Steven L Salzberg

List of Publications by Year
in descending order

334 papers	284,052 citations	323 141 h-index	198 321 g-index
419 all docs	419 docs citations	419 times ranked	237622 citing authors

#	Article	IF	Citations
1	PhyloCSF++: a fast and user-friendly implementation of PhyloCSF with annotation tools. Bioinformatics, 2022, 38, 1440-1442.	1.8	8
2	Genome‐wide association identifies candidate genes for drought tolerance in coast redwood and giant sequoia. Plant Journal, 2022, 109, 7-22.	2.8	17
3	The SAMBA tool uses long reads to improve the contiguity of genome assemblies. PLoS Computational Biology, 2022, 18, e1009860.	1.5	22
4	Assembled and annotated 26.5 Gbp coast redwood genome: a resource for estimating evolutionary adaptive potential and investigating hexaploid origin. G3: Genes, Genomes, Genetics, 2022, 12, .	0.8	28
5	A reference-quality, fully annotated genome from a Puerto Rican individual. Genetics, 2022, 220, .	1.2	7
6	The complete sequence of a human genome. Science, 2022, 376, 44-53.	6.0	1,222
7	Next-generation sequencing: insights to advance clinical investigations of the microbiome. Journal of Clinical Investigation, 2022, 132, .	3.9	116
8	High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. Nature Communications, 2022, 13, 2047.	5.8	30
9	Effects of transcriptional noise on estimates of gene and transcript expression in RNA sequencing experiments. Genome Research, 2021, 31, 301-308.	2.4	13
10	Liftoff: accurate mapping of gene annotations. Bioinformatics, 2021, 37, 1639-1643.	1.8	322
11	Dissecting the Polygenic Basis of Cold Adaptation Using Genome-Wide Association of Traits and Environmental Data in Douglas-fir. Genes, 2021, 12, 110.	1.0	14
12	Balrog: A universal protein model for prokaryotic gene prediction. PLoS Computational Biology, 2021, 17, e1008727.	1.5	21
13	Guillain-Barré Syndrome Outbreak in Peru 2019 Associated With <i>Campylobacter jejuni</i> Infection. Neurology: Neuroimmunology and NeuroInflammation, 2021, 8, .	3.1	20
14	Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. Genetics, 2021, 218, .	1.2	58
15	Rapidly fatal infection with Bacillus cereus/thuringiensis: genome assembly of the responsible pathogen and consideration of possibly contributing toxins. Diagnostic Microbiology and Infectious Disease, 2021, 101, 115534.	0.8	1
16	Releasing the Kraken. Frontiers in Bioinformatics, 2021, 1, .	1.0	7
17	Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics, 2020, 36, 1303-1304.	1.8	255
18	Genomic basis of white pine blister rust quantitative disease resistance and its relationship with qualitative resistance. Plant Journal, 2020, 104, 365-376.	2.8	32

#	Article	IF	Citations
19	The tuatara genome reveals ancient features of amniote evolution. Nature, 2020, 584, 403-409.	13.7	105
20	A Reference Genome Sequence for Giant Sequoia. G3: Genes, Genomes, Genetics, 2020, 10, 3907-3919.	0.8	67
21	Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. Microbiome, 2020, 8, 124.	4.9	146
22	Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies. Genetics, 2020, 216, 599-608.	1.2	34
23	Microbial Diagnostics for Cancer: A Step Forward but Not Prime Time Yet. Cancer Cell, 2020, 37, 625-627.	7.7	1
24	Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. Genome Biology, 2020, 21, 115.	3.8	130
25	High-quality chromosome-scale assembly of the walnut (Juglans regia L.) reference genome. GigaScience, 2020, 9, .	3.3	83
26	Assembly and annotation of an Ashkenazi human reference genome. Genome Biology, 2020, 21, 129.	3.8	42
27	The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLoS Computational Biology, 2020, 16, e1007981.	1.5	172
28	Pan-genomics in the human genome era. Nature Reviews Genetics, 2020, 21, 243-254.	7.7	191
29	Genome assembly and characterization of a complex zfBED-NLR gene-containing disease resistance locus in Carolina Gold Select rice with Nanopore sequencing. PLoS Genetics, 2020, 16, e1008571.	1.5	112
30	SkewIT: The Skew Index Test for large-scale GC Skew analysis of bacterial genomes. PLoS Computational Biology, 2020, 16, e1008439.	1.5	40
31	The genome of the American groundhog, Marmota monax. F1000Research, 2020, 9, 1137.	0.8	3
32	The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. , 2020, 16, e1007981.		0
33	The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. , 2020, 16, e1007981.		0
34	The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. , 2020, 16, e1007981.		0
35	The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. , 2020, 16, e1007981.		0
36	The Terabase Search Engine: a large-scale relational database of short-read sequences. Bioinformatics, 2019, 35, 665-670.	1.8	8

