Sophie Schbath

List of Publications by Year
in descending order

42 papers	1,640 citations	430874 18 h-index	315739 38 g-index
45 all docs	45 docs citations	45 times ranked	1605 citing authors

#	Article	IF	Citations
1	The MatP/matS Site-Specific System Organizes the Terminus Region of the E. coli Chromosome into a Macrodomain. Cell, 2008, 135, 475-485.	28.9	252
2	Probabilistic and Statistical Properties of Words: An Overview. Journal of Computational Biology, 2000, 7, 1-46.	1.6	207
3	Adaptive estimation for Hawkes processes; application to genome analysis. Annals of Statistics, 2010, 38, .	2.6	144
4	Temperate Phages Acquire DNA from Defective Prophages by Relaxed Homologous Recombination: The Role of Rad52-Like Recombinases. PLoS Genetics, 2014, 10, e1004181.	3.5	105
5	Multiple comparative metagenomics using multiset k-mer counting. PeerJ Computer Science, 0, 2, e94.	4.5	88
6	Compound Poisson and Poisson Process Approximations for Occurrences of Multiple Words in Markov Chains. Journal of Computational Biology, 1998, 5, 223-253.	1.6	81
7	Exceptional Motifs in Different Markov Chain Models for a Statistical Analysis of DNA Sequences. Journal of Computational Biology, 1995, 2, 417-437.	1.6	79
8	Characteristics of Chi distribution on different bacterial genomes. Research in Microbiology, 1999, 150, 579-587.	2.1	79
9	Mapping Reads on a Genomic Sequence: An Algorithmic Overview and a Practical Comparative Analysis. Journal of Computational Biology, 2012, 19, 796-813.	1.6	73
10	Identification of DNA Motifs Implicated in Maintenance of Bacterial Core Genomes by Predictive Modeling. PLoS Genetics, 2007, 3, e153.	3.5	63
11	An Efficient Statistic to Detect Over- and Under-represented Words in DNA Sequences. Journal of Computational Biology, 1997, 4, 189-192.	1.6	58
12	Compound Poisson approximation of word counts in DNA sequences. ESAIM - Probability and Statistics, 1997, 1, 1-16.	0.5	58
13	DNA motifs that sculpt the bacterial chromosome. Nature Reviews Microbiology, 2011, 9, 15-26.	28.6	52
14	Numerical Comparison of Several Approximations of the Word Count Distribution in Random Sequences. Journal of Computational Biology, 2001, 8, 349-359.	1.6	42
15	FADO: A Statistical Method to Detect Favored or Avoided Distances between Occurrences of Motifs using the Hawkes' Model. Statistical Applications in Genetics and Molecular Biology, 2005, 4, Article24.	0.6	36
16	An Overview on the Distribution of Word Counts in Markov Chains. Journal of Computational Biology, 2000, 7, 193-201.	1.6	33
17	Statistical tests to compare motif count exceptionalities. BMC Bioinformatics, 2007, 8, 84.	2.6	28
18	Improved compound Poisson approximation for the number of occurrences of any rare word family in a stationary markov chain. Advances in Applied Probability, 2007, 39, 128-140.	0.7	22

#	Article	IF	Citations
19	SIGffRid: A tool to search for sigma factor binding sites in bacterial genomes using comparative approach and biologically driven statistics. BMC Bioinformatics, 2008, 9, 73.	2.6	19
20	Assessing the Exceptionality of Coloured Motifs in Networks. Eurasip Journal on Bioinformatics and Systems Biology, 2009, 2009, 616234.	1.4	15
21	How Evolution of Genomes Is Reflected in Exact DNA Sequence Match Statistics. Molecular Biology and Evolution, 2015, 32, 524-535.	8.9	13
22	Coverage Processes in Physical Mapping by Anchoring Random Clones. Journal of Computational Biology, 1997, 4, 61-82.	1.6	12
23	R'MES: A Tool to Find Motifs with a Significantly Unexpected Frequency in Biological Sequences. Science, Engineering, and Biology Informatics, 2011, , 25-64.	0.1	12
24	Assessing the Exceptionality of Coloured Motifs in Networks. Eurasip Journal on Bioinformatics and Systems Biology, 2009, 2009, 1-9.	1.4	11
25	Exploring short k-mer profiles in cells and mobile elements from Archaea highlights the major influence of both the ecological niche and evolutionary history. BMC Genomics, 2021, 22, 186.	2.8	8
26	Large compound Poisson approximations for occurrences of multiple words. Lecture Notes-monograph Series / Institute of Mathematical Statistics, 1999, 33, 257-275.	1.0	6
27	Improved compound Poisson approximation for the number of occurrences of any rare word family in a stationary markov chain. Advances in Applied Probability, 2007, 39, 128-140.	0.7	6
28	Separating Significant Matches from Spurious Matches in DNA Sequences. Journal of Computational Biology, 2012, 19, 1-12.	1.6	6
29	Robustness Assessment of Whole Bacterial Genome Segmentations. Journal of Computational Biology, 2011, 18, 1155-1165.	1.6	5
30	SimkaMin: fast and resource frugal *de novo* comparative metagenomics. Bioinformatics, 2020, 36, 1275-1276.	4.1	5
31	The Effect of Nonhomogeneous Clone Length Distribution on the Progress of an STS Mapping Project. Journal of Computational Biology, 2000, 7, 47-57.	1.6	4
32	Poisson Approximation for the Number of Repeats in a Stationary Markov Chain. Journal of Applied Probability, 2008, 45, 440-455.	0.7	3
33	Occurrence of structured motifs in random sequences: Arbitrary number of boxes. Discrete Applied Mathematics, 2011, 159, 826-831.	0.9	3
34	Comparing the Statistical Fate of Paralogous and Orthologous Sequences. Genetics, 2016, 204, 475-482.	2.9	3
35	Poisson Approximation for the Number of Repeats in a Stationary Markov Chain. Journal of Applied Probability, 2008, 45, 440-455.	0.7	2
36	Model€based biclustering for overdispersed count data with application in microbial ecology. Methods in Ecology and Evolution, 2021, 12, 1050-1061.	5.2	2

#	Article	IF	Citations
37	Statistical Significance of Threading Scores. Journal of Computational Biology, 2012, 19, 13-29.	1.6	1
38	DUGMO: tool for the detection of unknown genetically modified organisms with high-throughput sequencing data for pure bacterial samples. BMC Bioinformatics, 2020, 21, 284.	2.6	1
39	Assessing the Robustness of Complete Bacterial Genome Segmentations. Lecture Notes in Computer Science, 2010, , 173-187.	1.3	1
40	How Can Pattern Statistics Be Useful for DNA Motif Discovery?. , 2009, , 319-350.		1
41	Statistical modelling of bacterial promoter sequences for regulatory motif discovery with the help of transcriptome data: application to <i>Listeria monocytogenes</i>. Journal of the Royal Society Interface, 2020, 17, 20200600.	3.4	1
42	Using graph modularity analysis to identify transcription factor binding sites. , 2010, , .		0