Cibtech Journal of Bio-Protocols ISSN: 2319–3840 (Online)
An Open Access, Online International Journal Available at http://www.cibtech.org/cjbp.htm
2015 Vol. 4 (1) January-April, pp.1-7/
Research Article

A PROTOCOL TO EVALUATE POPULATION GENETICS PAPERS

***Clyde Winters**
Uthman dan Fodio Institute
Chicago, Illinois 60643
**Author for Correspondence*

ABSTRACT
In this article, we provide a framework for analyzing and interpreting population genetics articles. Bayesian Statistics combines prior beliefs and sample DNA information to make inferences about the sample based on the researchers prior beliefs. In traditional research the researcher makes a hypothesis and performs an experiment to test the relationship between two or more variables. Population genetics research does not use Bayesian statistics to test the association of variables; but it can be used to study the difference between population groups, e.g. admixture rates between and among populations. Use of Bayesian statistics to predict past population events is problematic, unless the researcher has a firm knowledge of the prehistory of a region. The purpose of this checklist is to help the researcher understand the design of the research article under review and make sure archaeogenetic methods are being employed. The checklist will allow graduate students, anthropologists and professional geneticists to evaluate the essential design features of the research paper especially the interpretation and reporting of results. The checklist looks at the foundation of the scientist's thinking to determine the strengths and weaknesses of the research to find out if the research conclusions are accurate, valid and based on archaeogenetic measures. . Use of this checklist will expedite the evaluator's evaluation of the phases of the article being evaluated to insure the literature is valid and therefore will make a significant contribution to population genetics literature.

Keywords: *Hypothesis, Bayesian Statistics, Prehistory, Variables, Archaeogenetics*

INTRODUCTION
The most important activity of the researcher is reading research articles. Although many population genetics articles are published graduate students and professional geneticists need a tool to evaluate the articles to determine their relevance. Here we provide an evaluation tool researchers can use to evaluate genetics research articles. There are tens of articles published each year in population genetics. These articles are must reading for anthropologist and molecular geneticists interested in migration and population genetics.

Using Bayesian statistics molecular geneticists make inferences about prehistoric demographic events, relating to various ethnic populations. To reconcile their genomic evidence with prehistoric and historical information some population geneticists use archaeological, linguistic and paleoanthropological data to corroborate their DNA findings. The use of use archaeological, linguistic and paleoanthropological data to support molecular genetics is called archaeogenetics (Renfrew and Boyle, 2000).

Geneticists and anthropologists use Archaeogenetics to explain and discuss past population events. Archaeogenetics can be defined as the use of prehistoric and historical events to determine by archaeology, genetics and linguistics in concert with the DNA of various ethnic groups to infer the ethnic identity of ancient populations and/or the ancient migration of one population to another geographical location.

MATERIALS AND METHODS
Method
The research design used in this paper is a literature based research methodology. We sampled the archaeogenetics literature base. The literature was analyzed to determine what criterions explain best the relationship between contemporary DNA, ancient DNA human remains and past population events within

the context of the "wave of advance" model. The top criterions used to evaluate research papers were used to construct this Checklist.

RESULTS AND DISCUSSION

Results

In population genetics the researcher usually uses the "wave of advance" model to explain demographic movements in the past. The "wave of advance" model was used to explain the spread of advantageous genes within a population (Ackland *et al.*, 2007; Renfrew, 2001). This theory was adapted to explain why an advantageous technology that may appear in one population spreads (and or taken) to another population living in a different geographical area (Ackland *et al.*, 2007).

Although archaeogenetics is the norm for many molecular geneticists, most researchers believe that Bayesian statistics alone have sufficient power to demonstrate the validity of their research, and fail to corrobate the DNA data with corresponding archaeological, linguistic and paleoanthropological evidence.

Many people don't know how to evaluate population genetics articles, because they are expost facto research based on "statistical infererences" or the beliefs of the researcher supported by statistics. As a result, researchers cannot judge the validity and reliability of the research. One must assume the research is correct based solely on the Bayesian statistical inferences—not the interactions between an independent variable and dependent variable(s).

In research there are two variables: variables that can be manipulated and variables that cannot be manipulated.

A variable that can be manipulated is a variable that can be changed for example, your ability to perform a particular task can be influenced by the amount of training you receive in performing the task.

A variable that cannot be manipulated cannot be changed. For example, right now you are a particular age, it cannot be manipulated. You are either Black or white, race cannot change.

