

Tak-Wah Lam

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/9420297/publications.pdf>

Version: 2024-02-01

130
papers

18,755
citations

218592

26
h-index

22808

112
g-index

151
all docs

151
docs citations

151
times ranked

26964
citing authors

#	ARTICLE	IF	CITATIONS
1	MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct <i>de Bruijn</i> graph. <i>Bioinformatics</i> , 2015, 31, 1674-1676.	1.8	4,864
2	SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. <i>GigaScience</i> , 2012, 1, 18.	3.3	4,510
3	SOAP2: an improved ultrafast tool for short read alignment. <i>Bioinformatics</i> , 2009, 25, 1966-1967.	1.8	3,329
4	MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. <i>Methods</i> , 2016, 102, 3-11.	1.9	1,174
5	The sequence and de novo assembly of the giant panda genome. <i>Nature</i> , 2010, 463, 311-317.	13.7	1,058
6	SOAPdenovo-Trans: <i>de novo</i> transcriptome assembly with short RNA-Seq reads. <i>Bioinformatics</i> , 2014, 30, 1660-1666.	1.8	826
7	Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. <i>GigaScience</i> , 2013, 2, 10.	3.3	582
8	SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. <i>Bioinformatics</i> , 2012, 28, 878-879.	1.8	200
9	COPE: an accurate <i>k</i> -mer-based pair-end reads connection tool to facilitate genome assembly. <i>Bioinformatics</i> , 2012, 28, 2870-2874.	1.8	145
10	SOAP3-dp: Fast, Accurate and Sensitive GPU-Based Short Read Aligner. <i>PLoS ONE</i> , 2013, 8, e65632.	1.1	104
11	A multi-task convolutional deep neural network for variant calling in single molecule sequencing. <i>Nature Communications</i> , 2019, 10, 998.	5.8	102
12	SOAPsplice: Genome-Wide ab initio Detection of Splice Junctions from RNA-Seq Data. <i>Frontiers in Genetics</i> , 2011, 2, 46.	1.1	89
13	Exploring the limit of using a deep neural network on pileup data for germline variant calling. <i>Nature Machine Intelligence</i> , 2020, 2, 220-227.	8.3	87
14	Compressed indexes for dynamic text collections. <i>ACM Transactions on Algorithms</i> , 2007, 3, 21.	0.9	69
15	Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. <i>Scientific Reports</i> , 2011, 1, 55.	1.6	67
16	Scheduling for Speed Bounded Processors. <i>Lecture Notes in Computer Science</i> , 2008, , 409-420.	1.0	65
17	A Space and Time Efficient Algorithm for Constructing Compressed Suffix Arrays. <i>Algorithmica</i> , 2007, 48, 23-36.	1.0	59
18	An Even Faster and More Unifying Algorithm for Comparing Trees via Unbalanced Bipartite Matchings. <i>Journal of Algorithms</i> , 2001, 40, 212-233.	0.9	50

