

# Pavel A Pevzner

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/2594793/publications.pdf>

Version: 2024-02-01

116  
papers

44,675  
citations

30047

54  
h-index

26591

107  
g-index

132  
all docs

132  
docs citations

132  
times ranked

46606  
citing authors

#	ARTICLE	IF	CITATIONS
1	SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. <i>Journal of Computational Biology</i> , 2012, 19, 455-477.	0.8	20,193
2	Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. <i>Nature Biotechnology</i> , 2016, 34, 828-837.	9.4	2,802
3	metaSPAdes: a new versatile metagenomic assembler. <i>Genome Research</i> , 2017, 27, 824-834.	2.4	2,779
4	Assembly of long, error-prone reads using repeat graphs. <i>Nature Biotechnology</i> , 2019, 37, 540-546.	9.4	2,730
5	Assembling Single-Cell Genomes and Mini-Metagenomes From Chimeric MDA Products. <i>Journal of Computational Biology</i> , 2013, 20, 714-737.	0.8	1,235
6	The complete sequence of a human genome. <i>Science</i> , 2022, 376, 44-53.	6.0	1,222
7	MS-GF+ makes progress towards a universal database search tool for proteomics. <i>Nature Communications</i> , 2014, 5, 5277.	5.8	945
8	PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. <i>Analytical Chemistry</i> , 2005, 77, 964-973.	3.2	615
9	American Gut: an Open Platform for Citizen Science Microbiome Research. <i>MSystems</i> , 2018, 3, .	1.7	604
10	Telomere-to-telomere assembly of a complete human X chromosome. <i>Nature</i> , 2020, 585, 79-84.	13.7	549
11	InsPecT: Identification of Posttranslationally Modified Peptides from Tandem Mass Spectra. <i>Analytical Chemistry</i> , 2005, 77, 4626-4639.	3.2	546
12	De Novo Peptide Sequencing via Tandem Mass Spectrometry. <i>Journal of Computational Biology</i> , 1999, 6, 327-342.	0.8	529
13	How to apply de Bruijn graphs to genome assembly. <i>Nature Biotechnology</i> , 2011, 29, 987-991.	9.4	470
14	plasmidSPAdes: assembling plasmids from whole genome sequencing data. <i>Bioinformatics</i> , 2016, 32, 3380-3387.	1.8	468
15	<scp>hybrid</scp>SPA<scp>des</scp>: an algorithm for hybrid assembly of short and long reads. <i>Bioinformatics</i> , 2016, 32, 1009-1015.	1.8	463
16	metaFlye: scalable long-read metagenome assembly using repeat graphs. <i>Nature Methods</i> , 2020, 17, 1103-1110.	9.0	430
17	Spectral Probabilities and Generating Functions of Tandem Mass Spectra: A Strike against Decoy Databases. <i>Journal of Proteome Research</i> , 2008, 7, 3354-3363.	1.8	426
18	Short read fragment assembly of bacterial genomes. <i>Genome Research</i> , 2008, 18, 324-330.	2.4	371

#	ARTICLE	IF	CITATIONS
19	Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2003, 100, 7672-7677.	3.3	322
20	Genome Rearrangements in Mammalian Evolution: Lessons From Human and Mouse Genomes. <i>Genome Research</i> , 2003, 13, 37-45.	2.4	302
21	Identification of post-translational modifications by blind search of mass spectra. <i>Nature Biotechnology</i> , 2005, 23, 1562-1567.	9.4	247
22	l-Tuple DNA Sequencing: Computer Analysis. <i>Journal of Biomolecular Structure and Dynamics</i> , 1989, 7, 63-73.	2.0	243
23	Clustering Millions of Tandem Mass Spectra. <i>Journal of Proteome Research</i> , 2008, 7, 113-122.	1.8	230
24	Assembly of long error-prone reads using de Bruijn graphs. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2016, 113, E8396-E8405.	3.3	230
25	De novo fragment assembly with short mate-paired reads: Does the read length matter?. <i>Genome Research</i> , 2009, 19, 336-346.	2.4	220
26	Dereplication of microbial metabolites through database search of mass spectra. <i>Nature Communications</i> , 2018, 9, 4035.	5.8	220
27	Complete genomic and epigenetic maps of human centromeres. <i>Science</i> , 2022, 376, eabl4178.	6.0	204
28	De Novo Repeat Classification and Fragment Assembly. <i>Genome Research</i> , 2004, 14, 1786-1796.	2.4	195
29	Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. <i>Nature Genetics</i> , 2007, 39, 1361-1368.	9.4	192
30	Characterization of Cyanobacterial Hydrocarbon Composition and Distribution of Biosynthetic Pathways. <i>PLoS ONE</i> , 2014, 9, e85140.	1.1	190
31	Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. <i>Genome Research</i> , 2004, 14, 2245-2252.	2.4	184
32	Dereplication of peptidic natural products through database search of mass spectra. <i>Nature Chemical Biology</i> , 2017, 13, 30-37.	3.9	184
33	Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of <i>Bacillus subtilis</i> . <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2010, 107, 16286-16290.	3.3	179
34	Fragment assembly with short reads. <i>Bioinformatics</i> , 2004, 20, 2067-2074.	1.8	166
35	Protein identification by spectral networks analysis. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2007, 104, 6140-6145.	3.3	157
36	<scp>Metaviral</scp> <scp>SPAdes</scp>: assembly of viruses from metagenomic data. <i>Bioinformatics</i> , 2020, 36, 4126-4129.	1.8	149