#	Article	IF	Citations
37	Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology, 2019, 37, 907-915.	9.4	6,567
38	Whole Genome Sequencing Identifies CRISPLD2 as a Lung Function Gene in Children With Asthma. Chest, 2019, 156, 1068-1079.	0.4	5
39	Human contamination in bacterial genomes has created thousands of spurious proteins. Genome Research, 2019, 29, 954-960.	2.4	111
40	Next-generation genome annotation: we still struggle to get it right. Genome Biology, 2019, 20, 92.	3.8	216
41	Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. Journal of Allergy and Clinical Immunology, 2019, 143, AB296.	1.5	0
42	Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. Nature Communications, 2019, 10, 880.	5.8	71
43	Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biology, 2019, 20, 278.	3.8	897
44	Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nature Genetics, 2019, 51, 30-35.	9.4	276
45	Genomic architecture of complex traits in loblolly pine. New Phytologist, 2019, 221, 1789-1801.	3.5	60
46	A review of methods and databases for metagenomic classification and assembly. Briefings in Bioinformatics, 2019, 20, 1125-1136.	3.2	368
47	Cretaceous dinosaur bone contains recent organic material and provides an environment conducive to microbial communities. ELife, 2019, 8, .	2.8	38
48	Genomic characterization of chromosome translocations in patients with T/myeloid mixed-phenotype acute leukemia. Leukemia and Lymphoma, 2018, 59, 1231-1238.	0.6	8
49	KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biology, 2018, 19, 198.	3.8	283
50	CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. Genome Biology, 2018, 19, 208.	3.8	263
51	Genomic Variation Among and Within Six *Juglans* Species. G3: Genes, Genomes, Genetics, 2018, 8, 2153-2165.	0.8	73
52	Removing contaminants from databases of draft genomes. PLoS Computational Biology, 2018, 14, e1006277.	1.5	85
53	Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. Nature Communications, 2018, 9, 2606.	5.8	79
54	Development and Optimization of Metagenomic Next-Generation Sequencing Methods for Cerebrospinal Fluid Diagnostics. Journal of Clinical Microbiology, 2018, 56, .	1.8	65

#	Article	IF	Citations
55	Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. Nature Communications, 2018, 9, 3391.	5.8	140
56	Identifying Corneal Infections in Formalin-Fixed Specimens Using Next Generation Sequencing. , 2018, 59, 280.		51
57	Open questions: How many genes do we have?. BMC Biology, 2018, 16, 94.	1.7	127
58	MUMmer4: A fast and versatile genome alignment system. PLoS Computational Biology, 2018, 14, e1005944.	1.5	1,412
59	Short Read Mapping: An Algorithmic Tour. Proceedings of the IEEE, 2017, 105, 436-458.	16.4	63
60	The novel fusion transcript NR5A2–KLHL29FT is generated by an insertion at the KLHL29 locus. Cancer, 2017, 123, 1507-1515.	2.0	4
61	Hybrid assembly of the large and highly repetitive genome of <i>Aegilops tauschii</i>, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Research, 2017, 27, 787-792.	2.4	382
62	Horizontal gene transfer is not a hallmark of the human genome. Genome Biology, 2017, 18, 85.	3.8	55
63	16GT: a fast and sensitive variant caller using a 16-genotype probabilistic model. GigaScience, 2017, 6, 1-4.	3.3	11
64	First Draft Genome Sequence of the Pathogenic Fungus <i>Lomentospora prolificans</i> (Formerly) Tj ETQq0 0 0 rgBT /Overlock 10 Tf 5	0.8	11
65	The Douglas-Fir Genome Sequence Reveals Specialization of the Photosynthetic Apparatus in Pinaceae. G3: Genes, Genomes, Genetics, 2017, 7, 3157-3167.	0.8	103
66	Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. Nature, 2017, 551, 498-502.	13.7	563
67	Presence of Human Hepegivirus-1 in a Cohort of People Who Inject Drugs. Annals of Internal Medicine, 2017, 167, 1.	2.0	28
68	The first near-complete assembly of the hexaploid bread wheat genome, Triticum aestivum. GigaScience, 2017, 6, 1-7.	3.3	224
69	An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. GigaScience, 2017, 6, 1-4.	3.3	71
70	First Draft Assembly and Annotation of the Genome of a California Endemic Oak<i>Quercus lobata</i>Née (Fagaceae). G3: Genes, Genomes, Genetics, 2016, 6, 3485-3495.	0.8	95
71	Reminder to deposit DNA sequences. Nature, 2016, 533, 179-179.	13.7	8
72	Reminder to deposit DNA sequences. Science, 2016, 352, 780-780.	6.0	24

#	Article	IF	Citations
73	Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols, 2016, 11, 1650-1667.	5.5	4,743
74	Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Research, 2016, 26, 1721-1729.	2.4	1,025
75	Sequence of the Sugar Pine Megagenome. Genetics, 2016, 204, 1613-1626.	1.2	169
76	The walnut (<i>Juglans regia</i>) genome sequence reveals diversity in genes coding for the biosynthesis of non‑structural polyphenols. Plant Journal, 2016, 87, 507-532.	2.8	233
77	Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system. Neurology: Neuroimmunology and NeuroInflammation, 2016, 3, e251.	3.1	142
78	Use and mis-use of supplementary material in science publications. BMC Bioinformatics, 2015, 16, 237.	1.2	36
79	Single molecule real-time sequencing of Xanthomonas oryzae genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. Microbial Genomics, 2015, 1, .	1.0	86
80	Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. Nature, 2015, 517, 381-385.	13.7	469
81	StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology, 2015, 33, 290-295.	9.4	8,385
82	HISAT: a fast spliced aligner with low memory requirements. Nature Methods, 2015, 12, 357-360.	9.0	16,262
83	Ballgown bridges the gap between transcriptome assembly and expression analysis. Nature Biotechnology, 2015, 33, 243-246.	9.4	716
84	The genomes of two key bumblebee species with primitive eusocial organization. Genome Biology, 2015, 16, 76.	3.8	330
85	Genomic signatures of evolutionary transitions from solitary to group living. Science, 2015, 348, 1139-1143.	6.0	357
86	Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. Genome Research, 2015, 25, 1401-1409.	2.4	91
87	Re-analysis of metagenomic sequences from acute flaccid myelitis patients reveals alternatives to enterovirus D68 infection. F1000Research, 2015, 4, 180.	0.8	9
88	Re-analysis of metagenomic sequences from acute flaccid myelitis patients reveals alternatives to enterovirus D68 infection. F1000Research, 2015, 4, 180.	0.8	6
89	Arioc: high-throughput read alignment with GPU-accelerated exploration of the seed-and-extend search space. PeerJ, 2015, 3, e808.	0.9	33
90	A new rhesus macaque assembly and annotation for next-generation sequencing analyses. Biology Direct, 2014, 9, 20.	1.9	165

