG in the GST (green-colored nodes in Fig. 1). By taking the bitwise OR operation of the bitvectors contained by the respective nodes, i.e., '100' for AG and '011' for GG, it is immediately established that RG is contained by all three sequences. In later steps, longer words that have RG as a prefix are considered. Words of length 3 that will be considered are the words RGA, RGG, RGR and RGN. Note that e.g. RGC is never considered as no such path exists in the GST. Consequently, words that contain RGC as a prefix are also never considered. This branch-and-bound condition significantly reduces the number of words to be considered compared to exhaustively scoring all possible existing words over the IUPAC alphabet.

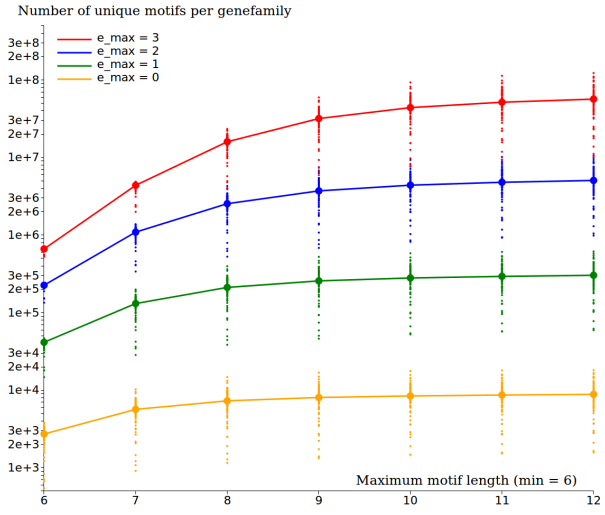

Fig. 2. Cumulative number of words in the restricted IUPAC alphabet (y-axis, log-scale) as a function of $k_{\max}$ (x-axis) and $e_{\max}$ (different curves) for 100 gene families. The solid lines indicate the average number of words over 100 gene families.

The search space can be further restricted if a certain word is found to be conserved with a BLS lower than the lowest BLS threshold used. In that case, the word does not need to be extended as such 'extended' words will have at most an equal BLS. For example, the word AG occurs only in sequence 1 (BLS = 0%). Therefore, there no need to visit e.g. AGA or AGAT.

Note that the GST can be truncated at a depth of $k_{\max}$ and contains only ACGT characters. To limit the memory requirements, words over the IUPAC alphabet are encoded using four bits per character.

1.1.2 Complexity and runtime Given $l_{\min}$, $l_{\max}$ and $e_{\max}$, an upper bound to the number of motifs considered in this algorithm is given by

$$\sum_{k=k_{\min}}^{k_{\max}} \sum_{e=0}^{e_{\max}} \binom{k}{e} 4^{k-e} d^e \qquad (1)$$

where d is the number of degenerate characters in the alphabet, i.e., $d = 1$ when using the ACGT+N alphabet, $d = 7$ when using the restricted IUPAC alphabet ACGT+RYSWKM+N and $d = 11$ for the full IUPAC alphabet. In reality, the number of words that is generated is lower because a) not all words exist in the input sequences and b) words with a too low BLS are never considered (see Suppl. Methods 1.1.1). Fig. 2 shows the cumulative number of words for a single gene family as a function of $k_{\max}$ and $e_{\max}$ for the restricted IUPAC alphabet. Fig. 3 shows that the runtime is largely proportional to the number of words to enumerate. Interactive versions of these graphs are available online[1,2].

[1] http://bioinformatics.intec.ugent.be/blsspeller/NumberOfMotifs.html

[2] http://bioinformatics.intec.ugent.be/blsspeller/MotifDiscoveryTime.html

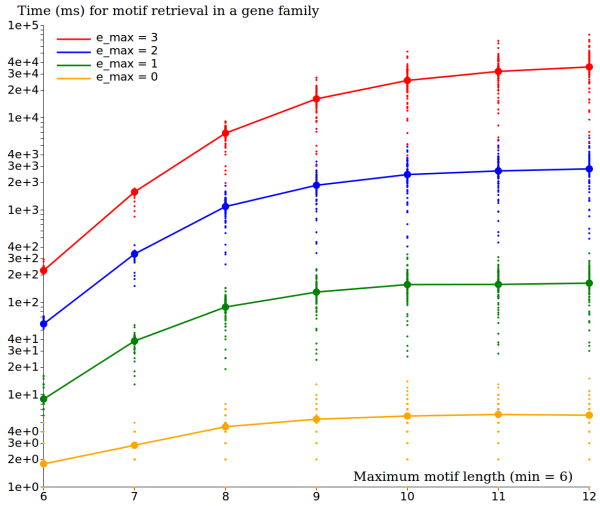

Fig. 3. Runtime (ms) to generate all words in the restricted IUPAC alphabet (y-axis, log-scale) as a function of $k_{\max}$ (x-axis) and $e_{\max}$ (different curves) for 100 gene families. The solid lines indicate the average runtime over 100 gene families. Runs were performed on a single core of an Intel Core i7-4610M CPU @ 3.00 GHz.

Table 1. Computational requirements of BLSSpeller using both alignment-free (AF) and alignment-based (AB) discovery on the Monocot dataset.

Computational requirements	AF	AB
Number of nodes (m1.xlarge)	20	20
Intrafamily step (Map phase) (hours)	33	10
Interfamily step (Reduce phase) (hours)	11	2
Map output records ($\times 10^9$)	537	82
Map output size (TByte)	3.77	0.53
Permutation groups	48 505	48 505
Reduce output records ($\times 10^9$)	6.6	6.3
Reduce output size (TByte)	0.46	0.41

1.2 Experimentally profiled open chromatin regions and transcription factor DNA binding affinity

DNase I hypersensitive (DH) sites for rice were determined as follows. The reads from DNase-seq were downloaded from the Gene Expression Omnibus, accession number GSE26610 and were aligned to the rice genome (TIGR release 6.1) using BWA (Li and Durbin, 2009). Only the reads that mapped to a unique position of the rice genome were used for further analysis. We used F-seq (Boyle *et al.*, 2008) to identify DNase I hypersensitive (DH) sites with 200 bp bandwidth. To detect the false discovery rate (FDR) of identified DH sites, we randomly generated 10 data sets, each containing the same amount of reads as the data set from DNase-seq. The FDR was calculated as the ratio of the number of DH sites identified from random data sets and the number of DH sites obtained from DNase-seq. We set a cutoff in F-seq to ensure that the FDR < 0.05. The callus and seedling datasets were merged using the BEDTools mergeBed function. This ensured a global picture of chromatin accessibility in the rice genome.

Regions with transcription factor binding sites interred through protein-binding microarrays in rice and maize were delineated as follows. The PWMs for rice and maize were downloaded from the CIS-BP website version 1.01 (Weirauch *et al.*, 2014). Only TFs with a directly obtained PWM or the best inferred PWM were used. The PWMs were mapped to the 2kb upstream region of the rice and maize genome using Matrixscan (Thomas-Chollier *et al.*, 2008) with a p-value cutoff of 1e-5.

The DH sites dataset and predicted transcription factor binding sites were formatted as BED files and the overlap with conserved motifs, also formatted as BED files, was determined using the BEDTools function intersectBed with -u parameter and the -f parameter set to 1. This means that a conserved motif region was considered to be in a DH site if it was completely overlapping with a DH site. The overlap analysis for the set of experimentally predicted transcription factor binding sites was performed in the same way with the exception that we now required that the experimentally predicted binding site was completely overlapping with a conserved motif region. The expected amount of conserved motifs in DH sites or overlapping with predicted transcription factor binding sites was determined by shuffling the conserved motif dataset 1000 times using shuffleBed with the -noOverlapping option enabled across the 2kb upstream regions. The overlap was determined for each shuffled file and the median number of conserved motifs over 1000 shuffled files was used as a measure for the expected presence of conserved motifs in DH sites or overlapping with predicted transcription factor binding sites. This estimation was used to calculate the fold enrichment, defined as the ratio between observed overlap and expected overlap by chance.

2 SUPPLEMENTARY RESULTS

2.1 Exhaustive motif discovery in four monocot species

BLSSpeller was run on this dataset using both the alignment-free (AF) and the alignment-based (AB) discovery mode on the Amazon Web Services (Elastic MapReduce) cloud infrastructure using 20 nodes of the type m1.xlarge. On every node, 7 map tasks and 2 reduce tasks were run in parallel. The computational requirements are listed in Table 1.

