Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Marco Marra

BC Cancer Agency; University of British Columbia, Genome Sciences Centre

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

15	Citing papers mapped
15	Citation edges
3	Home papers mapped
205	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.

100.0% independent of 11 classified citing papers

Citation type	Count
Independent	11
Self-citation	0
Co-author	0
Same-institution	0

4 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 contributed to the initial sequencing and analysis of the human genome, a foundational achievement published in Nature in 2001 that has garnered over 26,000 citations.

The researcher's primary contribution rests on the seminal 2001 Nature paper titled 'Initial sequencing and analysis of the human genome.' This work represents a major milestone in genomic science, establishing a comprehensive reference for human genetic structure. The titles indicate that this publication served as a definitive resource for the scientific community, rather than part of a narrow, incremental series of follow-up studies by the same author.

This line of work appears to address the fundamental challenge of mapping the human genetic code. By producing the initial sequence, the researcher helped bridge the gap between theoretical genetics and practical genomic analysis. The absence of follow-up papers by the same researcher suggests that this contribution stands as a singular, high-impact event that defined the field's baseline rather than a continuing niche project.

The significance of this work is evidenced by its extensive citation record, with over 26,000 citations indicating widespread adoption. Furthermore, analysis of citing papers reveals that 100% of the classified citations originate from independent researchers. This high degree of independent uptake underscores the work's broad utility and its status as a critical, field-defining resource relied upon by the global scientific community.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Initial sequencing and analysis of the human genome

2001 · Nature · 26,555 citations (GS)

Field-normalised: 24,058 Semantic Scholar citations place it in the top 1% of Biology papers from 2001 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)	Frederick National Laboratory for Cancer Research	United States	—
2	Reactive oxygen species, toxicity, oxidative stress, and antioxidants: chronic diseases and aging	Constantine the Philosopher University in Nitra, King Saud University, Slovak University of Technology	Czech Republic, Saudi Arabia, Slovakia	—
3	DNA methylation: a historical perspective	Max Planck Institute for Molecular Genetics	Germany	—
4	Gene regulation by long non-coding RNAs and its biological functions	Center for Applied Medical Research, University of Navarra, University of the Chinese Academy of Sciences	China, Spain	—
5	Highly accurate protein structure prediction for the human proteome (2021)	DeepMind, EMBL-EBI, European Molecular Biology Laboratory	United Kingdom	—

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).

Contribution 2

Claim — Contribution 2

The researcher led the seminal analysis of the Arabidopsis thaliana genome sequence, establishing a foundational reference for plant genomics that has been widely adopted by independent scientists.

CLAIM: The researcher's primary contribution is the comprehensive analysis of the genome sequence of the flowering plant Arabidopsis thaliana, published in 2000. This work serves as the cornerstone of the provided evidence, representing a major milestone in plant molecular biology.

ORIGINALITY: The titles indicate that this work addressed the critical need for a complete genomic reference for a model plant species. By sequencing and analyzing the Arabidopsis genome, the researcher provided a fundamental resource that enabled subsequent functional genomics studies, filling a significant gap in the understanding of plant genetic architecture.

SIGNIFICANCE: The core paper has accumulated over 11,000 citations, demonstrating its enduring impact on the field. Furthermore, citation analysis reveals that 100% of the classified citing papers originate from independent researchers, confirming that the work has been broadly adopted and utilized by the global scientific community outside the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 3

> **CORE PAPER**
>
> **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana**
>
> 2000 · 11,748 citations (GS)
>
> Field-normalised: 9,042 Semantic Scholar citations place it in the top 1% of Biology papers from 2000 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Breeding crops to feed 10 billion (2019)	Chinese Academy of Sciences, InterGrain Pty Ltd, John Innes Centre	Australia, China, Saudi Arabia	—
2	Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome (2019)	Agricultural Genomics Institute, Chinese Academy of Agricultural Sciences, Baylor College of Medicine, Dana-Farber Cancer Institute	China, Germany, United States	—
3	Unlocking plant genetics with telomere-to-telomere genome assemblies (2024)	BGI-Shenzhen, Helmholtz Zentrum München, John Innes Centre	Australia, China, Germany	—

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).

