Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

J Gregory Caporaso

Northern Arizona University

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Criterion 5 (original contributions of major significance). It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

29	29	3	83
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

93.1% independent of 29 classified citing papers

Citation type	Count
Independent	27
Self-citation	1
Co-author	0
Same-institution	1

0 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher developed QIIME, a foundational software framework that enabled the analysis of high-throughput community sequencing data, establishing a standard tool for microbiome research.

CLAIM: The researcher's primary contribution is the development of QIIME, a software framework designed to allow the analysis of high-throughput community sequencing data, as detailed in their seminal 2010 paper. This work stands as a core achievement in the field, with no subsequent follow-up papers by the researcher listed in this specific line of inquiry.

ORIGINALITY: The titles indicate that this work addressed the critical need for accessible tools to process complex, high-throughput sequencing data. By providing a dedicated framework for community sequencing analysis, the researcher appears to have filled a significant methodological gap, enabling researchers to interpret large-scale biological datasets that were previously difficult to manage.

SIGNIFICANCE: The impact of this contribution is evidenced by its extensive citation record, with the core paper accumulating tens of thousands of citations. Furthermore, analysis of citing literature reveals that the vast majority of citations originate from independent researchers, suggesting that QIIME has been widely adopted across the global scientific community as a standard tool rather than being limited to the researcher's immediate network.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 9 **· 1 flagged influential by Semantic Scholar**

CORE PAPER

QIIME allows analysis of high-throughput community sequencing data

2010 · 38,776 citations (GS)

Field-normalised: 32,386 Semantic Scholar citations place it in the top 1% of Biology papers from 2010 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	microeco: an R package for data mining in microbial community ecology (2021)	Chengdu Institute of Biology, Chinese Academy of Sciences, Fujian Agriculture and Forestry University, Henan University of Technology	China	—
2	The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update (2018)	Albert-Ludwigs-University, Albert-Ludwigs-University Freiburg, Cleveland Clinic	France, Germany, Netherlands	**Methodology**
3	The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. (2019)	Global Biodiversity Information Facility, Jacobs University Bremen and MPI for Marine Microbiology, National Museum of Natural History, Smithsonian Institution	Denmark, Estonia, Germany	Background
4	Gut-microbiota-targeted diets modulate human immune status (2021)	Chan Zuckerberg Biohub, Stanford School of Medicine, Stanford University	United States	—
5	Next-Generation Sequencing Technology: Current Trends and Advancements (2023)	miBiome Therapeutics, UMass Chan Medical School	India, United States	**Methodology**

No.	Citing paper	Citing institution(s)	Country	S2
6	The human skin microbiome (2018)	National Human Genome Research Institute, National Institutes of Health, National Institute of Allergy and Infectious Diseases, National Institutes of Health	United States	—
7	Oral administration of Blautia wexlerae ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota (2022)	Shunan City Shinnanyo Hospital, Waseda University	Japan	—
8	Diversifying crop rotation increases food production, reduces net greenhouse gas emissions and improves soil health (2024)	Agriculture and Agri-Food Canada, China Agricultural University, Chinese Academy of Agricultural Sciences	Australia, Canada, China	**Methodology**
9	Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients (2018)	European Institute of Oncology IRCCS, Gustave Roussy Cancer Campus, INRAe	France, Italy, Japan	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

> **METHODOLOGY** The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update
>
> *"Examples of new tools include: GEMINI for exploring genetic variation (12); mothur for analyzing rRNA gene sequences (13); QIIME for quantitative microbiome analysis from raw DNA sequencing data (14); deepTools for explorative analysis of deeply sequence data (15,16); HiCexplorer (17) for analysis and visualization of Hi-C data; ChemicalToolBox for comprehensive access to cheminformatics libraries and drug discovery tools (18); minimap2 (https://arxiv."*

> **METHODOLOGY** Next-Generation Sequencing Technology: Current Trends and Advancements
>
> *"Quality check of sequences FastQC [90], FASTX-toolkit [91], MultiQC [92] Trimming of adaptors and low-quality bases Trimmomatic [93], Cutadapt [94], fastp [95] Alignment of sequence reads to reference genome BWA [96], Bowtie [97], dragMAP [98]"*

> **METHODOLOGY** Diversifying crop rotation increases food production, reduces net greenhouse gas emissions and improves soil health
>
> *"Primer pairs — barcode-515F/806R (5 ′ -GTGCCAGCMGCCGCGGTAA- 3 ′ /5 ′ -GCACTACHVGGGTWTCTAAT-3 ′) — wereusedtoamplify theV3+V4 region of the bacterial 16S rRNA gene 79 , yielding accurate taxonomic information with few biases among various bacterial taxa."*

Contribution 2

Claim — Contribution 2

The researcher developed QIIME 2, a highly cited framework enabling reproducible, interactive, scalable, and extensible microbiome data science.