2.2 Estimation of the False Discovery Rate (FDR)

2.2.1 Limitations The false discovery rate (FDR) was estimated by the ratio of the number of motifs identified by BLSSpeller in the randomized dataset and a real dataset. However, it should be noted that both in the real and randomized dataset, correlations exist between motifs, i.e., for a given genome-wide conserved motif, many highly similar motifs (e.g., slightly more degenerate) that correspond to the same TF binding site also appear in the output. When assessing the FDR, we assume that the degree of correlation between motifs is comparable in real and randomized datasets.

2.2.2 Randomized datasets using higher-order Markov models Randomized datasets were generated using RSAT (Thomas-Chollier *et al.*, 2008) for a zeroth-, first- and second-order Markov model. Whereas as zeroth-order Markov model preserves only the relative mononucleotide (A, C, G, T) occurrences, a first-order Markov model preserves also relative dinucleotide composition. Similarly,

a third-order Markov models preserves both mono-, di- as well as trinucleotide composition. FDR tabels are provided in Fig. 4, 5 and 6. The use of higher-order Markov models results in an increased FDR, especially for relaxed settings of BLS, C and F thresholds. However, even for the second-order Markov model, the FDR for the datasets we used in the enrichment analysis is respectively 1.11% and 2.05%, which is still very low. The same remark holds for the motifs in the KN1 analysis.

2.2.3 False discovery rate as a function of motif length and degeneracy Additional FDR analysis can be performed as function of motif length k and degeneracy s. Here, s is defined as the total number of exact words that are implied by the degenerate word, i.e., $s = 2^{d_2} \cdot 4^{d_4}$, where d_2, and d_4 denote the number of two-fold and four-fold degenerate characters in a word, respectively.

Fig. 7 shows the number of motifs and FDR for $C_{\text{thres}} = 0.7$, $F_{\text{thres}} = 20$ and BLS $\geq 15\%$ as a function of motif length k and degeneracy s. This illustrates that the FDR is under control for all lengths and degeneracies. An interactive version of this graph can be explored online[3].

2.3 Comparison to Fastcompare

Fastcompare (Elemento and Tavazoie, 2005, 2007) is similar to BLSSpeller in that sense that it scores words (*k*-mers) that are conserved in homologous promoter sequences of related species, in an alignment-free and exhaustive manner. As opposed to BLSSpeller, Fastcompare is limited to pairwise comparisons between species and limited to the ACGT alphabet. Words are ranked according to the hypergeometric *p*-value given by

$$P(X \geq i) = \sum_{x=F}^{\min(s_1,s_2)} \frac{\binom{s_1}{x}\binom{N-s_1}{s_2-x}}{\binom{N}{s_2}} \qquad (2)$$

where s_1 and s_2 denote the number of gene families in which the word is present in the promoter region in the first and second species, respectively, F is the number of gene families in which the word is present in both first and second species and N denotes the total number of gene families. The *p*-value reflects the probability of observing conservation in at least i gene families by chance.

We ran Fastcompare on three pairwise species combinations: *Zea mays* (zma) vs. *Sorghum bicolor* (sbi), *Zea mays* vs. *Brachypodium distachyon* (bdi) and *Zea mays* vs. *Oryza sativa ssp. indica* (osa). Each time, the input consists of all pairs of orthologous promoter sequences from both species, i.e., 26 366 for zma vs. sbi; 26 099 for zma vs. bdi; 24 966 for zma vs. osa. Fastcompare was run with $k = 12$ and produces as output a ranked list of motifs from which ga2ox1-like KN1 motifs were filtered. Motif variants that are conserved in at least one gene family are listed in Table 2. Only few variants were conserved in more than one gene family, again illustrating the fact that degeneracy in the motif model is essential for a sensitive detection of motifs in diverged species. Most variants were found to be conserved in zma and sbi, the two most closely related species in the dataset. In total, over all species combinations, 36 unique maize genes were identified in which a ga2ox1-like KN1 motif was conserved, 10 of which overlap with the experimentally profiled maize genes. Note that no multiple hypothesis correction was applied to the *p*-values.

[3] http://bioinformatics.intec.ugent.be/blsspeller/AFABHistograms.html

Table 2. List of 25 ga2ox1-like KN1 motif variants retrieved by Fastcompare using alignment-free discovery in zma vs. sbi, zma vs. bdi and zma vs. osa. Here, F denotes the number of gene families in which the motif variant is conserved. $\mathcal{M}_{\text{FC}}$ denotes the number of (unique) maize genes contained in the gene families while M_{inters} denotes the intersection $M_{\text{FC}} \cap M_{\text{ChIP}}$ with experimentally profiled maize genes.

KN1 motif variant	species	F	p-value	rank	$\mathcal{M}_{\text{FC}}$	$\mathcal{M}_{\text{inters}}$
TGACTGACTGAC	zma-sbi	5	6.52e-08	78705	5	2
TGATGGATGGAT	zma-sbi	5	4.73e-06	182102	4	2
TGACAGACTGAC	zma-sbi	2	2.89e-05	218551	2	1
TGACCGACTGAC	zma-sbi	2	4.19e-05	233416	2	0
TGATTGATTGAT	zma-sbi	5	0.000144	284771	5	0
TGATCGACAGAT	zma-sbi	1	0.000303	365985	1	1
TGACCGACAGAC	zma-sbi	1	0.000303	376590	1	0
TGACAGACGGAC	zma-sbi	1	0.000379	409007	1	0
TGACCGATGGAC	zma-sbi	1	0.000569	468802	1	0
TGATAGACAGAT	zma-sbi	1	0.00102	564663	1	0
TGACTGATTGAT	zma-sbi	1	0.00121	597227	1	0
TGATGGACGGAC	zma-sbi	1	0.0019	680266	1	1
TGACAGATTGAC	zma-sbi	1	0.00227	707550	1	0
TGACAGATGGAT	zma-sbi	1	0.00273	735887	1	0
TGATTGATGGAC	zma-sbi	1	0.00545	822388	1	1
TGATTGACTGAT	zma-sbi	1	0.00583	828114	1	0
TGATGGATGGAC	zma-sbi	1	0.00816	856784	1	0
TGATTGACAGAT	zma-sbi	1	0.00907	865156	1	0
TGATCGATGGAT	zma-sbi	1	0.00982	870353	1	0
TGACTGACTGAT	zma-sbi	1	0.0163	897439	1	0
TGATTGATGGAT	zma-sbi	1	0.0185	902941	1	1
TGACTGACTGAT	zma-bdi	1	0.00855	63787	1	1
TGACTGACTGAC	zma-bdi	1	0.102	77976	1	1
TGATGGATGGAT	zma-osa	2	0.0234	80796	2	1
TGACTGACTGAC	zma-osa	1	0.0615	86863	1	1
Union (all variants)	–	39	–	–	36	10

REFERENCES

Boyle, A. P., Guinney, J., *et al.* (2008). F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics*, **24**(21), 2537–2538.

Elemento, O. and Tavazoie, S. (2005). Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. *Genome Biology*, **6**(2), R18+.

Elemento, O. and Tavazoie, S. (2007). Fastcompare: a nonalignment approach for genome-scale discovery of DNA and mRNA regulatory elements using network-level conservation. *Methods Mol Biol*, **395**, 349–366.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with BurrowsWheeler transform. *Bioinformatics*, **25**(14), 1754–1760.

Marsan, L. and Sagot, M. F. (2000). Algorithms for extracting structured motifs using a suffix tree with an application to promoter and regulatory site consensus identification. *Journal of computational biology*, **7**(3-4), 345–362.

Marschall, T. and Rahmann, S. (2009). Efficient exact motif discovery. *Bioinformatics (Oxford, England)*, **25**(12), 356–364.

Thomas-Chollier, M., Sand, O., *et al.* (2008). RSAT: regulatory sequence analysis tools. *Nucleic Acids Research*, **36**(suppl 2), W119–W127.

Weirauch, M. T., Yang, A., *et al.* (2014). Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell*, **158**(6), 1431–1443.