#	ARTICLE	IF	CITATIONS
37	Automated Genome Mining of Ribosomal Peptide Natural Products. <i>ACS Chemical Biology</i> , 2014, 9, 1545-1551.	1.6	133
38	Mutation-Tolerant Protein Identification by Mass Spectrometry. <i>Journal of Computational Biology</i> , 2000, 7, 777-787.	0.8	128
39	Automated de novo protein sequencing of monoclonal antibodies. <i>Nature Biotechnology</i> , 2008, 26, 1336-1338.	9.4	114
40	Plasmid detection and assembly in genomic and metagenomic data sets. <i>Genome Research</i> , 2019, 29, 961-968.	2.4	108
41	ExSPANder: a universal repeat resolver for DNA fragment assembly. <i>Bioinformatics</i> , 2014, 30, i293-i301.	1.8	103
42	Thousands of previously unknown phages discovered in whole-community human gut metagenomes. <i>Microbiome</i> , 2021, 9, 78.	4.9	101
43	Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. <i>Nature Biotechnology</i> , 2022, 40, 711-719.	9.4	99
44	Fragment assembly with double-barreled data. <i>Bioinformatics</i> , 2001, 17, S225-S233.	1.8	97
45	Spectral archives: extending spectral libraries to analyze both identified and unidentified spectra. <i>Nature Methods</i> , 2011, 8, 587-591.	9.0	86
46	Dereplication and de novo sequencing of nonribosomal peptides. <i>Nature Methods</i> , 2009, 6, 596-599.	9.0	81
47	NRPquest: Coupling Mass Spectrometry and Genome Mining for Nonribosomal Peptide Discovery. <i>Journal of Natural Products</i> , 2014, 77, 1902-1909.	1.5	81
48	Shotgun Protein Sequencing. <i>Molecular and Cellular Proteomics</i> , 2007, 6, 1123-1134.	2.5	80
49	Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative CrAss-like phages with unique genomic features. <i>Nature Communications</i> , 2021, 12, 1044.	5.8	80
50	DRIMM-Synteny: decomposing genomes into evolutionary conserved segments. <i>Bioinformatics</i> , 2010, 26, 2509-2516.	1.8	78
51	Computing Has Changed Biology—Biology Education Must Catch Up. <i>Science</i> , 2009, 325, 541-542.	6.0	76
52	Interpreting Top-Down Mass Spectra Using Spectral Alignment. <i>Analytical Chemistry</i> , 2008, 80, 2499-2505.	3.2	71
53	Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. <i>Nature Microbiology</i> , 2018, 3, 319-327.	5.9	71
54	Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. <i>Genome Research</i> , 2020, 30, 898-909.	2.4	68

