Chapter 12

Sequencing, Assembling, and Annotating: A Genomic Approach to Text Genealogy*

Wendy J. Phillips-Rodríguez

Introduction

Originally, the discipline of stemmatics was created as a method to edit texts that exist in many versions. As such, and despite its limitations, it is still probably the preferred practice among textual scholars due to its cogency.[1] The information revealed by a *stemma codicum* is considered useful mainly at the beginning of the process of editing, for it is precisely such a stemma that will make it possible for the editor(s) to choose readings in a genealogically informed manner. Nevertheless, it seems that the use of new technologies developed in other fields (i.e., evolutionary biology) besides helping textual scholars to achieve some of their traditional goals (editing), has also brought about the possibility of using stemmatology for purposes it was not initially meant for.

The grouping of manuscripts according to evolutionary characteristics (Howe *et al.*), detecting certain patterns of contamination (Windram *et al.* 2006) and even spotting some scribal practices such as "change of exemplar" (Phillips-Rodriguez *et al.* 2009), which can now be visualized in ways that were not possible in the past, may reveal important information about the text itself and the environments in which it was created and transmitted. Such

* It is a serendipitous circumstance that I came to conclude this work in Vienna, a place that has surely meant much for Dominik. This is not the first time that I have this impression, as several times in the past, in terms of themes or views (digital humanities, building communities of work, the study and preservation of manuscripts), I have felt that he has been there before and that his work paved many roads. To him, with much appreciation. The writing of this work was made possible by the generous support of a scholarship (PASPA 2022) from DGAPA (Dirección General de Asuntos del Personal Académico) of the National Autonomous University of Mexico (UNAM).

1 In this paper I take a similar approach to stemmatics as Jürgen Hanneder: "If we speak of stemmatics we mean the method that was described by [Paul] Maas including its later modifications" (Hanneder 2009–2010: 7). An analogous perspective is offered by Paolo Chiesa: "The genealogical method is also called 'Lachmann's method,' from the name of Karl Lachmann (1793–1851), a German classical scholar who was considered to be its creator or architect. In fact, this method was constructed over a rather long timespan (from the last decades of the eighteenth century to the early twentieth century) thanks to the contributions of many scholars, sometimes working in connection with one another, sometimes working autonomously" (2020: 74).

information will remain of first value to editors; however, it also calls for attention as an asset in itself, for it is a window to the way in which textual artifacts evolve, or, as William Robins has put it, "how texts perform their cultural work."[2]

This paper is, in a sense, a return, a reflection on themes that have been pressing methodological issues for me when approaching the ancient literature from India. It is against this background that I will not sum up results that I came to conclude from previous works, but I will revisit some of the core questions that I have been pondering ever since I got interested in Sanskrit textual studies.

1. What is the standing of the critical edition of the *Mahābhārata* against the current state of research in stemmatology?

If we rephrase the question in the following fashion: "Has the text of the critical edition of the *Mahābhārata*, reconstituted by traditional means, become obsolete now that we can use sophisticated tools to study the relationships between manuscripts?" the quick answer is "Not at all." Indeed, computer-aided stemmatics allow us some deeper insights into the text's history than traditional methods do, mainly due to its capacity to put information into graphics. However, it must be clear that the traditional information and the methods are not made redundant by the new tools. Most facts that can be observed now at a glance, by means of diagrams, were perceived too by the editors through their continuous work with the manuscripts, for, as V. S. Sukthankar clearly affirmed: "The study of the manuscripts themselves must first teach us what their interrelationship is."[3] However, much of the information that made itself available to the editors through their sustained interaction with the manuscripts and that informed their decisions is often not made explicit to the user of the edition, simply because the traditional layout of a printed critical edition does not allow for it. Only the prolegomena and the introductions are places to discuss such information, and then again, such discussion must be carried out succinctly, as too long an explanation or too many examples would defeat the very purpose of the edition. What new tools make possible is to grasp some of that insider knowledge in a more direct manner. In other words, the information revealed by means of computerized tools is not different in nature to that revealed by traditional methods, but often it can be finer-grained, more easily transmittable, and amenable to further analysis.