Alignment-free discovery					Alignment-based discovery				
		BLS thresholds T_i used					BLS threshold T_i used		
C_{thres}	F_{thres}	$T_1, \ldots, T_6$	$T_4, \ldots, T_6$	T_6 only	C_{thres}	F_{thres}	$T_1, \ldots, T_6$	$T_4, \ldots, T_6$	T_6 only
≥ 0.5	≥ 1	6.62E9 (4.09E9)	2.56E9 (4.32E8)	7.92E8 (4.57E7)	≥ 0.5	≥ 1	6.26E9 (3.77E8)	1.95E9 (3.47E6)	6.61E8 (1.04E5)
	≥ 5	2.24E9 (4.29E8)	2.90E8 (3.28E7)	5.47E7 (3.09E6)		≥ 5	1.25E9 (1.26E7)	1.19E8 (1.51E5)	2.50E7 (4.86E3)
	≥ 10	1.08E9 (9.24E7)	1.39E8 (5.68E6)	2.74E7 (6.21E5)		≥ 10	4.34E8 (2.19E6)	3.68E7 (1.73E4)	7.23E6 (34)
	≥ 20	5.34E8 (1.05E7)	7.55E7 (4.62E5)	1.57E7 (3.69E4)		≥ 20	1.47E8 (1.38E5)	1.33E7 (1.40E3)	2.54E6 (2)
	≥ 50	2.24E8 (7.87E4)	3.73E7 (1.34E3)	8.46E6 (100)		≥ 50	3.86E7 (197)	3.61E6 (1)	6.60E5 (0)
≥ 0.6	≥ 1	5.55E9 (3.33E9)	2.43E9 (3.77E8)	7.55E8 (3.75E7)	≥ 0.6	≥ 1	5.45E9 (3.47E8)	1.89E9 (3.11E6)	6.36E8 (9.53E4)
	≥ 5	1.77E9 (2.79E8)	2.38E8 (1.99E7)	4.46E7 (1.56E6)		≥ 5	9.68E8 (7.64E6)	1.01E8 (9.91E4)	2.10E7 (3.52E3)
	≥ 10	7.76E8 (3.73E7)	1.08E8 (1.49E6)	2.08E7 (9.84E4)		≥ 10	3.04E8 (9.11E5)	2.78E7 (4.19E3)	5.41E6 (28)
	≥ 20	3.74E8 (2.93E6)	5.52E7 (5.22E4)	1.16E7 (2.89E3)		≥ 20	9.50E7 (1.96E4)	9.53E6 (98)	1.74E6 (0)
	≥ 50	1.51E8 (4.89E3)	2.66E7 (39)	5.90E6 (0)		≥ 50	2.29E7 (4)	2.48E6 (0)	4.42E5 (0)
≥ 0.7	≥ 1	4.98E9 (2.95E9)	2.36E9 (3.53E8)	7.31E8 (3.42E7)	≥ 0.7	≥ 1	5.07E9 (3.32E8)	1.86E9 (2.95E6)	6.22E8 (9.10E4)
	≥ 5	1.11E9 (1.15E8)	1.75E8 (6.37E6)	3.30E7 (4.58E5)		≥ 5	6.32E8 (3.31E6)	7.54E7 (2.54E4)	1.56E7 (590)
	≥ 10	5.01E8 (1.55E7)	7.48E7 (6.50E5)	1.40E7 (3.77E4)		≥ 10	1.89E8 (2.73E5)	1.99E7 (1.15E3)	3.66E6 (15)
	≥ 20	2.23E8 (1.15E6)	3.64E7 (6.61E3)	7.63E6 (63)		≥ 20	5.16E7 (3.20E3)	6.17E6 (3)	1.12E6 (0)
	≥ 50	8.72E7 (244)	1.69E7 (1)	3.58E6 (0)		≥ 50	1.19E7 (0)	1.60E6 (0)	2.84E5 (0)
≥ 0.8	≥ 1	4.72E9 (2.82E9)	2.32E9 (3.47E8)	7.17E8 (3.33E7)	≥ 0.8	≥ 1	4.92E9 (3.28E8)	1.84E9 (2.91E6)	6.15E8 (9.01E4)
	≥ 5	7.77E8 (8.46E7)	1.39E8 (4.19E6)	2.52E7 (3.04E5)		≥ 5	4.86E8 (2.69E6)	6.44E7 (1.68E4)	1.28E7 (197)
	≥ 10	2.69E8 (6.43E6)	4.43E7 (1.69E5)	8.37E6 (3.19E3)		≥ 10	9.13E7 (7.33E4)	1.24E7 (285)	2.39E6 (4)
	≥ 20	1.08E8 (4.70E5)	2.11E7 (1.25E3)	4.29E6 (1)		≥ 20	2.40E7 (537)	3.59E6 (0)	7.06E5 (0)
	≥ 50	4.30E7 (2)	8.85E6 (0)	1.83E6 (0)		≥ 50	5.64E6 (0)	9.49E5 (0)	1.87E5 (0)
≥ 0.9	≥ 1	4.55E9 (2.76E9)	2.30E9 (3.45E8)	7.04E8 (3.30E7)	≥ 0.9	≥ 1	4.82E9 (3.26E8)	1.83E9 (2.90E6)	6.09E8 (8.99E4)
	≥ 5	1.90E8 (1.58E7)	5.76E7 (6.31E5)	1.23E7 (1.66E4)		≥ 5	1.29E8 (4.64E5)	2.92E7 (1.30E3)	6.71E6 (5)
	≥ 10	9.50E7 (2.64E6)	2.16E7 (4.16E4)	4.16E6 (141)		≥ 10	3.79E7 (3.59E4)	6.81E6 (10)	1.34E6 (0)
	≥ 20	3.85E7 (1.53E5)	8.71E6 (249)	1.77E6 (0)		≥ 20	8.73E6 (67)	1.89E6 (0)	3.70E5 (0)
	≥ 50	1.56E7 (0)	3.49E6 (0)	7.36E5 (0)		≥ 50	2.46E6 (0)	5.28E5 (0)	1.03E5 (0)

Legend

- $25\% \leq$ FDR
- $10\% \leq$ FDR $< 25\%$
- $5\% \leq$ FDR $< 10\%$
- $1\% \leq$ FDR $< 5\%$
- FDR $< 1\%$

Fig. 4. Number of genome-wide conserved motifs for both alignment-based and alignment-free discovery for different values of C_{thres} and F_{thres} and different subsets of the six BLS thresholds T_i ($T_1 = 15\%, T_2 = 50\%, T_3 = 60\%, T_4 = 70\%, T_5 = 90\%$ and $T_6 = 95\%$). Top number: real Monocot dataset; bottom number between brackets: random dataset generated using a zeroth-order Markov model (conservation of 1-mer frequencies). The colors represent the false discovery rate (see legend).