#	ARTICLE	IF	CITATIONS
91	Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. Genetics, 2014, 196, 875-890.	1.2	286
92	Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology, 2014, 15, R46.	13.9	3,540
93	Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biology, 2014, 15, R59.	13.9	424
94	DIAMUND : Direct Comparison of Genomes to Detect Mutations. Human Mutation, 2014, 35, 283-288.	1.1	9
95	Genomic Features of a Bumble Bee Symbiont Reflect Its Host Environment. Applied and Environmental Microbiology, 2014, 80, 3793-3803.	1.4	53
96	Unique Features of the Loblolly Pine (*Pinus taeda* L.) Megagenome Revealed Through Sequence Annotation. Genetics, 2014, 196, 891-909.	1.2	207
97	Unexpected cross-species contamination in genome sequencing projects. PeerJ, 2014, 2, e675.	0.9	145
98	TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology, 2013, 14, R36.	13.9	11,713
99	The MaSuRCA genome assembler. Bioinformatics, 2013, 29, 2669-2677.	1.8	1,127
100	GAGE-B: an evaluation of genome assemblers for bacterial organisms. Bioinformatics, 2013, 29, 1718-1725.	1.8	135
101	Hawkeye and AMOS: visualizing and assessing the quality of genome assemblies. Briefings in Bioinformatics, 2013, 14, 213-224.	3.2	54
102	NIH funding: It does support innovators. Nature, 2013, 493, 26-26.	13.7	1
103	Sequestration: inadvertently killing biomedical research to score political points. Genome Biology, 2013, 14, 109.	13.9	2
104	Open access to tree genomes: the path to a better forest. Genome Biology, 2013, 14, 120.	3.8	46
105	Genome-Guided Transcriptome Assembly in the Age of Next-Generation Sequencing. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2013, 10, 1234-1240.	1.9	16
106	The COMBREX Project: Design, Methodology, and Initial Results. PLoS Biology, 2013, 11, e1001638.	2.6	54
107	EDGE-pro: Estimated Degree of Gene Expression in Prokaryotic Genomes. Evolutionary Bioinformatics, 2013, 9, EBO.S11250.	0.6	141
108	Insights into the Loblolly Pine Genome: Characterization of BAC and Fosmid Sequences. PLoS ONE, 2013, 8, e72439.	1.1	46

#	ARTICLE	IF	CITATIONS
109	Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. F1000Research, 2013, 2, 188.	0.8	372
110	Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. F1000Research, 2013, 2, 188.	0.8	293
111	Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. Nucleic Acids Research, 2012, 40, e9-e9.	6.5	150
112	The Perils of Gene Patents. Clinical Pharmacology and Therapeutics, 2012, 91, 969-971.	2.3	7
113	Genome interpretation and assembly—recent progress and next steps. Nature Biotechnology, 2012, 30, 1081-1083.	9.4	10
114	The International Nucleotide Sequence Database Collaboration. Nucleic Acids Research, 2012, 40, D33-D37.	6.5	327
115	Thousands of missed genes found in bacterial genomes and their analysis with COMBREX. Biology Direct, 2012, 7, 37.	1.9	28
116	Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols, 2012, 7, 562-578.	5.5	11,433
117	Fast gapped-read alignment with Bowtie 2. Nature Methods, 2012, 9, 357-359.	9.0	42,357
118	Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature, 2012, 487, 94-98.	13.7	1,086
119	GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Research, 2012, 22, 557-567.	2.4	597
120	Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics, 2012, 13, 36-46.	7.7	1,382
121	Mis-Assembled "Segmental Duplications" in Two Versions of the Bos taurus Genome. PLoS ONE, 2012, 7, e42680.	1.1	22
122	Two New Complete Genome Sequences Offer Insight into Host and Tissue Specificity of Plant Pathogenic Xanthomonas spp. Journal of Bacteriology, 2011, 193, 5450-5464.	1.0	189
123	TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biology, 2011, 12, R72.	13.9	670
124	Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. Genome Biology, 2011, 12, R51.	13.9	56
125	Closure of the NCBI SRA and implications for the long-term future of genomics data storage. Genome Biology, 2011, 12, 402.	3.8	29
126	Genome Assembly Has a Major Impact on Gene Content: A Comparison of Annotation in Two Bos Taurus Assemblies. PLoS ONE, 2011, 6, e21400.	1.1	65