#	ARTICLE	IF	CITATIONS
19	Speed Scaling Functions for Flow Time Scheduling Based on Active Job Count. Lecture Notes in Computer Science, 2008, , 647-659.	1.0	49
20	Predicting RNA Secondary Structures with Arbitrary Pseudoknots by Maximizing the Number of Stacking Pairs. Journal of Computational Biology, 2003, 10, 981-995.	0.8	48
21	RENET: A Deep Learning Approach for Extracting Gene-Disease Associations from Literature. Lecture Notes in Computer Science, 2019, , 272-284.	1.0	45
22	A Decomposition Theorem for Maximum Weight Bipartite Matchings. SIAM Journal on Computing, 2001, 31, 18-26.	0.8	42
23	Finding Connected Components in $O(\log n \log \log n)$ Time on the EREW PRAM. Journal of Algorithms, 1995, 18, 378-402.	0.9	37
24	Optimizing throughput and energy in online deadline scheduling. ACM Transactions on Algorithms, 2009, 6, 1-22.	0.9	27
25	Deadline scheduling and power management for speed bounded processors. Theoretical Computer Science, 2010, 411, 3587-3600.	0.5	27
26	From Peer-Reviewed to Peer-Reproduced in Scholarly Publishing: The Complementary Roles of Data Models and Workflows in Bioinformatics. PLoS ONE, 2015, 10, e0127612.	1.1	27
27	Optimal Edge Ranking of Trees in Linear Time. Algorithmica, 2001, 30, 12-33.	1.0	25
28	Competitive non-migratory scheduling for flow time and energy. , 2008, , .		23
29	SOAPfusion: a robust and effective computational fusion discovery tool for RNA-seq reads. Bioinformatics, 2013, 29, 2971-2978.	1.8	23
30	GLProbs: Aligning Multiple Sequences Adaptively. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2015, 12, 67-78.	1.9	23
31	Compressed Indexes for Approximate String Matching. Algorithmica, 2010, 58, 263-281.	1.0	22
32	Succinct Text Indexing with Wildcards. Lecture Notes in Computer Science, 2009, , 39-50.	1.0	22
33	Compressed Index for Dictionary Matching. , 2008, , .		21
34	Structural Alignment of RNA with Complex Pseudoknot Structure. Journal of Computational Biology, 2011, 18, 97-108.	0.8	21
35	Continuous Monitoring of Distributed Data Streams over a Time-Based Sliding Window. Algorithmica, 2012, 62, 1088-1111.	1.0	21
36	RECONSTRUCTING AN ULTRAMETRIC CALLED PHYLOGENETIC NETWORK FROM A DISTANCE MATRIX. Journal of Bioinformatics and Computational Biology, 2006, 04, 807-832.	0.3	19

#	ARTICLE	IF	CITATIONS
37	Nonclairvoyant Speed Scaling for Flow and Energy. <i>Algorithmica</i> , 2011, 61, 507-517.	1.0	19
38	A Linear Size Index for Approximate Pattern Matching. <i>Lecture Notes in Computer Science</i> , 2006, , 49-59.	1.0	19
39	FaSD-somatic: a fast and accurate somatic SNV detection algorithm for cancer genome sequencing data. <i>Bioinformatics</i> , 2014, 30, 2498-2500.	1.8	18
40	Space Efficient Indexes for String Matching with Donâ€™t Cares. , 2007, , 846-857.		17
41	Approximate String Matching Using Compressed Suffix Arrays. <i>Lecture Notes in Computer Science</i> , 2004, , 434-444.	1.0	16
42	A Decomposition Theorem for MaximumWeight Bipartite Matchings with Applications to Evolutionary Trees. <i>Lecture Notes in Computer Science</i> , 1999, , 438-449.	1.0	16
43	Non-clairvoyant Speed Scaling for Weighted Flow Time. <i>Lecture Notes in Computer Science</i> , 2010, , 23-35.	1.0	16
44	BALSA: integrated secondary analysis for whole-genome and whole-exome sequencing, accelerated by GPU. <i>PeerJ</i> , 2014, 2, e421.	0.9	16
45	Potential utility of metagenomic sequencing for improving etiologic diagnosis of infective endocarditis. <i>Future Cardiology</i> , 2019, 15, 411-424.	0.5	15
46	A linear size index for approximate pattern matching. <i>Journal of Discrete Algorithms</i> , 2011, 9, 358-364.	0.7	14
47	MICA: A fast short-read aligner that takes full advantage of Many Integrated Core Architecture (MIC). <i>BMC Bioinformatics</i> , 2015, 16, S10.	1.2	14
48	Performance guarantee for edf under overload. <i>Journal of Algorithms</i> , 2004, 52, 193-206.	0.9	13
49	Improved Approximate String Matching Using Compressed Suffix Data Structures. <i>Algorithmica</i> , 2008, 51, 298-314.	1.0	13
50	Serine peptidase inhibitor Kazal type 1 (SPINK1) as novel downstream effector of the cadherin-17/12-catenin axis in hepatocellular carcinoma. <i>Cellular Oncology (Dordrecht)</i> , 2017, 40, 443-456.	2.1	13
51	Building a Chinese pan-genome of 486 individuals. <i>Communications Biology</i> , 2021, 4, 1016.	2.0	13
52	Improved multi-processor scheduling for flow time and energy. <i>Journal of Scheduling</i> , 2012, 15, 105-116.	1.3	12
53	Succinct Index for Dynamic Dictionary Matching. <i>Lecture Notes in Computer Science</i> , 2009, , 1034-1043.	1.0	12
54	Results on communication complexity classes. <i>Journal of Computer and System Sciences</i> , 1992, 44, 324-342.	0.9	11