Alignment-free discovery					Alignment-based discovery				
		BLS thresholds T_i used					BLS threshold T_i used		
C_{thres}	F_{thres}	$T_1, \ldots, T_6$	$T_4, \ldots, T_6$	T_6 only	C_{thres}	F_{thres}	$T_1, \ldots, T_6$	$T_4, \ldots, T_6$	T_6 only
≥ 0.5	≥ 1	6.62E9 (4.51E9)	2.56E9 (4.80E8)	7.92E8 (5.99E7)	≥ 0.5	≥ 1	6.26E9 (3.89E8)	1.95E9 (3.55E6)	6.61E8 (1.12E5)
	≥ 5	2.24E9 (7.97E8)	2.90E8 (8.25E7)	5.47E7 (1.36E7)		≥ 5	1.25E9 (1.88E7)	1.19E8 (2.77E5)	2.50E7 (9.13E3)
	≥ 10	1.08E9 (3.43E8)	1.39E8 (4.34E7)	2.74E7 (9.25E6)		≥ 10	4.34E8 (6.40E6)	3.68E7 (7.77E4)	7.23E6 (920)
	≥ 20	5.34E8 (1.45E8)	7.55E7 (2.37E7)	1.57E7 (5.42E6)		≥ 20	1.47E8 (1.66E6)	1.33E7 (3.80E4)	2.54E6 (455)
	≥ 50	2.24E8 (5.48E7)	3.73E7 (1.28E7)	8.46E6 (2.93E6)		≥ 50	3.86E7 (3.12E5)	3.61E6 (6.84E3)	6.60E5 (84)
≥ 0.6	≥ 1	5.55E9 (3.77E9)	2.43E9 (4.22E8)	7.55E8 (4.93E7)	≥ 0.6	≥ 1	5.45E9 (3.58E8)	1.89E9 (3.18E6)	6.36E8 (1.02E5)
	≥ 5	1.77E9 (5.46E8)	2.38E8 (5.60E7)	4.46E7 (8.32E6)		≥ 5	9.68E8 (1.16E7)	1.01E8 (1.88E5)	2.10E7 (6.56E3)
	≥ 10	7.76E8 (1.80E8)	1.08E8 (2.56E7)	2.08E7 (5.31E6)		≥ 10	3.04E8 (3.13E6)	2.78E7 (2.88E4)	5.41E6 (517)
	≥ 20	3.74E8 (6.41E7)	5.52E7 (1.20E7)	1.16E7 (2.96E6)		≥ 20	9.50E7 (3.88E5)	9.53E6 (8.79E3)	1.74E6 (66)
	≥ 50	1.51E8 (1.86E7)	2.66E7 (5.83E6)	5.90E6 (1.35E6)		≥ 50	2.29E7 (5.13E4)	2.48E6 (850)	4.42E5 (34)
≥ 0.7	≥ 1	4.98E9 (3.31E9)	2.36E9 (3.87E8)	7.31E8 (4.28E7)	≥ 0.7	≥ 1	5.07E9 (3.41E8)	1.86E9 (2.99E6)	6.22E8 (9.65E4)
	≥ 5	1.11E9 (2.69E8)	1.75E8 (2.80E7)	3.30E7 (4.00E6)		≥ 5	6.32E8 (6.06E6)	7.54E7 (6.42E4)	1.56E7 (1.88E3)
	≥ 10	5.01E8 (7.62E7)	7.48E7 (1.18E7)	1.40E7 (2.31E6)		≥ 10	1.89E8 (1.01E6)	1.99E7 (1.02E4)	3.66E6 (396)
	≥ 20	2.23E8 (1.92E7)	3.64E7 (4.30E6)	7.63E6 (1.20E6)		≥ 20	5.16E7 (7.30E4)	6.17E6 (1.07E3)	1.12E6 (1)
	≥ 50	8.72E7 (3.30E6)	1.69E7 (1.63E6)	3.58E6 (4.07E5)		≥ 50	1.19E7 (3.37E3)	1.60E6 (87)	2.84E5 (1)
≥ 0.8	≥ 1	4.72E9 (3.11E9)	2.32E9 (3.70E8)	7.17E8 (3.95E7)	≥ 0.8	≥ 1	4.92E9 (3.34E8)	1.84E9 (2.92E6)	6.15E8 (9.45E4)
	≥ 5	7.77E8 (1.76E8)	1.39E8 (1.60E7)	2.52E7 (1.80E6)		≥ 5	4.86E8 (4.72E6)	6.44E7 (3.60E4)	1.28E7 (534)
	≥ 10	2.69E8 (2.45E7)	4.43E7 (3.53E6)	8.37E6 (6.00E5)		≥ 10	9.13E7 (1.71E5)	1.24E7 (3.44E3)	2.39E6 (113)
	≥ 20	1.08E8 (3.10E6)	2.11E7 (8.20E5)	4.29E6 (2.75E5)		≥ 20	2.40E7 (8.54E3)	3.59E6 (31)	7.06E5 (0)
	≥ 50	4.30E7 (1.89E5)	8.85E6 (1.49E5)	1.83E6 (4.83E4)		≥ 50	5.64E6 (2)	9.49E5 (0)	1.87E5 (0)
≥ 0.9	≥ 1	4.55E9 (2.98E9)	2.30E9 (3.61E8)	7.04E8 (3.80E7)	≥ 0.9	≥ 1	4.82E9 (3.30E8)	1.83E9 (2.89E6)	6.09E8 (9.40E4)
	≥ 5	1.90E8 (3.08E7)	5.76E7 (2.36E6)	1.23E7 (2.69E5)		≥ 5	1.29E8 (5.84E5)	2.92E7 (3.63E3)	6.71E6 (42)
	≥ 10	9.50E7 (5.66E6)	2.16E7 (4.68E5)	4.16E6 (4.28E4)		≥ 10	3.79E7 (5.02E4)	6.81E6 (123)	1.34E6 (2)
	≥ 20	3.85E7 (3.55E5)	8.71E6 (2.27E4)	1.77E6 (6.75E3)		≥ 20	8.73E6 (392)	1.89E6 (0)	3.70E5 (0)
	≥ 50	1.56E7 (668)	3.49E6 (625)	7.36E5 (556)		≥ 50	2.46E6 (0)	5.28E5 (0)	1.03E5 (0)

Legend

- $25\% \leq$ FDR
- $10\% \leq$ FDR $< 25\%$
- $5\% \leq$ FDR $< 10\%$
- $1\% \leq$ FDR $< 5\%$
- FDR $< 1\%$

Fig. 5. Number of genome-wide conserved motifs for both alignment-based and alignment-free discovery for different values of C_{thres} and F_{thres} and different subsets of the six BLS thresholds T_i ($T_1 = 15\%, T_2 = 50\%, T_3 = 60\%, T_4 = 70\%, T_5 = 90\%$ and $T_6 = 95\%$). Top number: real Monocot dataset; bottom number between brackets: random dataset generated using a first-order Markov model (conservation of 1-mer and 2-mer frequencies). The colors represent the false discovery rate (see legend).

Alignment-free discovery					Alignment-based discovery				
		BLS thresholds T_i used					BLS threshold T_i used		
C_{thres}	F_{thres}	$T_1, \ldots, T_6$	$T_4, \ldots, T_6$	T_6 only	C_{thres}	F_{thres}	$T_1, \ldots, T_6$	$T_4, \ldots, T_6$	T_6 only
	≥ 1	6.62E9 (4.51E9)	2.56E9 (4.88E8)	7.92E8 (6.25E7)		≥ 1	6.26E9 (3.94E8)	1.95E9 (3.62E6)	6.61E8 (1.22E5)
	≥ 5	2.24E9 (8.59E8)	2.90E8 (8.87E7)	5.47E7 (1.55E7)		≥ 5	1.25E9 (2.08E7)	1.19E8 (2.99E5)	2.50E7 (9.47E3)
≥ 0.5	≥ 10	1.08E9 (3.94E8)	1.39E8 (4.89E7)	2.74E7 (9.90E6)	≥ 0.5	≥ 10	4.34E8 (7.06E6)	3.68E7 (9.15E4)	7.23E6 (1.00E3)
	≥ 20	5.34E8 (1.75E8)	7.55E7 (2.89E7)	1.57E7 (6.22E6)		≥ 20	1.47E8 (2.21E6)	1.33E7 (4.87E4)	2.54E6 (353)
	≥ 50	2.24E8 (6.87E7)	3.73E7 (1.52E7)	8.46E6 (3.63E6)		≥ 50	3.86E7 (5.11E5)	3.61E6 (1.05E4)	6.60E5 (82)
	≥ 1	5.55E9 (3.77E9)	2.43E9 (4.32E8)	7.55E8 (5.19E7)		≥ 1	5.45E9 (3.63E8)	1.89E9 (3.26E6)	6.36E8 (1.13E5)
	≥ 5	1.77E9 (5.97E8)	2.38E8 (6.27E7)	4.46E7 (1.04E7)		≥ 5	9.68E8 (1.34E7)	1.01E8 (2.07E5)	2.10E7 (6.66E3)
≥ 0.6	≥ 10	7.76E8 (2.23E8)	1.08E8 (3.00E7)	2.08E7 (6.13E6)	≥ 0.6	≥ 10	3.04E8 (3.73E6)	2.78E7 (3.90E4)	5.41E6 (646)
	≥ 20	3.74E8 (8.38E7)	5.52E7 (1.61E7)	1.16E7 (3.59E6)		≥ 20	9.50E7 (6.66E5)	9.53E6 (1.50E4)	1.74E6 (48)
	≥ 50	1.51E8 (2.78E7)	2.66E7 (7.79E6)	5.90E6 (1.91E6)		≥ 50	2.29E7 (1.15E5)	2.48E6 (2.26E3)	4.42E5 (19)
	≥ 1	4.98E9 (3.31E9)	2.36E9 (3.96E8)	7.31E8 (4.48E7)		≥ 1	5.07E9 (3.46E8)	1.86E9 (3.06E6)	6.22E8 (1.07E5)
	≥ 5	1.11E9 (3.13E8)	1.75E8 (3.45E7)	3.30E7 (5.53E6)		≥ 5	6.32E8 (6.89E6)	7.54E7 (7.40E4)	1.56E7 (1.96E3)
≥ 0.7	≥ 10	5.01E8 (9.98E7)	7.48E7 (1.49E7)	1.40E7 (3.05E6)	≥ 0.7	≥ 10	1.89E8 (1.44E6)	1.99E7 (1.31E4)	3.66E6 (434)
	≥ 20	2.23E8 (2.91E7)	3.64E7 (6.64E6)	7.63E6 (1.60E6)		≥ 20	5.16E7 (1.27E5)	6.17E6 (2.57E3)	1.12E6 (1)
	≥ 50	8.72E7 (6.89E6)	1.69E7 (2.79E6)	3.58E6 (7.23E5)		≥ 50	1.19E7 (1.56E4)	1.60E6 (291)	2.84E5 (1)
	≥ 1	4.72E9 (3.09E9)	2.32E9 (3.79E8)	7.17E8 (4.08E7)		≥ 1	4.92E9 (3.39E8)	1.84E9 (2.98E6)	6.15E8 (1.05E5)
	≥ 5	7.77E8 (2.05E8)	1.39E8 (2.10E7)	2.52E7 (2.80E6)		≥ 5	4.86E8 (4.96E6)	6.44E7 (4.15E4)	1.28E7 (596)
≥ 0.8	≥ 10	2.69E8 (3.28E7)	4.43E7 (5.21E6)	8.37E6 (1.07E6)	≥ 0.8	≥ 10	9.13E7 (2.91E5)	1.24E7 (4.49E3)	2.39E6 (151)
	≥ 20	1.08E8 (5.62E6)	2.11E7 (1.57E6)	4.29E6 (4.65E5)		≥ 20	2.40E7 (1.94E4)	3.59E6 (162)	7.06E5 (0)
	≥ 50	4.30E7 (6.95E5)	8.85E6 (4.68E5)	1.83E6 (1.38E5)		≥ 50	5.64E6 (286)	9.49E5 (14)	1.87E5 (0)
	≥ 1	4.55E9 (2.94E9)	2.30E9 (3.68E8)	7.04E8 (3.84E7)		≥ 1	4.82E9 (3.35E8)	1.83E9 (2.95E6)	6.09E8 (1.05E5)
	≥ 5	1.90E8 (3.66E7)	5.76E7 (3.31E6)	1.23E7 (4.31E5)		≥ 5	1.29E8 (6.67E5)	2.92E7 (3.38E3)	6.71E6 (61)
≥ 0.9	≥ 10	9.50E7 (7.30E6)	2.16E7 (9.14E5)	4.16E6 (1.50E5)	≥ 0.9	≥ 10	3.79E7 (5.82E4)	6.81E6 (211)	1.34E6 (1)
	≥ 20	3.85E7 (4.28E5)	8.71E6 (8.91E4)	1.77E6 (3.64E4)		≥ 20	8.73E6 (312)	1.89E6 (1)	3.70E5 (0)
	≥ 50	1.56E7 (8.15E3)	3.49E6 (7.75E3)	7.36E5 (4.68E3)		≥ 50	2.46E6 (0)	5.28E5 (0)	1.03E5 (0)