#	ARTICLE	IF	CITATIONS
127	The genome of woodland strawberry (Fragaria vesca). Nature Genetics, 2011, 43, 109-116.	9.4	1,091
128	PhymmBL expanded: confidence scores, custom databases, parallelization and more. Nature Methods, 2011, 8, 367-367.	9.0	97
129	Improving pan-genome annotation using whole genome multiple alignment. BMC Bioinformatics, 2011, 12, 272.	1.2	38
130	Detection of lineage-specific evolutionary changes among primate species. BMC Bioinformatics, 2011, 12, 274.	1.2	19
131	COMBREX: a project to accelerate the functional annotation of prokaryotic genomes. Nucleic Acids Research, 2011, 39, D11-D14.	6.5	47
132	FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 2011, 27, 2957-2963.	1.8	11,573
133	*Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108, 5027-5032.	3.3	152
134	Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics, 2011, 27, 334-342.	1.8	444
135	Assembly of large genomes using second-generation sequencing. Genome Research, 2010, 20, 1165-1173.	2.4	390
136	Clustering metagenomic sequences with interpolated Markov models. BMC Bioinformatics, 2010, 11, 544.	1.2	90
137	Probing the pan-genome of Listeria monocytogenes: new insights into intraspecific niche expansion and genomic diversification. BMC Genomics, 2010, 11, 500.	1.2	74
138	Cloud computing and the DNA data race. Nature Biotechnology, 2010, 28, 691-693.	9.4	242
139	Mind the gaps. Nature Methods, 2010, 7, 105-106.	9.0	10
140	Genome Sequence of the Dioxin-Mineralizing Bacterium *Sphingomonas wittichii* RW1. Journal of Bacteriology, 2010, 192, 6101-6102.	1.0	93
141	Multi-Platform Next-Generation Sequencing of the Domestic Turkey (Meleagris gallopavo): Genome Assembly and Analysis. PLoS Biology, 2010, 8, e1000475.	2.6	348
142	Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology, 2010, 28, 511-515.	9.4	13,805
143	Quake: quality-aware detection and correction of sequencing errors. Genome Biology, 2010, 11, R116.	13.9	470
144	Do-it-yourself genetic testing. Genome Biology, 2010, 11, 404.	3.8	16

#	Article	IF	Citations
145	Detection and correction of false segmental duplications caused by genome mis-assembly. Genome Biology, 2010, 11, R28.	13.9	96
146	Genes and genomes, an imperfect world: comparison of gene annotations of two Bos taurus draft assemblies. Genome Biology, 2010, 11, P13.	13.9	1
147	Between a chicken and a grape: estimating the number of human genes. Genome Biology, 2010, 11, 206.	13.9	182
148	Recent advances in RNA sequence analysis. F1000 Biology Reports, 2010, 2, 64.	4.0	13
149	EFFORTS TO DEREGULATE RAINBOW PAPAYA IN JAPAN: MOLECULAR CHARACTERIZATION OF TRANSGENE AND VECTOR INSERTS. Acta Horticulturae, 2010, , 235-240.	0.1	0
150	2009 Swine-Origin Influenza A (H1N1) Resembles Previous Influenza Isolates. PLoS ONE, 2009, 4, e6402.	1.1	31
151	OperonDB: a comprehensive database of predicted operons in microbial genomes. Nucleic Acids Research, 2009, 37, D479-D482.	6.5	83
152	Insignia: a DNA signature search web server for diagnostic assay development. Nucleic Acids Research, 2009, 37, W229-W234.	6.5	37
153	Open Access: The Sooner the Better. Science, 2009, 325, 266-266.	6.0	0
154	Genome Sequence of the *Wolbachia* Endosymbiont of *Culex quinquefasciatus* JHB. Journal of Bacteriology, 2009, 191, 1725-1725.	1.0	44
155	The Complete Genome Sequence of Bacillus anthracis Ames â€œAncestorâ€. Journal of Bacteriology, 2009, 191, 445-446.	1.0	88
156	Efficient oligonucleotide probe selection for pan-genomic tiling arrays. BMC Bioinformatics, 2009, 10, 293.	1.2	25
157	The genome of the blood fluke Schistosoma mansoni. Nature, 2009, 460, 352-358.	13.7	945
158	How to map billions of short reads onto genomes. Nature Biotechnology, 2009, 27, 455-457.	9.4	257
159	Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nature Methods, 2009, 6, 673-676.	9.0	526
160	TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 2009, 25, 1105-1111.	1.8	11,265
161	Searching for SNPs with cloud computing. Genome Biology, 2009, 10, R134.	13.9	437
162	A whole-genome assembly of the domestic cow, Bos taurus. Genome Biology, 2009, 10, R42.	13.9	1,005

