

Steven L Salzberg

List of Articles by Year in descending order

Source: [//exaly.com/author-pdf/6635057/publications.pdf](https://exaly.com/author-pdf/6635057/publications.pdf)

Version: 2025-02-01

276

PR articles

195,125

PR citations

697

115

PR h-index

565

257

g-index

330

documents

222588

doc citations

452

134

h-index

294779

citing authors

#	ARTICLE	IF	CITATIONS
1	PSAURON: a tool for assessing protein annotation across a broad range of species. <i>NAR Genomics and Bioinformatics</i> , 2025, 7, .	2.2	18
2	Conservation assessment of human splice site annotation based on a 470-genome alignment. <i>Nucleic Acids Research</i> , 2025, 53, .	15.5	0
3	Reply to Austin and Korem, "Compositional transformations can reasonably introduce phenotype-associated values into sparse features". <i>MSystems</i> , 2025, 10, .	4.4	0
4	Discovering Intron Gain Events in Humans Through Large-Scale Evolutionary Comparisons. <i>Genome Biology and Evolution</i> , 2025, 17, .	2.4	0
5	Phenotype to genotype: A new and rapid approach using whole-genome sequencing. <i>PLoS Genetics</i> , 2025, 21, e1011702.	3.2	1
6	Implementing governmental oversight of enhanced potential pandemic pathogen research. <i>Journal of Virology</i> , 2024, 98, .	3.6	3
7	A genome sequence for the threatened whitebark pine. <i>G3: Genes, Genomes, Genetics</i> , 2024, 14, .	1.9	10
8	Genomic variability in Zika virus in GBS cases in Colombia. <i>PLoS ONE</i> , 2024, 19, e0313545.	2.3	0
9	Upstream open reading frames may contain hundreds of novel human exons. <i>PLoS Computational Biology</i> , 2024, 20, e1012543.	3.1	3
10	JASPER: A fast genome polishing tool that improves accuracy of genome assemblies. <i>PLoS Computational Biology</i> , 2023, 19, e1011032.	3.1	11
11	Investigating open reading frames in known and novel transcripts using ORFanage. <i>Nature Computational Science</i> , 2023, 3, 700-708.	11.6	23
12	The complete sequence of a human Y chromosome. <i>Nature</i> , 2023, 621, 344-354.	37.9	362
13	Major data analysis errors invalidate cancer microbiome findings. <i>MBio</i> , 2023, 14, .	4.4	151
14	PhyloCSF++: a fast and user-friendly implementation of PhyloCSF with annotation tools. <i>Bioinformatics</i> , 2022, 38, 1440-1442.	4.7	16
15	Genome-wide association identifies candidate genes for drought tolerance in coast redwood and giant sequoia. <i>Plant Journal</i> , 2022, 109, 7-22.	6.2	36
16	The SAMBA tool uses long reads to improve the contiguity of genome assemblies. <i>PLoS Computational Biology</i> , 2022, 18, e1009860.	3.1	45
17	Assembled and annotated 26.5 Gbp coast redwood genome: a resource for estimating evolutionary adaptive potential and investigating hexaploid origin. <i>G3: Genes, Genomes, Genetics</i> , 2022, 12, .	1.9	48
18	A reference-quality, fully annotated genome from a Puerto Rican individual. <i>Genetics</i> , 2022, 220, .	4.2	17

#	ARTICLE	IF	CITATIONS
19	Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. <i>Nature Genetics</i> , 2022, 54, 263-273.	25.2	308
20	The complete sequence of a human genome. <i>Science</i> , 2022, 376, 44-53.	36.2	2,665
21	Next-generation sequencing: insights to advance clinical investigations of the microbiome. <i>Journal of Clinical Investigation</i> , 2022, 132, .	10.6	344
22	High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. <i>Nature Communications</i> , 2022, 13, .	13.7	61
23	The Human "Contaminome" and Understanding Infectious Disease. <i>New England Journal of Medicine</i> , 2022, 387, 943-946.	34.6	16
24	Metagenome analysis using the Kraken software suite. <i>Nature Protocols</i> , 2022, 17, 2815-2839.	14.4	583
25	Semi-automated assembly of high-quality diploid human reference genomes. <i>Nature</i> , 2022, 611, 519-531.	37.9	176
26	3D-Beacons: decreasing the gap between protein sequences and structures through a federated network of protein structure data resources. <i>GigaScience</i> , 2022, 11, .	3.2	27
27	Effects of transcriptional noise on estimates of gene and transcript expression in RNA sequencing experiments. <i>Genome Research</i> , 2021, 31, 301-308.	4.6	31
28	Liftoff: accurate mapping of gene annotations. <i>Bioinformatics</i> , 2021, 37, 1639-1643.	4.7	812
29	Dissecting the Polygenic Basis of Cold Adaptation Using Genome-Wide Association of Traits and Environmental Data in Douglas-fir. <i>Genes</i> , 2021, 12, 110.	2.5	29
30	Balrog: A universal protein model for prokaryotic gene prediction. <i>PLoS Computational Biology</i> , 2021, 17, e1008727.	3.1	30
31	Guillain-Barré Syndrome Outbreak in Peru 2019 Associated With <i>Campylobacter jejuni</i> Infection. <i>Neurology: Neuroimmunology and NeuroInflammation</i> , 2021, 8, .	6.7	30
32	Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. <i>Nature</i> , 2021, 590, 290-299.	37.9	1,952
33	Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. <i>Genetics</i> , 2021, 218, .	4.2	69
34	Genome sequencing unveils a regulatory landscape of platelet reactivity. <i>Nature Communications</i> , 2021, 12, .	13.7	46
35	Identification of microbial agents in tissue specimens of ocular and periocular sarcoidosis using a metagenomics approach. <i>F1000Research</i> , 2021, 10, 820.	0.5	6
36	Rapidly fatal infection with <i>Bacillus cereus/thuringiensis</i> : genome assembly of the responsible pathogen and consideration of possibly contributing toxins. <i>Diagnostic Microbiology and Infectious Disease</i> , 2021, 101, 115534.	1.6	8