Legend

- $25\% \leq$ FDR
- $10\% \leq$ FDR $< 25\%$
- $5\% \leq$ FDR $< 10\%$
- $1\% \leq$ FDR $< 5\%$
- FDR $< 1\%$

Fig. 6. Number of genome-wide conserved motifs for both alignment-based and alignment-free discovery for different values of C_{thres} and F_{thres} and different subsets of the six BLS thresholds T_i ($T_1 = 15\%, T_2 = 50\%, T_3 = 60\%, T_4 = 70\%, T_5 = 90\%$ and $T_6 = 95\%$). Top number: real Monocot dataset; bottom number between brackets: random dataset generated using a second-order Markov model (conservation of 1-mer, 2-mer and 3-mer frequencies). The colors represent the false discovery rate (see legend).

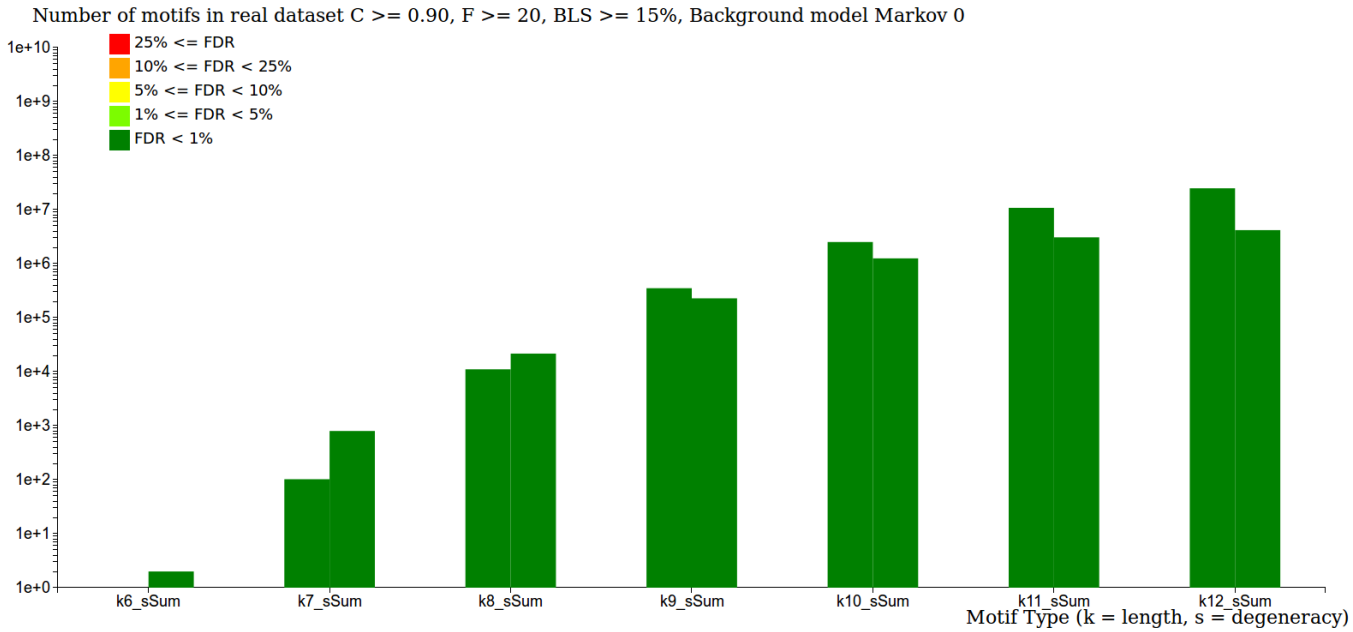

Fig. 7. Number of motifs (y-axis, log-scale) as a function of motif length k (x-axis) for both alignment-free (left bar) and alignment-based (right bar) discovery on the Monocot dataset with $C \geq 0.9$; $F \geq 20$; BLS $\geq 15\%$. The colors represents the false discovery rate (FDR).

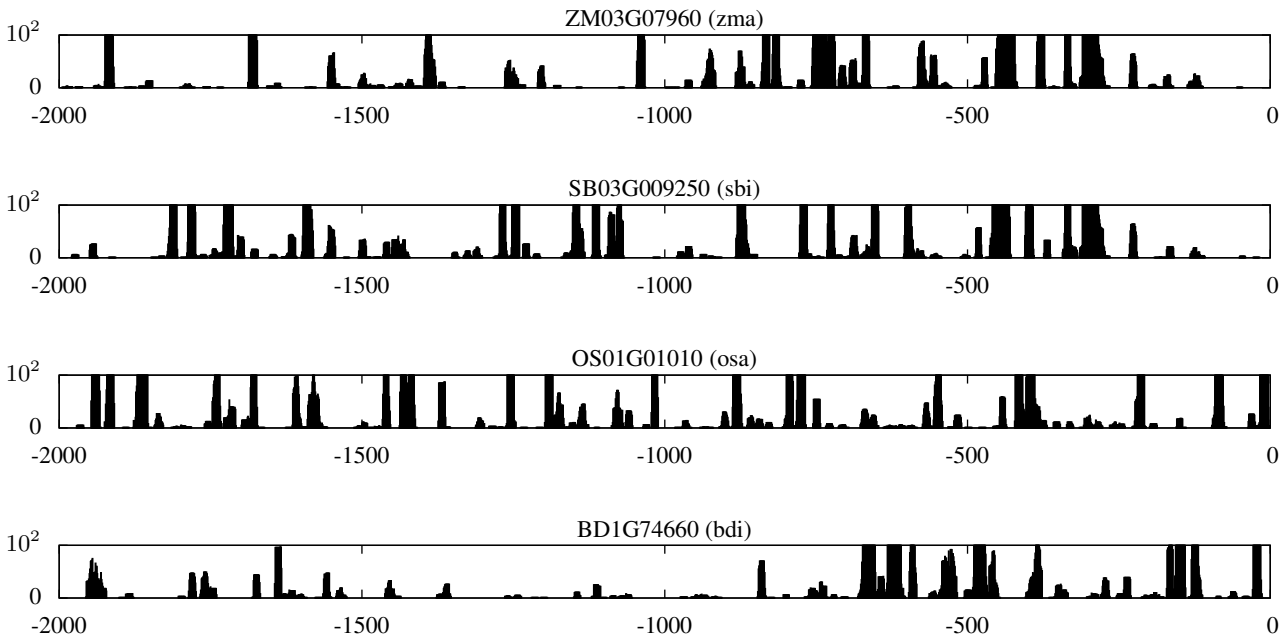

Fig. 8. Conserved regions in the promoters of the genes in gene family iORTHO00001 corresponding to motif instances with BLS $\geq$ 15%, $F \geq 20$ and $C \geq 0.9$, i.e., high-scoring motifs that are conserved in at least two species. The height of the bars corresponds to the number of distinct motif variants that map to that location. Note that the y-axis has been truncated at 100: certain loci in this gene family are covered with up to 18 418 distinct motif variants.