Because of that, at least in the foreseeable future, I do not see any chance for the critical edition of the *Mahābhārata* to be rendered obsolete by the new methods. Its lifespan does not get shortened by the use of computational tools. If at all, it gets lengthened, as the material it contains can now be mined (perhaps even supplemented) to respond to unexplored lines of enquiry. That opens room to novel approaches to text genealogy beyond the purpose of editing,

2 Robins 2006: 114.

3 Sukthankar 1933: LXXXII.

as several aspects of the transmission, migration, and adaptation of the text to different times and places can be potentially scrutinized.

2. What can we learn from this case study, and how can we use it to portray a bigger picture of textual evolution?

Thus, far from an appeal to re-editing using the new tools, my proposal would be on the perspective that the biological metaphor could lend to the work of the traditional stemmatologists and the results that can be achieved through that. At least, in view of the data provided by the Dyūtaparvan, I suggest that such a perspective could help us incorporate the previous efforts into a coherent view in which computerized tools are in no way against traditional stemmatology. On the contrary, they stand in a continuous line as efforts upon which further scholarship can be built.

Both, biological evolution and textual criticism – together with quite a few other disciplines that require for their study what Robert O'Hara (1996) calls "trees of history" – deal with the phenomenon of "descent with modification." Thus, there is a chance that they can cross-pollinate with each other while working on their own challenges. For instance, it is worth noticing that the basic principle for cladistic analysis, the principle of *apomorphy*, which says that only shared innovations provide evidence of common ancestry, was first understood in stemmatics and historical linguistics.[4] O'Hara even suggested that "if there had been more cross-fertilization among these fields the cladistic revolution in systematics might well have taken place in the nineteenth century."[5] However, it was only after the advent of molecular sequencing that evolutionary biologists started thinking on ways to organize their huge amounts of historical data.

Certainly, some of the tools evolutionary biologists are coming up with may be useful to textual scholars, but it is perhaps their way of dealing with the model which can be even more productive. Often, advancement in a discipline is propelled or hindered by the model it has as its core.[6] In other words, the narrative of what scholars are looking for shapes the kind of efforts that they invest. For a long time, textual scholarship has been driven by the quest of reconstructing a text that could best explain the state of all available witnesses. At the same time, it has been expected that such text, by being a possible ancestor, would have had a

4 Willi Hennig, the founder of phylogenetics, developed the term "apomorphic" through his early theoretical publications from 1947 to 1966 (Richter & Meier 1994). However, quite some decades earlier, in the area of textual criticism, several editors had already taken the approach of considering that only derived readings ("common errors") were useful to map shared descent between witnesses (Trovato 2020: 110). For instance, according to Reeves, in 1872 with his edition of *Vie de Saint Alexis*, Gaston Paris was "the first scholar to have applied systematically the principle that only shared errors establish families of textual witnesses" (2011: 68).

5 O'Hara 1996: 81.

6 Nersessian 2022.

concrete existence in the past. Consequently, the realization that a genealogical reconstruction, though useful in evolutionary terms, cannot be taken as a factual reality seems to deter some scholars from using it in all its capacities and even make them question the usefulness of the whole critical undertaking. Wendy Doniger, for instance, has succinctly expressed this lack of historical dimension by considering the text of the critical edition of the *Mahābhārata* "nobody's version,"[7] as opposed to the text witnessed by the individual manuscripts, which belongs to a particular time and place. However, can a critically reconstructed text be asked for a concrete existence? More importantly, would textual scholarship benefit from such demand? Perhaps evolutionary biology, by means of the human genome, can bear an example of the possibilities offered by an abstract model whose factual existence is not called for.