#	ARTICLE	IF	CITATIONS
37	Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. <i>Bioinformatics</i> , 2020, 36, 1303-1304.	4.7	453
38	Inherited causes of clonal haematopoiesis in 97,691 whole genomes. <i>Nature</i> , 2020, 586, 763-768.	37.9	655
39	Genomic basis of white pine blister rust quantitative disease resistance and its relationship with qualitative resistance. <i>Plant Journal</i> , 2020, 104, 365-376.	6.2	48
40	The tuatara genome reveals ancient features of amniote evolution. <i>Nature</i> , 2020, 584, 403-409.	37.9	124
41	A Reference Genome Sequence for Giant Sequoia. <i>G3: Genes, Genomes, Genetics</i> , 2020, 10, 3907-3919.	1.9	85
42	Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. <i>Microbiome</i> , 2020, 8, .	11.5	233
43	Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies. <i>Genetics</i> , 2020, 216, 599-608.	4.2	49
44	Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. <i>Genome Biology</i> , 2020, 21, .	8.1	217
45	High-quality chromosome-scale assembly of the walnut (<i>Juglans regia</i>) genome. <i>Genome Biology</i> , 2020, 21, .	3.2	120
46	Assembly and annotation of an Ashkenazi human reference genome. <i>Genome Biology</i> , 2020, 21, .	8.1	54
47	The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. <i>PLoS Computational Biology</i> , 2020, 16, e1007981.	3.1	368
48	Pan-genomics in the human genome era. <i>Nature Reviews Genetics</i> , 2020, 21, 243-254.	47.0	310
49	De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2020, 117, 2560-2569.	7.5	105
50	Genome assembly and characterization of a complex zFBED-NLR gene-containing disease resistance locus in Carolina Gold Select rice with Nanopore sequencing. <i>PLoS Genetics</i> , 2020, 16, e1008571.	3.2	138
51	SkewIT: The Skew Index Test for large-scale GC Skew analysis of bacterial genomes. <i>PLoS Computational Biology</i> , 2020, 16, e1008439.	3.1	60
52	The genome of the American groundhog, <i>Marmota monax</i> . <i>F1000Research</i> , 2020, 9, 1137.	0.5	4
53	The Terabase Search Engine: a large-scale relational database of short-read sequences. <i>Bioinformatics</i> , 2019, 35, 665-670.	4.7	8
54	Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. <i>Nature Biotechnology</i> , 2019, 37, 907-915.	29.8	12,746

#	ARTICLE	IF	CITATIONS
55	Whole Genome Sequencing Identifies CRISPLD2 as a Lung Function Gene in Children With Asthma. <i>Chest</i> , 2019, 156, 1068-1079.	1.0	6
56	Human contamination in bacterial genomes has created thousands of spurious proteins. <i>Genome Research</i> , 2019, 29, 954-960.	4.6	150
57	Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. <i>Nature Communications</i> , 2019, 10, .	13.7	89
58	Transcriptome assembly from long-read RNA-seq alignments with StringTie2. <i>Genome Biology</i> , 2019, 20, .	8.1	1,757
59	Genomic architecture of complex traits in loblolly pine. <i>New Phytologist</i> , 2019, 221, 1789-1801.	8.1	71
60	A review of methods and databases for metagenomic classification and assembly. <i>Briefings in Bioinformatics</i> , 2019, 20, 1125-1136.	6.6	486
61	CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. <i>Genome Biology</i> , 2018, 19, .	8.1	312
62	Genomic Variation Among and Within Six <i>Juglans</i> Species. <i>G3: Genes, Genomes, Genetics</i> , 2018, 8, 2153-2165.	1.9	89
63	Removing contaminants from databases of draft genomes. <i>PLoS Computational Biology</i> , 2018, 14, e1006277.	3.1	119
64	Development and Optimization of Metagenomic Next-Generation Sequencing Methods for Cerebrospinal Fluid Diagnostics. <i>Journal of Clinical Microbiology</i> , 2018, 56, .	4.0	95
65	Identifying Corneal Infections in Formalin-Fixed Specimens Using Next Generation Sequencing. , 2018, 59, 280.		63
66	Assembly of a pan-genome from deep sequencing of 910 humans of African descent. <i>Nature Genetics</i> , 2018, 51, 30-35.	25.2	358
67	MUMmer4: A fast and versatile genome alignment system. <i>PLoS Computational Biology</i> , 2018, 14, e1005944.	3.1	2,317
68	Short Read Mapping: An Algorithmic Tour. <i>Proceedings of the IEEE</i> , 2017, 105, 436-458.	9.5	70
69	The novel fusion transcript NR5A2â€KLHL29FT is generated by an insertion at the KLHL29 locus. <i>Cancer</i> , 2017, 123, 1507-1515.	4.0	6
70	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. <i>Genome Research</i> , 2017, 27, 787-792.	4.6	474
71	16GT: a fast and sensitive variant caller using a 16-genotype probabilistic model. <i>GigaScience</i> , 2017, 6, .	3.2	12
72	First Draft Genome Sequence of the Pathogenic Fungus <i>Lomentospora prolificans</i> (Formerly) Tj ETQq0 0 0 rgBT /Overlock 10 Tf 50 57 Td (<i>Scedosporium prolificans</i>) Tj ETQq0 0 0 rgBT /Overlock 10 Tf 50 57 Td (