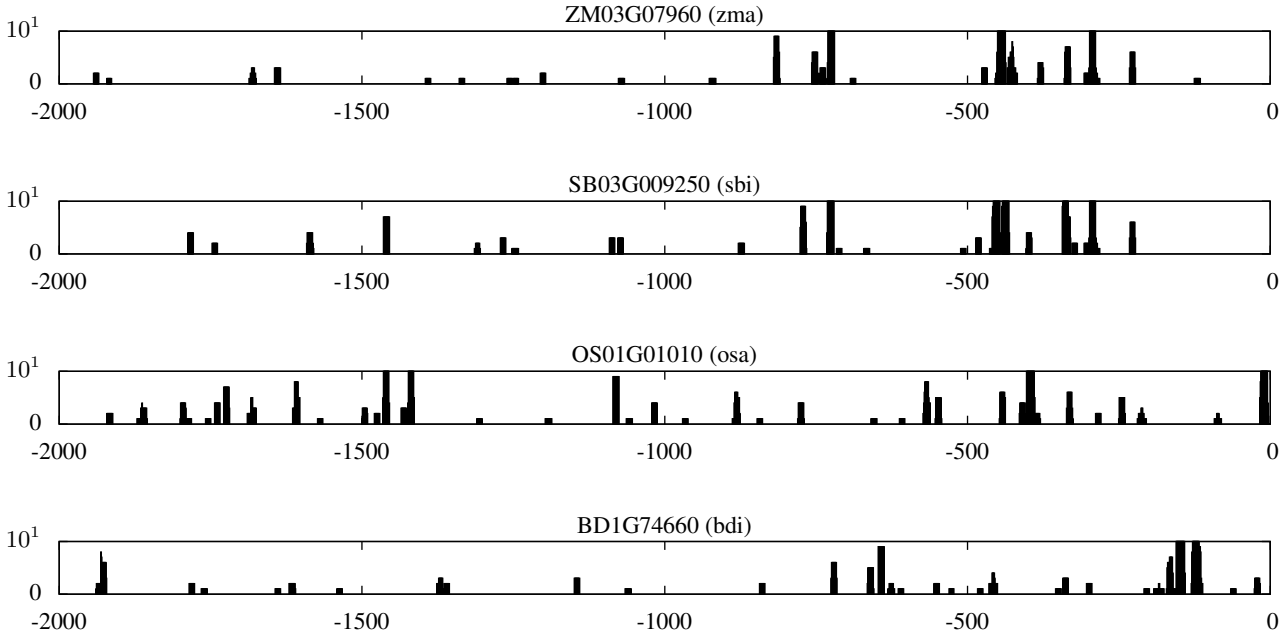

Fig. 9. Conserved regions in the promoters of the genes in gene family iORTHO00001 corresponding to motifs instances with BLS $\geq 95\%$, $F(95\%) \geq 50$ and $C(95\%) \geq 0.9$, i.e., motifs conserved in all four species. The height of the bars corresponds to the number of distinct motif variants that map to that location. Note that the y-axis has been truncated at 10: certain loci in this gene family are covered with up to 568 distinct motif variants.

Table 3: List of 165 gene families in which the ga2ox1-like KN1 motif was found to be genome-wide conserved by BLSSpeller (alignment-free discovery). For every gene family, genes that contain at least one ga2ox1-like KN1 instance are shown, along with the position(s) and strand(s) of the instance(s) relative to the translation start site. The symbols † and ‡ denote occurrences that are aligned in the multiple sequence alignment of the promoter sequences of the gene family. Most instances are not aligned.

ID	Gene (family) name	Position(s) and strand(s) of the occurrence(s) relative to the TSS
1	**iORTHO000066**	
	BD2G00740	-106 (-)
	SB03G008560	-706 (-)
2	**iORTHO000132**	
	BD2G01387	-575 (-)
	OS01G02920	-553 (-)
3	**iORTHO000360**	
	OS01G05810	-614 (-)
	ZM08G03630	-1871 (+)
4	**iORTHO000361**	
	OS01G05810	-614 (-)
	ZM08G03630	-1871 (+)
5	**iORTHO000738**	
	BD2G06240	-904 (+)
	OS01G10440	-340 (+)
6	**iORTHO000769**	
	BD3G09030	-205 (+)
	ZM04G37150	-330 (+)
7	**iORTHO001141**	
	BD2G09300	-645 (-)
	SB03G009900	-373 (-)
	ZM03G08480	-1414 (-)
8	**iORTHO001263**	
	OS01G16950	-468 (+), -464 (+)
	SB03G011120	-226 (-)
	ZM03G09610	-503 (+), -958 (-)
9	**iORTHO001881**	
	BD2G03437	-218 (-), -214 (-)
	OS01G28474	-793 (+), -789 (+), -785 (+), -296 (+), -260 (+), -256 (+)
10	**iORTHO002372**	
	OS01G38610	-169 (-)
	SB03G025570	-1037 (-)
11	**iORTHO002426**	
	SB03G025860	-39 (-)
	ZM08G22710	-354 (+)
12	**iORTHO002539**	
	SB03G026570	-1894 (-)
	ZM03G38660	-75 (-)
13	**iORTHO002680**	
	BD2G43380	-353 (-)
	SB03G027770	-365 (+)
14	**iORTHO003303**	
	BD2G47747	-825 (+)
	OS01G51010	-1186 (+)
15	**iORTHO003312**	
	OS01G51140	-1195 (+), -1185 (-)†
	SB03G032520	-246 (-)‡
	ZM03G33650	-913 (-)†, -269 (-)‡
16	**iORTHO003358**	
	SB03G032780	-907 (+)†
	ZM08G26610	-531 (+)†
17	**iORTHO004635**	
	SB03G042700	-90 (-)†
	ZM03G24140	-97 (-)†
18	**iORTHO004906**	

	OS01G70790	-241 (-)
	SB03G045000	-124 (-)
19	**iORTHO004971**	
	BD2G60497	-1193 (+)
	ZM03G21270	-1202 (-)
20	**iORTHO005073**	
	BD2G61340	-1824 (-)
	ZM10G18990	-333 (+), -211 (+)
21	**iORTHO005089**	
	OS01G72990	-1113 (+)
	ZM03G28480	-1443 (-)
22	**iORTHO005141**	
	SB03G047060	-803 (-)†
	ZM03G20160	-164 (-)†
23	**iORTHO005254**	
	OS02G01280	-721 (+)
	ZM04G41680	-139 (-)
24	**iORTHO005621**	
	OS02G05744	-1568 (-)
	SB04G003690	-60 (+)†
	ZM05G19730	-62 (+)†
25	**iORTHO005696**	
	OS06G46770	-89 (+)
	SB10G026870	-603 (+)
26	**iORTHO005726**	
	OS02G07030	-631 (+)
	SB03G045430	-1888 (+), -1884 (+), -1321 (+)
27	**iORTHO006041**	
	SB04G007190	-814 (-)†
	ZM05G27890	-781 (-)†
28	**iORTHO006250**	
	BD3G09030	-205 (+)
	ZM04G37150	-330 (+)
29	**iORTHO006414**	
	SB04G009820	-1859 (-)
	ZM04G36170	-1018 (+)
30	**iORTHO007022**	
	BD1G74760	-893 (-)
	OS02G27200	-464 (+)
31	**iORTHO007681**	
	OS02G36990	-1474 (+)
	SB04G024060	-745 (+)
32	**iORTHO008364**	
	OS02G45920	-1494 (-)
	SB04G031640	-772 (+)
33	**iORTHO008372**	
	SB04G031570	-503 (+)†
	ZM04G21560	-282 (+)†
34	**iORTHO008383**	
	SB04G031500	-1352 (-)†
	ZM05G36890	-1362 (-)†
35	**iORTHO008518**	
	OS02G47810	-238 (-)
	ZM05G38060	-497 (-)
36	**iORTHO008527**	
	BD3G52950	-1136 (+)
	OS02G47900	-107 (-)
37	**iORTHO008608**	
	BD1G52280	-440 (+)
	OS02G48790	-637 (-)
38	**iORTHO008936**	
	BD4G32407	-449 (-)
	OS02G52630	-842 (+)