#	ARTICLE	IF	CITATIONS
55	General techniques for comparing unrooted evolutionary trees. , 1997, , .		11
56	Nonmigratory Online Deadline Scheduling on Multiprocessors. SIAM Journal on Computing, 2005, 34, 669-682.	0.8	11
57	MegaGTA: a sensitive and accurate metagenomic gene-targeted assembler using iterative de Bruijn graphs. BMC Bioinformatics, 2017, 18, 408.	1.2	11
58	Improved Approximate String Matching Using Compressed Suffix Data Structures. Lecture Notes in Computer Science, 2005, , 339-348.	1.0	10
59	Cavity Matchings, Label Compressions, and Unrooted Evolutionary Trees. SIAM Journal on Computing, 2000, 30, 602-624.	0.8	9
60	Extra processors versus future information in optimal deadline scheduling. , 2002, , .		9
61	Improving the efficiency of parallel minimum spanning tree algorithms. Discrete Applied Mathematics, 2003, 126, 33-54.	0.5	9
62	database.bio: a web application for interpreting human variations. Bioinformatics, 2015, 31, 4035-4037.	1.8	9
63	Trade-offs between communication and space. Journal of Computer and System Sciences, 1992, 45, 296-315.	0.9	8
64	Automatic construction of online catalog topologies. IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, 2002, 32, 382-391.	3.3	8
65	AC-DIAMOND v1: accelerating large-scale DNAâ€“protein alignment. Bioinformatics, 2018, 34, 3744-3746.	1.8	8
66	Identification of Cooperative Gene Regulation Among Transcription Factors, LncRNAs, and MicroRNAs in Diabetic Nephropathy Progression. Frontiers in Genetics, 2020, 11, 1008.	1.1	8
67	Predicting RNA secondary structures with arbitrary pseudoknots by maximizing the number of stacking pairs. , 2001, , .		7
68	On-line stream merging in a general setting. Theoretical Computer Science, 2003, 296, 27-46.	0.5	7
69	Nonmigratory Multiprocessor Scheduling for Response Time and Energy. IEEE Transactions on Parallel and Distributed Systems, 2008, 19, 1527-1539.	4.0	7
70	Cache-oblivious index for approximate string matching. Theoretical Computer Science, 2011, 412, 3579-3588.	0.5	7
71	ECNano: A cost-effective workflow for target enrichment sequencing and accurate variant calling on 4800 clinically significant genes using a single MinION flowcell. BMC Medical Genomics, 2022, 15, 43.	0.7	7
72	Online Speed Scaling Based on Active Job Count to Minimize Flow Plus Energy. Algorithmica, 2013, 65, 605-633.	1.0	6