#	Article	IF	Citations
163	Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology, 2009, 10, R25.	13.9	19,212
164	Genome-Wide Analysis of Repetitive Elements in Papaya. Tropical Plant Biology, 2008, 1, 191-201.	1.0	24
165	Characterization of Insertion Sites in Rainbow Papaya, the First Commercialized Transgenic Fruit Crop. Tropical Plant Biology, 2008, 1, 293-309.	1.0	25
166	The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature, 2008, 452, 991-996.	13.7	964
167	Comparative genomics of the neglected human malaria parasite Plasmodium vivax. Nature, 2008, 455, 757-763.	13.7	756
168	The contents of the syringe. Nature, 2008, 454, 160-161.	13.7	39
169	Genome sequence and rapid evolution of the rice pathogen Xanthomonas oryzae pv. oryzae PXO99A. BMC Genomics, 2008, 9, 204.	1.2	327
170	Genome sequence and rapid evolution of the rice pathogen Xanthomonas oryzae pv. oryzae PXO99A. BMC Genomics, 2008, 9, 534.	1.2	33
171	Bioinformatics challenges of new sequencing technology. Trends in Genetics, 2008, 24, 142-149.	2.9	438
172	Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biology, 2008, 9, R7.	13.9	2,484
173	What are decision trees?. Nature Biotechnology, 2008, 26, 1011-1013.	9.4	409
174	Gene-Boosted Assembly of a Novel Bacterial Genome from Very Short Reads. PLoS Computational Biology, 2008, 4, e1000186.	1.5	46
175	Re-Assembly of the Genome of Francisella tularensis Subsp. holarctica OSU18. PLoS ONE, 2008, 3, e3427.	1.1	9
176	Acquisition and Evolution of Plant Pathogenesis–Associated Gene Clusters and Candidate Determinants of Tissue-Specificity in Xanthomonas. PLoS ONE, 2008, 3, e3828.	1.1	89
177	Comprehensive DNA Signature Discovery and Validation. PLoS Computational Biology, 2007, 3, e98.	1.5	63
178	A Unified Model Explaining the Offsets of Overlapping and Near-Overlapping Prokaryotic Genes. Molecular Biology and Evolution, 2007, 24, 2091-2098.	3.5	11
179	Hawkeye: an interactive visual analytics tool for genome assemblies. Genome Biology, 2007, 8, R34.	13.9	66
180	Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biology, 2007, 8, R22.	13.9	432

#	ARTICLE	IF	CITATIONS
181	Genome re-annotation: a wiki solution?. Genome Biology, 2007, 8, 102.	3.8	127
182	Draft Genome Sequence of the Sexually Transmitted Pathogen Trichomonas vaginalis. Science, 2007, 315, 207-212.	6.0	731
183	Draft Genome of the Filarial Nematode Parasite <i>Brugia malayi</i>. Science, 2007, 317, 1756-1760.	6.0	571
184	Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics, 2007, 23, 673-679.	1.8	2,684
185	Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. Science, 2007, 316, 1718-1723.	6.0	1,025
186	Genome Analysis Linking Recent European and African Influenza (H5N1) Viruses. Emerging Infectious Diseases, 2007, 13, 713-718.	2.0	191
187	Evolution of genes and genomes on the Drosophila phylogeny. Nature, 2007, 450, 203-218.	13.7	1,886
188	A computational survey of candidate exonic splicing enhancer motifs in the model plant Arabidopsis thaliana. BMC Bioinformatics, 2007, 8, 159.	1.2	81
189	Minimus: a fast, lightweight genome assembler. BMC Bioinformatics, 2007, 8, 64.	1.2	354
190	Gemina: A Web-Based Epidemiology and Genomic Metadata System Designed to Identify Infectious Agents. , 2007, , 228-229.		1
191	Using Protein Domains to Improve the Accuracy of Ab Initio Gene Finding. Lecture Notes in Computer Science, 2007, , 208-215.	1.0	0
192	A phylogenetic generalized hidden Markov model for predicting alternatively spliced exons. Algorithms for Molecular Biology, 2006, 1, 14.	0.3	10
193	JIGSAW, GeneZilla, and GlimmerHMM: puzzling out the features of human genes in the ENCODE regions. Genome Biology, 2006, 7, S9.	13.9	61
194	Shared data are key to beating threat from flu. Nature, 2006, 440, 605-605.	13.7	5
195	Physiogenomic resources for rat models of heart, lung and blood disorders. Nature Genetics, 2006, 38, 234-239.	9.4	48
196	It is time to end the patenting of software. Bioinformatics, 2006, 22, 1416-1417.	1.8	2
197	Macronuclear Genome Sequence of the Ciliate Tetrahymena thermophila, a Model Eukaryote. PLoS Biology, 2006, 4, e286.	2.6	657
198	Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature, 2005, 437, 1162-1166.	13.7	419