Research studies include a number of variables. Variables which can be manipulated or not manipulated

Independent Variable (IV) any variable used to control for individual differences (this variable usually not manipulated)

Dependent variable (DV) any outcome measure which is effected by the IV. The effect of sex (IV) on reading achievement (DV).

Validity is testing the appropriateness, meaningfulness and usefulness of specific inferences made from test scores. In qualitative research the extent to which the research uses methods and procedures that ensure a high degree of research quality and rigor.

Internal Validity, we assume that whatever was manipulated produced a change in the dependent measure. IV insured by control of the extraneous variables: health, sex, race, SES, age, IQ, religion.

External Validity provides the ability to generalize the findings. In other words the IV produced a change in DV.

In normal scientific research the researcher states a hypothesis and uses the scientific method to test his/her hypothesis. The validity and reliability of the piece of research is then determined by statistical significance tests focused on the interaction between the independent and dependent variable.

In the traditional evaluation of a piece of research literature you look at the researcher's hypothesis, results and statistical methods s/he used to determine the statistical significance of the research. This is not the case in population genetics research; in this research you are evaluating statistical inferences based on ***the beliefs already held by the researcher*** about a set of data, instead of testing a hypothesis.

As a result, the research contained in a population genetics article, reflects the views and beliefs already held by the researcher. Thusly, the statistical inferences will automatically support the views and beliefs held by that researcher; and any outliners that fail to support the researcher's beliefs may not be mentioned in the resulting research article/paper.

Here we will ask the question: "How do you evaluate population genetics research if it is expost facto research that lacks an experimental design?" First, we will attempt to look at the doxa that may influence a geneticist's research and the constructs that should be considered when evaluating this knowledge base.

In reading any piece of research literature, we assume that any article or book written by an establishment member of the academe is reliable and valid. A piece of research full of valid scientific and/or historical truths--erudite scholarship and impeccable research based on the scientific method.

The scientific method is based on hypotheses testing. Hypotheses testing mean that a researcher forms a hypothesis and test the hypothesis using a series of quantitative or qualitative statistical methods to determine the statistical significance of the hypothesis being tested. The scientific method is based on experimentation to test a hypothesis.

Population geneticists usually do not test hypotheses. They make inferences about data based on Bayesian statistical inferences. They do not use statistical methods to determine the statistical significance of a hypothesis, they use statistics to describe data being reviewed by the researcher based on the beliefs the researcher already holds about the data being reviewed.

Population genetics is a type of Expost facto research. Expost facto research design is a quasi-experimental type of study examining how an independent variable, present prior to the research study, affects a dependent variable.

Whereas the subjects in experimental research are randomly selected, the participants in Expost facto research, are not randomly selected or assigned. The genome of the research subjects is examined to determine the haplotypes and haplogroups carried by the participants in the study.

In population genetics research the researcher uses the Bayesian inference method of statistical inference. The Bayesian statistical method, is a subjective research design/method that provides a rational method of updating the researcher's beliefs.

Since, the results of a Bayesian statistical analysis are a series of beliefs based on statistical inferences, the results cannot stand alone. This is due to the reality, that any results, reported by a researcher are only a series of inferences based on the researcher's belief about phenomena backed up by a series statistical results. If the results are published without corresponding evidence from archaeology, anthropology, linguistics and or craniometrics the inferences are pure conjecture, because they reflect the attitudes already held by the researcher, confirmed by data selected by the researcher to support his or her beliefs.

There is a sociological basis behind how a researcher interprets data. Sociological research indicates that there are unconscious cognitive structures within each individual. Cognitive structures that hold the idealistic view of members of the academe that determine how they perceive "reality". These structures are called doxa (Berlinerblau, 1999).

Commenting on these schema Berlinerblau (1999) noted that "These types of theories share the assumption that human beings know things that they do not even know that they know; that they "possess" knowledge about the world which exists in some sort of cognitive substrate, beyond the realm of discourse" (p.106).Wacquant (1995) says that doxa is " a realm of implicit and unstated beliefs".