39	**iORTHO009368**	
	BD3G56040	-900 (+)
	SB04G037760	-360 (-)
40	**iORTHO009560**	
	BD1G77840	-331 (-)
	SB01G049810	-1729 (+)
41	**iORTHO009594**	
	BD1G72410	-1963 (-)
	ZM01G00980	-1629 (-)
42	**iORTHO009853**	
	BD1G75050	-27 (+)
	OS03G05570	-979 (-)
43	**iORTHO010079**	
	OS03G08310	-552 (+), -548 (+)
	ZM01G05520	-103 (-)
44	**iORTHO010136**	
	SB01G044640	-99 (-)†
	ZM01G06190	-75 (-)†
45	**iORTHO010311**	
	SB01G043160	-213 (+)†
	ZM01G07750	-219 (+)†
46	**iORTHO010351**	
	BD1G69747	-667 (-)
	SB01G042730	-1495 (+)
47	**iORTHO010421**	
	BD1G69000	-1597 (-)
	ZM01G08940	-1069 (+)
48	**iORTHO010708**	
	BD1G66990	-1510 (+), -1148 (+)
	SB03G006370	-1979 (+)
49	**iORTHO010741**	
	SB01G039690	-786 (-)†
	ZM01G11210	-827 (-)†
50	**iORTHO010813**	
	OS03G17470	-1447 (+)
	ZM01G11850	-55 (-)
	ZM09G23450	-218 (-)
51	**iORTHO010818**	
	OS03G17520	-33 (-), -29 (-), -25 (-)
	SB01G038850	-217 (-), -213 (-)
52	**iORTHO011067**	
	BD1G64030	-1360 (-)
	SB01G037020	-1054 (+)†
	ZM01G13830	-1109 (+)†
53	**iORTHO011118**	
	SB01G036630	-630 (+)†
	ZM01G14120	-1155 (+)†
	ZM09G21840	-612 (+)†
54	**iORTHO011318**	
	SB01G034990	-1342 (-)
	ZM01G15850	-589 (+)
55	**iORTHO011576**	
	SB01G033320	-742 (-)†
	ZM01G17480	-657 (-)
	ZM09G19220	-611 (-)†
56	**iORTHO012330**	
	SB01G014100	-855 (-)†
	ZM01G46920	-564 (-)†
57	**iORTHO012402**	
	SB01G013520	-146 (-)
	ZM01G46450	-713 (-)
58	**iORTHO013064**	
	OS03G53110	-106 (-)

	SB01G008500	-1756 (-)
	ZM01G52330	-1663 (+)
59	**iORTHO013195**	
	BD1G07960	-452 (-)
	OS03G55180	-157 (+)
60	**iORTHO013295**	
	SB01G006690	-987 (-)
	ZM01G54560	-188 (+)
61	**iORTHO013500**	
	OS03G58590	-793 (+)
	SB01G004930	-1725 (+)
62	**iORTHO013743**	
	BD1G02790	-816 (+)
	SB01G002510	-1010 (+)†, -1006 (+)‡, -998 (+)
	ZM01G58590	-775 (+)†, -771 (+)‡
63	**iORTHO013780**	
	SB01G001980	-393 (-)
	ZM01G58990	-1176 (-), -1172 (-)
64	**iORTHO013805**	
	BD1G02160	-863 (-)
	OS03G62500	-1391 (+)
65	**iORTHO013930**	
	SB01G000450	-642 (-)
	ZM05G00190	-1426 (+)
66	**iORTHO013938**	
	OS03G64230	-837 (-)
	SB01G000365	-1181 (-)
	ZM05G00130	-254 (+)
67	**iORTHO013975**	
	BD5G02290	-1971 (-)
	OS04G01490	-1246 (-)
68	**iORTHO014236**	
	BD5G00530	-153 (-)
	SB06G001510	-175 (-)†
	ZM10G12690	-189 (-)†
69	**iORTHO014919**	
	BD4G01740	-976 (+)
	SB01G004110	-838 (+)
70	**iORTHO015326**	
	SB06G014280	-1120 (-)
	ZM02G17290	-1117 (-)
71	**iORTHO015334**	
	SB06G014360	-554 (+)†
	ZM02G17160	-552 (+)†, -468 (+)
72	**iORTHO015747**	
	SB06G018800	-122 (-), -118 (-)
	ZM02G13330	-199 (+), -130 (-)
73	**iORTHO016098**	
	BD5G15070	-1386 (+), -1507 (-), -1503 (-)†, -1499 (-)
	OS04G42700	-1695 (-)†
74	**iORTHO016109**	
	SB06G021990	-99 (-)†
	ZM02G10550	-77 (-)†
75	**iORTHO016228**	
	BD5G16257	-1616 (-)
	ZM10G20640	-191 (-)
76	**iORTHO016303**	
	BD3G50050	-371 (-)
	SB06G023770	-409 (-)†
	ZM10G20990	-379 (-)†
77	**iORTHO016397**	
	OS04G46560	-1187 (+)
	ZM02G07720	-20 (-)

78	**iORTHO016543**	
	OS04G48416	-36 (-)
	ZM10G21960	-33 (-)
79	**iORTHO016716**	
	BD5G20420	-236 (-)
	SB06G027382	-1994 (+)
80	**iORTHO016813**	
	SB06G028270	-1174 (+)†
	ZM02G04710	-1168 (+)†
81	**iORTHO017121**	
	BD5G24650	-502 (-)
	SB06G031300	-1277 (+)†
	ZM02G01880	-1088 (+)†
82	**iORTHO017169**	
	SB06G031840	-54 (+), -50 (+)
	ZM10G25780	-54 (+)
83	**iORTHO017267**	
	SB06G032750	-325 (+)
	ZM02G00630	-439 (+)
84	**iORTHO017349**	
	SB06G033540	-1374 (-)
	ZM10G26820	-46 (-)
85	**iORTHO017366**	
	SB06G033740	-506 (-)†
	ZM10G26970	-539 (-)†
86	**iORTHO017390**	
	SB06G033920	-1649 (-)
	ZM10G27030	-445 (+), -441 (+), -437 (+)
87	**iORTHO017393**	
	BD5G27390	-440 (+)
	SB06G033960	-1052 (-)
88	**iORTHO017512**	
	OS05G02200	-12 (-)
	ZM06G04960	-1864 (-)
89	**iORTHO017570**	
	BD2G37630	-791 (+)
	OS05G02870	-64 (-)
90	**iORTHO017863**	
	BD2G30490	-1595 (-)
	ZM02G22780	-1514 (-)
91	**iORTHO018425**	
	OS05G16300	-399 (-)†
	SB09G010410	-438 (-)†
	ZM06G21160	-434 (-)†
92	**iORTHO018592**	
	BD2G27080	-1869 (-)
	OS05G23260	-182 (-)
93	**iORTHO018619**	
	SB01G048660	-30 (+)†
	ZM01G01930	-30 (+)†
94	**iORTHO019311**	
	BD2G24880	-1290 (-)
	OS05G35060	-394 (+)
95	**iORTHO019826**	
	BD2G20710	-1872 (-)†
	OS05G42190	-1218 (-)
	ZM08G18870	-1872 (-)†
96	**iORTHO019970**	
	OS05G44560	-1889 (-)
	ZM08G19500	-427 (+)
97	**iORTHO020108**	
	SB03G032520	-246 (-)†