#	Article	IF	Citations
199	Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. Nature, 2005, 438, 1151-1156.	13.7	1,272
200	Efficient decoding algorithms for generalized hidden Markov model gene finders. BMC Bioinformatics, 2005, 6, 16.	1.2	34
201	The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. BMC Biology, 2005, 3, 20.	1.7	158
202	Bioinformatics Methods for Microbial Detection and Forensic Diagnostic Design. , 2005, , 313-353.		0
203	Beware of mis-assembled genomes. Bioinformatics, 2005, 21, 4320-4321.	1.8	165
204	Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. Genome Research, 2005, 15, 1284-1291.	2.4	73
205	JIGSAW: integration of multiple sources of evidence for gene prediction. Bioinformatics, 2005, 21, 3596-3603.	1.8	147
206	Efficient implementation of a generalized pair hidden Markov model for comparative gene finding. Bioinformatics, 2005, 21, 1782-1788.	1.8	38
207	Whole-Genome Analysis of Human Influenza A Virus Reveals Multiple Persistent Lineages and Reassortment among Recent H3N2 Viruses. PLoS Biology, 2005, 3, e300.	2.6	340
208	The Transcriptional Landscape of the Mammalian Genome. Science, 2005, 309, 1559-1563.	6.0	3,227
209	The Genome Sequence of Trypanosoma cruzi, Etiologic Agent of Chagas Disease. Science, 2005, 309, 409-415.	6.0	1,273
210	Comparative Genomics of Trypanosomatid Parasitic Protozoa. Science, 2005, 309, 404-409.	6.0	713
211	The Genome of the Basidiomycetous Yeast and Human Pathogen Cryptococcus neoformans. Science, 2005, 307, 1321-1324.	6.0	664
212	Genome Sequence of Theileria parva, a Bovine Pathogen That Transforms Lymphocytes. Science, 2005, 309, 134-137.	6.0	309
213	The Genome of the African Trypanosome Trypanosoma brucei. Science, 2005, 309, 416-422.	6.0	1,496
214	New names for old strains? Wolbachia wSim is actually wRi. Genome Biology, 2005, 6, 401.	13.9	11
215	Serendipitous discovery of Wolbachia genomes in multiple Drosophila species. Genome Biology, 2005, 6, R23.	13.9	126
216	Automated correction of genome sequence errors. Nucleic Acids Research, 2004, 32, 562-569.	6.5	42

#	ARTICLE	IF	CITATIONS
217	TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics, 2004, 20, 2878-2879.	1.8	1,323
218	Comparative genome assembly. Briefings in Bioinformatics, 2004, 5, 237-248.	3.2	193
219	Genomic Insights into Methanotrophy: The Complete Genome Sequence of Methylococcus capsulatus (Bath). PLoS Biology, 2004, 2, e303.	2.6	275
220	DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics, 2004, 20, 3643-3646.	1.8	340
221	Gene synteny and evolution of genome architecture in trypanosomatids. Molecular and Biochemical Parasitology, 2004, 134, 183-191.	0.5	92
222	An empirical analysis of training protocols for probabilistic gene finders. BMC Bioinformatics, 2004, 5, 206.	1.2	23
223	Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. Genome Biology, 2004, 5, R61.	13.9	184
224	Versatile and open software for comparing large genomes. Genome Biology, 2004, 5, R12.	13.9	4,989
225	Insight into the genome of Aspergillus fumigatus: analysis of a 922kb region encompassing the nitrate assimilation gene cluster. Fungal Genetics and Biology, 2004, 41, 443-453.	0.9	55
226	The ENCODE (ENCyclopedia Of DNA Elements) Project. Science, 2004, 306, 636-640.	6.0	2,121
227	The Genome Assembly Archive: A New Public Resource. PLoS Biology, 2004, 2, e285.	2.6	30
228	The age of the Arabidopsis thaliana genome duplication. Plant Molecular Biology, 2003, 51, 859-866.	2.0	115
229	Statement on the consideration of biodefence and biosecurity. Nature, 2003, 421, 771-771.	13.7	31
230	Unrestricted free access works and must continue. Nature, 2003, 422, 801-801.	13.7	6
231	Yeast rises again. Nature, 2003, 423, 233-234.	13.7	14
232	Using MUMmer to Identify Similar Regions in Large Sequence Sets. Current Protocols in Bioinformatics, 2003, 00, Unit 10.3.	25.8	505
233	The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria. Nature, 2003, 423, 81-86.	13.7	760
234	The sequence and analysis of Trypanosoma brucei chromosome II. Nucleic Acids Research, 2003, 31, 4856-4863.	6.5	59

#	ARTICLE	IF	CITATIONS
235	Uncensored exchange of scientific results. Proceedings of the National Academy of Sciences of the United States of America, 2003, 100, 1464-1464.	3.3	37
236	In-Depth View of Structure, Activity, and Evolution of Rice Chromosome 10. Science, 2003, 300, 1566-1569.	6.0	245
237	Computational Discovery of Internal Micro-Exons. Genome Research, 2003, 13, 1216-1221.	2.4	62
238	GlimmerM, Exonomy and Unveil: three ab initio eukaryotic genefinders. Nucleic Acids Research, 2003, 31, 3601-3604.	6.5	60
239	Genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. Nucleic Acids Research, 2003, 31, 2134-2147.	6.5	266
240	Using GlimmerM to Find Genes in Eukaryotic Genomes. Current Protocols in Bioinformatics, 2003, 00, Unit 4.4.	25.8	5
241	Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research, 2003, 31, 5654-5666.	6.5	1,597
242	Hierarchical Scaffolding With Bambus. Genome Research, 2003, 14, 149-159.	2.4	172
243	Computational Gene Prediction Using Multiple Sources of Evidence. Genome Research, 2003, 14, 142-148.	2.4	110
244	A Comparison of Whole-Genome Shotgun-Derived Mouse Chromosome 16 and the Human Genome. Science, 2002, 296, 1661-1671.	6.0	344
245	Comparative Genome and Proteome Analysis ofAnopheles gambiaeandDrosophila melanogaster. Science, 2002, 298, 149-159.	6.0	531
246	Comparative Genome Sequencing for Discovery of Novel Polymorphisms in Bacillus anthracis. Science, 2002, 296, 2028-2033.	6.0	413
247	The Brucella suis genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99, 13148-13153.	3.3	422
248	Whole-Genome Comparison of Mycobacterium tuberculosis Clinical and Laboratory Strains. Journal of Bacteriology, 2002, 184, 5479-5490.	1.0	645
249	The Value of Complete Microbial Genome Sequencing (You Get What You Pay For). Journal of Bacteriology, 2002, 184, 6403-6405.	1.0	101
250	Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Research, 2002, 30, 2478-2483.	6.5	931
251	The Genome Sequence of the Malaria MosquitoAnopheles gambiae. Science, 2002, 298, 129-149.	6.0	1,859
252	Full-length messenger RNA sequences greatly improve genome annotation. Genome Biology, 2002, 3, research0029.1.	13.9	147