#	ARTICLE	IF	CITATIONS
73	The Douglas-Fir Genome Sequence Reveals Specialization of the Photosynthetic Apparatus in Pinaceae. <i>G3: Genes, Genomes, Genetics</i> , 2017, 7, 3157-3167.	1.9	120
74	Genome sequence of the progenitor of the wheat D genome <i>Aegilops tauschii</i> . <i>Nature</i> , 2017, 551, 498-502.	37.9	657
75	Presence of Human Hepatitis B Virus in a Cohort of People Who Inject Drugs. <i>Annals of Internal Medicine</i> , 2017, 167, 1-7.	9.7	33
76	The first near-complete assembly of the hexaploid bread wheat genome, <i>Triticum aestivum</i> . <i>GigaScience</i> , 2017, 6, .	3.2	251
77	An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. <i>GigaScience</i> , 2017, 6, .	3.2	97
78	First Draft Assembly and Annotation of the Genome of a California Endemic Oak <i>Quercus lobata</i> (Fagaceae). <i>G3: Genes, Genomes, Genetics</i> , 2016, 6, 3485-3495.	1.9	113
79	Reminder to deposit DNA sequences. <i>Nature</i> , 2016, 533, 179-179.	37.9	10
80	Centrifuge: rapid and sensitive classification of metagenomic sequences. <i>Genome Research</i> , 2016, 26, 1721-1729.	4.6	1,397
81	Sequence of the Sugar Pine Megagenome. <i>Genetics</i> , 2016, 204, 1613-1626.	4.2	196
82	The walnut (<i>Juglans regia</i>) genome sequence reveals diversity in genes coding for the biosynthesis of non-structural polyphenols. <i>Plant Journal</i> , 2016, 87, 507-532.	6.2	271
83	Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system. <i>Neurology: Neuroimmunology and Neuroinflammation</i> , 2016, 3, .	6.7	173
84	Single molecule real-time sequencing of <i>Xanthomonas oryzae</i> genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. <i>Microbial Genomics</i> , 2015, 1, .	2.0	95
85	Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. <i>Nature</i> , 2015, 517, 381-385.	37.9	500
86	StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. <i>Nature Biotechnology</i> , 2015, 33, 290-295.	29.8	12,805
87	HISAT: a fast spliced aligner with low memory requirements. <i>Nature Methods</i> , 2015, 12, 357-360.	24.6	22,430
88	Ballgown bridges the gap between transcriptome assembly and expression analysis. <i>Nature Biotechnology</i> , 2015, 33, 243-246.	29.8	841
89	The genomes of two key bumblebee species with primitive eusocial organization. <i>Genome Biology</i> , 2015, 16, .	12.2	376
90	Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. <i>Genome Research</i> , 2015, 25, 1401-1409.	4.6	107

#	ARTICLE	IF	CITATIONS
91	Re-analysis of metagenomic sequences from acute flaccid myelitis patients reveals alternatives to enterovirus D68 infection. F1000Research, 2015, 4, 180.	0.5	9
92	Re-analysis of metagenomic sequences from acute flaccid myelitis patients reveals alternatives to enterovirus D68 infection. F1000Research, 2015, 4, 180.	0.5	6
93	A new rhesus macaque assembly and annotation for next-generation sequencing analyses. Biology Direct, 2014, 9, .	4.3	176
94	Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. Genetics, 2014, 196, 875-890.	4.2	307
95	Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology, 2014, 15, .	8.1	4,386
96	Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biology, 2014, 15, .	8.1	463
97	DIAMUND : Direct Comparison of Genomes to Detect Mutations. Human Mutation, 2014, 35, 283-288.	4.5	9
98	Genomic Features of a Bumble Bee Symbiont Reflect Its Host Environment. Applied and Environmental Microbiology, 2014, 80, 3793-3803.	3.6	67
99	Unique Features of the Loblolly Pine (Pinus taeda L.) Megagenome Revealed Through Sequence Annotation. Genetics, 2014, 196, 891-909.	4.2	220
100	TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology, 2013, 14, .	8.1	12,644
101	The MaSuRCA genome assembler. Bioinformatics, 2013, 29, 2669-2677.	4.7	1,418
102	GAGE-B: an evaluation of genome assemblers for bacterial organisms. Bioinformatics, 2013, 29, 1718-1725.	4.7	139
103	Hawkeye and AMOS: visualizing and assessing the quality of genome assemblies. Briefings in Bioinformatics, 2013, 14, 213-224.	6.6	56
104	NIH funding: It does support innovators. Nature, 2013, 493, 26-26.	37.9	1
105	Sequestration: inadvertently killing biomedical research to score political points. Genome Biology, 2013, 14, 109.	12.2	2
106	Genome-Guided Transcriptome Assembly in the Age of Next-Generation Sequencing. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2013, 10, 1234-1240.	2.9	18
107	The COMBREX Project: Design, Methodology, and Initial Results. PLoS Biology, 2013, 11, e1001638.	5.0	55
108	EDGE-pro: Estimated Degree of Gene Expression in Prokaryotic Genomes. Evolutionary Bioinformatics, 2013, 9, .	1.3	157