	ZM03G33650	-913 (-), -269 (-)†
98	**iORTHO020121**	
	OS05G46510	-897 (+)
	ZM06G30000	-135 (-), -131 (-)
99	**iORTHO020492**	
	BD2G14420	-1754 (-)
	SB09G030600	-619 (-), -615 (-), -611 (-)
100	**iORTHO020607**	
	BD1G50230	-697 (-)
	SB10G001060	-411 (+)
101	**iORTHO020725**	
	BD1G72410	-1963 (-)
	SB10G001800	-1515 (-)
102	**iORTHO021593**	
	BD1G44170	-14 (+)
	OS06G14420	-1880 (+)
103	**iORTHO022984**	
	BD1G35970	-837 (-)
	OS06G41090	-669 (-)
104	**iORTHO023338**	
	OS06G45940	-1034 (-)
	SB10G026960	-25 (-)
105	**iORTHO023387**	
	SB10G027230	-1653 (+), -1649 (+)†
	ZM05G12560	-1616 (+)†
106	**iORTHO024148**	
	SB01G013520	-146 (-)
	ZM01G46450	-713 (-)
107	**iORTHO024686**	
	BD1G03820	-1708 (-)
	SB09G022670	-1109 (-)
108	**iORTHO025078**	
	BD1G52500	-1526 (-)
	ZM07G07090	-1485 (-)
109	**iORTHO025230**	
	OS07G26930	-440 (+)
	SB02G010640	-1313 (-)
110	**iORTHO025443**	
	SB02G033690	-1405 (+)
	ZM07G22120	-1227 (-)
111	**iORTHO025869**	
	OS07G37400	-692 (+)
	SB02G036360	-255 (+)
112	**iORTHO026060**	
	OS07G39810	-806 (+)
	SB02G037870	-1353 (+)
113	**iORTHO026103**	
	BD1G22810	-849 (-)
	OS07G40310	-96 (+)
114	**iORTHO026164**	
	OS07G41200	-1596 (-)
	ZM02G38700	-660 (-)
	ZM07G26600	-614 (+)
115	**iORTHO026604**	
	SB02G042280	-361 (+)†, -357 (+)‡, -353 (+)
	ZM07G29690	-332 (+)†, -328 (+)‡
116	**iORTHO026797**	
	OS08G01030	-24 (-)
	SB01G006170	-565 (+)
117	**iORTHO026829**	
	OS08G01390	-140 (+), -1381 (-)
	ZM04G09100	-1748 (+)
118	**iORTHO026863**	

	SB07G000990	-201 (-)†
	ZM04G08930	-163 (-)†
119	**iORTHO026935**	
	OS08G02640	-787 (+)
	SB02G023310	-1320 (+)
	ZM06G01330	-913 (+)
120	**iORTHO027453**	
	OS08G10020	-68 (-)†
	SB07G005660	-61 (-)†
121	**iORTHO027639**	
	OS08G14320	-1618 (+)
	ZM10G08940	-1238 (-)
122	**iORTHO027703**	
	BD3G19077	-42 (+)†
	OS08G15230	-30 (+)†
	SB07G008090	-30 (+)†
123	**iORTHO028360**	
	OS08G29510	-1665 (-)
	ZM04G12900	-271 (+)
124	**iORTHO028732**	
	BD4G30980	-536 (+)
	SB07G022100	-706 (+), -702 (+)
125	**iORTHO028993**	
	BD3G39070	-846 (+), -842 (+)†, -838 (+)‡
	OS08G38320	-632 (+), -628 (+)
	ZM07G17310	-1052 (+)†, -1048 (+)‡, -1044 (+)
126	**iORTHO029037**	
	SB02G027880	-615 (+)
	SB07G028560	-1088 (-)
	ZM01G31480	-680 (+)
127	**iORTHO029233**	
	OS08G41390	-229 (-)
	SB10G012970	-1022 (-)
128	**iORTHO029383**	
	BD3G42050	-1988 (+)
	SB07G025200	-1944 (+)
129	**iORTHO029455**	
	BD3G42610	-311 (+), -307 (+), -303 (+)
	OS08G44020	-247 (-)
130	**iORTHO029457**	
	BD3G42610	-311 (+), -307 (+), -303 (+)
	OS08G44020	-247 (-)
131	**iORTHO029500**	
	SB07G024020	-626 (-), -622 (-)†, -618 (-)‡, -614 (-), -610 (-), -606 (-), -602 (-)
	ZM01G36160	-582 (-)†, -578 (-)‡
132	**iORTHO029602**	
	BD4G08340	-1476 (-)
	OS09G02270	-153 (-), -149 (-), -145 (-)
133	**iORTHO030048**	
	SB02G020450	-1440 (+)
	ZM07G10780	-1341 (+), -1337 (+), -1333 (+)
134	**iORTHO030080**	
	OS09G13570	-1822 (+)
	OS09G13575	-1822 (+)
	ZM07G10870	-1213 (-), -1209 (-)
135	**iORTHO030153**	
	SB01G028610	-926 (-)†
	ZM09G17660	-812 (-)†
136	**iORTHO030510**	
	OS09G21450	-140 (-)
	ZM07G13170	-1276 (+), -1272 (+)
137	**iORTHO030779**	
	BD4G30840	-975 (-)†

	OS09G26144	-899 (-)†
138	**iORTHO030959**	
	OS09G28400	-1182 (-)
	SB02G026610	-1797 (+)
139	**iORTHO031268**	
	BD4G34490	-744 (-)
	OS09G32740	-770 (-)
140	**iORTHO031746**	
	OS09G39560	-100 (+)
	ZM02G27480	-260 (+)
141	**iORTHO031758**	
	BD4G38730	-34 (+)
	OS09G39670	-420 (+)
142	**iORTHO033229**	
	SB01G020080	-69 (+)†
	ZM01G43160	-60 (+)†
143	**iORTHO033481**	
	SB01G018180	-188 (+)†
	ZM05G09420	-447 (+)†
144	**iORTHO033548**	
	SB01G017560	-215 (-)†, -211 (-)‡
	ZM01G40560	-265 (-)†, -261 (-)‡
145	**iORTHO033592**	
	OS10G37180	-1566 (-)
	SB08G005210	-659 (+)
146	**iORTHO033714**	
	BD3G32010	-1328 (+), -1324 (+), -1320 (+), -1316 (+)
	OS10G38970	-752 (-)
147	**iORTHO034138**	
	BD4G44470	-26 (+)
	SB08G002750	-61 (+)
	ZM10G01470	-1680 (-)
148	**iORTHO034147**	
	BD4G44427	-387 (+)
	SB05G001070	-973 (-)
149	**iORTHO034400**	
	BD4G42600	-966 (-)
	ZM04G29680	-1595 (+)
150	**iORTHO035652**	
	OS11G30500	-851 (-)
	SB05G017940	-1006 (+), -82 (-)†
	ZM02G42380	-105 (-)†
151	**iORTHO035884**	
	BD4G16650	-1222 (-), -396 (-)
	SB05G021000	-198 (-)†
	ZM04G30220	-316 (-)†
152	**iORTHO036204**	
	SB05G024160	-296 (-)†
	ZM04G02430	-263 (-)†
153	**iORTHO036234**	
	BD4G13670	-1831 (+)
	OS11G39990	-337 (+)
154	**iORTHO036235**	
	BD4G13670	-1831 (+)
	OS11G39990	-337 (+)
155	**iORTHO036746**	
	BD4G44470	-26 (+)
	SB08G002750	-61 (+)
	ZM10G01470	-1680 (-)
156	**iORTHO036752**	
	BD4G44427	-387 (+)
	SB05G001070	-973 (-)
157	**iORTHO036991**	

	BD2G56000	-794 (-)
	ZM06G05540	-1605 (+)
158	**iORTHO037236**	
	BD4G40350	-268 (-)†
	ZM03G18620	-272 (-)†
159	**iORTHO038101**	
	SB08G015550	-1176 (+)†
	ZM10G04160	-623 (+)†, -943 (-)
160	**iORTHO038159**	
	SB08G016450	-1995 (-)
	ZM03G18130	-129 (-)
161	**iORTHO038542**	
	SB08G019150	-1522 (-)†, -196 (-)
	ZM03G16680	-1426 (-)†, -237 (-)
162	**iORTHO038658**	
	BD4G02740	-321 (+)
	OS12G40510	-207 (+)†, -203 (+)‡
	SB08G020190	-148 (+), -144 (+)†, -140 (+)‡
163	**iORTHO038838**	
	BD4G01250	-1790 (-)
	ZM03G13680	-184 (-)
164	**iORTHO038888**	
	BD4G00900	-71 (-), -67 (-)
	OS12G43640	-1698 (+), -1666 (+)
	SB08G022780	-165 (-)
165	**iORTHO038907**	
	BD4G00775	-144 (+), -140 (+)
	OS12G43880	-406 (+)