Given the research suggesting that doxa exist, support the view that some researchers allow their hatred of multiculturalism, ethnic prejudice and racism to define their discourse, teaching and writing about themes relating to groups "other", than their own cultural and ethnic group. Moreover, it suggest that when topics such as Eurasian and African haplogroups, Afrocentrism, African origins of the Dravidians and etc., is attacked by members of the academe, these academics are supported by the "establishment" without any reservation, or test of the validity of their claims. In fact, it appears that doxic assumptions relating to the validity of back migration of so-called Eurasian genes into Africa, recent African origin of Dravidians and Dravidian origin of the Indus Valley Civilization obviates critique of the academics that disparage these themes. Due to Doxa you can state a researcher's attitude toward a historical, genetic or anthropological concept and theorems without the statement being an ad hominem

Discussion

To evaluate research literature a student should know the varied research methods. A student evaluating a piece of population genetics' literature must understand that the researcher is conducting an expost facto method of research that does not involve hypotheses testing .Given the nature of Bayesian inferences, you can not determine the validity and reliability of a piece of genetics research literature based on the

statistical significance of the data. What you must do is look at the research article and ask yourself a series of questions regarding the article's validity and reliability.

To facilitate evaluation of genetics research literature I have created a check list: Checklist used to analyze a Population Genetics Papers, to evaluate research articles.

To use the Checklist you would perform the following task. The Evaluator should read the article twice. The first reading of the article is brief.

Next make a close reading of the article. The close read should involve the Evaluator in underlining key details in the article, while making annotations of important points in the text.

During the second reading of the text the Evaluator will assess the research article using **the Checklist used to analyze a Population Genetics Papers**. Since the Bayesian statistics used for the study will support the inferences of the Researcher the answers for the majority of the questions on the checklist will be yes.

The key question in determining the validity of the research will be question 17. If the researcher only has Bayesian statistical inferences supporting the research study, the inferences made in the research article, may not be representative of actual past population events.

I will use the Checklist to evaluate a recent Population genetics article. The paper is Chaubey and Endicott (2015). As mentioned earlier Bayesian statistics, since they are based on the author's belief, will just about always support the author's inference. Below are my responses to the article placed on the Checklist. The evaluation of this article revealed the following responses:

1-3 is yes
4. No
5. Yes
6. Yes
7. No
8. No
9. Yes
10. Yes
11. Yes
12. Yes
13. No
14. No. No discussion of Southeast Asian and mainland Indian archaeology.
15. Yes
16. No
17. No

Because the answer to Question 17, was no, demands that we check the archaeology literature to determine if the Bayesian statistical inferences can find support from the craniometric, and archaeological record for SEA and India, or if the results and conclusion are based solely on the doxa of the researchers.

Claim that the Onge, a mainland Munda group only recently came to India circa 26kya. This would place them in India after the alledged settlement of India by the Aryans. They wrote: "of the Andaman-specific mtDNA lineage M31a1 around 26 ka, while the ages of the diversification within M32 and M31a1 are estimated to fall within the Holocene, using whole-genome data in a Bayesian statistical setting (Barik *et al.*, 2008).

Because mtDNA divergence is anticipated to predate population divergence, collectively these estimates suggest that the Andamans were settled less than ~26 ka and that differentiation between the ancestors of the Onge and Great Andamanese commenced in the Terminal Pleistocene. Interestingly, this time frame is similar to the signal for population expansion found throughout ISEA (Guillot *et al.*, this issue) and represents the time of topographic transition from the vast expanses of Sundal and to the submerged Southeast Asian island chains of the Holocene. In conclusion, we find no support for the settlement of the Andaman Islands by a population descending from the initial out-of-Africa migration of humans, or their

immediate descendants in South Asia. It is clear that, overall, the Onge are more closely related to Southeast Asians than they are to present-day South Asians.

The similarity in proportions of the Onge genomes, attributed to the Melanesian, Malaysian (Jehai and Kensui), and South Asian ancestral components, combined with evidence for genetic drift, suggests that these constituent parts were present prior to their isolation from other parts of Southeast Asia".

Although this is the opinion of Chaubey and Endicott (2015), the Onge and other Munda populations were in India long before the Aryans. C Winters (2010) argues that Thangaraj et al using coalescence time and archaeological evidence illustrated that the TRMCA for mtDNA R8 which is found among Munda speakers have the following dates: R8 (41.7 kya), R8a (15.4 kya) and R8b (27.7 kya)13. The dating for mtDNA R8 indicates that this haplogroup and R7 are probably autochthonus to India.

The mtDNA of Munda speakers also includes deep rooted haplogroups from macrohaplogroup M. In addition to mtDNA haplogroup M2, we also find M58, M31, M6a2 and M42 among Munda speakers.