#	ARTICLE	IF	CITATIONS
55	Comment on "Protein Sequences from Mastodon and <i>Tyrannosaurus rex</i> Revealed by Mass Spectrometry". <i>Science</i> , 2008, 321, 1040-1040.	6.0	62
56	Paired de Bruijn Graphs: A Novel Approach for Incorporating Mate Pair Information into Genome Assemblers. <i>Journal of Computational Biology</i> , 2011, 18, 1625-1634.	0.8	60
57	Dereplication, sequencing and identification of peptidic natural products: from genome mining to peptidogenomics to spectral networks. <i>Natural Product Reports</i> , 2016, 33, 73-86.	5.2	59
58	SEQuel: improving the accuracy of genome assemblies. <i>Bioinformatics</i> , 2012, 28, i188-i196.	1.8	56
59	BiosyntheticSPAdes: reconstructing biosynthetic gene clusters from assembly graphs. <i>Genome Research</i> , 2019, 29, 1352-1362.	2.4	55
60	Single-molecule protein identification by sub-nanopore sensors. <i>PLoS Computational Biology</i> , 2017, 13, e1005356.	1.5	52
61	Top-down analysis of immunoglobulin G isotypes 1 and 2 with electron transfer dissociation on a high-field Orbitrap mass spectrometer. <i>Journal of Proteomics</i> , 2017, 159, 67-76.	1.2	47
62	Optimizing sequencing protocols for leaderboard metagenomics by combining long and short reads. <i>Genome Biology</i> , 2019, 20, 226.	3.8	47
63	MetaMiner: A Scalable Peptidogenomics Approach for Discovery of Ribosomal Peptide Natural Products with Blind Modifications from Microbial Communities. <i>Cell Systems</i> , 2019, 9, 600-608.e4.	2.9	46
64	Automated assembly of centromeres from ultra-long error-prone reads. <i>Nature Biotechnology</i> , 2020, 38, 1309-1316.	9.4	45
65	Single Cells within the Puerto Rico Trench Suggest Hadal Adaptation of Microbial Lineages. <i>Applied and Environmental Microbiology</i> , 2015, 81, 8265-8276.	1.4	43
66	IgRepertoireConstructor: a novel algorithm for antibody repertoire construction and immunoproteogenomics analysis. <i>Bioinformatics</i> , 2015, 31, i53-i61.	1.8	42
67	Multiplex de Bruijn graphs enable genome assembly from long, high-fidelity reads. <i>Nature Biotechnology</i> , 2022, 40, 1075-1081.	9.4	41
68	Assembling short reads from jumping libraries with large insert sizes. <i>Bioinformatics</i> , 2015, 31, 3262-3268.	1.8	40
69	TruSPAdes: barcode assembly of TruSeq synthetic long reads. <i>Nature Methods</i> , 2016, 13, 248-250.	9.0	40
70	TandemTools: mapping long reads and assessing/improving assembly quality in extra-long tandem repeats. <i>Bioinformatics</i> , 2020, 36, i75-i83.	1.8	40
71	Multiplex De Novo Sequencing of Peptide Antibiotics. <i>Journal of Computational Biology</i> , 2011, 18, 1371-1381.	0.8	39
72	Cycloquest: Identification of Cyclopeptides via Database Search of Their Mass Spectra against Genome Databases. <i>Journal of Proteome Research</i> , 2011, 10, 4505-4512.	1.8	38

#	ARTICLE	IF	CITATIONS
73	Detection and analysis of ancient segmental duplications in mammalian genomes. <i>Genome Research</i> , 2018, 28, 901-909.	2.4	38
74	Sequencing cyclic peptides by multistage mass spectrometry. <i>Proteomics</i> , 2011, 11, 3642-3650.	1.3	37
75	Reconstructing Antibody Repertoires from Error-Prone Immunosequencing Reads. <i>Journal of Immunology</i> , 2017, 199, 3369-3380.	0.4	37
76	Advanced Proteogenomic Analysis Reveals Multiple Peptide Mutations and Complex Immunoglobulin Peptides in Colon Cancer. <i>Journal of Proteome Research</i> , 2015, 14, 3555-3567.	1.8	36
77	Spatial Molecular Architecture of the Microbial Community of a <i>Peltigera</i> Lichen. <i>MSystems</i> , 2016, 1, .	1.7	36
78	A family of small, cyclic peptides buried in prealbumin since the Eocene epoch. <i>Plant Direct</i> , 2018, 2, e00042.	0.8	32
79	De Novo Sequencing of Peptides from Top-Down Tandem Mass Spectra. <i>Journal of Proteome Research</i> , 2015, 14, 4450-4462.	1.8	31
80	Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. <i>Nature Communications</i> , 2021, 12, 3225.	5.8	31
81	De Novo Peptide Sequencing Reveals Many Cyclopeptides in the Human Gut and Other Environments. <i>Cell Systems</i> , 2020, 10, 99-108.e5.	2.9	28
82	The string decomposition problem and its applications to centromere analysis and assembly. <i>Bioinformatics</i> , 2020, 36, i93-i101.	1.8	28
83	Multi-spectra peptide sequencing and its applications to multistage mass spectrometry. <i>Bioinformatics</i> , 2008, 24, i416-i423.	1.8	25
84	How bioinformatics and open data can boost basic science in countries and universities with limited resources. <i>Nature Biotechnology</i> , 2019, 37, 324-326.	9.4	25
85	De novo Inference of Diversity Genes and Analysis of Non-canonical V(DD)J Recombination in Immunoglobulins. <i>Frontiers in Immunology</i> , 2019, 10, 987.	2.2	22
86	A New Approach to Evaluating Statistical Significance of Spectral Identifications. <i>Journal of Proteome Research</i> , 2013, 12, 1560-1568.	1.8	20
87	A Multi-Omics Characterization of the Natural Product Potential of Tropical Filamentous Marine Cyanobacteria. <i>Marine Drugs</i> , 2021, 19, 20.	2.2	19
88	De novo repeat classification and fragment assembly. , 2004, , .		16
89	V(DD)J recombination is an important and evolutionarily conserved mechanism for generating antibodies with unusually long CDR3s. <i>Genome Research</i> , 2020, 30, 1547-1558.	2.4	15
90	CentromereArchitect: inference and analysis of the architecture of centromeres. <i>Bioinformatics</i> , 2021, 37, i196-i204.	1.8	15