#	ARTICLE	IF	CITATIONS
253	Contamination in the Draft of the Human Genome Masquerades As Lateral Gene Transfer. DNA Sequence, 2002, 13, 75-76.	0.7	15
254	A preliminary comparison of the mouse and human genomes. International Congress Series, 2002, 1246, 169-181.	0.2	2
255	Genome sequence assembly: algorithms and issues. Computer, 2002, 35, 47-54.	1.2	60
256	Sequence of Plasmodium falciparum chromosomes 2, 10, 11 and 14. Nature, 2002, 419, 531-534.	13.7	167
257	Genome sequence of the human malaria parasite Plasmodium falciparum. Nature, 2002, 419, 498-511.	13.7	3,881
258	Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. Nature, 2002, 419, 512-519.	13.7	666
259	Computational gene finding in plants. Plant Molecular Biology, 2002, 48, 39-48.	2.0	41
260	A Method to Improve the Performance of Translation Start Site Detection and Its Application for Gene Finding. Lecture Notes in Computer Science, 2002, , 210-219.	1.0	6
261	Genomics of Theileria Parva. World Class Parasites, 2002, , 85-92.	0.3	1
262	Computational gene finding in plants. , 2002, , 39-48.		1
263	Computational gene finding in plants. Plant Molecular Biology, 2002, 48, 39-48.	2.0	12
264	A clustering method for repeat analysis in DNA sequences. Genome Biology, 2001, 2, research0027.1.	13.9	114
265	Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae. Science, 2001, 293, 498-506.	6.0	1,281
266	Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss?. Science, 2001, 292, 1903-1906.	6.0	284
267	GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Research, 2001, 29, 1185-1190.	6.5	514
268	The Sequence of the Human Genome. Science, 2001, 291, 1304-1351.	6.0	12,623
269	Complete genome sequence of Caulobacter crescentus. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98, 4136-4141.	3.3	489
270	Sequence and analysis of the Arabidopsis genome. Current Opinion in Plant Biology, 2001, 4, 105-110.	3.5	66

#	Article	IF	Citations
271	Understanding the Adaptation of Halobacterium Species NRC-1 to Its Extreme Environment through Computational Analysis of Its Genome Sequence. Genome Research, 2001, 11, 1641-1650.	2.4	294
272	Prediction of operons in microbial genomes. Nucleic Acids Research, 2001, 29, 1216-1221.	6.5	255
273	A probabilistic method for identifying start codons in bacterial genomes. Bioinformatics, 2001, 17, 1123-1130.	1.8	180
274	Rice Bioinformatics. Analysis of Rice Sequence Data and Leveraging the Data to Other Plant Species. Plant Physiology, 2001, 125, 1166-1174.	2.3	46
275	Gene Index analysis of the human genome estimates approximately 120,000 genes. Nature Genetics, 2000, 25, 239-240.	9.4	262
276	Finding genes in Plasmodium falciparum. Nature, 2000, 404, 34-34.	13.7	18
277	Finding genes in Plasmodium falciparum. Nature, 2000, 404, 34-35.	13.7	17
278	DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. Nature, 2000, 406, 477-483.	13.7	1,723
279	Microbial genome sequencing. Nature, 2000, 406, 799-803.	13.7	164
280	Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana. Nature, 2000, 408, 816-820.	13.7	234
281	Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, 2000, 408, 796-815.	13.7	8,336
282	Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana. Nature, 2000, 408, 820-823.	13.7	188
283	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Research, 2000, 28, 1397-1406.	6.5	696
284	An optimized protocol for analysis of EST sequences. Nucleic Acids Research, 2000, 28, 3657-3665.	6.5	121
285	Prediction of transcription terminators in bacterial genomes 1 1Edited by F. E. Cohen. Journal of Molecular Biology, 2000, 301, 27-33.	2.0	231
286	Complete Genome Sequence of Neisseria meningitidis Serogroup B Strain MC58. Science, 2000, 287, 1809-1815.	6.0	1,083
287	Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biology, 2000, 1, research0011.1.	13.9	287
288	Alignment of whole genomes. Nucleic Acids Research, 1999, 27, 2369-2376.	6.5	766