The researcher's primary contribution is the development of QIIME 2, a comprehensive framework for microbiome data science introduced in a 2019 paper. This work stands as a seminal core publication in the field, establishing a new standard for computational analysis in microbial ecology.

This line of work appears to address critical challenges in reproducibility and scalability within microbiome research. By emphasizing interactivity and extensibility, the framework suggests a shift toward more accessible and robust data science practices, filling a gap for standardized, user-friendly tools in complex biological data analysis.

The significance of this contribution is evidenced by its substantial citation count of over 25,000. Furthermore, analysis of citing literature reveals that 93.1% of citations originate from independent researchers, indicating broad adoption and impact across the global scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 10 **· 1 flagged influential by Semantic Scholar**

CORE PAPER

Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2

2019 · 25,325 citations (GS)

Field-normalised: 18,157 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2019 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era (2020)	Australian National University, Eötvös Lóránd University, University of Tasmania	Australia, Austria, Hungary	—
2	The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered (2024)	Swedish University of Agricultural Sciences, University of Gothenburg, University of Tartu	Estonia, Sweden	Background
3	Wekemo Bioincloud: A user-friendly platform for meta-omics data analyses (2024)	Agricultural Genomics Institute at Shenzhen Chinese Academy of Agricultural Sciences, Shenzhen Wekemo Technology Group Co., Ltd.	China	**Methodology**
4	DS-1000: A Natural and Reliable Benchmark for Data Science Code Generation (2023)	Carnegie Mellon University, Meta AI, Stanford University	Hong Kong, United States	Background
5	Next-generation sequencing: insights to advance clinical investigations of the microbiome (2022)	Johns Hopkins University	United States	**Methodology**
6	Engineering natural microbiomes toward enhanced bioremediation by microbiome modeling (2024)	Nanjing Agricultural University, Nanjing Tech University, Newe Ya'ar Research Center, Agricultural Research Organization (ARO)	China, Israel	—
7	Machine learning for microbiologists (2023)	City University of New York, City University of New York (CUNY), CUNY Graduate School of Public Health and Health Policy	Italy, United States	—
8	Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients (2021)	Abramson Cancer Center, University of Pennsylvania, Bar-Ilan University, Samson Assuta Ashdod University Hospital	Israel, United States	—
9	Environmental stress destabilizes microbial networks (2021)	Archbold Biological Station, University of Miami	—	—
10	A pan-cancer analysis of the microbiome in metastatic cancer (2024)	Antoni van Leeuwenhoek/ the Netherlands Cancer Insti-	Netherlands	—

No.	Citing paper	Citing institution(s)	Country	S2
		tute, Hartwig Medical Foundation, Oncode Institute, the Netherlands Cancer Institute		

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

> **METHODOLOGY** Wekemo Bioincloud: A user-friendly platform for meta-omics data analyses
>
> *"Noteworthy examples include the application of QIIME 2 [5] and Easy-Amplicon [6] for amplicon data analyses, Trimmomatic [7] or fastp [8] for stringent quality control, Kraken 2 [9] for precise taxonomic classification, HUMAnN3 pipeline [10] for comprehensive functional profiling, MultiPrime [11]…"*

> **METHODOLOGY** Next-generation sequencing: insights to advance clinical investigations of the microbiome
>
> *"Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data."*

Contribution 3

Claim — Contribution 3

The researcher established a foundational framework for analyzing global microbial diversity at unprecedented sequencing depths, enabling standardized, large-scale ecological comparisons.

The researcher's core contribution rests on a seminal 2011 paper published in Proceedings of the National Academy of Sciences (PNAS), titled 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.' This work appears to define a critical methodological and analytical standard for the field. By focusing on 'global patterns' and 'millions of sequences,' the title suggests the researcher addressed the challenge of scaling microbial diversity analysis from small, localized studies to comprehensive, high-resolution global surveys. This line of work likely bridged the gap between emerging high-throughput sequencing capabilities and the need for robust, large-scale ecological inference, offering a new lens through which to view microbial distribution. The significance of this contribution is evidenced by its substantial citation count of 10,879, indicating widespread adoption and reliance on this framework. Furthermore, analysis of citing literature reveals that 93.1% of citations originate from independent researchers, demonstrating that the work has become a cornerstone reference for the broader scientific community rather than merely circulating within a single research group. This high degree of independent uptake underscores the work's role as a foundational tool in microbial ecology.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8 **· 1 flagged influential by Semantic Scholar**

