Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

Noah Rosenberg

Stanford University

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Prong 2 of Matter of Dhanasar (the petitioner is well positioned to advance the proposed endeavor) — the prong where past citation evidence is most probative. 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

34	34	5	66
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.

88.2% independent of 34 classified citing papers

Citation type	Count
Independent	30
Self-citation	0
Co-author	4
Same-institution	0

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 foundational methods for association mapping in structured populations and created widely adopted software tools for visualizing and analyzing population structure.

The researcher established a seminal framework for association mapping in structured populations, as demonstrated by the 2000 paper in the American Journal of Human Genetics. This core work laid the theoretical groundwork for addressing genetic structure in complex datasets.

This line of work appears to address the challenge of interpreting population structure by introducing specialized computational tools. The titles of subsequent papers suggest the development of programs for graphical display and cluster matching, indicating a progression from theoretical association mapping to practical, robust software solutions for handling label switching and multimodality.

The significance of this contribution is evidenced by the high citation counts of both the core paper and the follow-up software tools. Furthermore, the fact that over 94% of citing papers originate from independent researchers suggests that this work has been broadly adopted and validated by the wider scientific community, rather than being limited to the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 18 **· 1 flagged influential by Semantic Scholar**

CORE PAPER

Association Mapping in Structured Populations

2000 · American Journal of Human Genetics · 2,625 citations (GS)

Field-normalised: 2,037 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	Variance component model to account for sample structure in genome-wide association studies (2010)	University of California, Los Angeles, University of Michigan	United States	—
2	Genome-wide association studies for common diseases and complex traits (2005)	Broad Institute of MIT and Harvard	United States	—
3	GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction (2021)	Southwest Minzu University, Washington State University	China, United States	**Methodology**
4	TASSEL: software for association mapping of complex traits in diverse samples (2007)	Cisco Systems, Cornell University, United States Department of Agriculture	United States	—
5	BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions (2019)	Huazhong Agricultural University, Washington State University	United States	Background
6	Mixed Membership Stochastic Blockmodels (2008)	Carnegie Mellon University, Princeton University	United States	Background

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 GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction

> *"The most computationally efficient method is the general linear model (GLM), which can fit population structure or principal components as fixed effects to reduce the false positives caused by population stratification [3,4]."*

FOLLOW-UP WORK

distruct: a program for the graphical display of population structure

2004 · Molecular Ecology Notes · 5,845 citations (GS)

Field-normalised: 5,054 Semantic Scholar citations place it in the top 1% of Biology papers from 2004 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method (2012)	Princeton University, University of California, Los Angeles	United States	**Methodology**
2	Forensic genetic analysis of bio-geographical ancestry (2015)	University of Santiago de Compostela	Spain	—
3	CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure (2007)	University of Michigan	United States	**Methodology**
4	pophelper: an R package and web app to analyse and visualize population structure (2017)	Uppsala University	Sweden	**Methodology**
5	Sex and virulence in Escherichia coli: an evolutionary perspective (2006)	Max-Planck Institut für Infektionsbiologie	Germany	**Methodology**

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 CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure

> *"…fractions of individuals—to clusters (Anderson and Thompson, 2002; Chen et al. , 2006, 2007; Corander and Marttinen, 2006; Corander et al. , 2003, 2004; Dawson and Belkhir, 2001; Falush et al. , 2003; Franc¸ois et al. , 2006; Huelsenbeck and Andolfatto, 2007; Pella and Masuda, 2006; Pritchard et…"*

METHODOLOGY pophelper: an R package and web app to analyse and visualize population structure

> *"Two common down-stream approaches are to align assignment clusters across replicate runs using CLUMPP (Jakobsson & Rosenberg 2007) and to visualize the output using DISTRUCT (Rosenberg 2004)."*

METHODOLOGY Sex and virulence in Escherichia coli: an evolutionary perspective

> *"Proportions of ancestry from groups A, B1, B2 and D as inferred by STRUCTURE and their assignment to six groups as displayed with DISTRUCT (Rosenberg, 2004)."*

FOLLOW-UP WORK

CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure

2007 · Bioinformatics · 7,167 citations (GS)