The Munda y-chromosome is O2a (M95). Kumar reports a coalescent rate of 65kya for Indian M953. There is a clear distinction of Indian Munda and Southeast Asian (SEA) Mon-Khmer speakers. The predominate SEA O clades are O3 and O1a. If SEA males had carried the y-chromosome O haplogroup to India there should be evidence of these clades among the Munda speakers—but they are nil8. On the otherhand, SEA males carry Indian y-chromosomes such as F, H, K2 (T) and etc8.

This indicates an early migration of Munda speakers to SEA. It suggests that Munda spread mtDNA R7 and y-chromosome haplogroup O to SEA.

Many Indians carry Munda haplogroups. The spread of Munda haplogroups are probably the result of conquest and intermarriage. The mythology of some Indian populations supports this proposition. In other words, instead of the Munda originating in SEA, they probably migrated to the region from India.

Chaubey *et al.,* (2013) based his conclusion on the research on Endicott *et al.,* (2006). Endicott *et al.,* (2006) argue that without comprehensive data from Myanmar it is not possible to identify whether the Andaman M31a1 arrived from India or if the Indian M31a2 came from South-East Asia. But either scenario casts serious doubts on the concept that the Andaman Islands were settled at the time of the migrations out of Africa carrying the current Eurasian mtDNA diversity".

Endicott *et al.,* (2006), admit that their conclusions should be preliminary because: "Without comprehensive data from Myanmar it is not possible to identify whether the Andaman M31a1 arrived from India or if the Indian M31a2 came from South-East Asia. But either scenario casts serious doubts on the concept that the Andaman Islands were settled at the time of the migrations out of Africa carrying the current Eurasian mtDNA diversity".

It is obvious that Endicott *et al.,* (2006), could not answer this question because they did not know much about Southeast Asian history. If they knew the archaeology of Southeast Asia they would have been able to answer this question. They would have known that the Dravidians who carry M31a2 probably carry the haplogroup as a result of migration of Dravidian back to South India from Myanmar. Winters (2010), I explain0 that, many Dravidian speakers in India formerly lived in Southeast Asia. Formerly intimate relations existed between South Indians and Southeast Asian people (Kanakasahai, 1966). The Tamilian form of Saivism is known as Agamas, the esoteric and ritualistic parts of Agama are non-Vedic (not of Indo-European origin). Agama was also the Southeast Asian form of Hinduism (Winters, 1985).

The Proto-Tamil speakers in Central Asia and China were called the Yakshas in Indian literature (Yuehchih by the Chinese) and Kosars (Kushana in Chinese literature). They were forced from China due to first the classical Mongoloids who founded Shang-Yin, then the Zhou and succeeding mongoloid Chinese and Thai populations that invaded Indo-China. This forced the Proto-Tamil speaking Kosars and Yakshas to later invade southern India in search of a new homeland in addition to Southeast Asia (Winters, 2011). In Southeast Asia Dravidian speakers probably encountered Proto-Andamanese carrying M31 and M32 who may have been the original settlers of the area.

The archaeological and genetic evidence indicate that Dravidian speakers lived in Southeast Asia (Kanakasabhai, 1966; Winters, 1985). It indicates that the first civilizations in Southeast Asia were founded by Dravidian speakers (Kanakasabhai, 1966). The Khmer introduced various aspects of

civilization in this region which precede the advent of the Thai speakers into this region. Upon their arrival in Indo-China, the Thai- Vietnamese people conquered the blacks learned their culture and continued to perpetuate the same cultural traits (Winters, 1985). Thusly, we see that both the Vietnamese and Thai peoples learned their culture, architecture, religion and writing from the Khmers and other Indo-African people.

While the Dravidians lived in Southeast they probably mated with the inhabitants related to the Andamanese (Winters, 2011). This mating pattern probably led to M31a2 entering the Dravidian gene pool when the Kamboja settled in Sengal and South India (Kanakasabhai, 1966).

Conclusion

In summary we can reject the research of Gyaneshwer Chaubey and Phillip Endicott, based on question 17 of the Checklist used to analyze a Population Genetics Papers, because it is unreliable and lacks validity because the researchers failed to study the archaeology and history of SEA. If they had, they would have known that Dravidian speakers formerly lived in SEA, until the advance of the Classical mongoloid people 2.5kya.

In summary, the validity and reliability of a piece of genetics research literature does not demand the Evaluator of a piece of literature to provide counter evidence all they need to do is evaluate the research using the checklist (see Appendix). If the answer to most of these questions is no, the research is unreliable and lacks any validity.