Thus, in the next sections, I will deal with what I believe has been achieved in *Mahābhārata* textual studies by means of the critical edition from a perspective akin to the biological sciences and will try to suggest in which direction further paths may be developed.

3. Another way of framing the question as a whole, or what the critical edition of the *Mahābhārata* means in bioinformatic terms:

3.1. Sequencing: What collators did

In bioinformatics, the process of discovering the order in which nucleotides (the four building blocks of DNA, expressed by the letters A, G, C and T) are combined is called sequencing. However, the full length of a genome – even of smaller organisms, as "(t)here is a remarkable lack of correspondence between genome size and organism complexity" – is extremely long.[8] For this reason, the sequencing must be done bit by bit, using genomic discrete regions which are called "reads." For the study of evolutionary biology, sequencing is extremely important because it determines the primary structure of what will become the representative genome for a certain species: the string of characters against which all other genomes of the same or similar species will be compared.

Sequencing *de novo*, as it is called when no reference sequence is available, is an extremely time-consuming task (even when computers are at work). Therefore, the sequencing of the human genome is considered "one of the most ambitious and important scientific endeavors in human history."[9] A very large, well-organized, and highly collaborative international effort was made to sequence different parts of the genome, then assembling them together to form one single string of text and finally making it available to the whole scientific community. The project, which lasted from 1990 to 2003,

7 Ghosh 2022, New Books Network podcast: https://newbooksnetwork.com/after-the-war

8 Pray 2008: 96.

9 *Human Genome Project Fact Sheet*.

> involved researchers from 20 separate universities and research centers across the United States, United Kingdom, France, Germany, Japan, and China. The groups in these countries became known as the International Human Genome Sequencing Consortium.[10]

This sounds not completely unlike the Bhandarkar Oriental Research Institute's project that was taken over by different scholars (and their teams) who gave themselves to the task of editing the different volumes of the *Mahābhārata*. In this, however, the project lasted forty-eight years to reach its end. This is not a long time if one considers that sequencing the *Mahābhārata* was done by hand, not using the powerful algorithms developed by biologists to help them compare and organize their material.

> The collation is done by a permanent staff of specially trained Shastris (Northern as well as Southern) and University graduates. For the purposes of collation, each Mahābhārata stanza (according to the Bombay edition of Ganpat Krishnaji, Śaka 1799) is first written out, in bold characters, on the top line of a standard, horizontally and vertically ruled foolscap sheet. The variant readings are entered by the collator horizontally along a line allotted to the manuscripts collated, akṣara by akṣara, in the appropriate column, vertically bellow the corresponding portion of the original reading of the "Vulgate." On the right of each of these collation sheets, there is a column four inches wide reserved for remarks (regarding corrections, marginal additions etc.), and for "additional" stanzas found in the manuscripts collated, either immediately before or after the stanza in question. Very long "additions" are written out on separate "śodhapatras" and attached to the collation sheets. The collations are regularly checked by a batch of collators different from the one which did the collation in the first instance, before they are handed over to the editor for the constitution of the text.[11]

Unfortunately, none of those famous collation sheets has been located to bear testimony of how it looks to sequence by means of pencil and paper:[12] a feat that any modern evolutionary biologist would consider titanic. In fact, titanic it was, for as we are told by S. K. Belvalkar, Sukthankar's successor as the editor-in-chief, after the first *maṅgalācaraṇa śloka* was inscribed in the year 1919,

> for the first 7 or 8 years thereafter, the entire staff of Graduates, Under-graduates and our multi-lingual Shastris did nothing but collation work day after day. And

10 *Human Genome Project Fact Sheet*.

11 Sukthankar 1933: IV–V.

12 In 2003, I visited the BORI to look at the material kept there from the making of the *Mahābhārata* Critical Edition, however no collation sheet was located at that time.