#	ARTICLE	IF	CITATIONS
91	Pathset Graphs: A Novel Approach for Comprehensive Utilization of Paired Reads in Genome Assembly. <i>Journal of Computational Biology</i> , 2013, 20, 359-371.	0.8	14
92	What is the difference between the breakpoint graph and the de Bruijn graph?. <i>BMC Genomics</i> , 2014, 15, S6.	1.2	14
93	Top-down analysis of protein samples by <i>de novo</i> sequencing techniques. <i>Bioinformatics</i> , 2016, 32, 2753-2759.	1.8	14
94	SpectroGene: A Tool for Proteogenomic Annotations Using Top-Down Spectra. <i>Journal of Proteome Research</i> , 2016, 15, 144-151.	1.8	14
95	GENOME-WIDE ANALYSIS OF BACTERIAL PROMOTER REGIONS. , 2002, , .		14
96	Paired de Bruijn Graphs: A Novel Approach for Incorporating Mate Pair Information into Genome Assemblers. <i>Lecture Notes in Computer Science</i> , 2011, , 238-251.	1.0	13
97	Trace Reconstruction Problems in Computational Biology. <i>IEEE Transactions on Information Theory</i> , 2021, 67, 3295-3314.	1.5	13
98	viralFlye: assembling viruses and identifying their hosts from long-read metagenomics data. <i>Genome Biology</i> , 2022, 23, 57.	3.8	11
99	Automated annotation of human centromeres with HORmon. <i>Genome Research</i> , 2022, , gr.276362.121.	2.4	11
100	Joint Analysis of Long and Short Reads Enables Accurate Estimates of Microbiome Complexity. <i>Cell Systems</i> , 2018, 7, 192-200.e3.	2.9	9
101	Automated analysis of immunosequencing datasets reveals novel immunoglobulin D genes across diverse species. <i>PLoS Computational Biology</i> , 2020, 16, e1007837.	1.5	9
102	The Antibody Repertoire of Colorectal Cancer. <i>Molecular and Cellular Proteomics</i> , 2017, 16, 2111-2124.	2.5	8
103	Immunoglobulin Classification Using the Colored Antibody Graph. <i>Journal of Computational Biology</i> , 2016, 23, 483-494.	0.8	7
104	Variations in antibody repertoires correlate with vaccine responses. <i>Genome Research</i> , 2022, 32, 791-804.	2.4	7
105	EULER-PCR: FINISHING EXPERIMENTS FOR REPEAT RESOLUTION. , 2001, , .		6
106	Pathset Graphs: A Novel Approach for Comprehensive Utilization of Paired Reads in Genome Assembly. <i>Lecture Notes in Computer Science</i> , 2012, , 200-212.	1.0	3
107	ORFograph: search for novel insecticidal protein genes in genomic and metagenomic assembly graphs. <i>Microbiome</i> , 2021, 9, 149.	4.9	3
108	MosaicFlye: Resolving Long Mosaic Repeats Using Long Reads. <i>Lecture Notes in Computer Science</i> , 2020, , 226-228.	1.0	1

#	ARTICLE	IF	CITATIONS
109	DNA STATISTICS, OVERLAPPING WORD PARADOX AND CONWAY EQUATION. , 1993, , .		0
110	Educating Biologists in the 21st Century: Bioinformatics Scientists versus Bioinformatics Technicians. , 2007, , 1-5.		0
111	Preface: 2 <sup>nd</sup> Satellite Meeting on Bioinformatics Education, Research in Computational Molecular Biology (RECOMB-BE 2010). Journal of Computational Biology, 2011, 18, 865-865.	0.8	0
112	Keynote: De novo sequencing of novel peptide antibiotics by tandem mass spectrometry. , 2012, , .		0
113	Are There Rearrangement Hotspots in the Human Genome?. PLoS Computational Biology, 2005, preprint, e209.	1.5	0
114	PROTEIN IDENTIFICATION VIA SPECTRAL NETWORKS ANALYSIS. , 2007, , .		0
115	Identifying protease cleavage sites by label-free mass spectrometry. FASEB Journal, 2010, 24, 905.4.	0.2	0
116	Michael Waterman's Contributions to Computational Biology and Bioinformatics. Journal of Computational Biology, 0, , .	0.8	0