The key question on the checklist is question 17. To confirm the validity of the archaeological, craniometric and etc., data, the Evaluator should be knowledgeable about the archaeology of the area where the population movement has been inferred to have taken place. In this way you can determine if the Bayesian inferences correspond to the archaeological, craniometric, and linguistic data associated with the geographical area where the population movement is alleged to have occurred.

The major problem with most genetics literature which invalidates the research dealing with ancient population movements is that it is not supported by the ancient DNA, archaeological and/ or craniometric data.

This is why many of theories about the ancient populations of Europe and alledged back migrations are usually over turned once researchers examine the ancient DNA.

REFERENCES

Ackland GJ, Markus Signitzer, Kevin Stratford and Morrel H Cohen (2007). Cultural hitchhiking on the wave of advance of beneficial technologies. *PNAS* **104**(21) 8714-8719, published ahead of print May 16, 2007, Available: doi:10.1073/pnas.0702469104, Retrieved 2/6/2015 from: http://www.pnas.org/content/104/21/8714.full.

Berlinerblau J (1999). *Heresy in the University: The Black Athena Controversy and the Responsibilities of American Intellectuals* (Rutgers University Press).

Chaubey G and Phillip Endicott (2013). *The Andaman Islanders in a Regional Genetic Context: Reexamining the Evidence for an Early Peopling of the Archipelago from South Asia*. Retrieved 3/6/2015 from: http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=2055&context=humbiol.

Endicott P, Metspalu M, Stringer C, Macaulay V and Cooper A *et al.*, (2006). Multiplexed SNP Typing of Ancient DNA Clarifies the Origin of Andaman mtDNA Haplogroups amongst South Asian Tribal Populations. *PLoS ONE* **1**(1) e81, Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000081.

Kanakasabhai V (1966). The Tamil Eighteen Hundred Years ago.

Renfrew C (2001). From molecular genetics to archaeogenetics. *PNAS* **98**(9) 4830-4832, Available: doi:10.1073/pnas.091084198, http://www.pnas.org/content/98/9/4830.full.

Winters C (1985). The Far Eastern Origin of the Dravidians. *Journal of Tamil Studies* 66-92.

Winters C (2010). Munda Speakers are the Oldest Population in India. *The Internet Journal of Biological Anthropology* **4**(2), Available: https://ispub.com/IJBA/4/2/5591.

Winters C (2011). Comment: A back migration from Southeast Asia accounts for M31a2 in South India. Available: http://www.plosone.org/annotation/listThread.action?root=609.

Appendix

Checklist used to analyze a Population Genetics Papers

Answer the following questions relating to this research article below, or on a separate sheet of paper.

1. What was the rationale for the study, that is, what led up to it? Yes on page___, paragraph____ _, lines_____ No_____
2. Why do the authors believe that this problem is significant? Yes on page___, paragraph____ _, lines_____ No_____
3. What was the purpose of the study, that is, what did it intend to accomplish? Yes on page___, paragraph____ _, lines_____ No_____
4. What was the hypothesis of the study? Yes on page___, paragraph____ _, lines_____ No_____
5. What were the participant's major characteristics? Yes on page___, paragraph____ _, lines_____ No_____
6. Does the review of literature indicate previous research in the area associated with the article? Yes on page___, paragraph____ _, lines_____ No_____
7. What type of study is reported in this article? Yes on page___, paragraph____ _, lines_____ No_____
8. Was the sample randomly selected? Yes on page___, paragraph____ _, lines_____ No_____
9. What was the instrument? Yes on page___, paragraph____ _, lines_____ No_____
10. What were the major steps involved in the treatment? Yes on page___, paragraph____ _, lines_____ No_____
11. How were the variables tested? Yes on page___, paragraph____ _, lines_____ No_____
12. According to the author(s) how successful was the treatment? Yes on page___, paragraph____ _, lines_____ No_____
13. What factors could equally account for the student tests results? Yes on page___, paragraph____ _, lines_____ No_____
14. What problems, if any, do you detect in the study? Yes on page___, paragraph____ _, lines_____ No_____
15. Do the results of analysis agree with the author's objectives and expectations? Yes on page___, paragraph____ _, lines_____ No_____
16. What other interpretations could be made from the data? Yes on page___, paragraph____ _, lines_____ No_____
17. Is there archaeological, craniometric and or linguistic evidence that supports the research findings yes on page___, paragraph____ _, lines_____ No_____