> even that did not suffice: the Institute had to open at its own expense special collation centers, one at Santiniketan and the other at Tanjore, where manuscripts. written in the scripts of Bangal and South India could be more quickly assembled and collated.[13]

No wonder the task of comparing manuscripts reached high proportions as the formidable length of the text was witnessed by multiple copies in each of its sections:

> [t]he *entire* Mahābhārata stands now collated from a *minimum* of ten manuscripts; many parvans have been completely collated from twenty manuscripts; some from thirty; a few from as many as forty; while the first two adhyāyas of the Ādi, which have special importance for the critical constitution of the text of the entire epic, were collated from no less than sixty manuscripts.[14]


WANTED

20 Volunteers

to participate in the

Human Genome Project

a very large international scientific research effort.

The goal is to decode the human hereditary information (*human blueprint*) that determines all individual traits inherited from parents. The outcome of the project will have tremendous impact on future progress of medical science and lead to improved diagnosis and treatment of hereditary diseases.

Volunteers will receive information about the project from the Clinical Genetics Service at Roswell Park, and sign a consent form before participating.

No personal information will be maintained or transferred.

Volunteers will provide a one-time donation of a small blood specimen. A small monetary reimbursement will be provided to the participants for their time and effort.

Individuals must be at least 18 years of age.
Persons who have undergone chemotherapy are not eligible.

ROSWELL PARK CANCER INSTITUTE

For more information please contact the
Clinical Genetics Service
845-5720 (9:00 am - 3:00 pm)
March 24 - 26, 1997

Figure 12.1. A 1997 Buffalo, New York newspaper advertisement recruiting volunteers to provide blood samples and DNA for the Human Genome Project (HGP Fact Sheet).

The human genome project also gathered information of many individuals, though at the end only the data of a few was used:

13 Belvalkar 1954: 308.
14 Sukthankar 1933: IV.

> The project researchers used a thoughtful process to recruit volunteers, acquire their informed consent, and collect their blood samples. Most of the human genome sequence generated by the Human Genome Project came from blood donors in Buffalo, New York; specifically, 93% from 11 donors, and 70% from one donor (HGP fact sheet).[15]

3.2. Assembling: What general editors did

If sequencing requires cutting text into small pieces, assembling means to put all those pieces together again in a meaningful manner. In bioinformatics, sequence assembly refers to aligning and merging fragments from a longer DNA sequence to reconstruct the original sequence. In simple terms, one sequences "reads" (relatively short strings of nucleotides) which then need to be put together (assembled) by merging them in such a way to obtain the full/almost full concatenation of characters that describes an organism

The sequence of a genome is useful only in as much as it is coherently assembled, just as we could

> [i]magine cutting up many copies of a long poem into strips only a few words long, mixing them up and trying to put the poem back together based on the strips' overlapping ends. Final assemblies can leave regions out, put in too many copies of a repeating sequence, assemble the pieces in the wrong order or put them in backward.[16]

Scientist may agree or disagree with the accuracy of any chosen "read," but its importance comes from it being placed in the *right* order of the chain or, as a matter of fact, in any fixed order that would make it accessible time and again. The process of assembling (piecing together the whole string of characters) is, by far, the most time-consuming task of the whole procedure and it is considered "one of the central problems of bioinformatics."[17]

In editing, this means setting up a version that is believed by the editor to be the best available choice according to his or her knowledge of the manuscript material. Not everyone may agree with each choice made by the editor(s), but what makes the assembling of a representative text so crucial is that it allows the sequence to be seen as a whole. This can make us reflect on the frequently voiced criticism about the text of the critical edition not bearing a historical reality. At the inception of the project, some scholars even suggested that there was no need to constitute a text, as that of the Vulgate or of any other historical manuscript would do, with the *variae lectiones* of other manuscripts supplied in the footnotes. However, the editors opted for an

15 *Human Genome Project Fact Sheet*.
16 Yandell 2012.
17 Henson *et al.* 2012: 901.