#	ARTICLE	IF	CITATIONS
73	Compressing Dictionary Matching Index via Sparsification Technique. <i>Algorithmica</i> , 2015, 72, 515-538.	1.0	6
74	BASE: a practical de novo assembler for large genomes using long NGS reads. <i>BMC Genomics</i> , 2016, 17, 499.	1.2	6
75	Dictionary Matching with a Bounded Gap in Pattern or in Text. <i>Algorithmica</i> , 2018, 80, 698-713.	1.0	6
76	MegaPath: sensitive and rapid pathogen detection using metagenomic NGS data. <i>BMC Genomics</i> , 2020, 21, 500.	1.2	6
77	Non-shared edges and nearest neighbor interchanges revisited. <i>Information Processing Letters</i> , 2004, 91, 129-134.	0.4	5
78	Sleep Management on Multiple Machines for Energy and Flow Time. <i>Lecture Notes in Computer Science</i> , 2011, , 219-231.	1.0	5
79	RNASAlign: RNA Structural Alignment System. <i>Bioinformatics</i> , 2011, 27, 2151-2152.	1.8	5
80	Scheduling for weighted flow time and energy with rejection penalty. <i>Theoretical Computer Science</i> , 2013, 470, 93-104.	0.5	5
81	Improving multiple sequence alignment by using better guide trees. <i>BMC Bioinformatics</i> , 2015, 16, S4.	1.2	5
82	Transcriptome Analysis of Acute Phase Liver Graft Injury in Liver Transplantation. <i>Biomedicines</i> , 2018, 6, 41.	1.4	5
83	CONNET: Accurate Genome Consensus in Assembling Nanopore Sequencing Data via Deep Learning. <i>IScience</i> , 2020, 23, 101128.	1.9	5
84	Detecting structural variations with precise breakpoints using low-depth WGS data from a single oxford nanopore MinION flowcell. <i>Scientific Reports</i> , 2022, 12, 4519.	1.6	5
85	APPROXIMATING THE NEAREST NEIGHBOR INTERCHARGE DISTANCE FOR NON-UNIFORM-DEGREE EVOLUTIONARY TREES. <i>International Journal of Foundations of Computer Science</i> , 2001, 12, 533-550.	0.8	4
86	SUBTREE TRANSFER DISTANCE FOR DEGREE-D PHYLOGENIES. <i>International Journal of Foundations of Computer Science</i> , 2004, 15, 893-909.	0.8	4
87	Extra Processors versus Future Information in Optimal Deadline Scheduling. <i>Theory of Computing Systems</i> , 2004, 37, 323-341.	0.7	4
88	FINDING ALTERNATIVE SPLICING PATTERNS WITH STRONG SUPPORT FROM EXPRESSED SEQUENCES ON INDIVIDUAL EXONS/INTRONS. <i>Journal of Bioinformatics and Computational Biology</i> , 2008, 06, 1021-1033.	0.3	4
89	Nonclairvoyant sleep management and flow-time scheduling on multiple processors. , 2013, , .		4
90	Scheduling with Gaps: New Models and Algorithms. <i>Lecture Notes in Computer Science</i> , 2015, , 114-126.	1.0	4

#	ARTICLE	IF	CITATIONS
91	RENET2: high-performance full-text gene-disease relation extraction with iterative training data expansion. <i>NAR Genomics and Bioinformatics</i> , 2021, 3, lqab062.	1.5	4
92	Construction of online catalog topologies using decision trees. , 0, , .		3
93	Approximating Frequent Items in Asynchronous Data Stream over a Sliding Window. <i>Algorithms</i> , 2011, 4, 200-222.	1.2	3
94	PnpProbs: a better multiple sequence alignment tool by better handling of guide trees. <i>BMC Bioinformatics</i> , 2016, 17, 285.	1.2	3
95	A simple and economical method for improving whole genome alignment. <i>BMC Genomics</i> , 2017, 18, 362.	1.2	3
96	Compressed Indexes for Approximate String Matching. <i>Lecture Notes in Computer Science</i> , 2006, , 208-219.	1.0	3
97	Tradeoff between energy and throughput for online deadline scheduling. <i>Sustainable Computing: Informatics and Systems</i> , 2011, 1, 189-195.	1.6	2
98	Approximating Frequent Items in Asynchronous Data Stream over a Sliding Window. <i>Lecture Notes in Computer Science</i> , 2010, , 49-61.	1.0	2
99	A MEMORY EFFICIENT ALGORITHM FOR STRUCTURAL ALIGNMENT OF RNAs WITH EMBEDDED SIMPLE PSEUDOKNOTS. , 2007, , .		2
100	Adjacent Nucleotide Dependence in ncRNA and Order-1 SCFG for ncRNA Identification. <i>PLoS ONE</i> , 2010, 5, e12848.	1.1	2
101	Approximating the Nearest Neighbor Interchange Distance for Evolutionary Trees with Non-uniform Degrees. <i>Lecture Notes in Computer Science</i> , 1999, , 61-70.	1.0	2
102	HKG: an open genetic variant database of 205 Hong Kong cantonese exomes. <i>NAR Genomics and Bioinformatics</i> , 2022, 4, lqac005.	1.5	2
103	On the speed requirement for optimal deadline scheduling in overloaded systems. , 0, , .		1
104	A 5-competitive on-line scheduler for merging video streams. , 0, , .		1
105	Competitive Deadline Scheduling via Additional or Faster Processors. <i>Journal of Scheduling</i> , 2003, 6, 213-223.	1.3	1
106	On-line Stream Merging with Max Span and Min Coverage. <i>Theory of Computing Systems</i> , 2005, 38, 461-479.	0.7	1
107	A Tighter Extra-Resource Analysis of Online Deadline Scheduling. <i>Journal of Combinatorial Optimization</i> , 2005, 9, 157-165.	0.8	1
108	Extra Unit-Speed Machines Are Almost as Powerful as Speedy Machines for Flow Time Scheduling. <i>SIAM Journal on Computing</i> , 2008, 37, 1595-1612.	0.8	1