#	ARTICLE	IF	CITATIONS
109	Insights into the Loblolly Pine Genome: Characterization of BAC and Fosmid Sequences. PLoS ONE, 2013, 8, e72439.	2.3	48
110	Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. F1000Research, 2013, 2, 188.	0.5	462
111	Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. F1000Research, 2013, 2, 188.	0.5	320
112	Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. Nucleic Acids Research, 2012, 40, e9-e9.	15.5	182
113	Thousands of missed genes found in bacterial genomes and their analysis with COMBREX. Biology Direct, 2012, 7, 37.	4.3	28
114	Fast gapped-read alignment with Bowtie 2. Nature Methods, 2012, 9, 357-359.	24.6	54,882
115	GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Research, 2012, 22, 557-567.	4.6	634
116	Mis-Assembled "Segmental Duplications" in Two Versions of the Bos taurus Genome. PLoS ONE, 2012, 7, e42680.	2.3	23
117	Two New Complete Genome Sequences Offer Insight into Host and Tissue Specificity of Plant Pathogenic Xanthomonas spp. Journal of Bacteriology, 2011, 193, 5450-5464.	2.9	199
118	TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biology, 2011, 12, .	8.1	739
119	Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. Genome Biology, 2011, 12, .	8.1	62
120	Genome Assembly Has a Major Impact on Gene Content: A Comparison of Annotation in Two Bos Taurus Assemblies. PLoS ONE, 2011, 6, e21400.	2.3	72
121	PhymmBL expanded: confidence scores, custom databases, parallelization and more. Nature Methods, 2011, 8, 367-367.	24.6	107
122	Improving pan-genome annotation using whole genome multiple alignment. BMC Bioinformatics, 2011, 12, .	3.0	39
123	Detection of lineage-specific evolutionary changes among primate species. BMC Bioinformatics, 2011, 12, .	3.0	20
124	COMBREX: a project to accelerate the functional annotation of prokaryotic genomes. Nucleic Acids Research, 2011, 39, D11-D14.	15.5	46
125	FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 2011, 27, 2957-2963.	4.7	14,629
126	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.	7.5	163

#	ARTICLE	IF	CITATIONS
127	Mugsy: fast multiple alignment of closely related whole genomes. <i>Bioinformatics</i> , 2011, 27, 334-342.	4.7	477
128	Repetitive DNA and next-generation sequencing: computational challenges and solutions. <i>Nature Reviews Genetics</i> , 2011, 13, 36-46.	47.0	1,590
129	Assembly of large genomes using second-generation sequencing. <i>Genome Research</i> , 2010, 20, 1165-1173.	4.6	425
130	Clustering metagenomic sequences with interpolated Markov models. <i>BMC Bioinformatics</i> , 2010, 11, .	3.0	98
131	Probing the pan-genome of <i>Listeria monocytogenes</i> : new insights into intraspecific niche expansion and genomic diversification. <i>BMC Genomics</i> , 2010, 11, .	3.3	82
132	Multi-Platform Next-Generation Sequencing of the Domestic Turkey (<i>Meleagris gallopavo</i>): Genome Assembly and Analysis. <i>PLoS Biology</i> , 2010, 8, e1000475.	5.0	371
133	Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. <i>Nature Biotechnology</i> , 2010, 28, 511-515.	29.8	15,291
134	Do-it-yourself genetic testing. <i>Genome Biology</i> , 2010, 11, .	8.1	18
135	Detection and correction of false segmental duplications caused by genome mis-assembly. <i>Genome Biology</i> , 2010, 11, .	8.1	103
136	Between a chicken and a grape: estimating the number of human genes. <i>Genome Biology</i> , 2010, 11, 206.	12.2	196
137	The genome of woodland strawberry (<i>Fragaria vesca</i>). <i>Nature Genetics</i> , 2010, 43, 109-116.	25.2	1,165
138	Recent advances in RNA sequence analysis. <i>F1000 Biology Reports</i> , 2010, 2, .	2.2	13
139	EFFORTS TO DEREGULATE RAINBOW PAPAYA IN JAPAN: MOLECULAR CHARACTERIZATION OF TRANSGENE AND VECTOR INSERTS. <i>Acta Horticulturae</i> , 2010, , 235-240.	0.3	0
140	Between a chicken and a grape: estimating the number of human genes. <i>Genome Biology</i> , 2010, 11, 11.	12.2	0
141	2009 Swine-Origin Influenza A (H1N1) Resembles Previous Influenza Isolates. <i>PLoS ONE</i> , 2009, 4, e6402.	2.3	33
142	OperonDB: a comprehensive database of predicted operons in microbial genomes. <i>Nucleic Acids Research</i> , 2009, 37, D479-D482.	15.5	86
143	Insignia: a DNA signature search web server for diagnostic assay development. <i>Nucleic Acids Research</i> , 2009, 37, W229-W234.	15.5	41
144	Efficient oligonucleotide probe selection for pan-genomic tiling arrays. <i>BMC Bioinformatics</i> , 2009, 10, .	3.0	25