> eclectic but cautious utilization of all manuscript classes. Since all categories of manuscripts have their strong points and week points, each variant must be judged on its own merits.[18]

Thus, as plainly stated by Sukthankar, the constituted text offered by the critical edition is an inferred version that "cannot be accurately dated, nor labelled as pertaining to any particular place or personality."[19]

In terms of the human genome, it is interesting to notice that

> [t]he sequence of the human genome generated by the Human Genome Project was not from a single person. Rather, it reflects a patchwork from multiple people whose identities were intentionally made anonymous to protect their privacy.[20]

The term "patchwork" is the key concept here. The usefulness of the human genome does not come from belonging to any human in the flesh, but from representing a parameter against which other variants (of the same species) can be compared. Once the representative sequence is made, however "biased" or "idiosyncratic" that may be, it is then possible (and necessary) to diversify the material by means of resequencing. Indeed, the efforts of the Human Genome Diversity Project or others such as the Human Pangenome Project are directed to capture such human genetic diversity by comparing the DNA of specific populations to understand how different it is from the reference. Something important to keep in mind is that resequencing can only take place when a representative genome already exists, as only such parts that are different from the pre-existing reference need to be aligned and sequenced.

> Resequencing for comparison with the reference genome generally doesn't involve any assembly, because this has already been done for the reference genome. Instead alignment is used. This means that the sections of DNA or "reads" produced after sequencing are compared to the reference genomes and placed alongside their most similar (ideally identical) counterpart. Once all the sections are aligned, it is then possible to look for differences between the individual sequence and the reference sequence.[21]

Thus, when challenging or questioning readings from the constituted text, we are, so to say, resequencing. A task to be done at much less expense, as we are focusing only on comparatively short parts of the text. Departing from the readings offered by the constituted text of the critical edition was something expected by the General Editors, who never claimed infallibility:

18 Sukthankar 1933: LXXXVI.
19 Sukthankar 1933: LXXXVI.
20 *Human Genome Project Fact Sheet*.
21 http://tinyurl.com/yjtebec2 (last accessed June 16, 2023).

"For, who and what is to prevent him [the Critical Reader] from constructing his own text from this critical edition?"[22] They understood that once a constituted text was produced, the task of "resequencing" could be carried out whenever necessary with relative ease. That is the whole point of the critical apparatus.

> Since all divergent readings of any importance will be given in the critical notes, printed at the foot of the page, this edition will, for the first time, render it possible for the reader to have before him the entire significant manuscript evidence for each individual passage. The value of this method for scientific investigation of the epic is obvious.[23]

On the one hand, the text above the line is the *Mahābhārata* genome, so to say, whose value – just as in the case of the human genome, being of a composite matter – does not come from being historical, but representative. On the other hand, the material below the line, which constitutes a veritable "thesaurus of the Mahābhārata tradition," provides for some diversity.[24] Furthermore, besides the material already available in the critical apparatus, it is not discardable the possibility that other manuscript material could be aligned to the critical text, should that become available.

3.3. Annotating: What is there to do

> The genome sequence of an organism is an information resource unlike any that biologists have previously had access to. But the value of the genome is only as good as its annotation. It is the annotation that bridges the gap from the sequence to the biology of the organism.[25]

In bioinformatics, annotating a genome means several specific things, such as giving clues about the structural or functional role of portions of the text in order to render it more useful for future research. Since now there is no need to sequence from scratch, it is possible to focus on specific regions. As expected, sustained attention on short amounts of the string often lead to deeper understanding, and thus scientists may feel the need to leave "a note" attached to that section for future readers.

In the human genome some of those annotations may appear along the string, as one of the characteristics of digital data is to be easily supplemented with metainformation. However, when we think of a printed critical edition, it is evident that such annotations will have to be supplied otherwise. As a matter of fact, the constituted text of the critical edition started to

22 Sukthankar 1933: LXXXIV.
23 Sukthankar 1933: IV.
24 Sukthankar 1933: IV.
25 Stein 2001: 493.

be annotated as soon as it was made available, whenever scholars felt the need to focus on particular readings or episodes and decided to write about them. Only those annotations do not coexist in the same universe as the printed edition and therefore they are not able to be read along the constituted text. If we want to know about them, it is necessary to consult them in other locations (papers, articles, books, etc.).