#	ARTICLE	IF	CITATIONS
109	Detection of novel tandem duplication with next-generation sequencing. , 2011, , .		1
110	Non-clairvoyant weighted flow time scheduling with rejection penalty. , 2012, , .		1
111	GLProbs. , 2013, , .		1
112	Non-clairvoyant Weighted Flow Time Scheduling on Different Multi-processor Models. Theory of Computing Systems, 2015, 56, 82-95.	0.7	1
113	AC-DIAMOND: Accelerating Protein Alignment via Better SIMD Parallelization and Space-Efficient Indexing. Lecture Notes in Computer Science, 2016, , 426-433.	1.0	1
114	Draft Genome Sequence of Helicobacter cinaedi, Compiled by Direct Whole-Genome Sequencing of a Blood Culture-Positive Isolate in Hong Kong. Microbiology Resource Announcements, 2018, 7, .	0.3	1
115	Scheduling with gaps: new models and algorithms. Journal of Scheduling, 2021, 24, 381-403.	1.3	1
116	Dictionary Matching with Uneven Gaps. Lecture Notes in Computer Science, 2015, , 247-260.	1.0	1
117	Energy-Efficient Due Date Scheduling. Lecture Notes in Computer Science, 2011, , 69-80.	1.0	1
118	Online Flow Time Scheduling in the Presence of Preemption Overhead. Lecture Notes in Computer Science, 2012, , 85-97.	1.0	1
119	Non-clairvoyant Weighted Flow Time Scheduling on Different Multi-processor Models. Lecture Notes in Computer Science, 2012, , 137-149.	1.0	1
120	Aggressive Online Deadline Scheduling. Electronic Notes in Theoretical Computer Science, 2004, 91, 148-157.	0.9	0
121	Compressed index for dynamic text. , 0, , .		0
122	Algorithms for pseudoknot classification. , 2011, , .		0
123	Efficient SNP-sensitive alignment and database-assisted SNP calling for low coverage samples. , 2012, , .		0
124	Guest Editors Foreword. Algorithmica, 2015, 73, 621-622.	1.0	0
125	A simple and economical method for improving whole-genome alignment. , 2015, , .		0
126	Unbalanced and Hierarchical Bipartite Matchings with Applications to Labeled Tree Comparison. Lecture Notes in Computer Science, 2000, , 479-490.	1.0	0

#	ARTICLE	IF	CITATIONS
127	Improved Phylogeny Comparisons: Non-shared Edges, Nearest Neighbor Interchanges, and Subtree Transfers. Lecture Notes in Computer Science, 2000, , 527-538.	1.0	0
128	Extra unit-speed machines are almost as powerful as speedy machines for competitive flow time scheduling. , 2006, , .		0
129	LCR_Finder: A de Novo Low Copy Repeat Finder for Human Genome. Lecture Notes in Computer Science, 2013, , 125-136.	1.0	0
130	Translocator. , 2020, , .		0