#	ARTICLE	IF	CITATIONS
145	The genome of the blood fluke <i>Schistosoma mansoni</i> . <i>Nature</i> , 2009, 460, 352-358.	37.9	981
146	How to map billions of short reads onto genomes. <i>Nature Biotechnology</i> , 2009, 27, 455-457.	29.8	262
147	Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. <i>Nature Methods</i> , 2009, 6, 673-676.	24.6	551
148	TopHat: discovering splice junctions with RNA-Seq. <i>Bioinformatics</i> , 2009, 25, 1105-1111.	4.7	11,795
149	A whole-genome assembly of the domestic cow, <i>Bos taurus</i> . <i>Genome Biology</i> , 2009, 10, R42.	12.2	1,057
150	Genome-Wide Analysis of Repetitive Elements in Papaya. <i>Tropical Plant Biology</i> , 2008, 1, 191-201.	1.3	24
151	Characterization of Insertion Sites in Rainbow Papaya, the First Commercialized Transgenic Fruit Crop. <i>Tropical Plant Biology</i> , 2008, 1, 293-309.	1.3	29
152	The draft genome of the transgenic tropical fruit tree papaya (<i>Carica papaya</i> Linnaeus). <i>Nature</i> , 2008, 452, 991-996.	37.9	1,017
153	Comparative genomics of the neglected human malaria parasite <i>Plasmodium vivax</i> . <i>Nature</i> , 2008, 455, 757-763.	37.9	792
154	Genome sequence and rapid evolution of the rice pathogen <i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO99A. <i>BMC Genomics</i> , 2008, 9, 204.	3.3	343
155	Genome sequence and rapid evolution of the rice pathogen <i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO99A. <i>BMC Genomics</i> , 2008, 9, 534.	3.3	39
156	Bioinformatics challenges of new sequencing technology. <i>Trends in Genetics</i> , 2008, 24, 142-149.	9.8	454
157	Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. <i>Genome Biology</i> , 2008, 9, R7.	12.2	3,675
158	What are decision trees?. <i>Nature Biotechnology</i> , 2008, 26, 1011-1013.	29.8	616
159	Gene-Boosted Assembly of a Novel Bacterial Genome from Very Short Reads. <i>PLoS Computational Biology</i> , 2008, 4, e1000186.	3.1	46
160	Re-Assembly of the Genome of <i>Francisella tularensis</i> Subsp. <i>holarctica</i> OSU18. <i>PLoS ONE</i> , 2008, 3, e3427.	2.3	9
161	Acquisition and Evolution of Plant Pathogenesis-Associated Gene Clusters and Candidate Determinants of Tissue-Specificity in <i>Xanthomonas</i> . <i>PLoS ONE</i> , 2008, 3, e3828.	2.3	100
162	Comprehensive DNA Signature Discovery and Validation. <i>PLoS Computational Biology</i> , 2007, 3, e98.	3.1	65

#	ARTICLE	IF	CITATIONS
163	A Unified Model Explaining the Offsets of Overlapping and Near-Overlapping Prokaryotic Genes. <i>Molecular Biology and Evolution</i> , 2007, 24, 2091-2098.	4.7	12
164	Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. <i>Genome Biology</i> , 2007, 8, R22.	12.2	471
165	Draft Genome of the Filarial Nematode Parasite <i>Brugia malayi</i> . <i>Science</i> , 2007, 317, 1756-1760.	36.2	591
166	Identifying bacterial genes and endosymbiont DNA with Glimmer. <i>Bioinformatics</i> , 2007, 23, 673-679.	4.7	2,994
167	Genome Analysis Linking Recent European and African Influenza (H5N1) Viruses. <i>Emerging Infectious Diseases</i> , 2007, 13, 713-718.	3.8	199
168	Evolution of genes and genomes on the <i>Drosophila</i> phylogeny. <i>Nature</i> , 2007, 450, 203-218.	37.9	1,992
169	A computational survey of candidate exonic splicing enhancer motifs in the model plant <i>Arabidopsis thaliana</i> . <i>BMC Bioinformatics</i> , 2007, 8, .	3.0	85
170	A phylogenetic generalized hidden Markov model for predicting alternatively spliced exons. <i>Algorithms for Molecular Biology</i> , 2006, 1, .	1.2	11
171	Title is missing!. <i>Genome Biology</i> , 2006, 7, S9.	12.2	62
172	Shared data are key to beating threat from flu. <i>Nature</i> , 2006, 440, 605-605.	37.9	5
173	Physiogenomic resources for rat models of heart, lung and blood disorders. <i>Nature Genetics</i> , 2006, 38, 234-239.	25.2	49
174	It is time to end the patenting of software. <i>Bioinformatics</i> , 2006, 22, 1416-1417.	4.7	2
175	Macronuclear Genome Sequence of the Ciliate <i>Tetrahymena thermophila</i> , a Model Eukaryote. <i>PLoS Biology</i> , 2006, 4, e286.	5.0	702
176	Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. <i>Nature</i> , 2005, 437, 1162-1166.	37.9	438
177	Genomic sequence of the pathogenic and allergenic filamentous fungus <i>Aspergillus fumigatus</i> . <i>Nature</i> , 2005, 438, 1151-1156.	37.9	1,325
178	Beware of mis-assembled genomes. <i>Bioinformatics</i> , 2005, 21, 4320-4321.	4.7	172
179	Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. <i>Genome Research</i> , 2005, 15, 1284-1291.	4.6	74
180	JIGSAW: integration of multiple sources of evidence for gene prediction. <i>Bioinformatics</i> , 2005, 21, 3596-3603.	4.7	151