Contribution 3

Claim — Contribution 3

The researcher developed Circos, a visualization framework that established a new information aesthetic for comparative genomics, enabling the clear display of complex genomic relationships.

CLAIM: The researcher's primary contribution is the development of Circos, a visualization tool introduced in a 2009 paper that established a distinct information aesthetic for comparative genomics. This work stands as a seminal core contribution in the field.

ORIGINALITY: The title suggests the work addressed the challenge of visually representing complex genomic data by introducing a novel aesthetic approach. By focusing on an 'information aesthetic,' the researcher appears to have shifted the paradigm from purely functional data display to one that emphasizes clarity and visual structure in comparative genomic analysis.

SIGNIFICANCE: The core paper has accumulated over 11,000 citations, indicating widespread adoption and foundational importance. Analysis of citing literature reveals that 100% of classified citations originate from independent researchers, demonstrating that the tool has been embraced broadly across the global scientific community rather than being limited to the researcher's immediate network.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 3 **· 1 flagged influential by Semantic Scholar**

> **CORE PAPER**
>
> ### Circos: an information aesthetic for comparative genomics
>
> 2009 · 11,588 citations (GS)
>
> Field-normalised: 9,927 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2009 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	JBrowse 2: a modular genome browser with views of synteny and structural variation (2023)	Ontario Institute for Cancer Research, Oregon Health and Science University, Oregon State University	Canada, United States	**Methodology**
2	A comprehensive map of molecular drug targets (2017)	European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI), The Institute of Cancer Research, University of New Mexico	United Kingdom, United States	—
3	Personalized bacteriophage therapy outcomes for 100 consecutive cases: a multicentre, multinational, retrospective observational study (2024)	Eliava Institute of Bacteriophages, Microbiology and Virology, KU Leuven, Queen Astrid Military Hospital	Belgium, Georgia	**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** JBrowse 2: a modular genome browser with views of synteny and structural variation
>
> *"The Circular View displays annotations in a circular format as popularized by Circos [8]."*

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
University of Georgia	United States	SCImago #597 · THE 351–400 · QS 525	2
John Innes Centre	United Kingdom	—	2
Chinese Academy of Sciences	China	SCImago #2	1
Slovak University of Technology	Slovakia	SCImago #6145	1
King Saud University	Saudi Arabia	SCImago #264 · THE 251–300 · QS 143	1
University of Hradec Kralove	Czech Republic	SCImago #5405 · THE 1001–1200 · QS 1001-1200	1
Max Planck Institute of Molecular Cell Biology and Genetics	Germany	SCImago #951	1
Saarland University; Max Planck Institute for Informatics	Germany	—	1
The University of Queensland	Australia	SCImago #126 · THE =80 · QS =42	1
Murdoch University	Australia	SCImago #2328 · THE 401–500 · QS =423	1
Oregon Health and Science University	United States	SCImago #689 · THE 351–400	1
European Molecular Biology Laboratory	United Kingdom	—	1
University of California, Berkeley	United States	SCImago #95 · THE 9 · QS =17	1
KU Leuven	Belgium	SCImago #180 · THE 46 · QS 60	1
Pacific Biosciences	United States	—	1

Geographic distribution of citing authors

Country	Citing papers
United States	6
China	4
United Kingdom	4
Germany	3
Australia	2
Saudi Arabia	2
Slovakia	1
Spain	1
Canada	1
Czech Republic	1
Belgium	1
Georgia	1

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	Initial sequencing and analysis of the human genome	5	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	Analysis of the genome sequence of the flowering plant Arabidopsis thaliana	3	8 CFR 204.5(h)(3)(v) — Criterion 5

Contribution	Core paper	Indep. cites	Supports
Contribution 3	Circos: an information aesthetic for comparative genomics	3	8 CFR 204.5(h)(3)(v) — Criterion 5