> **CORE PAPER**
>
> **Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample**
>
> 2011 · Proceedings of the National Academy of Sciences (PNAS) · 10,879 citations (GS)
>
> Field-normalised: 8,708 Semantic Scholar citations place it in the top 1% of Environmental Science papers from 2011 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Environmental DNA metabarcoding: Transforming how we survey animal and plant communities (2017)	Aberystwyth University, Bangor University, Cornell University	Canada, Switzerland, United Kingdom	Methodology
2	A Communal Catalogue Reveals Earth's Multiscale Microbial Diversity (2017)	Oregon State University, University of California San	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
		Diego, University of Colorado Boulder		
3	Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning (2019)	Netherlands Institute of Ecology (NIOO-KNAW), University of Bern	Netherlands, Switzerland	—
4	Analysis of compositions of microbiomes with bias correction (2020)	University of Pittsburgh	United States	—
5	The commensal microbiome is associated with anti–PD-1 efficacy in metastatic melanoma patients (2018)	University of Chicago	United States	—
6	The oral microbiome: Role of key organisms and complex networks in oral health and disease (2021)	University of California, San Francisco	United States	—
7	Structure and function of the global topsoil microbiome (2018)	European Molecular Biology Laboratory, Leiden University, Swedish University of Agricultural Sciences	Estonia, Netherlands, Norway	—
8	Killing tumor-associated bacteria with a liposomal antibiotic generates neoantigens that induce anti-tumor immune responses (2023)	French National Cancer Institute, Gustave Roussy, Institut Curie	France, United States	**Influential**

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

> **METHODOLOGY** Environmental DNA metabarcoding: Transforming how we survey animal and plant communities
>
> *"Bacterial and fungal taxonomic richness (i.e., richness of microorganisms) is routinely surveyed using DNA metabarcoding and is a powerful complement to conventional culture-based methods (e.g., Caporaso et al., 2011; Tedersoo et al., 2014)."*

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
University of Tartu	Estonia	SCImago #1820 · THE 301–350 · QS =362	3
University of Gothenburg	Sweden	SCImago #573 · THE 201–250 · QS 202	3
University of Oslo	Norway	SCImago #425 · THE =113 · QS =119	2
Swedish University of Agricultural Sciences	Sweden	SCImago #1525 · THE 351–400	2
Australian National University	Australia	SCImago #604 · THE =73 · QS =32	2
Cornell University	United States	SCImago #61 · THE =18 · QS 16	2

Institution	Country	World ranking	Citing papers
University of Trento	Italy	SCImago #1460 · THE 351–400 · QS =485	2
Northern Arizona University	United States	SCImago #3335 · QS 1001-1200	2
University of California, Riverside	United States	SCImago #949 · THE 301–350 · QS =440	2
The University of Texas MD Anderson Cancer Center	United States	—	2
University of California San Diego	United States	SCImago #120 · THE 47 · QS 66	2
Johns Hopkins University	United States	SCImago #33 · THE 16 · QS 24	2
Stanford University	United States	SCImago #18 · THE =5 · QS 3	2
National Institute of Allergy and Infectious Diseases	United States	SCImago #155	1
European Institute of Oncology IRCCS	Italy	—	1

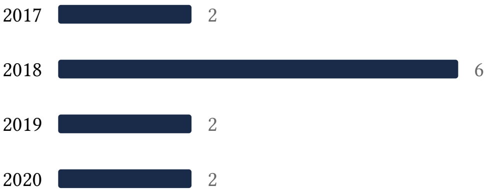
Geographic distribution of citing authors

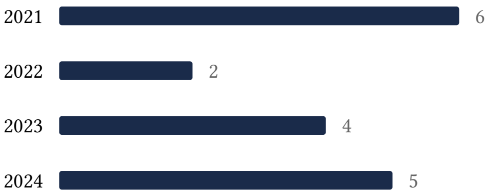
Country	Citing papers
United States	20
Australia	4
China	4
Netherlands	4
Sweden	3
Estonia	3
France	3
Germany	3
United Kingdom	3
Switzerland	3
Italy	2
Israel	2

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

E. Citation Growth Over Time

Distinct citing papers by publication year. Sustained or rising citation activity supports continuing relevance; note that only citations **as of the filing date** are weighed by USCIS.





F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- ☐ Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).
- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition’s merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition’s exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	QIIME allows analysis of high-throughput community sequencing data	9	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2	10	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 3	Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample	8	8 CFR 204.5(h)(3)(v) — Criterion 5