#	ARTICLE	IF	CITATIONS
181	Efficient implementation of a generalized pair hidden Markov model for comparative gene finding. <i>Bioinformatics</i> , 2005, 21, 1782-1788.	4.7	39
182	Whole-Genome Analysis of Human Influenza A Virus Reveals Multiple Persistent Lineages and Reassortment among Recent H3N2 Viruses. <i>PLoS Biology</i> , 2005, 3, e300.	5.0	353
183	Title is missing!. <i>Genome Biology</i> , 2005, 6, R23.	12.2	131
184	Automated correction of genome sequence errors. <i>Nucleic Acids Research</i> , 2004, 32, 562-569.	15.5	42
185	TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. <i>Bioinformatics</i> , 2004, 20, 2878-2879.	4.7	1,813
186	Comparative genome assembly. <i>Briefings in Bioinformatics</i> , 2004, 5, 237-248.	6.6	199
187	Genomic Insights into Methanotrophy: The Complete Genome Sequence of <i>Methylococcus capsulatus</i> (Bath). <i>PLoS Biology</i> , 2004, 2, e303.	5.0	294
188	DAGchainer: a tool for mining segmental genome duplications and synteny. <i>Bioinformatics</i> , 2004, 20, 3643-3646.	4.7	372
189	Gene synteny and evolution of genome architecture in trypanosomatids. <i>Molecular and Biochemical Parasitology</i> , 2004, 134, 183-191.	1.3	94
190	Title is missing!. <i>BMC Bioinformatics</i> , 2004, 5, 206.	3.0	25
191	Title is missing!. <i>Genome Biology</i> , 2004, 5, R61.	12.2	186
192	Insight into the genome of <i>Aspergillus fumigatus</i> : analysis of a 922kb region encompassing the nitrate assimilation gene cluster. <i>Fungal Genetics and Biology</i> , 2004, 41, 443-453.	2.1	63
193	The Genome Assembly Archive: A New Public Resource. <i>PLoS Biology</i> , 2004, 2, e285.	5.0	30
194	Title is missing!. <i>Plant Molecular Biology</i> , 2003, 51, 859-866.	3.2	117
195	Unrestricted free access works and must continue. <i>Nature</i> , 2003, 422, 801-801.	37.9	7
196	Using MUMmer to Identify Similar Regions in Large Sequence Sets. <i>Current Protocols in Bioinformatics</i> , 2003, 00, .	3.3	552
197	The genome sequence of <i>Bacillus anthracis</i> Ames and comparison to closely related bacteria. <i>Nature</i> , 2003, 423, 81-86.	37.9	779
198	The sequence and analysis of <i>Trypanosoma brucei</i> chromosome II. <i>Nucleic Acids Research</i> , 2003, 31, 4856-4863.	15.5	59

#	ARTICLE	IF	CITATIONS
199	Computational Discovery of Internal Micro-Exons. <i>Genome Research</i> , 2003, 13, 1216-1221.	4.6	69
200	GlimmerM, Economy and Unveil: three ab initio eukaryotic gene finders. <i>Nucleic Acids Research</i> , 2003, 31, 3601-3604.	15.5	63
201	Genome sequence of <i>Chlamydomonas reinhardtii</i> (Chlamydomonas reinhardtii GPC): examining the role of niche-specific genes in the evolution of the Chlamydomonadales. <i>Nucleic Acids Research</i> , 2003, 31, 2134-2147.	15.5	267
202	Using GlimmerM to Find Genes in Eukaryotic Genomes. <i>Current Protocols in Bioinformatics</i> , 2003, 00, .	3.3	5
203	Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. <i>Nucleic Acids Research</i> , 2003, 31, 5654-5666.	15.5	2,200
204	Hierarchical Scaffolding With Bambus. <i>Genome Research</i> , 2003, 14, 149-159.	4.6	175
205	Computational Gene Prediction Using Multiple Sources of Evidence. <i>Genome Research</i> , 2003, 14, 142-148.	4.6	115
206	A Comparison of Whole-Genome Shotgun-Derived Mouse Chromosome 16 and the Human Genome. <i>Science</i> , 2002, 296, 1661-1671.	36.2	349
207	Comparative Genome and Proteome Analysis of <i>Anopheles gambiae</i> and <i>Drosophila melanogaster</i> . <i>Science</i> , 2002, 298, 149-159.	36.2	542
208	The <i>Brucella suis</i> genome reveals fundamental similarities between animal and plant pathogens and symbionts. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2002, 99, 13148-13153.	7.5	433
209	Whole-Genome Comparison of <i>Mycobacterium tuberculosis</i> Clinical and Laboratory Strains. <i>Journal of Bacteriology</i> , 2002, 184, 5479-5490.	2.9	662
210	Fast algorithms for large-scale genome alignment and comparison. <i>Nucleic Acids Research</i> , 2002, 30, 2478-2483.	15.5	989
211	The Genome Sequence of the Malaria Mosquito <i>Anopheles gambiae</i> . <i>Science</i> , 2002, 298, 129-149.	36.2	1,916
212	Title is missing!. <i>Genome Biology</i> , 2002, 3, research0029.1.	12.2	156
213	Contamination in the Draft of the Human Genome Masquerades As Lateral Gene Transfer. <i>DNA Sequence</i> , 2002, 13, 75-76.	0.5	18
214	A preliminary comparison of the mouse and human genomes. <i>International Congress Series</i> , 2002, 1246, 169-181.	0.2	2
215	Genome sequence assembly: algorithms and issues. <i>Computer</i> , 2002, 35, 47-54.	0.7	62
216	Sequence of <i>Plasmodium falciparum</i> chromosomes 2, 10, 11 and 14. <i>Nature</i> , 2002, 419, 531-534.	37.9	170