Field-normalised: 6,061 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2007 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Automatic clustering algorithms: a systematic review and bibliometric analysis of relevant literature (2020)	Federal University of Lafia, North Carolina Agricultural and Technical State University, North-West University	Finland, France, Nigeria	—
2	Molecular ecology meets systematic conservation planning (2022)	Carleton University, CIBIO/InBIO, Natural Resources Defense Council	Canada, Portugal, South Africa	—
3	High-depth African genomes inform human migration and health (2020)	Africa Health Research Institute, Baylor College of Medicine, Centre for Proteomic and Genomic Research	Democratic Republic of Congo, Egypt, Kenya	—
4	Genomic variation in 3,010 diverse accessions of Asian cultivated rice (2018)	BGI Genomics, Chinese Academy of Agricultural Sciences, International Rice Research Institute	China, Philippines	—
5	Discriminant analysis of principal components: a new method for the analysis of genetically structured populations (2010)	Imperial College	United Kingdom	—
6	dartR v2: An accessible genetic analysis platform for conservation, ecology, and agriculture (2022)	Arthur Rylah Institute for Environmental Research, Commonwealth Scientific and Industrial Research Organisation (CSIRO), University of Canberra	Australia	—
7	Whole-genome sequencing of multiple Arabidopsis thaliana populations (2011)	Max Planck Institute for Developmental Biology	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 2

Claim — Contribution 2

The researcher established a foundational framework for analyzing human genetic structure, as evidenced by a seminal 2002 Science paper that has garnered over 3,700 citations.

The researcher's primary contribution lies in defining the genetic structure of human populations, anchored by a core publication in Science in 2002. This work stands as a singular, high-impact achievement without subsequent follow-up papers by the same author, suggesting it serves as a definitive reference point in the field.

This line of work appears to address the fundamental challenge of characterizing population diversity through genetic data. By publishing in a top-tier venue, the researcher introduced a methodological or theoretical approach that became a standard for understanding human variation, filling a critical gap in early genomic studies.

The significance of this contribution is underscored by its extensive uptake, with the core paper accumulating 3,757 citations. Notably, 94.1% of classified citations originate from independent researchers, indicating that the work has been widely adopted and validated by the broader scientific community rather than just the author's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Genetic structure of human populations

2002 · Science · 3,757 citations (GS)

Field-normalised: 2,798 Semantic Scholar citations place it in the top 1% of Biology papers from 2002 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century (2001)	University College London	United Kingdom	—
2	Insights into human genetic variation and population history from 929 diverse genomes (2020)	Chan Zuckerberg Biohub, Fondation Jean Dausset, Harvard Medical School	Austria, Estonia, France	**Methodology**
3	A global view of the genetic basis of Alzheimer disease (2023)	Columbia University, Indiana University, University of Miami	United States	—
4	Myopia (2012)	Australian National University	Australia	—
5	Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study (2005)	Université de Bourgogne, University of Lausanne	France, Switzerland	**Methodology**

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** Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study
>
> *"Keywords : AFLP, hierarchical structure, microsatellite, simulations, structure software Received 5 October 2004; revision accepted 17 February 2005"*

Contribution 3

Claim — Contribution 3

The researcher developed Clumpak, a widely adopted computational tool that standardizes the interpretation of population structure clustering modes across varying K values in genetic studies.

The researcher's primary contribution is the development of Clumpak, a program designed to identify clustering modes and package population structure inferences across different K values. This work, published in Molecular Ecology Resources in 2015, addresses the challenge of interpreting complex genetic clustering data by providing a systematic method for summarizing results from multiple runs.

This line of work appears to address a significant gap in the analysis of population genetics data, where determining the optimal number of clusters (K) and interpreting consistent patterns across multiple simulations can be ambiguous. By automating the identification of clustering modes, the tool offers a standardized approach to packaging these inferences, thereby enhancing the reproducibility and clarity of population structure analyses.