Certainly, it would be more efficient to have that information linked to the text, directly to the individual passages where scholars have found something worth pointing out. That, however, could only be accomplished by bringing on board the digital humanities. It is at this point that another mighty effort in terms of *Mahābhārata* scholarship comes to my mind: the electronic format of the constituted text, transcribed by Muneo Tokunaga and then revised and maintained by John Smith.[26] The existence of this electronic version has the potential to propel the usefulness of the printed critical text to new heights as it makes it possible to link it to new tools for its mining and enable alternative forms of display in which hypertext could be of much use. Certainly, much collective work will have to be done to build the interfaces. However, with so useful resources within our reach, we find ourselves in a truly privileged position. All those resources were constructed collaboratively through decades of work, and they also require a joint effort from present-day and future scholars to put them to good use. In that regard too, *Mahābhārata* textual scholars are in a similar standing as evolutionary biologists of today: as more of them come to rely on annotation, it will become more important for the scholarly community as a whole to contribute to this continuing process.[27]

4. Conclusions

If we stick to the biological analogy, the constituted text of the critical edition of the *Mahābhārata* means for *Mahābhārata* studies what the sequencing of the human genome does for human genetics: it is a monumental amount of groundwork that enables the continuity of the project. Just as the first sequencing of the human genome in 2003 marked the starting point of genomic studies, the availability of the critical edition constitutes the footing for many future efforts. Perhaps rather than thinking of it as an edition, it would help to consider it as a bank of data in the making, the first step to what could become a genomic archive of the written *Mahābhārata*. The good electronic version of the text as constituted in the critical edition and the information on variants in the footnotes of the printed one are two further readily available resources from which much profit could be made. That is, if we allow ourselves to partake of some of the objectives that currently drive the research in genomics: understanding of variation, patterns of circulation, and interactions of the biological organism (in our case, cultural organism) with the environment.

26 https://bombay.indology.info/mahabharata/statement.html (Last accessed June 16, 2023).
27 Cf. Stein 2001: 501–502.

This speaks about the long-term usefulness to keep critically editing other Sanskrit texts, in spite of the difficulties and the costs (human and material) of the task. Critical editions are things to build upon and they can take research a long way, particularly if they are seen as open-ended projects, where further material, meta information and new views can be incorporated.

As mentioned earlier, the only change this paper proposes is in the outlook of the material we already have available. By doing so, it seeks to build a narrative in which the efforts of the past can be incorporated in a continuum where the textual history of the *Mahābhārata* may be not only documented but also pored over to understand aspects of it that go beyond the interest of solely editing.

Possibly because they were dealing with a similar kind of material, the efforts of the makers of the critical edition of the *Mahābhārata* bear resemblance to the efforts done in the biological sciences. Guided by knowledge and intuitiveness, they did the first sequencing and assembling and therefore provided future scholars with, if we may, a *Mahābhārata* genome. It is now for future research to work on resequencing and annotating, opening the way to enlarge the bank of data and providing diversity and nuance where it is needed.

References

Belvalkar, S. K. (1954). Statement Presented by the General Editor of the *Mahābhārata*. *Annals of the Bhandarkar Oriental Research Institute* 35(1/4), 305–310. http://www.jstor.org/stable/41784969

Chiesa, P. (2020). Principles and Practice. In Roelli, P. (Ed.), *Handbook of Stemmatology. History, Methodology, Digital Approaches*. Berlin, Boston: De Gruyter, p. 74–87.