#	ARTICLE	IF	CITATIONS
217	Genome sequence of the human malaria parasite <i>Plasmodium falciparum</i> . <i>Nature</i> , 2002, 419, 498-511.	37.9	4,102
218	Genome sequence and comparative analysis of the model rodent malaria parasite <i>Plasmodium yoelii yoelii</i> . <i>Nature</i> , 2002, 419, 512-519.	37.9	676
219	Title is missing!. <i>Plant Molecular Biology</i> , 2002, 48, 39-48.	3.2	42
220	Title is missing!. <i>Genome Biology</i> , 2001, 2, research0027.1.	12.2	123
221	GeneSplicer: a new computational method for splice site prediction. <i>Nucleic Acids Research</i> , 2001, 29, 1185-1190.	15.5	560
222	The Sequence of the Human Genome. <i>Science</i> , 2001, 291, 1304-1351.	36.2	13,090
223	Complete genome sequence of <i>Caulobacter crescentus</i> . <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2001, 98, 4136-4141.	7.5	507
224	Sequence and analysis of the <i>Arabidopsis</i> genome. <i>Current Opinion in Plant Biology</i> , 2001, 4, 105-110.	7.2	68
225	Understanding the Adaptation of <i>Halobacterium</i> Species NRC-1 to Its Extreme Environment through Computational Analysis of Its Genome Sequence. <i>Genome Research</i> , 2001, 11, 1641-1650.	4.6	309
226	Prediction of operons in microbial genomes. <i>Nucleic Acids Research</i> , 2001, 29, 1216-1221.	15.5	261
227	A probabilistic method for identifying start codons in bacterial genomes. <i>Bioinformatics</i> , 2001, 17, 1123-1130.	4.7	184
228	Rice Bioinformatics. Analysis of Rice Sequence Data and Leveraging the Data to Other Plant Species. <i>Plant Physiology</i> , 2001, 125, 1166-1174.	5.5	47
229	Gene Index analysis of the human genome estimates approximately 120,000 genes. <i>Nature Genetics</i> , 2000, 25, 239-240.	25.2	265
230	Finding genes in <i>Plasmodium falciparum</i> . <i>Nature</i> , 2000, 404, 34-34.	37.9	19
231	DNA sequence of both chromosomes of the cholera pathogen <i>Vibrio cholerae</i> . <i>Nature</i> , 2000, 406, 477-483.	37.9	1,802
232	Microbial genome sequencing. <i>Nature</i> , 2000, 406, 799-803.	37.9	170
233	Sequence and analysis of chromosome 1 of the plant <i>Arabidopsis thaliana</i> . <i>Nature</i> , 2000, 408, 816-820.	37.9	251
234	Sequence and analysis of chromosome 3 of the plant <i>Arabidopsis thaliana</i> . <i>Nature</i> , 2000, 408, 820-823.	37.9	200