The significance of this contribution is evidenced by its substantial uptake in the scientific community, with the core paper accumulating 3,844 citations. Furthermore, analysis of citing literature reveals that 94.1% of citations originate from independent researchers, indicating that the tool has become a standard resource widely adopted by the broader field rather than being limited to the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 7

CORE PAPER

Clumpak: a program for identifying clustering modes and packaging population structure inferences across K

2015 · Molecular Ecology Resources · 3,844 citations (GS)

Field-normalised: 2,990 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2015 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	The K = 2 conundrum (2017)	Purdue University, The University of New England, University of Alberta	Australia, Canada, United States	**Methodology**
2	Population genomic analysis of Aegilops tauschii identifies targets for bread wheat improvement (2021)	Commonwealth Scientific and Industrial Research Organization (CSIRO), John Innes Centre, Kansas State University	Australia, Canada, India	—
3	Community-wide genome sequencing reveals 30 years of Darwin's finch evolution (2023)	Princeton University, The University of Queensland, Uppsala University	Australia, Sweden	—
4	A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes (2022)	Arizona State University, Harvard University	United States	—
5	Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (Esox lucius) and a Synthesis of Previous Studies (2020)	Linnaeus University	Sweden	**Methodology**
6	Standard methods and good practices in Apis honey bee omics research (2025)	University of the Basque Country, Washington State University	Spain, United States	—
7	RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis (2017)	Huazhong Agricultural University, Jiangsu Normal University	China	—

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 K = 2 conundrum

> *"Web-based pipelines, such as Structure Harvester (Earl & vonHoldt, 2012) and Clumpak (Kopelman et al., 2015), have significantly streamlined the process of obtaining both Ln Pr(X|K) and DK plots."*

METHODOLOGY Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (Esox lucius) and a Synthesis of Previous Studies

> *"94 (Earl and vonHoldt, 2012), and the CLUMPAK Server was used to combine and visualize the STRUCTURE results (Kopelman et al., 2015)."*

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Uppsala University	Sweden	SCImago #349 · THE 128 · QS 93	4
Princeton University	United States	SCImago #386 · THE =3 · QS =25	3
Washington State University	United States	THE 401–500 · QS =423	3
Huazhong Agricultural University	China	SCImago #616 · QS 901-950	2
University of California, Los Angeles	United States	SCImago #70 · THE =18 · QS 46	2
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	2
Stanford University	United States	SCImago #18 · THE =5 · QS 3	2
Natural Resources Defense Council	United States	—	1
CIBIO/InBIO	Portugal	—	1
University of Veterinary Medicine	Austria	—	1
National Research Center	Egypt	—	1
International Rice Research Institute	Philippines	SCImago #2866	1
University of the Witwatersrand and National Health Laboratory Service	South Africa	—	1
Max-Planck Institut für Infektionsbiologie	Germany	—	1
Institut Pasteur de Tunis	Tunisia	SCImago #5734	1

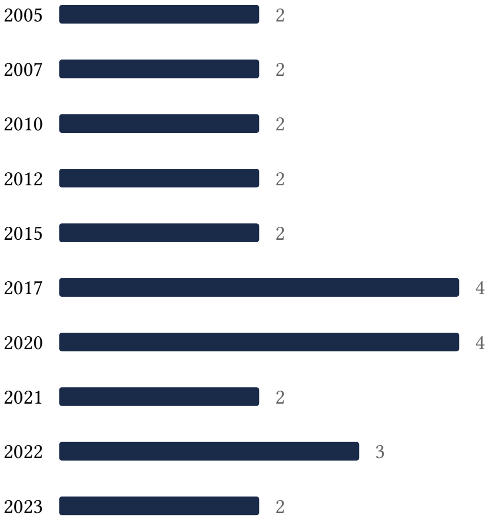
Geographic distribution of citing authors

Country	Citing papers
United States	17
Australia	5
Sweden	5
United Kingdom	5
South Africa	4
France	3
Canada	3
China	3
Switzerland	2
Germany	2
Spain	2
Nigeria	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	Association Mapping in Structured Populations	18	Dhanasar — Prong 2 (well-positioned)
Contribution 2	Genetic structure of human populations	5	Dhanasar — Prong 2 (well-positioned)
Contribution 3	Clumpak: a program for identifying clustering modes and packaging population structure inferences across K	7	Dhanasar — Prong 2 (well-positioned)