Ghosh, U. (Host). (2022, August 2). An interview with Wendy Doniger "After the War: The Last Books of the Mahabharata" [Audio podcast episode]. In *New Books in Indian Religions*. New books network. https://newbooksnetwork.com/after-the-war

Hanneder, J. (2009–2010). Introduction. In Hannder, J. & Maas, P. A. (Guest Eds.), *Text Genealogy, Textual Criticism and Editorial Critique*. = *Wiener Zeitschrift für die Kunde Südasiens* 52–53, 5–16.

Henson, Joe, *et al*. (2012). Next-generation Sequencing and Large Genome Assemblies. *Pharmacogenomics* 13.8, 901–915. https://doi.org/10.2217/pgs.12.72

How do you put a genome back together after sequencing? (2021, July 21). *YourGenome*. Wellcome Sanger Institute UK. http://tinyurl.com/yjtebec2

Howe C. J. *et al.* (2001). Manuscript Evolution. *Endeavour* 25.3, 121–126.

Human Genome Project Fact Sheet, National Human Genome Research Institute, Last updated August 4, 2022. http://tinyurl.com/yjtebec2

Nersessian, N. J. (2022). Interdisciplinarity in the Making: Models and Methods in Frontier Science. Cambridge, Mass. London MIT Press. http://tinyurl.com/bdhfv9cc

O'Hara, R. J. (1996). Trees of History in Systematics and Philology. *Memorie della Società italiana di scienze naturali e del museo civico di storia naturale di Milano*, 27.1, 81–88. Available at SSRN: https://ssrn.com/abstract=2558818

Phillips-Rodriguez, W. J., Howe, C. J., Windram, H. F. (2009). Chi-Squares and the Phenomenon of "Change of Exemplar" in the *Dyūtaparvan*. In Huet, G., Kulkarni, A. & Scharf, P. (Eds.), *Sanskrit Computational Linguistics*. ISCLS ISCLS 2007 2008. Lecture Notes in Computer Science, vol. 5402. Berlin, Heidelberg: Springer, p. 380–390. https://doi.org/10.1007/978-3-642-00155-0_20

Pray, L. (2008). Eukaryotic Genome Complexity. *Nature Education* 1(1). https://www.nature.com/scitable/topicpage/eukaryotic-genome-complexity-437/

Richter, S., & Meier, R. (1994). The Development of Phylogenetic Concepts in Hennig's Early Theoretical Publications (1947–1966). *Systematic Biology* 43.2, 212–221.

Robinson, P. & O'Hara, R. (1996). Cladistic Analysis of an Old Norse Manuscript Tradition. *Research in Humanities Computing* 4, 115–137. https://rjohara.net/cv/1996-rhc

Robins, W. (2006). Editing and Evolution. *Literature Compass* 4, 89–120. https://doi.org/10.1111/j.1741-4113.2006.00391.x

Stein, L. (2001). Genome Annotation. From Sequence to Biology. *Nat Rev Genet* 2, 493–503. https://doi.org/10.1038/35080529

Sukthankar, V. S. (1933). Prolegomena. In Sukthankar, V. S. (Ed.), *The Mahābhārata. For the First Time Critically Edited*. Vol. 1 *Ādiparvan*. Poona: Bhandarkar Oriental Research Institute, p. I–CX.

Trovato, P. (2020). Neo-Lachmannism: A New Synthesis? In Roelli, P. (Ed.), *Handbook of Stemmatology. History, Methodology, Digital Approaches*. Berlin, Boston: De Gruyter, 109–138.

Windram, H. F., Spencer, M. & Howe, C. J. (2006). Phylogenetic Analysis of Manuscript Traditions, and the Problem of Contamination. In Macè, C. *et al*. (Eds.), *The Evolution of Texts: Confronting Stemmatological and Genetical Methods. Proceedings of the International Workshop Held in Louvain-la-Neuve on September 1–2, 2004*. = *Linguistica Computazionale* 24–25, 141–156.

Yandell, K. (2012, May 1). Solving the Genomic Jigsaw Puzzle. *Scienceline. A Project of NYU's Science, Health and Environmental Reporting Program*. http://tinyurl.com/4zkmuxyc