#	ARTICLE	IF	CITATIONS
235	Genome sequences of <i>Chlamydia trachomatis</i> MoPn and <i>Chlamydia pneumoniae</i> AR39. <i>Nucleic Acids Research</i> , 2000, 28, 1397-1406.	15.5	708
236	An optimized protocol for analysis of EST sequences. <i>Nucleic Acids Research</i> , 2000, 28, 3657-3665.	15.5	121
237	Prediction of transcription terminators in bacterial genomes 1 Edited by F. E. Cohen. <i>Journal of Molecular Biology</i> , 2000, 301, 27-33.	4.1	238
238	Title is missing!. <i>Genome Biology</i> , 2000, 1, research0011.1.	12.2	302
239	Alignment of whole genomes. <i>Nucleic Acids Research</i> , 1999, 27, 2369-2376.	15.5	798
240	Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of <i>Thermotoga maritima</i> . <i>Nature</i> , 1999, 399, 323-329.	37.9	1,428
241	Sequence and analysis of chromosome 2 of the plant <i>Arabidopsis thaliana</i> . <i>Nature</i> , 1999, 402, 761-768.	37.9	736
242	DNA uptake signal sequences in naturally transformable bacteria. <i>Research in Microbiology</i> , 1999, 150, 603-616.	3.0	120
243	Gene discovery in DNA sequences. <i>IEEE Intelligent Systems</i> , 1999, 14, 44-48.	0.0	10
244	Improved microbial gene identification with GLIMMER. <i>Nucleic Acids Research</i> , 1999, 27, 4636-4641.	15.5	2,359
245	Interpolated Markov Models for Eukaryotic Gene Finding. <i>Genomics</i> , 1999, 59, 24-31.	2.8	192
246	Optimized Multiplex PCR: Efficiently Closing a Whole-Genome Shotgun Sequencing Project. <i>Genomics</i> , 1999, 62, 500-507.	2.8	132
247	A probabilistic framework for memory-based reasoning. <i>Artificial Intelligence</i> , 1998, 104, 287-311.	2.8	35
248	Skewed oligomers and origins of replication. <i>Gene</i> , 1998, 217, 57-67.	2.3	172
249	Complete Genome Sequence of <i>Treponema pallidum</i> , the Syphilis Spirochete. <i>Science</i> , 1998, 281, 375-388.	36.2	1,025
250	A Decision Tree System for Finding Genes in DNA. <i>Journal of Computational Biology</i> , 1998, 5, 667-680.	1.5	108
251	Microbial gene identification using interpolated Markov models. <i>Nucleic Acids Research</i> , 1998, 26, 544-548.	15.5	921
252	A method for identifying splice sites and translational start sites in eukaryotic mRNA. <i>Bioinformatics</i> , 1997, 13, 365-376.	4.7	80

#	ARTICLE	IF	CITATIONS
253	Finding Genes in DNA with a Hidden Markov Model. <i>Journal of Computational Biology</i> , 1997, 4, 127-141.	1.5	205
254	Genomic sequence of a Lyme disease spirochaete, <i>Borrelia burgdorferi</i> . <i>Nature</i> , 1997, 390, 580-586.	37.9	2,074
255	Title is missing!. , 1997, 11, 343-370.		20
256	Title is missing!. <i>Data Mining and Knowledge Discovery</i> , 1997, 1, 317-328.	2.7	678
257	Testing simple polygons. <i>Computational Geometry: Theory and Applications</i> , 1997, 8, 97-114.	0.5	2
258	Testing orthogonal shapes. <i>Computational Geometry: Theory and Applications</i> , 1995, 5, 33-49.	0.5	3
259	Decision trees for automated identification of cosmic-ray hits in Hubble Space Telescope images. <i>Publications of the Astronomical Society of the Pacific</i> , 1995, 107, 279.	7.1	61
260	Locating Protein Coding Regions in Human DNA Using a Decision Tree Algorithm. <i>Journal of Computational Biology</i> , 1995, 2, 473-485.	1.5	68
261	Best-case results for nearest-neighbor learning. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 1995, 17, 599-608.	11.2	20
262	A weighted nearest neighbor algorithm for learning with symbolic features. <i>Machine Learning</i> , 1993, 10, 57-78.	1.7	431
263	Title is missing!. <i>Machine Learning</i> , 1993, 10, 57-78.	1.7	342
264	Predicting protein secondary structure with a nearest-neighbor algorithm. <i>Journal of Molecular Biology</i> , 1992, 227, 371-374.	4.1	84
265	A nearest hyperrectangle learning method. <i>Machine Learning</i> , 1991, 6, 251-276.	1.7	224
266	Title is missing!. <i>Machine Learning</i> , 1991, 6, 251-276.	1.7	181
267	Managing information for concurrent engineering: Challenges and barriers. <i>Research in Engineering Design - Theory, Applications, and Concurrent Engineering</i> , 1990, 2, 35-52.	1.4	34
268	Cretaceous dinosaur bone contains recent organic material and provides an environment conducive to microbial communities. <i>ELife</i> , 0, 8, .	1.6	52
269	Bracken: estimating species abundance in metagenomics data. <i>PeerJ Computer Science</i> , 0, 3, e104.	0.0	1,775
270	Unexpected cross-species contamination in genome sequencing projects. <i>PeerJ</i> , 0, 2, e675.	0.0	166

#	ARTICLE	IF	CITATIONS
271	Arioc: high-throughput read alignment with GPU-accelerated exploration of the seed-and-extend search space. PeerJ, 0, 3, e808.	0.0	38
272	Structure-guided isoform identification for the human transcriptome. ELife, 0, 11, .	1.6	36
273	OpenSpliceAI provides an efficient modular implementation of SpliceAI enabling easy retraining across nonhuman species. ELife, 0, 14, .	1.6	1
274	Translon: a single term for translated regions. Nature Methods, 0, 22, 2002-2006.	24.6	4
275	Comprehensive analysis of microbial content in whole-genome sequencing samples from The Cancer Genome Atlas project. Science Translational Medicine, 0, 17, .	12.5	16
276	OpenSpliceAI provides an efficient modular implementation of SpliceAI enabling easy retraining across nonhuman species. ELife, 0, 14, .	1.6	0