#	ARTICLE	IF	CITATIONS
289	Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima. Nature, 1999, 399, 323-329.	13.7	1,397
290	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. Nature, 1999, 402, 761-768.	13.7	724
291	Genome Sequence of the Radioresistant Bacterium Deinococcus radiodurans R1. Science, 1999, 286, 1571-1577.	6.0	879
292	DNA uptake signal sequences in naturally transformable bacteria. Research in Microbiology, 1999, 150, 603-616.	1.0	117
293	Gene discovery in DNA sequences. IEEE Intelligent Systems, 1999, 14, 44-48.	0.2	10
294	Improved microbial gene identification with GLIMMER. Nucleic Acids Research, 1999, 27, 4636-4641.	6.5	2,282
295	Interpolated Markov Models for Eukaryotic Gene Finding. Genomics, 1999, 59, 24-31.	1.3	184
296	Optimized Multiplex PCR: Efficiently Closing a Whole-Genome Shotgun Sequencing Project. Genomics, 1999, 62, 500-507.	1.3	131
297	A probabilistic framework for memory-based reasoning. Artificial Intelligence, 1998, 104, 287-311.	3.9	35
298	Skewed oligomers and origins of replication. Gene, 1998, 217, 57-67.	1.0	171
299	Complete Genome Sequence of Treponema pallidum, the Syphilis Spirochete. , 1998, 281, 375-388.		969
300	A Decision Tree System for Finding Genes in DNA. Journal of Computational Biology, 1998, 5, 667-680.	0.8	102
301	Microbial gene identification using interpolated Markov models. Nucleic Acids Research, 1998, 26, 544-548.	6.5	896
302	Chromosome 2 Sequence of the Human Malaria Parasite Plasmodium falciparum. , 1998, 282, 1126-1132.		419
303	A tutorial introduction to computation for biologists. New Comprehensive Biochemistry, 1998, , 11-27.	0.1	0
304	An introduction to biological sequence analysis. New Comprehensive Biochemistry, 1998, 32, 29-42.	0.1	3
305	Decision trees and Markov chains for gene finding. New Comprehensive Biochemistry, 1998, , 187-203.	0.1	4
306	A method for identifying splice sites and translational start sites in eukaryotic mRNA. Bioinformatics, 1997, 13, 365-376.	1.8	78

#	Article	IF	Citations
307	Finding Genes in DNA with a Hidden Markov Model. Journal of Computational Biology, 1997, 4, 127-141.	0.8	198
308	Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. Nature, 1997, 390, 580-586.	13.7	1,985
309	A Teaching Strategy for Memory-Based Control. , 1997, 11, 343-370.		20
310	On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach. , 1997, 1, 317-328.		653
311	Testing simple polygons. Computational Geometry: Theory and Applications, 1997, 8, 97-114.	0.3	2
312	A Teaching Strategy for Memory-Based Control. , 1997, , 343-370.		9
313	Learning nested concept classes with limited storage. Journal of Experimental and Theoretical Artificial Intelligence, 1996, 8, 129-147.	1.8	2
314	Chapter 18 Committees of decision trees. Advances in Psychology, 1996, 113, 305-317.	0.1	12
315	Testing orthogonal shapes. Computational Geometry: Theory and Applications, 1995, 5, 33-49.	0.3	2
316	Decision trees for automated identification of cosmic-ray hits in Hubble Space Telescope images. Publications of the Astronomical Society of the Pacific, 1995, 107, 279.	1.0	60
317	Locating Protein Coding Regions in Human DNA Using a Decision Tree Algorithm. Journal of Computational Biology, 1995, 2, 473-485.	0.8	64
318	Best-case results for nearest-neighbor learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17, 599-608.	9.7	20
319	Towards a Better Understanding of Memory-Based Reasoning Systems. , 1994, , 242-250.		27
320	Title is missing!. Machine Learning, 1994, 16, 235-240.	3.4	61
321	Learning to Catch: Applying Nearest Neighbor Algorithms to Dynamic Control Tasks. Lecture Notes in Statistics, 1994, , 321-328.	0.1	17
322	A weighted nearest neighbor algorithm for learning with symbolic features. Machine Learning, 1993, 10, 57-78.	3.4	415
323	A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features. Machine Learning, 1993, 10, 57-78.	3.4	339
324	Predicting protein secondary structure with a nearest-neighbor algorithm. Journal of Molecular Biology, 1992, 227, 371-374.	2.0	82

#	Article	IF	Citations
325	A nearest hyperrectangle learning method. Machine Learning, 1991, 6, 251-276.	3.4	215
326	A Nearest Hyperrectangle Learning Method. Machine Learning, 1991, 6, 251-276.	3.4	177
327	Distance metrics for instance-based learning. Lecture Notes in Computer Science, 1991, , 399-408.	1.0	18
328	Managing information for concurrent engineering: Challenges and barriers. Research in Engineering Design - Theory, Applications, and Concurrent Engineering, 1990, 2, 35-52.	1.2	33
329	Learning with Nested Generalized Exemplars. Kluwer International Series in Engineering and Computer Science, 1990, , .	0.2	38
330	Nested hyper-rectangles for exemplar-based learning. Lecture Notes in Computer Science, 1989, , 184-201.	1.0	3
331	Finding Repeats in Genome Sequences. , 0, , 197-233.		3
332	Foreword by Steven Salzberg. , 0, , xi-xii.		0
333	Identification of microbial agents in tissue specimens of ocular and periocular sarcoidosis using a metagenomics approach. F1000Research, 0, 10, 820.	0.8	2
334	Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science, 0, 3, e104.	2.7	928